TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Nachrichtentechnik

Low-Precision Quantizer Design for Communication Problems

Georg Christoph Zeitler

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor–Ingenieurs

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr.-Ing. Wolfgang Utschick
Prüfer der Dissertation:	1. Univ.-Prof. Dr. sc. techn. Gerhard Kramer 2. Prof. Andrew C. Singer, Ph.D. University of Illinois at Urbana-Champaign, Urbana, USA 3. Prof. Upamanyu Madhow, Ph.D. University of California, Santa Barbara, USA

Die Dissertation wurde am 22.11.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 06.02.2012 angenommen.

Preface

I thank Professor Gerhard Kramer for his support and guidance during the second phase of my time as an assistant at TUM's Institute for Communications Engineering (LNT). When Gerhard took over the position as head of LNT, he gave me the freedom to continue working in the field I have been working on, and he greatly influenced the contributions of this thesis by giving valuable and extensive advice.

I am deeply grateful to Professor Andrew Singer for his continuous support and for giving me the opportunity to visit the University of Illinois at Urbana-Champaign as a visiting researcher in the spring of 2010. Meetings and discussions with Andy shaped my thinking about research and provided the basis for important parts of this work; these meetings also kept me fit because they occasionally happened cycling along the Isar river or running in the park.

At the beginning of the work leading to this thesis, I had the honor and luck to work with the late Professor Ralf Kötter. Thank you Ralf, for giving me the opportunity to pursue a Dr.-Ing. degree at TUM, and for sparking my interest in quantization for communication problems.

Special thanks go to Professor Upamanyu Madhow for acting as a co-referee of my dissertation. During a project carried out with DOCOMO Euro-Labs GmbH, I had the chance to closely and pleasantly collaborate with Professor Gerhard Bauch, Jörg Widmer, and Guido Dietl. Among the many colleagues and students whose company I enjoyed at work, off work, and on many trips abroad I would like to single out my office-mate Joschi Brauchle for being a good friend and for his patience with various computer-related problems, and Andrew Bean for making my visit at UIUC a very enjoyable one. I would also like to thank Johannes Brehmer for contributing to Chapter 4 of this thesis.

I am most indebted to my family, my mother Elisabeth, my father Georg, my sister Katharina, my brother Maximilian, and my uncle Friedrich. The last word of thanks goes to Viktoria for her love and support.

München, November 2011 Georg Zeitler

Contents

Zusammenfassung

Diese Arbeit untersucht den Entwurf von Quantisierern mit geringer Auflösung für zwei ausgewählte Probleme in der Kommunikationstechnik: zum einen den Relais-Kanal mit Mehrfachzugriff und Komprimierung der empfangenen Information am Relais, und zum anderen den Punkt-zu-Punkt Kanal mit Intersymbolinterferenz, additivem Rauschen und Analog-Digital-Wandlung am Empfänger. Für den Relais-Kanal mit Mehrfachzugriff werden skalare und zweidimensionale Quantisierer für die am Relais vorliegende Information entworfen, wobei der Fokus auf einem Verfahren geringer Komplexität liegt. Außerdem wird die Zuteilung der Kompressionsraten für maximale Summenrate untersucht, wobei das Protokoll am Relais darin besteht, die Empfangssequenzen mittels Vektorquantisierung zu komprimieren und diese dann an die Senke weiterzuleiten. Anschließend wird der Entwurf von Analog-Digital-Wandlern für Interferenz-Kanäle betrachtet, mit dem Ziel, die erreichbare Informationsrate zu maximieren. Für rauschfreie Kanäle wird die minimal nötige Alphabetgröße des Analog-Digital-Wandlers für maximale Informationsrate in Abhängigkeit von der Größe des Sendealphabets bestimmt. Außerdem werden skalare und zweidimensionale Analog-Digital-Wandler für verrauschte Kanäle entworfen. Abschließend wird das vorgestellte Verfahren zum Entwurf von Analog-Digital-Wandlern ergänzt durch die Kanalschätzung mittels eines adaptiv regelbaren Quantisierers mit nur einem Bit Auflösung. Neben unterer Schranken für den mittleren quadratischen Fehler werden Schätz- und Adaptionsverfahren entwickelt, welche die berechneten Schranken annähernd erreichen.

Abstract

In this work, the design of low-precision quantizers for two selected problems in communications is addressed: the multiple-access relay channel with compression of the received signals at the relay, and the point-to-point link with intersymbol-interference, additive noise, and analog-to-digital conversion at the receiver. For the multiple-access relay channel, scalar and two-dimensional quantizers are designed for log-likelihood ratios at the relay yielding a low complexity scheme. The sum-rate optimal allocation of compression rates using a compress-and-forward strategy is also considered. The low-precision analog-to-digital converter design problem for intersymbol-interference channels is studied next, where the focus is on maximizing the information rate over such channels. The smallest possible size of the analog-to-digital converter alphabet yielding maximal information rate is derived for noiseless channels as a function of the transmit alphabet size, and scalar and two-dimensional converters are designed for noisy channels. Finally, the analog-to-digital converter design problem is complemented by studying channel estimation using a single-bit adaptively dithered quantizer. Lower bounds on the mean squared error are derived, and dither and estimation schemes are proposed that are shown to closely approach the lower bounds.

1

Introduction

Quantization is the process of assigning an element from a discrete set to each element from a larger set. Often, the smaller set is finite, while the larger set may be infinite and possibly uncountable. A particularly simple example of quantization is rounding of a real-valued number to the nearest integer. Quantization may be formally represented by a quantization function that is many-to-one: many elements from the larger set may be mapped to the same element of the smaller, discrete set. Since this process cannot be reversed, there is an unrecoverable loss of information associated with quantization.

In modern digital communication systems, quantization plays a pivotal role for two main reasons: first, the signals to be communicated are often analog in nature, and they need to be digitized at the transmitting side in order for the advantages of digital technology to be leveraged [Pro00, Section 3.4], for example easy storage and error-correction coding, to name a few. Second, the received waveform as a continuous-time and analog signal is sampled and quantized by an analog-to-digital converter at the front-end of a digital receiver to allow the application of sophisticated digital signal processing algorithms (e.g., equalization for channels with intersymbol-interference [BLM04, Chapters 8 and 9]) thereby enhancing the quality of detection and decoding. Beyond digital communication links, quantization is also used in a host of other applications such as wireless sensor networks [SMZ07, RG06a, RG06b], remote sensing [Cam02], and biomedical applications [YS06].

For decades, the design of analog-to-digital converters for communication systems has been driven by metrics that permit system designers to proceed with their design and refinement unaware of the applications for which they will eventually be used. As a result, metrics such as spurious free dynamic range or total harmonic distortion tend to dominate the design of such systems [Kes05, Section 2-3]. However, if the precision of quantization

is reduced (to as low as one bit per sample in the most extreme case), such a system-agnostic design of the quantization step can have a severe impact on system performance. The goal of this thesis is to explicitly design quantizers for communication problems. In lieu of employing traditional metrics for that design, we use figures of merit suitable for communications such as mutual information for coded systems, we develop algorithms for the design of such quantizers, and we derive performance bounds taking the quantization step into account. This thesis is organized as follows:

Chapter 2 reviews rate-distortion theory and introduces a tradeoff between rate and relevant information contained in a quantization. Algorithms are given to compute the rate-distortion curve as well as the rate-information tradeoff.

In **Chapter 3**, the multiple-access relay channel with two sources, a single relay, and one destination is considered. For cases when the relay cannot decode without error, we propose a framework for designing one- and two-dimensional quantizers for quantizing log-likelihood ratios (or soft information) at the relay. These quantizers are mutual-information preserving. Simulation results show that a) mutual-information preserving quantization outperforms techniques for which the soft information is forwarded in an analog fashion to the destination, b) two-dimensional quantization outperforms one-dimensional quantization for source–relay links of different quality, and c) a diversity order of two can be gained in block Rayleigh fading channels by having the relay adaptively select a two-dimensional quantizer from a fixed set of quantizers shared with the destination, depending on the channel state on the source–relay links.

We continue to consider the orthogonal multiple-access relay channel in **Chapter 4**, but now with many sources and a compress-and-forward protocol, for which we address the source coding rate allocation problem at the relay. In case of Gaussian codebooks at the sources and Gaussian channels, we show that the sum-rate-optimal assignment of source coding rate at the relay is given by water-filling. For general modulation alphabets at the sources and finite-alphabet discrete memoryless channels, the source coding rate allocation problem is formulated using the tradeoff between rate and relevant information for this system, based on which we appropriately modify a standard cutting-plane algorithm to numerically compute an optimal source coding rate vector at the relay. We also study a variant of the compress-and-forward protocol without binning, known as noisy network coding.

Analog-to-digital converters that maximize the information rate between the quantized channel output sequence and the channel input sequence are designed in **Chapter 5** for discrete-time channels with intersymbol-interference, additive noise, and for independent and identically distributed signaling. It is shown that optimized scalar quantizers with Λ regions achieve the full information rate of $\log_2(\Lambda)$ bits per channel use with a transmit alphabet of size Λ at infinite signal-to-noise ratio; these quantizers, however, are not necessarily uniform quantizers. Low precision scalar and two-dimensional analog-to-digital converters are designed at finite signal-to-noise ratio, and an upper bound on the information rate is derived. Simulation results demonstrate the effectiveness of the optimized quantizers over conventional quantizers. The advantage of the new quantizers is further emphasized by an example of a channel for which a simple slicer and a carefully opti-

mized channel input with memory fail to achieve a rate of one bit per channel use at high signal-to-noise ratio, in contrast to memoryless binary signaling and an optimized quantizer.

In **Chapter 6**, the Bayesian parameter estimation problem using a single-bit dithered quantizer is considered. This problem arises, e.g., for channel estimation under low-precision analog-to-digital conversion at the receiver as considered in Chapter 5. Based on the Bayesian Cramér-Rao lower bound, bounds on the mean squared error are derived that hold for all dither strategies with strictly causal adaptive processing of the quantizer output sequence. In particular, any estimator using the binary quantizer output sequence is asymptotically (in the sequence length) at least 1.96 dB worse than the minimum mean squared error estimator using continuous observations, for any dither strategy. Moreover, dither strategies are designed that are shown by simulation to closely approach the derived lower bounds, and are compared to existing approaches to dithering and estimation.

Finally, **Chapter 7** summarizes the results and discusses open research problems that are related to the work in this thesis.

Throughout, we use standard notation for probabilities, random variables, expectation, entropies, and other mathematical expressions. Appendix F summarizes the mathematical notation and contains a list of abbreviations.

2

Preliminaries

2.1. Rate-distortion theory

Let $Y_1, Y_2, \ldots, Y_n$ be a string of independent and identically distributed (i.i.d.) discrete random variables distributed according to P_Y, so that Y_i takes on values in the finite set $\mathcal{Y}$ of size $|\mathcal{Y}|$. Assume that the string is to be transmitted over a noisy channel, and that the channel supports only a rate smaller than $H(Y)$, the entropy of the source emitting the string $Y_1, Y_2, \ldots, Y_n$. Therefore, by Shannon's source coding theorem [Sha48], we cannot perfectly reconstruct the source sequence at the receiver with high probability. But how well can one do? To answer this type of question, one must first quantify the quality of the reproduction. This is accomplished by defining a distortion function $d : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}_0^+$ between Y and its representation $Z \in \mathcal{Z}$. The maximum tolerable value of the average distortion then specifies the fidelity criterion of the source coding system. Following [Sha59, Ber71, CT06], for a bounded distortion function satisfying $d_{\max} = \max_{y,z} d(y,z) < \infty$, the minimum rate required to be able to reconstruct the source with an average distortion no larger than D is given by the *rate-distortion function*

$$R(D) = \min_{P_{Z|Y}} I(Y;Z), \tag{2.1}$$

where the minimum is over all conditional distributions $P_{Z|Y}$ that satisfy the constraint on the average distortion given by

$$\mathrm{E}\left[d(Y,Z)\right] = \sum_{y,z} P_{Z|Y}(z|y) P_Y(y) d(y,z) \leq D. \tag{2.2}$$

The function $R(D)$ has the following properties (cf. [CT06, Chapter 10] and [Yeu02, Chapter 9]):

1. The function $R(D)$ is a convex ($\cup$) function of D.
2. $R(D)$ is non-increasing in D.
3. $R(D) = 0$ for $D \geq D_{\max}$, where $D_{\max} = \min_{z \in \mathcal{Z}} \mathrm{E}[d(Y, z)]$.
4. $R(0) \leq H(Y)$.
5. Properties 1 and 2 imply that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{\max}$ if $R(0) > 0$. Property 2 implies that $R(D) = 0$ for all $D \geq 0$ if $R(0) = 0$.

Example 2.1. The rate-distortion function for a binary source and reconstruction alphabet ($\mathcal{Y} = \mathcal{Z} = \{0, 1\}$) with $P_Y(0) = \gamma$ and Hamming distortion

$$d(y, z) = \begin{cases} 0 & \text{if } y = z \\ 1 & \text{otherwise} \end{cases} \tag{2.3}$$

is given by [CT06, Chapter 10]

$$R(D) = \begin{cases} H_{\mathrm{b}}(\gamma) - H_{\mathrm{b}}(D) & \text{if } 0 \leq D \leq \min\{\gamma, 1-\gamma\} \\ 0 & \text{if } D > \min\{\gamma, 1-\gamma\}, \end{cases} \tag{2.4}$$

where $H_{\mathrm{b}}(x) = -x \log_2(x) - (1-x) \log_2(1-x)$ denotes the binary entropy function. We plot $R(D)$ for $\gamma = 0.11$ in Figure 2.1. As expected, $R(0) \leq H(Y) = 0.5$, and $R(D) = 0$ for $D \geq D_{\max} = \gamma$.

For most other rate-distortion problems of interest, a closed-form solution for $R(D)$ is not available; however, Arimoto [Ari72] and Blahut [Bla72] independently developed an algorithm to numerically compute the rate-distortion function. The algorithm is widely known as the Blahut-Arimoto algorithm (BAA), and can also be used to numerically calculate the capacity of an arbitrary discrete memoryless channel (DMC). We will briefly review the algorithm for the computation of the rate-distortion function.

To derive the algorithm, we will restrict ourselves to $R(0) > 0$, because otherwise, $R(D) = 0$ for all $D \geq 0$. For $R(0) > 0$, the function $R(D)$ is a convex and strictly decreasing function of D for $0 \leq D \leq D_{\max}$. Let $\lambda_{\max}$ and $\lambda_{\min}$ be the negative slope of the $R(D)$ curve at $D = 0$ and $D = D_{\max}$, respectively. Then for any λ, $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$, the tangent to the $R(D)$ curve at the point $(D_\lambda, R(D_\lambda))$ has slope equal to $-\lambda$, and we denote $R(D_\lambda)$ by R_λ. At $R(D_\lambda) + \lambda D_\lambda$, the tangent intersects with the ordinate. We illustrate the tangent with $\lambda = 2$ to $(D_\lambda, R_\lambda) = (0.2, 0.278)$ in Figure 2.2. For a given $P_{Z|Y} = P'_{Z|Y}$, we denote

$$P'_Z(z) = \sum_y P_Y(y) P'_{Z|Y}(z|y) \tag{2.5}$$

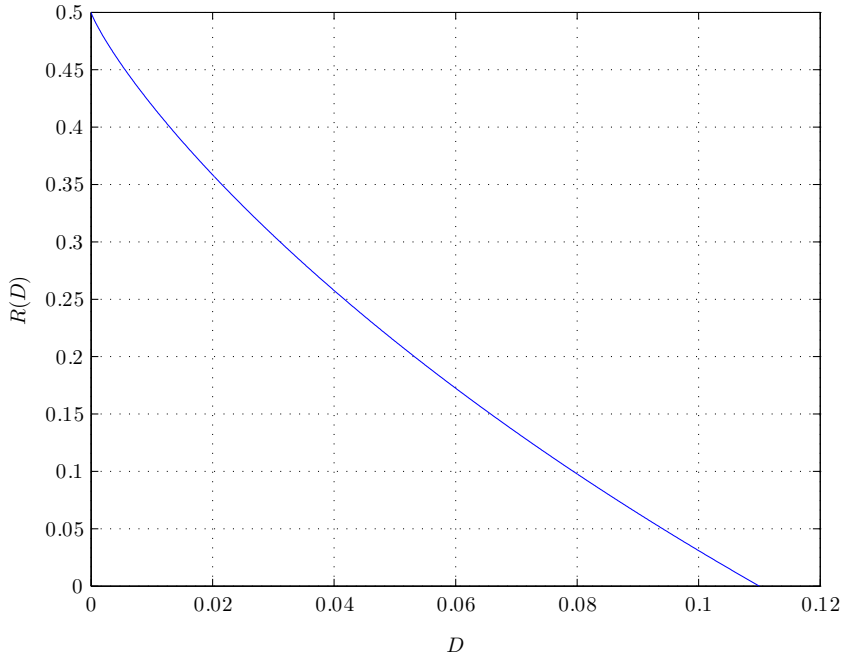


Figure 2.1.: $R(D)$ for a binary source with $\gamma = 0.11$ and Hamming distortion.

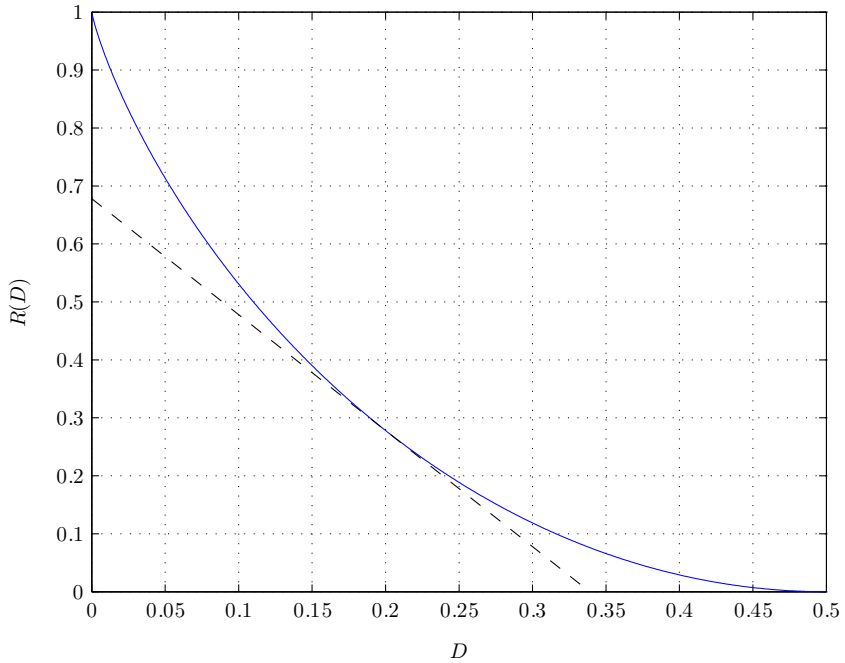


Figure 2.2.: $R(D)$ for a binary source with $\gamma = 0.5$ and Hamming distortion; the tangent to $(D_\lambda, R_\lambda) = (0.2, 0.278)$ and $\lambda = 2$ is also shown.

$$I_{P'_{Z|Y}} = \sum_{y,z} P_Y(y) P'_{Z|Y}(z|y) \log_2 \left(\frac{P'_{Z|Y}(z|y)}{P'_Z(z)} \right) \tag{2.6}$$

$$D_{P'_{Z|Y}} = \sum_{y,z} P'_{Z|Y}(z|y) P_Y(y) d(y,z), \tag{2.7}$$

and the point $(D_{P'_{Z|Y}}, I_{P'_{Z|Y}})$ lies on or above the rate-distortion curve $R(D)$. The line of slope $-\lambda$ which passes through $(D_{P'_{Z|Y}}, I_{P'_{Z|Y}})$ has axis intercept with the ordinate of $I_{P'_{Z|Y}} + \lambda D_{P'_{Z|Y}}$. Therefore, for any λ, $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$, we have

$$R(D_\lambda) + \lambda D_\lambda = \min_{P_{Z|Y}} \left\{ I_{P_{Z|Y}} + \lambda D_{P_{Z|Y}} \right\}. \tag{2.8}$$

Based on [CT06, Lemma 10.8.1], we can expand the definition of the rate-distortion function as a double minimization and insert the definition of mutual information to obtain

$$R(D_\lambda) + \lambda D_\lambda = \min_{P_Z} \min_{P_{Z|Y}} \left\{ \sum_{y,z} P_Y(y) P_{Z|Y}(z|y) \log_2 \left(\frac{P_{Z|Y}(z|y)}{P_Z(z)} \right) + \lambda \mathrm{E}\left[d(Y,Z) \right] \right\}. \tag{2.9}$$

The BAA now applies the method of alternating minimization [CT84] to (2.9). Starting with an initial distribution $P^{(0)}_{Z|Y}$, we find the output distribution

$$P^{(0)}_Z(z) = \sum_y P_Y(y) P^{(0)}_{Z|Y}(z|y), \tag{2.10}$$

which, according to [CT06, Lemma 10.8.1], is the output distribution minimizing the mutual information. For that $P^{(0)}_Z$ and a particular $\lambda \geq 0$, the conditional distribution minimizing the mutual information subject to the distortion constraint is

$$P^{(1)}_{Z|Y}(z|y) = \frac{P^{(0)}_Z(z) 2^{-\lambda d(y,z)}}{\sum_{z'} P^{(0)}_Z(z') 2^{-\lambda d(y,z')}}, \tag{2.11}$$

with which another iteration can be performed. Ciszár [Csi74] proved that the iterative procedure converges, so that

$$\left(D_{P^{(k)}_{Z|Y}}, I_{P^{(k)}_{Z|Y}} \right) \to (D_\lambda, R_\lambda) \tag{2.12}$$

as $k \to \infty$, and the entire rate-distortion curve can be covered by varying $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$. We summarize the algorithm in Algorithm 2.1.

Algorithm 2.1 The Blahut-Arimoto algorithm.

Input: $P_Y(y)$, $\mathcal{Y}$, $\mathcal{Z}$, $d(y,z)$, $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$, $\epsilon > 0$
Initialization: randomly choose $P_{Z|Y}^{(0)}(z|y)$, $k \leftarrow 1$
$P_Z^{(0)}(z) \leftarrow \sum_y P_Y(y) P_{Z|Y}^{(0)}(z|y)$
$P_{Z|Y}^{(1)}(z|y) \leftarrow \dfrac{P_Z^{(0)}(z) 2^{-\lambda d(y,z)}}{\sum_{z'} P_Z^{(0)}(z') 2^{-\lambda d(y,z')}}$
while $\sum_{y,z} \left| P_{Z|Y}^{(k)}(z|y) - P_{Z|Y}^{(k-1)}(z|y) \right| / (|\mathcal{Y}| \cdot |\mathcal{Z}|) \geq \epsilon$ **do**
    $P_Z^{(k)}(z) \leftarrow \sum_y P_Y(y) P_{Z|Y}^{(k)}(z|y)$
    $P_{Z|Y}^{(k+1)}(z|y) \leftarrow \dfrac{P_Z^{(k)}(z) 2^{-\lambda d(y,z)}}{\sum_{z'} P_Z^{(k)}(z') 2^{-\lambda d(y,z')}}$
    $k \leftarrow k + 1$
end while
$P_{Z|Y}(z|y) \leftarrow P_{Z|Y}^{(k)}(z|y)$, $P_Z(z) \leftarrow \sum_y P_Y(y) P_{Z|Y}(z|y)$
$D_\lambda \leftarrow \sum_{y,z} P_Y(y) P_{Z|Y}(z|y) d(y,z)$
$R_\lambda \leftarrow \sum_{y,z} P_Y(y) P_{Z|Y}(z|y) \log_2 \left(\dfrac{P_{Z|Y}(z|y)}{P_Z(z)} \right)$

2.2. The tradeoff between mutual information and rate

2.2.1. Problem formulation

One key requirement for determining a rate-distortion function is to choose a distortion function *a priori* as a measure of goodness of the reproduction Z; for many applications, the choice for d is made in favor of analytic tractability rather than perceptual meaningfulness for the problem at hand [Gra90, Chapter 2.4]. By definition of the distortion function, we also note that the fidelity criterion is based on a measure of closeness between Y and Z. But what if one was not interested in "closely" representing Y in the reproduction Z, but was instead aiming at extracting relevant features from Y about a third variable, say X? This conceptually different question was asked by Tishby *et al.* in [TPB99], where they formalized the problem and provided an algorithm in the spirit of the BAA to solve it. A related problem was studied much earlier in [WW75], where the authors are concerned with the properties of the problem and its application to selected topics in information theory. The approach of [WW75] is as follows.

Let (X, Y) be a pair of random variables with joint probability mass function P_{XY}, so

that X and Y take on values in the finite sets $\mathcal{X}$ and $\mathcal{Y}$ of size $|\mathcal{Y}| = n$ and $|\mathcal{X}| = m$, respectively. The variable Y is to be mapped into a random variable $Z \in \mathcal{Z}$ such that the conditional entropy $H(X|Z)$ is minimal while the conditional entropy $H(Y|Z)$ is no smaller than s. Furthermore, we require that X and Z are conditionally independent given Y, i.e., $X \leftrightarrow Y \leftrightarrow Z$ forms a Markov chain. Then, for $0 \leq s \leq H(Y)$, the function $F(s)$ is defined as [WW75]

$$F(s) = \min_{P_{Z|Y}} H(X|Z) \quad \text{s.t.} \quad H(Y|Z) \geq s, \tag{2.13}$$

where the minimum is over all distributions $P_{Z|Y}$ such that $H(Y|Z) \geq s$, and the aforementioned Markov condition is satisfied since the optimization is over $P_{Z|Y}$. Here, the definition of $F(s)$ is in terms of conditional entropies; we now introduce a function related to $F(s)$ which is in terms of mutual information expressions.

The information-rate function

We define the *information-rate function* $I(R)$ for $0 \leq R \leq H(Y)$ as

$$I(R) \triangleq \max_{P_{Z|Y}} I(X;Z) \quad \text{s.t.} \quad I(Y;Z) \leq R, \tag{2.14}$$

where the maximum is over all conditional distributions $P_{Z|Y}$ such that the mutual information $I(Y;Z)$ is no larger than R. Since $I(X;Z) = H(X) - H(X|Z)$ and $I(Y;Z) = H(Y) - H(Y|Z)$, and since $H(Y)$ and $H(X)$ are fixed, $I(R)$ can be readily recovered from $F(s)$ by writing

$$I(R) = H(X) - F(H(Y) - R). \tag{2.15}$$

Although (2.14) bears obvious resemblance to (2.1), there is no *operational* meaning of R as a rate here, in contrast to $R(D)$, which is an operational rate-distortion function. Nevertheless, we still refer to $I(R)$ as an information-rate function.

The function $I(R)$ has a number of interesting properties, which can be readily obtained from properties of $F(s)$ derived in [WW75].

1. Based on [WW75, Theorem 2.3], we can conclude that $I(R)$ is a concave ($\cap$) function of R, for $0 \leq R \leq H(Y)$.

2. The maximum in (2.14) is attainable with Z taking at most $n+1$ values, i.e., $|\mathcal{Z}|$ need not be larger than $n+1$. This also follows from [WW75, Theorem 2.3].

3. The function $I(R)$ is monotonically non-decreasing in R [WW75, Theorem 2.5]. By the same theorem, we have that the constraint in (2.14) is binding, i.e., it can be replaced by $I(Y;Z) = R$.

4. Following [WW75, Theorem 2.6], $I(R) \leq R$.

5. From [WW75, Theorem 4.1], we have $I(R = 0) = 0$ and $I(R = H(Y)) = I(X;Y)$.

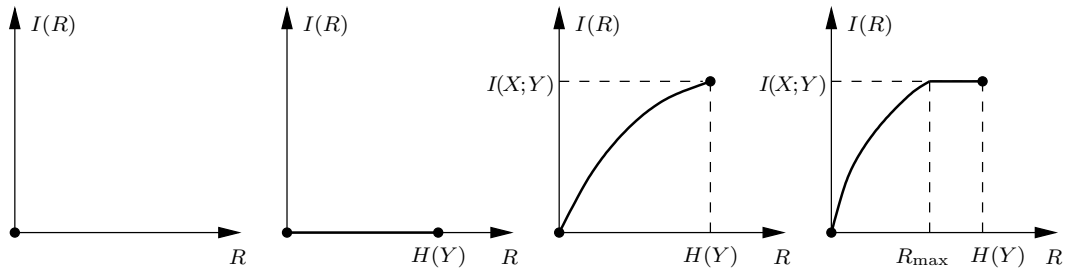


Figure 2.3.: Sketches of $I(R)$ for a) $H(Y) = I(X;Y) = 0$, b) $H(Y) > 0$, $I(X;Y) = 0$, c) $H(Y) > 0$, $I(X;Y) > 0$, $R_{\max} = H(Y)$, and d) $H(Y) > 0$, $I(X;Y) > 0$, $R_{\max} < H(Y)$.

6. Combining Properties 3 and 5 yields that $I(R) = 0$ for all $0 \leq R \leq H(Y)$ if $I(X;Y) = 0$. Moreover, with the definition of

$$R_{\max} \triangleq \min R \quad \text{s.t.} \quad I(R) = I(X;Y), \tag{2.16}$$

the function $I(R)$ is strictly increasing for $0 \leq R \leq R_{\max}$ if $I(X;Y) > 0$.

We sketch the shapes of $I(R)$ for various cases in Figure 2.3.

The rate-information function

It is also useful to introduce a dual formulation of the information-rate tradeoff, namely a *rate-information function*. We have the following proposition.

Proposition 2.1. Define for $0 \leq I \leq I(X;Y)$ the rate-information function as

$$R(I) \triangleq \min_{P_{Z|Y}} I(Y;Z) \quad \text{s.t.} \quad I(X;Z) \geq I, \tag{2.17}$$

where the minimum is over all conditional distributions $P_{Z|Y}$ such that the mutual information between X and Z is at least I. Then $R(I)$ is the inverse function of $I(R)$ restricted to the interval $0 \leq R \leq R_{\max}$, i.e., if $I^* = I(R^*)$, then $R(I^*) = R^*$ for all $0 \leq R^* \leq R_{\max}$.

The proof of Proposition 2.1 is given in Appendix A. Since $R(I)$ is the inverse function of $I(R)$ for $0 \leq R \leq R_{\max}$, we can deduct its properties in a straightforward manner from the properties of $I(R)$:

1. The function $R(I)$ is a convex ($\cup$) function of I, for $0 \leq I \leq I(X;Y)$.

2. The minimum in (2.17) is attainable with Z taking at most $n + 1$ values.

3. The function $R(I)$ is monotonically non-decreasing in I, and the inequality constraint is binding, so that it can be replaced by $I(X;Z) = I$.

4. We have $R(I) \geq I$ for $0 \leq I \leq I(X;Y)$.

5. The function $R(I)$ satisfies $R(I=0) = 0$ and $R(I = I(X;Y)) = R_{\max}$.

6. If $R_{\max} > 0$ and $I(X;Y) > 0$, the function $R(I)$ is strictly increasing; otherwise, $R(I) = 0$.

2.2.2. Computation of the information-rate function

The information-rate function $I(R)$ can be computed in two different ways, the first one being based on [WW75, Theorem 4.1], and the second using the iterative information bottleneck algorithm from [TPB99].

We begin with the method of [WW75]. Denote by $\boldsymbol{q}$ the probability mass function of Y, i.e., $\boldsymbol{q} = [P_Y(1), P_Y(2), \ldots, P_Y(n)]^{\mathrm{T}}$, and let $\boldsymbol{T}$ be the fixed $m \times n$ matrix with entries $[\boldsymbol{T}]_{i,j} = P_{X|Y}(i|j)$. We denote by Δ_n the probability simplex of dimension $(n-1)$, and for $\boldsymbol{p} \in \Delta_n$, the function $h_n(\boldsymbol{p})$ is the entropy function, i.e., $h_n(\boldsymbol{p}) = -\sum_{j=1}^{n} p_j \log_2(p_j)$. Following [WW75, Theorem 4.1], we define $\phi(\boldsymbol{p}, \lambda) \triangleq h_m(\boldsymbol{T}\boldsymbol{p}) - \lambda h_n(\boldsymbol{p})$, and let $\psi(\cdot, \lambda)$ be the lower convex envelope on Δ_n of $\phi(\cdot, \lambda)$. Then,

$$I(R) = h_m(\boldsymbol{T}\boldsymbol{q}) - \max_{0 \leq \lambda \leq 1} \left\{ \psi(\boldsymbol{q}, \lambda) + \lambda(h_n(\boldsymbol{q}) - R) \right\}. \tag{2.18}$$

Hence, to compute $I(R)$ for general $\boldsymbol{T}$ and $\boldsymbol{q}$, one has to find the lower convex envelope on Δ_n of $\phi(\cdot, \lambda)$, which seems hard to find in closed form for general $\boldsymbol{T}$ and $\boldsymbol{q}$. The function is available in closed form for some special cases [WW75, Section IV] only.

Example 2.2 ([WW75, Section IV-A])**.** Let X, Y be such that X is the result of applying Y to a binary symmetric channel (BSC) with error probability $0 < \epsilon < 0.5$, and let $P_Y(0) = q$. Consequently,

$$\boldsymbol{T} = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix} \tag{2.19}$$

and

$$\boldsymbol{q} = \begin{bmatrix} q \\ 1-q \end{bmatrix}. \tag{2.20}$$

Then, we have

$$H(Y) = H_{\mathrm{b}}(q) \tag{2.21}$$

$$H(X) = H_{\mathrm{b}}(q + \epsilon - 2q\epsilon), \tag{2.22}$$

and with (2.15) and [WW75], the information-rate function is given by

$$I(R) = H_{\mathrm{b}}(q + \epsilon - 2q\epsilon) - H_{\mathrm{b}}\left(\epsilon + (1-2\epsilon) H_{\mathrm{b}}^{-1}\left(H_{\mathrm{b}}(q) - R\right)\right) \tag{2.23}$$

for $0 \leq R \leq H_{\mathrm{b}}(q)$.

Given the difficulties involving the computation of $I(R)$ in closed form for general cases, we next describe an algorithm to directly solve the optimization problem defining the information-rate function. The algorithm was proposed in [TPB99] and is known as the *information bottleneck iterative algorithm*. In order to be able to relate the information bottleneck iterative algorithm to the previously introduced BAA, we work with the rate-information function in the following. To establish the algorithm, it is useful to expand the constraint on $I(X;Z)$ in the definition of $R(I)$, exploiting the Markov condition $X \leftrightarrow Y \leftrightarrow Z$, yielding

$$I(X;Z) = I(X;Y) - I(X;Y|Z). \tag{2.24}$$

Hence, we obtain

$$R(I) = \min_{P_{Z|Y}} I(Y;Z) \quad \text{s.t.} \quad I(X;Y|Z) \leq I(X;Y) - I, \tag{2.25}$$

where $I(X;Y)$ does not depend on $P_{Z|Y}$. Next, again exploiting the Markov condition, we write

$$\begin{aligned}
I(X;Y|Z) &= \sum_{x,y,z} P_{XYZ}(x,y,z) \log_2 \left(\frac{P_{XY|Z}(x,y|z)}{P_{X|Z}(x|z) P_{Y|Z}(y|z)} \right) && (2.26)\\
&= \sum_{y,z} P_{YZ}(y,z) \sum_{x} P_{X|YZ}(x|y,z) \log_2 \left(\frac{P_{Y|Z}(y|z) P_{X|YZ}(x|y,z)}{P_{X|Z}(x|z) P_{Y|Z}(y|z)} \right) && (2.27)\\
&= \sum_{y,z} P_{YZ}(y,z) \sum_{x} P_{X|Y}(x|y) \log_2 \left(\frac{P_{X|Y}(x|y)}{P_{X|Z}(x|z)} \right) && (2.28)\\
&= \sum_{y,z} P_{YZ}(y,z) D_{\mathrm{KL}} \left(P_{X|Y}(\cdot|y) || P_{X|Z}(\cdot|z) \right) && (2.29)\\
&= \mathrm{E} \left[D_{\mathrm{KL}} \left(P_{X|Y}(\cdot|Y) || P_{X|Z}(\cdot|Z) \right) \right], && (2.30)
\end{aligned}$$

where

$$D_{\mathrm{KL}} \left(P_{X|Y}(\cdot|y) || P_{X|Z}(\cdot|z) \right) = \sum_{x} P_{X|Y}(x|y) \log_2 \left(\frac{P_{X|Y}(x|y)}{P_{X|Z}(x|z)} \right) \tag{2.31}$$

denotes the *relative entropy* or *Kullback–Leibler distance* between the distributions $P_{X|Y}(\cdot|y)$ and $P_{X|Z}(\cdot|z)$. Inserting (2.30) into (2.25) yields

$$R(I) = \min_{P_{Z|Y}} I(Y;Z) \quad \text{s.t.} \quad \mathrm{E} \left[D_{\mathrm{KL}} \left(P_{X|Y}(\cdot|Y) || P_{X|Z}(\cdot|Z) \right) \right] \leq I(X;Y) - I, \tag{2.32}$$

and by defining the function

$$\bar{d}(y,z) \triangleq D_{\mathrm{KL}} \left(P_{X|Y}(\cdot|y) || P_{X|Z}(\cdot|z) \right), \tag{2.33}$$

we observe that

$$R(I) = \min_{P_{Z|Y}} I(Y;Z) \quad \text{s.t.} \quad \mathrm{E} \left[\bar{d}(Y,Z) \right] \leq I(X;Y) - I, \tag{2.34}$$

is in the form of (2.1), with a minimization of $I(Y;Z)$ and an upper bound for the average value of $\bar{d}$. The key differences between (2.34) and (2.1), however, are the following:

- ▷ In Problem (2.34), the function $\bar{d}(y,z)$ is dependent on the choice of $P_{Z|Y}$, since $P_{X|Z}$ is a function of $P_{Z|Y}$, while in Problem (2.1), the distortion function $d(y,z)$ is independent of the choice of $P_{Z|Y}$.

- ▷ Problem (2.1) is a convex minimization problem since $I(Y;Z)$ is convex in $P_{Z|Y}$ [CT06, Chapter 2], and since $\mathrm{E}\left[d(Y,Z)\right]$ is linear in $P_{Z|Y}$, and therefore convex. While in (2.17), both the objective function $I(Y;Z)$ and $I(X;Z)$ are convex in $P_{Z|Y}$, the non-convexity of the constraint $I(X;Z) \geq I$ renders (2.17) a non-convex problem.

Keeping the above in mind, we nevertheless view $\bar{d}(y,z)$ as the right "distortion" function emerging from placing a constraint on $I(X;Z)$ in (2.17).

The motivation for the information bottleneck iterative algorithm now is very similar to the one for the BAA. Since for $R_{\max} > 0$ and $I(X;Y) > 0$, $R(I)$ is a convex and strictly increasing function of I, the tangent of slope $\beta > 0$ through $R_\beta = R(I_\beta)$ has intercept with the ordinate of $R(I_\beta) - \beta I_\beta$, and

$$R(I_\beta) - \beta I_\beta = \min_{P_{Z|Y}} \Big\{ I(Y;Z) - \beta I(X;Z) \Big\} \tag{2.35}$$

$$= \min_{P_{Z|Y}} \Big\{ I(Y;Z) - \beta \left[I(X;Y) - I(X;Y|Z) \right] \Big\} \tag{2.36}$$

$$= \min_{P_{Z|Y}} \Big\{ I(Y;Z) + \beta \mathrm{E}\left[\bar{d}(Y,Z) \right] \Big\} - \beta I(X;Y). \tag{2.37}$$

The information bottleneck iterative algorithm [TPB99] proceeds in a very similar manner to Section 2.1. We begin with an initial mapping $P_{Z|Y}^{(0)}(z|y)$, for which we obtain

$$P_Z^{(0)} = \sum_y P_Y(y) P_{Z|Y}^{(0)}(z|y). \tag{2.38}$$

Assuming $P_Z^{(0)}(z) \neq 0$ for all z, we are now in a position to calculate the "distortion" function, by computing

$$P_{X|Z}^{(0)}(x|z) = \frac{1}{P_Z^{(0)}(z)} \sum_y P_{XY}(x,y) P_{Z|Y}^{(0)}(z|y) \tag{2.39}$$

$$\bar{d}^{(0)}(y,z) = D_{\mathrm{KL}}\left(P_{X|Y}(\cdot|y) || P_{X|Z}^{(0)}(\cdot|z) \right). \tag{2.40}$$

Given $\bar{d}^{(0)}(y,z)$ and $\beta > 0$, we get the next mapping

$$P_{Z|Y}^{(1)}(z|y) = \frac{P_Z^{(0)}(z) 2^{-\beta \bar{d}^{(0)}(y,z)}}{\sum_{z'} P_Z^{(0)}(z') 2^{-\beta \bar{d}^{(0)}(y,z')}}, \tag{2.41}$$

serving as a starting point for the next iteration. The entire information bottleneck iterative algorithm is summarized in Algorithm 2.2. Due to the non-convexity of the problem, only local convergence can be guaranteed [TPB99]. Therefore, the algorithm concludes with estimates $(\hat{I}_\beta, \hat{R}_\beta)$ of (I_β, R_β), and one can execute the algorithm repeatedly with a different initial mapping $P_{Z|Y}^{(0)}$ until a satisfactory solution is obtained.

Algorithm 2.2 The information bottleneck iterative algorithm [TPB99].

Input: $P_{XY}(x,y)$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, $\beta > 0$, $\epsilon > 0$
Initialization: randomly choose a valid mapping $P_{Z|Y}^{(0)}(z|y)$, $k \leftarrow 1$
$P_Z^{(0)}(z) \leftarrow \sum_y P_Y(y) P_{Z|Y}^{(0)}(z|y)$
$P_{X|Z}^{(0)}(x|z) \leftarrow \left(1/P_Z^{(0)}(z)\right) \sum_y P_{XY}(x,y) P_{Z|Y}^{(0)}(z|y)$
$\bar{d}^{(0)}(y,z) \leftarrow D_{\mathrm{KL}}\left(P_{X|Y}(\cdot|y) || P_{X|Z}^{(0)}(\cdot|z)\right)$.
$P_{Z|Y}^{(1)}(z|y) \leftarrow \dfrac{P_Z^{(0)}(z) 2^{-\beta \bar{d}^{(0)}(y,z)}}{\sum_{z'} P_Z^{(0)}(z') 2^{-\beta \bar{d}^{(0)}(y,z')}}$
while $\sum_{y,z} \left| P_{Z|Y}^{(k)}(z|y) - P_{Z|Y}^{(k-1)}(z|y) \right| / (|\mathcal{Y}| \cdot |\mathcal{Z}|) \geq \epsilon$ **do**
  $P_Z^{(k)}(z) \leftarrow \sum_y P_Y(y) P_{Z|Y}^{(k)}(z|y)$
  $P_{X|Z}^{(k)}(x|z) \leftarrow \left(1/P_Z^{(k)}(z)\right) \sum_y P_{XY}(x,y) P_{Z|Y}^{(k)}(z|y)$
  $\bar{d}^{(k)}(y,z) \leftarrow D_{\mathrm{KL}}\left(P_{X|Y}(\cdot|y) || P_{X|Z}^{(k)}(\cdot|z)\right)$
  $P_{Z|Y}^{(k+1)}(z|y) \leftarrow \dfrac{P_Z^{(k)}(z) 2^{-\beta \bar{d}^{(k)}(y,z)}}{\sum_{z'} P_Z^{(k)}(z') 2^{-\beta \bar{d}^{(k)}(y,z')}}$
  $k \leftarrow k + 1$
end while
$P_{Z|Y}(z|y) \leftarrow P_{Z|Y}^{(k)}(z|y)$
$P_Z(z) \leftarrow \sum_y P_Y(y) P_{Z|Y}(z|y)$
$P_{X|Z}(x|z) \leftarrow (1/P_Z(z)) \sum_y P_{XY}(x,y) P_{Z|Y}(z|y)$
$\hat{I}_\beta \leftarrow \sum_{x,z} P_Z(z) P_{X|Z}(x|z) \log_2 \left(\dfrac{P_{X|Z}(x|z)}{P_X(x)} \right)$
$\hat{R}_\beta \leftarrow \sum_{y,z} P_Y(y) P_{Z|Y}(z|y) \log_2 \left(\dfrac{P_{Z|Y}(z|y)}{P_Z(z)} \right)$

Example 2.3. We apply Algorithm 2.2 to the scenario of Example 2.2 with binary Y and X resulting from applying Y to a BSC with error probability ϵ; we also choose $|\mathcal{Z}| = 3$,

$q = 0.5$, and $\epsilon = 0.11$. In Figure 2.4, we plot the resulting information-rate curves obtained analytically from Example 2.2, and the result of the numerical optimization using Algorithm 2.2. Both curves coincide remarkably well, and we can also observe $I(R = 1) = I(X;Y) = 0.5$, which is the capacity of a BSC with $\epsilon = 0.11$.

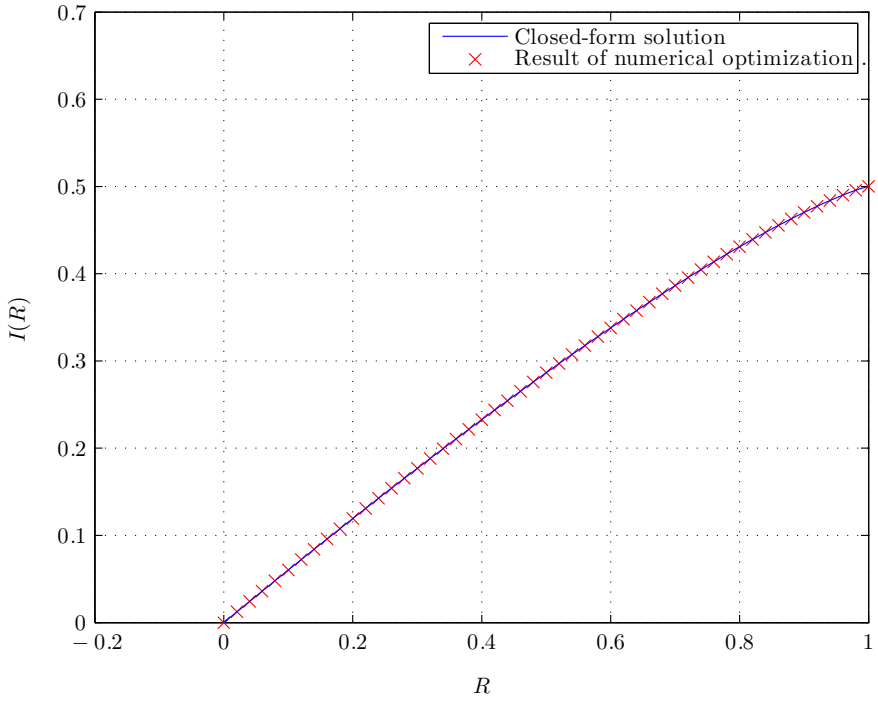


Figure 2.4.: The function $I(R)$ for Example 2.2 and 2.3.

Example 2.4. As another illustration for an information-rate curve, let $\mathcal{X} = \{-1, 0, 1\}$, and let $P_X(x) = 1/3$ for all $x \in \mathcal{X}$. The random variable Y' is given by

$$Y' = X + N, \tag{2.42}$$

where $N \sim \mathcal{N}(0, \sigma^2)$ is independent of X. We choose $\sigma^2 = 0.1$. The discrete random variable Y is obtained from Y' by regular sampling from its distribution such that $|\mathcal{Y}| = 14$. We measure $I(X;Y) = 1.309$ and $H(Y) = 3.406$. For that pair (X, Y) and $|\mathcal{Z}| = |\mathcal{Y}| + 1 = 15$, the information-rate function obtained with the information bottleneck iterative algorithm is shown in Figure 2.5, as well as the information-rate tradeoff computed for $|\mathcal{Z}| = 3$ and $|\mathcal{Z}| = 2$ in order to illustrate that $I(R)$ cannot be swept out completely if $|\mathcal{Z}|$ is small.

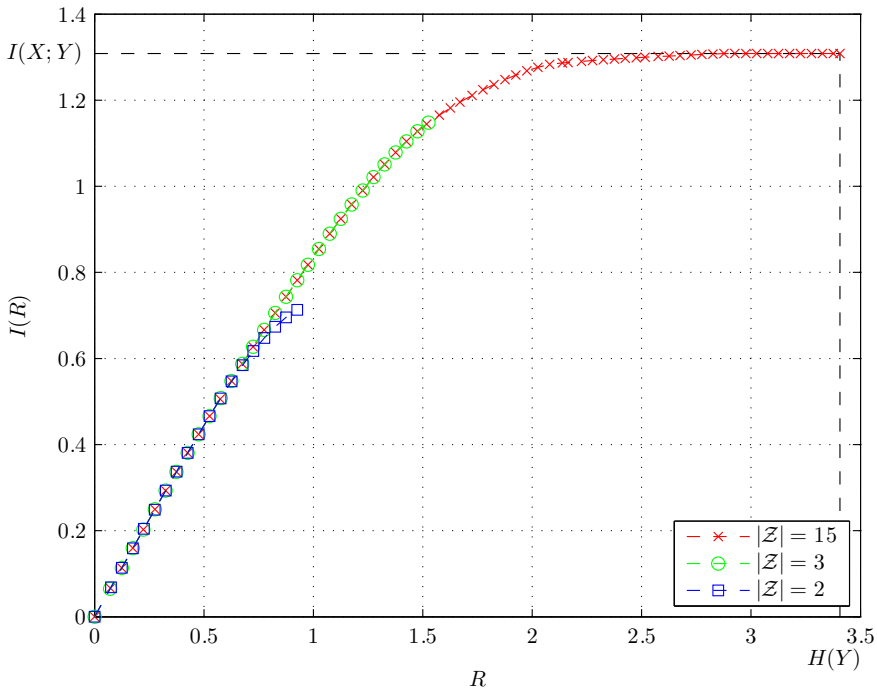


Figure 2.5.: The function $I(R)$ for Example 2.4.

3

Quantize-and-forward schemes for the multiple-access relay channel

Diversity techniques have been widely studied as an effective means to combat multipath fading effects inherent in wireless communication channels. Multiple transmit and/or receive antennas can often provide a form of spatial diversity whenever the application of simpler time or frequency diversity techniques is precluded due to delay or bandwidth constraints. However, due to size limitations on the mobile devices of, e.g., a cellular communication network, the placement of multiple antennas at such mobile terminals is not always a feasible option. Cooperative diversity, first proposed in [SEA03a, SEA03b], introduces spatial diversity without interfering with the size limitations of the terminals by allowing nodes to cooperate in facilitating their transmissions. In some cases, cooperation is achieved by employing a relay node whose sole purpose is to facilitate the transmissions of other nodes. Besides the gain in reliability, relays are also envisioned to provide coverage extension for cell-edge users of cellular networks [YHXM09] at reasonable cost, since relayed transmission alleviates path-loss effects, thereby providing prolonged life time for battery-powered nodes.

Due to technological difficulties [CJS+10, JCK+11] related to nodes receiving and transmitting simultaneously in the same frequency band, orthogonal relayed transmission, where the channels of the sources and the relay are orthogonal either in frequency or in time, received a lot of attention recently [LTW04, MY04, Hau09]. However, the extra resources allocated to the relay result in a loss of spectral efficiency, a loss that can be reduced by allowing several users to share one relay for joint processing. We therefore focus on the orthogonal multiple-access relay channel (MARC) [KvW00] in this chapter, where two sources transmit independent information to a common destination via a single relay.

In related work, diversity achieving schemes are proposed for distributed antenna systems [CKL06], and for the MARC using low-density parity-check codes [HSOB05] and distributed turbo codes [Hau09], combined with network coding [ACLY00] by performing joint network channel coding at the relay. Since the relay performs some form of (joint) re-encoding of the source messages in such a decode-and-forward scheme [CEG79,LTW04], it is common to that work that the relay node is required to decode the source messages perfectly. However, even if the relay fails at fully recovering the source messages, the information available at the relay can still be beneficial for decoding at the destination if forwarded properly. For example, amplify-and-forward [LTW04] or soft-decode-and-forward schemes [SV05] may be used; these methods have the additional advantage that they avoid error propagation that occurs if residual bit errors remain at the relay after a hard decision about the information sequence.

In more recent work [YK07], the authors combine soft decoding and analog forwarding of beliefs from the relay with network coding in the MARC, in that they form and transmit the beliefs about the network coded code bits of both users at the relay, thereby achieving notable gains in symmetric additive white Gaussian noise (AWGN) channels. However, the block of beliefs about the network coded code bits is sent to the destination in an analog manner; furthermore, forming the beliefs about the network coded code bits turns out to be disadvantageous especially in situations where the source–relay channels are of different quality. In this chapter, we aim to compensate these disadvantages:

- ▷ Building on the system in [YK07], we propose a framework for designing scalar quantizers for the soft information at the relay. The framework is related to the information bottleneck method in that we formulate the optimization problem for designing such a quantizer as a tradeoff between quantization rate and obtained mutual information (cf. Section 2.2), taking into account that the available rate on the relay–destination link is (potentially severely) limited.

- ▷ Noting that forming the beliefs about the network coded bits is particularly disadvantageous if the soft information of the two users at the relay has different reliability, the proposed framework is extended to the design of two-dimensional quantizers operating directly on the soft information of both users, without going through the intermediate step of computing the likelihoods of the network coded message. In doing so, the available rate on the relay–destination link is appropriately divided among the two users, according to the quality of the soft information that is at hand for each user at the relay. Moreover, by employing quantization at the relay, digital transmission with its well-known advantages can be leveraged on the relay–destination link.

This chapter is structured as follows. The system model is introduced in Section 3.1, based on which we design quantizers that maximize mutual information in Section 3.2. Numerical results for various channel models are shown in Section 3.3, before we end with concluding remarks in Section 3.4.

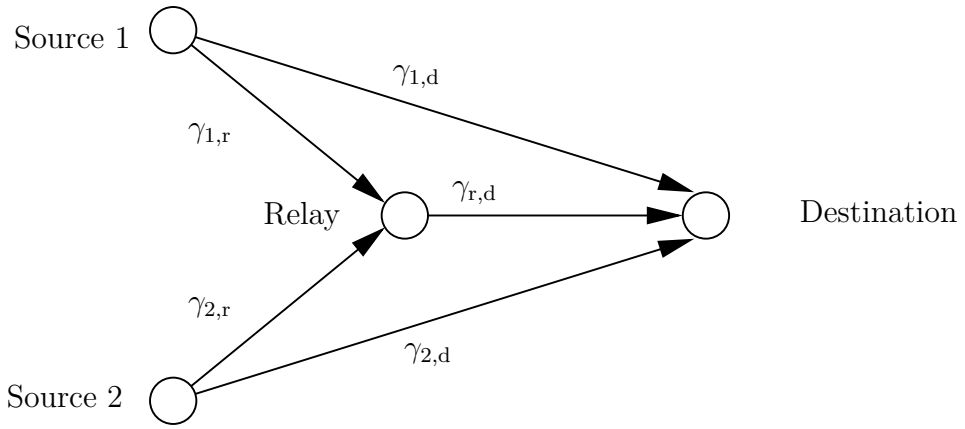


Figure 3.1.: The multiple-access relay channel. The relay only knows $\gamma_{1,\mathrm{r}}$ and $\gamma_{2,\mathrm{r}}$.

3.1. System model

3.1.1. Sources

At each source $i \in \{1, 2\}$, a block of independent information bits $\boldsymbol{U}_i \in \{0, 1\}^{k_i}$ is encoded with a channel code of rate k_i/n_i to a block of code bits $\boldsymbol{X}_i \in \{0, 1\}^{n_i}$, which is then modulated to the channel symbols $\boldsymbol{S}_i \in \mathbb{M}_i^{m_i}$, where $\mathbb{M}_i$ is the modulation alphabet of size M_i at the i-th source. In the rest of the chapter, we assume that the number of code bits satisfies $n = n_1 = n_2$.

3.1.2. Channel model

The channel model is shown in Figure 3.1. In our model, the transmissions from the sources and the relay are assumed to be orthogonalized either in frequency or in time. Despite the suboptimality of this constraint, the restriction to orthogonal channels eases practical implementation. Note that the restriction to orthogonal channels also includes a half-duplex constraint often imposed on the relay for implementation reasons, so that the relay cannot transmit and receive simultaneously in the same frequency band. Without loss of generality, we assume the orthogonality to be guaranteed by time division; consequently, a first slot is assigned to source 1, a second slot to source 2, and a third slot to the relay. Let $\boldsymbol{S}_{\mathrm{r}} \in \mathbb{M}_{\mathrm{r}}^{m_{\mathrm{r}}}$ be the transmitted vector from the modulation alphabet $\mathbb{M}_{\mathrm{r}}$ at the relay. For a path-loss coefficient α and distances $d_{i,\mathrm{r}}, d_{i,\mathrm{d}}$, and $d_{\mathrm{r,d}}$ between the terminals, the received signals at the relay and at the destination are given by

$$\boldsymbol{Y}_{\mathrm{r},i} = \frac{H_{i,\mathrm{r}}}{\sqrt{d_{i,\mathrm{r}}^{\alpha}}} \boldsymbol{S}_i + \boldsymbol{N}_{\mathrm{r},i}, \quad i \in \{1, 2\} \tag{3.1}$$

$$\boldsymbol{Y}_{\mathrm{d},i} = \frac{H_{i,\mathrm{d}}}{\sqrt{d_{i,\mathrm{d}}^{\alpha}}} \boldsymbol{S}_i + \boldsymbol{N}_{\mathrm{d},i}, \quad i \in \{1, 2\} \tag{3.2}$$

in time slot one and two, and by

$$\boldsymbol{Y}_{\mathrm{d,r}} = \frac{H_{\mathrm{r,d}}}{\sqrt{d_{\mathrm{r,d}}^{\alpha}}}\boldsymbol{S}_{\mathrm{r}} + \boldsymbol{N}_{\mathrm{d,r}} \tag{3.3}$$

in the last slot. Here, $H_{i,\mathrm{r}}$, $H_{i,\mathrm{d}}$, and $H_{\mathrm{r,d}}$ are the complex channel fading coefficients satisfying $\mathrm{E}[|H_{i,\mathrm{r}}|^2] = \mathrm{E}[|H_{i,\mathrm{d}}|^2] = \mathrm{E}[|H_{\mathrm{r,d}}|^2] = 1$, and the additive noise variables are independent circularly symmetric complex Gaussian random variables with zero mean and variance normalized to unity. The average values of the signal-to-noise ratio (SNR) are given as $\rho_{i,\mathrm{r}} = P_i/d_{i,\mathrm{r}}^{\alpha}$, $\rho_{i,\mathrm{d}} = P_i/d_{i,\mathrm{d}}^{\alpha}$, and $\rho_{\mathrm{r,d}} = P_{\mathrm{r}}/d_{\mathrm{r,d}}^{\alpha}$, where P_i and P_{r} are the powers of the sources and the relay. Throughout, we make the common assumption that the receivers know the instantaneous SNR values $\gamma_{i,\mathrm{r}} = |h_{i,\mathrm{r}}|^2\rho_{i,\mathrm{r}}$, $\gamma_{i,\mathrm{d}} = |h_{i,\mathrm{d}}|^2\rho_{i,\mathrm{d}}$, and $\gamma_{\mathrm{r,d}} = |h_{\mathrm{r,d}}|^2\rho_{\mathrm{r,d}}$ of their channels, and that the transmitters only possess knowledge about the average SNR. In particular, the relay is assumed to lack instantaneous channel state information (CSI) of the source–destination channels due to their fading nature and limited signaling from the destination to the relay.

3.1.3. Relay operations

The operations at the relay considered in our work are restricted to methods generating and transforming soft information about the coded bits of each user for transmission to the destination.

Generation of soft information

Upon reception of $\boldsymbol{y}_{\mathrm{r},1}$ and $\boldsymbol{y}_{\mathrm{r},2}$, the relay's first option is to invoke soft demappers to compute log-likelihood ratios (LLRs) $\boldsymbol{\ell}_i = \boldsymbol{\ell}_i^{(\mathrm{dem})} \in \mathbb{R}^n$, $i = 1, 2$, about the coded bits, where, for $m = 1, 2, \ldots, n$,

$$\ell_{i,m}^{(\mathrm{dem})} = \ln\left(\frac{P_{X_{i,m}|Y_{\mathrm{r},i,j}}\left(x_{i,m} = 0|y_{\mathrm{r},i,j}\right)}{P_{X_{i,m}|Y_{\mathrm{r},i,j}}\left(x_{i,m} = 1|y_{\mathrm{r},i,j}\right)}\right), \quad j = \lceil m/\log_2(M_i)\rceil. \tag{3.4}$$

Alternatively, the relay performs soft decoding to calculate $\boldsymbol{\ell}_i = \boldsymbol{\ell}_i^{(\mathrm{dec})} \in \mathbb{R}^n$, $i = 1, 2$, where

$$\ell_{i,m}^{(\mathrm{dec})} = \ln\left(\frac{P_{X_{i,m}|\boldsymbol{Y}_{\mathrm{r},i}}\left(x_{i,m} = 0|\boldsymbol{y}_{\mathrm{r},i}\right)}{P_{X_{i,m}|\boldsymbol{Y}_{\mathrm{r},i}}\left(x_{i,m} = 1|\boldsymbol{y}_{\mathrm{r},i}\right)}\right), \quad m = 1, 2, \ldots, n. \tag{3.5}$$

Processing of soft information

The first strategy for processing the soft information $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)$ is the one of [YK07], where the relay computes soft information about the network coded code bits based on $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)$. Specifically, the relay first interleaves $\boldsymbol{\ell}_2$ to avoid short cycles in the factor graph associated with the iterative decoder introduced in Section 3.1.4, yielding the block $\boldsymbol{\ell}_2' \in \mathbb{R}^n$ carrying

soft information about $\boldsymbol{x}_2'$, which is the interleaved version of $\boldsymbol{x}_2$. Then, the relay forms the soft information $\boldsymbol{\ell} \in \mathbb{R}^n$ about $\boldsymbol{x} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2'$, where [HOP96]

$$\begin{aligned} \ell_m &= \ell_{1,m} \boxplus \ell_{2,m}' && (3.6)\\ &\triangleq \ln\left(\frac{1+e^{\ell_{1,m}+\ell_{2,m}'}}{e^{\ell_{1,m}}+e^{\ell_{2,m}'}}\right) && (3.7)\\ &\approx \operatorname{sign}(\ell_{1,m})\operatorname{sign}(\ell_{2,m}')\min\left\{|\ell_{1,m}|,|\ell_{2,m}'|\right\}. && (3.8) \end{aligned}$$

The second strategy combines the computation of soft information about the network coded code bits from [YK07] with scalar deterministic quantization of $\boldsymbol{\ell}$. More formally, the relay employs a quantizer with quantization rule $q(\ell)$, $q : \mathbb{R} \to \mathcal{Z}$, yielding the compressed version $\boldsymbol{z} \in \mathcal{Z}^n$, where $\mathcal{Z} = \{0, 1, \ldots, N-1\}$ refers to the quantizer index set. Since the quantizer is invariant for the entire block, the m-th component z_m of $\boldsymbol{z}$ is given by $z_m = q(\ell_m)$, $m = 1, 2, \ldots, n$. Transforming the quantization rule into a probability mass function

$$P_{Z|L}(z|\ell) = \begin{cases} 1 & \text{if } q(\ell) = z \\ 0 & \text{otherwise,} \end{cases} \tag{3.9}$$

we have the mass function $P_{Z|X}(z|x)$ associated with the quantization given as

$$P_{Z|X}(z|x) = \int_{-\infty}^{\infty} P_{Z|L}(z|\ell) p_{L|X}(\ell|x) \mathrm{d}\ell, \tag{3.10}$$

where $p_{L|X}$ is the density of the soft information L given X, which is assumed to be known or obtained from measurement (cf. Section 3.2).

By investigating (3.8), we note that the reliability of the soft information $\boldsymbol{\ell}$ is dominated by $\min\left\{|\ell_{1,m}|, |\ell_{2,m}'|\right\}$, so that $|\ell_m|$ is limited by the weaker user at the relay. Such a scenario occurs, e.g., if the source–relay links have different SNR. Therefore, in the third proposed strategy for processing soft information at the relay, the exclusive or (XOR) computation is omitted. Instead, the relay performs two-dimensional quantization of $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2'$, which is described by the quantization rule $q(\ell_1, \ell_2)$, $q : \mathbb{R}^2 \to \mathcal{Z}$, where again, $\mathcal{Z}$ is the index set of the quantizer. As before, the quantizer is assumed invariant for the entire block, so that the m-th element of $\boldsymbol{z} \in \mathcal{Z}^n$ is given by $z_m = q(\ell_{1,m}, \ell_{2,m}')$, $m = 1, 2, \ldots, n$. Defining

$$P_{Z|L_1L_2}(z|\ell_1, \ell_2) = \begin{cases} 1 & \text{if } q(\ell_1, \ell_2) = z \\ 0 & \text{otherwise,} \end{cases} \tag{3.11}$$

and writing $p_{L_1L_2|X_1X_2}(\ell_1, \ell_2|x_1, x_2)$ for the conditional density of the soft information at the relay, the mass function $P_{Z|X_1X_2}(z|x_1, x_2)$ is obtained as

$$P_{Z|X_1X_2}(z|x_1, x_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_{Z|L_1L_2}(z|\ell_1, \ell_2) p_{L_1L_2|X_1X_2}(\ell_1, \ell_2|x_1, x_2) \mathrm{d}\ell_1 \mathrm{d}\ell_2, \tag{3.12}$$

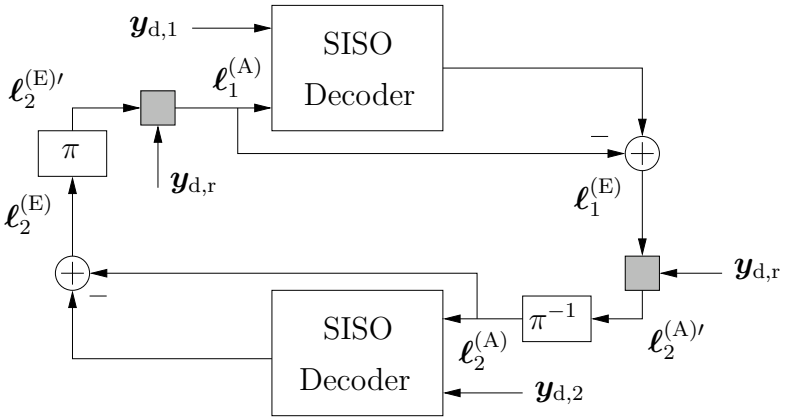


Figure 3.2.: Iterative decoder.

where the density $p_{L_1L_2|X_1X_2}$ is known or approximated by measurement.

Transmission from the relay

Utilizing a quantizer at the relay does not necessarily result in equiprobable quantization indices $\{0, 1, \ldots, N-1\}$ of the quantizer index set $\mathcal{Z}$. Hence, the sequence $\boldsymbol{z}$ needs to be source encoded, yielding the block of bits $\boldsymbol{u}_{\mathrm{r}} \in \{0,1\}^{k_{\mathrm{r}}}$, which is then channel encoded using a channel code of rate $R_{\mathrm{r}} = k_{\mathrm{r}}/n_{\mathrm{r}}$ to the code bits $\boldsymbol{x}_{\mathrm{r}} \in \{0,1\}^{n_{\mathrm{r}}}$ before modulation to the symbols $\boldsymbol{s}_{\mathrm{r}} \in \mathbb{M}_{\mathrm{r}}^{m_{\mathrm{r}}}$.

3.1.4. Destination

The destination uses the iterative receiver [YK07] shown in Figure 3.2. It contains two soft-in/soft-out (SISO) decoders using the received words $\boldsymbol{y}_{\mathrm{d},1}$ and $\boldsymbol{y}_{\mathrm{d},2}$ from the direct links. Furthermore, since the code bits of the two users are coupled by the joint processing at the relay, these two SISO decoders are connected by *relay check nodes* using $\boldsymbol{y}_{\mathrm{d,r}}$ from the relay, drawn as gray squares in Figure 3.2. The relay check nodes allow exchange of soft information between the component SISO decoders, so that the overall decoder resembles a turbo decoder, in contrast to which, however, *code bits* of two *independent* sources are coupled.

The operations of the relay check nodes of course depend on how the soft information at the relay is processed and transmitted. Figure 3.3 shows a summary. In case of scalar quantization at the relay, the destination first needs to recover an estimate $\hat{\boldsymbol{z}}$ of the quantizer output at the relay depending on the success of decoding $\boldsymbol{u}_{\mathrm{r}}$. At this point, we presume the existence of a cyclic redundancy check (CRC) in $\boldsymbol{u}_{\mathrm{r}}$, which we assume to be perfect in the sequel. Then, in order to avoid catastrophic error propagation through the source decoder in case of residual errors in $\hat{\boldsymbol{u}}_{\mathrm{r}}$, the entire transmission from the relay is discarded in that case, so that there is no exchange of soft information between the component decoders. Otherwise, we can assume the source decoder output $\hat{\boldsymbol{z}}$ to correspond to the quantization

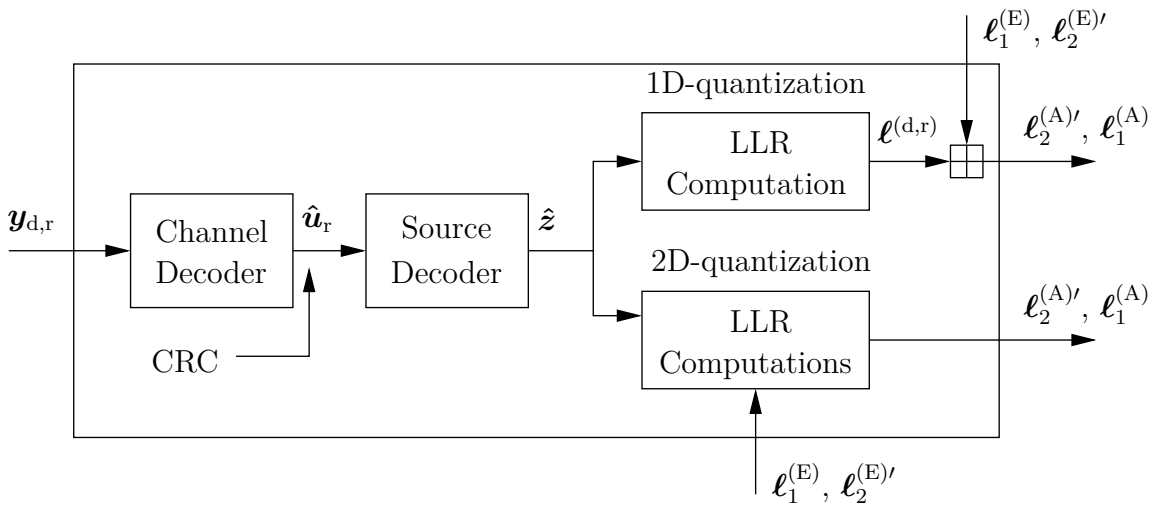


Figure 3.3.: Relay check node for quantized transmission from the relay.

indices obtained at the relay node, i.e., $\hat{\boldsymbol{z}} = \boldsymbol{z}$. For one-dimensional quantization of $\boldsymbol{\ell}$ at the relay, the sequence $\boldsymbol{z}$ specifies the probability distribution $P_{X|Z}(x_m|z_m)$, $m = 1, 2, \ldots, n$, from which we obtain

$$\ell_m^{(\mathrm{d,r})} = \ln\left(\frac{P_{X|Z}(x_m = 0|z_m)}{P_{X|Z}(x_m = 1|z_m)}\right), \quad m = 1, 2, \ldots, n, \tag{3.13}$$

which are the input to a check node (cf. Figure 3.4) with indicator function $\mathbb{1}_{\{x_m = x_{1,m} \oplus x'_{2,m}\}}$, so that we have

$$\ell_{2,m}^{(\mathrm{A})\prime} = \ell_{1,m}^{(\mathrm{E})} \boxplus \ell_m^{(\mathrm{d,r})} \tag{3.14}$$

$$\ell_{1,m}^{(\mathrm{A})} = \ell_{2,m}^{(\mathrm{E})\prime} \boxplus \ell_m^{(\mathrm{d,r})}, \tag{3.15}$$

for $m = 1, 2, \ldots, n$.

For two-dimensional quantization of $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}'_2$ at the relay, the situation is different in that the soft information about $\boldsymbol{x}$ is not formed at the relay. Nevertheless, the quantization at the relay specifies the distribution $P_{Z|X_1X_2}(z_m|x_{1,m}, x'_{2,m})$, $m = 1, 2, \ldots, n$, which is available given perfect reconstruction of $\boldsymbol{z}$ at the destination. Consequently, the coupling of the two component decoders in the factor graph of the iterative decoder at the destination occurs through the function nodes specified by $P_{Z|X_1X_2}(z_m|x_{1,m}, x'_{2,m})$, a section of which is shown in Figure 3.5. For convenience, we describe the operations of that function node in terms of likelihood ratios. To that end, define for $\xi \in \{0, 1\}$

$$\ell(x_{1,m}, x'_{2,m} = \xi|z_m) \triangleq \ln\left(\frac{P_{Z|X_1X_2}(z_m|x_{1,m} = 0, x'_{2,m} = \xi)}{P_{Z|X_1X_2}(z_m|x_{1,m} = 1, x'_{2,m} = \xi)}\right) \tag{3.16}$$

$$\ell(x_{1,m} = \xi, x'_{2,m}|z_m) \triangleq \ln\left(\frac{P_{Z|X_1X_2}(z_m|x_{1,m} = \xi, x'_{2,m} = 0)}{P_{Z|X_1X_2}(z_m|x_{1,m} = \xi, x'_{2,m} = 1)}\right) \tag{3.17}$$

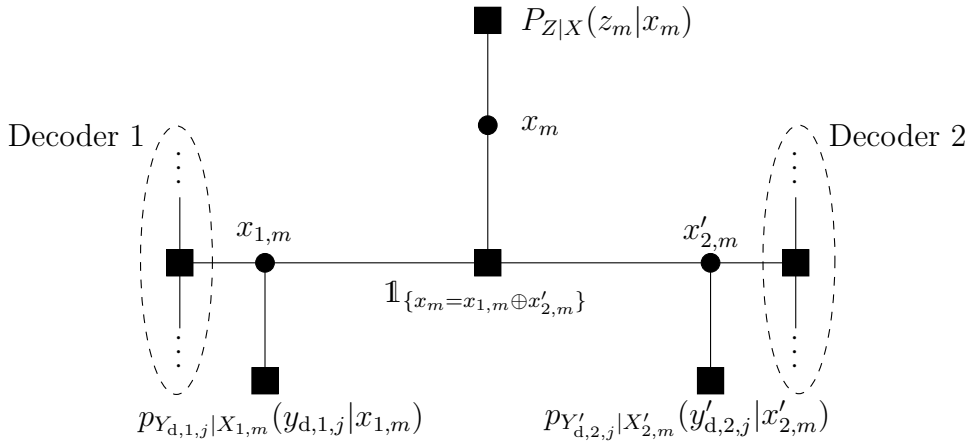


Figure 3.4.: One section of the factor graph at the destination for scalar quantization, where $j = \lceil m/\log_2(M_i)\rceil$.

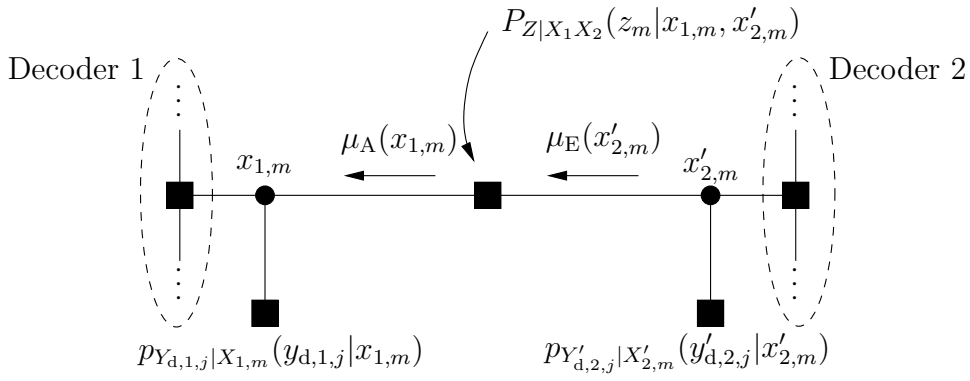


Figure 3.5.: One section of the factor graph at the destination for two-dimensional quantization with messages $\mu_{\mathrm{A}}(x_{1,m})$ and $\mu_{\mathrm{E}}(x'_{2,m})$, where $j = \lceil m/\log_2(M_i)\rceil$.

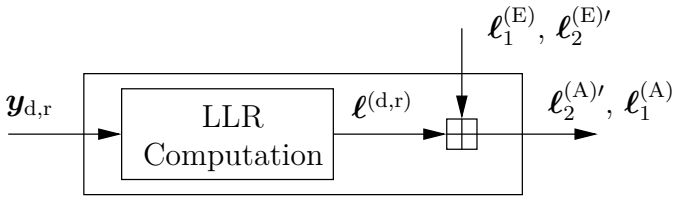


Figure 3.6.: Relay check node for analog and soft bit transmission from the relay.

$$\ell(x_{1,m}, x'_{2,m}|z_m) \triangleq \ln\left(\frac{P_{Z|X_1X_2}(z_m|x_{1,m}=0, x'_{2,m}=1)}{P_{Z|X_1X_2}(z_m|x_{1,m}=1, x'_{2,m}=0)}\right). \tag{3.18}$$

Then, for $m = 1, 2, \ldots, n$, $\ell_{1,m}^{(\mathrm{A})}$ and $\ell_{2,m}^{(\mathrm{A})\prime}$ are given by

$$\ell_{1,m}^{(\mathrm{A})} = \ln\left(\frac{1 + e^{\ell_{2,m}^{(\mathrm{E})\prime}} e^{\ell(x_{1,m}=0, x'_{2,m}|z_m)}}{e^{\ell_{2,m}^{(\mathrm{E})\prime}} e^{-\ell(x_{1,m}, x'_{2,m}|z_m)} + e^{-\ell(x_{1,m}, x'_{2,m}=1|z_m)}}\right) \tag{3.19}$$

$$\ell_{2,m}^{(\mathrm{A})\prime} = \ln\left(\frac{1 + e^{\ell_{1,m}^{(\mathrm{E})}} e^{\ell(x_{1,m}, x'_{2,m}=0|z_m)}}{e^{\ell_{1,m}^{(\mathrm{E})}} e^{\ell(x_{1,m}, x'_{2,m}|z_m)} + e^{-\ell(x_{1,m}=1, x'_{2,m}|z_m)}}\right). \tag{3.20}$$

Appendix B.1 contains a derivation of (3.19) and (3.20).

3.1.5. Reference schemes

In addition to point-to-point links without the use of the relay, we also consider analog transmission of the soft information from the relay and transmission as soft bit as reference schemes. For analog transmission [YK07], we have $\boldsymbol{s}_{\mathrm{r}} = a\boldsymbol{\ell}$, where a is chosen such that the power constraint at the relay is satisfied, i.e., $\mathrm{E}[|S_{\mathrm{r},m}|^2] \leq P_{\mathrm{r}}$. We also include transmission as soft bit [LVWD06], so that $s_{\mathrm{r},m} = a\tanh(\ell_m/2)$. Again, the normalization factor a ensures that the power constraint at the relay is met. For the sake of completeness, Figure 3.6 shows the corresponding operations for those reference schemes at the destination.

3.2. Quantizer design

In this section, we study the design of quantizers for application at the relay to allow maximal exchange of soft information between the component decoders at the destination. Throughout, we restrict the design framework to the case where the quantizer output can be perfectly recovered at the destination.

3.2.1. One-dimensional quantizers

The most common distortion function for the design of quantizers and for rate-distortion problems involving real-valued random variables is the squared-error distortion for its simplicity and convenience in analysis, despite its lack of perceptual meaningfulness for some problems [Gra90]. As already pointed out in Section 2.2, finding the right distortion function for a particular problem can be a difficult and controversial task. It seems to be equally hard to choose a distortion function for the problem we consider here, namely quantizing the soft information at the relay node. Therefore, we follow the approach taken by Tishby *et al.* presented in Section 2.2, where they deal with the rate-distortion problem in a different way using the notion of *relevance through another variable*. Instead of putting the constraint on the average distortion for some distortion function chosen *a-priori*, the constraint is that the quantization $q(L)$ should contain some minimum level of information about a third variable, the *relevant* variable, which, in our case, is the random variable $X = X_1 \oplus X_2$. Given a random variable L representing the soft information at the relay, the goal is to find a quantized version $q(L)$ that contains as much *relevant* information as possible, which is information about X. That is, instead of forcing, e.g., the squared error between L and $q(L)$ to be small, the goal is to preserve as much information as possible in $q(L)$ about X.

We now formalize these ideas. In order to design a quantization function q with N quantization regions, we aim at solving the optimization problem

$$I^* = \sup_{q:\mathbb{R}\to\mathcal{Z}} I(X; q(L)) \quad \text{s.t. } |\mathcal{Z}| = N. \tag{3.21}$$

In order to compute an approximation of (3.21) in the following, we make a number of simplifying assumptions. First, finely quantize the range of the continuous random variable L with density $p_L(\ell)$, yielding a random variable $\bar{L}$ with finite range and mass function $P_{\bar{L}}(\ell)$, so that $\bar{L}$ is from a finite set $\mathcal{L}$. The optimization problem at hand now is

$$\bar{q}^* = \operatorname*{argmax}_{\bar{q}:\mathcal{L}\to\mathcal{Z}} I(X; \bar{q}(\bar{L})) \quad \text{s.t. } |\mathcal{Z}| = N. \tag{3.22}$$

The second step comprises a transformation of the mapping $\bar{q}$, similar to before, into a conditional mass function $P_{Z|\bar{L}}(z|\ell) = \mathbb{1}_{\{\bar{q}(\ell)=z\}}$ allowing us to rewrite (3.22) as

$$P^*_{Z|\bar{L}} = \operatorname*{argmax}_{P_{Z|\bar{L}}\in\mathcal{P}_1} I(X; Z), \tag{3.23}$$

where the constraint set

$$\mathcal{P}_1 = \Big\{P_{Z|\bar{L}} : P_{Z|\bar{L}}(z|\ell) \in \{0,1\} \ \forall\ z \in \mathcal{Z}, \forall\ \ell \in \mathcal{L}, \sum_z P_{Z|\bar{L}}(z|\ell) = 1 \ \forall\ \ell \in \mathcal{L}, \\ |\mathcal{Z}| = N\Big\} \tag{3.24}$$

ensures that the mapping $P_{Z|\bar{L}}$ is a valid conditional mass function, and that it represents

a scalar deterministic quantizer with N quantization regions.

Before proceeding, consider the related problem

$$\max_{P_{Z|\bar{L}} \in \mathcal{P}_1'} I(X;Z), \tag{3.25}$$

with

$$\mathcal{P}_1' = \Big\{ P_{Z|\bar{L}} : P_{Z|\bar{L}}(z|\ell) \geq 0 \, \forall z \in \mathcal{Z}, \forall \ell \in \mathcal{L}, \sum_z P_{Z|\bar{L}}(z|\ell) = 1 \ \forall \ell \in \mathcal{L}, |\mathcal{Z}| = N \Big\}. \tag{3.26}$$

The set $\mathcal{P}_1'$ is a polyhedron, and therefore convex [BSS06]; also, $\mathcal{P}_1'$ is bounded and closed, and therefore compact [BSS06, Section 2.2]. Further, $I(X;Z)$ is convex in the distribution $P_{Z|\bar{L}}$, for fixed P_X and $P_{\bar{L}|X}$; this is because $I(X;Z)$ is convex in $P_{Z|X}$ for fixed P_X [CT06], and

$$P_{Z|X}(z|x) = \sum_\ell P_{Z|\bar{L}}(z|\ell) P_{\bar{L}|X}(\ell|x) \tag{3.27}$$

is linear in $P_{Z|\bar{L}}$ [BV04, Chapter 3.2.2]. Consequently, Problem (3.25) is a convex *maximization* over a compact polyhedral set, whose maximum is therefore attained at an extreme point of $\mathcal{P}_1'$ [BSS06, Theorem 3.4.7]. However, to solve (3.25) with global optimality, one still needs to search all extreme points of $\mathcal{P}_1'$, which is prohibitively complex since there are $N^{|\mathcal{L}|}$ of those. Since all the extreme points of $\mathcal{P}_1'$ correspond to a distribution $P_{Z|\bar{L}}(z|\ell) \in \{0,1\}$, $\forall z \in \mathcal{Z}, \forall \ell \in \mathcal{L}$, this also shows that scalar *deterministic* quantization (as considered in (3.23)) is optimal, i.e., the mutual information $I(X;Z)$ cannot be improved by allowing randomized quantization, so that (3.23) and (3.25) have the same maximizer.

In the following, we restrict the optimization to finding a locally optimal solution to Problem (3.23).

Proposition 3.1. Solving Problem (3.23) is equivalent to solving

$$P^*_{Z|\bar{L}} = \operatorname*{argmin}_{P_{Z|\bar{L}} \in \mathcal{P}_1} \mathrm{E}\left[D_{\mathrm{KL}}\left(P_{X|\bar{L}}(\cdot|\bar{L}) || P_{X|Z}(\cdot|Z) \right) \right]. \tag{3.28}$$

Proof. By the chain rule for mutual information [CT06], we have

$$I(X;\bar{L},Z) = I(X;Z) + I(X;\bar{L}|Z) = I(X;\bar{L}) + I(X;Z|\bar{L}). \tag{3.29}$$

Since $X \leftrightarrow \bar{L} \leftrightarrow Z$ forms a Markov chain, we have $P_{X|\bar{L}Z}(x|\ell,z) = P_{X|\bar{L}}(x|\ell)$ as well as $I(X;Z|\bar{L}) = 0$, so that by (3.29) one obtains

$$I(X;Z) = I(X;\bar{L}) + \underbrace{I(X;Z|\bar{L})}_{=0} - I(X;\bar{L}|Z) \tag{3.30}$$

$$= I(X;\bar{L}) - \sum_{x,\ell,z} P_{X\bar{L}Z}(x,\ell,z) \log_2 \left(\frac{P_{X\bar{L}|Z}(x,\ell|z)}{P_{X|Z}(x|z) P_{\bar{L}|Z}(\ell|z)} \right) \tag{3.31}$$

$$= I(X;\bar{L}) - \sum_{\ell,z} P_{\bar{L}Z}(\ell,z) \sum_{x} P_{X|\bar{L}}(x|\ell) \log_2 \left(\frac{P_{X|\bar{L}}(x|\ell)}{P_{X|Z}(x|z)} \right) \tag{3.32}$$

$$= I(X;\bar{L}) - \sum_{\ell,z} P_{\bar{L}Z}(\ell,z) D_{\mathrm{KL}}\Big(P_{X|\bar{L}}(\cdot|\ell) || P_{X|Z}(\cdot|z) \Big) \tag{3.33}$$

$$= I(X;\bar{L}) - \mathrm{E}\left[D_{\mathrm{KL}}\Big(P_{X|\bar{L}}(\cdot|\bar{L}) || P_{X|Z}(\cdot|Z) \Big) \right]. \tag{3.34}$$

For $I(X;\bar{L})$ is fixed for a given $P_{X\bar{L}}$, the maximization in (3.23) is equivalent to minimizing the expected relative entropy $\mathrm{E}\left[D_{\mathrm{KL}}\Big(P_{X|\bar{L}}(\cdot|\bar{L}) || P_{X|Z}(\cdot|Z) \Big) \right]$. ■

Hence, the relative entropy between $P_{X|\bar{L}}(\cdot|\ell)$ and $P_{X|Z}(\cdot|z)$ can be seen as the distortion function

$$d(\ell,z) = D_{\mathrm{KL}}\left(P_{X|\bar{L}}(\cdot|\ell) || P_{X|Z}(\cdot|z) \right) = \sum_{x} P_{X|\bar{L}}(x|\ell) \log_2 \left(\frac{P_{X|\bar{L}}(x|\ell)}{P_{X|Z}(x|z)} \right) \tag{3.35}$$

for the problem at hand. Note that relative entropy *emerged* as the right distortion function for the quantizer design problem by posing it as an optimization of relevant information.

In Problem (3.28), the probability distribution $P_{X\bar{L}}$ is fixed and known, and has to be obtained numerically. Hence, $P_{Z|\bar{L}}$ is the only free variable. This is because $P_{X\bar{L}}$ is fixed, and P_Z and $P_{X|Z}$ are fully determined by $P_{Z|\bar{L}}$. Although the optimal distribution $P^*_{Z|\bar{L}}$ cannot be obtained in closed form, we propose an iterative optimization algorithm that can be shown to converge to a Karush-Kuhn-Tucker (KKT) point [BSS06, Section 4.3] of (3.25). The algorithm is given in Algorithm 3.1. Convergence of the algorithm follows since the update of the mapping $P_{Z|\bar{L}}$ in line 11 of the algorithm is chosen such that the average distortion of the new mapping is no worse than that of the previous mapping, and from the concavity of $I(X;\bar{L}|Z)$ with respect to $P_{Z|\bar{L}}$.

In essence, this algorithm is reminiscent of the Lloyd algorithm [Llo82], where however, in our algorithm, the distortion function $d(\ell,z)$ is given by the relative entropy (3.35), and therefore depends on the mapping $P_{Z|\bar{L}}$. This is reflected in the update of line 10 of the algorithm. Algorithm 3.1 is also related to the iterative information bottleneck algorithm [TPB99] (cf. Section 2.2), where Algorithm 3.1 can be recovered by choosing the parameter β of [TPB99] to be $\beta \gg 0$ to ensure that the mapping $P_{Z|\bar{L}}(z|\ell) \in \{0,1\}$, $\forall z \in \mathcal{Z}, \forall \ell \in \mathcal{L}$, corresponds to a deterministic quantizer. As highlighted in Chapter 2, the iterative algorithm of [TPB99] in turn is reminiscent of the celebrated BAA [Bla72] for computing channel capacities and rate-distortion functions, with the main difference that the algorithm for computing the mapping in the information bottleneck setting updates the distribution $P_{X|Z}$ and also incorporates the dependency of the distortion $d(\ell,z)$ on the mapping $P_{Z|\bar{L}}$ to be optimized. Relative entropy as a distortion function for vector quantizer design was also employed in [DFA+10]; however, the algorithm in [DFA+10] is explicitly formulated for quantizing LLRs, whereas the Algorithm 3.1 can be used to design a quantizer for maximum mutual information irrespective of whether $\bar{L}$ is an LLR or not, as long as the joint probability mass function $P_{X\bar{L}}$ or an estimate thereof is available. Further, from Algorithm 3.1, the connection to the information bottleneck principle [TPB99] is

Algorithm 3.1 Algorithm to compute $P_{Z|\bar{L}}$.

Input: $P_{X\bar{L}}(x,\ell), \mathcal{Z}, \epsilon > 0$
Initialization: randomly choose a valid mapping $P^{(0)}_{Z|\bar{L}}(z|\ell) \in \{0,1\}$, $k \leftarrow 1$
$P^{(0)}_Z(z) \leftarrow \sum_\ell P_{\bar{L}}(\ell) P^{(0)}_{Z|\bar{L}}(z|\ell)$
$P^{(0)}_{X|Z}(x|z) \leftarrow \left(1/P^{(0)}_Z(z)\right) \sum_\ell P_{X\bar{L}}(x,\ell) P^{(0)}_{Z|\bar{L}}(z|\ell)$
$d^{(0)}(\ell,z) \leftarrow D_{\mathrm{KL}}(P_{X|\bar{L}}(\cdot|\ell)||P^{(0)}_{X|Z}(\cdot|z))$
find, for each ℓ, $z^*_\ell = \mathrm{argmin}_z\, d^{(0)}(\ell,z)$, and set $P^{(1)}_{Z|\bar{L}}(z|\ell) \leftarrow \mathbb{1}_{z=z^*_\ell}$
while $\sum_{\ell,z} \left| P^{(k)}_{Z|\bar{L}}(z|\ell) - P^{(k-1)}_{Z|\bar{L}}(z|\ell) \right| / (|\mathcal{L}| \cdot N) \geq \epsilon$ **do**
    $P^{(k)}_Z(z) \leftarrow \sum_\ell P_{\bar{L}}(\ell) P^{(k)}_{Z|\bar{L}}(z|\ell)$
    $P^{(k)}_{X|Z}(x|z) \leftarrow \left(1/P^{(k)}_Z(z)\right) \sum_\ell P_{X\bar{L}}(x,\ell) P^{(k)}_{Z|\bar{L}}(z|\ell)$
    $d^{(k)}(\ell,z) \leftarrow D_{\mathrm{KL}}(P_{X|\bar{L}}(\cdot|\ell)||P^{(k)}_{X|Z}(\cdot|z))$
    find, for each ℓ, $z^*_\ell = \mathrm{argmin}_z\, d^{(k)}(\ell,z)$, and set $P^{(k+1)}_{Z|\bar{L}}(z|\ell) \leftarrow \mathbb{1}_{z=z^*_\ell}$
    $k \leftarrow k+1$
end while
$P_{Z|\bar{L}}(z|\ell) \leftarrow P^{(k)}_{Z|\bar{L}}(z|\ell)$
$P_Z(z) \leftarrow \sum_\ell P_{\bar{L}}(\ell) P_{Z|\bar{L}}(z|\ell)$
$P_{X|Z}(x|z) \leftarrow (1/P_Z(z)) \sum_\ell P_{X\bar{L}}(x,\ell) P_{Z|\bar{L}}(z|\ell)$
$I(X;Z) \leftarrow \sum_{x,z} P_Z(z) P_{X|Z}(x|z) \log_2 \left(\frac{P_{X|Z}(x|z)}{P_X(x)} \right)$
$H(Z) \leftarrow -\sum_z P_Z(z) \log_2(P_Z(z))$

evident, which is a general framework for the tradeoff between rate and relevant mutual information. Quantization of log-likelihood ratios was also considered in [Rav09] under the assumption of them being conditionally Gaussian distributed.

Since the resulting mapping $P_{Z|\bar{L}}$ represents a scalar quantizer, the mutual information is $I(\bar{L};Z) = H(Z)$, which is also the rate of the resulting quantizer. Fixing N, the rate of the quantizer is therefore upper bounded by $\log_2(N)$. Since Problem (3.23) is a convex maximization problem, Algorithm 3.1 is only guaranteed to find a solution satisfying the necessary conditions for local optimality. In our attempt to find a good mutual-information preserving quantizer, the algorithm is therefore repeatedly carried out with random starting conditions until a satisfactory solution is obtained.

3.2.2. Two-dimensional quantizers

As for one-dimensional quantization of $\boldsymbol{\ell}$, the framework for designing a two-dimensional quantizer for $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2'$ will be established using mutual information as a figure of merit, but with a different expression as relevant information, whose choice will be motivated by the following two arguments.

Inspection of the iterative decoding process at the destination

During that process, cf. Figure 3.2, the component decoders produce random variables $\boldsymbol{L}_1^{(\mathrm{E})}$ and $\boldsymbol{L}_2^{(\mathrm{E})}$ with mutual information $I(X_1; L_1^{(\mathrm{E})})$ and $I(X_2; L_2^{(\mathrm{E})})$, which is the input information to the corresponding relay check node. At this point, again assuming error-free transmission of the quantizer output $\boldsymbol{Z}$, we note that the relay check node in the receiver processes $\boldsymbol{Z}$ and $\boldsymbol{L}_i^{(\mathrm{E})}$ to produce *a-priori* information for the corresponding component decoder. Therefore, we would like the quantizer at the relay to be such that $I(X_i; Z, L_j^{(\mathrm{E})})$, $i, j \in \{1, 2\}$, $i \neq j$, is maximal, both for the information exchange from decoder 1 to decoder 2, and vice versa. Since

$$I(X_i; Z, L_j^{(\mathrm{E})}) = I(X_i; L_j^{(\mathrm{E})}) + I(X_i; Z|L_j^{(\mathrm{E})}) \tag{3.36}$$

$$= I(X_i; Z|L_j^{(\mathrm{E})}), \tag{3.37}$$

where it is assumed that $I(X_i; L_j^{(\mathrm{E})}) = 0$ for $i \neq j$, we are left with maximizing $I(X_i; Z|L_j^{(\mathrm{E})})$, which is, however, hard to maximize due to its dependence on the variable $L_j^{(\mathrm{E})}$ that changes its statistics with an increasing number of iterations. We therefore propose the following. Observing that the extrinsic information $L_j^{(\mathrm{E})}$ from component decoder j being perfectly reliable corresponds to X_j being given, we optimize the mutual information $I(X_i; Z|X_j)$ instead of $I(X_i; Z|L_j^{(\mathrm{E})})$, thereby removing the dependency on the changing statistics of $L_j^{(\mathrm{E})}$. Consequently, to allow maximal information transfer from decoder 1 to decoder 2, $I(X_2; Z|X_1)$ should be maximized, and analogously, for decoder 1 to receive maximal information from decoder 2, $I(X_1; Z|X_2)$ should be as large as possible. Various combinations of these information expressions can be taken to form the relevant information term. We propose to take the sum of $I(X_1; Z|X_2)$ and $I(X_2; Z|X_1)$ as the relevant information term, i.e., $I_{\mathrm{rel}} \triangleq I(X_1; Z|X_2) + I(X_2; Z|X_1)$.

Connection with one-dimensional quantization of soft information about XORed code bits

In addition to the motivation above, we also establish a connection between the proposed cost function I_{rel} for two-dimensional quantization with the cost function employed for the design of one-dimensional quantizers in the following proposition.

Proposition 3.2. Let L be the soft information about $X = X_1 \oplus X_2$ of two independent and equally likely bits X_1 and X_2, and let q_1 be the quantization function of a quantizer

processing L, so that $Z = q_1(L)$. Further, assume that $p_{L_i|X_i}(\ell_i|x_i)$ satisfies the symmetry condition $p_{L_i|X_i}(\ell_i|0) = p_{L_i|X_i}(-\ell_i|1)$. Then, $I(X; q_1(L)) = I(X_1, X_2; q_1(L))$ and $I(X_1; q_1(L)) = I(X_2; q_1(L)) = 0$.

The proof is relegated to Appendix B.2. From Proposition 3.2 we see that maximizing $I(X; q_1(L))$ corresponds to maximizing $I(X_1, X_2; q_1(L))$, subject to $I(X_1; q_1(L)) = I(X_2; q_1(L)) = 0$. For two-dimensional quantization using a quantization function $q_2 : \mathbb{R} \times \mathbb{R} \to \mathcal{Z}$, we relax the condition $I(X_1; q_2(L_1, L_2)) = I(X_2; q_2(L_1, L_2)) = 0$, but seek to choose the quantization function q_2 such that $q_2(L_1, L_2)$ carries as much information about the pair (X_1, X_2) as possible, while carrying little information about X_1 and X_2 alone. This is reflected in the choice of

$$I_{\text{rel}} = I(X_1; q_2(L_1, L_2)|X_2) + I(X_2; q_2(L_1, L_2)|X_1) \tag{3.38}$$

$$= 2I(X_1, X_2; q_2(L_1, L_2)) - I(X_1; q_2(L_1, L_2)) - I(X_2; q_2(L_1, L_2)). \tag{3.39}$$

To compute a two-dimensional quantizer, we finely quantize the ranges of the continuous random variables L_1 and L_2 with densities $p_{L_1}(\ell_1)$ and $p_{L_2}(\ell_2)$ to obtain discrete variables $\bar{L}_1 \in \mathcal{L}_1$ and $\bar{L}_2 \in \mathcal{L}_2$ with probability mass functions $P_{\bar{L}_1}(\ell_1)$ and $P_{\bar{L}_2}(\ell_2)$, where both $\mathcal{L}_1$ and $\mathcal{L}_2$ are finite sets. Writing the mapping $\bar{q} : \mathcal{L}_1 \times \mathcal{L}_2 \to \mathcal{Z}$ as $P_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2) = \mathbb{1}_{\{\bar{q}(\ell_1,\ell_2)=z\}}$, we pose the optimization problem we wish to solve as

$$P^*_{Z|\bar{L}_1\bar{L}_2} = \underset{P_{Z|\bar{L}_1\bar{L}_2} \in \mathcal{P}_2}{\operatorname{argmax}} \; I_{\text{rel}}, \tag{3.40}$$

where

$$\mathcal{P}_2 = \Bigg\{ P_{Z|\bar{L}_1\bar{L}_2} : P_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2) \in \{0, 1\}, \forall (z, \ell_1, \ell_2) \in (\mathcal{Z} \times \mathcal{L}_1 \times \mathcal{L}_2)$$

$$\sum_z P_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2) = 1, \forall (\ell_1, \ell_2) \in (\mathcal{L}_1 \times \mathcal{L}_2), |\mathcal{Z}| = N \Bigg\}. \tag{3.41}$$

Proposition 3.3. Solving Problem (3.40) is equivalent to solving

$$\underset{P_{Z|\bar{L}_1\bar{L}_2} \in \mathcal{P}_2}{\operatorname{argmin}} \Bigg\{ \mathrm{E}\left[2D_{\text{KL}}\left(P_{X_1X_2|\bar{L}_1,\bar{L}_2}(\cdot,\cdot|\bar{L}_1, \bar{L}_2)||P_{X_1X_2|Z}(\cdot,\cdot|Z)\right)\right] \tag{3.42}$$

$$-\mathrm{E}\left[D_{\text{KL}}\left(P_{X_1|\bar{L}_1}(\cdot|\bar{L}_1)||P_{X_1|Z}(\cdot|Z)\right)\right] - \mathrm{E}\left[D_{\text{KL}}\left(P_{X_2|\bar{L}_2}(\cdot|\bar{L}_2)||P_{X_2|Z}(\cdot|Z)\right)\right] \Bigg\}.$$

Proof. Similar to the proof of Proposition 3.1, $(X_1, X_2) \leftrightarrow (\bar{L}_1, \bar{L}_2) \leftrightarrow Z$ forms a Markov chain and $I(X_1, X_2; \bar{L}_1, \bar{L}_2)$ is fixed, so that maximizing I_{rel} is equivalent to minimizing $2I(X_1, X_2; \bar{L}_1, \bar{L}_2|Z) - I(X_1; \bar{L}_1|Z) - I(X_2; \bar{L}_2|Z)$, which can be expressed in terms of relative entropies as in (3.42). ■

To compute an approximate solution to (3.42), we can use an appropriately modified version of Algorithm 3.1. The algorithm for the two-dimensional quantizer design is given in

Type	γ_r	N	Decision Border(s)	$\ln\left(\frac{P_{X\mid Z}(0\mid z)}{P_{X\mid Z}(1\mid z)}\right)$	$I(X;\bar{L})$	$I(X;Z)$	$H(Z)$
MIC	0 dB	2	$[0]$	$[-5.61, 5.61]$	0.985	0.964	1
MIC	−2 dB	3	$[-1.45, 1.45]$	$[-3.55, 0, 3.55]$	0.762	0.714	1.425
LM	−2 dB	3	$[-3.25, 3.25]$	$[-4.91, 0, 4.91]$	0.762	0.634	1.585
UF	−2 dB	3	$[-4.75, 4.75]$	$[-6.05, 0, 6.05]$	0.762	0.484	1.495
MIC	−3 dB	5	$[-2.30, -0.71, 0.71, 2.30]$	$[-3.35, -1.30, 0, 1.30, 3.35]$	0.505	0.485	2.286

Table 3.1.: Comparison of quantizer characteristics.

Algorithm 3.2. Similarly to Section 3.2.1, the mass functions $P_{X_1\bar{L}_1}$ and $P_{X_2\bar{L}_2}$ are obtained numerically, and we run Algorithm 3.2 with different starting conditions to ensure that a good two-dimensional quantizer is found. For a deterministic quantizer with quantization rule $q(\ell_1, \ell_2)$ we then have $I(\bar{L}_1, \bar{L}_2; Z) = H(Z)$ as the rate of the resulting quantizer.

3.2.3. Examples of quantizers

In this section, we present some examples for quantizers obtained with Algorithms 3.1 and 3.2. In all cases, the underlying channel codes are recursive convolutional codes with generator

$$G(D) = \left(1, \frac{1 + D^4}{1 + D + D^2 + D^3 + D^4}\right) \tag{3.43}$$

and information blocklength $k = k_1 = k_2 = 1996$. Binary phase shift keying (BPSK) modulation is employed at the sources.

The first set of examples involves some illustration on how the designed quantizers look for different values of the source–relay SNR $\gamma_{i,\mathrm{r}}$ and alphabet sizes N, in case of one-dimensional quantizers for $\boldsymbol{\ell}$ and BCJR soft decoders [BCJR74] at the relay. Here, we assume symmetric source–relay channels, i.e., $\gamma_\mathrm{r} = \gamma_{1,\mathrm{r}} = \gamma_{2,\mathrm{r}}$. Numerical characteristics of the quantizers designed with the mutual information criterion (MIC) are summarized in Table 3.1. As the source–relay SNR decreases, the number of quantization regions required to achieve a mutual information $I(X;Z)$ close to the limit $I(X;\bar{L})$ increases, leading to an increase in rate on the relay–destination link. To highlight the effectiveness of the proposed quantizer design framework using mutual information as a figure of merit compared to other well-known quantization methods, we also show the parameters of the Lloyd-Max (LM) [Llo82] and uniform (UF) quantizer in Table 3.1 for $\gamma_\mathrm{r} = -2\,\mathrm{dB}$ and $N = 3$. Evidently, both the LM and the UF quantizer require a higher rate than the quantizer designed with the proposed algorithm, while preserving less relevant mutual information.

In the second set of examples, the relay performs soft demapping only. Figure 3.7 depicts the partitioning of the (ℓ_1, ℓ_2)-plane into quantization regions as obtained by the iterative optimization algorithm. Each of the resulting regions is color coded, with each color corresponding to one symbol of the quantizer alphabet $\mathcal{Z}$. Using $N = 3$ regions, the quantizer is shown in Figure 3.7(a) for symmetric source–relay links at $\gamma_{1,\mathrm{r}} = \gamma_{2,\mathrm{r}} = 4\,\mathrm{dB}$. Note that this partition effectively mimics one-dimensional quantization of the soft

Algorithm 3.2 Algorithm to compute $P_{Z|\bar{L}_1\bar{L}_2}$.

Input: $P_{X_1X_2\bar{L}_1\bar{L}_2}(x_1, x_2, \ell_1, \ell_2), \mathcal{Z}, \epsilon > 0$

Initialization: randomly choose a valid mapping $P^{(0)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2) \in \{0, 1\}$, $k \leftarrow 1$

$P^{(0)}_Z(z) \leftarrow \sum_{\ell_1,\ell_2} P_{\bar{L}_1\bar{L}_2}(\ell_1, \ell_2) P^{(0)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2)$

$P^{(0)}_{X_1X_2|Z}(x_1, x_2|z) \leftarrow \left(1/P^{(0)}_Z(z)\right) \sum_{\ell_1,\ell_2} P_{X_1X_2\bar{L}_1\bar{L}_2}(x_1, x_2, \ell_1, \ell_2) P^{(0)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2)$

$P^{(0)}_{X_1|Z}(x_1|z) \leftarrow \sum_{x_2} P^{(0)}_{X_1X_2|Z}(x_1, x_2|z)$

$P^{(0)}_{X_2|Z}(x_2|z) \leftarrow \sum_{x_1} P^{(0)}_{X_1X_2|Z}(x_1, x_2|z)$

$d^{(0)}(\ell_1, \ell_2, z) \leftarrow 2D_{\mathrm{KL}}\left(P_{X_1X_2|\bar{L}_1\bar{L}_2}(\cdot,\cdot|\ell_1, \ell_2)||P^{(0)}_{X_1X_2|Z}(\cdot,\cdot|z)\right)$
$\quad -D_{\mathrm{KL}}\left(P_{X_1|\bar{L}_1}(\cdot|\ell_1)||P^{(0)}_{X_1|Z}(\cdot|z)\right) - D_{\mathrm{KL}}\left(P_{X_2|\bar{L}_2}(\cdot|\ell_2)||P^{(0)}_{X_2|Z}(\cdot|z)\right)$

find, for each (ℓ_1, ℓ_2), $z^*_{\ell_1,\ell_2} = \mathrm{argmin}_z\, d^{(0)}(\ell_1, \ell_2, z)$,
and set $P^{(1)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2) \leftarrow \mathbb{1}_{z=z^*_{\ell_1,\ell_2}}$

while $\sum_{\ell_1,\ell_2,z} \left|P^{(k)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2) - P^{(k-1)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2)\right| / (|\mathcal{L}_1| \cdot |\mathcal{L}_2| \cdot N) \geq \epsilon$ **do**

$\quad P^{(k)}_Z(z) \leftarrow \sum_{\ell_1,\ell_2} P_{\bar{L}_1\bar{L}_2}(\ell_1, \ell_2) P^{(k)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2)$

$\quad P^{(k)}_{X_1X_2|Z}(x_1, x_2|z) \leftarrow \left(1/P^{(k)}_Z(z)\right) \sum_{\ell_1,\ell_2} P_{X_1X_2\bar{L}_1\bar{L}_2}(x_1, x_2, \ell_1, \ell_2) P^{(k)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2)$

$\quad P^{(k)}_{X_1|Z}(x_1|z) \leftarrow \sum_{x_2} P^{(k)}_{X_1X_2|Z}(x_1, x_2|z)$

$\quad P^{(k)}_{X_2|Z}(x_2|z) \leftarrow \sum_{x_1} P^{(k)}_{X_1X_2|Z}(x_1, x_2|z)$

$\quad d^{(k)}(\ell_1, \ell_2, z) \leftarrow 2D_{\mathrm{KL}}\left(P_{X_1X_2|\bar{L}_1\bar{L}_2}(\cdot,\cdot|\ell_1, \ell_2)||P^{(k)}_{X_1X_2|Z}(\cdot,\cdot|z)\right)$
$\quad\quad -D_{\mathrm{KL}}\left(P_{X_1|\bar{L}_1}(\cdot|\ell_1)||P^{(k)}_{X_1|Z}(\cdot|z)\right) - D_{\mathrm{KL}}\left(P_{X_2|\bar{L}_2}(\cdot|\ell_2)||P^{(k)}_{X_2|Z}(\cdot|z)\right)$

$\quad$ find, for each (ℓ_1, ℓ_2), $z^*_{\ell_1,\ell_2} = \mathrm{argmin}_z\, d^{(k)}(\ell_1, \ell_2, z)$,
$\quad$ and set $P^{(k+1)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2) \leftarrow \mathbb{1}_{z=z^*_{\ell_1,\ell_2}}$

$\quad k \leftarrow k + 1$

end while

$P_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2) \leftarrow P^{(k)}_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2)$

$P_Z(z) \leftarrow \sum_{\ell_1,\ell_2} P_{\bar{L}_1\bar{L}_2}(\ell_1, \ell_2) P_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2)$

$P_{X_1X_2|Z}(x|z) \leftarrow (1/P_Z(z)) \sum_{\ell_1,\ell_2} P_{X_1X_2\bar{L}_1\bar{L}_2}(x_1, x_2, \ell_1, \ell_2) P_{Z|\bar{L}_1\bar{L}_2}(z|\ell_1, \ell_2)$

$P_{X_1|Z}(x_1|z) \leftarrow \sum_{x_2} P_{X_1X_2|Z}(x_1, x_2|z)$

$P_{X_2|Z}(x_2|z) \leftarrow \sum_{x_1} P_{X_1X_2|Z}(x_1, x_2|z)$

$I(X_1, X_2; Z) \leftarrow \sum_{x_1,x_2,z} P_Z(z) P_{X_1X_2|Z}(x_1, x_2|z) \log_2 \left(\frac{P_{X_1X_2|Z}(x_1, x_2|z)}{P_{X_1X_2}(x_1, x_2)}\right)$

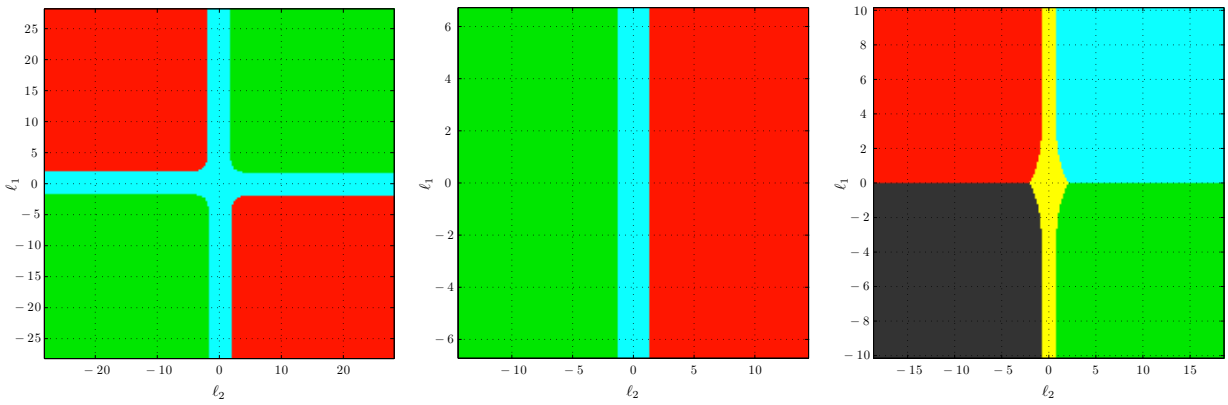
Algorithm 3.2 (continued)

24: $I(X_1;Z) \leftarrow \sum_{x_1,z} P_Z(z) P_{X_1|Z}(x_1|z) \log_2 \left(\frac{P_{X_1|Z}(x_1|z)}{P_{X_1}(x_1)} \right)$

25: $I(X_2;Z) \leftarrow \sum_{x_2,z} P_Z(z) P_{X_2|Z}(x_2|z) \log_2 \left(\frac{P_{X_2|Z}(x_2|z)}{P_{X_2}(x_2)} \right)$

26: $I_{\mathrm{rel}} \leftarrow 2I(X_1, X_2; Z) - I(X_1; Z) - I(X_2; Z)$

27: $H(Z) \leftarrow -\sum_z P_Z(z) \log_2(P_Z(z))$



(a) $\gamma_{1,\mathrm{r}} = \gamma_{2,\mathrm{r}} = 4\,\mathrm{dB}$. (b) $\gamma_{1,\mathrm{r}} = -8\,\mathrm{dB}, \gamma_{2,\mathrm{r}} = -1\,\mathrm{dB}$. (c) $\gamma_{1,\mathrm{r}} - 4\,\mathrm{dB}, \gamma_{2,\mathrm{r}} = 1\,\mathrm{dB}$.

Figure 3.7.: Quantizers obtained for two-dimensional quantization of $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}'_2$ at the relay. The quantizer in (c) is also suitable for $\gamma_{1,\mathrm{r}} = -1\,\mathrm{dB}, \gamma_{2,\mathrm{r}} = 4\,\mathrm{dB}$, and QPSK modulation with soft demapping.

information $\boldsymbol{\ell}$ about the XOR-coded bits. In contrast, if channel conditions on the source–relay links are profoundly different, then the relay should preferably allocate more of the rate available on the relay–destination channel to the stronger user, and this is exactly achieved with the two-dimensional formulation of the quantization problem at the relay, as shown in Figure 3.7(b) for again $N = 3$ regions, where $\gamma_{1,\mathrm{r}} = -8$ dB and $\gamma_{2,\mathrm{r}} = -1$ dB. Note that in this rather extreme case, all the quantization rate is allocated to the second user, whose soft information at the relay is vastly more reliable than the one of the first user. Finally, Figure 3.7(c) displays a typical quantization mapping obtained for $\gamma_{1,\mathrm{r}} = -4$ dB, $\gamma_{2,\mathrm{r}} = 1$ dB, and $N = 5$ regions.

3.3. Simulation results

3.3.1. Additive white Gaussian noise channels

We first show bit error rate (BER) results for AWGN channels, i.e., $h_{i,\mathrm{r}} = h_{i,\mathrm{d}} = h_{\mathrm{r,d}} = 1$, and symmetric links, for which $\gamma_{\mathrm{d}} = \gamma_{1,\mathrm{d}} = \gamma_{2,\mathrm{d}}$ as well as $\gamma_{\mathrm{r}} = \gamma_{1,\mathrm{r}} = \gamma_{2,\mathrm{r}}$. Both sources

employ BPSK modulation. In the reference system without the aid of the relay, a recursive convolutional code with generator [Bla03, Chapter 9.1]

$$G(D) = \left(1, \frac{1 + D + D^3 + D^4}{1 + D^2 + D^4}, \frac{1 + D + D^2 + D^3 + D^4}{1 + D^2 + D^4}\right) \tag{3.44}$$

is used at the sources with $k = k_1 = k_2 = 996$ and $n = 3000$, yielding 6000 total uses of the channel. In the system with the relay, the sources use the recursive convolutional code with generator given in (3.43), again with $k = k_1 = k_2 = 996$ information bits and $n = 2000$, so that a fair comparison with the reference system is guaranteed. Throughout, $\gamma_\mathrm{r} = 4\,\mathrm{dB}$, and the relay employs soft demodulation to obtain the soft information. The scalar quantizer used at the relay is one with $N = 3$ quantization regions, for which $H(Z) = 1.257 < \log_2(3)$. Source coding at the relay is therefore beneficial to exploit the additional redundancy in $\boldsymbol{Z}$, and is performed with an arithmetic code [Ris76]. We also employ the corresponding two-dimensional quantizer (shown in Figure 3.7(a)) for comparison. Taking $\gamma_{\mathrm{r,d}} = 3.5\,\mathrm{dB}$ ensures reliable transmission of $\boldsymbol{z}$ with a turbo code of appropriate rate as specified in the Universal Mobile Telecommunication System (UMTS) [Eur01] standard, and 8-phase shift keying (PSK) modulation at the relay. Note that $\gamma_{\mathrm{r,d}} = 3.5\,\mathrm{dB}$ is kept constant for analog and soft bit transmission as well. For comparison, we also show the performance of a scheme with the resource allocation of the system including the relay (i.e., $k = 996$ and $n = 2000$), but in which the information obtained from the relay is not employed for decoding at the destination. The corresponding BER curves are shown in Figure 3.8, from which we observe that the schemes with the relay considerably outperform the reference point-to-point link; furthermore, quantized transmission provides a gain of roughly 1 dB over analog transmission in the waterfall region of the BER curve.

Next, we show a comparison between soft demapping and soft decoding at the relay in Figure 3.9, for symmetric source-relay channels. The system parameters are identical to before, except that now $\gamma_\mathrm{r} = 0\,\mathrm{dB}$ and $\gamma_{\mathrm{r,d}} = 5\,\mathrm{dB}$ (at $\gamma_\mathrm{r} = 0\,\mathrm{dB}$, the rate of the quantizer for soft demapping is $H(Z) = 1.56$, so that a higher SNR is needed on the relay–destination link for reliable transmission). We observe a significant gain from soft decoding at the relay especially if the source–destination SNR is small. For larger values of γ_d, the point-to-point link eventually outperforms the schemes including the relay, which can be explained by noting that the fraction of resources allocated to both sources and the relay is constant and equal to 1/3; an optimization of the resource allocation is beyond the scope of this work. Also note the prominent error floor occurring for the schemes with soft decoding at the relay, which can be explained as follows. Consider the output of the SISO decoder operating on $\boldsymbol{y}_{\mathrm{d},1}$ (cf. Figure 3.2), and assume a high source–destination SNR γ_d and a sufficient number of iterations. Further suppose that the magnitude of $\ell_{1,m}^{(\mathrm{E})}$ is sufficiently large, so that we have

$$\ell_{2,m}^{(\mathrm{A})\prime} \approx \operatorname{sign}\left(\ell_{1,m}^{(\mathrm{E})}\right) \operatorname{sign}\left(\ell_m^{(\mathrm{d,r})}\right) \left|\ell_m^{(\mathrm{d,r})}\right|. \tag{3.45}$$

Therefore, the reliability of $\ell_{2,m}^{(\mathrm{A})\prime}$ is dominated by the reliability of $\ell_m^{(\mathrm{d,r})}$. The crucial point, however, is that $\ell_{2,m}^{(\mathrm{A})\prime}$ is fed into a SISO decoder as *a-priori* information without taking into

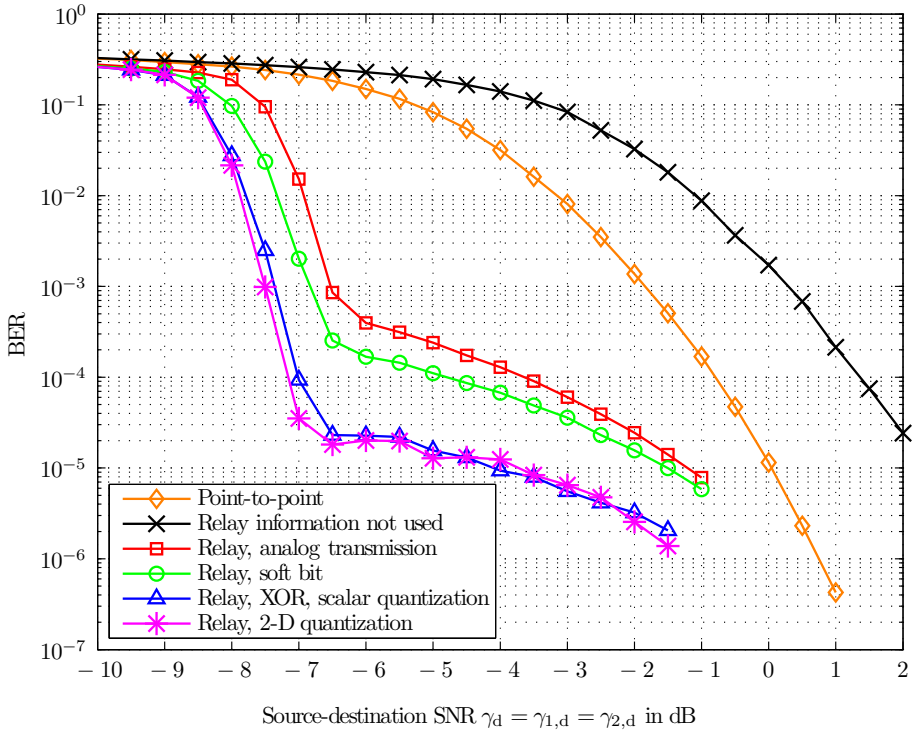


Figure 3.8.: BERs for symmetric source–relay channels, soft demapping, $\gamma_r = 4\,\text{dB}$, $N = 3$.

account that $\ell_m^{(d,r)}$ was obtained by soft decoding at the relay. Note that a similar error floor as in Figure 3.9 was also observed in [ALYM11] for soft decode-and-forward in the relay channel with a single source. As a heuristic to alleviate the error floor while retaining the benefits from soft decoding at the relay at small values of γ_d, one could rescale $\ell_m^{(d,r)}$, $m = 1, 2, \ldots, n$, by multiplication with a factor $a < 1$, which is not pursued further in this work.

In the symmetric scenario, the gain of two-dimensional quantization over scalar quantization is marginal (cf. Figure 3.8); however, the picture changes for asymmetric source–relay links, for which the error rates are shown in Figure 3.10. Again, $k = k_1 = k_2 = 996$, and the sources employ the recursive convolutional code with generator in (3.43) and quaternary phase shift keying (QPSK) modulation, yielding $m_1 = m_2 = 1000$. We assume that $\gamma_{1,r} = \gamma_d + 3\,\text{dB}$ and $\gamma_{2,r} = \gamma_d + 8\,\text{dB}$, and that $\gamma_{r,d} = 18\,\text{dB}$. The relay performs soft demapping of its received signals followed by quantization with $N = 5$ regions and an arithmetic encoder for source coding. Although some of the quantizers used in this scheme have $H(Z)$ very close to the limit of $\log_2(5)$ and hence, the redundancy in $\boldsymbol{Z}$ is small, source coding is used here to obtain a binary representation of $\boldsymbol{z}$. On the relay–destination link, we use the UMTS turbo code of appropriate rate and 256-quadrature amplitude modulation (QAM) with $m_r = 1000$. Note that for $\gamma_d = -4\,\text{dB}$, the quantizer used at the relay is given in Figure 3.7(c). In the point-to-point link, the sources have $k = k_1 = k_2 = 996$ information bits, and use the convolutional code with generator in (3.44) with QPSK modulation, so that

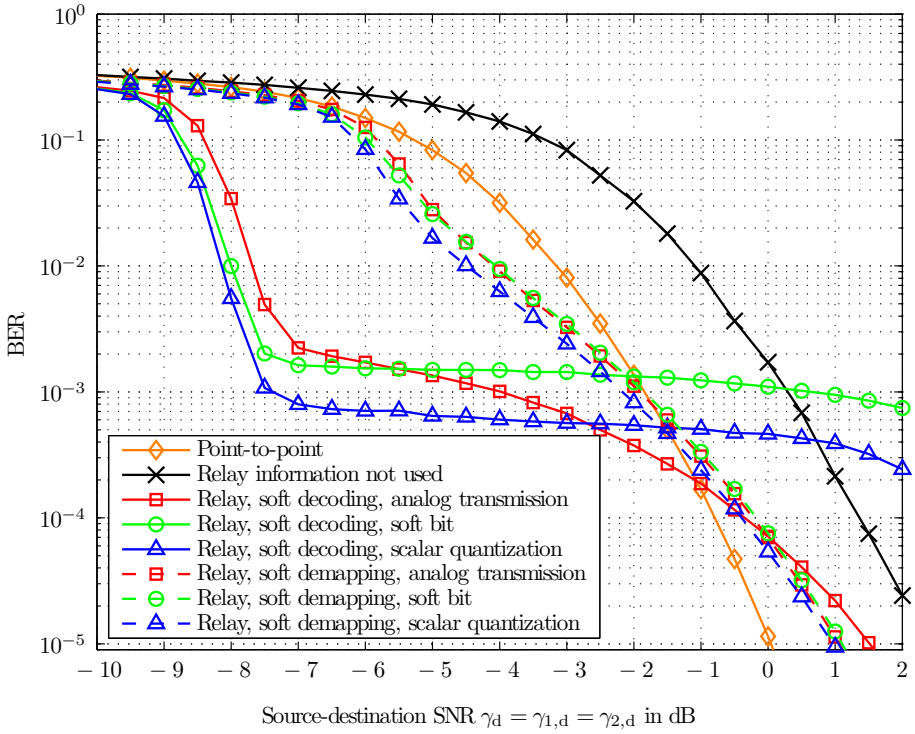


Figure 3.9.: BERs for symmetric source–relay channels, $\gamma_r = 0\,\text{dB}$, $N = 3$.

$m_1 = m_2 = 1500$. As expected, two-dimensional quantization considerably outperforms one-dimensional quantization of the soft information about the XOR-coded bits where the source–relay links have different SNR. The gains compared to the point-to-point link are most pronounced for small source–destination SNR, with user 2 as the stronger user at the relay doing clearly better than user 1.

3.3.2. Block fading channels

We now turn our attention to the performance of the proposed schemes in Rayleigh block fading channels, where we assume that the fading variables $H_{i,r}$, $H_{i,d}$, and $H_{r,d}$ are mutually independent and each distributed according to $\mathcal{CN}(0,1)$. Throughout, we assume a symmetric network, i.e., $d_r = d_{1,r} = d_{2,r}$ and $d_d = d_{1,d} = d_{2,d}$. Further, the relay is placed closer to the destination than to the sources. In particular, we set $\alpha = 3.52$ [HT10] and consider two cases:

- ▷ Case 1: the relay is placed between the sources and the destination, with $d_r = (9/10)d_d$. Consequently, $\rho_r = \rho_d + 1.61\,\text{dB}$, so that the source–relay SNR is only a little larger than the source–destination SNR.

- ▷ Case 2: the relay is placed behind the destination, with $d_r = (3/2)d_d$, so that $\rho_r = \rho_d - 6.20\,\text{dB}$. The relaying scheme turns out to be useful even in this case in which

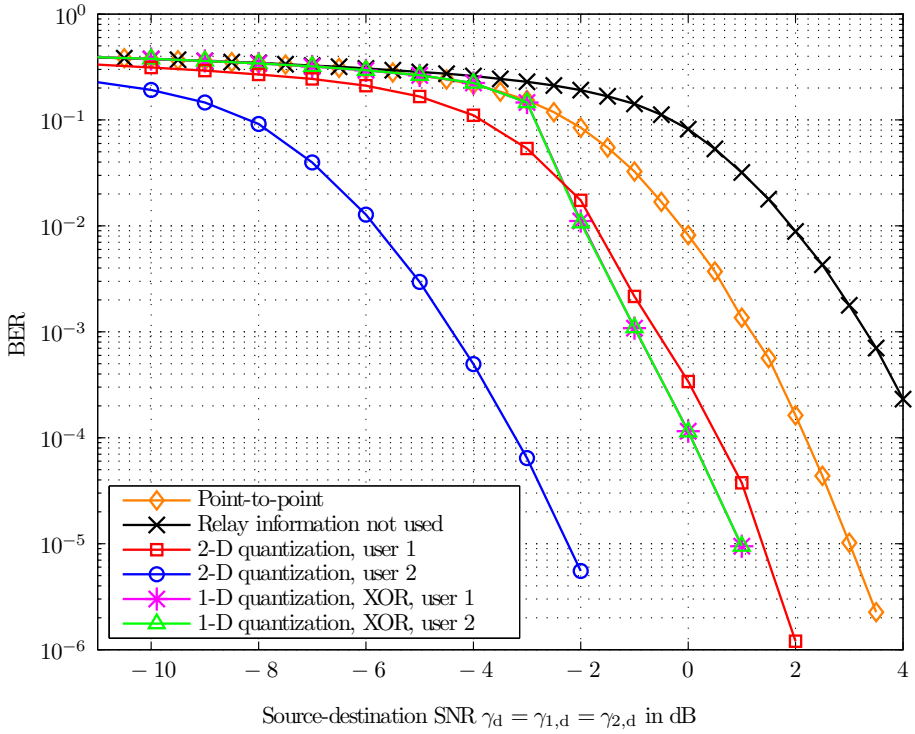


Figure 3.10.: BERs for asymmetric source–relay channels, $\gamma_{1,r} = \gamma_d + 3\,\text{dB}$, $\gamma_{2,r} = \gamma_d + 8\,\text{dB}$, $N = 5$.

the source–relay SNR is smaller than the source–destination SNR. This is in contrast to decode-and-forward schemes, in which the requirement that the relay can decode reliably requires a fairly high SNR on the source–relay links [Hau09].

Due to the proximity of the destination and the relay, we take $\rho_{r,d} = \rho_d + 15\,\text{dB}$.

If the relay is present, the sources have $k = k_1 = k_2 = 2000$ information bits to transmit, but now use the UMTS turbo code [Eur01] of rate 1/2 and BPSK modulation, so that $m_1 = m_2 = 4000$. The relay and the destination share a set $\mathcal{Q}$ of two-dimensional quantizers designed with the framework introduced in Section 3.2.2. Given the realizations of the received sequences $(\boldsymbol{y}_{r,1}, \boldsymbol{y}_{r,2})$ and of $\gamma_{1,r}$ and $\gamma_{2,r}$, the relay selects the proper quantizer in $\mathcal{Q}$, computes the sequence $\boldsymbol{z}$, source encodes that sequence with an arithmetic encoder, and channel encodes using the UMTS turbo code of appropriate rate, yielding the sequence $\boldsymbol{s}_r \in \mathbb{M}_r^{m_r}$ with $m_r = 4000$, where the modulation alphabet $\mathbb{M}_r$ at the relay is chosen to be 16-QAM. In this example, the entropy coding step is useful both to exploit the additional redundancy in the quantized sequence (depending on the actual choice of the quantizer), and to perform the mapping to a binary string efficiently. We compare two sets of quantizers. The first set $\mathcal{Q}_1$ contains one quantizer with $N = 5$ quantization regions for every pair of instantaneous SNR values $\gamma_{1,r}$ and $\gamma_{2,r}$ in the set $\mathcal{S}_1 = \{-9\,\text{dB}, -8\,\text{dB}, -7\,\text{dB}, \ldots, 7\,\text{dB}\}$. For this choice of $\mathcal{S}_1$, there are $|\mathcal{Q}_1| = |\mathcal{S}_1|^2 = 289$ quantizers available at the relay. Consequently, signaling the relay's quantizer choice to

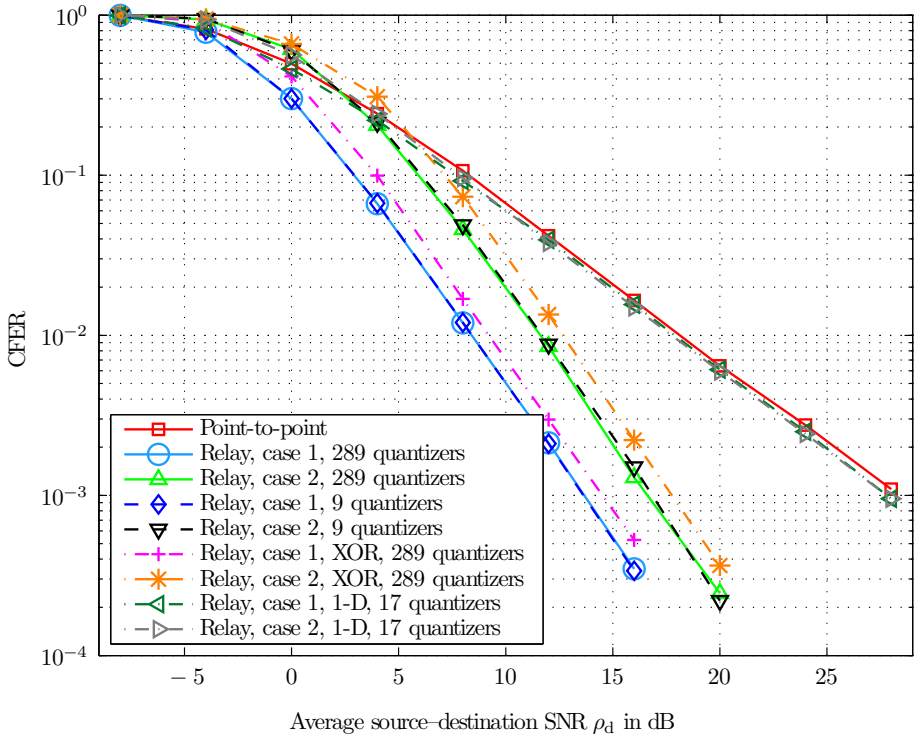


Figure 3.11.: Frame error rates for the MARC and the point-to-point link.

the destination requires at most 9 bits; we assume this signaling to be perfect in the sequel. The other set $\mathcal{Q}_2$ of quantizers consists of one quantizer with $N = 5$ regions for every pair of $\gamma_{1,\mathrm{r}}$ and $\gamma_{2,\mathrm{r}}$ in the set $\mathcal{S}_2 = \{-9\,\mathrm{dB}, 0\,\mathrm{dB}, 6\,\mathrm{dB}\}$, so that $|\mathcal{Q}_2| = 9$, resulting in 4 bits to signal the quantizer choice.

In the reference point-to-point system, the sources employ the UMTS turbo code [Eur01] of rate 1/3 and $k = k_1 = k_2 = 2000$ with BPSK modulation, yielding $m_1 = m_2 = 6000$. Our second reference system includes the relay, so that $k = k_1 = k_2 = 2000$ and $m_1 = m_2 = m_\mathrm{r} = 4000$; here, the relay does perform one-dimensional quantization of $\boldsymbol{\ell}$, the soft information about $\boldsymbol{x} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2'$. In particular, $\mathcal{S}_\mathrm{XOR} = \mathcal{S}_1$, so that the quantizer set shared by the relay and the destination contains $|\mathcal{Q}_\mathrm{XOR}| = 289$ quantizers with $N = 5$ quantization regions each. The third reference system under consideration includes the relay as well, so that $k = k_1 = k_2 = 2000$ and $m_1 = m_2 = m_\mathrm{r} = 4000$. However, instead of two-dimensional quantization, the relay performs one-dimensional quantization of the soft information $\boldsymbol{\ell}_j$ of the user j with the stronger source–relay channel, while excluding the weaker user in the cooperation. For the stronger user j, there is one quantizer with $N = 5$ for each instantaneous SNR value $\gamma_{j,\mathrm{r}}$ in $\mathcal{S}_1$. Including the bit required to signal the index of the stronger user at the relay, at most 6 signaling bits are needed. Note that the destination performs maximum-ratio combining for that user included in the cooperation.

The simulation results in Figure 3.11 show the common frame error rate (CFER) of the reference systems and the system with the relay, for both network geometries. Based on

these curves we observe that the schemes involving the relay achieve second order diversity for both case 1 and case 2, since the CFER decays proportional to ρ_{d}^{-2}. However, if the relay processes $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2')$ directly without going through the intermediate step of computing the likelihood ratios of the XOR of the coded vectors, considerably better performance is obtained, which is because the reliability of the XOR at the relay is undesirably dominated by the weaker source–relay channel – a disadvantage avoided by joint quantization of the soft information at the relay. Particularly, the system with the proposed two-dimensional quantizers at the relay gains more than 10 dB compared to the point-to-point link at relevant CFERs of 10^{-3}. More importantly, simple one-dimensional quantization of the stronger user at the relay is not sufficient to achieve second order diversity. It is also important to note that the scheme involving joint quantization at the relay does not require the set of available quantizers at the relay to be prohibitively large. In fact, we observe that the system with 9 quantizers shared at the relay performs only marginally poorer than the one with a set of 289 quantizers, at considerable lower signaling overhead.

3.4. Discussion

In this chapter, we studied the MARC with two users and noisy source–relay links preventing successful decoding at the relay, so that the operations at the relay are limited to schemes generating and processing soft information. One- and two-dimensional deterministic quantizers were designed for the soft information at the relay based on the notion of relevant information, leading to an improvement over analog transmission methods from the relay. Simulation results further suggest that two-dimensional quantization at the relay outperforms schemes based on network coding the soft values in case of unequal channel quality on the source–relay channels. To perform the quantization, the relay does not require CSI about the source–destination links, a fact especially advantageous in wireless fading channels where this information may not always be available at the relay. In a Rayleigh block fading environment, the relay chooses a suitable quantizer from a fixed set based on the SNR on the incoming links, and forwards its compressed estimate of the received sequences to destination. We observe from numerical results that full diversity order of two can be gained with this scheme. The scheme incurs small delay, since no (soft) decoding is required at the relay node to achieve these gains. Further, we remark that the only overhead created through cooperation is due to the signaling of the quantizer choice at the relay to the destination, since the choice of the quantizer depends on the source–relay SNRs, which are assumed to be unknown at the destination. An efficient low-complexity implementation of two-dimensional quantization might be to approximate the boundaries of the two-dimensional quantizers by hyperplanes, so that the quantization can be found by comparing the vector of likelihoods $(\ell_{1,m}, \ell_{2,m}')$ to be quantized with a number of hyperplanes.

4

Source coding rate allocation in orthogonal compress-and-forward relay networks

In Chapter 3, we designed scalar and two-dimensional *symbol-by-symbol* quantizers for the relay node of a MARC with two sources, where the emphasis was on practical schemes of low complexity. In this chapter, we generalize the work of Chapter 3 by considering an arbitrary number of users transmitting information via a single relay, and by employing vector quantization (with or without binning) at the relay. The goal is to optimize the quantization at the relay to maximize the achievable sum-rate.

Capacity results for the relay channel with full- and half-duplex relays go back to [vdM77] and [CEG79]; more recent work includes [LTW04, HZ05] and [SKM04a, KGG05] for the MARC. In this chapter, we focus on the orthogonal MARC with M sources and compress-and-forward (CF) [CEG79] at the relay, where the relay compresses its received values before forwarding the estimates to the destination. CF methods are useful when the relay cannot decode the source messages reliably [SKM04b, SKM04c], e.g., if the relay is geographically placed closer to the destination than to the source(s). If the relay's power resources are limited, we show that determining which users to include in the cooperation, and at which source coding rate, is critical to achieve a good sum-rate. Intuitively, the relay should spend source coding rate only for those users with a sufficiently strong signal at the relay and a weak direct link to the destination, allowing them to benefit from cooperation. Specifically, for CF, we obtain the following results:

- ▷ A water-filling source coding rate assignment at the relay maximizes the achievable sum-rate for Gaussian modulation at the sources and Gaussian channels. This result

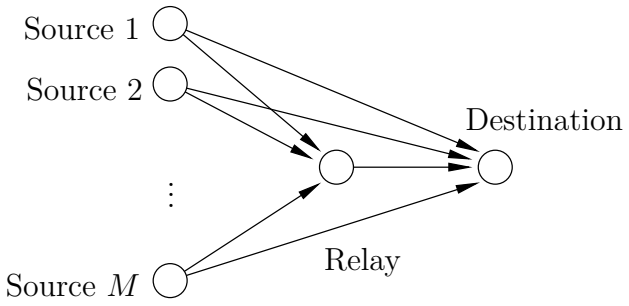


Figure 4.1.: The multiple-access relay channel with M sources.

is similar to the one in [SY08] for the powers allocated by the relay for cooperation with the different sources.

- ▷ For general modulation alphabets at the sources and DMCs, we formulate the sum-rate maximization problem as a tradeoff between mutual information and source coding rate (cf. Section 2.2). We adapt the iterative information bottleneck algorithm so that the information-rate tradeoff can be computed for each user individually. Given this tradeoff, we show that the sum-rate maximization problem for M users is a convex optimization problem, which we solve numerically with a cutting-plane algorithm.

- ▷ We compare the CF rate with the rate of a scheme in which the relay omits the binning step, a strategy known as noisy network coding (NNC) [LKEGC11]. The sum-rate optimal CF rate allocation is demonstrated to be sum-rate optimal for NNC as well.

In Section 4.1 of this chapter, we summarize the system model and the cooperation protocol for CF. Focusing on CF, the rate allocation problem is solved for Gaussian modulation and Gaussian channels in Section 4.2, and for general finite inputs and DMCs in Section 4.3. NNC is considered in Section 4.4, and Section 4.5 concludes the chapter.

4.1. System model

4.1.1. Channel model

The system is shown in Figure 4.1. Like in Chapter 3, the relay is limited by a half-duplex constraint, i.e., the relay cannot receive and transmit simultaneously in the same frequency band. Further, the transmissions from the sources and the relay are assumed to be orthogonal either in frequency or in time. Without loss of generality, we assume the orthogonality to be guaranteed by time division, so that the first M time slots of length $\alpha_i n$ each, $\alpha_i > 0$, $\sum_{i=1}^{M} \alpha_i < 1$, are assigned to the sources, and the last time slot of length $(1 - \sum_{i=1}^{M} \alpha_i)n = \bar{\alpha} n$ to the relay. Let $\mathcal{X}_i$ and $\mathcal{X}_\mathrm{r}$ be the modulation alphabets at

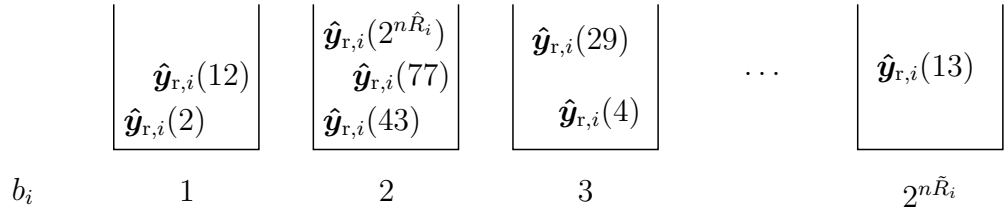


Figure 4.2.: Random binning.

source i and at the relay, respectively, so that $\boldsymbol{X}_i \in \mathcal{X}_i^{\alpha_i n}$ is the transmitted vector from an i.i.d. codebook of rate R_i at source i, and $\boldsymbol{X}_{\mathrm{r}} \in \mathcal{X}_{\mathrm{r}}^{\bar{\alpha} n}$ is the transmitted vector from an i.i.d. codebook of rate R_{r} at the relay. Then, the received signals at the relay and at the destination are given by $\boldsymbol{Y}_{\mathrm{r},i} \in \mathcal{Y}_{\mathrm{r},i}^{\alpha_i n}$, $\boldsymbol{Y}_{\mathrm{d},i} \in \mathcal{Y}_{\mathrm{d},i}^{\alpha_i n}$, and $\boldsymbol{Y}_{\mathrm{d,r}} \in \mathcal{Y}_{\mathrm{d,r}}^{\bar{\alpha} n}$, respectively; the orthogonality assumption requires the channel transition probability to factor as

$$\begin{aligned} &P_{Y_{\mathrm{r},1}\ldots Y_{\mathrm{r},M} Y_{\mathrm{d},1}\ldots Y_{\mathrm{d},M} Y_{\mathrm{d,r}}|X_1\ldots X_M X_{\mathrm{r}}}(y_{\mathrm{r},1},\ldots,y_{\mathrm{r},M},y_{\mathrm{d},1},\ldots,y_{\mathrm{d},M},y_{\mathrm{d,r}}|x_1,\ldots,x_M,x_{\mathrm{r}}) \\ &\quad = P_{Y_{\mathrm{d,r}}|X_{\mathrm{r}}}(y_{\mathrm{d,r}}|x_{\mathrm{r}}) \prod_{i=1}^{M} P_{Y_{\mathrm{r},i}Y_{\mathrm{d},i}|X_i}(y_{\mathrm{r},i},y_{\mathrm{d},i}|x_i) \qquad (4.1) \\ &\quad = P_{Y_{\mathrm{d,r}}|X_{\mathrm{r}}}(y_{\mathrm{d,r}}|x_{\mathrm{r}}) \prod_{i=1}^{M} P_{Y_{\mathrm{r},i}|X_i}(y_{\mathrm{r},i}|x_i) P_{Y_{\mathrm{d},i}|X_i}(y_{\mathrm{d},i}|x_i). \qquad (4.2) \end{aligned}$$

4.1.2. Cooperation protocol and achievable rates

The strategy at the relay considered in Sections 4.2 and 4.3 of this chapter is CF [CEG79, Section VI]. In this protocol, the relay performs Wyner–Ziv coding [WZ76, Wyn78] to compress its received sequences exploiting the side information available at the destination for reconstruction. In our model, source coding is performed *separately* for each received vector $\boldsymbol{Y}_{\mathrm{r},i}$, however, the allocation of source coding rate to user i at the relay is done *jointly*. Specifically, the relay compresses the vector $\boldsymbol{Y}_{\mathrm{r},i}$ to an estimate $\hat{\boldsymbol{Y}}_{\mathrm{r},i}(s_i) \in \hat{\mathcal{Y}}_{\mathrm{r},i}^{\alpha_i n}$, $s_i \in \{1,2,\ldots,2^{n\hat{R}_i}\}$, where each $\hat{\boldsymbol{Y}}_{\mathrm{r},i}$ is generated i.i.d. according to the distribution

$$P_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}) = \sum_{y_{\mathrm{r},i}} P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) P_{Y_{\mathrm{r},i}}(y_{\mathrm{r},i}). \qquad (4.3)$$

Using a random uniform distribution of the indices $s_i \in \{1,2,\ldots,2^{n\hat{R}_i}\}$ among $2^{n\tilde{R}_i}$ bins, $\tilde{R}_i \leq \hat{R}_i$, the relay determines the index of the bin $b_i(s_i), b_i \in \{1,2,\ldots,2^{n\tilde{R}_i}\}$, to which s_i belongs, cf. Figure 4.2. Here, $\tilde{R}_i$ is the source coding rate with side information for user i. Then, the relay sends a corresponding codeword $\boldsymbol{X}_{\mathrm{r}}(b_1(s_1), b_2(s_2), \ldots, b_M(s_M))$ from a codebook with $2^{nR_{\mathrm{r}}}$ elements, where $R_{\mathrm{r}} = \sum_{i=1}^{M} \tilde{R}_i$. After successful decoding of $\boldsymbol{X}_{\mathrm{r}}$ from $\boldsymbol{Y}_{\mathrm{d,r}}$, the destination uses the side information $\boldsymbol{Y}_{\mathrm{d},i}$ and $b_i(s_i)$ to resolve the remaining uncertainty within bin $b_i(s_i)$ about s_i, yielding the estimate $\hat{\boldsymbol{Y}}_{\mathrm{r},i}$. Then, $\boldsymbol{Y}_{\mathrm{d},i}$ and $\hat{\boldsymbol{Y}}_{\mathrm{r},i}$ are used to jointly decode the message of user i.

Achievable rates for the relay channel with CF are derived in [CEG79, Theorem 6], [KGG05]. We specialize these rates to the M-user orthogonal MARC with CF in the following proposition.

Proposition 4.1. The rate vector $\boldsymbol{R} = [R_1, R_2, \ldots, R_M]^{\mathrm{T}}$ is achievable in the M-user orthogonal MARC with CF at the relay if

$$R_i < \alpha_i I(X_i; Y_{\mathrm{d},i}, \hat{Y}_{\mathrm{r},i}) = \alpha_i \left[I(X_i; Y_{\mathrm{d},i}) + I(X_i; \hat{Y}_{\mathrm{r},i} | Y_{\mathrm{d},i}) \right], \quad i = 1, 2, \ldots, M, \tag{4.4}$$

and

$$\sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i} | Y_{\mathrm{d},i}) \leq \bar{\alpha} I(X_{\mathrm{r}}; Y_{\mathrm{d,r}}). \tag{4.5}$$

We denote the set of rate vectors $\boldsymbol{R}$ satisfying (4.4) and (4.5) by $\mathcal{R}_{\mathrm{CF}}$.

Proof. By [WZ76, CEG79], $\alpha_i I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i} | Y_{\mathrm{d},i})$ is the source coding rate for user i at the relay with side information at the destination. Satisfying (4.5) ensures that the total source coding rate for all users at the relay does not exceed the available rate $\bar{\alpha} I(X_{\mathrm{r}}; Y_{\mathrm{d,r}})$ on the relay–destination link, so that $\hat{\boldsymbol{Y}}_{\mathrm{r},i}$, $i = 1, 2, \ldots, M$, can be reconstructed reliably at the destination. Then, the destination has $\boldsymbol{Y}_{\mathrm{d},i}$ and $\hat{\boldsymbol{Y}}_{\mathrm{r},i}$ available to decode the message from user i, so that $\alpha_i I(X_i; Y_{\mathrm{d},i}, \hat{Y}_{\mathrm{r},i})$ is the rate bound for user i. ■

4.2. Optimal allocation of source coding rate for Gaussian modulation

In this section, we consider Gaussian channels and Gaussian modulation at the sources and at the relay, i.e., $\boldsymbol{X}_i$ is from a Gaussian codebook of rate R_i with power P_i, and $\boldsymbol{X}_{\mathrm{r}}$ is from a Gaussian codebook of rate R_{r} with power P_{r}. The received signals $\boldsymbol{Y}_{\mathrm{r},i} \in \mathbb{C}^{\alpha_i n}$, $\boldsymbol{Y}_{\mathrm{d},i} \in \mathbb{C}^{\alpha_i n}$, and $\boldsymbol{Y}_{\mathrm{d,r}} \in \mathbb{C}^{\bar{\alpha} n}$ at the relay and at the destination are given by

$$\boldsymbol{Y}_{\mathrm{r},i} = H_{i,\mathrm{r}} \boldsymbol{X}_i + \boldsymbol{N}_{\mathrm{r},i} \tag{4.6}$$

$$\boldsymbol{Y}_{\mathrm{d},i} = H_{i,\mathrm{d}} \boldsymbol{X}_i + \boldsymbol{N}_{\mathrm{d},i} \tag{4.7}$$

in time slot i, and

$$\boldsymbol{Y}_{\mathrm{d,r}} = H_{\mathrm{r,d}} \boldsymbol{X}_{\mathrm{r}} + \boldsymbol{N}_{\mathrm{d,r}} \tag{4.8}$$

in the last time slot, where $H_{i,\mathrm{r}}$, $H_{i,\mathrm{d}}$, and $H_{\mathrm{r,d}}$ are complex channel fading coefficients satisfying $\mathrm{E}[|H_{i,\mathrm{r}}|^2] = \mathrm{E}[|H_{i,\mathrm{d}}|^2] = \mathrm{E}[|H_{\mathrm{r,d}}|^2] = 1$, and the additive noise vectors $\boldsymbol{N}_{\mathrm{r},i}$, $\boldsymbol{N}_{\mathrm{d},i}$, and $\boldsymbol{N}_{\mathrm{d,r}}$ are independent proper complex Gaussian with zero mean and unit variance. We assume that the receivers know the instantaneous SNR values $\gamma_{i,\mathrm{r}} = |h_{i,\mathrm{r}}|^2 P_i$, $\gamma_{i,\mathrm{d}} = |h_{i,\mathrm{d}}|^2 P_i$, and $\gamma_{\mathrm{r,d}} = |h_{\mathrm{r,d}}|^2 P_{\mathrm{r}}$ of their channels. Additionally, the relay knows the instantaneous SNRs of all source–destination channels and of the relay–destination channel.

The rate region of Proposition 4.1 is specialized to the Gaussian setting as follows. For Gaussian codebooks at the sources and Gaussian channels, we choose $\hat{Y}_{\mathrm{r},i} = Y_{\mathrm{r},i} + \bar{N}_{\mathrm{r},i}$, where $\bar{N}_{\mathrm{r},i} \sim \mathcal{CN}(0, \sigma_i^2)$ is independent of $Y_{\mathrm{r},i}$ [HZ05]. We thus have

$$I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = h(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) - h(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}, Y_{\mathrm{d},i}) \tag{4.9}$$

$$= h(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) - h(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}). \tag{4.10}$$

The conditional variance of $\hat{Y}_{\mathrm{r},i}$ conditioned on $Y_{\mathrm{d},i} = y_{\mathrm{d},i}$ is

$$\mathrm{Var}(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i} = y_{\mathrm{d},i}) = \mathrm{Var}(h_{i,\mathrm{r}}X_i + N_{\mathrm{r},i} + \bar{N}_{\mathrm{r},i}|Y_{\mathrm{d},i} = y_{\mathrm{d},i}) \tag{4.11}$$

$$= |h_{i,\mathrm{r}}|^2\mathrm{Var}(X_i|Y_{\mathrm{d},i} = y_{\mathrm{d},i}) + \mathrm{Var}(N_{\mathrm{r},i}|Y_{\mathrm{d},i} = y_{\mathrm{d},i}) + \mathrm{Var}(\bar{N}_{\mathrm{r},i}|Y_{\mathrm{d},i} = y_{\mathrm{d},i})$$

$$= |h_{i,\mathrm{r}}|^2 \frac{P_i \frac{1}{|h_{i,\mathrm{d}}|^2}}{P_i + \frac{1}{|h_{i,\mathrm{d}}|^2}} + 1 + \sigma_i^2, \tag{4.12}$$

and consequently, we have

$$h(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = \log_2\left(\pi e\left(1 + \sigma_i^2 + \frac{\gamma_{i,\mathrm{r}}}{1+\gamma_{i,\mathrm{d}}}\right)\right). \tag{4.13}$$

Together with

$$h(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}) = \log_2(\pi e\sigma_i^2), \tag{4.14}$$

(4.13) yields

$$I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = \log_2\left(1 + \frac{1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}}}{\sigma_i^2(1+\gamma_{i,\mathrm{d}})}\right), \tag{4.15}$$

and setting $\tilde{R}_i = \alpha_i I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i})$ gives

$$\sigma_i^2 = \frac{1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}}}{(1+\gamma_{i,\mathrm{d}})(2^{\tilde{R}_i/\alpha_i}-1)}. \tag{4.16}$$

Therefore (cf. [HZ05, Proposition 3] for the relay channel with a single source), the rate vector $\boldsymbol{R}$ is achievable if

$$R_i < \alpha_i \log_2\left(1 + \gamma_{i,\mathrm{d}} + \frac{\gamma_{i,\mathrm{r}}}{1+\sigma_i^2}\right) \tag{4.17}$$

$$= \alpha_i \log_2\left(1+\gamma_{i,\mathrm{d}}\right) + \alpha_i \log_2\left(\frac{2^{\tilde{R}_i/\alpha_i}(1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}})}{2^{\tilde{R}_i/\alpha_i}(1+\gamma_{i,\mathrm{d}})+\gamma_{i,\mathrm{r}}}\right) \tag{4.18}$$

and

$$\sum_{i=1}^{M} \tilde{R}_i \leq \bar{\alpha} I(X_{\mathrm{r}}; Y_{\mathrm{d,r}}) = \bar{\alpha} \log_2(1+\gamma_{\mathrm{r,d}}). \tag{4.19}$$

Defining $\tilde{\boldsymbol{R}} = [\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_M]^{\mathrm{T}}$, the rate allocation at the relay maximizing the achievable sum-rate is the solution of the optimization problem

$$\begin{aligned} \max_{\tilde{\boldsymbol{R}}} & \sum_{i=1}^{M} \alpha_i \log_2 \left(\frac{2^{\tilde{R}_i/\alpha_i}(1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}})}{2^{\tilde{R}_i/\alpha_i}(1+\gamma_{i,\mathrm{d}})+\gamma_{i,\mathrm{r}}} \right) \\ \text{s.t.} \quad & \sum_{i=1}^{M} \tilde{R}_i \leq \bar{\alpha} \log_2(1+\gamma_{\mathrm{r,d}}), \quad \tilde{R}_i \geq 0,\ i=1,\ldots,M. \end{aligned} \tag{4.20}$$

The solution to (4.20) is given in the following theorem.

Theorem 4.2. Let α_i, $h_{i,\mathrm{r}}$, $h_{i,\mathrm{d}}$, $h_{\mathrm{r,d}}$, P_i, and P_{r} be fixed, and define

$$\xi_i \triangleq \frac{1+\gamma_{i,\mathrm{d}}}{\gamma_{i,\mathrm{r}}}. \tag{4.21}$$

The sum-rate optimal source coding rate assignment $\tilde{\boldsymbol{R}}^*$ at the relay node of an orthogonal CF MARC with M users and Gaussian modulation satisfies

$$\tilde{R}_i^* = \begin{cases} \alpha_i \log_2 \left(\tau \dfrac{\gamma_{i,\mathrm{r}}}{1+\gamma_{i,\mathrm{d}}} \right) & \text{if } \tau \geq \xi_i \\ 0 & \text{if } \tau < \xi_i, \end{cases} \tag{4.22}$$

where τ is chosen such that

$$\sum_{i=1}^{M} \tilde{R}_i^* = \bar{\alpha} \log_2(1+\gamma_{\mathrm{r,d}}). \tag{4.23}$$

Proof. Problem (4.20) is a convex program. Using Lagrange multipliers $\lambda \geq 0$ and $\nu_i \geq 0$, $i = 1, \ldots, M$, we construct the functional

$$\begin{aligned} J(\tilde{\boldsymbol{R}}, \lambda, \boldsymbol{\nu}) = & -\sum_{i=1}^{M} \alpha_i \log_2 \left(\frac{2^{\tilde{R}_i/\alpha_i}(1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}})}{2^{\tilde{R}_i/\alpha_i}(1+\gamma_{i,\mathrm{d}})+\gamma_{i,\mathrm{r}}} \right) \\ & +\lambda \left(\sum_{i=1}^{M} \tilde{R}_i - \bar{\alpha} \log_2(1+\gamma_{\mathrm{r,d}}) \right) - \sum_{i=1}^{M} \nu_i \tilde{R}_i, \end{aligned} \tag{4.24}$$

and differentiation with respect to $\tilde{\boldsymbol{R}}$ and setting to zero gives

$$\frac{\partial J}{\partial \tilde{R}_i} = -\frac{\gamma_{i,\mathrm{r}}}{2^{\tilde{R}_i/\alpha_i}(1+\gamma_{i,\mathrm{d}})+\gamma_{i,\mathrm{r}}} + \lambda - \nu_i = 0, \tag{4.25}$$

which implies

$$\lambda \geq \frac{\gamma_{i,\mathrm{r}}}{2^{\tilde{R}_i^*/\alpha_i}(1+\gamma_{i,\mathrm{d}})+\gamma_{i,\mathrm{r}}}, \quad i=1,\ldots,M. \tag{4.26}$$

Further, the KKT conditions [BV04, Chapter 5.5.3] require that, for $i=1,\ldots,M$,

$$\nu_i \tilde{R}_i^* = \left(\lambda - \frac{\gamma_{i,\mathrm{r}}}{2^{\tilde{R}_i^*/\alpha_i}(1+\gamma_{i,\mathrm{d}})+\gamma_{i,\mathrm{r}}}\right)\tilde{R}_i^* = 0. \tag{4.27}$$

Now, if $\lambda < \gamma_{i,\mathrm{r}}/(1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}})$, (4.26) can hold only if $\tilde{R}_i^* > 0$, so that (4.27) implies

$$\tilde{R}_i^* = \alpha_i \log_2\left(\frac{1-\lambda}{\lambda}\frac{\gamma_{i,\mathrm{r}}}{1+\gamma_{i,\mathrm{d}}}\right). \tag{4.28}$$

Alternatively, if $\lambda \geq \gamma_{i,\mathrm{r}}/(1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}})$, then $\tilde{R}_i^* = 0$ by (4.26). Setting $\tau = (1-\lambda)/\lambda$, we find that

$$\tilde{R}_i^* = \left(\alpha_i \log_2\left(\tau\frac{\gamma_{i,\mathrm{r}}}{1+\gamma_{i,\mathrm{d}}}\right)\right)^+ \tag{4.29}$$

is the optimal assignment maximizing the achievable sum-rate, where $(x)^+$ denotes the positive part of x, and τ is chosen so that

$$\sum_{i=1}^{M} \tilde{R}_i^* = \bar{\alpha}\log_2(1+\gamma_{\mathrm{r,d}}). \tag{4.30}$$

■

The solution in the theorem is a water-filling solution. Given the available rate on the relay–destination link, the constant τ is chosen such that $\sum_{i=1}^{M}\tilde{R}_i = \bar{\alpha}I(X_{\mathrm{r}};Y_{\mathrm{d,r}})$. Further, user i is included in the source coding at the relay only if $\tau \geq \xi_i$; no resources of the relay are assigned to users for which $\tau < \xi_i$. An example is illustrated in Figure 4.3. Informally stated, by examining ξ_i, source coding rate is allocated to those users whose source–relay SNR $\gamma_{i,\mathrm{r}}$ is large, and whose source–destination SNR $\gamma_{i,\mathrm{d}}$ is small.

4.3. Rate allocation for arbitrary modulation alphabets and discrete memoryless channels

In this section, we study the rate allocation problem at the relay for general finite modulation alphabets $\mathcal{X}_i$ and $\mathcal{X}_{\mathrm{r}}$ at the sources and at the relay, respectively, and arbitrary DMCs with finite output alphabets $\mathcal{Y}_{\mathrm{r},i}$ and $\mathcal{Y}_{\mathrm{d},i}$. The distribution $P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$ specifying the mapping from $Y_{\mathrm{r},i} \in \mathcal{Y}_{\mathrm{r},i}$ to $\hat{Y}_{\mathrm{r},i} \in \hat{\mathcal{Y}}_{\mathrm{r},i}$ needs to be chosen carefully for each user to obtain a small source coding rate $\tilde{R}_i = \alpha_i I(Y_{\mathrm{r},i};\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i})$ and a large mutual information $I(X_i;\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i})$.

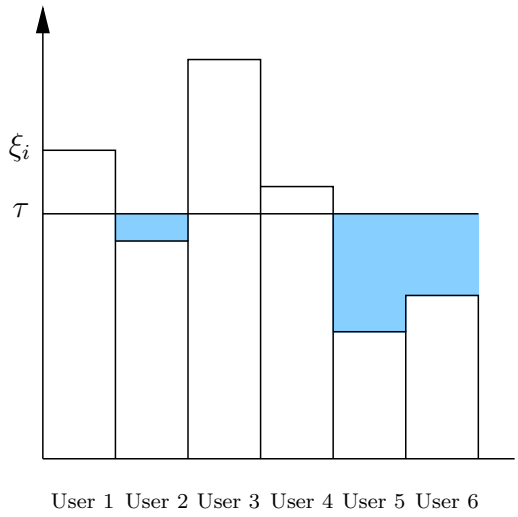


Figure 4.3.: Water-filling for $M = 6$ users.

We formulate the sum-rate maximization problem as

$$\begin{aligned} \max_{P_{\hat{Y}_{\mathrm{r},1}|Y_{\mathrm{r},1}}, P_{\hat{Y}_{\mathrm{r},2}|Y_{\mathrm{r},2}}, \ldots, P_{\hat{Y}_{\mathrm{r},M}|Y_{\mathrm{r},M}}} & \sum_{i=1}^{M} \alpha_i I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) \\ \text{s.t.} \quad & \sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) \leq \bar{\alpha} I(X_{\mathrm{r}}; Y_{\mathrm{d,r}}), \end{aligned} \tag{4.31}$$

where we regard P_{X_i}, $i \in \{1, 2, \ldots, M\}$, and $P_{X_{\mathrm{r}}}$ as fixed. Essentially, this is an optimization over a weighted sum of mutual information expressions, subject to a sum-rate constraint. To solve (4.31), we first study the tradeoff between relevant information and required source coding rate for each user separately. Based on that tradeoff we will then introduce a cutting-plane algorithm for computing an optimal solution to (4.31).

4.3.1. The information-rate tradeoff

The tradeoff between rate and mutual information presented in Chapter 2 can be extended to the tradeoff between $I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i})$ and $I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i})$. To that end, we define, for the i-th user and a fixed joint distribution of $(X_i, Y_{\mathrm{d},i}, Y_{\mathrm{r},i})$, the function

$$I_i(r_i) \triangleq \max_{P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}} I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) \quad \text{s.t.} \quad I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) \leq r_i, \tag{4.32}$$

where $0 \leq r_i \leq H(Y_{\mathrm{r},i}|Y_{\mathrm{d},i})$. Similar to Chapter 2, a Markov condition $X_i \leftrightarrow Y_{\mathrm{r},i} \leftrightarrow \hat{Y}_{\mathrm{r},i}$ is introduced by restricting the mapping to $\hat{Y}_{\mathrm{r},i}$ to be of the form $P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$. The function $I_i(r_i)$ is characterized by the following two theorems.

Theorem 4.3. The function $I_i(r_i)$ is concave and monotonically non-decreasing in the domain $0 \leq r_i \leq H(Y_{\mathrm{r},i}|Y_{\mathrm{d},i})$, and every point of $I_i(r_i)$ can be achieved with $\hat{Y}_{\mathrm{r},i}$ taking at most $|\mathcal{Y}_{\mathrm{r},i}| + 1$ values.

The proof of the theorem is based on applying the same techniques as in [WW75, Section II] to $H(X_i|\hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i})$ and $H(Y_{\mathrm{r},i}|\hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i})$. We state the proof in Appendix C.1.

Theorem 4.4. The function $I_i(r_i)$ further has the following properties:

a) We have $I_i(r_i = 0) = 0$ and $I_i(r_i = H(Y_{\mathrm{r},i}|Y_{\mathrm{d},i})) = I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i})$.

b) If $I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i}) = 0$, then $I_i(r_i) = 0$ for all $0 \leq r_i \leq H(Y_{\mathrm{r},i}|Y_{\mathrm{d},i})$.

c) Define

$$r_{i,\max} = \inf r_i \quad \text{s.t.} \quad I_i(r_i) = I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i}). \tag{4.33}$$

If $I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i}) > 0$, then $I_i(r_i)$ is strictly increasing for $0 \leq r_i \leq r_{i,\max}$.

Proof. See Appendix C.2. ■

4.3.2. Rate allocation for M users

Using the definition of $I_i(r_i)$, we can restate the sum-rate maximization problem given in (4.31) as

$$\max_{\boldsymbol{r} \geq \mathbf{0}} \ \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{I}(\boldsymbol{r}) \quad \text{s.t.} \quad \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{r} \leq I_{\mathrm{r,d}}, \tag{4.34}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_M]^{\mathrm{T}}$, $\boldsymbol{r} = [r_1, \ldots, r_M]^{\mathrm{T}}$, $\boldsymbol{I}(\boldsymbol{r}) = [I_1(r_1), \ldots, I_M(r_M)]^{\mathrm{T}}$, and $I_{\mathrm{r,d}} = \bar{\alpha} I(X_{\mathrm{r}}; Y_{\mathrm{d,r}})$. Since I_i is a concave function in r_i and the constraints are linear inequality constraints, Problem (4.34) is a convex optimization problem. If a method is available to compute the value and a subgradient of I_i at r_i, Problem (4.34) can be solved by standard convex optimization methods, such as cutting-plane methods [BSS06]. However, to evaluate I_i at r_i, we need to solve Problem (4.32).

4.3.3. Evaluation of the function I_i

Except for the conditioning on $Y_{\mathrm{d},i}$, the optimization in (4.32) defining I_i was studied in [TPB99] in the context of the information bottleneck method, and Tishby *et al.* also provided an iterative algorithm [TPB99] to solve the corresponding optimization, cf. Chapter 2. We will derive a version of that algorithm adapted to Problem (4.32) encountered here.

Throughout, suppose that $I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i}) > 0$, since otherwise $I_i(r_i) = 0$ for all $0 \leq r_i \leq H(Y_{\mathrm{r},i}|Y_{\mathrm{d},i})$ due to Theorem 4.4. Therefore, $I_i(r_i)$ is concave and strictly increasing, and the tangent of slope $1/\beta$, $\beta > 0$, through the point $I_i(r_i(\beta))$ has axis intercept with the ordinate of $I_i(r_i(\beta)) - (1/\beta) r_i(\beta)$. Moreover, we have

$$I_i(r_i(\beta)) - \frac{1}{\beta} r_i(\beta) = \max_{P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}} \left\{ I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) - \frac{1}{\beta} I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) \right\} \tag{4.35}$$

$$= \frac{1}{\beta} \min_{P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}} \left\{ I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) - \beta I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) \right\}. \tag{4.36}$$

By the chain rule of mutual information [CT06] and since $X_i \leftrightarrow Y_{\mathrm{r},i} \leftrightarrow \hat{Y}_{\mathrm{r},i}$ forms a Markov chain, we have

$$I(X_i; Y_{\mathrm{r},i}, \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) + I(X_i; Y_{\mathrm{r},i}|\hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}) \tag{4.37}$$

$$= I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i}) + \underbrace{I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}, Y_{\mathrm{d},i})}_{=0}. \tag{4.38}$$

Rewriting (4.37) and (4.38) yields

$$I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i}) - I(X_i; Y_{\mathrm{r},i}|\hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}), \tag{4.39}$$

where $I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i})$ does not dependent on $P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$. Similarly, by the chain rule of mutual information and since $Y_{\mathrm{d},i} \leftrightarrow Y_{\mathrm{r},i} \leftrightarrow \hat{Y}_{\mathrm{r},i}$ forms a Markov chain, we obtain the identity

$$I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}) - I(Y_{\mathrm{d},i}; \hat{Y}_{\mathrm{r},i}) = I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}) + H(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) - H(\hat{Y}_{\mathrm{r},i}). \tag{4.40}$$

By inserting (4.39) and (4.40) into (4.36), we have

$$I_i(r_i(\beta)) - \frac{1}{\beta} r_i(\beta) \tag{4.41}$$

$$= \frac{1}{\beta} \min_{P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}} \left\{ I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}) + \beta I(X_i; Y_{\mathrm{r},i}|\hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}) + H(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) - H(\hat{Y}_{\mathrm{r},i}) \right\} - I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i}).$$

To see that (4.41) is in a form which allows the application of an alternating minimization algorithm in the spirit of the information bottleneck iterative algorithm given in Algorithm 2.2, we define the function

$$d(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i}) \triangleq \beta \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) D_{\mathrm{KL}} \left(P_{X_i|Y_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|y_{\mathrm{r},i}, y_{\mathrm{d},i}) \middle\| P_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \right)$$
$$- \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) \log_2 \left(P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i}) \right) + \log_2 \left(P_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}) \right). \tag{4.42}$$

By inserting (4.42) into the minimization of (4.41), we observe that

$$\min_{P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}} \left\{ I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}) + \beta I(X_i; Y_{\mathrm{r},i}|\hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}) + H(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) - H(\hat{Y}_{\mathrm{r},i}) \right\}$$
$$= \min_{P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}} \left\{ I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}) + \mathrm{E}\left[d(Y_{\mathrm{r},i}, \hat{Y}_{\mathrm{r},i}) \right] \right\}. \tag{4.43}$$

Equation (4.43) is now in a similar form as (2.37), and it is straightforward to modify the corresponding alternating minimization algorithm. Beginning with an initial mapping

$P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$, we obtain the distributions

$$P^{(0)}_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}) = \sum_{y_{\mathrm{r},i}} P_{Y_{\mathrm{r},i}}(y_{\mathrm{r},i}) P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) \tag{4.44}$$

$$P^{(0)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) = \sum_{y_{\mathrm{r},i}} P_{Y_{\mathrm{r},i}Y_{\mathrm{d},i}}(y_{\mathrm{r},i}, y_{\mathrm{d},i}) P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) \tag{4.45}$$

$$P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i}) = \frac{1}{P_{Y_{\mathrm{d},i}}(y_{\mathrm{d},i})} P^{(0)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \tag{4.46}$$

$$P^{(0)}_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(x_i|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) = \frac{1}{P^{(0)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})} \sum_{y_{\mathrm{r},i}} P_{X_i Y_{\mathrm{r},i} Y_{\mathrm{d},i}}(x_i, y_{\mathrm{r},i}, y_{\mathrm{d},i}) P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}), \tag{4.47}$$

assuming that $P_{Y_{\mathrm{d},i}}(y_{\mathrm{d},i}) \neq 0$ and $P^{(0)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \neq 0$. Using (4.44) to (4.47), we compute

$$\begin{aligned} d^{(0)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i}) = \beta \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) D_{\mathrm{KL}} \left(P_{X_i|Y_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|y_{\mathrm{r},i}, y_{\mathrm{d},i}) \middle\| P^{(0)}_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \right) \\ - \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) \log_2 \left(P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i}) \right) + \log_2 \left(P^{(0)}_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}) \right). \end{aligned} \tag{4.48}$$

Given $d^{(0)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i})$ and $\beta > 0$, the next mapping is obtained by computing

$$P^{(1)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) = \frac{P^{(0)}_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}) 2^{-d^{(0)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i})}}{\sum\limits_{\hat{y}'_{\mathrm{r},i}} P^{(0)}_{\hat{Y}_{\mathrm{r},i}}(\hat{y}'_{\mathrm{r},i}) 2^{-d^{(0)}(y_{\mathrm{r},i}, \hat{y}'_{\mathrm{r},i})}}, \tag{4.49}$$

which is used as a starting point for the next iteration. The update (4.49) can be further simplified since the last addend in (4.48) cancels with $P^{(0)}_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i})$ after exponentiation in (4.49). Defining the function

$$\begin{aligned} \bar{d}^{(0)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i}) = \beta \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) D_{\mathrm{KL}} \left(P_{X_i|Y_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|y_{\mathrm{r},i}, y_{\mathrm{d},i}) \middle\| P^{(0)}_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \right) \\ - \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) \log_2 \left(P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i}) \right), \end{aligned} \tag{4.50}$$

where clearly

$$\bar{d}^{(0)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i}) \geq 0, \quad \forall (y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i}) \in \left(\mathcal{Y}_{\mathrm{r},i} \times \hat{\mathcal{Y}}_{\mathrm{r},i} \right), \tag{4.51}$$

the simplified update rule is

$$P^{(1)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) = \frac{2^{-\bar{d}^{(0)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i})}}{\sum\limits_{\hat{y}'_{\mathrm{r},i}} 2^{-\bar{d}^{(0)}(y_{\mathrm{r},i}, \hat{y}'_{\mathrm{r},i})}}. \tag{4.52}$$

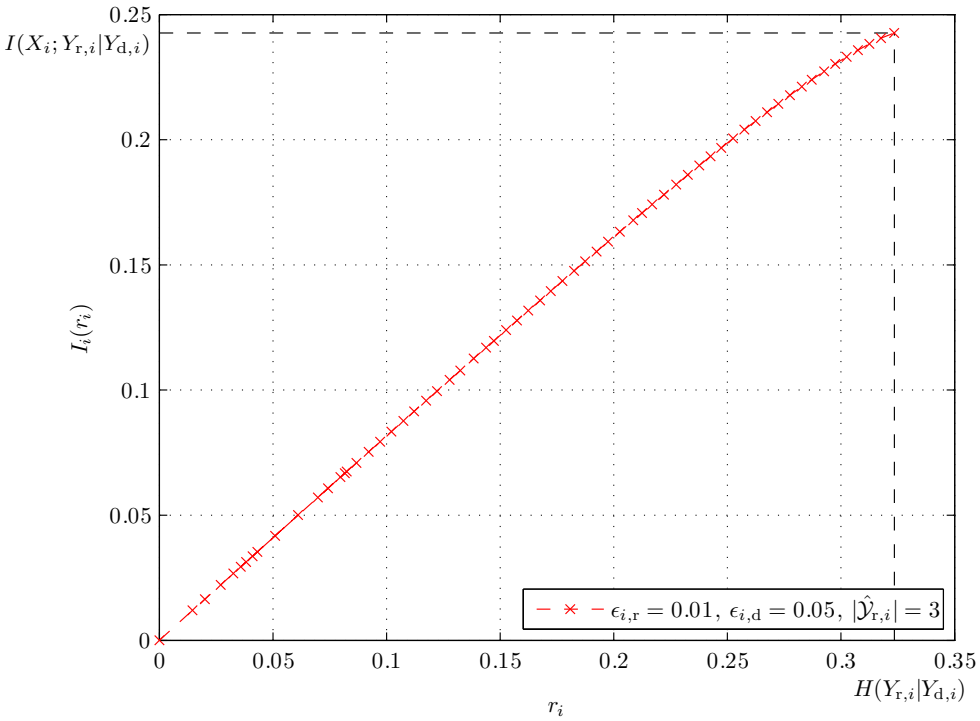


Figure 4.4.: The function $I_i(r_i)$ from Example 4.1.

We summarize the overall algorithm in Algorithm 4.1. As for the iterative information bottleneck algorithm, we repeatedly execute Algorithm 4.1 to ensure that $\bar{I}_i(r_i(\beta))$ is close to $I(r_i(\beta))$, and by varying $\beta > 0$, one can cover the entire $I_i(r_i)$ curve. In the following, it is assumed that $\bar{I}_i(r_i(\beta)) = I_i(r_i(\beta))$, i.e., given some parameter $\beta > 0$, we are in a position to compute the triple $(r_i(\beta), I_i(r_i(\beta)), I_i'(r_i(\beta)))$, where the slope $I_i'(r_i(\beta))$ at $r_i(\beta)$ is given by $1/\beta$.

Example 4.1. Suppose that both the source–relay channel and the source–destination channels are BSCs with uniform inputs and crossover probabilities given by $\epsilon_{i,\mathrm{r}} = 0.01$ and $\epsilon_{i,\mathrm{d}} = 0.05$, respectively. For such a setting, we have $H(Y_{\mathrm{r},i}|Y_{\mathrm{d},i}) = 0.324$ and $I(X_i;Y_{\mathrm{r},i}|Y_{\mathrm{d},i}) = 0.243$, and Figure 4.4 shows the curve of $I_i(r_i)$ obtained numerically with Algorithm 4.1.

4.3.4. Cutting-plane algorithm

By virtue of Algorithm 4.1, we have access to the function $I_i(r_i)$ only through a parametrization via β. However, in order to solve Problem (4.34) by standard methods, what is needed is a method to compute a pair $(I_i(r_i), I_i'(r_i))$ for a given value of r_i. Obviously, the missing link is a method to compute the value of β that corresponds to a given r_i. It follows straightforwardly from Theorems 4.3 and 4.4 that $r_i(\beta)$ is increasing in β; consequently, $\beta(r_i)$ can be found by bisection. In each step of the bisection, Algorithm 4.1 is called for a

Algorithm 4.1 Algorithm to compute $(r_i(\beta), \bar{I}_i(r_i(\beta))$.

Input: $P_{X_i Y_{\mathrm{d},i} Y_{\mathrm{r},i}}(x_i, y_{\mathrm{d},i}, y_{\mathrm{r},i})$, $\mathcal{X}_i$, $\mathcal{Y}_{\mathrm{d},i}$, $\mathcal{Y}_{\mathrm{r},i}$, $\beta > 0$, $\epsilon > 0$

Initialization: randomly choose a valid initial mapping $P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$, $k \leftarrow 1$

$$P^{(0)}_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}) \leftarrow \sum_{y_{\mathrm{r},i}} P_{Y_{\mathrm{r},i}}(y_{\mathrm{r},i}) P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$$

$$P^{(0)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \leftarrow \sum_{y_{\mathrm{r},i}} P_{Y_{\mathrm{r},i}Y_{\mathrm{d},i}}(y_{\mathrm{r},i}, y_{\mathrm{d},i}) P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$$

$$P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i}) \leftarrow \frac{1}{P_{Y_{\mathrm{d},i}}(y_{\mathrm{d},i})} P^{(0)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})$$

$$P^{(0)}_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(x_i|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \leftarrow \frac{1}{P^{(0)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})} \sum_{y_{\mathrm{r},i}} P_{X_i Y_{\mathrm{r},i} Y_{\mathrm{d},i}}(x_i, y_{\mathrm{r},i}, y_{\mathrm{d},i}) P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$$

$$\begin{aligned}\bar{d}^{(0)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i}) \leftarrow{}& \beta \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) D_{\mathrm{KL}}\left(P_{X_i|Y_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|y_{\mathrm{r},i}, y_{\mathrm{d},i}) \middle\| P^{(0)}_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})\right) \\ &- \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) \log_2\left(P^{(0)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i})\right)\end{aligned}$$

$$P^{(1)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) \leftarrow \frac{2^{-\bar{d}^{(0)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i})}}{\sum_{\hat{y}'_{\mathrm{r},i}} 2^{-\bar{d}^{(0)}(y_{\mathrm{r},i}, \hat{y}'_{\mathrm{r},i})}}$$

while $\sum_{y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i}} \left| P^{(k)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) - P^{(k-1)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) \right| / (|\mathcal{Y}_{\mathrm{r},i}| \cdot |\hat{\mathcal{Y}}_{\mathrm{r},i}|) \geq \epsilon$ **do**

$$P^{(k)}_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}) \leftarrow \sum_{y_{\mathrm{r},i}} P_{Y_{\mathrm{r},i}}(y_{\mathrm{r},i}) P^{(k)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$$

$$P^{(k)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \leftarrow \sum_{y_{\mathrm{r},i}} P_{Y_{\mathrm{r},i}Y_{\mathrm{d},i}}(y_{\mathrm{r},i}, y_{\mathrm{d},i}) P^{(k)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$$

$$P^{(k)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i}) \leftarrow \frac{1}{P_{Y_{\mathrm{d},i}}(y_{\mathrm{d},i})} P^{(k)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})$$

$$P^{(k)}_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(x_i|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \leftarrow \frac{1}{P^{(k)}_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})} \sum_{y_{\mathrm{r},i}} P_{X_i Y_{\mathrm{r},i} Y_{\mathrm{d},i}}(x_i, y_{\mathrm{r},i}, y_{\mathrm{d},i}) P^{(k)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$$

$$\begin{aligned}\bar{d}^{(k)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i}) \leftarrow{}& \beta \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) D_{\mathrm{KL}}\left(P_{X_i|Y_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|y_{\mathrm{r},i}, y_{\mathrm{d},i}) \middle\| P^{(k)}_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\cdot|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})\right) \\ &- \sum_{y_{\mathrm{d},i}} P_{Y_{\mathrm{d},i}|Y_{\mathrm{r},i}}(y_{\mathrm{d},i}|y_{\mathrm{r},i}) \log_2\left(P^{(k)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i})\right)\end{aligned}$$

$$P^{(k+1)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) \leftarrow \frac{2^{-\bar{d}^{(k)}(y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i})}}{\sum_{\hat{y}'_{\mathrm{r},i}} 2^{-\bar{d}^{(k)}(y_{\mathrm{r},i}, \hat{y}'_{\mathrm{r},i})}}$$

$k \leftarrow k + 1$

end while

Algorithm 4.1 (continued)

18: $P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) \leftarrow P^{(k)}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$

19: $P_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}) \leftarrow \sum_{y_{\mathrm{r},i}} P_{Y_{\mathrm{r},i}}(y_{\mathrm{r},i}) P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$

20: $P_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \leftarrow \sum_{y_{\mathrm{r},i}} P_{Y_{\mathrm{r},i}Y_{\mathrm{d},i}}(y_{\mathrm{r},i}, y_{\mathrm{d},i}) P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$

21: $P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i}) \leftarrow \dfrac{1}{P_{Y_{\mathrm{d},i}}(y_{\mathrm{d},i})} P_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})$

22: $P_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(x_i|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \leftarrow \dfrac{1}{P_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})} \sum_{y_{\mathrm{r},i}} P_{X_i Y_{\mathrm{r},i} Y_{\mathrm{d},i}}(x_i, y_{\mathrm{r},i}, y_{\mathrm{d},i}) P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})$

23: $r_i(\beta) \leftarrow \displaystyle\sum_{y_{\mathrm{r},i}, \hat{y}_{\mathrm{r},i}} P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i}) P_{Y_{\mathrm{r},i}}(y_{\mathrm{r},i}) \log_2 \left(\frac{P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{r},i})}{P_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i})} \right)$
$\qquad - \displaystyle\sum_{y_{\mathrm{d},i}, \hat{y}_{\mathrm{r},i}} P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i}) P_{Y_{\mathrm{d},i}}(y_{\mathrm{d},i}) \log_2 \left(\frac{P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}|y_{\mathrm{d},i})}{P_{\hat{Y}_{\mathrm{r},i}}(\hat{y}_{\mathrm{r},i})} \right)$

24: $\bar{I}_i(r_i(\beta)) \leftarrow \displaystyle\sum_{x_i, y_{\mathrm{d},i}, \hat{y}_{\mathrm{r},i}} P_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(x_i|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) P_{\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i}) \log_2 \left(\frac{P_{X_i|\hat{Y}_{\mathrm{r},i}Y_{\mathrm{d},i}}(x_i|\hat{y}_{\mathrm{r},i}, y_{\mathrm{d},i})}{P_{X_i|Y_{\mathrm{d},i}}(x_i|y_{\mathrm{d},i})} \right)$

particular β, yielding a triple $(r_i(\beta), I_i(r_i(\beta)), I_i'(r_i(\beta)))$. Clearly, each of these triples provides information about the function I_i. However, using a standard method, only the triple corresponding to the solution $\beta(r_i)$ of the bisection procedure is taken into account. Based on this observation, we propose a modified cutting-plane algorithm to solve Problem (4.34) that exploits all available information.

The proposed method is a variation of the standard outer linearization method (OLM) [BSS06], which is based on an outer approximation of the graph of $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{I}(\boldsymbol{r})$ by tangential hyperplanes. Let $\boldsymbol{r}^*$ denote a maximizer of (4.34), and let $\boldsymbol{r}(\boldsymbol{\beta}) = [r_1(\beta_1), \ldots, r_M(\beta_M)]^{\mathrm{T}}$, $f(\boldsymbol{r}) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{I}(\boldsymbol{r})$, and suppose $\delta > 0$ is the desired tolerance. The algorithm is as follows:

1. Choose

$$\boldsymbol{\beta}_{\min} = [\beta_{\min,1}, \beta_{\min,2}, \ldots, \beta_{\min,M}]^{\mathrm{T}} \tag{4.53}$$

$$\boldsymbol{\beta}_{\max} = [\beta_{\max,1}, \beta_{\max,2} \ldots, \beta_{\max,M}]^{\mathrm{T}} \tag{4.54}$$

such that $\boldsymbol{r}(\boldsymbol{\beta}_{\min}) \leq \boldsymbol{r}^* \leq \boldsymbol{r}(\boldsymbol{\beta}_{\max})$. Set $\boldsymbol{r}_{-2} = \boldsymbol{r}(\boldsymbol{\beta}_{\min})$ and $\boldsymbol{r}_{-1} = \boldsymbol{r}(\boldsymbol{\beta}_{\max})$.

2. Initialization: Run Algorithm 4.1 for an initial vector $\boldsymbol{\beta}_0 = 0.5(\boldsymbol{\beta}_{\max} + \boldsymbol{\beta}_{\min})$, yielding $\boldsymbol{r}_0 = \boldsymbol{r}_0(\boldsymbol{\beta}_0)$ and compute $f_0 = f(\boldsymbol{r}_0)$. The subgradient $\boldsymbol{g}_0 = \nabla f(\boldsymbol{r}_0)$ is given by $g_{0,i} = \alpha_i/\beta_{0,i}$. Set $b_0 = f_0 - \boldsymbol{g}_0^{\mathrm{T}}\boldsymbol{r}_0$, $f_{\mathrm{LB}} = -1$, and $k = 1$.

3. At iteration k, solve the linear program

$$\begin{aligned} \max_{(s,\boldsymbol{r})} \; s \quad \text{s.t.} \quad & \boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{r} + b_j \geq s, \quad j = 0, \ldots, k-1, \\ & \boldsymbol{r}(\boldsymbol{\beta}_{\min}) \leq \boldsymbol{r} \leq \boldsymbol{r}(\boldsymbol{\beta}_{\max}), \quad \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{r} \leq I_{\mathrm{r,d}}. \end{aligned} \tag{4.55}$$

   Let $(s_k^*, \boldsymbol{r}_k^*)$ be a maximizer of (4.55). Set $f_{\mathrm{UB}} = s_k^*$.

4. The standard OLM [BSS06] proceeds by evaluating f and ∇f at $\boldsymbol{r}_k^*$. As pointed out before, the function value $f(\boldsymbol{r}_k^*)$ and subgradient $\nabla f(\boldsymbol{r}_k^*)$ can only be computed by searching for the corresponding $\boldsymbol{\beta}_k^*$. Therefore, a vector $\boldsymbol{r}_k$ close to $\boldsymbol{r}_k^*$ is determined as follows. Let

$$r_{\mathrm{U},i} = \min\{r_{j,i} | r_{j,i} \geq r_{k,i}^*, j = -2, \ldots, k-1\} \tag{4.56}$$

$$r_{\mathrm{L},i} = \max\{r_{j,i} | r_{j,i} \leq r_{k,i}^*, j = -2, \ldots, k-1\}. \tag{4.57}$$

   Note that $\boldsymbol{r}_{\mathrm{L}} \leq \boldsymbol{r}_k^* \leq \boldsymbol{r}_{\mathrm{U}}$. Let $\boldsymbol{\beta}_{\mathrm{U}}$ and $\boldsymbol{\beta}_{\mathrm{L}}$ correspond to $\boldsymbol{r}_{\mathrm{U}}$ and $\boldsymbol{r}_{\mathrm{L}}$, respectively. Set $\boldsymbol{\beta}_k = 0.5(\boldsymbol{\beta}_{\mathrm{U}} + \boldsymbol{\beta}_{\mathrm{L}})$. Run Algorithm 4.1 using $\boldsymbol{\beta}_k$, yielding $\boldsymbol{r}_k = \boldsymbol{r}_k(\boldsymbol{\beta}_k)$, and compute $f_k = f(\boldsymbol{r}_k)$. If $f_k > f_{\mathrm{LB}}$ and $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{r}_k \leq I_{\mathrm{r,d}}$, set $f_{\mathrm{LB}} = f_k$ and $\boldsymbol{r}_{\mathrm{LB}} = \boldsymbol{r}_k$.

5. If $f_{\mathrm{UB}} - f_{\mathrm{LB}} \geq \delta$, compute $\boldsymbol{g}_k = \nabla f(\boldsymbol{r}_k)$ and $b_k = f_k - \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{r}_k$, increment k, and go to step 3). Otherwise, set $\boldsymbol{r}^* = \boldsymbol{r}_{\mathrm{LB}}$.

In contrast to the standard OLM, the iterates $\boldsymbol{r}_k^*$ may not change over multiple iterations, even if $\boldsymbol{r}_k^* \neq \boldsymbol{r}^*$. Still, in this case $\boldsymbol{\beta}_k$ converges to $\boldsymbol{\beta}_k^*$, which implies that the overall algorithm converges to the optimal solution $\boldsymbol{r}^*$ (under the previously stated assumption that $I_i(r_i(\beta)) = \bar{I}_i(r_i(\beta))$ holds).

4.3.5. Numerical example

One example is given in the following for $M = 4$ sources and BSCs as source–relay and source–destination channels with i.i.d. X_i, where the crossover probabilities are

$$\boldsymbol{\epsilon}_{\mathrm{d}} = [\epsilon_{1,\mathrm{d}}, \epsilon_{2,\mathrm{d}}, \epsilon_{3,\mathrm{d}}, \epsilon_{4,\mathrm{d}}]^{\mathrm{T}} = [0.1, 0.2, 0.05, 0.01]^{\mathrm{T}} \tag{4.58}$$

$$\boldsymbol{\epsilon}_{\mathrm{r}} = [\epsilon_{1,\mathrm{r}}, \epsilon_{2,\mathrm{r}}, \epsilon_{3,\mathrm{r}}, \epsilon_{4,\mathrm{r}}]^{\mathrm{T}} = [0.05, 0.01, 0.03, 0.02]^{\mathrm{T}}. \tag{4.59}$$

The time sharing variables are $\alpha_i = 0.2$ for all i, so that $\bar{\alpha} = 0.2$, and by choosing $I(X_{\mathrm{r}}; Y_{\mathrm{d,r}}) = 1$ we obtain $I_{\mathrm{r,d}} = 0.2$. Then, setting $|\hat{\mathcal{Y}}_{\mathrm{r},i}| = 3$, we obtain

$$\boldsymbol{r}_{\max} = [H(Y_{\mathrm{r},1}|Y_{\mathrm{d},1}), H(Y_{\mathrm{r},2}|Y_{\mathrm{d},2}), H(Y_{\mathrm{r},3}|Y_{\mathrm{d},3}), H(Y_{\mathrm{r},4}|Y_{\mathrm{d},4})]^{\mathrm{T}} \tag{4.60}$$

$$= [0.58, 0.73, 0.39, 0.19]^{\mathrm{T}}, \tag{4.61}$$

and $\boldsymbol{I}(\boldsymbol{r}_{\max}) = [0.30, 0.65, 0.20, 0.05]^{\mathrm{T}}$. The solution provided by the proposed modified OLM after 25 iterations is $\boldsymbol{r}^* = [0.18, 0.72, 0.10, 0]^{\mathrm{T}}$ and $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{I}(\boldsymbol{r}^*) = 0.1621$, with a tolerance of $\delta = 10^{-4}$. The individual tradeoff between relevant information and rate for each user

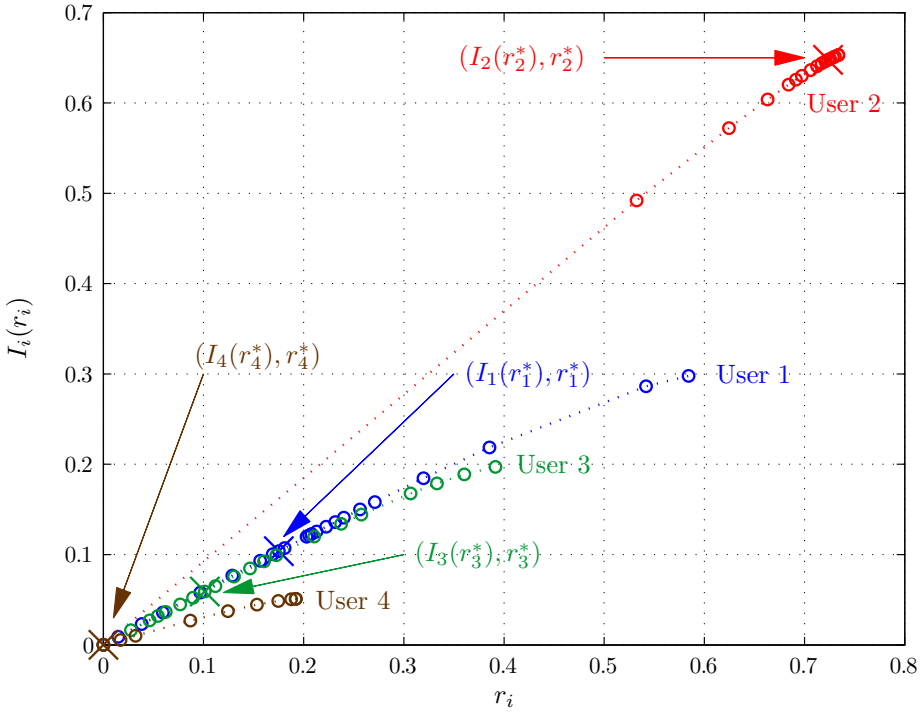


Figure 4.5.: Information-rate-tradeoff for each user in the numerical example.

is shown in Figure 4.5 with the iterates of the algorithm plotted as circles. Note that the algorithm computes an increased number of samples of $I_i(r_i)$ in proximity to the optimal points, which are also marked in Figure 4.5. As expected, those users with a strong source–relay link and a weak source–destination channel are assigned a large source coding rate. Comparing this with the uniform rate allocation $\boldsymbol{r}_{\text{unif}} = [0.25, 0.25, 0.25, 0.25]^{\text{T}}$ yields $\boldsymbol{\alpha}^{\text{T}}\boldsymbol{I}(\boldsymbol{r}_{\text{unif}}) = 0.1138$, a loss of roughly 30 % compared to $\boldsymbol{\alpha}^{\text{T}}\boldsymbol{I}(\boldsymbol{r}^*)$.

4.4. Relation to Noisy Network Coding

NNC [ADT11, LKEGC11] is a recently proposed scheme for communicating messages between multiple sources and destinations over a noisy network. Applied to the relay channel, NNC is variant of the CF protocol with quantization at the relay but without binning. For the relay channel with a single user, the rates achievable with CF and NNC are the same [LKEGC11]. For networks with multiple users and/or multiple relays, however, NNC may give a larger rate region than CF [LKEGC11]. In this section, we show that the rate region for the M-user orthogonal MARC with NNC is the same as the rate region of CF when $\prod_{i=1}^{M} P_{\hat{Y}_{\text{r},i}|Y_{\text{r},i}}$ is optimized for maximal sum-rate. Hence, the sum-rate optimization techniques presented in Sections 4.2 and 4.3 also apply to NNC.

4.4.1. Description of NNC and achievable rates

Like for the CF scheme, the relay compresses its received vectors $\boldsymbol{Y}_{\mathrm{r},i}$ to an estimate $\hat{\boldsymbol{Y}}_{\mathrm{r},i}(s_i) \in \hat{\mathcal{Y}}_{\mathrm{r},i}^{\alpha_i n}$, $s_i \in \{1, 2, \ldots, 2^{n\hat{R}_i}\}$, where each $\hat{\boldsymbol{Y}}_{\mathrm{r},i}$ is generated i.i.d. according to the distribution given in (4.3). In contrast to the CF protocol, however, the binning step is omitted, and the relay sends a corresponding codeword $\boldsymbol{X}_{\mathrm{r}}(s_1, s_2, \ldots, s_M)$ from a codebook with $2^{nR_{\mathrm{r}}}$ elements, where now $R_{\mathrm{r}} = \sum_{i=1}^{M} \hat{R}_i$. Instead of employing a two-step decoder that first recovers $\hat{\boldsymbol{Y}}_{\mathrm{r},i}(s_i)$, $i = 1, 2, \ldots, M$, and then uses $\hat{\boldsymbol{Y}}_{\mathrm{r},i}(s_i)$ and $\boldsymbol{Y}_{\mathrm{d},i}$ to recover $\boldsymbol{X}_i$, the destination performs decoding *jointly* on $\boldsymbol{Y}_{\mathrm{d,r}}$ and $\boldsymbol{Y}_{\mathrm{d},i}$, $i = 1, 2, \ldots, M$.

Achievable rates for NNC are given in [LKEGC11, Theorem 2], which we specialize to the orthogonal MARC in the following proposition.

Proposition 4.5. The rate vector $\boldsymbol{R}$ is achievable in the orthogonal MARC with M users, NNC, and separate compression of the received vectors $\boldsymbol{Y}_{\mathrm{r},i}$ at the relay if

$$\sum_{i\in\mathcal{I}} R_i < \sum_{i\in\mathcal{I}} \alpha_i I(X_i; \hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}) \tag{4.62}$$

$$\sum_{i\in\mathcal{I}} R_i < \sum_{i\in\mathcal{I}} \alpha_i I(X_i; Y_{\mathrm{d},i}) + \bar{\alpha} I(X_{\mathrm{r}}; Y_{\mathrm{d,r}}) - \sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i} | X_i, Y_{\mathrm{d},i}), \tag{4.63}$$

for all subsets $\mathcal{I} \subseteq \{1, 2, \ldots, M\}$. The set of rate vectors $\boldsymbol{R}$ satisfying (4.62) and (4.63) for all $\mathcal{I}$ is denoted by $\mathcal{R}_{\mathrm{NNC}}$.

Proof. Proposition 4.5 can be readily obtained from [LKEGC11, Theorem 2]. ■

4.4.2. NNC and CF

We have the result that any rate achievable with CF is also achievable with NNC.

Proposition 4.6. For the orthogonal MARC with M users and some fixed distribution $\prod_{i=1}^{M} P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$, we have $\mathcal{R}_{\mathrm{CF}} \subseteq \mathcal{R}_{\mathrm{NNC}}$.

Proof. See Appendix C.3. ■

At the sum-rate optimal distribution, the NNC rate region is characterized as follows.

Theorem 4.7. Consider the orthogonal MARC with M users and NNC. For the distribution $\prod_{i=1}^{M} P^{*}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$ that maximizes the total achievable sum-rate $\sum_{i=1}^{M} R_i$ we have

$$\bar{\alpha} I(X_{\mathrm{r}}; Y_{\mathrm{d,r}}) = \sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i} | Y_{\mathrm{d},i}). \tag{4.64}$$

Thus, for the sum-rate optimal distribution, the NNC rate region is given by

$$R_i < \alpha_i I(X_i; \hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}), \quad i = 1, 2, \ldots, M. \tag{4.65}$$

Proof. See Appendix C.4. ■

Since (4.64) holds, Condition (4.5) is satisfied with equality, and any rate vector satisfying (4.65) also satisfies (4.4). Thus, at the sum-rate optimal distribution, any rate achievable with NNC may also be achieved with CF. Theorem 4.7 also implies that the sum-rate maximizing distribution for NNC can be found by solving the optimization problem

$$\max_{P_{\hat{Y}_{\mathrm{r},1}|Y_{\mathrm{r},1}},P_{\hat{Y}_{\mathrm{r},2}|Y_{\mathrm{r},2}},\ldots,P_{\hat{Y}_{\mathrm{r},M}|Y_{\mathrm{r},M}}} \min\Bigg\{\sum_{i=1}^{M}\alpha_i I(X_i;\hat{Y}_{\mathrm{r},i},Y_{\mathrm{d},i}), \tag{4.66}$$
$$\sum_{i=1}^{M} I(X_i;\hat{Y}_{\mathrm{r},i},Y_{\mathrm{d},i}) + \bar{\alpha} I(X_{\mathrm{r}};Y_{\mathrm{d,r}}) - \sum_{i=1}^{M}\alpha_i I(Y_{\mathrm{r},i};\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i})\Bigg\},$$

which, in contrast to Problem (4.31), is an unconstrained optimization of a minimum of two terms. Due to Theorem 4.7, (4.31) and (4.66) have the same optimum and the same optimizer. Likewise, for the Gaussian scenario (cf. Section 4.2) and NNC, the sum-rate optimal vector $\boldsymbol{\sigma}^{2,*} = [\sigma_1^{2,*}, \sigma_2^{2,*}, \ldots, \sigma_M^{2,*}]^{\mathrm{T}}$ of variances of $\bar{N}_{\mathrm{r},i}$ is the solution of the optimization problem

$$\max_{\boldsymbol{\sigma}^2\geq\mathbf{0}} \min\Bigg\{\sum_{i=1}^{M}\alpha_i\log_2\left(1+\gamma_{i,\mathrm{d}}+\frac{\gamma_{i,\mathrm{r}}}{1+\sigma_i^2}\right), \tag{4.67}$$
$$\sum_{i=1}^{M}\alpha_i\log_2\left(1+\gamma_{i,\mathrm{d}}+\frac{\gamma_{i,\mathrm{r}}}{1+\sigma_i^2}\right) + \bar{\alpha}\log_2(1+\gamma_{\mathrm{r,d}}) - \sum_{i=1}^{M}\alpha_i\log_2\left(1+\frac{1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}}}{\sigma_i^2(1+\gamma_{i,\mathrm{d}})}\right)\Bigg\}.$$

Note that we write the optimization problem in terms of σ_i^2 here, since there is no notion of the rate $\tilde{R}_i$ in the NNC scheme. The solution to (4.67) can be readily obtained by inserting the solution provided in Theorem 4.2 into (4.16), yielding

$$\sigma_i^{2,*} = \begin{cases} \dfrac{1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}}}{\tau\gamma_{i,\mathrm{r}}-(1+\gamma_{i,\mathrm{d}})} & \text{if } \tau \geq \xi_i \\ \infty & \text{if } \tau < \xi_i, \end{cases} \tag{4.68}$$

where τ is chosen such that

$$\sum_{i=1}^{M}\alpha_i\log_2\left(1+\frac{1+\gamma_{i,\mathrm{r}}+\gamma_{i,\mathrm{d}}}{\sigma_i^{2,*}(1+\gamma_{i,\mathrm{d}})}\right) = \bar{\alpha}\log_2(1+\gamma_{\mathrm{r,d}}). \tag{4.69}$$

In summary, the sum-rate maximizing solutions to the source coding rate allocation problems in (4.20) and (4.31) for CF also provide sum-rate maximizing solutions for NNC.

Finally, we point out that for $M = 1$, any rate achievable with NNC can also be achieved with CF.

Corollary 4.8. Consider the orthogonal MARC with $M = 1$ user and any fixed distribution $P_{\hat{Y}_{\mathrm{r},1}|Y_{\mathrm{r},1}}$, and suppose that R_1 is in the NNC rate region. Then R_1 can also be

achieved with CF.

Corollary 4.8 follows from Theorem 4.7. However, an extension of Corollary 4.8 to $M > 1$ may not hold, as the following example shows.

Example 4.2. Suppose that $M = 2$, consider the Gaussian scenario of Section 4.2, and assume that $\gamma_{1,\mathrm{r}} = \gamma_{2,\mathrm{r}} = \gamma_{1,\mathrm{d}} = \gamma_{2,\mathrm{d}} = \gamma_{\mathrm{r,d}} = 10$. We further choose $\alpha_1 = \alpha_2 = \bar{\alpha} = 1/3$, and $\sigma_1^2 = \sigma_2^2 = 2/3$. With this choice, we obtain

$$1.1531 = \bar{\alpha}\log_2(1+\gamma_{\mathrm{r,d}}) < \sum_{i=1}^{2} \underbrace{\alpha_i I(Y_{\mathrm{r},i};\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i})}_{=0.65} = 1.3, \tag{4.70}$$

or, equivalently,

$$\bar{\alpha}\log_2(1+\gamma_{\mathrm{r,d}}) - \sum_{i=1}^{2} \alpha_i I(Y_{\mathrm{r},i};\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = -0.1468, \tag{4.71}$$

so that (4.5) does not hold. Moreover, for $i = 1, 2$, we have

$$\alpha_i I(X_i; Y_{\mathrm{d},i}) = \alpha_i \log_2(1+\gamma_{i,\mathrm{d}}) = 1.1531 \tag{4.72}$$

$$\alpha_i I(X_i;\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = \alpha_i \log_2\left(1 + \frac{\gamma_{i,\mathrm{r}}}{(1+\sigma_i^2)(1+\gamma_{i,\mathrm{d}})}\right) = 0.2093 \tag{4.73}$$

$$\alpha_i I(X_i;\hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}) = \alpha_i \log_2\left(1 + \gamma_{i,\mathrm{d}} + \frac{\gamma_{i,\mathrm{r}}}{(1+\sigma_i^2)}\right) = 1.3625. \tag{4.74}$$

Hence, we obtain the NNC rate region as

$$R_1 < \alpha_1 I(X_1;\hat{Y}_{\mathrm{r},1}, Y_{\mathrm{d},1}) = 1.3625 \tag{4.75}$$

$$R_2 < \alpha_2 I(X_2;\hat{Y}_{\mathrm{r},2}, Y_{\mathrm{d},2}) = 1.3625 \tag{4.76}$$

$$R_1 + R_2 < \sum_{i=1}^{M} \alpha_i I(X_i;\hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}) + \bar{\alpha}\log_2(1+\gamma_{\mathrm{r,d}}) - \sum_{i=1}^{2} \alpha_i I(Y_{\mathrm{r},i};\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = 2.5781, \tag{4.77}$$

which is illustrated in Figure 4.6. To apply the CF protocol to that system, one has to pick $(\hat{Y}'_{\mathrm{r},1}, \hat{Y}'_{\mathrm{r},2})$ and hence $\sigma_i^{2\prime} > \sigma_i^2$ for at least one i such that

$$\bar{\alpha}\log_2(1+\gamma_{\mathrm{r,d}}) \geq \sum_{i=1}^{2} \alpha_i I(Y_{\mathrm{r},i};\hat{Y}'_{\mathrm{r},i}|Y_{\mathrm{d},i}), \tag{4.78}$$

which necessarily implies

$$\alpha_i I(X_i;\hat{Y}'_{\mathrm{r},i}, Y_{\mathrm{d},i}) < \alpha_i I(X_i;\hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}) \tag{4.79}$$

for at least one i. Consequently, since the CF achievable region is rectangular for $M = 2$ as long as (4.78) holds, it cannot cover the entire NNC region. For instance, let $\sigma_1^{2\prime} = \sigma_1^2 =$

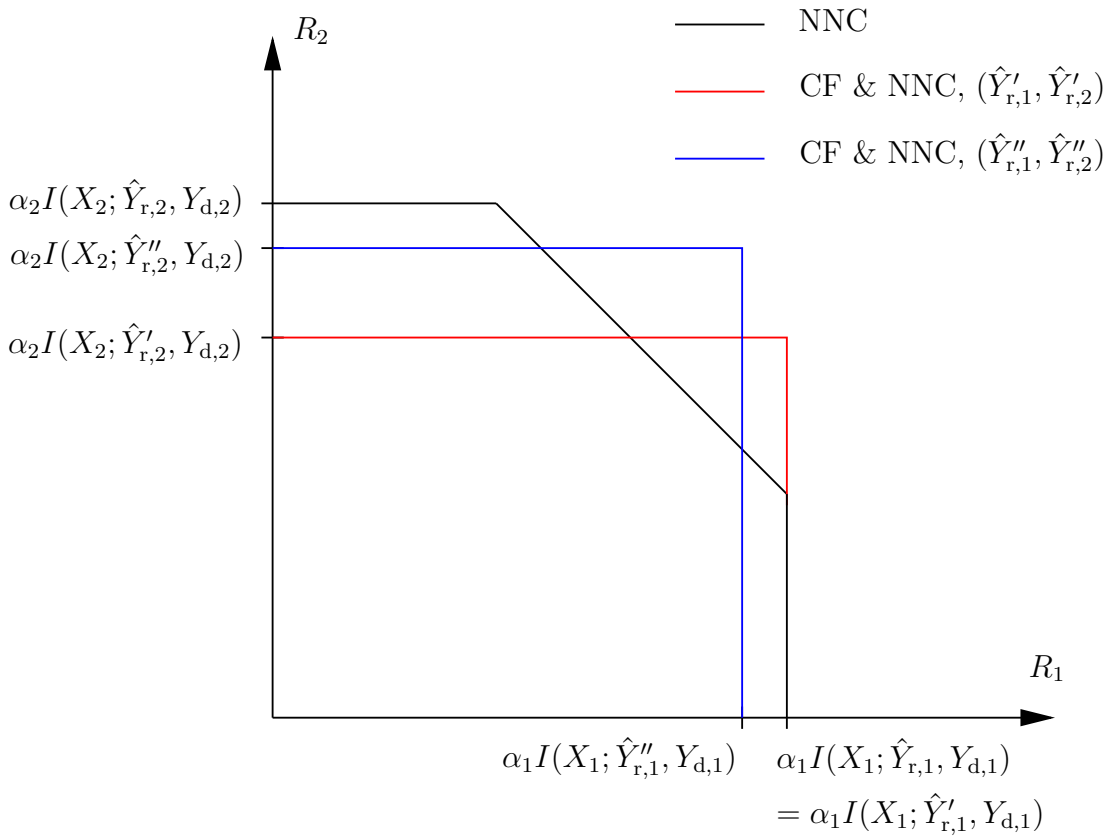


Figure 4.6.: Rate regions for Example 4.2.

2/3, and choose $\sigma_2^{2\prime}$ such that (4.78) holds with equality. We thus obtain $\sigma_2^{2\prime} = 1.0336$ and the CF bounds

$$R_1 < \alpha_1 I(X_1; \hat{Y}'_{\mathrm{r},1}, Y_{\mathrm{d},1}) = 1.3625 \tag{4.80}$$

$$R_2 < \alpha_2 I(X_2; \hat{Y}'_{\mathrm{r},2}, Y_{\mathrm{d},2}) = 1.3308, \tag{4.81}$$

which are also depicted in Figure 4.6. Note that (4.80) and (4.81) also provide the rate bounds for NNC with the above choice of $\sigma_1^{2\prime}$ and $\sigma_2^{2\prime}$. Alternatively, pick $(\hat{Y}''_{\mathrm{r},1}, \hat{Y}''_{\mathrm{r},2})$ with $\sigma_1^{2\prime\prime} = \sigma_2^{2\prime\prime} = 0.8241$. Again, (4.78) holds with equality. We obtain the CF bounds

$$R_1 < \alpha_1 I(X_1; \hat{Y}''_{\mathrm{r},1}, Y_{\mathrm{d},1}) = 1.3476 \tag{4.82}$$

$$R_2 < \alpha_2 I(X_2; \hat{Y}''_{\mathrm{r},2}, Y_{\mathrm{d},2}) = 1.3476. \tag{4.83}$$

The above choice of $\sigma_1^{2\prime\prime} = \sigma_2^{2\prime\prime} = 0.8241$ is also the sum-rate optimal choice for both CF and NNC.

4.5. Discussion

Source coding rate allocation for the orthogonal MARC with CF is considered. The optimal source coding rate assignment at the relay is given by a water-filling solution for Gaussian channels and modulation, where only those users are included for cooperation that exhibit a sufficiently strong source–relay link and a weak source–destination link. For general DMCs with finite input and output alphabets, an optimal solution can be found by first using a variant of the information bottleneck iterative algorithm to characterize the tradeoff between rate and relevant information for each user individually, and by then employing a modified outer linearization method based on that information-rate tradeoff. An example for BSCs confirms that source coding rate is assigned to those users gaining the most from cooperation, i.e., users with a strong source–relay link and a weak source–destination link. Finally, we demonstrate that the optimal source coding rate allocation for CF is also sum-rate optimal for NNC, a variant of CF without binning at the relay.

5

Low-precision A/D conversion for maximum information rate in channels with memory

Considerable effort is spent on optimizing the modulation, coding, and detection of modern communication systems by using performance metrics such as the BER or the rate at which reliable transmission is possible. Other critical components, however, are traditionally designed by incorporating figures of merit inherently more suited for signal reproduction tasks than for digital communications, e.g., the mean squared error (MSE) [Llo82] or the total harmonic distortion [Kes05, Section 2-3]. One example of such a component is the analog-to-digital converter (ADC) at the receiver, which is omnipresent in modern digital systems. At high data rate and high precision, the analog-to-digital (A/D) conversion step is power-hungry, costly, and time-critical [Wal99, LRRB05, Mur08], especially at converter resolutions of 6-12 bits commonly employed today. Examples of high-speed links are optical transceivers with electronic dispersion compensation for single-mode and multi-mode fiber [BAP+06, ACH+08], and chip-to-chip serial links [HWS+07].

Since decreasing the resolution of an ADC with flash architecture [Wal99], [Kes05, Section 3-2] by a single bit cuts the number of necessary comparators (which is directly proportional to the power consumption) in half, one remedy to reduce the power consumption of ADCs operating at high speed is to reduce their precision to a few bits, e.g., one to three bits. To minimize the impact on performance, such a scheme requires optimizing the ADC levels: this has been studied for the AWGN channel by using mutual information as a figure of merit [SDM09], and for the intersymbol-interference (ISI) channel with AWGN by using the BER of the link [LSS10]. The properties of flat fading channels under single-

bit output quantization are analyzed in [KF10]. Alternatively, several converters can be operated in parallel on B branches with different relative delays to form a time-interleaved ADC [VJ06]; in such a design, the timing constraints can be relaxed by providing a B-fold increase in conversion time for the samples. The relative sampling phases of the component converters of such a device are chosen to maximize mutual information in [SBC10].

In this chapter, we consider the ISI channel with AWGN under low-precision A/D conversion and target the information rate of the channel as a cost criterion for the optimization of the quantization step. Specifically, our contributions are as follows:

- ▷ We show that at infinite SNR and for memoryless signaling using an alphabet of size Λ, a uniform quantizer cannot achieve the information rate of $\log_2(\Lambda)$ bits per channel use for all channels with $\log_2(\Lambda)$-bit/sample output quantization. Our proof uses the concept of information losslessness of finite-state machines [Huf54]. We further provide a constructive proof that optimal $\log_2(\Lambda)$-bit/sample quantizers exist for all channels. For BPSK modulation, we provide numerical results demonstrating that uniform quantizers are frequently suboptimal when the channel coefficients exhibit a Gaussian distribution.

- ▷ We design scalar ADCs at finite SNRs by maximizing a lower bound on the information rate under output quantization. This design framework is extended to vector quantizers to better exploit the correlation in the received sequence. We also derive an upper bound on the channel information rate, which is numerically optimized over the quantizer.

- ▷ We infer from our simulation results that 2-bit/sample optimized ADCs perform close to the limit given by unquantized outputs. The numerical results also demonstrate the advantage of optimized single-bit quantization over conventional methods, they highlight the gain from vector quantization, and they are in accordance with our theoretical results derived for high SNR. Finally, we provide an example of a channel for which a simple slicer combined with a carefully optimized channel input with memory fails to achieve a rate of one bit per channel use at high SNR, in contrast to memoryless binary signaling and an optimized single-bit quantizer.

This chapter is organized as follows. In Section 5.1, we describe the system model. The information losslessness of finite-state machines and quantization at infinite SNR is studied in Section 5.2, while the design framework at finite SNRs and the upper bound on the information rate are presented in Sections 5.3-5.5. Numerical results are shown in Section 5.6, followed by a discussion in Section 5.7.

5.1. System model and achievable rates

Consider transmission over the discrete-time channel with ISI and AWGN, so that the channel output at time k is

$$Y_k = \sum_{\ell=1}^{L_h} h_\ell X_{k+1-\ell} + N_k, \quad k = 1, 2, \ldots, n, \tag{5.1}$$

where the channel of length L_h has fixed real coefficients h_ℓ, $h_1 \neq 0$, and is normalized to $\sum_{\ell=1}^{L_h} h_\ell^2 = 1$. For notational convenience, we form the vector $\boldsymbol{h} = [h_1, h_2, \ldots, h_L]^{\mathrm{T}}$. Note that the restriction to $h_1 \neq 0$ imposes no loss of generality since the channel model (5.1) can be shifted in time such that the first channel coefficient is non-zero. The channel input $X_k \in \mathcal{X}$, $|\mathcal{X}| = \Lambda$, is real, discrete, and of Markov order M_x, i.e.,

$$P_{X_k|X^{k-1}}(x_k|x^{k-1}) = P_{X_k|X_{k-M_x}^{k-1}}(x_k|x_{k-M_x}^{k-1}), \quad k > M_x, \tag{5.2}$$

and with $M = \max\{M_x, L_h - 1\}$, we define the state S_k of the channel (cf. [ALV+06, Section II]) as $S_k = f(X_{k-M+1}^k) \in \mathcal{S} = \{0, 1, \ldots, \Lambda^M - 1\}$, where $f : \mathcal{X}^M \to \mathcal{S}$ is a one-to-one mapping. The additive noise $N_k \sim \mathcal{N}(0, \sigma^2)$ satisfies $\mathrm{E}[N_k N_{k'}] = \sigma^2 \mathbb{1}_{k=k'}$, and with $\sigma^2 = N_0/2$ and $\|\boldsymbol{h}\| = 1$, the SNR is $E_s/N_0 = \mathrm{E}[X_k^2]/(2\sigma^2)$.

At the receiver, the channel output Y_k is quantized using an ADC that is fixed for the entire transmission. Scalar quantization with J quantization regions is modeled using a quantization function $Q_1 : \mathbb{R} \to \mathcal{Z}$, where $\mathcal{Z} = \{0, 1, \ldots, J-1\}$ is the finite set of quantization indices, so that $Z_k = Q_1(Y_k)$, $k = 1, 2, \ldots, n$, is the quantizer output. In Section 5.4, we will also consider two-dimensional A/D conversion, where a two-dimensional quantizer with J regions has the quantization function $Q_2 : \mathbb{R}^2 \to \mathcal{Z}$, yielding $Z_k = Q_2(Y_{2k}, Y_{2k-1})$, $k = 1, 2, \ldots, n/2$, as the quantizer output, supposing without loss of essential generality that n is even. The rate of a quantizer is defined as $\log_2(J)/d$ bit/sample, where d, $d = 1, 2$, is the quantizer's dimension. Throughout, we assume that the channel is fixed for the entire transmission, and that the receiver has perfect channel state information. We refer to Chapter 6 for the problem of estimating an ISI channel under low-precision output quantization. In the sequel, a uniform quantizer is defined as a regular quantizer [GG92, Chapter 5] characterized by its step size δ. We optimize δ for the particular probability density function of Y_k [Say00, Chapter 8.4] for our comparisons in Section 5.6.

The information rate of the channel (5.1) is defined as

$$I(X;Y) = \lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n | S_0) \tag{5.3}$$

$$= \lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n | S_0 = s_0), \tag{5.4}$$

where the second equality holds since the ISI channel is indecomposable [DG08], so that $I(X;Y)$ does not depend on the choice of the initial state s_0 [Gal68, Theorem 4.6.4]. Although the closed-form computation of $I(X;Y)$ remains intractable for most cases of practical interest, various upper and lower bounds on $I(X;Y)$ are available in the literature [SK90,SOW91,SL96]. To numerically compute the information rate, (5.3) can be expanded yielding

$$I(X;Y) = \lim_{n\to\infty} \frac{1}{n} \left[h(Y^n | S_0 = s_0) - h(Y^n | X^n, S_0 = s_0) \right], \tag{5.5}$$

where $h(Y^n|X^n, S_0 = s_0) = h(N^n) = (n/2)\log_2(2\pi e\sigma^2)$ since the noise is AWGN; for a fixed input alphabet and distribution, one can evaluate $h(Y^n|S_0 = s_0)$ efficiently with the forward recursions of the BCJR-algorithm on the trellis of the channel [ALV+06]. The definition of the information rate is readily extended to quantized channel outputs, with Y^n replaced by its quantized version, and the differential entropy replaced by entropy. For a fixed channel, signal alphabet and distribution, and a fixed quantizer, the information rate can likewise be computed numerically with the algorithm of [ALV+06]. Since $P_{Z_k|X_{k-L_h+1}^k}(z_k|x_{k-L_h+1}^k)$ needs to be fixed for that algorithm, it seems hard to optimize the quantizer with the method of [ALV+06]; therefore, we will use a method related to the information bottleneck iterative algorithm [TPB99] for that optimization in Section 5.3.

5.2. Scalar quantization in the limit of high SNR

In this section we study the behavior of $I(X; Q_1(Y))$ in the limit of high SNR and i.i.d. signaling, i.e., $M_x = 0$, and $P_{X_k|X^{k-1}}(x_k|x^{k-1}) = P_{X_k}(x_k) = 1/\Lambda$ for all $x_k \in \mathcal{X}$. Further, we restrict the transmit alphabet to be of the form $\mathcal{X} = \{\xi_0, \xi_1, \ldots, \xi_{\Lambda-1}\}$ with $\xi_m - \xi_{m-1} = \Delta > 0$, $m = 1, 2, \ldots, \Lambda - 1$, and $(1/\Lambda)\sum_{m=0}^{\Lambda-1} \xi_m^2 = 1$. The information rate for the channel (5.1) with continuous outputs is at most $\log_2(\Lambda)$ bit per channel use under these assumptions. As we now show, an information rate of $\log_2(\Lambda)$ bit per channel use can be achieved with $\log_2(\Lambda)$-bit/sample quantization, at high SNR. The resulting quantizer, however, is not necessarily given by a uniform quantizer. For our analysis, we utilize the theory of information lossless finite-state machines as introduced by Huffman [Huf54], and studied further in subsequent work by Even [Eve65].

Definition 5.1 (Finite-state machine representation of the ISI channel)**.** The finite-state machine $\mathcal{M}^{\mathcal{X}}(\boldsymbol{h})$ formed by i.i.d. signaling from a transmit alphabet $\mathcal{X}$ over the channel $\boldsymbol{h}$ is a quintuple $(\mathcal{S}, \mathcal{X}, \tilde{\mathcal{Y}}, \nu, \varphi)$, where $\mathcal{S}$, $\mathcal{X}$, and $\tilde{\mathcal{Y}}$ are the finite nonempty sets of states, inputs, and outputs, respectively; we have $|\mathcal{S}| = \Lambda^{L_h-1}$, $|\tilde{\mathcal{Y}}| = \Gamma$, and $\tilde{\mathcal{Y}} = \{\gamma_0, \gamma_1, \ldots, \gamma_{\Gamma-1}\}$. The function $\nu : \mathcal{S} \times \mathcal{X} \to \mathcal{S}$ determines the next state from the current state and the current input, i.e., $s_t = \nu(s_{t-1}, x_t) = \nu(f(x_{t-L_h+1}^{t-1}), x_t) = f(x_{t-L_h+2}^{t})$, and the function $\varphi : \mathcal{S} \times \mathcal{X} \to \tilde{\mathcal{Y}}$ specifies the output of the machine associated with the current state and the current input, i.e., $\tilde{y}_t = \varphi(s_{t-1}, x_t) = \varphi(f(x_{t-L_h+1}^{t-1}), x_t) = \sum_{\ell=1}^{L_h} h_\ell x_{t+1-\ell}$.

The machine $\mathcal{M}^{\mathcal{X}}(\boldsymbol{h})$ can be represented by a trellis section as shown in Figure 5.1 for $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$. In this example, we have $\mathcal{X} = \{\pm 1\}$, $\mathcal{S} = \{0, 1, 2, 3\}$, $|\mathcal{S}| = 4$, and $\tilde{\mathcal{Y}} = \{-5/3, -1, -1/3, 1/3, 1, 5/3\}$ with $\Gamma = 6$. Two states s_{t-1} and s_t are connected by a branch if $s_t = \nu(s_{t-1}, x_t)$ for some $x_t \in \mathcal{X}$, and we label the branch with $(\tilde{y}_t, x_t)$, i.e., the output of the machine $\tilde{y}_t \in \tilde{\mathcal{Y}}$ when the input is x_t. Since Λ branches originate from each state, the total number of branches is Λ^{L_h}. Note that several branches can have the same noise-free channel output, so that it is possible that $\Gamma < \Lambda^{L_h}$, as in our example.

Also note that

$$\varphi(i, \xi_m) - \varphi(i, \xi_{m-1}) = \Delta h_1 \neq 0, \quad \text{for all } i \in \mathcal{S},\ m = 1, 2, \ldots, \Lambda - 1, \tag{5.6}$$

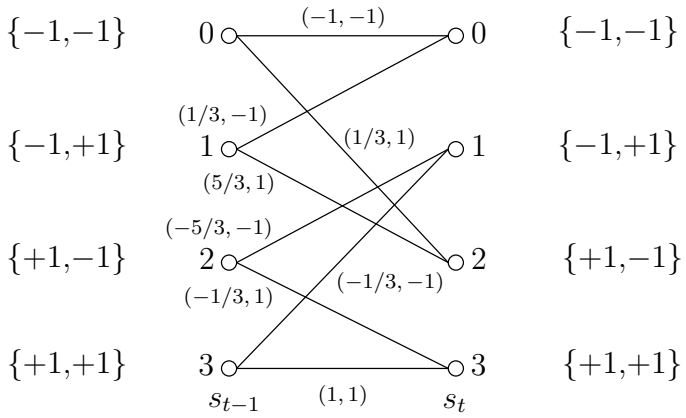


Figure 5.1.: Trellis section for the channel $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$ with $\mathcal{X} = \{\pm 1\}$. The corresponding x_{t-1}^{t} is also shown next to each state s_t.

since $h_1 \neq 0$ and $\Delta > 0$ by assumption, i.e., the outputs on different branches originating from the same state are distinct.

Definition 5.2. A finite-state machine $\mathcal{M}^{\mathcal{X}}(\boldsymbol{h})$ is called *information lossless* [Huf54] if there exists no integer $N \geq 0$, no two (not necessarily different) states $s_{t-1} = i$ and $s_{t+N} = e$ and no two different input sequences x_t^{t+N} and $\tilde{x}_t^{t+N}$ and output sequence $\tilde{y}_t^{t+N}$, such that both x_t^{t+N} and $\tilde{x}_t^{t+N}$ can lead from state $s_{t-1} = i$ to $s_{t+N} = e$, and both yield $\tilde{y}_t^{t+N}$. Likewise, a finite-state machine is called *information lossless of finite order* μ [KJ09, Chapter 14.4] if the initial state $s_{t-1} = i$ and the output sequence $\tilde{y}_t^{t+\mu-1}$ of length μ uniquely determine the channel input x_t, for all $i \in \mathcal{S}$.

We remark that the trellis associated with an information lossless machine is called *observable* [KS95].

Definition 5.3. The finite-state machine $\mathcal{M}_{Q_1}^{\mathcal{X}}(\boldsymbol{h})$ is obtained by concatenating $\mathcal{M}^{\mathcal{X}}(\boldsymbol{h}) = (\mathcal{S}, \mathcal{X}, \tilde{\mathcal{Y}}, \nu, \varphi)$ with a scalar quantizer Q_1, so that $\mathcal{M}_{Q_1}^{\mathcal{X}}(\boldsymbol{h})$ is given by the quintuple $(\mathcal{S}, \mathcal{X}, \mathcal{Z}, \nu, \varphi_{Q_1})$. The function $\varphi_{Q_1} : \mathcal{S} \times \mathcal{X} \to \mathcal{Z}$ specifies the output of $\mathcal{M}_{Q_1}^{\mathcal{X}}(\boldsymbol{h})$ associated with the current state and the current input, i.e., $\varphi_{Q_1}(s_{t-1}, x_t) = Q_1(\varphi(s_{t-1}, x_t))$.

We have the following theorem, which shows that an information rate of $I(X; Q_1(Y)) = \log_2(\Lambda)$ at $\sigma^2 = 0$ is closely related to $\mathcal{M}_{Q_1}^{\mathcal{X}}(\boldsymbol{h})$ being information lossless.

Theorem 5.1. Consider i.i.d. signaling with $\mathcal{X}$ over an ISI channel $\boldsymbol{h}$ with L_h coefficients, $\sigma^2 = 0$, and the use of a scalar quantizer with quantization function Q_1 at the channel output. We have $I(X; Q_1(Y)) = \log_2(\Lambda)$ if and only if $\mathcal{M}_{Q_1}^{\mathcal{X}}(\boldsymbol{h})$ is information lossless.

Proof. See Appendix D.1. ■

5.2.1. Optimal quantization with $J = \Lambda$ regions at infinite SNR

For quantization with $J = \Lambda$ regions at infinite SNR, we have the following theorem.

Theorem 5.2. Assume i.i.d. signaling using the alphabet $\mathcal{X}$ of size Λ, a noiseless channel with $\sigma^2 = 0$, and an impulse response $\boldsymbol{h}$ with $h_1 \neq 0$ and $L_h < \infty$. There exists a scalar quantization function $\tilde{Q}_1$ with $J = \Lambda$ regions such that $I(X; \tilde{Q}_1(Y)) = \log_2(\Lambda)$. Moreover, there exists at least one channel impulse response and alphabet $\mathcal{X}$ such that $I(X; Q_{\mathrm{u}}(Y)) < \log_2(\Lambda)$, where $Q_{\mathrm{u}}(y)$ denotes the quantization function of the uniform quantizer with $J = \Lambda$ regions.

We conclude from Theorem 5.2 that for memoryless BPSK, i.e., $\mathcal{X} = \{\pm 1\}$, it suffices to employ a single-bit quantizer to achieve an information rate of one bit per channel use at infinite SNR.

Proof. The key idea is to show that there is a mapping $\hat{Q} : \tilde{\mathcal{Y}} \to \{0, 1, \ldots, \Lambda - 1\}$, for all channels, such the finite-state machine $\mathcal{M}_{\hat{Q}}^{\mathcal{X}}(\boldsymbol{h})$ is information lossless of order $\mu = 1$, i.e., the input x_t can be uniquely recovered from $\hat{Q}(y_t)$ and knowledge of the previous state s_{t-1}. To make $\mathcal{M}_{\hat{Q}}^{\mathcal{X}}(\boldsymbol{h})$ information lossless of order $\mu = 1$, the mapping $\hat{Q}$ needs to be constructed such that, for any state i, the quantized noise-free outputs associated with branches leaving state i are different, i.e.,

$$\hat{Q}(\varphi(i, \xi_m)) \neq \hat{Q}(\varphi(i, \xi_{m'})), \quad \forall i \in \mathcal{S}, \quad m, m' \in \{0, 1, \ldots, \Lambda - 1\}, m \neq m'. \tag{5.7}$$

In principle, this should always be possible due to (5.6). However, one needs to proceed with care in order to adhere to constraints possibly imposed if $\Gamma < \Lambda^{L_h}$. In this case, there are noise-free outputs with multiplicity greater than one, i.e., there are $\tilde{y}_t = \gamma_k$, $k = 0, 1, \ldots, \Gamma - 1$, that can be generated by more than one input sequence $x_{t-L_h+1}^{t}$. These constraints can be taken into account by processing the γ_k's in ascending or descending order, depending on the sign of h_1, as done in the following.
Algorithm for constructing $\hat{Q}$:

1. If $h_1 < 0$, sort the elements of $\tilde{\mathcal{Y}}$ in descending order, so that $\gamma_k < \gamma_{k-1}$, $k = 1, 2, \ldots, \Gamma - 1$; if $h_1 > 0$, sort the elements of $\tilde{\mathcal{Y}}$ in ascending order.

2. Set $\hat{Q}(\gamma_0) = 0$, and set $k = 1$.

3. If $(\gamma_k - \Delta h_1) \in \tilde{\mathcal{Y}}$, then set $\hat{Q}(\gamma_k) = [\hat{Q}(\gamma_k - \Delta h_1) + 1] \bmod \Lambda$; if $(\gamma_k - \Delta h_1) \notin \tilde{\mathcal{Y}}$, set $\hat{Q}(\gamma_k) = \hat{Q}(\gamma_{k-1})$.

4. If $k < \Gamma - 1$, increment k and go to 3); if $k = \Gamma - 1$, return $\hat{Q}$.

To see that the aforementioned constraints can be resolved by the proposed algorithm, consider the third line of the algorithm. If $(\gamma_k - \Delta h_1) \in \tilde{\mathcal{Y}}$, then $\hat{Q}(\gamma_k - \Delta h_1)$ is already defined due to the ordered processing of the γ_k's, and by choosing $\hat{Q}(\gamma_k) = [\hat{Q}(\gamma_k - \Delta h_1) + 1] \bmod \Lambda$, (5.7) is not violated by the current mapping $\hat{Q}$. Alternatively, if $(\gamma_k - \Delta h_1) \notin$

$\tilde{\mathcal{Y}}$, which means that there are no constraints due to the previously processed noise-free outputs $\gamma_0, \gamma_1, \ldots, \gamma_{k-1}$, the choice of $\hat{Q}(\gamma_k)$ is not restricted; choosing $\hat{Q}(\gamma_k) = \hat{Q}(\gamma_{k-1})$ is advantageous since no decision boundary is needed between γ_{k-1} and γ_k. Also note that after execution of the algorithm, we have

$$\hat{Q}(\varphi(i, \xi_m)) = [\hat{Q}(\varphi(i, \xi_{m-1})) + 1] \bmod \Lambda, \quad \forall i \in \mathcal{S}, \forall m \in \{1, 2, \ldots, \Lambda - 1\}, \tag{5.8}$$

so that (5.7) is satisfied. As a consequence of (5.8), the input symbol x_t can be uniquely determined from $\hat{Q}(y_t)$, given the channel state s_{t-1}. Since $\sigma^2 = 0$, we have

$$I(X_t; \hat{Q}(Y_t)|S_{t-1}) = \log_2(\Lambda), \quad t = 1, 2, \ldots, n, \tag{5.9}$$

and $\mathcal{M}_{\hat{Q}}^{\mathcal{X}}(\boldsymbol{h})$ is information lossless of order $\mu = 1$; consequently, $\mathcal{M}_{\hat{Q}}^{\mathcal{X}}(\boldsymbol{h})$ is also information lossless, and we can conclude that $I(X; \hat{Q}(Y)) = \log_2(\Lambda)$ based on Theorem 5.1. It remains to construct a quantization function $\tilde{Q}_1$ from the discrete mapping $\hat{Q}$. Since the channel is noiseless, $\tilde{Q}_1$ can be constructed from $\hat{Q}$ by introducing a decision threshold at $(\gamma_{k-1}+\gamma_k)/2$ if $\hat{Q}(\gamma_{k-1}) \neq \hat{Q}(\gamma_k)$, $k = 1, 2, \ldots, \Gamma - 1$, so that we have $I(X; \tilde{Q}_1(Y)) = \log_2(\Lambda)$.

The proof of the second part of the theorem is relegated to Appendix D.2. ■

Example 5.1. Note that $\tilde{Q}_1$ corresponds to a quantizer whose quantization regions are possibly *discontiguous*. To see this, consider the application of the preceding algorithm to the design of a single-bit quantizer for $\mathcal{X} = \{\pm 1\}$ and $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$, yielding

$$\tilde{Q}_1(y) = \begin{cases} 0 & \text{if } y \in \Big\{(-\infty, -2/3] \bigcup [2/3, \infty)\Big\} \\ 1 & \text{if } y \in (-2/3, 2/3). \end{cases} \tag{5.10}$$

For the same channel and signaling alphabet, the single-bit uniform quantizer $Q_{\mathrm{u}}(y)$ is a slicer, and for $\sigma^2 = 0$, we numerically compute $I(X; Q_{\mathrm{u}}(Y)) = 0.762 < \log_2(\Lambda) = 1$.

5.2.2. Optimal 1-bit/sample quantization is often necessary for i.i.d. BPSK

Next, we show that the above choice of $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$ is not an isolated example that has rate loss at high SNR and i.i.d. BPSK signaling ($\mathcal{X} = \{\pm 1\}$) when a single-bit uniform quantizer – given by a slicer $Q_{\mathrm{s}}(y) \triangleq \mathbb{1}_{y \geq 0}$ for BPSK – is used at the receiver; in fact, that rate loss turns out be a common effect. For the simulation, we randomly generate real-valued impulse responses of lengths $L_h \in \{2, 3, 4, 5, 6, 7\}$, and investigate whether or not the corresponding finite-state machines are information lossless. More precisely, we generate the random vector $\tilde{\boldsymbol{H}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{1}_{L_h})$, and compute $\boldsymbol{H} = \tilde{\boldsymbol{H}}/\|\tilde{\boldsymbol{H}}\|$. Then, given a realization $\boldsymbol{h}$ of $\boldsymbol{H}$, we take the state diagram of $\mathcal{M}_{Q_{\mathrm{s}}}^{\mathrm{BPSK}}(\boldsymbol{h})$ and construct its testing graph[1] [Eve65], from which information losslessness can be readily verified [Eve65, Theorem 2]. If

[1] The definition of the testing graph of a finite-state machine is provided in the proof of Theorem 5.1 in Appendix D.1, which also includes an example.

the channel is information lossless, we can conclude that $\Upsilon_{\mathrm{s}}(\boldsymbol{h}) \triangleq I(X;Q_{\mathrm{s}}(Y)|\boldsymbol{H}=\boldsymbol{h}) = 1$ based on Theorem 5.1; otherwise, we resort to the algorithm of [ALV+06] to numerically compute $\Upsilon_{\mathrm{s}}(\boldsymbol{h}) < 1$.

Figure 5.2 shows the simulated cumulative distribution functions (CDFs) $F_{\Upsilon_{\mathrm{s}}(\boldsymbol{H})}(\upsilon) = \Pr[\Upsilon_{\mathrm{s}}(\boldsymbol{H}) \leq \upsilon]$, so that for a particular L_h, the height of the jump of the CDF at $\upsilon = 1$ is equal to the fraction of channels of that length which are information lossless. Channels of length $L_h = 2$ turn out to always be information lossless[2] in our model, and we observe that for channel length $L_h = 3$ only about 65 % of the channels have that property; additionally, the CDF of the achievable channel rates is a step-function, suggesting that only distinct information rates of approximately 0.76, 0.86, and 1 are achievable. With increasing length of the channel, the CDF gradually becomes a smooth function. Note, however, that for $L_h = 5$, only about 15 % percent of the channels are information lossless, which shows the advantage of the proposed single-bit quantizers that render $\mathcal{M}_{Q_1}^{\mathrm{BPSK}}(\boldsymbol{h})$ information lossless. In Section 5.6.4, we provide further numerical results suggesting that if $\boldsymbol{h}$ is such that $\mathcal{M}_{Q_{\mathrm{s}}}^{\mathrm{BPSK}}(\boldsymbol{h})$ is information lossy, that rate loss is hard to reduce, even if Λ and the input memory are increased, and the distribution $P_{X_k|X_{k-M_x}^{k-1}}(x_k|x_{k-M_x}^{k-1})$ is optimized.

5.3. Design of scalar A/D converters

5.3.1. Problem formulation

In this section we are interested in computing

$$\sup_{Q_1:\mathbb{R}\to\mathcal{Z}} I(X;Q_1(Y)) \tag{5.11}$$

for a fixed signaling alphabet $\mathcal{X}$, distribution, and channel. Suppose the input is i.i.d., i.e., $M_x = 0$ and $P_{X_k|X^{k-1}}(x_k|x^{k-1}) = P_{X_k}(x_k)$. We first obtain a lower bound on $I(X;Q_1(Y))$.

Lemma 5.3. Assume the channel has length L_h, let the channel input be i.i.d., and let $K \geq 0$ be an integer. Then we have

$$I(X;Q_1(Y)) \geq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(X_i; Z_i^{i+K}|X_{i-L_h+1}^{i-1}). \tag{5.12}$$

Proof. See Appendix D.3. ■

Based on Lemma 5.3, we pose the problem of computing a mutual information preserving

[2]The only impulse response of length $L_h = 2$ we were able to find that is not information lossless with a slicer has the form $h_1 = \pm(1/\sqrt{2})$, $h_2 = \pm(1/\sqrt{2})$, which, however, has zero probability in our probabilistic channel model. The channel $\boldsymbol{h} = [1,-1]^{\mathrm{T}}/\sqrt{2}$ is the normalized version of the Dicode channel often encountered in magnetic recording [KP75].

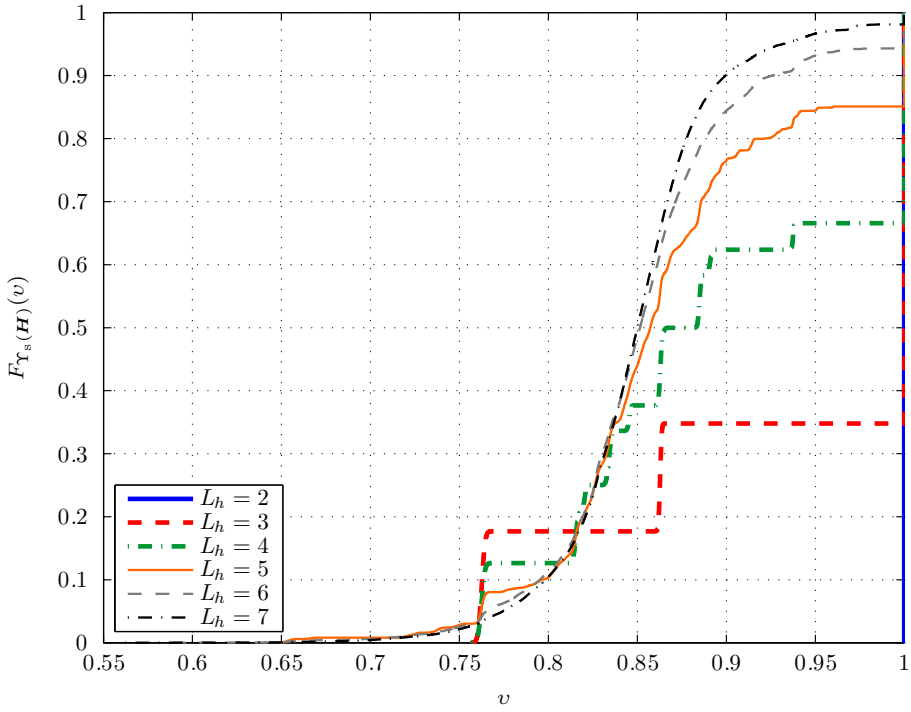


Figure 5.2.: The simulated CDF $F_{\Upsilon_s(\boldsymbol{H})}(v) = \Pr[\Upsilon_s(\boldsymbol{H}) \leq v]$ of information rates for various channel lengths.

quantizer with J regions as

$$I^* = \sup_{Q_1:\mathbb{R}\to\mathcal{Z}} I(X_i; Z_i^{i+K} | X_{i-L_h+1}^{i-1}) \quad \text{s.t. } |\mathcal{Z}| = J. \tag{5.13}$$

The lower bound of Lemma 5.3 becomes increasingly tight with increasing K; we shall see in Section 5.6 that choosing K on the order of L_h suffices in many cases.

5.3.2. Design algorithm

Problem (5.13) seems hard to solve since it is a functional optimization over the quantization function Q_1. We derive an algorithm to solve (5.13) approximately. First, the channel output $Y_i \in \mathbb{R}$ is discretized with high resolution, yielding a discrete approximation $\bar{Y}_i \in \mathcal{Y}$, where $\mathcal{Y}$ is a finite set. With $\bar{Z}_i = \bar{Q}_1(\bar{Y}_i)$, an approximation of (5.13) becomes

$$\bar{I}^* = \max_{\bar{Q}_1:\mathcal{Y}\to\mathcal{Z}} I(X_i; \bar{Z}_i^{i+K} | X_{i-L_h+1}^{i-1}) \quad \text{s.t. } |\mathcal{Z}| = J. \tag{5.14}$$

Next, the mapping $\bar{Q}_1$ is written as the conditional probability mass function

$$q(z|y) = \begin{cases} 1 & \text{if } z = \bar{Q}_1(y) \\ 0 & \text{otherwise,} \end{cases} \tag{5.15}$$

so that the maximization (5.14) changes to

$$\begin{aligned} q^*(z|y) = \underset{q(z|y)}{\operatorname{argmax}} \; I(X_i; \bar{Z}_i^{i+K} | X_{i-L_h+1}^{i-1}) \quad \text{s.t.} \quad & q(z|y) \in \{0,1\}, \forall z \in \mathcal{Z}, \forall y \in \mathcal{Y} \\ & \sum_{z \in \mathcal{Z}} q(z|y) = 1, \forall y \in \mathcal{Y} \\ & z \in \{0, 1, \ldots, J-1\}. \end{aligned} \tag{5.16}$$

By definition, $X_{i-L_h+1}^{i} \leftrightarrow \bar{Y}_i^{i+K} \leftrightarrow \bar{Z}_i^{i+K}$ forms a Markov chain; therefore, we have

$$I(X_{i-L_h+1}^{i}; \bar{Z}_i^{i+K} | \bar{Y}_i^{i+K}) = I(X_{i-L_h+1}^{i-1}; \bar{Z}_i^{i+K} | \bar{Y}_i^{i+K}) = 0, \tag{5.17}$$

so that

$$I(X_{i-L_h+1}^{i}; \bar{Z}_i^{i+K}) = I(X_{i-L_h+1}^{i}; \bar{Y}_i^{i+K}) - I(X_{i-L_h+1}^{i}; \bar{Y}_i^{i+K} | \bar{Z}_i^{i+K}) \tag{5.18}$$

$$I(X_{i-L_h+1}^{i-1}; \bar{Z}_i^{i+K}) = I(X_{i-L_h+1}^{i-1}; \bar{Y}_i^{i+K}) - I(X_{i-L_h+1}^{i-1}; \bar{Y}_i^{i+K} | \bar{Z}_i^{i+K}). \tag{5.19}$$

Consequently, we have

$$I(X_i; \bar{Z}_i^{i+K} | X_{i-L_h+1}^{i-1}) = I(X_{i-L_h+1}^{i}; \bar{Z}_i^{i+K}) - I(X_{i-L_h+1}^{i-1}; \bar{Z}_i^{i+K}) \tag{5.20}$$

$$= I(X_{i-L_h+1}^{i}; \bar{Y}_i^{i+K}) - I(X_{i-L_h+1}^{i}; \bar{Y}_i^{i+K} | \bar{Z}_i^{i+K}) \tag{5.21}$$

$$- I(X_{i-L_h+1}^{i-1}; \bar{Y}_i^{i+K}) + I(X_{i-L_h+1}^{i-1}; \bar{Y}_i^{i+K} | \bar{Z}_i^{i+K}), \tag{5.22}$$

and since $I(X_{i-L_h+1}^{i}; \bar{Y}_i^{i+K})$ and $I(X_{i-L_h+1}^{i-1}; \bar{Y}_i^{i+K})$ are not subject to the optimization over $q(z|y)$, the maximization in (5.16) is equivalent to minimizing

$$I(X_{i-L_h+1}^{i}; \bar{Y}_i^{i+K} | \bar{Z}_i^{i+K}) - I(X_{i-L_h+1}^{i-1}; \bar{Y}_i^{i+K} | \bar{Z}_i^{i+K}). \tag{5.23}$$

Defining

$$r_{j'}^{j}(x_{j'}^{j} | y_i^{i+K}) = P_{X_{j'}^{j} | \bar{Y}_i^{i+K}}(x_{j'}^{j} | y_i^{i+K}) \tag{5.24}$$

$$t_{j'}^{j}(x_{j'}^{j} | z_i^{i+K}) = P_{X_{j'}^{j} | \bar{Z}_i^{i+K}}(x_{j'}^{j} | z_i^{i+K}) \tag{5.25}$$

and

$$\begin{aligned} & d(y_i^{i+K}, z_i^{i+K}) \\ & = D_{\mathrm{KL}}\left(r_{i-L_h+1}^{i}(\cdot|y_i^{i+K}) || t_{i-L_h+1}^{i}(\cdot|z_i^{i+K})\right) - D_{\mathrm{KL}}\left(r_{i-L_h+1}^{i-1}(\cdot|y_i^{i+K}) || t_{i-L_h+1}^{i-1}(\cdot|z_i^{i+K})\right), \end{aligned} \tag{5.26}$$

where $D_{\mathrm{KL}}(\cdot||\cdot)$ denotes relative entropy, we obtain

$$I(X_{i-L_h+1}^{i};\bar{Y}_i^{i+K}|\bar{Z}_i^{i+K}) - I(X_{i-L_h+1}^{i-1};\bar{Y}_i^{i+K}|\bar{Z}_i^{i+K})$$
$$= \sum_{y_i,z_i} q(z_i|y_i)P_{\bar{Y}_i}(y_i) \cdot \sum_{\substack{y_{i+K},\ldots,y_{i+1}\\ z_{i+K},\ldots,z_{i+1}}} P_{\bar{Y}_{i+1}^{i+K}|\bar{Y}_i}(y_{i+1}^{i+K}|y_i)d(y_i^{i+K},z_i^{i+K}) \prod_{j=i+1}^{i+K} q(z_j|y_j). \quad (5.27)$$

Based on (5.27) we use Algorithm 5.1 to compute a mapping $q(z|y)$ that yields a large mutual information $I(X_i;\bar{Z}_i^{i+K}|X_{i-L_h+1}^{i-1})$. Our algorithm can be viewed as a modification of the information bottleneck algorithm [TPB99] accounting for the conditioning on $X_{i-L_h+1}^{i-1}$ in (5.16) and the parameter K. However, since $I(X_i;\bar{Z}_i^{i+K}|X_{i-L_h+1}^{i-1})$ is non-convex in the mapping $q(z|y)$, the algorithm is repeatedly carried out with different initializations to yield a satisfactory solution.

Algorithm 5.1 Design algorithm for scalar quantizers.

Notation: $\boldsymbol{X} = [X_i,\ldots,X_{i-L_h+1}]^{\mathrm{T}}$, $\tilde{\boldsymbol{X}} = [X_{i-1},\ldots,X_{i-L_h+1}]^{\mathrm{T}}$, $\bar{\boldsymbol{Y}} = [\bar{Y}_{i+K},\ldots,\bar{Y}_i]^{\mathrm{T}}$, $\bar{\boldsymbol{Z}} = [\bar{Z}_{i+K},\ldots,\bar{Z}_i]^{\mathrm{T}}$

Input: $P_{\boldsymbol{X}\bar{\boldsymbol{Y}}}(\boldsymbol{x},\boldsymbol{y})$, J, K, $\epsilon > 0$

Initialization: randomly choose a valid mapping $q^{(\mathrm{old})}(z_i|y_i)$, $T \leftarrow 1$,
$r(\boldsymbol{x}|\boldsymbol{y}) \leftarrow P_{\boldsymbol{X}\bar{\boldsymbol{Y}}}(\boldsymbol{x},\boldsymbol{y})/P_{\bar{\boldsymbol{Y}}}(\boldsymbol{y})$, $\tilde{r}(\tilde{\boldsymbol{x}}|\boldsymbol{y}) \leftarrow \sum_{x_i} r(\boldsymbol{x}|\boldsymbol{y})$

loop

$P_{\bar{\boldsymbol{Z}}|\bar{\boldsymbol{Y}}}(\boldsymbol{z}|\boldsymbol{y}) \leftarrow \prod_{j=i}^{i+K} q^{(\mathrm{old})}(z_j|y_j)$

$P_{\bar{\boldsymbol{Z}}}(\boldsymbol{z}) \leftarrow \sum_{\boldsymbol{y}} P_{\bar{\boldsymbol{Y}}}(\boldsymbol{y})P_{\bar{\boldsymbol{Z}}|\bar{\boldsymbol{Y}}}(\boldsymbol{z}|\boldsymbol{y})$

$t(\boldsymbol{x}|\boldsymbol{z}) \leftarrow \Big(\sum_{\boldsymbol{y}} P_{\boldsymbol{X}\bar{\boldsymbol{Y}}}(\boldsymbol{x},\boldsymbol{y})P_{\bar{\boldsymbol{Z}}|\bar{\boldsymbol{Y}}}(\boldsymbol{z}|\boldsymbol{y})\Big)/P_{\bar{\boldsymbol{Z}}}(\boldsymbol{z})$, $\tilde{t}(\tilde{\boldsymbol{x}}|\boldsymbol{z}) \leftarrow \sum_{x_i} t(\boldsymbol{x}|\boldsymbol{z})$

if $T = 0$ **then**

return $q^{(\mathrm{old})}$

end if

$D(\boldsymbol{y},\boldsymbol{z}) \leftarrow D_{\mathrm{KL}}(r(\cdot|\boldsymbol{y})||t(\cdot|\boldsymbol{z})) - D_{\mathrm{KL}}(\tilde{r}(\cdot|\boldsymbol{y})||\tilde{t}(\cdot|\boldsymbol{z}))$

$d(y_i,z_i) \leftarrow \sum_{\substack{y_{i+K},\ldots,y_{i+1}\\ z_{i+K},\ldots,z_{i+1}}} P_{\bar{Y}_{i+1}^{i+K}|\bar{Y}_i}(y_{i+1}^{i+K}|y_i)D(\boldsymbol{y},\boldsymbol{z}) \prod_{j=i+1}^{i+K} q^{(\mathrm{old})}(z_j|y_j)$

find, for each y_i, $z^*(y_i) = \mathrm{argmin}_{z_i} d(y_i,z_i)$,
and set $q^{(\mathrm{new})}(z_i|y_i) \leftarrow \mathbb{1}_{z_i=z^*(y_i)}$

if $\sum_{y_i,z_i} \left|q^{(\mathrm{new})}(z_i|y_i) - q^{(\mathrm{old})}(z_i|y_i)\right| / (J \cdot |\mathcal{Y}|) < \epsilon$ **then**

$T \leftarrow 0$

end if

$q^{(\mathrm{old})}(z_i|y_i) \leftarrow q^{(\mathrm{new})}(z_i|y_i)$

end loop

5.4. Design of two-dimensional A/D converters

Depending on the length L_h of the channel, the source $\{Y_k\}$ to be quantized at the receiver may exhibit considerable correlation; that correlation, however, cannot be exploited by a scalar quantizer [GN98, Section IV-E]. Therefore, we consider vector ADCs, and restrict ourselves to two-dimensional quantization for the sake of simplicity. Again, we assume the channel input to consist of i.i.d. symbols.

Analogous to the derivations in Section 5.3.1, the goal is to compute a two-dimensional quantization function such that a lower bound on the achievable information rate is maximized. Along the lines of the proof of Lemma 5.3 we obtain

$$\begin{aligned} I(X;Q_2(Y)) &= \lim_{n\to\infty} \frac{1}{n} I(X^n; Z^{\frac{n}{2}}|S_0) \\ &\geq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n/2} I(X_{2i-1}^{2i}; Z_i^{i+K}|X_{2i-L_h}^{2i-2}), \end{aligned} \tag{5.28}$$

based on which we pose the quantizer design problem as

$$I^* = \sup_{Q_2:\mathbb{R}^2\to\mathcal{Z}} I(X_{2i-1}^{2i}; Z_i^{i+K}|X_{2i-L_h}^{2i-2}) \quad \text{s.t. } |\mathcal{Z}| = J. \tag{5.29}$$

We proceed as in Section 5.3 and discretize Y_i with high resolution yielding $\bar{Y}_i$ from the finite set $\mathcal{Y}$, and write $\bar{Z}_i = \bar{Q}_2(\bar{Y}_{2i}, \bar{Y}_{2i-1})$. Defining

$$q(z|\boldsymbol{y}) = \begin{cases} 1 & \text{if } z = \bar{Q}_2(\boldsymbol{y}) \\ 0 & \text{otherwise,} \end{cases} \tag{5.30}$$

for $\boldsymbol{y} \in (\mathcal{Y}\times\mathcal{Y})$, we arrive at the approximation of (5.29) as

$$\begin{aligned} q^*(z|\boldsymbol{y}) = \underset{q(z|\boldsymbol{y})}{\operatorname{argmax}} \; I(X_{2i-1}^{2i}; \bar{Z}_i^{i+K}|X_{2i-L_h}^{2i-2}) \quad \text{s.t.} \quad & q(z|\boldsymbol{y}) \in \{0,1\}, \forall z \in \mathcal{Z}, \forall \boldsymbol{y} \in (\mathcal{Y}\times\mathcal{Y}) \\ & \sum_{z\in\mathcal{Z}} q(z|\boldsymbol{y}) = 1, \forall \boldsymbol{y} \in (\mathcal{Y}\times\mathcal{Y}) \\ & z \in \{0,1,\ldots,J-1\}, \end{aligned} \tag{5.31}$$

which is solved with an appropriately modified version of the design algorithm for scalar quantizers to incorporate the altered cost function and the two-dimensional nature of the quantization mapping. We show the algorithm in Algorithm 5.2.

We remark that the best two-dimensional quantizer can perform no worse than the optimal scalar quantizer while keeping the number of bits per sample constant. To see this, let Q_1^* be the optimizer of (5.13), and construct Q_2 with J^2 regions as the product quantizer [Gra90, Section V-B] of Q_1^*, so that applying Q_2 to (Y_{2k}, Y_{2k-1}) is equivalent to two successive applications of Q_1^* to the sequence Y_{2k-1}^{2k}.

Algorithm 5.2 Design algorithm for two-dimensional quantizers.

Notation: $\boldsymbol{X} = [X_{2i}, \dots, X_{2i-L_h}]^{\mathrm{T}}$, $\tilde{\boldsymbol{X}} = [X_{2i-2}, \dots, X_{2i-L_h}]^{\mathrm{T}}$, $\bar{\boldsymbol{Y}} = [\bar{Y}_{2i+2K}, \dots, \bar{Y}_{2i-1}]^{\mathrm{T}}$, $\bar{\boldsymbol{Z}} = [\bar{Z}_{i+K}, \dots, \bar{Z}_i]^{\mathrm{T}}$, $\bar{\boldsymbol{Y}}_i = [\bar{Y}_{2i}, \bar{Y}_{2i-1}]^{\mathrm{T}}$
Input: $P_{\boldsymbol{X}\bar{\boldsymbol{Y}}}(\boldsymbol{x}, \boldsymbol{y})$, J, K, $\epsilon > 0$
Initialization: randomly choose a valid mapping $q^{(\mathrm{old})}(z_i|\boldsymbol{y}_i)$, $T \leftarrow 1$,
$r(\boldsymbol{x}|\boldsymbol{y}) \leftarrow P_{\boldsymbol{X}\bar{\boldsymbol{Y}}}(\boldsymbol{x}, \boldsymbol{y}) / P_{\bar{\boldsymbol{Y}}}(\boldsymbol{y})$, $\quad \tilde{r}(\tilde{\boldsymbol{x}}|\boldsymbol{y}) \leftarrow \sum_{x_{2i}, x_{2i-1}} r(\boldsymbol{x}|\boldsymbol{y})$
loop
  $P_{\bar{\boldsymbol{Z}}|\bar{\boldsymbol{Y}}}(\boldsymbol{z}|\boldsymbol{y}) \leftarrow \prod_{j=i}^{i+K} q^{(\mathrm{old})}(z_j|\boldsymbol{y}_j)$
  $P_{\bar{\boldsymbol{Z}}}(\boldsymbol{z}) \leftarrow \sum_{\boldsymbol{y}} P_{\bar{\boldsymbol{Y}}}(\boldsymbol{y}) P_{\bar{\boldsymbol{Z}}|\bar{\boldsymbol{Y}}}(\boldsymbol{z}|\boldsymbol{y})$
  $t(\boldsymbol{x}|\boldsymbol{z}) \leftarrow \Big(\sum_{\boldsymbol{y}} P_{\boldsymbol{X}\bar{\boldsymbol{Y}}}(\boldsymbol{x}, \boldsymbol{y}) P_{\bar{\boldsymbol{Z}}|\bar{\boldsymbol{Y}}}(\boldsymbol{z}|\boldsymbol{y}) \Big) / P_{\bar{\boldsymbol{Z}}}(\boldsymbol{z}), \quad \tilde{t}(\tilde{\boldsymbol{x}}|\boldsymbol{z}) \leftarrow \sum_{x_{2i}, x_{2i-1}} t(\boldsymbol{x}|\boldsymbol{z})$
  if $T = 0$ **then**
    return $q^{(\mathrm{old})}$
  end if
  $D(\boldsymbol{y}, \boldsymbol{z}) \leftarrow D_{\mathrm{KL}}(r(\cdot|\boldsymbol{y}) || t(\cdot|\boldsymbol{z})) - D_{\mathrm{KL}}(\tilde{r}(\cdot|\boldsymbol{y}) || \tilde{t}(\cdot|\boldsymbol{z}))$
  $d(\boldsymbol{y}_i, z_i) \leftarrow \sum_{\substack{y_{2i+2K}, \dots, y_{2i+1} \\ z_{i+K}, \dots, z_{i+1}}} P_{\bar{Y}_{2i+1}^{2i+2K}|\bar{\boldsymbol{Y}}_i}(y_{2i+1}^{2i+2K}|\boldsymbol{y}_i) D(\boldsymbol{y}, \boldsymbol{z}) \prod_{j=i+1}^{i+K} q^{(\mathrm{old})}(z_j|\boldsymbol{y}_j)$
  find, for each $\boldsymbol{y}_i$, $z^*(\boldsymbol{y}_i) = \mathrm{argmin}_{z_i} d(\boldsymbol{y}_i, z_i)$,
  and set $q^{(\mathrm{new})}(z_i|\boldsymbol{y}_i) \leftarrow \mathbb{1}_{z_i = z^*(\boldsymbol{y}_i)}$
  if $\sum_{\boldsymbol{y}_i, z_i} \left| q^{(\mathrm{new})}(z_i|\boldsymbol{y}_i) - q^{(\mathrm{old})}(z_i|\boldsymbol{y}_i) \right| / (J \cdot |\mathcal{Y}|^2) < \epsilon$ **then**
    $T \leftarrow 0$
  end if
  $q^{(\mathrm{old})}(z_i|\boldsymbol{y}_i) \leftarrow q^{(\mathrm{new})}(z_i|\boldsymbol{y}_i)$
end loop

5.5. Upper bound on the information rate

To better assess the performance of the scalar quantizers designed in Section 5.3.2, we derive an upper bound on the information rate for i.i.d. inputs.

Lemma 5.4. For i.i.d. channel inputs and a length-L_h channel, we have

$$\sup_{Q_1 : \mathbb{R} \to \mathcal{Z}} I(X; Q_1(Y)) \leq \sup_{Q_1 : \mathbb{R} \to \mathcal{Z}} I(X_i; Z_i^{i+L_h-1} | X_{i-L_h+1}^{i-1}, X_{i+1}^{i+L_h-1}). \tag{5.32}$$

Proof. See Appendix D.3. ■

We compute an approximation to (5.32) by first discretizing Y_i at high resolution and writing Q_1 as a conditional mass function $q(z|y)$; then, we use an appropriately adapted

version of Algorithm 5.1 to seek the mass function $q(z|y)$ that minimizes

$$I(X_{i-L_h+1}^{i+L_h-1}; \bar{Y}_i^{i+L_h-1}|\bar{Z}_i^{i+L_h-1}) - I(X_{i-L_h+1}^{i-1}, X_{i+1}^{i+L_h-1}; \bar{Y}_i^{i+L_h-1}|\bar{Z}_i^{i+L_h-1}), \tag{5.33}$$

and therefore maximizes $I(X_i; \bar{Z}_i^{i+L_h-1}|X_{i-L_h+1}^{i-1}, X_{i+1}^{i+L_h-1})$. The only modifications to Algorithm 5.1 are to define $\boldsymbol{X} = [X_{i+L_h-1}, \ldots, X_{i-L_h+1}]^{\mathrm{T}}$, $\tilde{\boldsymbol{X}} = [X_{i+L_h-1}, \ldots, X_{i+1}, X_{i-1}, \ldots, X_{i-L_h+1}]^{\mathrm{T}}$, $\bar{\boldsymbol{Y}} = [\bar{Y}_{i+L_h-1}, \ldots, \bar{Y}_i]^{\mathrm{T}}$, and $\bar{\boldsymbol{Z}} = [\bar{Z}_{i+L_h-1}, \ldots, \bar{Z}_i]^{\mathrm{T}}$. A tighter upper bound on $I(X; Q_1(Y))$ than the one in Lemma 5.4 can be derived by not conditioning on X_{i+1}^n in (D.37) in Appendix D.3, but only a subset thereof; however, the subsequent optimization over Q_1 then becomes increasingly computationally complex, especially for large L_h and $|\mathcal{Y}|$.

5.6. Simulation results

5.6.1. Examples of quantizers

Consider BPSK modulation ($\mathcal{X} = \{\pm 1\}$) with i.i.d. symbols for the channel $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$. Table 5.1 shows the characteristics of scalar 1-bit/sample ($J = 2$) and 2-bit/sample ($J = 4$) quantizers for various SNRs and different values of K. In addition to the quantizer threshold(s), the table also shows the quantization indices corresponding to each region and the achievable rate $I(X; Q_1(Y))$. First consider the 1-bit/sample quantizers, for which the algorithm produces a quantizer with a single threshold at zero at $E_s/N_0 = 0\,\mathrm{dB}$, for $K \in \{0, 1, 2, 3\}$. This quantizer is the 1-bit LM and the UF quantizer. In contrast, for $E_s/N_0 = 10\,\mathrm{dB}$ and $K \in \{0, 1, 2, 3\}$, a splitting of the quantization regions occurs, so that the resulting quantizer is not characterized by a single threshold. To complete the discussion, consider the resulting quantizers at $E_s/N_0 = 4\,\mathrm{dB}$, where the result of the optimization depends on K. Since the tightness of the lower bound for which the quantizer is optimized increases with K, it is the quantizer computed with $K \in \{2, 3\}$ that outperforms the one designed for $K \in \{0, 1\}$ at $E_s/N_0 = 4\,\mathrm{dB}$ in terms of channel rate $I(X; Q_1(Y))$, cf. Section 5.6.2. We remark that the choice of $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$ is not a construed example to show the splitting of the regions at high SNR; in fact, channels of the form $\boldsymbol{h'} = [1, -0.5, \alpha]^{\mathrm{T}}/\sqrt{1.25 + \alpha^2}$, $0.5 < \alpha < 1.5$, require discontiguous quantization regions for $J = 2$ to achieve one bit per channel use at high SNR (choosing $\alpha = 1$ gives $\boldsymbol{h'} = [2/3, -1/3, 2/3]^{\mathrm{T}}$).

We observe a similar effect for the 2-bit/sample quantizers with characteristics in Table 5.1: while the optimal quantization regions are contiguous at $E_s/N_0 = 0\,\mathrm{dB}$ and discontiguous at $E_s/N_0 = 10\,\mathrm{dB}$ for $K \in \{0, 1, 2, 3\}$, the result depends on K at $E_s/N_0 = 6\,\mathrm{dB}$.

As a second set of examples, consider the 1-bit/sample ($J = 4$, $K = 1$) two-dimensional quantizers shown in Figure 5.3, in which the quantization regions are color-coded. Similar to the scalar case, we observe that the quantization regions are contiguous at low SNR, whereas they are not for high SNR. Also observe that the decision boundaries of the proposed quantizer are not hyperplanes in general, in contrast to the MSE-minimizing

Table 5.1.: Scalar ADCs for $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$.

J	E_s/N_0	K	Threshold(s)	Region indices	$I(X;Q_1(Y))$
2	0 dB	$\in \{0,1,2,3\}$	0	$[0,1]$	0.375
2	4 dB	$\in \{0,1\}$	$[-0.923, 0.923]$	$[0,1,0]$	0.440
2	4 dB	$\in \{2,3\}$	0	$[0,1]$	0.497
2	10 dB	$\in \{0,1,2,3\}$	$[-0.8, 0.8]$	$[0,1,0]$	0.758
4	0 dB	$\in \{0,1,2,3\}$	$[-1.03, 0, 1.03]$	$[0,1,2,3]$	0.563
4	6 dB	$\in \{0,1\}$	$[-1.12, -0.364, 0.364, 1.12]$	$[0,1,2,3,4,0]$	0.898
4	6 dB	$\in \{2,3\}$	$[-0.941, 0, 0.941]$	$[0,1,2,3]$	0.913
4	10 dB	$\in \{0,1,2,3\}$	$[-1.03, -0.338, 0.338, 1.03]$	$[0,1,2,3,0]$	0.997

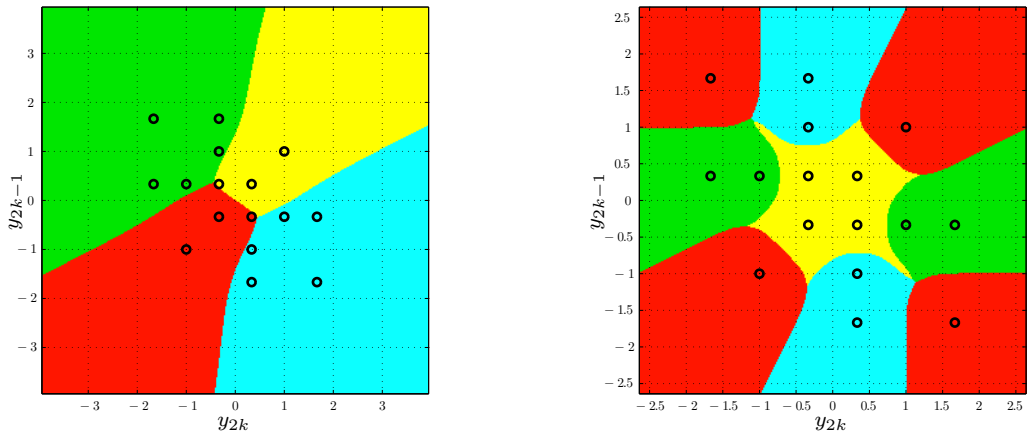


Figure 5.3.: Two-dimensional 1-bit/sample ($J = 4$, $K = 1$) quantizers for $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$. Left: $E_s/N_0 = 1\,\mathrm{dB}$, right: $E_s/N_0 = 9\,\mathrm{dB}$.

LM vector quantizer, whose decision boundaries are given by hyperplanes [GG92, Chapter 10.4].

5.6.2. Achievable rates

Figure 5.4 shows achievable information rates for the channel $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$ under scalar and two-dimensional 1-bit/sample output quantization. Additionally, we plot the rates for the continuous-output channel, and the upper bound from Section 5.5. Observe that the information rates achievable with the LM and UF quantizers saturate at a maximum rate of approximately 0.76, while the quantizers designed with the proposed framework eventually achieve a rate of one bit per channel use at high SNR. Furthermore, the curves for scalar quantizers demonstrate the effect of increasing the parameter K; choosing $K = 4$ resulted in no further gain in terms of rate.

For completeness, we also show information rates for $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$ and 2-bit/sample quantization in Figure 5.5. Here, the performance difference between the pro-

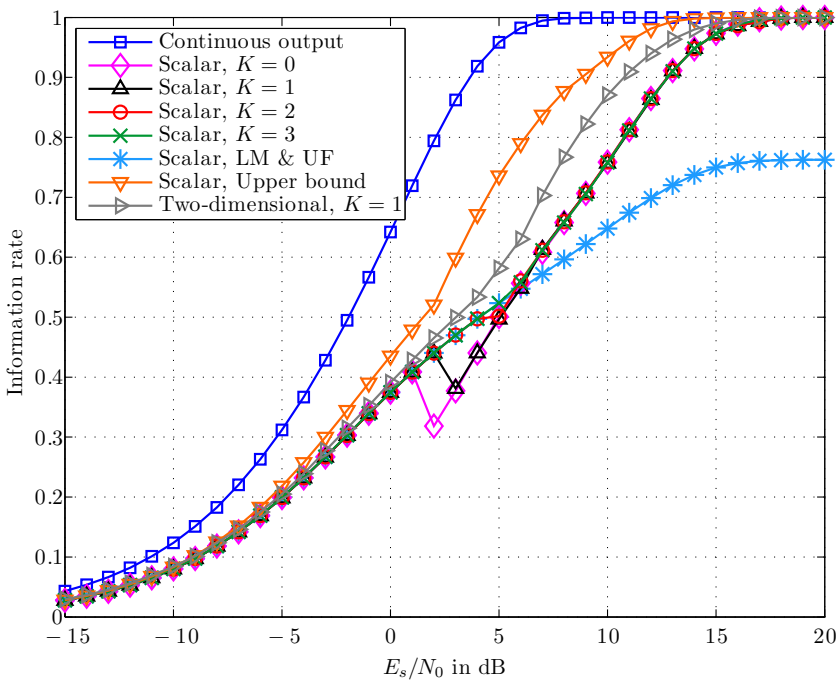


Figure 5.4.: Information rates for $h = [2/3, -1/3, 2/3]^{\mathrm{T}}$ and i.i.d. BPSK signaling. All quantizers have 1 bit/sample.

posed quantizers and the LM and UF quantizer is fairly small. More importantly, the performance loss compared to the channel with unquantized outputs is at most about 0.1 bits per channel use in this example, suggesting that optimized 2-bit/sample quantization can perform close to the limit given by the continuous-output channel.

5.6.3. Error rates

Frame error rates (FERs) obtained for $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$, $n = 10000$, and i.i.d. BPSK are shown in Figure 5.6. In agreement with the proof of Theorem 5.2 and Figure 5.4, the FER of the single-bit LM and UF quantizer remains exactly one, even at very high SNR. In contrast, the 1-bit/sample scalar ($K = 3$) and two-dimensional ($K = 1$) quantizers designed for maximum information rate (MIR) have a loss of approximately 2 dB compared to the 2-bit/sample LM and UF quantizer at an FER of 10^{-5}, and the proposed 2-bit/sample scalar ($K = 3$) quantizer shows a loss of roughly 3.5 dB compared to the continuous-output channel.

5.6.4. Optimization of input alphabet and distribution

In this section, we shift our attention to the choice of the channel input for a fixed quantizer at the output. In particular, a natural question is whether there exists an alphabet $\mathcal{X}$ and an input distribution (5.2) that gives an information rate of one bit per channel use at high

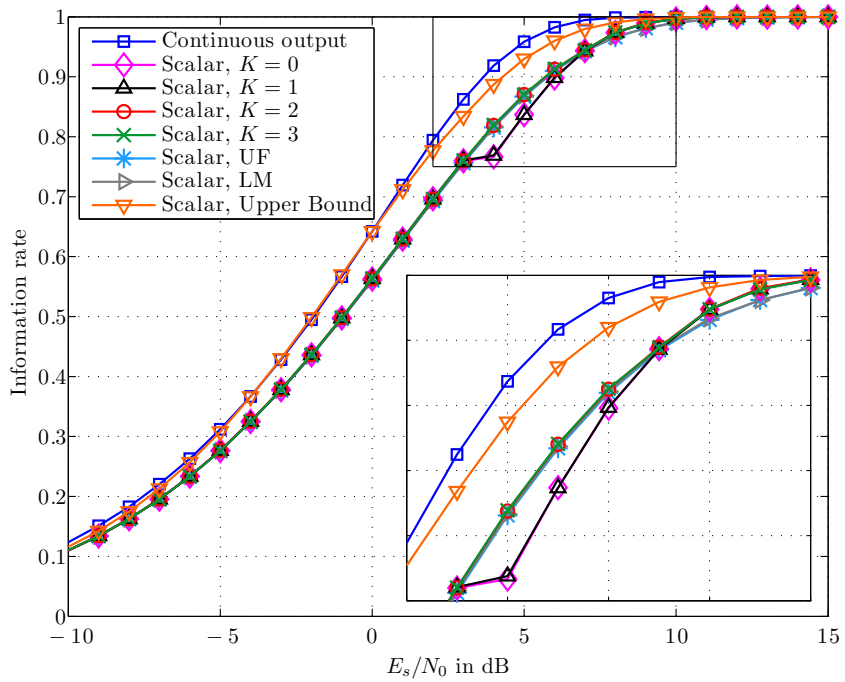


Figure 5.5.: Information rates for $h = [2/3, -1/3, 2/3]^{\mathrm{T}}$ and i.i.d. BPSK signaling. All quantizers have 2 bit/sample.

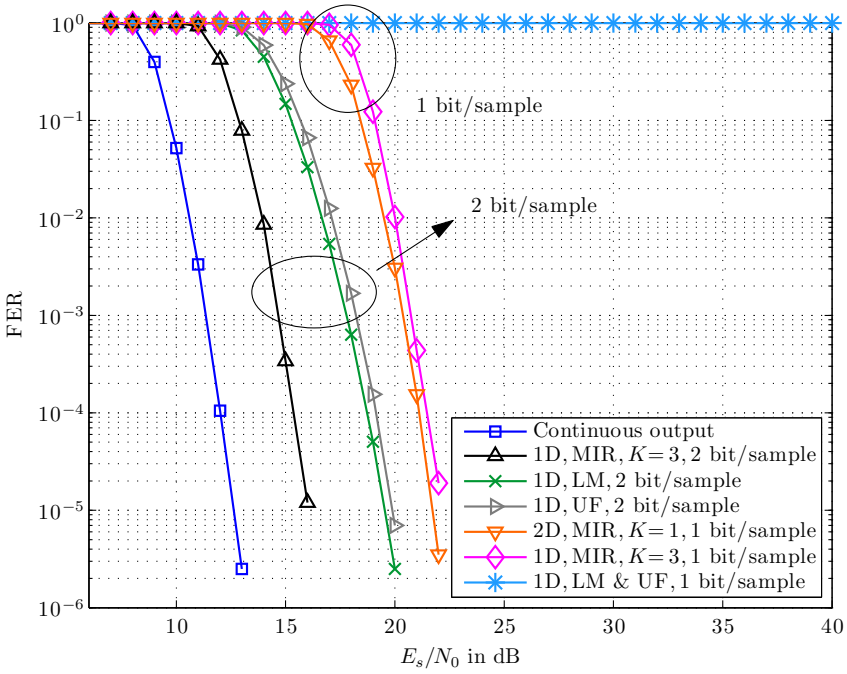


Figure 5.6.: FERs for $h = [2/3, -1/3, 2/3]^{\mathrm{T}}$ and i.i.d. BPSK signaling.

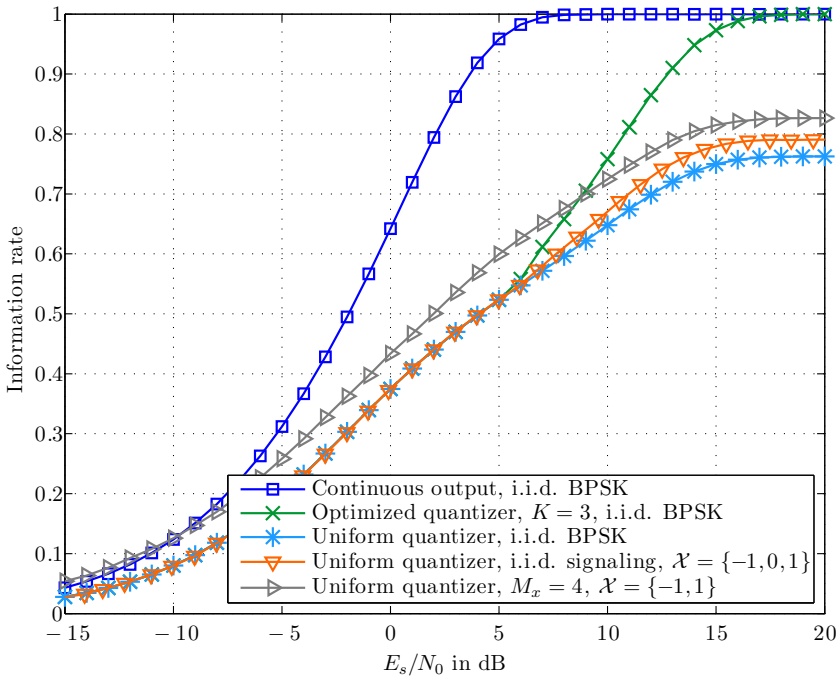


Figure 5.7.: Information rates for $h = [2/3, -1/3, 2/3]^{\mathrm{T}}$. All quantizers have 1 bit/sample.

SNR for a simple slicer with a single threshold at zero, instead of the more complicated design using Algorithm 5.1. The results of our numerical study for $\boldsymbol{h} = [2/3, -1/3, 2/3]^{\mathrm{T}}$ are shown in Figure 5.7. Increasing the size of the input alphabet from BPSK to $\mathcal{X} = \{-1, 0, 1\}$ and subsequent optimization of $P_{X_k}(x_k)$ (i.i.d. signaling) subject to a power constraint using the generalized BAA [VKAL08] gives a marginal gain at high SNR; further, increasing the input memory for BPSK (i.e., the power constraint is met) to $M_x = 4$ and optimizing $P_{X_k|X_{k-M_x}^{k-1}}(x_k|x_{k-M_x}^{k-1})$ (again using the generalized BAA) shows gains over the entire SNR range. However, both schemes exhibit a saturation of the information rate at high SNR, while i.i.d. BPSK signaling using the optimized quantizer clearly performs better at considerably lower encoding and decoding complexity due to the memoryless modulation and a small number of channel states. These numerical findings demonstrate that in contrast to the AWGN channel without ISI under single-bit symmetric output quantization, for which BPSK is optimal [SDM09, Theorem 2], high order modulation schemes can provide a rate gain when the channel has memory.

5.7. Discussion

The information rate of ISI channels under output quantization was studied. For i.i.d. signaling using a transmit alphabet of size Λ and $\log_2(\Lambda)$-bit/sample quantization, we leveraged the theory of information lossless finite-state machines to demonstrate that uniform quantization is suboptimal, and gave a constructive proof that there exists an information-

rate optimal $\log_2(\Lambda)$-bit/sample quantizer for all channels. We then provided a framework for the design of scalar and two-dimensional quantizers (with low precision) at finite SNRs to maximize the information rate over ISI channels, and derived an upper bound on the rate for quantized channel outputs. The quantizer performance is evaluated in terms of information rate as well as in terms of error rate for uncoded transmission. Further, a specific example suggests that a simple slicer cannot achieve an information rate of one bit per channel use at high SNR, even if one increases the memory or alphabet size of the input.

6

Bayesian parameter estimation using single-bit dithered quantization

In Chapter 5, ADCs are designed to maximize the information rate over channels with ISI. However, that chapter and other work on the design of ADCs for receivers of communication systems (e.g., [SDM09, LSS10, SBC10]) assumes perfect CSI at the receiver, and an important question is how to reliably estimate the channel under low-precision output quantization. For such estimation problems, dithered quantizers turn out to be particularly useful [PWO01]. For example, for a Gaussian prior on the channel coefficients, using the linear minimum mean squared error (MMSE) estimate of the channel as a dither signal is shown to work well in practice [DM10]. In this chapter, we also assume the channel to be estimated to be random, i.e., in contrast to [PWO01] we consider a Bayesian parameter estimation problem as in [DM10]. Specifically, our contributions are as follows.

- ▷ For a single-bit adaptively dithered quantizer and a Gaussian prior[1], we first derive lower bounds on the MSE of the channel estimate for a finite number n of quantized received symbols. While the first bound we derive appears to be almost tight for large n, we also derive lower bounds on the MSE that are tighter for small n.

- ▷ We show that the MSE that results from any dither strategy is asymptotically (in the quantizer-output sequence length) at least $10\log_{10}(\pi/2) \approx 1.96\,\text{dB}$ worse than the MSE of the MMSE estimator based on unquantized observations.

[1] We remark that the *algorithms* we present can also be used for unknown parameters by assigning some prior over the parameter space. However, several of our *bounds* are based on Gaussian priors and may not apply more generally.

- ▷ Dither and estimation strategies are designed that closely approach the derived lower bounds over a broad SNR range and that perform well for any number of observations. Among these schemes, the best results are achieved by one that uses an approximated MMSE estimate for estimation combined with an optimized dither signal minimizing the expected error at the next time step, given the observations processed so far.

- ▷ Through simulation, we compare our approach both with the derived lower bounds on the MSE, and with other dither and estimation schemes proposed in the literature.

This chapter is organized as follows. In Section 6.1, we describe the system model. Lower bounds on the MSE are derived in Sections 6.2 and 6.3. Dither strategies and simulation results are presented in Section 6.4 and Section 6.5, respectively. Concluding remarks appear in Section 6.6.

6.1. System model

Consider transmission over the discrete-time channel with ISI and AWGN, so that the channel output at time t is given by

$$Y_t = \sum_{\ell=1}^{L_h} H_\ell X_{t+1-\ell} + N_t, \quad t = 1, 2, \ldots, \tag{6.1}$$

where in contrast to Section 5.1, the channel of length L_h has independent *random* coefficients $H_\ell \sim \mathcal{N}(0, \sigma_h^2)$. The channel input $X_t \in \mathcal{X}$ is real and discrete, and the additive noise $N_t \sim \mathcal{N}(0, \sigma^2)$ satisfies $\mathrm{E}[N_t N_{t'}] = \sigma^2 \mathbb{1}_{t=t'}$. If the length L_h of the channel is known at the receiver, one may use a simple periodic training sequence with period L_h to decompose the channel estimation problem for the channel (6.1) into L_h parallel estimation problems [DM10] for each H_ℓ, $\ell = 1, 2, \ldots, L_h$. More precisely, let

$$X_t = \begin{cases} 1 & \text{if } (t-1) \bmod L_h = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{6.2}$$

Then we have

$$Y_t = H_{1+((t-1) \bmod L_h)} + N_t, \tag{6.3}$$

so that Y_t only depends on $H_{1+((t-1) \bmod L_h)}$, and is independent of all other H_ℓ, $\ell \neq 1+((t-1) \bmod L_h)$. Consequently, the problem of estimating H_ℓ, $\ell = 1, 2, \ldots, L_h$, consists of L_h independent parallel estimation problems, as illustrated in Figure 6.1. Therefore, we only consider one of those L_h estimation problems in the sequel.

Suppose that the real-valued random parameter H is corrupted by AWGN, and the receiver observes

$$Y_i = H + N_i, \tag{6.4}$$

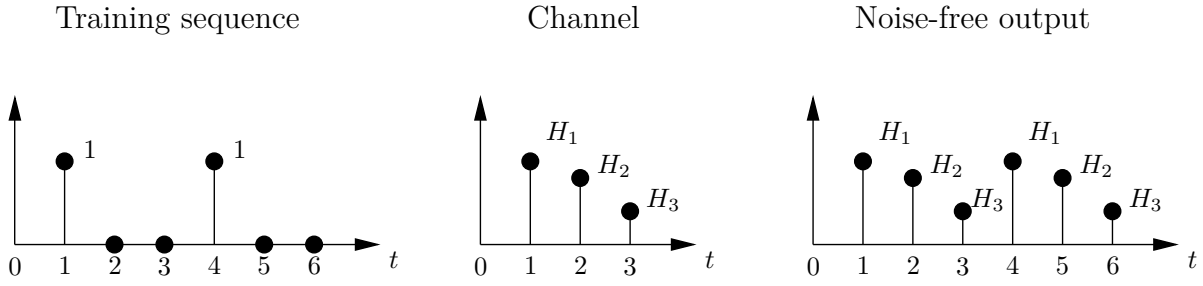


Figure 6.1.: The effect of using a bursty periodic training sequence on a channel of length $L_h = 3$.

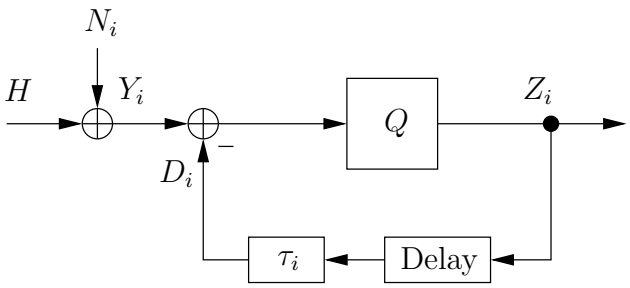


Figure 6.2.: System model.

where $H \sim \mathcal{N}(0, \sigma_h^2)$, $N_i \sim \mathcal{N}(0, \sigma_n^2)$, and the sequence $\{N_i\}$ is i.i.d. and independent of H. The quantizer output Z_i at time i is

$$Z_i = Q(Y_i - D_i), \tag{6.5}$$

where D_i is a real-valued dither signal to be designed, and the quantization function $Q : \mathbb{R} \to \{0, 1\}$ satisfies

$$Q(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 & \text{if } y \geq 0. \end{cases} \tag{6.6}$$

The SNR is defined as σ_h^2/σ_n^2.

It remains to design the dither signal D_i. We permit the receiver to use a strictly causal and adaptive dither signal $D_i = \tau_i(Z^{i-1})$, for some function $\tau_i : \{0, 1\}^{i-1} \to \mathbb{R}$. The system is depicted in Figure 6.2.

Let $\hat{h}(Z^n)$ denote an estimate of H based on the output sequence Z^n of the quantizer. Throughout this chapter, the estimation performance is quantified by the MSE between the estimate and the random parameter H; for an estimate $\hat{h}(Z^n)$, the MSE is $\mathrm{E}[(H - \hat{h}(Z^n))^2]$.

We remark that the model in (6.4) and (6.5), and therefore all the results that follow, apply to a much larger class of estimation problems than the channel estimation problem for channels with temporal ISI. For instance, consider the 2×2 Multiple-Input/Multiple-

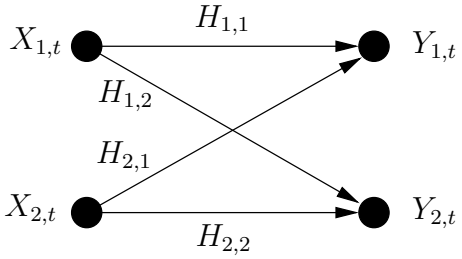


Figure 6.3.: The 2×2 MIMO channel.

Output (MIMO) channel with spatial interference shown in Figure 6.3, where the independent channel coefficients $H_{1,1}$, $H_{1,2}$, $H_{2,1}$, and $H_{2,2}$ shall be estimated using an adaptively dithered single-bit quantizer. By choosing $X_{1,t} = 1$ if t is odd and zero otherwise, and by choosing $X_{2,t} = 1$ if t is even and zero otherwise, the problem of estimating $H_{1,1}$, $H_{1,2}$, $H_{2,1}$, and $H_{2,2}$ can be decomposed into four parallel estimation problems, each of the form of (6.4) and (6.5).

6.2. Bayesian Cramér-Rao lower bounds

The celebrated Cramér-Rao lower bound (CRLB) [Cra46, Rao45], [VT68, p. 66] provides a lower bound on the variance of any unbiased estimate of a *nonrandom* parameter, which is not applicable in our setting due to the random nature of H. However, for the *random* (Bayesian) parameter estimation problem, a similar bound on the MSE known as the Bayesian Cramér-Rao lower bound (BCRLB) [VT68, p. 72] holds under some mild regularity conditions. We state the BCRLB in the following theorem.

Theorem 6.1 (The BCRLB [VT68])**.** Let Θ be a random variable and $\boldsymbol{Y} \in \mathbb{R}^k$ an observation vector, let $p_{\Theta\boldsymbol{Y}}$ be the joint density of Θ and $\boldsymbol{Y}$, and let $\hat{\theta}(\boldsymbol{Y})$ be an estimator of Θ. Suppose the following conditions hold:

1. $\dfrac{\partial p_{\Theta\boldsymbol{Y}}(\theta, \boldsymbol{y})}{\partial \theta}$ exists and is absolutely integrable with respect to θ and $\boldsymbol{y}$.
2. $\dfrac{\partial^2 p_{\Theta\boldsymbol{Y}}(\theta, \boldsymbol{y})}{\partial \theta^2}$ exists and is absolutely integrable with respect to θ and $\boldsymbol{y}$.
3. The conditional expectation of the error, given $\Theta = \theta$, is

$$B(\theta) = \int_{\mathbb{R}^k} \left[\hat{\theta}(\boldsymbol{y}) - \theta\right] p_{\boldsymbol{Y}|\Theta}(\boldsymbol{y}|\theta) \mathrm{d}\boldsymbol{y}. \tag{6.7}$$

We have

$$\lim_{\theta \to \infty} B(\theta) p_{\Theta}(\theta) = 0 \tag{6.8}$$

$$\lim_{\theta \to -\infty} B(\theta) p_\Theta(\theta) = 0. \tag{6.9}$$

Then the MSE of $\hat{\theta}(\boldsymbol{Y})$ satisfies the inequality

$$\mathrm{E}\left[(\Theta - \hat{\theta}(\boldsymbol{Y}))^2\right] \geq \left\{ -\mathrm{E}\left[\frac{\partial^2 \ln p_{\Theta \boldsymbol{Y}}(\Theta, \boldsymbol{Y})}{\partial \Theta^2} \right] \right\}^{-1}. \tag{6.10}$$

Theorem 6.1 holds for continuous observation vectors. We next state a version of the BCRLB that holds for discrete observations, and therefore applies to the estimation problem considered in this chapter.

Theorem 6.2 (The BCRLB for discrete observations)**.** Let Θ be a random variable and $Z^n \in \mathcal{Z}^n$ an observation sequence, where $\mathcal{Z}$ is a finite set. Let $p_\Theta(\theta)$ be the density of Θ, let $P_{Z^n|\Theta}(z^n|\theta)$ denote the probability of z^n given θ, and let $\hat{\theta}(Z^n)$ be an estimator of Θ. Suppose the following conditions hold:

1. The conditional expectation of the error, given $\Theta = \theta$, is

$$\bar{B}(\theta) = \sum_{z^n} \left[\hat{\theta}(z^n) - \theta\right] P_{Z^n|\Theta}(z^n|\theta). \tag{6.11}$$

We have

$$\lim_{\theta \to \infty} \bar{B}(\theta) p_\Theta(\theta) = 0 \tag{6.12}$$

$$\lim_{\theta \to -\infty} \bar{B}(\theta) p_\Theta(\theta) = 0. \tag{6.13}$$

2. The first derivative of $p_\Theta(\theta)$ exists and satisfies

$$\lim_{\theta \to \infty} \frac{\partial p_\Theta(\theta)}{\partial \theta} = 0 \tag{6.14}$$

$$\lim_{\theta \to -\infty} \frac{\partial p_\Theta(\theta)}{\partial \theta} = 0. \tag{6.15}$$

Then the MSE of $\hat{\theta}(z^n)$ satisfies the inequality

$$\mathrm{E}\left[(\Theta - \hat{\theta}(Z^n))^2\right] \geq \left\{ -\mathrm{E}\left[\frac{\partial^2 \ln \left(P_{Z^n|\Theta}(Z^n|\Theta) p_\Theta(\Theta) \right)}{\partial \Theta^2} \right] \right\}^{-1}, \tag{6.16}$$

assuming that the right-hand side of (6.16) exists.

The proof of Theorem 6.2 is a straightforward modification of the proof of Theorem 6.1 given in [VT68]. For the sake of completeness, we give the proof in Appendix E.1.

6.3. Lower bounds on the MSE using single-bit dithered quantization

In this section, performance bounds on the MSE for single-bit dithered quantizers are derived.

6.3.1. The BCRLB for parameter estimation with quantized observations

We apply Theorem 6.2 to the problem of estimating H in AWGN with a single-bit adaptively dithered quantizer, and obtain the following result.

Theorem 6.3. Suppose that $|\hat{h}(z^n)| < \infty$ for any z^n. For any dither signal $D_i = \tau_i(Z^{i-1})$, $i = 1, 2, \ldots, n$, the MSE of $\hat{h}(Z^n)$ is lower bounded by

$$\mathrm{E}\left[(H - \hat{h}(Z^n))^2\right] \geq \frac{\sigma_h^2}{1 + n\frac{2}{\pi}\frac{\sigma_h^2}{\sigma_n^2}}. \tag{6.17}$$

The proof of Theorem 6.3 is given in Appendix E.2. The corollary below relates the MSE using adaptively dithered single-bit quantization to the MSE achievable with unquantized observations. Let $\hat{h}_{\mathrm{MMSE}}(Y^n) = \mathrm{E}[H|Y^n]$, i.e., $\hat{h}_{\mathrm{MMSE}}(Y^n)$ is the MMSE estimate of H based on Y^n; the MSE of $\hat{h}_{\mathrm{MMSE}}(Y^n)$ is given by [Poo94, Example IV.B.2]

$$\mathrm{E}\left[(H - \hat{h}_{\mathrm{MMSE}}(Y^n))^2\right] = \frac{\sigma_h^2}{1 + n\frac{\sigma_h^2}{\sigma_n^2}}. \tag{6.18}$$

Corollary 6.4. Suppose that $|\hat{h}(z^n)| < \infty$ for any z^n. The estimate $\hat{h}(Z^n)$ and any dither signal $D_i = \tau_i(Z^{i-1})$, $i = 1, 2, \ldots, n$, satisfy

$$\lim_{n\to\infty} \frac{\mathrm{E}\left[(H - \hat{h}(Z^n))^2\right]}{\mathrm{E}\left[(H - \hat{h}_{\mathrm{MMSE}}(Y^n))^2\right]} \geq \lim_{n\to\infty} \frac{1 + n\frac{\sigma_h^2}{\sigma_n^2}}{1 + n\frac{2}{\pi}\frac{\sigma_h^2}{\sigma_n^2}} = \frac{\pi}{2}. \tag{6.19}$$

Based on Corollary 6.4, any estimator using quantized observations asymptotically loses at least $10\log_{10}(\pi/2) \approx 1.96\,\mathrm{dB}$ in MSE compared to the MMSE estimator employing unquantized observations.

An asymptotic loss of $\pi/2$ for single-bit adaptively dithered quantization compared to unquantized observations was also derived in [PWO01, Section III-C] for the case of estimating a bounded non-random parameter using an unbiased estimator. Note, however, that the bound of [PWO01] does not apply in our setting due to the random nature of H and since H is unbounded. Moreover, for Theorem 6.3 to hold, the estimate $\hat{h}(Z^n)$ need not be unbiased.

We remark that a factor of $\pi/2$ also arises as a multiplicative factor relating the low-SNR capacity of the binary-input AWGN channel with single-bit symmetric output quantization and with continuous output. The corresponding capacities are given by [VO79, (3.4.19)]

$$C_{\text{AWGN}} \approx \frac{E_{\text{s}}}{N_0}, \quad E_{\text{s}}/N_0 \ll 1, \tag{6.20}$$

for unquantized channel outputs, and by [VO79, (3.4.20)]

$$C_{\text{Q}} \approx \frac{2}{\pi}\frac{E_{\text{s}}}{N_0}, \quad E_{\text{s}}/N_0 \ll 1, \tag{6.21}$$

for single-bit symmetric output quantization, where E_{s}/N_0 denotes the SNR. Hence, the use of hard decisions obtained from a symmetric quantizer causes a power loss of roughly 2 dB at low SNR. Recently, it was shown that this power loss can be removed if asymmetric signaling and asymmetric quantization is employed [KL11].

6.3.2. Tightening the BCRLB for short observations

The simulation results in Section 6.5 suggest that the BCRLB of Theorem 6.3 is almost tight for large values of n, whereas it is loose for small n. We state an alternative version of the BCRLB in the next theorem.

Theorem 6.5. Suppose that $|\hat{h}(z^n)| < \infty$ for any z^n. Then for any dither signal $D_i = \tau_i(Z^{i-1})$, $i = 1, 2, \ldots, n$, the MSE of $\hat{h}(Z^n)$ is lower bounded by

$$\mathrm{E}\left[(H - \hat{h}(Z^n))^2\right] \geq \begin{cases} \sigma_h^2 - \dfrac{2}{\pi\sigma_h^2}\gamma^2(\sigma_n^2, \sigma_h^2) & \text{if } n = 1 \\ \dfrac{\sigma_h^2}{1 + \dfrac{\sigma_h}{2\pi\sigma_n^2\sqrt{2\pi}}\left(\bar{\gamma}(\sigma_n^2, \sigma_h^2) + (2^n - 2)\bar{\bar{\gamma}}(\sigma_n^2)\right)} & \text{if } n \geq 2, \end{cases} \tag{6.22}$$

where

$$\gamma(\sigma_n^2, \sigma_h^2) \triangleq \int_{-\infty}^{\infty} h\mathrm{Q}\left(\frac{h}{\sigma_n}\right) e^{-h^2/(2\sigma_h^2)}\mathrm{d}h \tag{6.23}$$

$$\bar{\gamma}(\sigma_n^2, \sigma_h^2) \triangleq \int_{-\infty}^{\infty} \frac{e^{-h^2/\sigma_n^2} e^{-h^2/(2\sigma_h^2)}}{\mathrm{Q}\left(\frac{h}{\sigma_n}\right)\left[1 - \mathrm{Q}\left(\frac{h}{\sigma_n}\right)\right]}\mathrm{d}h, \tag{6.24}$$

and

$$\bar{\bar{\gamma}}(\sigma_n^2) \triangleq \int_{-\infty}^{\infty} \frac{e^{-h^2/\sigma_n^2}}{\mathrm{Q}\left(\frac{h}{\sigma_n}\right)\left[1 - \mathrm{Q}\left(\frac{h}{\sigma_n}\right)\right]}\mathrm{d}h. \tag{6.25}$$

The proof of Theorem 6.5 is given in Appendix E.3.

In order to evaluate the bound of Theorem 6.5, the integrals in (6.23), (6.24), and (6.25) must be evaluated. Unfortunately, there appears to be no closed-form solution for these integrals; we therefore resort to numerical integration techniques. The simulation results in Section 6.5 show that especially at medium and high SNRs, Theorem 6.5 provides a better lower bound on the MSE than Theorem 6.3 for small n, whereas it is loose for large n because the bound decreases exponentially in n.

To avoid numerical integration, we can bound $\gamma^2(\sigma_n^2, \sigma_h^2)$ and apply Lemma E.1 from Appendix E.2 to the integrands of $\bar{\gamma}(\sigma_n^2, \sigma_h^2)$ and $\bar{\bar{\gamma}}(\sigma_n^2)$, respectively, leading to the following Theorem.

Theorem 6.6. Suppose that $|\hat{h}(z^n)| < \infty$ for any z^n. Then for any dither signal $D_i = \tau_i(Z^{i-1})$, $i = 1, 2, \ldots, n$, the MSE of $\hat{h}(Z^n)$ is lower bounded by

$$\mathrm{E}\left[(H - \hat{h}(Z^n))^2\right] \geq \begin{cases} \left(1 - \frac{2}{\pi}\right)\sigma_h^2 + \frac{2\sigma_h^2\sigma_n^2}{\pi\sigma_n^2 + 4\sigma_h^2} & \text{if } n = 1 \\ \dfrac{\sigma_h^2}{1 + \dfrac{2}{\sqrt{2\pi(\pi-2)\frac{\sigma_n^2}{\sigma_h^2} + \pi^2\frac{\sigma_n^4}{\sigma_h^4}}} + (2^n - 2)\sqrt{\dfrac{2}{\pi(\pi-2)\frac{\sigma_n^2}{\sigma_h^2}}}} & \text{if } n \geq 2. \end{cases} \tag{6.26}$$

A proof of Theorem 6.6 is given in Appendix E.4.

Plots in Section 6.5 show that the bound of Theorem 6.6 is only slightly worse than the bound of Theorem 6.5.

6.4. Design of estimators and dither strategies

In this section, we design estimators and dither sequences $d_i = \tau_i(z^{i-1})$. We first summarize the approach of [DM10], where both d_i and the estimator for H are chosen as the linear MMSE estimate of H based on z^{i-1}. Next, we derive two other dither and estimation schemes that considerably outperform those of [DM10] at high SNR.

6.4.1. The linear MMSE estimate as estimator and dither signal

The linear MMSE estimator of H given $\boldsymbol{z}_{i-1} = [z_{i-1}, z_{i-2}, \ldots, z_1]^{\mathrm{T}}$ is

$$\hat{h}_{\mathrm{lin}}(\boldsymbol{z}_{i-1}) = \boldsymbol{w}_i^{\mathrm{T}} \boldsymbol{z}_{i-1}, \tag{6.27}$$

where $\boldsymbol{w}_i = \boldsymbol{R}_i^{-1}\boldsymbol{r}_i$, with

$$\boldsymbol{R}_i = \mathrm{E}[\boldsymbol{Z}_{i-1}\boldsymbol{Z}_{i-1}^{\mathrm{T}}] \tag{6.28}$$

$$\boldsymbol{r}_i = \mathrm{E}[H\boldsymbol{Z}_{i-1}]. \tag{6.29}$$

The strategy of [DM10] is to use $\hat{h}_{\text{lin}}(\boldsymbol{z}_{i-1})$ as the dither at time i, i.e., $d_i = \hat{h}_{\text{lin}}(\boldsymbol{z}_{i-1})$. Since a closed-form solution is not available for $\boldsymbol{R}_i$ and $\boldsymbol{r}_i$ due to the non-linearity of the quantizer and the feedback of the dither signal, both $\boldsymbol{R}_i$ and $\boldsymbol{r}_i$ are calculated by Monte-Carlo simulations over the statistics of H and the noise [DM10], yielding an approximation of $\hat{h}_{\text{lin}}(\boldsymbol{z}_{i-1})$.

Having received $\boldsymbol{z}_n$, the approximation of $\hat{h}_{\text{lin}}(\boldsymbol{z}_n)$ is used as an estimator of H. The linear MMSE scheme therefore consists of both estimation and dithering using $\hat{h}_{\text{lin}}$.

6.4.2. The MMSE estimate as estimator and dither signal

Given $\boldsymbol{R}_i$ and $\boldsymbol{r}_i$, the linear MMSE estimator $\hat{h}_{\text{lin}}(\boldsymbol{z}_{i-1})$ can be computed efficiently. However, for a *fixed* (non-adaptive) dither sequence, the (non-linear) MMSE estimator

$$\hat{h}_{\text{MMSE}}(z^{i-1}) = \mathrm{E}[H|Z^{i-1} = z^{i-1}] = \int_{-\infty}^{\infty} h p_{H|Z^{i-1}}(h|z^{i-1})\mathrm{d}h \tag{6.30}$$

is the optimal estimator of H, since we are using square-error cost in this chapter. Since the overall system is non-linear and non-Gaussian due to the single-bit quantizer, it is intuitive that $\hat{h}_{\text{MMSE}}(z^{i-1})$ may considerably outperform the linear MMSE estimate $\hat{h}_{\text{lin}}(\boldsymbol{z}_{i-1})$. The strategy of employing $\hat{h}_{\text{MMSE}}(z^{i-1})$ as an estimator of H is combined with also using $\hat{h}_{\text{MMSE}}(z^{i-1})$ as the dither signal at the next time instance, i.e., by choosing $d_i = \hat{h}_{\text{MMSE}}(z^{i-1})$. We next discuss how to compute $\hat{h}_{\text{MMSE}}(z^{i-1})$ efficiently.

The integration over

$$h \cdot p_{H|Z^{i-1}}(h|z^{i-1}) = h \cdot \frac{P_{Z^{i-1}|H}(z^{i-1}|h)p_H(h)}{P_{Z^{i-1}}(z^{i-1})} \tag{6.31}$$

in (6.30) cannot be solved in closed form, since $P_{Z^{i-1}|H}(z^{i-1}|h)$ is a product of $(i-1)$ terms involving the Q-function. But an approximation $\bar{h}_{\text{MMSE}}(z^{i-1})$ of $\hat{h}_{\text{MMSE}}(z^{i-1})$ can be computed based on a recursively updated *discrete* approximation of $p_{H|Z^{i-1}}(h|z^{i-1})$ combined with interpolation. To that end, we form a discrete random variable $\bar{H}$ by sampling from the distribution of H. The variable $\bar{H}$ takes on values in $\mathcal{H}^{(0)} = \{-\Delta, -\Delta(1-\frac{2}{B}), -\Delta(1-2\frac{2}{B}), -\Delta(1-3\frac{2}{B}), \ldots, \Delta\}$, where $|\mathcal{H}^{(0)}| = B$, so that $\bar{H}$ has probability mass function $P_{\bar{H}}(h)$, i.e., $\sum_{h\in\mathcal{H}^{(0)}} P_{\bar{H}}(h) = 1$. The parameter Δ, $\Delta > 0$, is chosen such that $\Pr[-\Delta \leq H \leq \Delta] = 0.99995$; since $H \sim \mathcal{N}(0, \sigma_h^2)$, we get $\Delta = \sqrt{2}\sigma_h \mathrm{erf}^{-1}(0.99995) \approx 4.056\,\sigma_h$. Defining $P^{(0)}(h|z^0) \triangleq P_{\bar{H}}(h)$, $h \in \mathcal{H}^{(0)}$, we can recursively compute an approximation of $p_{H|Z^i}(h|z^i)$ and of $\hat{h}_{\text{MMSE}}(z^i)$ by the following algorithm.

1. Update step: For $h \in \mathcal{H}^{(i-1)}$, compute

$$P^{(i)}(h, z_i|z^{i-1}) = P^{(i-1)}(h|z^{i-1}) \cdot \begin{cases} \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right) & \text{if } z_i = 0 \\ 1 - \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right) & \text{if } z_i = 1. \end{cases} \tag{6.32}$$

2. Interval expansion and interpolation step: Let $p_{\max}^{(i)} = \max_{h\in\mathcal{H}^{(i-1)}} P^{(i)}(h, z_i|z^{i-1})$, suppose ϵ satisfies $0 < \epsilon < 1$, and let

$$h_{\min}^{(i)} = \operatorname*{argmin}_{h\in\mathcal{H}^{(i-1)}} P^{(i)}(h, z_i|z^{i-1}) \quad \text{s.t. } P^{(i)}(h, z_i|z^{i-1}) > \epsilon p_{\max}^{(i)} \tag{6.33}$$

$$h_{\max}^{(i)} = \operatorname*{argmax}_{h\in\mathcal{H}^{(i-1)}} P^{(i)}(h, z_i|z^{i-1}) \quad \text{s.t. } P^{(i)}(h, z_i|z^{i-1}) > \epsilon p_{\max}^{(i)}. \tag{6.34}$$

Defining $\eta^{(i)} = (h_{\max}^{(i)} - h_{\min}^{(i)})/B$, the set $\mathcal{H}^{(i)}$ of size B is given by

$$\mathcal{H}^{(i)} = \{h_{\min}^{(i)}, h_{\min}^{(i)} + \eta^{(i)}, h_{\min}^{(i)} + 2\eta^{(i)}, \ldots, h_{\max}^{(i)}\}. \tag{6.35}$$

Next, compute $\bar{P}^{(i)}(h, z_i|z^{i-1})$, $h \in \mathcal{H}^{(i)}$, from $P^{(i)}(h, z_i|z^{i-1})$, $h \in \mathcal{H}^{(i-1)}$ by interpolating $P^{(i)}(h, z_i|z^{i-1})$ at the B points $h \in \mathcal{H}^{(i)}$; then we obtain an approximation $P^{(i)}(z_i|z^{i-1})$ of the conditional probability $P_{Z_i|Z^{i-1}}(z_i|z^{i-1})$ through

$$P^{(i)}(z_i|z^{i-1}) = \sum_{h\in\mathcal{H}^{(i)}} \bar{P}^{(i)}(h, z_i|z^{i-1}) \tag{6.36}$$

and finally,

$$P^{(i)}(h|z^i) = \frac{\bar{P}^{(i)}(h, z_i|z^{i-1})}{P^{(i)}(z_i|z^{i-1})}, \tag{6.37}$$

where the normalization with $P^{(i)}(z_i|z^{i-1})$ ensures that $\sum_{h\in\mathcal{H}^{(i)}} P^{(i)}(h|z^i) = 1$.

3. The estimate of $\hat{h}_{\mathrm{MMSE}}(z^i)$ is given by

$$\bar{h}_{\mathrm{MMSE}}(z^i) = \sum_{h\in\mathcal{H}^{(i)}} hP^{(i)}(h|z^i). \tag{6.38}$$

In a practical implementation (cf. Section 6.5), choosing B on the order of 100 and ϵ around 10^{-5} yields excellent performance at reasonable computational complexity since only a small number of samples describing $p_{H|Z^i}(h|z^i)$ needs to be updated at each time step.

Note that for the (non-linear) MMSE scheme, both estimation and dithering are performed using $\hat{h}_{\mathrm{MMSE}}$.

6.4.3. The dither signal that minimizes the MSE

While the MMSE estimator $\hat{h}_{\mathrm{MMSE}}(z^n)$ minimizes the MSE of estimating H for a fixed dither sequence d^n, it is not clear if employing $d_i = \hat{h}_{\mathrm{MMSE}}(z^{i-1})$ is an optimal dither strategy. Instead of dithering using $\hat{h}_{\mathrm{MMSE}}(z^{i-1})$, the dither signal should be selected for optimal MSE performance at time n, i.e., given z^{i-1}, $i \leq n$, the MSE-optimal dither

sequence $d^n_{\mathrm{opt},i}$ is formally given by

$$d^n_{\mathrm{opt},i} = \operatorname*{argmin}_{d_i^n} \mathrm{E}\left[(H - \hat{h}_{\mathrm{MMSE}}(Z^n, d_i^n))^2 | Z^{i-1} = z^{i-1}\right], \tag{6.39}$$

where we make the dependency of $\hat{h}_{\mathrm{MMSE}}(z^n, d_i^n)$ on d_i^n explicit. For complexity reasons, we will not attempt to solve Problem (6.39) for $n > i$; instead, we solve (6.39) for $n = i$ only, i.e., the dither signal d_i is chosen based on z^{i-1} such that the MSE at the next time instance, i.e., at time step i, is minimized. We refer to this dither as the "one-step look-ahead" (OSLA) dither signal.

Suppose that $Z^{i-1} = z^{i-1}$, so that the conditional distribution $p_{H|Z^{i-1}}(h|z^{i-1})$ is fixed. Then, the dither signal d_i^* is given by

$$d_i^* = \operatorname*{argmin}_{d_i} \mathrm{E}\left[(H - \hat{h}_{\mathrm{MMSE}}(Z^i, d_i))^2 | Z^{i-1} = z^{i-1}\right]. \tag{6.40}$$

Given z^{i-1} and some d_i, we can view $\mathrm{E}\left[(H - \hat{h}_{\mathrm{MMSE}}(Z^i, d_i))^2 | Z^{i-1} = z^{i-1}\right]$ as the MSE of estimating the random parameter H with prior $p_{H|Z^{i-1}}(h|z^{i-1})$ using the estimator $\hat{h}_{\mathrm{MMSE}}(z^i, d_i)$. Due to the properties of MMSE estimation [Poo94, Section IV.B], we have

$$\begin{aligned}\mathrm{E}&\left[(H - \hat{h}_{\mathrm{MMSE}}(Z^i, d_i))^2 | Z^{i-1} = z^{i-1}\right] \\ &= \mathrm{E}\left[H^2 | Z^{i-1} = z^{i-1}\right] - \mathrm{E}\left[\hat{h}^2_{\mathrm{MMSE}}(Z^i, d_i) | Z^{i-1} = z^{i-1}\right],\end{aligned} \tag{6.41}$$

Equation (6.41) can be readily verified using the orthogonality principle [Poo94, Proposition V.C.2]: since $\hat{h}_{\mathrm{MMSE}}(z^i, d_i)$ is the conditional mean of H, we have

$$\mathrm{E}\left[(H - \hat{h}_{\mathrm{MMSE}}(Z^i, d_i))\hat{h}_{\mathrm{MMSE}}(Z^i, d_i) | Z^{i-1} = z^{i-1}\right] = 0. \tag{6.42}$$

Rearranging the identity

$$\begin{aligned}\mathrm{E}&\left[H^2 | Z^{i-1} = z^{i-1}\right] \\ &= \mathrm{E}\left[\left((H - \hat{h}_{\mathrm{MMSE}}(Z^i, d_i)) + \hat{h}_{\mathrm{MMSE}}(Z^i, d_i)\right)^2 \middle| Z^{i-1} = z^{i-1}\right] && (6.43) \\ &= \mathrm{E}\left[(H - \hat{h}_{\mathrm{MMSE}}(Z^i, d_i))^2 | Z^{i-1} = z^{i-1}\right] + \mathrm{E}\left[\hat{h}^2_{\mathrm{MMSE}}(Z^i, d_i) | Z^{i-1} = z^{i-1}\right] && (6.44) \\ &\quad + 2\mathrm{E}\left[(H - \hat{h}_{\mathrm{MMSE}}(Z^i, d_i))\hat{h}_{\mathrm{MMSE}}(Z^i, d_i) | Z^{i-1} = z^{i-1}\right]\end{aligned}$$

and using (6.42) yields (6.41). Consequently, since $\mathrm{E}\left[H^2 | Z^{i-1} = z^{i-1}\right]$ is fixed for a given z^{i-1}, the minimization in (6.40) is equivalent to

$$d_i^* = \operatorname*{argmax}_{d_i} \mathrm{E}\left[\hat{h}^2_{\mathrm{MMSE}}(Z^i, d_i) | Z^{i-1} = z^{i-1}\right]. \tag{6.45}$$

Problem (6.45) seems hard to solve since $\hat{h}_{\mathrm{MMSE}}(z^i, d_i)$ cannot be found in closed form

as a function of z^i and d_i; we therefore work with a *discrete* approximation $\tilde{h}_{\mathrm{MMSE}}(z^i, d_i)$ of $\hat{h}_{\mathrm{MMSE}}(z^i, d_i)$, similar to Section 6.4.2. As before, let $P^{(i-1)}(h|z^{i-1})$, $h \in \mathcal{H}^{(i-1)}$ be the approximation of the conditional density $p_{H|Z^{i-1}}(h|z^{i-1})$ obtained from the received sequence z^{i-1}. Based on $P^{(i-1)}(h|z^{i-1})$ and

$$\tilde{P}^{(i)}(z_i|z^{i-1}) = \sum_{h \in \mathcal{H}^{(i-1)}} P_{Z_i|HZ^{i-1}}(z_i|h, z^{i-1}) P^{(i-1)}(h|z^{i-1}), \tag{6.46}$$

we can compute the estimate $\tilde{h}_{\mathrm{MMSE}}(z^i, d_i)$ (omitting the interval expansion and interpolation step) of $\hat{h}_{\mathrm{MMSE}}(z^i, d_i)$, given by

$$\tilde{h}_{\mathrm{MMSE}}(z^i, d_i) = \frac{\sum\limits_{h \in \mathcal{H}^{(i-1)}} h P_{Z_i|HZ^{i-1}}(z_i|h, z^{i-1}) P^{(i-1)}(h|z^{i-1})}{\tilde{P}^{(i)}(z_i|z^{i-1})}, \tag{6.47}$$

where the right-hand side depends on d_i through

$$P_{Z_i|HZ^{i-1}}(z_i|h, z^{i-1}) = \delta(z_i)\mathrm{Q}\left(\frac{h - d_i}{\sigma_n}\right) + \delta(z_i - 1)\left(1 - \mathrm{Q}\left(\frac{h - d_i}{\sigma_n}\right)\right). \tag{6.48}$$

Consequently, the cost function approximating the one in (6.45) is

$$V(d_i, z^{i-1}) \triangleq \sum_{z_i} \tilde{P}^{(i)}(z_i|z^{i-1}) \tilde{h}^2_{\mathrm{MMSE}}(z^i, d_i), \tag{6.49}$$

and the approximate solution to Problem (6.45) is

$$\tilde{d}_i^* = \operatorname*{argmax}_{d_i} \; V(d_i, z^{i-1}). \tag{6.50}$$

In all the simulations that we performed, we observed that $V(d_i, z^{i-1})$ is a strictly quasi-concave function [BSS06, Definition 3.5.5] of d_i, for any z^{i-1}. However, we were unable to formally verify this observation. Nevertheless, we solve Problem (6.50) with a gradient descent algorithm [BV04, Chapter 9.3] combined with backtracking line search [BV04, Chapter 9.2], observing excellent convergence behavior. For such a gradient descent method, we need the first derivative of $V(d_i, z^{i-1})$ with respect to d_i, which is

$$\begin{aligned} &\frac{\partial V(d_i, z^{i-1})}{\partial d_i} \\ &= \sum_{z_i} \tilde{h}_{\mathrm{MMSE}}(z^i, d_i) \left[2\tilde{P}^{(i)}(z_i|z^{i-1}) \frac{\partial \tilde{h}_{\mathrm{MMSE}}(z^i, d_i)}{\partial d_i} + \tilde{h}_{\mathrm{MMSE}}(z^i, d_i) \frac{\partial \tilde{P}^{(i)}(z_i|z^{i-1})}{\partial d_i}\right]. \end{aligned} \tag{6.51}$$

Given

$$\frac{\partial P_{Z_i|HZ^{i-1}}(z_i|h, z^{i-1})}{\partial d_i} = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-(h-d_i)^2/(2\sigma_n^2)} (\delta(z_i) - \delta(z_i - 1)), \tag{6.52}$$

it is straightforward to compute

$$\frac{\partial \tilde{P}^{(i)}(z_i|z^{i-1})}{\partial d_i} = \sum_{h\in\mathcal{H}^{(i-1)}} P^{(i-1)}(h|z^{i-1}) \frac{\partial P_{Z_i|HZ^{i-1}}(z_i|h,z^{i-1})}{\partial d_i} \tag{6.53}$$

and

$$\begin{aligned}\frac{\partial \tilde{h}_{\mathrm{MMSE}}(z^i,d_i)}{\partial d_i} =& \frac{1}{\tilde{P}^{(i)}(z_i|z^{i-1})} \sum_{h\in\mathcal{H}^{(i-1)}} h \frac{\partial P_{Z_i|HZ^{i-1}}(z_i|h,z^{i-1})}{\partial d_i} P^{(i-1)}(h|z^{i-1}) \qquad (6.54)\\ &- \frac{\partial \tilde{P}^{(i)}(z_i|z^{i-1})}{\partial d_i} \frac{\sum_{h\in\mathcal{H}^{(i-1)}} h P_{Z_i|HZ^{i-1}}(z_i|h,z^{i-1}) P^{(i-1)}(h|z^{i-1})}{[\tilde{P}^{(i)}(z_i|z^{i-1})]^2}\\ =& \frac{1}{\tilde{P}^{(i)}(z_i|z^{i-1})} \sum_{h\in\mathcal{H}^{(i-1)}} h \frac{\partial P_{Z_i|HZ^{i-1}}(z_i|h,z^{i-1})}{\partial d_i} P^{(i-1)}(h|z^{i-1}) \qquad (6.55)\\ &- \frac{\partial \tilde{P}^{(i)}(z_i|z^{i-1})}{\partial d_i} \frac{\tilde{h}_{\mathrm{MMSE}}(z^i,d_i)}{\tilde{P}^{(i)}(z_i|z^{i-1})}.\end{aligned}$$

Inserting (6.53) and (6.55) into (6.51), we have

$$\begin{aligned}\frac{\partial V(d_i,z^{i-1})}{\partial d_i} = \sum_{z_i} \tilde{h}_{\mathrm{MMSE}}(z^i,d_i) \Bigg[& 2 \sum_{h\in\mathcal{H}^{(i-1)}} h \frac{\partial P_{Z_i|HZ^{i-1}}(z_i|h,z^{i-1})}{\partial d_i} P^{(i-1)}(h|z^{i-1}) \qquad (6.56)\\ & - \tilde{h}_{\mathrm{MMSE}}(z^i,d_i) \frac{\partial \tilde{P}^{(i)}(z_i|z^{i-1})}{\partial d_i} \Bigg].\end{aligned}$$

The approximation $\bar{h}_{\mathrm{MMSE}}(z^{i-1})$ of $\hat{h}_{\mathrm{MMSE}}(z^{i-1})$ (cf. Section 6.4.2) is chosen as a starting point for the gradient descent method solving Problem (6.50).

We remark that the OSLA scheme consists of dithering using the OSLA-optimal dither given by (6.50), and of estimation using $\hat{h}_{\mathrm{MMSE}}$.

6.4.4. The maximum likelihood estimate as estimator and dither signal

Next, we consider the maximum likelihood (ML) estimate $\hat{h}_{\mathrm{ML}}(z^{i-1})$ as an estimator and dither signal, which is given by

$$\hat{h}_{\mathrm{ML}}(z^{i-1}) = \operatorname*{argmax}_{h\in\mathbb{R}} P_{Z^{i-1}|H}(z^{i-1}|h) = \operatorname*{argmax}_{h\in\mathbb{R}} \ln\left(P_{Z^{i-1}|H}(z^{i-1}|h)\right), \tag{6.57}$$

where the last equality holds since $\ln(x)$ is strictly increasing in x for $x > 0$. Note that the prior on H is not required for computing the ML estimate. Since $P_{Z^{i-1}|H}(z^{i-1}|h)$ is a product of $i-1$ terms involving the Q-function, $\hat{h}_{\mathrm{ML}}(z^{i-1})$ cannot be computed in closed

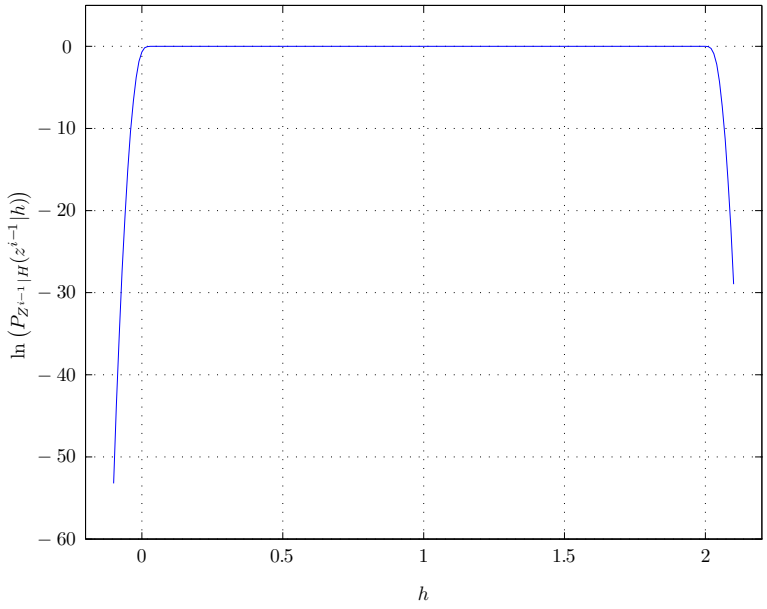


Figure 6.4.: A graph of $\ln(P_{Z^{i-1}|H}(z^{i-1}|h))$ for $z^3 = \{1, 0, 0\}$ at an SNR of 40 dB. Numerical optimization yields $\hat{h}_{\mathrm{ML}}(z^{i-1}) = 1.01$, however, for $0.5 \leq h \leq 1.5$, we have[2] $|\partial \ln(P_{Z^{i-1}|H}(z^{i-1}|h))/(\partial h)| < 10^{-541}$.

form. However, the optimization problem defining $\hat{h}_{\mathrm{ML}}(z^{i-1})$ can be efficiently solved numerically, since $P_{Z^{i-1}|H}(z^{i-1}|h)$ is log-concave in h (cf. [RG06a, Proposition 2]), for any z^{i-1}, i.e., $\ln(P_{Z^{i-1}|H}(z^{i-1}|h))$ is concave.

Proposition 6.7. For any $z^{i-1} \in \{0, 1\}^{i-1}$ and $d^{i-1} \in \mathbb{R}^{i-1}$, the function $P_{Z^{i-1}|H}(z^{i-1}|h)$ is log-concave in h.

Proof. See Appendix E.5. ■

In order to compute $\hat{h}_{\mathrm{ML}}(z^{i-1})$ numerically, we use the Newton method given in [BV04, Chapter 9.5.2] combined with backtracking line search [BV04, Chapter 9.2]. For the Newton method, we need the first and second derivative of $\ln(P_{Z^{i-1}|H}(z^{i-1}|h))$ as contained in (E.37) and (E.38). At high SNR and short observations z^{i-1}, however, the Newton method performs poorly since the first derivative of $\ln(P_{Z^{i-1}|H}(z^{i-1}|h))$ is extremely flat in a rather wide interval around $\hat{h}_{\mathrm{ML}}(z^{i-1})$. This is illustrated in Figure 6.4. Therefore, at high SNR and for short observations, we first perform a bisection algorithm on $\partial \ln(P_{Z^{i-1}|H}(z^{i-1}|h))/(\partial h)$ to determine an interval $[h_1, h_2]$ of width $(h_2 - h_1) \leq 2 \cdot 10^{-3}$ containing $\hat{h}_{\mathrm{ML}}(z^{i-1})$. Then $(h_2 - h_1)/2$ serves as a starting point for the subsequent Newton method.

[2] Such small numbers can be computed using the GNU Multiple Precision Arithmetic Library, available online: `http://gmplib.org/`

We remark that the computation of $\hat{h}_{\mathrm{ML}}(z^{i-1})$ can be simplified if $z^{i-1} = \{0, 0, \ldots, 0\}$ or $z^{i-1} = \{1, 1, \ldots, 1\}$. To see this, suppose that $z^{i-1} = \{0, 0, \ldots, 0\}$, so that

$$\ln\left(P_{Z^{i-1}|H}(z^{i-1}|h)\right) = \sum_{k=1}^{i-1} \ln\left(\mathrm{Q}\left(\frac{h - d_k}{\sigma_n}\right)\right). \tag{6.58}$$

Since we have

$$\frac{\partial}{\partial x} \ln\left(\mathrm{Q}\left(x\right)\right) = -\frac{1}{\sqrt{2\pi}} \frac{e^{-x^2/2}}{\mathrm{Q}\left(x\right)} < 0, \tag{6.59}$$

the function $\ln(P_{Z^{i-1}|H}(z^{i-1}|h))$ is decreasing in h. Strictly speaking, we therefore have $\hat{h}_{\mathrm{ML}}(z^{i-1}) = -\infty$ if $z^{i-1} = \{0, 0, \ldots, 0\}$; however, we set $\hat{h}_{\mathrm{ML}}(z^{i-1}) = -\Delta \approx -4.056\,\sigma_h$ (cf. Section 6.4.2) in that case. Likewise, if $z^{i-1} = \{1, 1, \ldots, 1\}$, $\ln(P_{Z^{i-1}|H}(z^{i-1}|h))$ is increasing in h, and we set $\hat{h}_{\mathrm{ML}}(z^{i-1}) = \Delta$.

Note that $\hat{h}_{\mathrm{ML}}$ is used for both estimation and dithering in the ML scheme.

6.4.5. The maximum a posteriori estimate as estimator and dither signal

Finally, we also consider the maximum a posteriori (MAP) scheme, in which the MAP estimate $\hat{h}_{\mathrm{MAP}}(z^{i-1})$ is employed as an estimator and dither signal. The MAP estimate is given by

$$\hat{h}_{\mathrm{MAP}}(z^{i-1}) = \operatorname*{argmax}_{h \in \mathbb{R}} P_{Z^{i-1}|H}(z^{i-1}|h) p_H(h) = \operatorname*{argmax}_{h \in \mathbb{R}} \ln\left(P_{Z^{i-1}|H}(z^{i-1}|h) p_H(h)\right), \tag{6.60}$$

which cannot be solved in closed-form due to the Q-function appearing in $P_{Z^{i-1}|H}(z^{i-1}|h)$. However, just like the optimization problem defining $\hat{h}_{\mathrm{ML}}(z^{i-1})$, Problem (6.60) can be solved numerically since $P_{Z^{i-1}|H}(z^{i-1}|h)p_H(h)$ has the following property.

Proposition 6.8. For any $z^{i-1} \in \{0, 1\}^{i-1}$ and any $d^{i-1} \in \mathbb{R}^{i-1}$, $P_{Z^{i-1}|H}(z^{i-1}|h)p_H(h)$ is log-concave in h.

Proof. By Proposition 6.7, the function $P_{Z^{i-1}|H}(z^{i-1}|h)$ is log-concave. Moreover,

$$\ln(p_H(h)) = -\frac{1}{2}\ln(2\pi\sigma_h) - \frac{h^2}{2\sigma_h^2} \tag{6.61}$$

is concave, so that $p_H(h)$ is log-concave. Since the product of log-concave functions is log-concave [BV04, Chapter 3.5.2], we have that $P_{Z^{i-1}|H}(z^{i-1}|h)p_H(h)$ is log-concave. ■

In order to compute $\hat{h}_{\mathrm{MAP}}(z^{i-1})$ numerically, we again use a Newton method with backtracking line search.

6.5. Simulation results

Simulation results for various SNRs are shown in Figures 6.5 to 6.9, where $\sigma_h^2 = 1$ without loss of generality. In addition to the BCRLBs of Theorems 6.3, 6.5, and 6.6, we also show the MSE of the MMSE estimator $\hat{h}_{\mathrm{MMSE}}(Y^n)$ (cf. (6.18)) using unquantized observations Y^n. The performance of the dither and estimation schemes presented in Section 6.4 is determined by means of simulation for the linear/non-linear MMSE, the OSLA, the ML, and the MAP scheme.

At an SNR of 0 dB, the performance of all dither strategies except the ML scheme is almost indistinguishable, and very close to the lower bound provided by the BCRLB. The non-linear MMSE and OSLA schemes continue to perform close to the BCRLB for higher SNRs, while the linear feedback strategy exhibits a considerable performance gap if the SNR is 20 dB or higher. While the performance difference between the non-linear MMSE and OSLA scheme is not large, it is most pronounced at very high SNR. Throughout the SNR range, the ML scheme performs poorly for small n because $\hat{h}_{\mathrm{ML}}(z_1) = \pm\Delta$. For longer observations, however, the ML scheme approaches the performance of the MMSE scheme. The MAP scheme performs well for small SNR, whereas a considerable performance degradation can be observed for SNRs larger than 10 dB. This performance degradation is due to those realizations of $H = h$ with magnitude large enough such that $|d_i| = |\hat{h}_{\mathrm{MAP}}(z^{i-1})| < |h|$ for all $i - 1 \leq n$. Suppose for example that the SNR equals 40 dB and that $H = 1.5$; then, due to the small variance of the noise, we observe $z^{50} = \{1, 1, \ldots, 1\}$, and the sequence of $\hat{h}_{\mathrm{MAP}}(z^{i-1})$ is an increasing positive sequence with $\hat{h}_{\mathrm{MAP}}(z_1) \approx 0.04$ and $\hat{h}_{\mathrm{MAP}}(z^{50}) \approx 1.44 < 1.5$, i.e., the dither sequence $\{d_i\}$ is not such that $Z_i = 0$ with large enough probability. The effect is even stronger for realizations of H with larger magnitude than 1.5. Since $\Pr[H \geq 1.5]$ is sufficiently high if $H \sim \mathcal{N}(0, 1)$, those realizations of H contribute to the bad performance of the MAP scheme at sufficiently high SNR.

6.6. Discussion

We studied the parameter estimation problem using a single-bit dithered quantizer. By bounding the BCRLB for this problem, we derived lower bounds on the MSE that hold for all dither strategies. We showed that the performance of single-bit dithered parameter estimation cannot approach the performance of estimation using unquantized observations; in particular, the estimator based on continuous observations outperforms the estimator based on the quantized observations asymptotically by at least 1.96 dB. We also designed dither sequences that are computed by strictly causal processing of the quantizer output sequence. Through simulations, we showed that the derived bounds on the MSE can be closely approached.

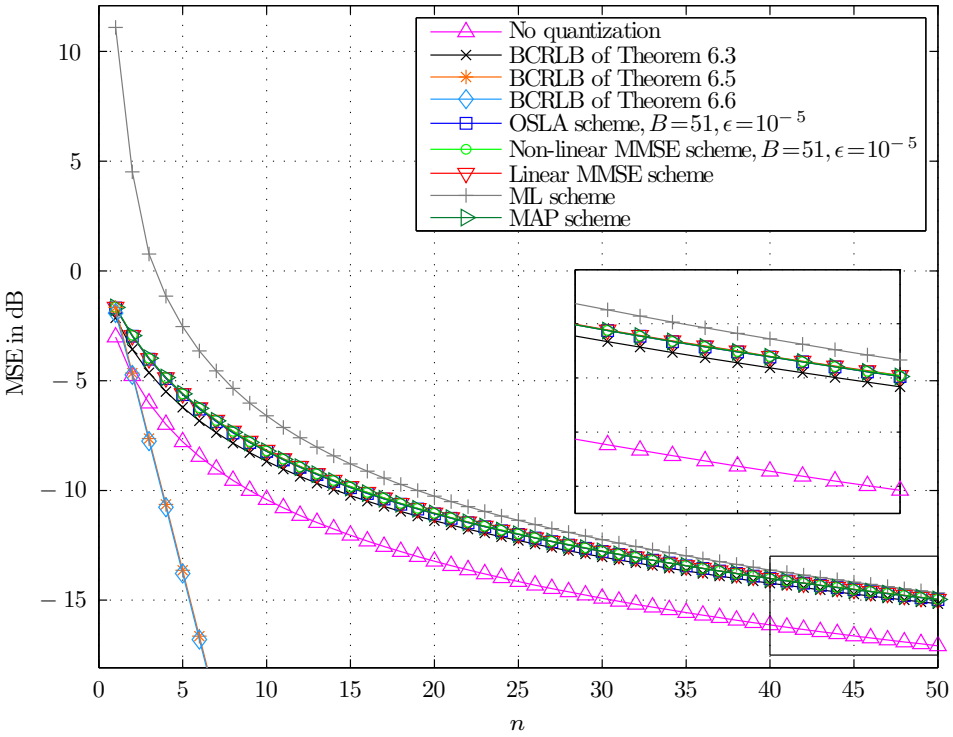


Figure 6.5.: MSE for an SNR of 0 dB.

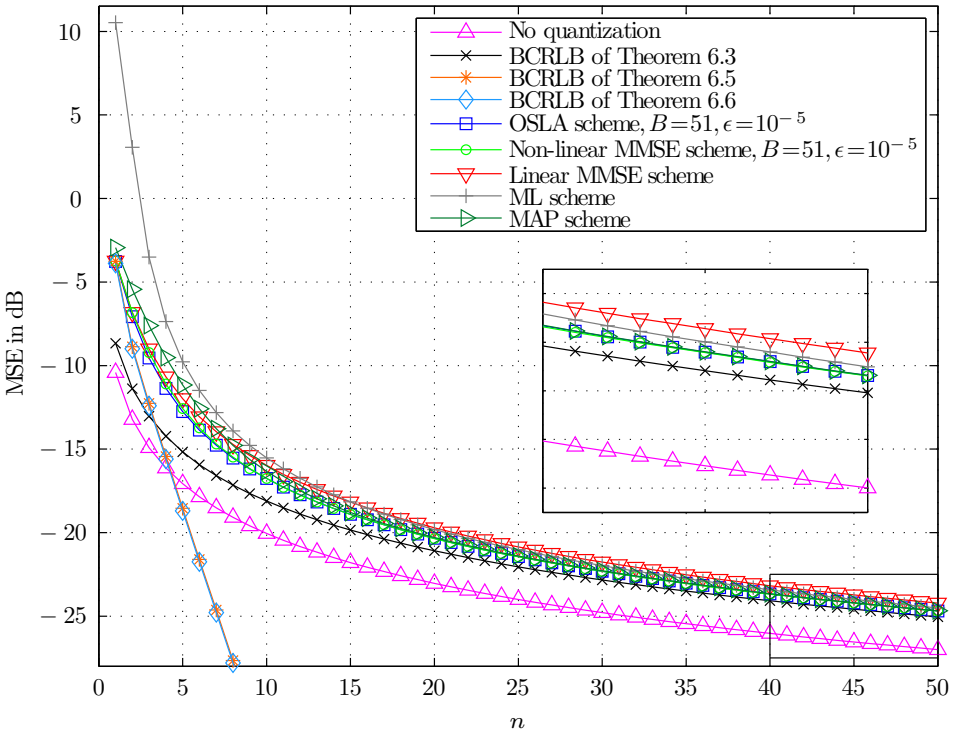


Figure 6.6.: MSE for an SNR of 10 dB.

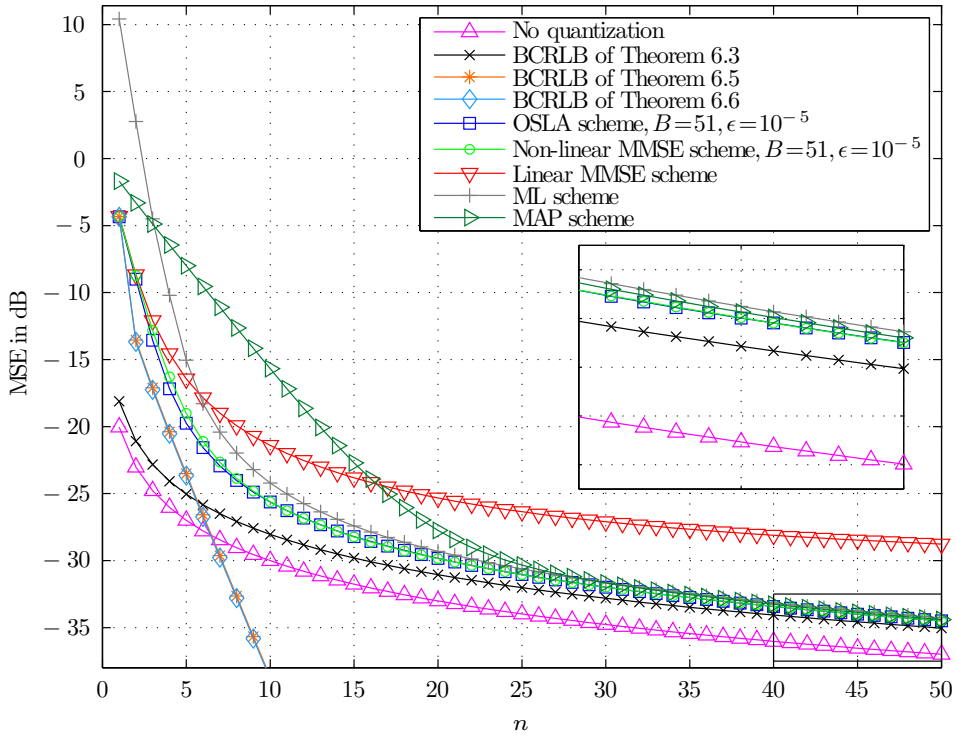


Figure 6.7.: MSE for an SNR of 20 dB.

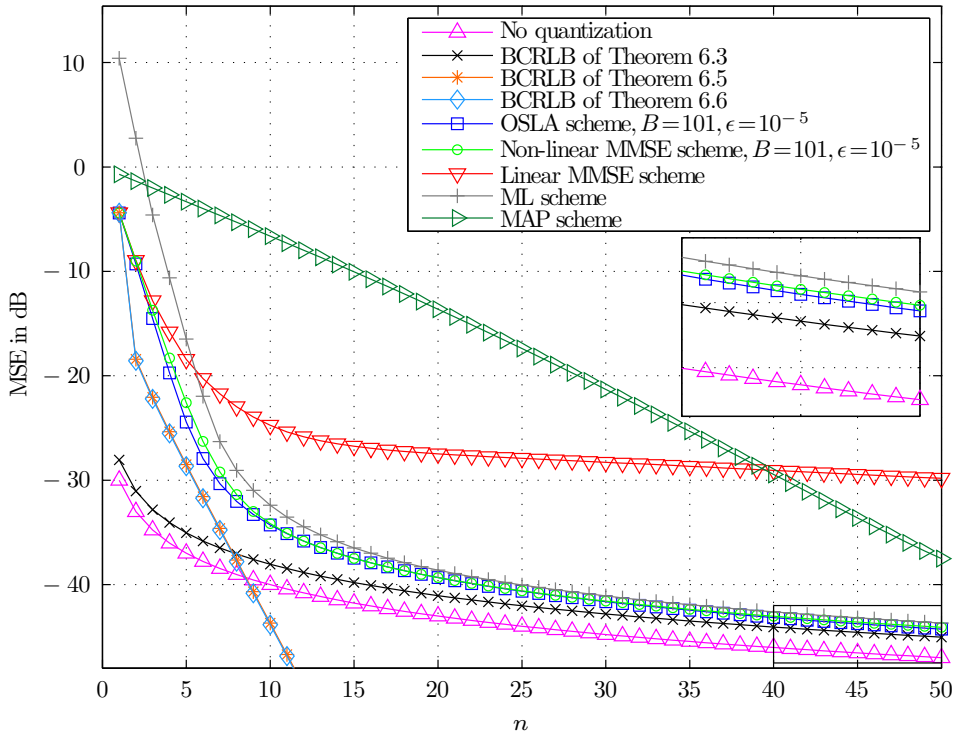


Figure 6.8.: MSE for an SNR of 30 dB.

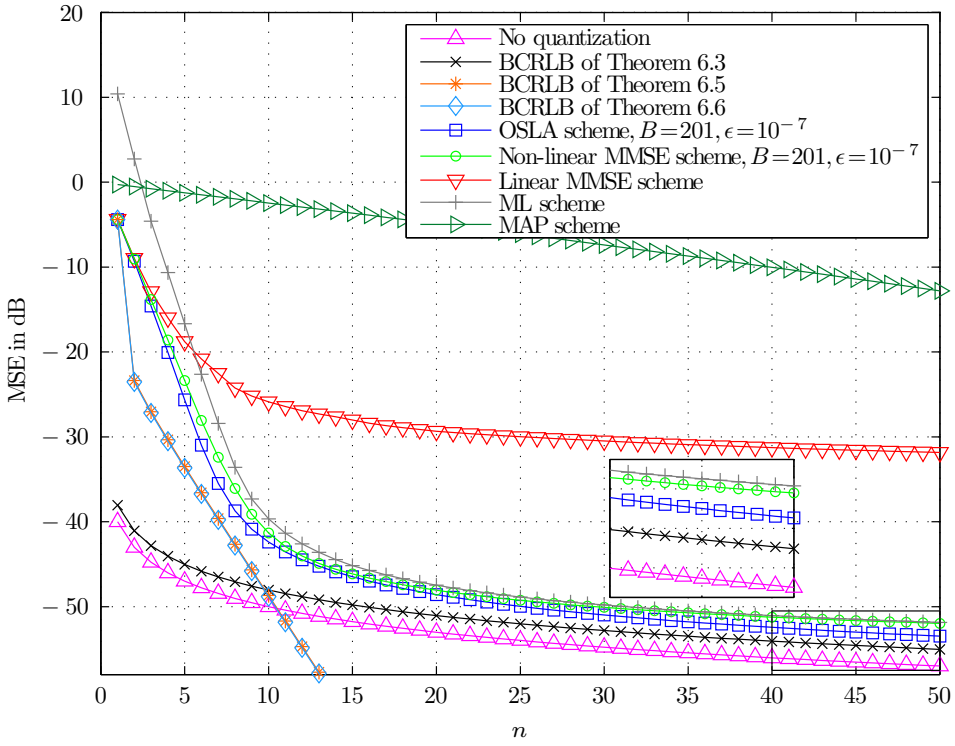


Figure 6.9.: MSE for an SNR of 40 dB.

7

Conclusions

We have studied the design of quantizers for two problems in communications: the orthogonal MARC with compression of the received sequences at the relay, and the point-to-point link with ISI, additive noise, and A/D conversion at the receiver. Both systems share the need for low-precision quantization, the MARC because in wireless systems, the relay–destination link is not only of finite capacity, but possibly of very limited capacity, and the quantized ISI channel after A/D conversion becomes both power-hungry and costly at high precision with increasing communication speed.

For the MARC, we have restricted our study to orthogonal channels. In addition to optimizing the compression rate at the relay for maximal sum-rate, we addressed the problem of designing scalar and two-dimensional quantizers for a practical scheme with low complexity. In this context, there are a number of interesting questions and open problems:

- ▷ For the sum-rate optimal rate allocation problem presented in Chapter 4, one can also optimize the time-sharing variables α_i and powers P_i and P_{r}, given fixed realizations of the channel coefficients. This brings into question whether or not the new optimization problem is still convex in all of its variables.

- ▷ As a further generalization of the CF or NNC strategies considered in Chapter 4, one can address the problem of optimizing the quantization for a scheme in which CF is combined with partial decode-and-forward [CEG79, Theorem 7]. How well does amplify-and-forward perform?

For ISI channels with additive noise and i.i.d. inputs, we designed low-precision scalar and two-dimensional ADCs that maximize the information rate. We also showed that quantization with Λ regions suffices if the channel is noiseless and if the transmitter employs

a signaling alphabet of size Λ. Since the design of such ADCs relies on sufficient CSI at the receiver, we also addressed the problem of obtaining an estimate of the channel coefficients using a single-bit adaptively dithered quantizer. In this line of work, the following interesting questions are open:

- ▷ The algorithm for the design of scalar ADCs assumes the channel to be fixed for the entire transmission. An algorithm that updates the ADC thresholds adaptively could be applied to account for such time-varying channels.
- ▷ Since the boundaries of information-rate optimal two-dimensional ADCs are not hyperplanes in general, it is not clear whether or not such an ADC can be efficiently implemented in hardware.
- ▷ Instead of using a fixed quantizer for the entire transmission, one can consider dithered (single-bit) quantizers for ISI channels. What is the right algorithm to adaptively design the dither signal?
- ▷ By selecting the training sequence as a bursty periodic sequence, the channel estimation problem for an ISI channel was shown to be decomposable into several parallel independent sub-problems. But are there training sequences that do not yield such a decomposition and yet still outperform the approach employed in this work of estimating each channel coefficient independently? This question can be extended to finding lower bounds on the MSE and to designing dither/estimation schemes approaching the bounds for such a scheme.
- ▷ In Chapter 6, the MSE decreases slowly ($\sim 1/n$) with the number n of quantized observations. Can an algorithm with feedback to the transmitter speed up the estimation process (cf. [WKN$^+$11])?
- ▷ One could also study whether or not a vector approach to estimating an ISI channel offers any gains. For instance, again for a bursty periodic training sequence, the task of estimating the real-valued coefficients H_1 and H_2 could be transformed into estimation of the amplitude $|H|$ and the phase $\arg(H)$ of the complex number $H = H_1 + j\, H_2$.
- ▷ Recently, the use of very large antenna arrays (massive MIMO) at base stations of a cellular communication system has been studied as a means of carrying the exponentially growing wireless data traffic [Mar10, HtBD11]. For a large number of antennas, it might not be feasible to quantize the received signal at each antenna element with more than a few bits per sample, so that both the ADC design problem and the channel estimation problem under low-precision output quantization arise.

A

Proofs for Chapter 2

For $0 \leq y \leq I(X;Y)$, define the function

$$G(y) \triangleq \min_{P_{Z|Y}} I(Y;Z) \quad \text{s.t.} \quad I(X;Z) = y. \tag{A.1}$$

Let $I^* = I(R^*)$, for some $0 \leq R^* \leq R_{\max}$, and let $P^*_{Z|Y}$ be the corresponding optimizer, i.e., the mutual information under the mapping $P_{Z|Y} = P^*_{Z|Y}$ satisfies

$$I_{P^*_{Z|Y}}(X;Z) = I^* \tag{A.2}$$

$$I_{P^*_{Z|Y}}(Y;Z) = R^*. \tag{A.3}$$

Since there exists $P_{Z|Y} = P^*_{Z|Y}$ such that (A.2) and (A.3) hold, we consequently have

$$G(I^*) = \min_{P_{Z|Y}: I(X;Z)=I^*} I(Y;Z) \tag{A.4}$$

$$\leq R^*. \tag{A.5}$$

Now consider $G(I^*)$, and let $P'_{Z|Y}$ be the corresponding optimizer, i.e., the mutual information under the mapping $P_{Z|Y} = P'_{Z|Y}$ satisfies

$$I_{P'_{Z|Y}}(X;Z) = I^* \tag{A.6}$$

$$I_{P'_{Z|Y}}(Y;Z) = G(I^*). \tag{A.7}$$

Therefore, we have

$$I(G(I^*)) = \max_{P_{Z|Y}: I(Y;Z) \leq G(I^*)} I(X;Z) \tag{A.8}$$

$$= \max_{P_{Z|Y}: I(Y;Z) = G(I^*)} I(X;Z) \tag{A.9}$$

$$\geq I^* \tag{A.10}$$

$$= I(R^*). \tag{A.11}$$

The fact that $I(R)$ is a non-decreasing function implies

$$G(I^*) \geq R^*. \tag{A.12}$$

Combining (A.5) and (A.12) gives $G(I^*) = R^*$, so that $G(y)$ is the inverse function of $I(R)$ for $0 \leq R \leq R_{\max}$. Consequently, $G(y)$ is convex and non-decreasing, so that we can relax the equality constraint in (A.1) to $I(X;Z) \geq y$. Therefore, writing

$$R(I) = \min_{P_{Z|Y}} I(Y;Z) \quad \text{s.t.} \quad I(X;Z) \geq I, \tag{A.13}$$

the function $R(I)$ is the inverse function of $I(R)$, for $0 \leq R \leq R_{\max}$. ■

B

Proofs for Chapter 3

B.1. Message passing rules

To derive the message passing rules given in (3.19) and (3.20), consider Figure 3.5. The message passing rules for function nodes are applied in the following to compute

$$\ell_{1,m}^{(\mathrm{A})} = \ln\left(\frac{\mu_{\mathrm{A}}(x_{1,m}=0)}{\mu_{\mathrm{A}}(x_{1,m}=1)}\right). \tag{B.1}$$

Since

$$\mu_{\mathrm{A}}(x_{1,m}) = \sum_{x_2} P_{Z|X_1X_2}(z_m|x_{1,m},x_2)\mu_{\mathrm{E}}(x_2), \tag{B.2}$$

one obtains

$$\ell_{1,m}^{(\mathrm{A})} \tag{B.3}$$

$$= \ln\left(\frac{P_{Z|X_1X_2}(z_m|x_{1,m}=0,x'_{2,m}=0)\mu_{\mathrm{E}}(x'_{2,m}=0)+P_{Z|X_1X_2}(z_m|x_{1,m}=0,x'_{2,m}=1)\mu_{\mathrm{E}}(x'_{2,m}=1)}{P_{Z|X_1X_2}(z_m|x_{1,m}=1,x'_{2,m}=0)\mu_{\mathrm{E}}(x'_{2,m}=0)+P_{Z|X_1X_2}(z_m|x_{1,m}=1,x'_{2,m}=1)\mu_{\mathrm{E}}(x'_{2,m}=1)}\right)$$

$$= \ln\left(\frac{1+\frac{P_{Z|X_1X_2}(z_m|x_{1,m}=0,x'_{2,m}=0)}{P_{Z|X_1X_2}(z_m|x_{1,m}=0,x'_{2,m}=1)}\frac{\mu_{\mathrm{E}}(x'_{2,m}=0)}{\mu_{\mathrm{E}}(x'_{2,m}=1)}}{\frac{P_{Z|X_1X_2}(z_m|x_{1,m}=1,x'_{2,m}=0)}{P_{Z|X_1X_2}(z_m|x_{1,m}=0,x'_{2,m}=1)}\frac{\mu_{\mathrm{E}}(x'_{2,m}=0)}{\mu_{\mathrm{E}}(x'_{2,m}=1)}+\frac{P_{Z|X_1X_2}(z_m|x_{1,m}=1,x'_{2,m}=1)}{P_{Z|X_1X_2}(z_m|x_{1,m}=0,x'_{2,m}=1)}}\right), \tag{B.4}$$

which with the definitions in (3.16)-(3.18) yields (3.19). Along the same lines, one can also verify (3.20).

B.2. Proof of Proposition 3.2

The proof of Proposition 3.2 consists of two parts. First, observe that

$$\begin{aligned} I(X_1, X_2; q_1(L)) &= H(q_1(L)) - H(q_1(L)|X_1, X_2) && \text{(B.5)}\\ &= H(q_1(L)) - H(q_1(L)|X_1, X_2, X) && \text{(B.6)}\\ &\geq H(q_1(L)) - H(q_1(L)|X) && \text{(B.7)}\\ &= I(X; q_1(L)). && \text{(B.8)} \end{aligned}$$

where (B.6) holds since $X = X_1 \oplus X_2$ is a deterministic function of X_1 and X_2, and (B.7) holds since conditioning does not increase entropy. Next, we show that $(X_1, X_2) \leftrightarrow X \leftrightarrow L$ forms a Markov chain, i.e., that L is independent of (X_1, X_2) given X. To that end, define

$$g(\lambda_1, \lambda_2) \triangleq \ln\left(\frac{1 + e^{\lambda_1+\lambda_2}}{e^{\lambda_1} + e^{\lambda_2}}\right), \qquad \lambda_1 \in \mathbb{R}, \lambda_2 \in \mathbb{R}, \tag{B.9}$$

and observe that $g(-\lambda_1, -\lambda_2) = g(\lambda_1, \lambda_2)$. Using the independence of X_1 and X_2 and the symmetry assumption on $p_{L_i|X_i}(\ell_i|x_i)$, and substituting $\lambda_i' = -\lambda_i$, we obtain for the conditional CDF of L that

$$\begin{aligned} F_{L|X_1X_2}(\ell|0,0) &= \Pr\left[L \leq \ell | X_1 = 0, X_2 = 0\right] && \text{(B.10)}\\ &= \iint\limits_{(\lambda_1,\lambda_2):g(\lambda_1,\lambda_2)\leq\ell} p_{L_1L_2|X_1X_2}(\lambda_1, \lambda_2|0,0)\mathrm{d}\lambda_1\mathrm{d}\lambda_2 && \text{(B.11)}\\ &= \iint\limits_{(\lambda_1,\lambda_2):g(\lambda_1,\lambda_2)\leq\ell} p_{L_1L_2|X_1X_2}(-\lambda_1, -\lambda_2|1,1)\mathrm{d}\lambda_1\mathrm{d}\lambda_2 && \text{(B.12)}\\ &= \iint\limits_{(\lambda_1',\lambda_2'):g(\lambda_1',\lambda_2')\leq\ell} p_{L_1L_2|X_1X_2}(\lambda_1', \lambda_2'|1,1)\mathrm{d}\lambda_1'\mathrm{d}\lambda_2' && \text{(B.13)}\\ &= F_{L|X_1X_2}(\ell|1,1). && \text{(B.14)} \end{aligned}$$

Along the same lines, one can show that $F_{L|X_1,X_2}(\ell|0,1) = F_{L|X_1,X_2}(\ell|1,0)$, so that L is independent of (X_1, X_2) given $X = X_1 \oplus X_2$. Since $(X_1, X_2) \leftrightarrow X \leftrightarrow L$ forms a Markov chain, also $(X_1, X_2) \leftrightarrow X \leftrightarrow L \leftrightarrow q_1(L)$ forms a Markov chain; consequently, $I(X; q_1(L)) \geq I(X_1, X_2; q_1(L))$ by the data processing inequality, which together with (B.8) gives $I(X; q_1(L)) = I(X_1, X_2; q_1(L))$. For the second part of the proof, we write

$$\begin{aligned} I(X_i; q_1(L)) &= H(X_i) - H(X_i|q_1(L)) && \text{(B.15)}\\ &\leq H(X_i) - H(X_i|X, q_1(L)) && \text{(B.16)}\\ &= H(X_i) - H(X_i|X) && \text{(B.17)}\\ &= 0, && \text{(B.18)} \end{aligned}$$

which, together with the non-negativity of mutual information, gives $I(X_i; q_1(L)) = 0$. ■

C

Proofs for Chapter 4

C.1. Proof of Theorem 4.3

To avoid numerous indices i, we prove Theorem 4.3 for a very general setup. To that end, let (X, Y, W) be a triple of discrete random variables with fixed joint probability mass function P_{XYW}, and let X, Y, and W take on values in the finite sets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{W}$, respectively. Furthermore, we require that $W \leftrightarrow X \leftrightarrow Y$ forms a Markov chain. Let Z be a discrete random variable[1] taking on values from the finite set $\mathcal{Z}$, and define, for $0 \leq s \leq H(Y|W)$, the function

$$J(s) \triangleq \inf H(X|W, Z) \tag{C.1}$$

subject to the constraints

$$\begin{aligned} &H(Y|W, Z) = s, \\ &Z \text{ and } X \text{ are conditionally independent given } Y. \end{aligned} \tag{C.2}$$

The last constraint in the definition of $J(s)$ is equivalent to requiring that $X \leftrightarrow Y \leftrightarrow Z$ forms a Markov chain. Note that the definition of $J(s)$ is very similar to the one of $F(s)$ in [WW75], except for the conditioning[2] on W in (C.1) and (C.2). Therefore, the proof of Theorem 4.3 closely follows [WW75].

[1] We will later set $W = Y_{\mathrm{d},i}$, $X = X_i$, $Y = Y_{\mathrm{r},i}$, and $Z = \hat{Y}_{\mathrm{r},i}$, respectively, to recover the system model of Section 4.3.

[2] Note that the labels for the random variables involved are chosen to be consistent with Chapter 2; therefore, our notation differs from the one in [WW75].

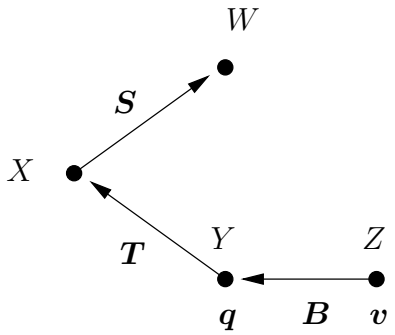


Figure C.1.: Random variables and their distributions.

In order to simplify the exposition, we use the following compact notation. Suppose that $|\mathcal{Y}| = n$, $|\mathcal{X}| = m$, and $|\mathcal{W}| = \ell$. Denote the distribution P_Y by a vector $\boldsymbol{q} = [q_1, q_2, \ldots, q_n]^{\mathrm{T}}$ in the probability simplex Δ_n with $q_i = P_Y(i)$, let $\boldsymbol{T}$ be an $m \times n$ stochastic matrix with $[\boldsymbol{T}]_{i,j} = P_{X|Y}(i|j)$, and let $\boldsymbol{S}$ be an $\ell \times m$ stochastic matrix with $[\boldsymbol{S}]_{i,j} = P_{W|X}(i|j)$. Let Z be a k-ary random variable, $k > 0$, with distribution $\boldsymbol{v} = [v_1, v_2, \ldots, v_k]^{\mathrm{T}}$, $\boldsymbol{v} \in \Delta_k$, where $v_i = P_Z(i)$, and let $\boldsymbol{B}$ be the $n \times k$ stochastic matrix having $\boldsymbol{b}_a$ for its a-th column, $a = 1, 2, \ldots, k$, where $\boldsymbol{b}_a \in \Delta_n$. The situation is summarized in Figure C.1.

Now apply the following concatenation of three channels to Z, for any choice of $\boldsymbol{v}$ and $\boldsymbol{B}$. The first channel has transition matrix $\boldsymbol{B}$, producing the random variable Y' with marginal distribution

$$\boldsymbol{p} = \boldsymbol{B}\boldsymbol{v} = \sum_{a=1}^{k} v_a \boldsymbol{b}_a. \tag{C.3}$$

The second channel has transition matrix $\boldsymbol{T}$, producing the random variable X' with marginal distribution $\boldsymbol{T}\boldsymbol{p}$, and the third channel has transition matrix $\boldsymbol{S}$, producing the random variable W' with marginal distribution $\boldsymbol{S}\boldsymbol{T}\boldsymbol{p}$. Note that the Markov condition from (C.2) is satisfied, and if $\boldsymbol{p} = \boldsymbol{q}$, then (Y', X', W') have the same joint distribution as (Y, X, W).

For any choice of $\boldsymbol{v}$ and the $\boldsymbol{b}_a$, we compute the distribution (C.3) and the quantities

$$\chi_1^{(\boldsymbol{T},\boldsymbol{S})}(\boldsymbol{b}_a) \triangleq \sum_w P_{W'|Z}(w|a) H(Y'|W' = w, Z = a) \tag{C.4}$$

and

$$\chi_2^{(\boldsymbol{T},\boldsymbol{S})}(\boldsymbol{b}_a) \triangleq \sum_w P_{W'|Z}(w|a) H(X'|W' = w, Z = a), \tag{C.5}$$

where $\chi_1^{(\boldsymbol{T},\boldsymbol{S})} : \Delta_n \to \mathbb{R}$ and $\chi_2^{(\boldsymbol{T},\boldsymbol{S})} : \Delta_n \to \mathbb{R}$. We observe that for fixed $\boldsymbol{T}$ and $\boldsymbol{S}$, $\chi_1^{(\boldsymbol{T},\boldsymbol{S})}$ and $\chi_2^{(\boldsymbol{T},\boldsymbol{S})}$ are bounded from above, and we denote these upper bounds by $\chi_{1,\max}^{(\boldsymbol{T},\boldsymbol{S})}$ and $\chi_{2,\max}^{(\boldsymbol{T},\boldsymbol{S})}$. Using (C.4) and (C.5), the quantities needed in the definition of $J(s)$ are

compactly written as

$$\begin{aligned}\xi &= H(Y'|W', Z)\\ &= \sum_{w,z} P_{W'Z}(w,z) H(Y'|W'=w, Z=z)\\ &= \sum_{z} P_Z(z) \sum_{w} P_{W'|Z}(w|z) H(Y'|W'=w, Z=z)\\ &= \sum_{a=1}^{k} v_a \chi_1^{(\boldsymbol{T},\boldsymbol{S})}(\boldsymbol{b}_a), \end{aligned} \tag{C.6}$$

and

$$\begin{aligned}\eta &= H(X'|W', Z)\\ &= \sum_{w,z} P_{W'Z}(w,z) H(X'|W'=w, Z=z)\\ &= \sum_{z} P_Z(z) \sum_{w} P_{W'|Z}(w|z) H(X'|W'=w, Z=z)\\ &= \sum_{a=1}^{k} v_a \chi_2^{(\boldsymbol{T},\boldsymbol{S})}(\boldsymbol{b}_a). \end{aligned} \tag{C.7}$$

Those choices of Z satisfying (C.2) correspond to those choices of k, $\boldsymbol{v}$, and the $\boldsymbol{b}_a$ for which (C.3), (C.6), and (C.7) yield $\boldsymbol{p} = \boldsymbol{q}$ and $\xi = s$.

Next, for $\boldsymbol{b} \in \Delta_n$, we are interested in the mapping

$$\boldsymbol{b} \to \left(\boldsymbol{b}, \chi_1^{(\boldsymbol{T},\boldsymbol{S})}(\boldsymbol{b}), \chi_2^{(\boldsymbol{T},\boldsymbol{S})}(\boldsymbol{b})\right). \tag{C.8}$$

Since Δ_n is the $(n-1)$ dimensional probability simplex, the product $\Delta_n \times [0, \chi_{1,\max}^{(\boldsymbol{T},\boldsymbol{S})}] \times [0, \chi_{2,\max}^{(\boldsymbol{T},\boldsymbol{S})}]$ is an $(n+1)$-dimensional convex polytope. The mapping (C.8) assigns a point of this $(n+1)$-dimensional polytope to every point $\boldsymbol{b} \in \Delta_n$. Denote by $\mathcal{S}$ the set of all such points $(\boldsymbol{b}, \chi_1^{(\boldsymbol{T},\boldsymbol{S})}(\boldsymbol{b}), \chi_2^{(\boldsymbol{T},\boldsymbol{S})}(\boldsymbol{b}))$; hence, $\mathcal{S}$ is the image of Δ_n under the mapping (C.8). The set $\mathcal{S}$ is compact and connected, since it is the continuous image[3] of the compact connected set Δ_n. Let $\mathcal{C}$ denote the convex hull of $\mathcal{S}$; owing to the compactness of $\mathcal{S}$, the convex hull $\mathcal{C}$ is also compact. We have the following two lemmas.

Lemma C.1 ([WW75])**.** The set of all triples $(\boldsymbol{p}, \xi, \eta)$ determined by (C.3), (C.6), and (C.7), for all integers $k > 0$, $\boldsymbol{v} \in \Delta_k$, $\boldsymbol{b}_a \in \Delta_n$, $a = 1, 2, \ldots, k$, is precisely $\mathcal{C}$.

Lemma C.2 ([WW75])**.** Every point of $\mathcal{C}$ can be obtained by (C.3), (C.6), and (C.7) with $k \le n+1$, that is, it suffices to consider random variables Z taking at most $n+1$ values.

The proof of both lemmas follows along the lines of the proof Lemmas 2.1 and 2.2 in [WW75]; the proof of Lemma C.2 is based on the strengthening of Carathéodory's Theorem [Wit80, Section III] for connected sets.

[3]Continuity of $\chi_1^{(\boldsymbol{T},\boldsymbol{S})}$ and $\chi_2^{(\boldsymbol{T},\boldsymbol{S})}$ follows from [Yeu02, Chapter 2.3].

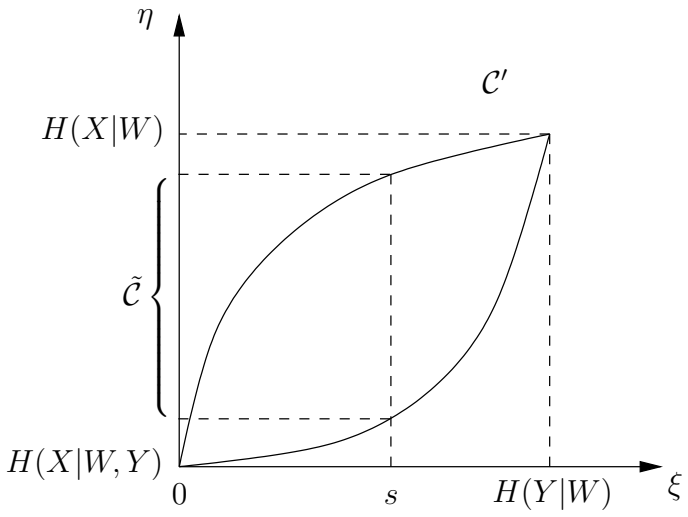


Figure C.2.: Illustration of the sets $\tilde{\mathcal{C}}$ and $\mathcal{C}'$.

Lemma C.3. Provided that $0 \leq s \leq H(Y|W)$, the infimum in the definition of $J(s)$ is a minimum, and can be attained with Z taking at most $n+1$ values.

Proof. Let $\tilde{\mathcal{C}} = \{\eta | (\boldsymbol{q}, s, \eta) \in \mathcal{C}\}$, i.e., $\tilde{\mathcal{C}}$ is obtained by intersecting $\mathcal{C}$ with the straight line determined by the conditions $\boldsymbol{p} = \boldsymbol{q}$ and $\xi = s$, and projecting the result of the intersection onto the η-axis, as illustrated in Figure C.2. Since $\mathcal{C}$ is compact, so is the interval $\tilde{\mathcal{C}}$; therefore, one can replace the infimum in the definition of $J(s)$ with a minimum, unless the intersection is empty. For $0 \leq s \leq H(Y|W)$, however, the intersection is never empty, because

a) for $k = 1$, $\boldsymbol{b}_1 = \boldsymbol{q}$ gives $\boldsymbol{p} = \boldsymbol{q}$ and $\xi = H(Y|W)$,

b) for $k = n$, $\boldsymbol{v} = \boldsymbol{q}$, $\boldsymbol{B} = \boldsymbol{I}_n$ gives $\boldsymbol{p} = \boldsymbol{q}$ and $\xi = 0$,

c) and due to the convexity of $\mathcal{C}$, the set $\mathcal{C}$ contains points with $\boldsymbol{p} = \boldsymbol{q}$ and $\xi = s$ for all $0 \leq s \leq H(Y|W)$.

Let $\eta^* = \min_\eta \tilde{\mathcal{C}}$. Thus, $(\boldsymbol{q}, s, \eta^*) \in \mathcal{C}$, and due to Lemma C.2, the minimum can be attained with Z taking no more than $n+1$ values. ■

Lemma C.4. The function $J(s)$ is a convex function for $0 \leq s \leq H(Y|W)$.

Proof. Consider the set $\mathcal{C}' = \{(\xi, \eta) | (\boldsymbol{q}, \xi, \eta) \in \mathcal{C}\}$, which is the projection on the (ξ, η)-plane of the intersection of $\mathcal{C}$ with the two-dimensional plane determined by $\boldsymbol{p} = \boldsymbol{q}$. The two-dimensional plane determined by $\boldsymbol{p} = \boldsymbol{q}$ in $(n+1)$ dimensions can represented by the intersection of $(n-1)$ hyperplanes; thus, that two-dimensional plane is convex. Since $\mathcal{C}$ is convex, its intersection with the convex two-dimensional plane determined by $\boldsymbol{p} = \boldsymbol{q}$ is also convex [BV04, Chapter 2.3.1], and as the projection of a convex set onto some of its coordinates is convexity-preserving [BV04, Chapter 2.3.2], the set $\mathcal{C}'$ is convex. The

function $J(s)$ is precisely the lower boundary of the plane convex set $\mathcal{C}'$; due to the convexity of $\mathcal{C}'$, the convexity of $J(s)$ follows. ■

Lemma C.5. The function $J(s)$ is monotonically non-decreasing in s. Therefore, the constraints (C.2) can be replaced by

$$\begin{aligned} &H(Y|W,Z) \geq s, \\ &Z \text{ and } X \text{ are conditionally independent given } Y. \end{aligned} \tag{C.9}$$

Proof. By the chain rule for mutual information, we have

$$I(X;Y,Z|W) = I(X;Z|W) + I(X;Y|Z,W) \tag{C.10}$$
$$= I(X;Y|W) + I(X;Z|Y,W). \tag{C.11}$$

Due to the Markov condition, $I(X;Z|Y,W) = 0$, so that

$$I(X;Y|W) = I(X;Z|W) + I(X;Y|Z,W) \tag{C.12}$$
$$\geq I(X;Z|W) \tag{C.13}$$

because of the non-negativity of mutual information. Therefore, (C.13) implies

$$H(X|W,Z) \geq H(X|W,Y). \tag{C.14}$$

Let W, X, Y, Z satisfy (C.2). Then, $J(s) \geq H(X|W,Y)$ by (C.14). Also, if Z is chosen as $Z \equiv Y$, then $J(0) = H(X|W,Y)$. Therefore, $J(s) \geq J(0)$ for all $0 \leq s \leq H(Y|W)$, and with the convexity of $J(s)$, the monotonicity of $J(s)$ and the lemma follows. ■

Summarizing the above, we have for $0 \leq s \leq H(Y|W)$

$$\begin{aligned} J(s) = \min\ & H(X|W,Z) \quad \text{s.t.} \quad H(Y|W,Z) \geq s, \\ & Z \text{ and } X \text{ are conditionally independent given } Y, \end{aligned} \tag{C.15}$$

where $J(s)$ is convex and monotonically non-decreasing in s, and the minimum in the definition of $J(s)$ can be achieved with Z taking at most $n+1$ values.

Next, define the function

$$\bar{J}(r) = \max_{P_{Z|Y}} I(X;Z|W) \quad \text{s.t.} \quad I(Y;Z|W) \leq r, \tag{C.16}$$

for $0 \leq r \leq H(Y|W)$, where the Markov condition is captured by maximizing over the distribution $P_{Z|Y}$ only. Since $\bar{J}(r)$ can be recovered from $J(s)$ by

$$\bar{J}(r) = H(X|W) - J(H(Y|W) - r), \tag{C.17}$$

the function $\bar{J}(r)$ is concave and monotonically non-decreasing in r, and the maximum in the definition of $\bar{J}(r)$ can be achieved with Z taking no more than $n+1$ values.

Finally, the connection with Theorem 4.3 is established by setting $W = Y_{\mathrm{d},i}$, $X = X_i$, $Y = Y_{\mathrm{r},i}$, and $Z = \hat{Y}_{\mathrm{r},i}$, respectively; the properties of $I_i(r_i)$ follow from the properties of $\bar{J}(r)$ defined in (C.16). ■

C.2. Proof of Theorem 4.4

a) Let $P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$ be such that $r_i = I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = 0$. By the data processing inequality, we have

$$I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) \leq I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = 0. \tag{C.18}$$

Combining (C.18) with the non-negativity of mutual information yields $I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = 0$, and therefore $I_i(r_i = 0) = 0$. Next, suppose that $P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$ is such that $\hat{Y}_{\mathrm{r},i} \equiv Y_{\mathrm{r},i}$; consequently, we have

$$I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i}) \tag{C.19}$$

$$I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) = H(Y_{\mathrm{r},i}|Y_{\mathrm{d},i}), \tag{C.20}$$

and we conclude that $I_i(r_i = H(Y_{\mathrm{r},i}|Y_{\mathrm{d},i})) = I(X_i; Y_{\mathrm{r},i}|Y_{\mathrm{d},i})$.

b) This follows from the concavity of $I_i(r_i)$ and Theorem 4.4a).

c) This follows since $I_i(r_i)$ is concave and non-decreasing for $0 \leq r_i \leq r_{i,\max}$. ■

C.3. Proof of Proposition 4.6

To begin the proof, note that we have

$$\sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|X_i, Y_{\mathrm{d},i}) \tag{C.21}$$

$$= \sum_{i=1}^{M} \alpha_i \left[H(\hat{Y}_{\mathrm{r},i}|X_i, Y_{\mathrm{d},i}) - H(\hat{Y}_{\mathrm{r},i}|X_i, Y_{\mathrm{d},i}, Y_{\mathrm{r},i}) \right] \tag{C.22}$$

$$= \sum_{i=1}^{M} \alpha_i \left[H(\hat{Y}_{\mathrm{r},i}|X_i, Y_{\mathrm{d},i}) - H(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}, Y_{\mathrm{r},i}) \right] \tag{C.23}$$

$$= \sum_{i=1}^{M} \alpha_i \left[H(\hat{Y}_{\mathrm{r},i}|X_i, Y_{\mathrm{d},i}) - H(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}, Y_{\mathrm{r},i}) + H(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) - H(\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) \right] \tag{C.24}$$

$$= \sum_{i=1}^{M} \alpha_i \left[I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) - I(X_i; \hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) \right]. \tag{C.25}$$

With (C.25), the conditions in (4.62) and (4.63) become

$$\sum_{i\in\mathcal{I}} R_i < \sum_{i\in\mathcal{I}} \alpha_i I(X_i;\hat{Y}_{\mathrm{r},i},Y_{\mathrm{d},i}) \tag{C.26}$$

$$\sum_{i\in\mathcal{I}} R_i < \sum_{i\in\mathcal{I}} \alpha_i I(X_i;\hat{Y}_{\mathrm{r},i},Y_{\mathrm{d},i}) + \sum_{i\in\mathcal{I}^c} \alpha_i I(X_i;\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}) + \bar{\alpha} I(X_{\mathrm{r}};Y_{\mathrm{d,r}}) - \sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i};\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{d},i}), \tag{C.27}$$

for all subsets $\mathcal{I}\subseteq\{1,2,\ldots,M\}$.

Let $\boldsymbol{R}$ be a rate vector satisfying (4.4) and (4.5) for some fixed distribution $\prod_{i=1}^{M} P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$. Since (4.5) holds, the right-hand side of (C.27) is not smaller than the right-hand side of (C.26), for any $\mathcal{I}$. Hence, the NNC rate bounds reduce to

$$\sum_{i\in\mathcal{I}} R_i < \sum_{i\in\mathcal{I}} \alpha_i I(X_i;\hat{Y}_{\mathrm{r},i},Y_{\mathrm{d},i}), \tag{C.28}$$

for all subsets $\mathcal{I}\subseteq\{1,2,\ldots,M\}$. But (C.28) holds since (4.4) is satisfied. We thus have $\mathcal{R}_{\mathrm{CF}}\subseteq\mathcal{R}_{\mathrm{NNC}}$. ■

C.4. Proof of Theorem 4.7

To prove the theorem, we show that at the optimal distribution $\prod_{i=1}^{M} P^{*}_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$ maximizing the total sum-rate $\sum_{i=1}^{M} R_i$ we must have

$$\sum_{i=1}^{M} \alpha_i I(X_i;\hat{Y}_{\mathrm{r},i},Y_{\mathrm{d},i}) = \sum_{i=1}^{M} \alpha_i I(X_i;Y_{\mathrm{d},i}) + \bar{\alpha} I(X_{\mathrm{r}};Y_{\mathrm{d,r}}) - \sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i};\hat{Y}_{\mathrm{r},i}|X_i,Y_{\mathrm{d},i}). \tag{C.29}$$

In order to show (C.29) at the optimal distribution, suppose that $\prod_{i=1}^{M} P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$ is the optimum, and that we have

$$\sum_{i=1}^{M} \alpha_i I(X_i;\hat{Y}_{\mathrm{r},i},Y_{\mathrm{d},i}) > \sum_{i=1}^{M} \alpha_i I(X_i;Y_{\mathrm{d},i}) + \bar{\alpha} I(X_{\mathrm{r}};Y_{\mathrm{d,r}}) - \sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i};\hat{Y}_{\mathrm{r},i}|X_i,Y_{\mathrm{d},i}) \tag{C.30}$$

at that optimum. Now, let $\hat{Y}'_{\mathrm{r},i}=\hat{Y}_{\mathrm{r},i}$ with probability $1-\epsilon_i$, $\epsilon_i\in[0,1]$, and let $\hat{Y}'_{\mathrm{r},i}=\emptyset$ with probability ϵ_i. Consequently, $Y_{\mathrm{r},i}\leftrightarrow\hat{Y}_{\mathrm{r},i}\leftrightarrow\hat{Y}'_{\mathrm{r},i}$ forms a Markov chain, so that we have

$$I(X_i;\hat{Y}'_{\mathrm{r},i},Y_{\mathrm{d},i}) \le I(X_i;\hat{Y}_{\mathrm{r},i},Y_{\mathrm{d},i}) \tag{C.31}$$

$$I(Y_{\mathrm{r},i};\hat{Y}'_{\mathrm{r},i}|X_i,Y_{\mathrm{d},i}) \le I(Y_{\mathrm{r},i};\hat{Y}_{\mathrm{r},i}|X_i,Y_{\mathrm{d},i}) \tag{C.32}$$

by the data processing inequality. Moreover, $I(X_i;\hat{Y}'_{\mathrm{r},i},Y_{\mathrm{d},i})$ and $I(Y_{\mathrm{r},i};\hat{Y}'_{\mathrm{r},i}|X_i,Y_{\mathrm{d},i})$ are decreasing in ϵ_i, so that the left-hand side of (C.30) is decreasing, and the right-hand side

of (C.30) is increasing as the ϵ_i's increase. Thus there are ϵ_i^*, $i = 1, 2, \ldots, M$, such that

$$\sum_{i=1}^{M} \alpha_i I(X_i; \hat{Y}'_{\mathrm{r},i}, Y_{\mathrm{d},i}) = \sum_{i=1}^{M} \alpha_i I(X_i; Y_{\mathrm{d},i}) + \bar{\alpha} I(X_{\mathrm{r}}; Y_{\mathrm{d,r}}) - \sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i}; \hat{Y}'_{\mathrm{r},i} | X_i, Y_{\mathrm{d},i}), \quad \text{(C.33)}$$

and the bound on the total sum-rate using $\prod_{i=1}^{M} P_{\hat{Y}'_{\mathrm{r},i}|Y_{\mathrm{r},i}}$ is larger than the bound on the total sum-rate using $\prod_{i=1}^{M} P_{\hat{Y}_{\mathrm{r},i}|Y_{\mathrm{r},i}}$, which implies (C.29). Moreover, (C.29) and (C.25) imply that we have

$$\bar{\alpha} I(X_{\mathrm{r}}; Y_{\mathrm{d,r}}) = \sum_{i=1}^{M} \alpha_i I(Y_{\mathrm{r},i}; \hat{Y}_{\mathrm{r},i} | Y_{\mathrm{d},i}) \quad \text{(C.34)}$$

at the optimal distribution maximizing the total sum-rate. Since (C.34) holds, the NNC rate region of (C.26) and (C.27) is equivalent to

$$R_i < \alpha_i I(X_i; \hat{Y}_{\mathrm{r},i}, Y_{\mathrm{d},i}), \quad i = 1, 2, \ldots, M, \quad \text{(C.35)}$$

at the sum-rate optimal distribution. ■

D

Proofs for Chapter 5

D.1. Proof of Theorem 5.1

First assume that $\mathcal{M}_{Q_1}^{\mathcal{X}}(\boldsymbol{h})$ is an information lossless finite-state machine, and recall that $S_n = f(X_{n-L_h+2}^n)$. We have

$$
\begin{aligned}
I(X;Q_1(Y)) &= \lim_{n\to\infty} \frac{1}{n} I(X^n; Z^n | S_0 = s_0) && \text{(D.1)}\\
&= \lim_{n\to\infty} \frac{1}{n} [\underbrace{I(X_{n-L_h+2}^n; Z^n | S_0 = s_0)}_{\geq 0} + I(X^{n-L_h+1}; Z^n | S_0 = s_0, S_n)] && \text{(D.2)}\\
&\geq \lim_{n\to\infty} \frac{1}{n} I(X^{n-L_h+1}; Z^n | S_0 = s_0, S_n) && \text{(D.3)}\\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{s_n} P_{S_n}(s_n) \Big[H(X^{n-L_h+1} | S_0 = s_0, S_n = s_n) && \text{(D.4)}\\
&\qquad\qquad - H(X^{n-L_h+1} | Z^n, S_0 = s_0, S_n = s_n) \Big],
\end{aligned}
$$

where $P_{S_n}(s_n) = \Lambda^{1-L_h}$ since i.i.d. signaling is assumed. Since $\mathcal{M}_{Q_1}^{\mathcal{X}}(\boldsymbol{h})$ is information lossless, we have

$$
H(X^{n-L_h+1} | Z^n, S_0 = s_0, S_n = s_n) = 0, \quad \forall s_n, \tag{D.5}
$$

so that

$$
I(X;Q_1(Y)) \geq \lim_{n\to\infty} \frac{1}{n} \sum_{s_n} P_{S_n}(s_n) H(X^{n-L_h+1} | S_0 = s_0, S_n = s_n) \tag{D.6}
$$

$$= \lim_{n\to\infty} \frac{n - L_h + 1}{n} \log_2(\Lambda) \tag{D.7}$$

$$= \log_2(\Lambda). \tag{D.8}$$

Combining (D.8) with

$$I(X; Q_1(Y)) \leq \lim_{n\to\infty} \frac{1}{n} H(X^n) = \log_2(\Lambda) \tag{D.9}$$

implies $I(X; Q_1(Y)) = \log_2(\Lambda)$ if $\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h})$ is information lossless.

In order to prove the "only if" part of Theorem 5.1, we use properties of the testing graph [Eve65] $\mathcal{T}(\mathcal{M}^{\mathcal{X}}(\boldsymbol{h}))$ of the finite-state machine $\mathcal{M}^{\mathcal{X}}(\boldsymbol{h})$. The construction of the testing graph relies on the notion of compatible states[1] of $\mathcal{M}^{\mathcal{X}}(\boldsymbol{h})$.

Definition D.1 (Compatible states [Eve65])**.** Let $\mathcal{M}^{\mathcal{X}}(\boldsymbol{h})$ be a finite-state machine. An unordered pair of states (i, j), $i, j \in \mathcal{S}$, is said to be compatible if there exists a state k such that there exists a branch leading from $s_{t-1} = k$ to $s_t = i$ producing the output $\tilde{y}_t$ and a branch leading from $s_{t-1} = k$ to $s_t = j$ producing the same output $\tilde{y}_t$. A pair of states (i, j) is also called compatible if there exists a compatible pair of states (k, ℓ) such that there exist branches leading from $s_{t-1} = k$ to $s_t = i$ and from $s_{t-1} = \ell$ to $s_t = j$, both producing the same output $\tilde{y}_t$.

Definition D.2 (Testing graph [Eve65])**.** The testing graph $\mathcal{T}(\mathcal{M}^{\mathcal{X}}(\boldsymbol{h}))$ of a finite-state machine $\mathcal{M}^{\mathcal{X}}(\boldsymbol{h})$ is a directed graph defined in the following way:

1. The nodes of $\mathcal{T}(\mathcal{M}^{\mathcal{X}}(\boldsymbol{h}))$ correspond to compatible pairs of states of $\mathcal{M}^{\mathcal{X}}(\boldsymbol{h})$.

2. If there exists a branch leading from $s_{t-1} = k$ to $s_t = i$ and a branch leading from $s_{t-1} = \ell$ to $s_t = j$ producing the same output $\tilde{y}_t$, and if (k, ℓ) is a compatible pair of states, then $\mathcal{T}(\mathcal{M}^{\mathcal{X}}(\boldsymbol{h}))$ has a directed arc leading from the node corresponding to (k, ℓ) into the node corresponding to (i, j). That arc is labeled with $\tilde{y}_t$.

Example D.1. Consider $\mathcal{X} = \{\pm 1\}$, $\boldsymbol{h} = [1, -1]^{\mathrm{T}}/\sqrt{2}$, and $Q_1(y) = \mathbb{1}_{y\geq 0}$. The trellis section for $\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h})$ and the testing graph $\mathcal{T}(\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h}))$ are shown in Figure D.1.

Now suppose that $\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h})$ is not information lossless. By [Eve65, Theorem 2], the testing graph $\mathcal{T}(\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h}))$ must contain at least one pair (j, j), $j \in \mathcal{S}$, that has the same vertex repeated, since $\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h})$ is assumed to be information lossy. Now suppose that the first pair generated during the construction of $\mathcal{T}(\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h}))$ that has the same vertex repeated is (e, e). Then, by definition, there is a finite positive integer N', (at least) two different input sequences $\bar{x}_t^{t+N'-1}$ and $\bar{\bar{x}}_t^{t+N'-1}$ of length N', and states $s_{t-1} = i$ and $s_{t+N'-1} = e$, yielding the same output of the machine $\tilde{z}_t^{t+N'-1}$. Further, since the state $s_{t+N'-1} = f(x_{t+N'-L_h+1}^{t+N'-1})$ is determined by the $L_h - 1$ previous inputs to the machine, we must have $N' \geq L_h$ so that $\bar{x}_t^{t+N'-1}$ and $\bar{\bar{x}}_t^{t+N'-1}$ differ.

[1] Compatibility and the testing graph are defined in terms of the vertices of a general coding graph in [Eve65]. We explicitly define both for finite-state machines here.

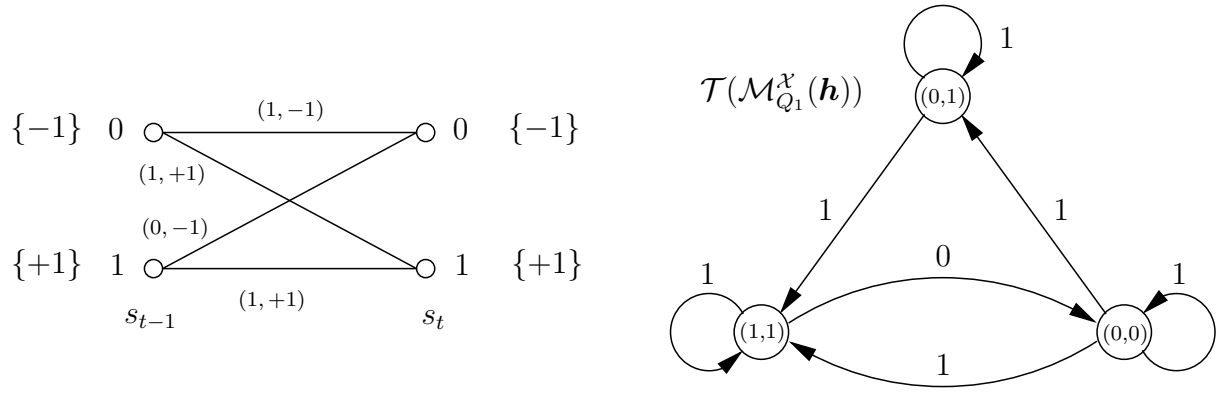


Figure D.1.: Trellis section for $\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h})$, where $\boldsymbol{h} = [1, -1]^{\mathrm{T}}/\sqrt{2}$, $\mathcal{X} = \{\pm 1\}$, and $Q_1(y) = \mathbb{1}_{y \geq 0}$. The testing graph $\mathcal{T}(\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h}))$ is also shown.

Example D.1 (continued)**.** In $\mathcal{T}(\mathcal{M}^{\mathcal{X}}_{Q_1}(\boldsymbol{h}))$, there exists a path of length 2 of the form $(0,0) \xrightarrow{1} (0,1) \xrightarrow{1} (1,1)$. Therefore, there are two distinct input sequences $\bar{x}_t^{t+1} = \{-1, +1\}$ and $\bar{\bar{x}}_t^{t+1} = \{+1, +1\}$ of length $N' = 2$ that lead from state $s_{t-1} = 0$ via $s_t = 0$ or $s_t = 1$ to state $s_{t+1} = 1$, and produce the same output $\tilde{z}_t^{t+1} = \{1, 1\}$.

Next, assume that S_0 is uniformly distributed, and that $n = aN'$ is an integer multiple of N', so that

$$I(X^n; Z^n|S_0) = H(X^n|S_0) - \sum_{i=1}^{a} H(X_{N'(i-1)+1}^{N'i}|Z^n, X^{N'(i-1)}, S_0) \tag{D.10}$$

$$= H(X^n|S_0) - \sum_{i=1}^{a} H(X_{N'(i-1)+1}^{N'i}|Z_{N'(i-1)+1}^n, S_{N'(i-1)}) \tag{D.11}$$

$$\leq H(X^n|S_0) - \sum_{i=1}^{a} H(X_{N'(i-1)+1}^{N'i}|Z_{N'(i-1)+1}^n, S_{N'(i-1)}, X_{N'i-L_h+2}^n) \tag{D.12}$$

$$= H(X^n|S_0) - \sum_{i=1}^{a} H(X_{N'(i-1)+1}^{N'i}|Z_{N'(i-1)+1}^{N'i}, S_{N'(i-1)}, X_{N'i-L_h+2}^{N'i}) \tag{D.13}$$

$$= H(X^n|S_0) - \sum_{i=1}^{a} H(X_{N'(i-1)+1}^{N'i-L_h+1}|Z_{N'(i-1)+1}^{N'i}, S_{N'(i-1)}, S_{N'i}) \tag{D.14}$$

$$= n \log_2(\Lambda) - aH(X^{N'-L_h+1}|Z^{N'}, S_0, S_{N'}), \tag{D.15}$$

where (D.12) follows because conditioning does not increase entropy, (D.13) follows since $Z_{N'i+1}^n$ and $X_{N'i+1}^n$ are independent of $X_{N'(i-1)+1}^{N'i}$ given $X_{N'i-L_h+2}^{N'i}$, and (D.15) follows from stationarity. Expanding the conditional entropy yields

$$H(X^{N'-L_h+1}|Z^{N'}, S_0, S_{N'})$$
$$= \sum_{z^{N'}, s_0, s_{N'}} P_{Z^{N'} S_0 S_{N'}}(z^{N'}, s_0, s_{N'}) H(X^{N'-L_h+1}|Z^{N'} = z^{N'}, S_0 = s_0, S_{N'} = s_{N'}). \tag{D.16}$$

Since there are at least two different input sequences $\bar{x}^{N'}$ and $\bar{\bar{x}}^{N'}$, and states $s_0 = i$ and $s_{N'} = e$, yielding the same output of the machine $\tilde{z}^{N'}$, we have $H(X^{N'-L_h+1}|Z^{N'} = \tilde{z}^{N'}, S_0 = i, S_{N'} = e) > 0$ and $P_{Z^{N'}S_0S_{N'}}(\tilde{z}^{N'}, i, e) > 0$. Consequently, we have

$$H(X^{N'-L_h+1}|Z^{N'}, S_0, S_{N'}) > 0, \tag{D.17}$$

which implies

$$\begin{aligned} I(X;Q_1(Y)) &\leq \lim_{n\to\infty} \frac{1}{n}[n\log_2(\Lambda) - aH(X^{N'-L_h+1}|Z^{N'}, S_0, S_{N'})] && \text{(D.18)}\\ &= \log_2(\Lambda) - \frac{1}{N'}H(X^{N'-L_h+1}|Z^{N'}, S_0, S_{N'}) && \text{(D.19)}\\ &< \log_2(\Lambda). && \text{(D.20)} \end{aligned}$$

■

Example D.1 (continued)**.** For $s_0 = 0$, $s_2 = 1$, and $\tilde{z}^2 = \{1, 1\}$, we have $H(X_1|Z^2 = \tilde{z}^2, S_0 = 0, S_2 = 1) = 1$, and

$$\begin{aligned} &\Pr[Z^2 = \tilde{z}^2, S_0 = 0, S_2 = 1]\\ &\quad= \Pr[S_0 = 0]\Pr[Z_1 = 1|S_0 = 0]\Pr[S_2 = 1, Z_2 = 1|S_0 = 0, Z_1 = 1]\\ &\quad= \underbrace{\Pr[S_0 = 0]}_{=1/2}\ \underbrace{\Pr[Z_1 = 1|S_0 = 0]}_{=1}\ \underbrace{\Pr[X_2 = 1, Z_2 = 1|S_0 = 0, Z_1 = 1]}_{=1/2}\\ &\quad= 1/4. \end{aligned} \tag{D.21}$$

Therefore, $H(X_1|Z^2, S_0, S_2) \geq 1/4$, and by (D.19), the information rate is upper bounded by

$$\begin{aligned} I(X;Q_1(Y)) &\leq \log_2(\Lambda) - \frac{1}{N'}H(X_1|Z^2, S_0, S_2) && \text{(D.22)}\\ &\leq 1 - \frac{1}{2}\cdot\frac{1}{4} && \text{(D.23)}\\ &= \frac{7}{8}. && \text{(D.24)} \end{aligned}$$

D.2. Proof of Theorem 5.2

To prove the second part of Theorem 5.2, it suffices to provide an example of $\mathcal{X}$ and $\boldsymbol{h}$, and show that $I(X;Q_u(Y)) < \log_2(\Lambda)$ for that particular example. Consider $\Lambda = 2$, $\mathcal{X} = \{\pm 1\}$, and $\boldsymbol{h} = [2/3, -1/3, 2/3]^T$, and recall that the set of noise-free channel outputs is $\tilde{\mathcal{Y}} = \{-5/3, -1, -1/3, 1/3, 1, 5/3\}$ for that channel. Therefore, the single-bit uniform quantizer for that channel is a slicer with quantization function $Q_u(y) = \mathbb{1}_{y\geq 0}$. Writing $Z_{u,t} = Q_u(Y_t)$, it is tedious but straightforward to compute the conditional entropy for

$\sigma^2 = 0$, namely

$$H\Big(X_k\Big|Z_{\mathrm{u},k}^{k+2}, S_{k-1}, X_{k+1}^{k+2}\Big) = 1/8, \quad k = 1, 2, \ldots, n-2, \tag{D.25}$$

which implies

$$H\Big(X_{n-1}\Big|Z_{\mathrm{u},n-1}^{n}, S_{n-2}, X_n\Big) \geq 1/8 \tag{D.26}$$

$$H\Big(X_n\Big|Z_{\mathrm{u},n}, S_{n-1}\Big) \geq 1/8, \tag{D.27}$$

since conditioning does not increase entropy. We then have the following upper bound on the rate:

$$\begin{aligned}
I(X; Q_{\mathrm{u}}(Y)) &= \lim_{n\to\infty} \frac{1}{n}[H(X^n|S_0) - H(X^n|Z_{\mathrm{u}}^n, S_0)] && \text{(D.28)}\\
&= \lim_{n\to\infty} \frac{1}{n}\left[\sum_{i=1}^{n} H(X_i) - H\left(X_i\,|Z_{\mathrm{u}}^n, S_{i-1}\right)\right] && \text{(D.29)}\\
&= 1 - \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} H\left(X_i\,\Big|Z_{\mathrm{u},i}^n, S_{i-1}\right) && \text{(D.30)}\\
&\leq 1 - \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} H\left(X_i\,\Big|Z_{\mathrm{u},i}^n, S_{i-1}, X_{i+1}^n\right) && \text{(D.31)}\\
&\leq 7/8 && \text{(D.32)}\\
&< \log_2(\Lambda), && \text{(D.33)}
\end{aligned}$$

where

- ▷ (D.30) follows since X_i is independent of Z_{u}^{i-1}, given S_{i-1},
- ▷ (D.31) follows since conditioning does not increase entropy,
- ▷ and (D.32) follows from (D.25)-(D.27). ■

D.3. Proof of Lemmas 5.3 and 5.4

Since the channel input is assumed i.i.d., we have

$$I(X; Q_1(Y)) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} H(X_i) - H(X_i|Z_i^n, X_{i-L_h+1}^{i-1}). \tag{D.34}$$

Hence, for Lemma 5.3, we obtain

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} H(X_i) - H(X_i|Z_i^n, X_{i-L_h+1}^{i-1}) &\geq \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} H(X_i) - H(X_i|Z_i^{i+K}, X_{i-L_h+1}^{i-1}) && \text{(D.35)}\\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I(X_i; Z_i^{i+K}|X_{i-L_h+1}^{i-1}), && \text{(D.36)}
\end{aligned}$$

where the inequality holds for all $K \geq 0$ since conditioning does not increase entropy. To prove Lemma 5.4, note that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i) - H(X_i|Z_i^n, X_{i-L_h+1}^{i-1}) \\
&\qquad \leq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i) - H(X_i|Z_i^n, X_{i-L_h+1}^{i-1}, X_{i+1}^{n}) && \text{(D.37)} \\
&\qquad = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i) - H(X_i|Z_i^{i+L_h-1}, X_{i-L_h+1}^{i-1}, X_{i+1}^{i+L_h-1}) && \text{(D.38)} \\
&\qquad = I(X_i; Z_i^{i+L_h-1}|X_{i-L_h+1}^{i-1}, X_{i+1}^{i+L_h-1}), && \text{(D.39)}
\end{aligned}$$

where the inequality follows since conditioning does not increase entropy, and the last equality follows from stationarity. The latter bound is reminiscent of the bound employed in the proof of the matched filter bound [SOW91, Theorem 2], which holds for the ISI channel with i.i.d. signaling and continuous output. ■

E

Proofs for Chapter 6

E.1. Proof of Theorem 6.2

Multiplying both sides of (6.11) with $p_\Theta(\theta)$ and differentiating gives

$$\frac{\partial\, p_\Theta(\theta)\bar{B}(\theta)}{\partial\theta} = -\sum_{z^n} P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) + \sum_{z^n}\left[\hat{\theta}(z^n)-\theta\right]\frac{\partial\, P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}{\partial\theta}. \tag{E.1}$$

Integrating both sides with respect to θ yields

$$\begin{aligned} p_\Theta(\theta)\bar{B}(\theta)\Big|_{-\infty}^{\infty} &= -\int_{-\infty}^{\infty}\sum_{z^n} P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta + \int_{-\infty}^{\infty}\sum_{z^n}\left[\hat{\theta}(z^n)-\theta\right]\frac{\partial\, P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}{\partial\theta}\mathrm{d}\theta \\ &= -1 + \int_{-\infty}^{\infty}\sum_{z^n}\left[\hat{\theta}(z^n)-\theta\right]\frac{\partial\, P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}{\partial\theta}\mathrm{d}\theta, \end{aligned} \tag{E.2}$$

and the assumptions in (6.12) and (6.13) ensure that

$$p_\Theta(\theta)\bar{B}(\theta)\Big|_{-\infty}^{\infty} = 0, \tag{E.3}$$

so that we have

$$\sum_{z^n}\int_{-\infty}^{\infty}\left[\hat{\theta}(z^n)-\theta\right]\frac{\partial\, P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}{\partial\theta}\mathrm{d}\theta = 1. \tag{E.4}$$

Next, observe that

$$\frac{\partial\, P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}{\partial\theta} = \frac{\partial\, \ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)}{\partial\theta} P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta). \tag{E.5}$$

Substituting (E.5) into (E.4) and rewriting, we have

$$\sum_{z^n}\int_{-\infty}^{\infty}\left[\frac{\partial\, \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial\theta}\sqrt{P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}\right]\left[\left[\hat{\theta}(z^n)-\theta\right]\sqrt{P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}\right]\mathrm{d}\theta = 1,$$

and, by applying the Schwarz inequality to the integral in the summation, we have

$$\begin{aligned} 1 \leq \sum_{z^n}\Bigg\{ &\left(\int_{-\infty}^{\infty}\left[\frac{\partial\, \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial\theta}\right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta\right)^{\frac{1}{2}} \\ &\times\left(\int_{-\infty}^{\infty}\left[\hat{\theta}(z^n)-\theta\right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta\right)^{\frac{1}{2}}\Bigg\}. \end{aligned} \tag{E.6}$$

Next, apply the Cauchy inequality to the summation in (E.6) to obtain

$$\begin{aligned} 1 \leq &\left(\sum_{z^n}\int_{-\infty}^{\infty}\left[\frac{\partial\, \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial\theta}\right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta\right)^{\frac{1}{2}} \\ &\times\left(\sum_{z^n}\int_{-\infty}^{\infty}\left[\hat{\theta}(z^n)-\theta\right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta\right)^{\frac{1}{2}}, \end{aligned} \tag{E.7}$$

or, equivalently,

$$\begin{aligned} 1 \leq &\sum_{z^n}\int_{-\infty}^{\infty}\left[\frac{\partial\, \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial\theta}\right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta \\ &\times\sum_{z^n}\int_{-\infty}^{\infty}\left[\hat{\theta}(z^n)-\theta\right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta \end{aligned} \tag{E.8}$$

$$= \mathrm{E}\left[\left(\frac{\partial\, \ln\left(P_{Z^n|\Theta}(Z^n|\Theta)p_\Theta(\Theta)\right)}{\partial\Theta}\right)^2\right]\cdot\mathrm{E}\left[\left(\Theta-\hat{\theta}(Z^n)\right)^2\right]. \tag{E.9}$$

Rearranging of the above inequality yields the BCRLB in terms of the square of the first derivative of $\ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)$. To derive the BCRLB in terms of the second derivative

of $\ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)$, note that

$$\sum_{z^n} P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) = p_\Theta(\theta). \tag{E.10}$$

Next, differentiation of (E.10) twice on both sides with respect to θ yields

$$\sum_{z^n} \frac{\partial}{\partial\theta}\left[\frac{\partial\, P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}{\partial\theta}\right] = \frac{\partial^2 p_\Theta(\theta)}{\partial\theta^2}. \tag{E.11}$$

Now, substitution of (E.5) into the left-hand side of (E.11) yields

$$\begin{aligned}
&\sum_{z^n} \frac{\partial}{\partial\theta}\left[\frac{\partial\, P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)}{\partial\theta}\right] \\
&= \sum_{z^n} \frac{\partial}{\partial\theta}\left[\frac{\partial \ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)}{\partial\theta} P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right] \qquad \text{(E.12)} \\
&= \sum_{z^n} \left[\frac{\partial^2 \ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)}{\partial\theta^2} + \left(\frac{\partial \ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)}{\partial\theta}\right)^2\right] P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta) \\
&= \frac{\partial^2 p_\Theta(\theta)}{\partial\theta^2}, \qquad \text{(E.13)}
\end{aligned}$$

and by integrating with respect to θ, we have

$$\begin{aligned}
&\sum_{z^n} \int_{-\infty}^{\infty} \left[\frac{\partial^2 \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial\theta^2} + \left(\frac{\partial \ln(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta))}{\partial\theta}\right)^2\right] P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta \\
&= \int_{-\infty}^{\infty} \frac{\partial^2 p_\Theta(\theta)}{\partial\theta^2}\mathrm{d}\theta = \left.\frac{\partial p_\Theta(\theta)}{\partial\theta}\right|_{-\infty}^{\infty} = 0, \qquad \text{(E.14)}
\end{aligned}$$

where the last equality holds due to Condition 2) of Theorem 6.2. Therefore, we have

$$\mathrm{E}\left[\left(\frac{\partial \ln\left(P_{Z^n|\Theta}(Z^n|\Theta)p_\Theta(\Theta)\right)}{\partial\Theta}\right)^2\right] = \mathrm{E}\left[-\frac{\partial^2 \ln\left(P_{Z^n|\Theta}(Z^n|\Theta)p_\Theta(\Theta)\right)}{\partial\Theta^2}\right]. \tag{E.15}$$

Inserting (E.15) into (E.9), we have

$$1 \le \mathrm{E}\left[-\frac{\partial^2 \ln\left(P_{Z^n|\Theta}(Z^n|\Theta)p_\Theta(\Theta)\right)}{\partial\Theta^2}\right] \cdot \mathrm{E}\left[\left(\Theta - \hat{\theta}(Z^n)\right)^2\right], \tag{E.16}$$

which is Theorem 6.2.

Finally, we comment on the tightness of Theorem 6.2. Equality in the application of the

Schwarz inequality in (E.6) holds if and only if

$$\frac{\partial \ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)}{\partial\theta} = \left[\hat{\theta}(z^n) - \theta\right] c(z^n), \tag{E.17}$$

for all z^n and θ, where $c(z^n)$ is a function of z^n.

Furthermore, equality in (E.8) holds if and only if there is a constant $\tilde{c} \in \mathbb{R}$ such that

$$\left(\int_{-\infty}^{\infty} \left[\frac{\partial \ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)}{\partial\theta}\right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta\right)^{\frac{1}{2}} = \tilde{c}\left(\int_{-\infty}^{\infty} \left[\hat{\theta}(z^n) - \theta\right]^2 P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\mathrm{d}\theta\right)^{\frac{1}{2}}, \tag{E.18}$$

for all z^n. Combining (E.17) and (E.18), we see that equality holds in Theorem 6.2 if and only if

$$c(z^n) = \tilde{c} \quad \forall z^n, \tag{E.19}$$

i.e., if and only if there is a real-valued constant $\tilde{c}$ such that

$$\frac{\partial \ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)}{\partial\theta} = \left[\hat{\theta}(z^n) - \theta\right] \tilde{c}. \tag{E.20}$$

Differentiating (E.20) with respect to θ gives the condition

$$\frac{\partial^2 \ln\left(P_{Z^n|\Theta}(z^n|\theta)p_\Theta(\theta)\right)}{\partial\theta^2} = -\tilde{c}. \tag{E.21}$$

■

E.2. Proof of Theorem 6.3

E.2.1. Conditions for the applicability of Theorem 6.2

We first show that the conditions for the applicability of Theorem 6.2 are satisfied. To show that Condition 1 is satisfied, let $\hat{h}(z^n)$ be any estimator with $|\hat{h}(z^n)| < \infty$ for all z^n, and let

$$\bar{B}(h) = \sum_{z^n} \left[\hat{h}(z^n) - h\right] P_{Z^n|H}(z^n|h), \tag{E.22}$$

so that

$$\lim_{h\to\pm\infty} \bar{B}(h)p_H(h) = \sum_{z^n} \hat{h}(z^n) \lim_{h\to\pm\infty} P_{Z^n|H}(z^n|h)p_H(h) - \sum_{z^n} \lim_{h\to\pm\infty} hP_{Z^n|H}(z^n|h)p_H(h). \tag{E.23}$$

Since $0 \leq P_{Z^n|H}(z^n|h) \leq 1$ for any z^n and h, and since $H \sim \mathcal{N}(0, \sigma_h^2)$, we have

$$\lim_{h \to \pm\infty} P_{Z^n|H}(z^n|h) p_H(h) = 0, \quad \forall z^n \tag{E.24}$$

$$\lim_{h \to \pm\infty} h P_{Z^n|H}(z^n|h) p_H(h) = 0, \quad \forall z^n, \tag{E.25}$$

and therefore

$$\lim_{h \to \pm\infty} \bar{B}(h) p_H(h) = 0. \tag{E.26}$$

To check Condition 2, we compute

$$\frac{\partial p_H(h)}{\partial h} = -\frac{h}{\sigma_h^2} \frac{1}{\sqrt{2\pi}\sigma_h} e^{-h^2/(2\sigma_h^2)}, \tag{E.27}$$

so that clearly

$$\lim_{h \to \pm\infty} \frac{\partial p_H(h)}{\partial h} = 0. \tag{E.28}$$

E.2.2. Computation of the second derivative of $\ln\left(P_{Z^n|H}(z^n|h) p_H(h)\right)$

Let $p_N(a)$ denote the distribution of a zero-mean Gaussian random variable with variance σ_n^2, i.e.,

$$p_N(a) \triangleq \frac{1}{\sqrt{2\pi}\sigma_n} e^{-a^2/(2\sigma_n^2)}, \tag{E.29}$$

and let $\mathrm{Q}(x)$ be the Q-function, i.e.,

$$\mathrm{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} \mathrm{d}u, \tag{E.30}$$

whose derivative is

$$\frac{\partial \mathrm{Q}(x)}{\partial x} = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{E.31}$$

We also use the Kronecker delta function

$$\delta(x) \triangleq \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{E.32}$$

With

$$P_{Z^n|H}(z^n|h) = \prod_{i=1}^{n} P_{Z_i|HZ^{i-1}}(z_i|h, z^{i-1}) \tag{E.33}$$

$$= \prod_{i=1}^{n} \left[\delta(z_i) \mathrm{Q}\left(\frac{h - \tau_i(z^{i-1})}{\sigma_n} \right) + \delta(z_i - 1) \left(1 - \mathrm{Q}\left(\frac{h - \tau_i(z^{i-1})}{\sigma_n} \right) \right) \right], \tag{E.34}$$

we have

$$\ln\left(P_{Z^n|H}(z^n|h) p_H(h) \right) = \ln p_H(h) + \sum_{i=1}^{n} \ln P_{Z_i|HZ^{i-1}}(z_i|h, z^{i-1}) \tag{E.35}$$

$$= -\ln(\sqrt{2\pi}\sigma_h) - \frac{h^2}{2\sigma_h^2} \tag{E.36}$$

$$+ \sum_{i=1}^{n} \ln \left(\delta(z_i) \mathrm{Q}\left(\frac{h - d_i}{\sigma_n} \right) + \delta(z_i - 1) \left(1 - \mathrm{Q}\left(\frac{h - d_i}{\sigma_n} \right) \right) \right),$$

where we write $d_i = \tau_i(z^{i-1})$ for brevity, and keep in mind that d_i is a function of z^{i-1}. We have

$$\frac{\partial \ln\left(P_{Z^n|H}(z^n|h) p_H(h) \right)}{\partial h} = -\frac{h}{\sigma_h^2} + \sum_{i=1}^{n} p_N(h - d_i) \left(\frac{\delta(z_i - 1)}{1 - \mathrm{Q}\left(\frac{h-d_i}{\sigma_n} \right)} - \frac{\delta(z_i)}{\mathrm{Q}\left(\frac{h-d_i}{\sigma_n} \right)} \right), \tag{E.37}$$

and

$$\frac{\partial^2 \ln\left(P_{Z^n|H}(z^n|h) p_H(h) \right)}{\partial h^2} = -\frac{1}{\sigma_h^2} + \sum_{i=1}^{n} \left\{ p_N(h - d_i) \left[\delta(z_i) \left(\frac{\frac{h-d_i}{\sigma_n^2}}{\mathrm{Q}\left(\frac{h-d_i}{\sigma_n} \right)} - \frac{p_N(h - d_i)}{\mathrm{Q}\left(\frac{h-d_i}{\sigma_n} \right)^2} \right) \right. \right.$$

$$\left. \left. - \delta(z_i - 1) \left(\frac{\frac{h-d_i}{\sigma_n^2}}{1 - \mathrm{Q}\left(\frac{h-d_i}{\sigma_n} \right)} + \frac{p_N(h - d_i)}{\left(1 - \mathrm{Q}\left(\frac{h-d_i}{\sigma_n} \right) \right)^2} \right) \right] \right\}. \tag{E.38}$$

Since (E.38) is not a constant, we conclude based on (E.21) that there exists no estimator of H that achieves the BCRLB

$$\left\{ -\mathrm{E}\left[\frac{\partial^2 \ln\left(P_{Z^n|H}(Z^n|H) p_H(H) \right)}{\partial H^2} \right] \right\}^{-1} \tag{E.39}$$

with equality.

E.2.3. Evaluation of the BCLRB

Defining

$$G(z^n,h) \triangleq \frac{\partial^2 \ln\left(P_{Z^n|H}(z^n|h)p_H(h)\right)}{\partial h^2} \tag{E.40}$$

$$g(z^i,h) \triangleq p_N(h-d_i)\left[\delta(z_i)\left(\frac{\frac{h-d_i}{\sigma_n^2}}{\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)} - \frac{p_N(h-d_i)}{\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)^2}\right)\right. \tag{E.41}$$
$$\left. - \delta(z_i-1)\left(\frac{\frac{h-d_i}{\sigma_n^2}}{1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)} + \frac{p_N(h-d_i)}{\left(1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right)^2}\right)\right],$$

we have

$$G(z^n,h) = -\frac{1}{\sigma_h^2} + \sum_{i=1}^{n} g(z^i,h), \tag{E.42}$$

so that

$$-\mathrm{E}[G(Z^n,H)] \tag{E.43}$$
$$= \int_{-\infty}^{\infty} \sum_{z^n}\left(\frac{1}{\sigma_h^2} - \sum_{i=1}^{n} g(z^i,h)\right) P_{Z^n|H}(z^n|h)p_H(h)\mathrm{d}h \tag{E.44}$$
$$= \frac{1}{\sigma_h^2} - \sum_{i=1}^{n}\int_{-\infty}^{\infty}\sum_{z^n} g(z^i,h)P_{Z^n|H}(z^n|h)p_H(h)\mathrm{d}h \tag{E.45}$$
$$= \frac{1}{\sigma_h^2} - \sum_{i=1}^{n}\int_{-\infty}^{\infty}\sum_{z^i} g(z^i,h)P_{Z^i|H}(z^i|h)p_H(h)\underbrace{\left[\sum_{z_{i+1}^n} P_{Z_{i+1}^n|HZ^i}(z_{i+1}^n|h,z^i)\right]}_{=1\ \forall z^i,h}\mathrm{d}h \tag{E.46}$$
$$= \frac{1}{\sigma_h^2} - \sum_{i=1}^{n}\int_{-\infty}^{\infty}\sum_{z^{i-1}} P_{Z^{i-1}|H}(z^{i-1}|h)\left[\sum_{z_i} g(z^i,h)P_{Z_i|HZ^{i-1}}(z_i|h,z^{i-1})\right]p_H(h)\mathrm{d}h. \tag{E.47}$$

The expression in the inner square braces of (E.47) is equal to

$$\sum_{z_i} g(z^i,h)P_{Z_i|HZ^{i-1}}(z_i|h,z^{i-1})$$
$$= \sum_{z_i} g(z^i,h)\left(\delta(z_i)\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right) + \delta(z_i-1)\left(1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right)\right) \tag{E.48}$$
$$= -\frac{1}{2\pi\sigma_n^2}e^{-(h-d_i)^2/\sigma_n^2}\frac{1}{\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right]}. \tag{E.49}$$

Inserting (E.49) into (E.47), we obtain

$$-\mathrm{E}[G(Z^n,H)] = \frac{1}{\sigma_h^2} + \sum_{i=1}^{n} \int_{-\infty}^{\infty} \sum_{z^{i-1}} P_{Z^{i-1}|H}(z^{i-1}|h) \frac{e^{-(h-d_i)^2/\sigma_n^2}}{2\pi\sigma_n^2 \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right]} p_H(h)\mathrm{d}h, \tag{E.50}$$

which seems hard to solve in closed form since $d_i = \tau_i(z^{i-1})$ is a function of z^{i-1}. In order to find an upper bound on $-\mathrm{E}[G(Z^n,H)]$ and therefore a lower bound on $\{-\mathrm{E}[G(Z^n,H)]\}^{-1}$, we need the following lemma.

Lemma E.1. Let

$$\lambda(x) = \frac{e^{-x^2}}{\mathrm{Q}(x)\left[1-\mathrm{Q}(x)\right]}, \tag{E.51}$$

with $x \in \mathbb{R}$. We have $\lambda(x) \leq 4e^{-(1-2/\pi)x^2} \approx 4e^{-0.3634x^2}$.

Proof. First, we express $\lambda(x)$ in terms of the error function

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}\mathrm{d}u, \tag{E.52}$$

yielding

$$\lambda(x) = \frac{4e^{-x^2}}{1-\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)^2}. \tag{E.53}$$

To bound the square of the error function in (E.53), we employ an upper bound due to Williams [Wil46] and Pólya [Pól49], which was complemented with a lower bound by Chu [Chu55]; the bound is

$$\mathrm{erf}(x) \leq \sqrt{1-e^{-4x^2/\pi}}, \quad x \geq 0. \tag{E.54}$$

Consequently, since $1-\mathrm{erf}\left(x/\sqrt{2}\right)^2$ is symmetric around the origin, i.e., $1-\mathrm{erf}\left(-x/\sqrt{2}\right)^2 = 1-\mathrm{erf}\left(x/\sqrt{2}\right)^2$, we have

$$1-\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)^2 \geq e^{-2x^2/\pi}, \quad x \in \mathbb{R}, \tag{E.55}$$

and therefore

$$\lambda(x) \leq \frac{4e^{-x^2}}{e^{-2x^2/\pi}} = 4e^{-\left(1-\frac{2}{\pi}\right)x^2}, \quad x \in \mathbb{R}. \tag{E.56}$$

■

To find an upper bound on $-\mathrm{E}[G(Z^n, H)]$, we apply Lemma E.1 to obtain, for any d_i,

$$\frac{e^{-(h-d_i)^2/\sigma_n^2}}{\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right]} \leq 4e^{-\left(1-\frac{2}{\pi}\right)\frac{(h-d_i)^2}{\sigma_n^2}} \leq 4, \quad \forall h. \tag{E.57}$$

Inserting (E.57) into (E.50) yields

$$-\mathrm{E}[G(Z^n, H)] \leq \frac{1}{\sigma_h^2} + \sum_{i=1}^{n} \int_{-\infty}^{\infty} \underbrace{\sum_{z^{i-1}} P_{Z^{i-1}|H}(z^{i-1}|h)}_{=1\ \forall h} \frac{4}{2\pi\sigma_n^2} p_H(h)\mathrm{d}h \tag{E.58}$$

$$= \frac{1}{\sigma_h^2} + \sum_{i=1}^{n} \frac{2}{\pi\sigma_n^2} \underbrace{\int_{-\infty}^{\infty} p_H(h)\mathrm{d}h}_{=1} \tag{E.59}$$

$$= \frac{1}{\sigma_h^2} + n\frac{2}{\pi\sigma_n^2}. \tag{E.60}$$

Therefore, we have

$$\mathrm{E}\left[(H-\hat{h}(Z^n))^2\right] \geq \frac{1}{\frac{1}{\sigma_h^2} + n\frac{2}{\pi\sigma_n^2}} \tag{E.61}$$

$$= \frac{\sigma_h^2}{1 + n\frac{2}{\pi}\frac{\sigma_h^2}{\sigma_n^2}}, \tag{E.62}$$

which completes the proof of Theorem 6.3. ■

E.3. Proof of Theorem 6.5

To prove Theorem 6.5 for $n = 1$, we compute $\mathrm{E}[(H - \hat{h}_{\mathrm{MMSE}}(Z_1))^2]$ exactly without using the BCRLB. Due to the properties of MMSE estimation [Poo94, Section IV.B], we have

$$\mathrm{E}\left[(H-\hat{h}_{\mathrm{MMSE}}(Z_1))^2\right] = \mathrm{E}\left[H^2\right] - \mathrm{E}\left[\hat{h}_{\mathrm{MMSE}}^2(Z_1)\right] = \sigma_h^2 - \mathrm{E}\left[\hat{h}_{\mathrm{MMSE}}^2(Z_1)\right]. \tag{E.63}$$

Since H is a zero-mean random variable, the initial quantizer threshold is $d_1 = \tau_1(z^0) = 0$, so that $P_{Z_1}(1) = P_{Z_1}(0) = 1/2$, and the MMSE estimator has the symmetry property

$$\hat{h}_{\mathrm{MMSE}}(z_1 = 1) = \int_{-\infty}^{\infty} h\left(1 - \mathrm{Q}\left(\frac{h}{\sigma_n}\right)\right) \frac{p_H(h)}{P_{Z_1}(1)}\mathrm{d}h \tag{E.64}$$

$$= \frac{1}{P_{Z_1}(1)} \underbrace{\int_{-\infty}^{\infty} h p_H(h) \mathrm{d}h}_{=\mathrm{E}[H]=0} - \frac{1}{P_{Z_1}(0)} \int_{-\infty}^{\infty} h \mathrm{Q}\left(\frac{h}{\sigma_n}\right) p_H(h) \mathrm{d}h \tag{E.65}$$

$$= -\hat{h}_{\mathrm{MMSE}}(z_1 = 0). \tag{E.66}$$

Consequently, the expectation of $\hat{h}^2_{\mathrm{MMSE}}(Z_1)$ is given by

$$\mathrm{E}\left[\hat{h}^2_{\mathrm{MMSE}}(Z_1)\right] = \sum_{z_1} P_{Z_1}(z_1) \hat{h}^2_{\mathrm{MMSE}}(z_1) \tag{E.67}$$

$$= \hat{h}^2_{\mathrm{MMSE}}(z_1 = 0) \tag{E.68}$$

$$= \frac{2}{\pi \sigma_h^2} \left(\int_{-\infty}^{\infty} h \, \mathrm{Q}\left(\frac{h}{\sigma_n}\right) e^{-h^2/(2\sigma_h^2)} \mathrm{d}h \right)^2. \tag{E.69}$$

Defining

$$\gamma(\sigma_n^2, \sigma_h^2) \triangleq \int_{-\infty}^{\infty} h \, \mathrm{Q}\left(\frac{h}{\sigma_n}\right) e^{-h^2/(2\sigma_h^2)} \mathrm{d}h, \tag{E.70}$$

we have

$$\mathrm{E}\left[(H - \hat{h}_{\mathrm{MMSE}}(Z_1))^2\right] = \sigma_h^2 - \frac{2}{\pi \sigma_h^2} \gamma^2(\sigma_n^2, \sigma_h^2). \tag{E.71}$$

For $n \geq 2$, we begin the proof with (E.50), which is

$$-\mathrm{E}[G(Z^n, H)] = \frac{1}{\sigma_h^2} + \sum_{i=1}^{n} \int_{-\infty}^{\infty} \sum_{z^{i-1}} P_{Z^{i-1}|H}(z^{i-1}|h) \frac{e^{-(h-d_i)^2/\sigma_n^2}}{2\pi\sigma_n^2 \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\left[1 - \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right]} p_H(h) \mathrm{d}h. \tag{E.72}$$

First consider the term for $i = 1$ in the summation in (E.72); since z^{i-1} is the empty sequence for $i = 1$, that term is

$$\int_{-\infty}^{\infty} \frac{e^{-(h-d_1)^2/\sigma_n^2}}{2\pi\sigma_n^2 \mathrm{Q}\left(\frac{h-d_1}{\sigma_n}\right)\left[1 - \mathrm{Q}\left(\frac{h-d_1}{\sigma_n}\right)\right]} p_H(h) \mathrm{d}h = \int_{-\infty}^{\infty} \frac{e^{-h^2/\sigma_n^2}}{2\pi\sigma_n^2 \mathrm{Q}\left(\frac{h}{\sigma_n}\right)\left[1 - \mathrm{Q}\left(\frac{h}{\sigma_n}\right)\right]} p_H(h) \mathrm{d}h, \tag{E.73}$$

where the equality follows since $d_1 = \tau_1(z^0) = 0$. Since there seems to be no closed-form solution available for the integral in (E.73), we compute

$$\bar{\gamma}(\sigma_n^2, \sigma_h^2) \triangleq \int_{-\infty}^{\infty} \frac{e^{-h^2/\sigma_n^2} e^{-h^2/(2\sigma_h^2)}}{\mathrm{Q}\left(\frac{h}{\sigma_n}\right)\left[1 - \mathrm{Q}\left(\frac{h}{\sigma_n}\right)\right]} \mathrm{d}h \tag{E.74}$$

by numerical integration so that (E.73) becomes

$$\frac{\bar{\gamma}(\sigma_n^2, \sigma_h^2)}{2\pi\sigma_n^2\sqrt{2\pi}\sigma_h}. \tag{E.75}$$

Next, consider the summation in (E.72) for $i \geq 2$ and note that $P_{Z^{i-1}|H}(z^{i-1}|h) \leq 1$ for all z^{i-1} and h. Therefore, we have

$$\begin{aligned}\sum_{i=2}^{n} \int_{-\infty}^{\infty} \sum_{z^{i-1}} P_{Z^{i-1}|H}(z^{i-1}|h) \frac{e^{-(h-d_i)^2/\sigma_n^2}}{2\pi\sigma_n^2 \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right]} p_H(h)\mathrm{d}h \\ \leq \sum_{i=2}^{n} \sum_{z^{i-1}} \int_{-\infty}^{\infty} \frac{e^{-(h-d_i)^2/\sigma_n^2}}{2\pi\sigma_n^2 \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right]} p_H(h)\mathrm{d}h. \end{aligned} \tag{E.76}$$

Since $p_H(h) \leq 1/(\sqrt{2\pi}\sigma_h)$ for all h, we can upper bound the integral in (E.76) by

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{e^{-(h-d_i)^2/\sigma_n^2} p_H(h)}{2\pi\sigma_n^2 \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right]} \mathrm{d}h &\leq \int_{-\infty}^{\infty} \frac{e^{-(h-d_i)^2/\sigma_n^2}}{2\pi\sigma_n^2 \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right]} \frac{1}{\sqrt{2\pi}\sigma_h} \mathrm{d}h \quad \text{(E.77)} \\ &= \frac{1}{2\pi\sigma_n^2\sqrt{2\pi}\sigma_h} \int_{-\infty}^{\infty} \frac{e^{-h^2/\sigma_n^2}}{\mathrm{Q}\left(\frac{h}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h}{\sigma_n}\right)\right]} \mathrm{d}h, \quad \text{(E.78)}\end{aligned}$$

for any $d_i = \tau_i(z^{i-1})$. Unfortunately, a closed-form solution for the integral in (E.78) does not seem to be available. Therefore, we compute

$$\bar{\bar{\gamma}}(\sigma_n^2) \triangleq \int_{-\infty}^{\infty} \frac{e^{-h^2/\sigma_n^2}}{\mathrm{Q}\left(\frac{h}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h}{\sigma_n}\right)\right]} \mathrm{d}h \tag{E.79}$$

by numerical integration. Consequently, we have

$$\begin{aligned}\sum_{i=2}^{n} \int_{-\infty}^{\infty} \sum_{z^{i-1}} P_{Z^{i-1}|H}(z^{i-1}|h) \frac{e^{-(h-d_i)^2/\sigma_n^2}}{2\pi\sigma_n^2 \mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\left[1-\mathrm{Q}\left(\frac{h-d_i}{\sigma_n}\right)\right]} p_H(h)\mathrm{d}h \\ \leq \sum_{i=2}^{n} \sum_{z^{i-1}} \frac{\bar{\bar{\gamma}}(\sigma_n^2)}{2\pi\sigma_n^2\sqrt{2\pi}\sigma_h} \quad \text{(E.80)} \\ = \sum_{i=2}^{n} 2^{i-1} \frac{\bar{\bar{\gamma}}(\sigma_n^2)}{2\pi\sigma_n^2\sqrt{2\pi}\sigma_h} \quad \text{(E.81)} \\ = (2^n - 2) \frac{\bar{\bar{\gamma}}(\sigma_n^2)}{2\pi\sigma_n^2\sqrt{2\pi}\sigma_h}. \quad \text{(E.82)}\end{aligned}$$

Finally, by inserting (E.82) and (E.75) into (E.72), we have

$$-\mathrm{E}[G(Z^n, H)] \leq \frac{1}{\sigma_h^2} + \frac{1}{2\pi\sigma_n^2\sqrt{2\pi}\sigma_h}\left(\bar{\gamma}(\sigma_n^2, \sigma_h^2) + (2^n - 2)\bar{\bar{\gamma}}(\sigma_n^2)\right), \tag{E.83}$$

and for $n \geq 2$ the lower bound on the MSE becomes

$$\mathrm{E}\left[(H - \hat{h}(Z^n))^2\right] \geq \left(\frac{1}{\sigma_h^2} + \frac{1}{2\pi\sigma_n^2\sqrt{2\pi}\sigma_h}\left(\bar{\gamma}(\sigma_n^2, \sigma_h^2) + (2^n - 2)\bar{\bar{\gamma}}(\sigma_n^2)\right)\right)^{-1} \tag{E.84}$$

$$= \frac{\sigma_h^2}{1 + \frac{\sigma_h}{2\pi\sigma_n^2\sqrt{2\pi}}\left(\bar{\gamma}(\sigma_n^2, \sigma_h^2) + (2^n - 2)\bar{\bar{\gamma}}(\sigma_n^2)\right)}. \tag{E.85}$$

■

E.4. Proof of Theorem 6.6

We begin the proof for $n = 1$, and an upper bound for $\gamma^2(\sigma_n^2, \sigma_h^2)$, which is

$$\gamma^2(\sigma_n^2, \sigma_h^2) = \left(\int_{-\infty}^{\infty} h\,\mathrm{Q}\left(\frac{h}{\sigma_n}\right) e^{-h^2/(2\sigma_h^2)}\mathrm{d}h\right)^2 \tag{E.86}$$

$$= \left(\frac{1}{2}\int_{-\infty}^{\infty} h\left(1 - \mathrm{erf}\left(\frac{h}{\sqrt{2}\sigma_n}\right)\right) e^{-h^2/(2\sigma_h^2)}\mathrm{d}h\right)^2 \tag{E.87}$$

$$= \left(\int_{0}^{\infty} h\,\mathrm{erf}\left(\frac{h}{\sqrt{2}\sigma_n}\right) e^{-h^2/(2\sigma_h^2)}\mathrm{d}h\right)^2 \tag{E.88}$$

$$= \left(\int_{0}^{\infty} \left[\sqrt{h}\,\mathrm{erf}\left(\frac{h}{\sqrt{2}\sigma_n}\right) e^{-h^2/(4\sigma_h^2)}\right]\left[\sqrt{h}e^{-h^2/(4\sigma_h^2)}\right]\mathrm{d}h\right)^2, \tag{E.89}$$

where we exploit in (E.87) that

$$\int_{-\infty}^{\infty} h\, e^{-h^2/(2\sigma_h^2)}\mathrm{d}h = 0. \tag{E.90}$$

Applying the Schwarz inequality to (E.89) yields

$$\gamma^2(\sigma_n^2, \sigma_h^2) \leq \left(\int_{0}^{\infty} h\,\mathrm{erf}^2\left(\frac{h}{\sqrt{2}\sigma_n}\right) e^{-h^2/(2\sigma_h^2)}\mathrm{d}h\right)\left(\int_{0}^{\infty} h\, e^{-h^2/(2\sigma_h^2)}\mathrm{d}h\right) \tag{E.91}$$

$$= \sigma_h^2 \int_{0}^{\infty} h\,\mathrm{erf}^2\left(\frac{h}{\sqrt{2}\sigma_n}\right) e^{-h^2/(2\sigma_h^2)}\mathrm{d}h, \tag{E.92}$$

where the last equality follows by

$$\int_0^\infty x\, e^{-ax^2} \mathrm{d}x = \frac{1}{2a}, \quad a > 0. \tag{E.93}$$

Applying the bound in (E.54) to (E.92) yields

$$\gamma^2(\sigma_n^2, \sigma_h^2) \leq \sigma_h^2 \int_0^\infty h \left(1 - e^{-2h^2/(\pi\sigma_n^2)}\right) e^{-h^2/(2\sigma_h^2)} \mathrm{d}h \tag{E.94}$$

$$= \sigma_h^2 \left(\sigma_h^2 - \int_0^\infty h\, e^{-(\pi\sigma_n^2 + 4\sigma_h^2)h^2/(2\pi\sigma_h^2\sigma_n^2)} \mathrm{d}h\right) \tag{E.95}$$

$$= \sigma_h^2 \left(\sigma_h^2 - \frac{\pi\sigma_h^2\sigma_n^2}{\pi\sigma_n^2 + 4\sigma_h^2}\right). \tag{E.96}$$

Inserting (E.96) into (E.71), we have the bound

$$\mathrm{E}\left[(H - h_{\mathrm{MMSE}}(Z_1))^2\right] \geq \sigma_h^2 - \frac{2}{\pi}\left(\sigma_h^2 - \frac{\pi\sigma_h^2\sigma_n^2}{\pi\sigma_n^2 + 4\sigma_h^2}\right) \tag{E.97}$$

$$= \left(1 - \frac{2}{\pi}\right)\sigma_h^2 + \frac{2\sigma_h^2\sigma_n^2}{\pi\sigma_n^2 + 4\sigma_h^2}. \tag{E.98}$$

Next, consider the case $n \geq 2$ and apply Lemma E.1 to $\bar{\gamma}(\sigma_n^2, \sigma_h^2)$ to obtain the bound

$$\bar{\gamma}(\sigma_n^2, \sigma_h^2) = \int_{-\infty}^\infty \frac{e^{-h^2/\sigma_n^2} e^{-h^2/(2\sigma_h^2)}}{\mathrm{Q}\left(\frac{h}{\sigma_n}\right)\left[1 - \mathrm{Q}\left(\frac{h}{\sigma_n}\right)\right]} \mathrm{d}h \tag{E.99}$$

$$\leq 4 \int_{-\infty}^\infty e^{-(1-2/\pi)h^2/\sigma_n^2} e^{-h^2/(2\sigma_h^2)} \mathrm{d}h \tag{E.100}$$

$$= 4\sqrt{\frac{2\pi\sigma_n^2\sigma_h^2}{2(1-2/\pi)\sigma_h^2 + \sigma_n^2}}, \tag{E.101}$$

where we used

$$\int_{-\infty}^\infty e^{-ax^2} \mathrm{d}x = \sqrt{\frac{\pi}{a}}, \quad a > 0. \tag{E.102}$$

Likewise, we obtain the bound

$$\bar{\bar{\gamma}}(\sigma_n^2) = \int_{-\infty}^\infty \frac{e^{-h^2/\sigma_n^2}}{\mathrm{Q}\left(\frac{h}{\sigma_n}\right)\left[1 - \mathrm{Q}\left(\frac{h}{\sigma_n}\right)\right]} \mathrm{d}h \tag{E.103}$$

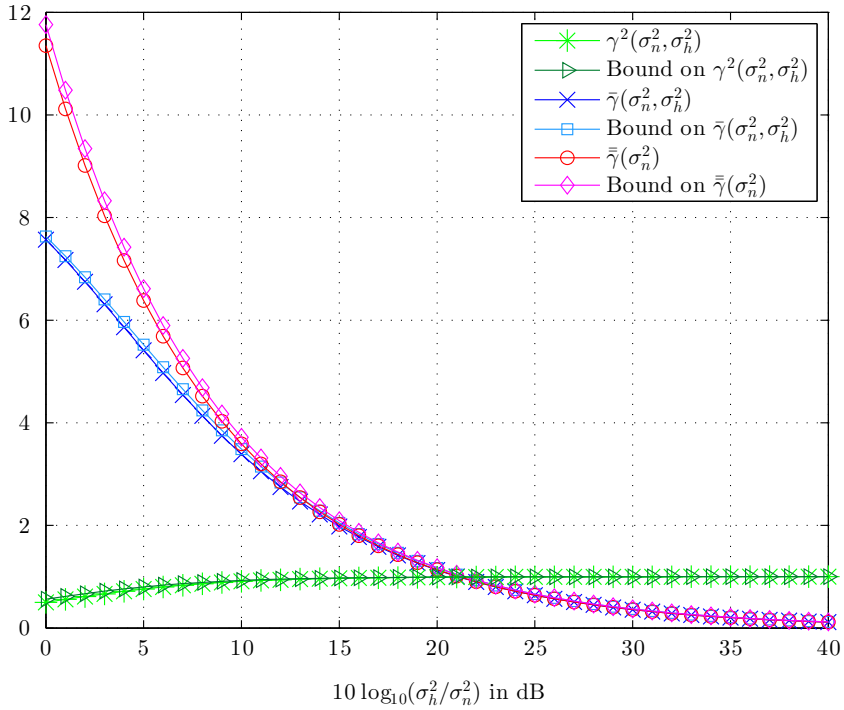


Figure E.1.: Comparison of $\gamma^2(\sigma_n^2,\sigma_h^2)$, $\bar{\gamma}(\sigma_n^2,\sigma_h^2)$, and $\bar{\bar{\gamma}}(\sigma_n^2)$ with the upper bounds of (E.96), (E.101), and (E.105) for $\sigma_h^2 = 1$.

$$\leq 4 \int_{-\infty}^{\infty} e^{-(1-2/\pi)h^2/\sigma_n^2} \mathrm{d}h \tag{E.104}$$

$$= 4\sqrt{\frac{\pi\sigma_n^2}{(1-2/\pi)}}. \tag{E.105}$$

Inserting (E.101) and (E.105) into the bound of Theorem 6.5 yields the bound of Theorem 6.6. The bounds in (E.96), (E.101), and (E.105) appear to be fairly tight bounds on $\gamma^2(\sigma_n^2,\sigma_h^2)$, $\bar{\gamma}(\sigma_n^2,\sigma_h^2)$ and $\bar{\bar{\gamma}}(\sigma_n^2)$, respectively, as Figure E.1 illustrates. ■

E.5. Proof of Proposition 6.7

The conditional probability $P_{Z^{i-1}|H}(z^{i-1}|h)$ is given as a product of $(i-1)$ terms involving the Q-function, cf. (E.34). The function $1-\mathrm{Q}(x) = \Phi(x)$, where $\Phi(x)$ is the CDF of a random variable with distribution $\mathcal{N}(0,1)$. The CDF $\Phi(x)$ is log-concave [BV04, Chapter 3.5]. The log-concavity of $\Phi(x)$ also follows from [BB05, Theorem 1], since the probability distribution function (PDF) of a Gaussian random variable is continuously differentiable and log-concave. Moreover, we have $\mathrm{Q}(x) = 1-\mathrm{Q}(-x) = \Phi(-x)$, so that $\mathrm{Q}(x)$ is also log-concave. Consequently, $P_{Z^{i-1}|H}(z^{i-1}|h)$ is a product of $i-1$ log-concave functions;

since log-concavity is preserved by multiplication [BV04, Chapter 3.5], $P_{Z^{n-1}|H}(z^{n-1}|h)$ is log-concave, for any z^{i-1} and d^{i-1}. ■

F

Mathematical Notation and Abbreviations

Mathematical Notation

$\boxplus$	boxplus operator
$F_X(\cdot)$	cumulative distribution function of the random variable X
$\delta(x)$	Kronecker delta function
$H(X)$	entropy of the random variable X
$H(X\|Y)$	entropy of X conditioned on Y
$h(X)$	differential entropy of the random variable X
$h(X\|Y)$	differential entropy of X conditioned on Y
$H_\mathrm{b}(x)$	binary entropy function
$\mathrm{erf}(x)$	error function
$\mathrm{E}[X]$	expectation of the random variable X
$\mathrm{E}[X\|A]$	expectation of X conditioned on event A
$\mathbb{1}_{\{\cdot\}}$	indicator function
$\ln(x)$	natural logarithm of x
$\log_2(x)$	logarithm to base 2
$I(X;Y)$	mutual information between X and Y
$I(X;Y\|Z)$	mutual information between X and Y conditioned on Z
∇	Nabla operator
$p_X(\cdot)$	probability density function of the random variable X
$p_{X\|Y}(\cdot)$	probability density function of X conditioned on Y

$P_X(\cdot)$	probability mass function of the random variable X
$P_{X\|Y}(\cdot)$	probability mass function of X conditioned on Y
$\Pr[A]$	probability of event A
$\Pr[A\|B]$	probability of event A conditioned on event B
$\mathrm{Q}(x)$	Gaussian Q-function
$D_{\mathrm{KL}}(\cdot\|\|\cdot)$	relative entropy or Kullback–Leibler distance
x_i^j	the sequence $\{x_i, x_{i+1}, \ldots, x_j\}$
x^n	short for x_1^n
$\mathrm{sign}(x)$	sign of x
$\mathrm{Var}[X]$	variance of the random variable X
x^+	positive part of x

List of Abbreviations

A/D	analog-to-digital
ADC	analog-to-digital converter
AWGN	additive white Gaussian noise
BAA	Blahut-Arimoto algorithm
BCJR	algorithm by Bahl, Cocke, Jelinek, Raviv
BCRLB	Bayesian Cramér-Rao lower bound
BER	bit error rate
BPSK	binary phase shift keying
BSC	binary symmetric channel
CDF	cumulative distribution function
CF	compress-and-forward
CFER	common frame error rate
CRC	cyclic redundancy check
CRLB	Cramér-Rao lower bound
CSI	channel state information
DMC	discrete memoryless channel
FER	frame error rate
i.i.d.	independent and identically distributed
ISI	intersymbol-interference
KKT	Karush-Kuhn-Tucker
LLR	log-likelihood ratio
LM	Lloyd-Max
MAP	maximum a posteriori
MARC	multiple-access relay channel
MIC	mutual information criterion
MIMO	Multiple-Input/Multiple-Output
MIR	maximum information rate
ML	maximum likelihood

MMSE	minimum mean squared error
MSE	mean squared error
NNC	noisy network coding
OLM	outer linearization method
OSLA	"one-step look-ahead"
PDF	probability distribution function
PSK	phase shift keying
QAM	quadrature amplitude modulation
QPSK	quaternary phase shift keying
SISO	soft-in/soft-out
SNR	signal-to-noise ratio
UF	uniform
UMTS	Universal Mobile Telecommunication System
XOR	exclusive or

Bibliography

[ACH+08] O. Agazzi, D. Crivelli, M. Hueda, H. Carrer, G. Luna, A. Nazemi, C. Grace, B. Kobeissy, C. Abidin, M. Kazemi, M. Kargar, C. Marquez, S. Ramprasad, F. Bollo, V. Posse, S. Wang, G. Asmanis, G. Eaton, N. Swenson, T. Lindsay, and P. Voois, "A 90nm CMOS DSP MLSD transceiver with integrated AFE for electronic dispersion compensation of multi-mode optical fibers at 10 Gb/s," in *Proceedings of the IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, USA, February 2008, pp. 232–609.

[ACLY00] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, April 2000.

[ADT11] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.

[ALV+06] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, August 2006.

[ALYM11] M. H. Azmi, J. Li, J. Yuan, and R. Malaney, "Soft decode-and-forward using LDPC coding in half-duplex relay channels," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, St. Petersburg, Russia, August 2011, pp. 1484–1488.

[Ari72] S. Arimoto, "An algorithm for calculating the capacity of an arbitrary discrete memoryless channel," *IEEE Transactions on Information Theory*, vol. IT-18, no. 1, pp. 14–20, January 1972.

[BAP+06] H.-M. Bae, J. B. Ashbrook, J. Park, N. R. Shanbhag, A. C. Singer, and S. Chopra, "An MLSE receiver for electronic dispersion compensation of OC-192 fiber links," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 11, pp. 2541–2554, November 2006.

[BB05] M. Bagnoli and T. Bergstrom, "Log-concave probability and its applications," *Economic Theory*, vol. 26, no. 2, pp. 445–469, August 2005.

[BCJR74] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. IT-20, no. 2, pp. 284–287, March 1974.

[Ber71] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[Bla72] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. IT-18, no. 4, pp. 460–473, July 1972.

[Bla03] R. E. Blahut, *Algebraic Codes for Data Transmission*. Cambridge: Cambridge University Press, 2003.

[BLM04] J. R. Barry, E. A. Lee, and D. G. Messerschmidt, *Digital Communication*, 3rd ed. Boston: Kluwer Academic Publishers, 2004.

[BSS06] M. S. Bazaara, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming*, 3rd ed. Hoboken: John Wiley & Sons, Inc., 2006.

[BV04] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.

[Cam02] J. B. Campbell, *Introduction to Remote Sensing*, 3rd ed. New York: The Guilford Press, 2002.

[CEG79] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. IT-25, no. 5, pp. 572–584, September 1979.

[Chu55] J. T. Chu, "On bounds for the normal integral," *Biometrika*, vol. 42, no. 1/2, pp. 263–265, June 1955.

[CJS+10] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proceedings of the 16th Annual International Conference on Mobile Computing and Networking (MobiCom)*, Chicago, IL, USA, September 2010.

[CKL06] Y. Chen, S. Kishore, and J. Li, "Wireless diversity through network coding," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, Las Vegas, NV, USA, April 2006, pp. 1681–1686.

[Cra46] H. Cramér, *Mathematical Methods of Statistics*. Princeton University Press, 1946.

[Csi74] I. Csiszár, "On the computation of rate distortion functions," *IEEE Transactions on Information Theory*, vol. IT-20, no. 1, pp. 122–124, January 1974.

[CT84] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statistics and Decisions,* Supplement Issue, vol. 1, pp. 205–237, 1984.

[CT06] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken: John Wiley & Sons, Inc., 2006.

[DFA+10] M. Danieli, S. Forchhammer, J. Andersen, L. Christensen, and S. Christensen, "Maximum mutual information vector quantization of log-likelihood ratios for memory efficient HARQ implementations," in *Proceedings of the Data Compression Conference (DCC)*, Snowbird, UT, USA, March 2010, pp. 30–39.

[DG08] R. Dabora and A. Goldsmith, "On the capacity of indecomposable finite-state channels with feedback," in *Proceedings of the 46th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, USA, September 2008.

[DM10] O. Dabeer and U. Madhow, "Channel estimation with low-precision analog-to-digital conversion," in *Proceedings of the IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010.

[Eur01] *Universal Mobile Telecommunications System (UMTS): Multiplexing and channel coding (FDD)*, European Telecommunications Standards Institute TS 125 212 V4.3.0, 2001.

[Eve65] S. Even, "On information lossless automata of finite order," *IEEE Transactions on Electronic Computers*, vol. EC-14, no. 4, pp. 561–569, August 1965.

[Gal68] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley & Sons, Inc., 1968.

[GG92] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Boston: Kluwer Academic Publishers, 1992.

[GN98] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, October 1998.

[Gra90] R. M. Gray, *Source Coding Theory.* Norwell: Kluwer Academic Publishers, 1990.

[Hau09] C. Hausl, "Joint network-channel coding for the multiple-access relay channel based on turbo codes," *European Transactions on Telecommunications*, vol. 20, no. 2, pp. 175–181, January 2009.

[HOP96] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, March 1996.

[HSOB05] C. Hausl, F. Schreckenbach, I. Oikonomidis, and G. Bauch, "Iterative network and channel decoding on a Tanner graph," in *Proceedings of the 43rd Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, USA, September 2005.

[HT10] H. Holma and A. Toskala, Eds., *WCDMA for UMTS - HSPA Evolution and LTE*, 5th ed. Chichester: John Wiley & Sons, Ltd., 2010.

[HtBD11] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?" in *Proceedings of the 49th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, USA, September 2011.

[Huf54] D. A. Huffman, "Canonical forms for information-lossless finite-state logical machines," *IRE Transactions on Circuit Theory*, vol. CT-6, no. 5, pp. 41–59, May 1954.

[HWS+07] M. Harwood, N. Warke, R. Simpson, T. Leslie, A. Amerasekera, S. Batty, D. Colman, E. Carr, V. Gopinathan, S. Hubbins, P. Hunt, A. Joy, P. Khandelwal, B. Killips, T. Krause, S. Lytollis, A. Pickering, M. Saxton, D. Sebastio, G. Swanson, A. Szczepanek, T. Ward, J. Williams, R. Williams, and T. Willwerth, "A 12.5 Gb/s SerDes in 65nm CMOS using a baud-rate ADC with digital receiver equalization and clock recovery," in *Proceedings of the IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, USA, February 2007, pp. 436–437.

[HZ05] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2020–2040, June 2005.

[JCK+11] M. Jain, J. I. Choi, T. Kim, D. Bharadia, K. Srinivasan, S. Seth, P. Levis, S. Katti, and P. Sinha., "Practical, real-time, full duplex wireless," in *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking (MobiCom)*, Las Vegas, NV, USA, September 2011.

[Kes05] W. Kester, Ed., *The Data Conversion Handbook*. Burlington: Elsevier, 2005.

[KF10] S. Krone and G. Fettweis, "Fading channels with 1-bit output quantization: Optimal modulation, ergodic capacity and outage probability," in *Proceedings of the IEEE Information Theory Workshop (ITW)*, Dublin, Ireland, August 2010.

[KGG05] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, September 2005.

[KJ09] Z. Kohavi and N. K. Jha, *Switching and Finite Automata Theory*, 3rd ed. New York: Cambridge University Press, 2009.

[KL11] T. Koch and A. Lapidoth, "Asymmetric quantizers are better at low SNR," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, St. Petersburg, Russia, August 2011, pp. 2697–2701.

[KP75] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Transactions on Communications*, vol. COM-23, no. 9, pp. 921–934, September 1975.

[KS95] F. R. Kschischang and V. Sorokine, "On the trellis structure of block codes," *IEEE Transactions on Information Theory*, vol. 41, no. 6, pp. 1924–1937, November 1995.

[KvW00] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Sorrento, Italy, June 2000, p. 40.

[LKEGC11] S. H. Lim, Y.-H. Kim, A. A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[Llo82] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. IT-28, no. 2, pp. 192–137, March 1982.

[LRRB05] B. Le, T. W. Rondeau, J. H. Reed, and C. W. Bostian, "Analog-to-digital converters," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 69–77, November 2005.

[LSS10] M. Lu, N. Shanbhag, and A. Singer, "BER-optimal analog-to-digital converters for communication links," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, Paris, France, May 2010, pp. 1029–1032.

[LTW04] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, December 2004.

[LVWD06] Y. Li, B. Vucetic, T. F. Wong, and M. Dohler, "Distributed turbo coding with soft information relaying in multihop relay networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2040–2050, November 2006.

[Mar10] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[Mur08] B. Murmann, "A/D converter trends: Power dissipation, scaling and digitally assisted architectures," in *Proceedings of the IEEE Custom Integrated Circuits Conference (CICC)*, San Jose, CA, USA, September 2008, pp. 105–112.

[MY04] I. Maric and R. D. Yates, "Bandwidth and power allocation for cooperative strategies in Gaussian relay networks," in *Proceedings of the 38th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, November 2004, pp. 1907– 1911.

[Pól49] G. Pólya, "Remarks on computing the probability integral in one and two dimensions," in *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA, August 1949, pp. 63–78.

[Poo94] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer, 1994.

[Pro00] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.

[PWO01] H. P. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 978–1002, March 2001.

[Rao45] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945.

[Rav09] W. Rave, "Quantization of log-likelihood ratios to maximize mutual information," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 283–286, April 2009.

[RG06a] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks - part I: Gaussian case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, March 2006.

[RG06b] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks - part II: unknown probability density function," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2784–2796, July 2006.

[Ris76] J. J. Rissanen, "Generalized Kraft inequality and arithmetic coding," *IBM Journal of Research and Development*, vol. 20, no. 3, pp. 198–203, May 1976.

[Say00] K. Sayood, *Introduction to Data Compression*, 2nd ed. San Francisco: Morgan Kaufmann Publishers, Inc., 2000.

[SBC10] A. Singer, A. Bean, and J. W. Choi, "Mutual information and time-interleaved analog-to-digital conversion," in *Proceedings of the Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, February 2010.

[SDM09] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Transactions on Communications*, vol. 57, no. 12, pp. 3629–3639, December 2009.

[SEA03a] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, part I: System description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, November 2003.

[SEA03b] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, part II: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, November 2003.

[Sha48] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[Sha59] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE National Convention Record, Part 4*, New York, NY, USA, 1959, pp. 142–163.

[SK90] S. Shamai (Shitz) and Y. Kofman, "On the capacity of binary and Gaussian channels with run-length limited inputs," *IEEE Transactions on Communications*, vol. 38, no. 5, pp. 584–594, May 1990.

[SKM04a] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Capacity theorems for the multiple-access relay channel," in *Proceedings of the 42nd Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, USA, September 2004.

[SKM04b] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Hierarchical sensor networks: Capacity bounds and cooperative strategies using the multiple-access relay channel model," in *Proceedings of the First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON)*, Santa Clara, CA, USA, October 2004, pp. 191–199.

[SKM04c] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Hierarchical wireless networks: Capacity bounds using the constrained multiple-access relay channel model," in *Proceedings of the 38th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, November 2004, pp. 1912–1916.

[SL96] S. Shamai (Shitz) and R. Laroia, "The intersymbol interference channel: Lower bounds on capacity and channel precoding loss," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1388–1404, September 1996.

[SMZ07] K. Sohraby, D. Minoli, and T. Znati, *Wireless Sensor Networks*. Hoboken: John Wiley & Sons, Inc., 2007.

[SOW91] S. Shamai (Shitz), L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Transactions on Information Theory*, vol. 37, no. 6, pp. 1527–1539, November 1991.

[SV05] H. H. Sneessens and L. Vandendorpe, "Soft decode and forward improves cooperative communications," in *Proceedings of the 6th IEEE International Conference on 3G and Beyond*, London, UK, November 2005, pp. 73–76.

[SY08] S. Serbetli and A. Yener, "Relay assisted F/TDMA ad hoc networks: Node classification, power allocation and relaying strategies," *IEEE Transactions on Communications*, vol. 56, no. 6, pp. 937–947, June 2008.

[TPB99] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proceedings of the 37th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, USA, September 1999, pp. 368–377.

[vdM77] E. C. van der Meulen, "A survey of multiway channels in information theory: 1961-1976," *IEEE Transactions on Information Theory*, vol. IT-23, no. 1, pp. 1–37, January 1977.

[VJ06] C. Vogel and H. Johansson, "Time-interleaved analog-to-digital converters: Status and future directions," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, Island of Kos, Greece, May 2006, pp. 3386–3389.

[VKAL08] P. O. Vontobel, A. Kavčić, D. M. Arnold, and H.-A. Loeliger, "A generalization of the Blahut-Arimoto algorithm to finite-state channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.

[VO79] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, Inc., 1979.

[VT68] H. L. Van Trees, *Detection, Estimation and Modulation Theory - Part I.* New York: John Wiley & Sons, Inc., 1968.

[Wal99] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, April 1999.

[Wil46] J. D. Williams, "An approximation to the probability integral," *The Annals of Mathematical Statistics*, vol. 17, no. 3, pp. 363–365, September 1946.

[Wit80] H. S. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Transactions on Information Theory*, vol. IT-26, no. 3, pp. 265–271, May 1980.

[WKN+11] P. A. Whiting, G. Kramer, C. J. Nuzman, A. Ashikhmin, A. J. van Wijngaarden, and M. Živković, "Analysis of inverse crosstalk channel estimation using SNR feedback," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1102–1115, March 2011.

[WW75] H. S. Witsenhausen and A. D. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Transactions on Information Theory*, vol. IT-21, no. 5, pp. 493–501, September 1975.

[Wyn78] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder II: General sources," *Information and Control*, vol. 38, no. 3, pp. 60–80, July 1978.

[WZ76] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 1–10, January 1976.

[Yeu02] R. W. Yeung, *A First Course in Information Theory.* New York: Kluwer Academic/Plenum Publishers, 2002.

[YHXM09] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMAX and LTE-advanced mobile systems," *IEEE Communications Magazine*, vol. 47, no. 10, pp. 100–105, October 2009.

[YK07] S. Yang and R. Koetter, "Network coding over a noisy relay: A belief propagation approach," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007, pp. 801–804.

[YS06] H. Y. Yangand and R. Sarpeshkar, "A bio-inspired ultra-energy-efficient analog-to-digital converter for biomedical applications," *IEEE Transactions on Circuits and Systems I*, vol. 53, no. 11, pp. 2349–2356, November 2006.