Ulm University

Institute of Telecommunications and Applied Information Theory

Bidirectional Wireless Communication via Relay Nodes using Lattices

Diploma Thesis

Christian Häger, *July 2011*

Abstract

The two-way relay channel (TRC) is a simple communication network in which two users wish to exchange independent messages with the help of a relay. This network has attracted a lot of research interest in recent years. In particular the separated two-way relay channel (sTRC), where no direct communication link exists between the users, has become a widely used example to demonstrate the potential benefits of network coding (NC) and physical-layer network coding (PNC) over conventional routing approaches. Moreover, information-theoretic results about achievable rates can provide insight and helpful benchmarks for the design of future communication networks.

Motivated by wireless applications, in this thesis some important relaying strategies for the Gaussian case of the sTRC are reviewed and compared. First, the topic is introduced with an example in order to illustrate important principles that arise in cooperative wireless networks. Then three relaying strategies, namely amplify-and-forward (AF), decode-and-forward (DF), and modulo-and-forward (MF), are described and compared for particular channel conditions. It is also shown that by using nested lattice codes in combination with superposition coding it is possible to characterize the capacity region of this network to within a constant gap for arbitrary channel conditions. After that, the network is modified by adding a second relay. We call the resulting setup the separated two-way two-relay channel (sTTRC) and show that the cut-set region is achievable for a (binary symmetric) finite field physical layer using nested linear codes. For the Gaussian case an achievable rate region is derived assuming certain channel conditions. The employed coding strategy is based on a lattice partition chain and the derived region is within 1/2 bit per dimension of the capacity region for each user.

Author:	Christian Häger
Course of Studies:	Electrical Engineering
Matrikelnummer:	589903
Examiner:	Dr. Aydin Sezgin
Catalog No.:	D/2011/AS/01

I, Christian Häger, declare that this thesis is my original work. I have never received any unauthorized help and citations are marked as such. I understand that, if at any time it is shown that I have significantly misrepresented material, any degree or credits awarded to me by the University of Ulm may be revoked.

Ulm, 14. July 2011

(Christian Häger)

Contents

List of Figures

CHAPTER 1

Introduction

The theme of this thesis is bidirectional communication of two wireless nodes, when so-called relay nodes assist in the message exchange. For the case with only one relay the setup is commonly referred to as the two-way relay channel (TRC) and this chapter provides a brief overview over important principles that arise in such cooperative wireless communication networks as opposed to conventional point-to-point communication.

The network setup itself may occur often in practical situations. As an example in Figure 1.1 two wireless users are depicted who wish to communicate (e.g. chat, file exchange, etc.) via a common relay station. Usually wireless transceivers are half-duplex (HD) constrained [1], which means that a device node can either transmit or receive but not both at the same time. Then, if the direct link between the two users is sufficiently

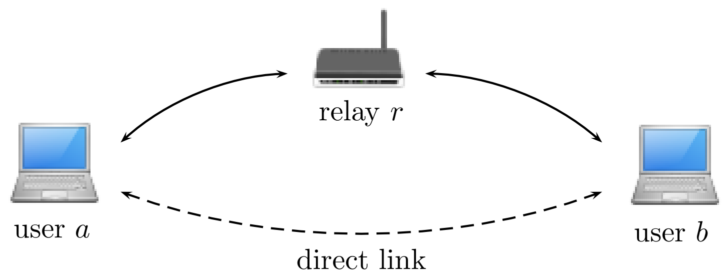


Figure 1.1.: A common practical example for bidirectional wireless communication via a relay r. The users a and b want to communicate (exchange independent messages).

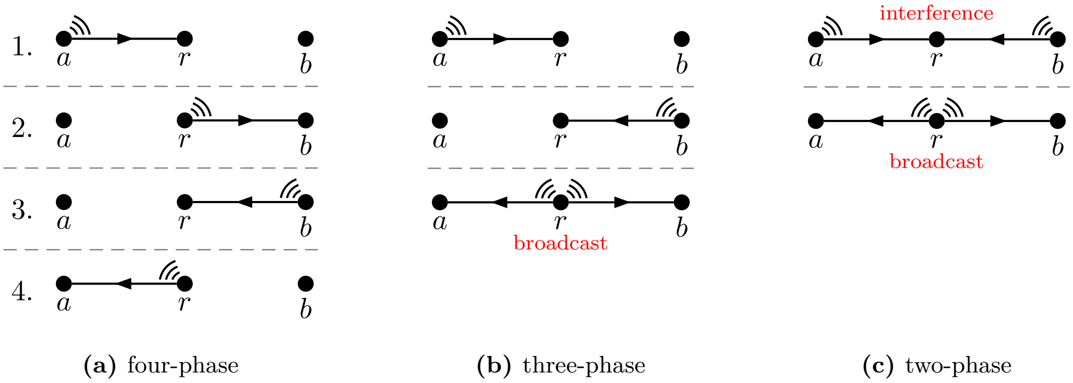


Figure 1.2.: Possible transmission protocols for the sTRC with HD constrained nodes.

strong, a very simple (and suboptimal) strategy would be that the users just take turns transmitting their data while ignoring the relay. However, in this thesis the direct link is assumed to be very weak and neglectable which may occur when the users are very far apart. In that case another, quite different, situation arises. Now the relay becomes the *enabler of communication* which means that no positive data rate can be established without it and any information flow has to pass through the relay. This setup is called the separated two-way relay channel (sTRC) [2] and substantial progress has been made in recent years in order to find an optimal transmission strategy for this network and to characterize the capacity region [2–7].

The usual paradigm in wireless networking is to avoid interference as much as possible, establish reliable point-to-point links between nodes at the physical layer and route the data packets through those links. With this approach a hypothetical exchange of two bits, say x_1 (direction $a \to b$) and x_2 (direction $b \to a$), would then demand for four transmission phases, which is depicted in Figure 1.2 (a). First user a transmits the bit x_1 to the relay r and the relay forwards the bit to b. Then similarly user b transmits the bit x_2 to r and r forwards the bit to a. However, this approach must not be optimal from an information-theoretic viewpoint. In fact, there has been a paradigm shift in networking, and in particular wireless networking, with the introduction of network coding (NC) by Ahlswede *et al.* [8]. The key insight is that messages which are intended for different nodes (or "flow" in opposite directions as in the case of bidirectional communication) should not be treated separately but can be mixed or combined at intermediate relay nodes[1]. Applying NC to this example may actually save one transmission phase [10]. The

[1] A more descriptive way to put this is that information flow through a network should not be treated as the flow of water through pipes [9].

idea is to first let both users transmit their respective message bit to the relay. The relay computes the XOR ($x_1 \oplus x_2$) of those bits, i.e. the parity-check bit, and then transmits this bit. Without violating the HD constraint, both users may receive the parity-check bit because the wireless channel allows for *broadcasting*, such that more than one node can overhear a particular transmission. Then, after successful reception, each user is able to extract the message of the other user by employing side information (SI), because the own message bit is known. User a computes $(x_1 \oplus x_2) \oplus x_1 = x_2$ while user b computes $(x_1 \oplus x_2) \oplus x_2 = x_1$, thereby recovering the message of the respective other user. This approach uses a three-phase transmission protocol and is depicted in Figure 1.2 (b).

In wireless communication, interference is regarded as harmful because an interfering signal acts as additive noise and therefore reduces the achievable data rates. An important characteristic of the four-phase and three-phase transmission protocol described above is that interference between transmitting signals is avoided entirely. For the sTRC, however, interference can only happen at the relay (because the two users are assumed to be separated) but the relay is not the intended destination for neither of the messages. Therefore there is no necessity for the relay to actually decode the messages at all[2]. Surprisingly then, interference can be harnessed in the form of physical-layer network coding (PNC) [11, 12] which may substantially increase the overall throughput for this network. An illustration of a two-phase protocol that involves interference in the first phase is depicted in Figure 1.2 (c). Here both users transmit their respective signals in the same time slot and consequently these signals interfere at the receiving antenna of the relay. The key idea is to exploit the fact that wireless signals interfere in an additive way and to use this additive nature of the wireless channel "for an equivalent network coding operation" [11]. This is the main motivation to study lattices in the context of cooperative wireless networks because they naturally fit this idea of PNC [5, 6, 12–15]. In order to illustrate the basic principle, assume now that conventional binary phase-shift keying (BPSK) is employed and both users map their message bits to signals according to $0 \to -1$ and $1 \to +1$. BPSK modulation can be regarded as a mapping of bits to points in a (translated) lattice, more precisely $\{0, 1\} \to 2\mathbb{Z} - 1$, where $\mathbb{Z}$ is the set of all integers or equivalently the one-dimensional integer lattice. The users then transmit the modulated signals in the same time slot. For an ideal, noiseless channel with perfect synchronization the relay receives either ± 2, when the bits of the users are identical, or 0, when the bits of the users are different. The relay can then apply the mapping $\pm 2 \to 0$ and $0 \to 1$ and one can check that the received signal at the relay indeed corresponds to a "physical-layer equivalent" of

[2]Note that in both the routing and the network coding approach described above, the relay is fully aware of both message bits.

the XOR of the two message bits. Of course, when the relay receives 0, it is impossible to extract the individual message of the users from the superimposed signal. However, this is not necessary. The relay can proceed as in the NC case (i.e. broadcast the parity-check bit) and rely on the fact that the two users have sufficient SI in order recover the original messages.
In summary, the following principles may arise in cooperative wireless communication networks such as the TRC and will be used or referred to extensively throughout the rest of this thesis:

Network Coding. Decoded messages can be combined locally at intermediate relay nodes [8]. In this thesis NC will simply mean computing the componentwise XOR of two binary message vectors.

Broadcasting. In a wireless network it is possible that more than one node can overhear a particular signal by a transmitting node.

Side Information. In a communication network nodes have knowledge of past received or transmitted signals and/or messages.

Interference. When two or more wireless nodes transmit a signal in the same time slot, these signals get linearly superimposed at a receiving node.

Physical-Layer Network Coding. The additive nature of simultaneously arriving wireless signals can be used for an equivalent network coding operation [11].

The research on this topic is ongoing and new material is published frequently. Therefore, the author would like to point out a very recent survey-style paper on this subject [16], which provides an excellent overview over different techniques and ideas that have been developed so far as well as many references.

1.1. Thesis Outline

In chapter 2 the notation is introduced and the so-called cut-set bound is applied to the Gaussian case of the sTRC yielding an upper bound on the achievable rate pairs. After that, two relaying strategies which to not rely on lattices, namely amplify-and-forward (AF) and decode-and-forward (DF), are briefly reviewed and serve as a comparison reference for the strategies described in chapter 3. Chapter 3 starts with a general introduction to lattices and the most important figure of merits regarding their usage in cooperative communication networks. Two relaying approaches based on lattices are then described in detail. First the modulo-and-forward (MF) strategy is considered which is proposed in [15]. In [15] a restrictive channel model is used, in the sense that the two communication links from the relay to the users are assumed to have equal signal-to-noise ratio (SNR). In this thesis the MF scheme is generalized by applying it to the Gaussian sTRC with arbitrary channel conditions and the achievable rates are provided. Then, it is shown that by using nested lattice codes in combination with rate splitting and superposition coding it is possible to characterize the capacity region of this network to within a gap of 1/2 bit per dimension for each user. Moreover, it is shown that the gap vanishes for one user with increasing uplink SNR. In chapter 4 we compare the described approaches under various channel conditions in terms of achievable sum rates. Finally, in chapter 5 a second relay is added to the network. It is assumed that that only two adjacent nodes can communicate with each other and we call this network the (fully) separated two-way two-relay channel (sTTRC). We derive the capacity region assuming a (binary symmetric) finite field physical layer and an achievable rate region for the Gaussian case under certain channel conditions.

CHAPTER 2

Preliminaries for the Separated Two-Way Relay Channel

In this chapter the cut-set bound [9] is used to obtain an upper bound on the achievable rate pairs for the Gaussian sTRC with full-duplex (FD) nodes and for the Gaussian TRC with HD nodes under the assumption of a two-phase transmission protocol. After that, two important relaying strategies which do not rely on lattices are briefly reviewed. While the focus of this thesis is on lattice-based strategies, a wide variety of different relaying strategies have been proposed for this network model (see e.g. [3, 17, 18]). The choice of these two particular strategies, namely amplify-and-forward (AF) and decode-and-forward (DF), is motivated by the fact that they constitute possible extremes regarding the knowledge about the messages of the two users at the relay node. Roughly, in the AF scheme the relay merely acts as a repeater, i.e. it rescales and forwards the signal it receives. In particular the relay does not attempt to decode the messages and therefore is unaware of them. In contrast to this, in the DF scheme the relay tries to fully decode both messages. This may or may not be useful, but there is however no necessity for the relay to actually decode both messages individually. Both strategies are not optimal in general for this communication network. In fact, the particular shortcoming of the DF scheme, which will be explained below, directly motivates the use of lattice codes in chapter 3. The upper bound as well as the achievable rates for the two strategies described in this chapter will then serve as a comparison reference for the lattice-based strategies in chapter 4.

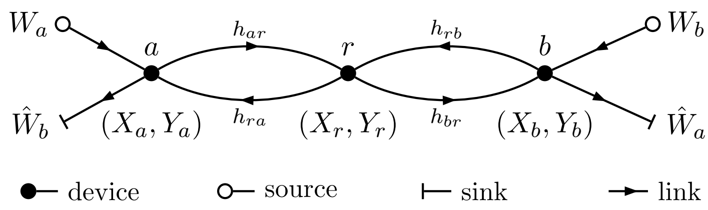


Figure 2.1.: The network model of the sTRC where no direct communication link is present between the users.

2.1. Network Model and Notation

In Figure 2.1 the network model of the sTRC is depicted[1]. It is assumed that all device nodes have FD capability and the case where the nodes are HD constrained is described further below. The two users a and b want to exchange independent messages, denoted by W_a (direction $a \to b$) and W_b (direction $b \to a$), at rates R_a and R_b (in bits per channel use) respectively. The messages are chosen uniformly and independently from the message sets $\mathcal{W}_a = \{1, 2, \cdots, \lfloor 2^{nR_a} \rfloor\}$ and $\mathcal{W}_b = \{1, 2, \cdots, \lfloor 2^{nR_b} \rfloor\}$, where n denotes the number of channel uses. The relay r has no message to transmit and is not the intended destination for neither of the users messages, i.e. no source or sink nodes are attached to the relay device node in the network model. For simplicity it is assumed that nR_a and nR_b are integers and therefore the messages can also be uniquely written as binary vectors denoted by $\boldsymbol{W}_a \in \{0,1\}^{nR_a}$ and $\boldsymbol{W}_b \in \{0,1\}^{nR_b}$. Each node is associated with an input and an output variable denoted by (X_a, Y_a), (X_r, Y_r) and (X_b, Y_b) respectively. For the Gaussian case of the sTRC these variables take on values in $\mathbb{R}$ and the channel is modeled by the following three equations:

$$Y_a = h_{ra} X_r + Z_a \tag{2.1a}$$

$$Y_r = h_{ar} X_a + h_{br} X_b + Z_r \tag{2.1b}$$

$$Y_b = h_{rb} X_r + Z_b, \tag{2.1c}$$

where Z_a, Z_r and Z_b are i.i.d. Gaussian random variables with zero mean and unit variance and the channel gains $h_{ar}, h_{br}, h_{ra}, h_{rb}$ are positive real numbers representing the signal attenuation between nodes. It is assumed that all channel gains can be different (no reciprocity) and are fixed for the whole message exchange (flat fading). Each node has full cannnel state information (CSI) and is therefore aware of all channel gains. Note that the equations (2.1) imply the absence of a direct communication link between the two users

[1]This graphical notation for cooperative communication networks is proposed in [1].

and in particular equation (2.1b) accounts for the interference that can happen at the relay and (2.1a) and (2.1c) reflect the broadcast nature of this network. The channel is assumed to be memoryless and for successive channel uses a vector notation will be used. Vectors with length n are be denoted by $\boldsymbol{X}_a = (X_{a,1}, \ldots, X_{a,n})$, where $X_{a,i}$ refers to the transmitted signal of user a at time i and $\boldsymbol{X}_a^i = (X_{a,1}, X_{a,2}, \ldots, X_{a,i})$ denotes all channel inputs for user a up to time i (and similarly for all other input and output variables). All transmitted signals are subject to an average power constraint according to

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[X_{j,i}^2\right] \leq P \qquad \forall j \in \{a, r, b\}, \tag{2.2}$$

where $\mathrm{E}\left[\,\cdot\,\right]$ denotes expectancy. For a particular node the transmitted signal at time i can be a function of the message of that node and also of the past received channel outputs:

$$X_{a,i} = f(W_a, \boldsymbol{Y}_a^{i-1}) \tag{2.3a}$$

$$X_{r,i} = f(\boldsymbol{Y}_r^{i-1}) \tag{2.3b}$$

$$X_{b,i} = f(W_b, \boldsymbol{Y}_b^{i-1}). \tag{2.3c}$$

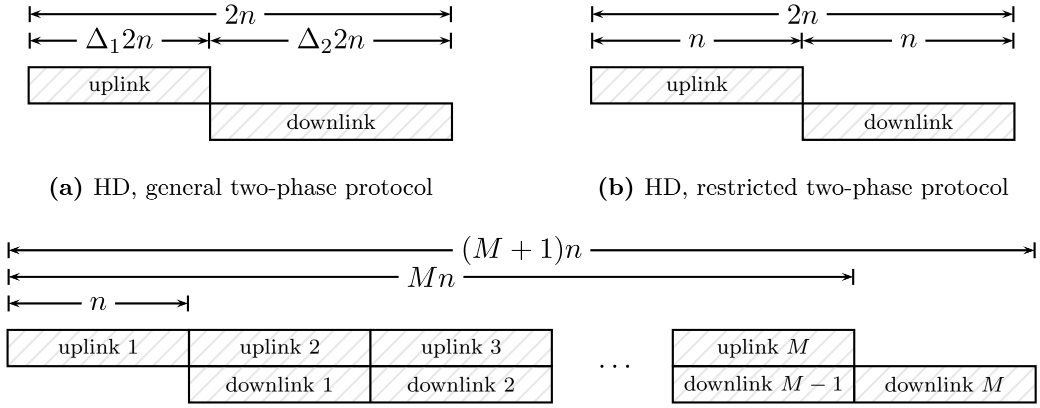
Recall that the relay has no message of its own and therefore $X_{r,i}$ is a function of the past received channel outputs only. The message estimates at the two users are computed based on all received channel outputs as well as the own transmitted message, i.e. $\hat{W}_a = g(\boldsymbol{Y}_b, W_b)$ and $\hat{W}_b = g(\boldsymbol{Y}_a, W_a)$. An error occurs if the transmitted message of one user is not equal to the corresponding estimate of the other user and the error probability is given by

$$p_e = \Pr(\{\hat{W}_a \neq W_a\} \cup \{\hat{W}_b \neq W_b\}). \tag{2.4}$$

A rate pair (R_a, R_b) is said to be achievable if encoding and decoding functions exists such that $p_e \to 0$ for $n \to \infty$. The capacity region $\mathcal{R}_c$ is the closure of all rate pairs that are achievable.

Half-duplex nodes

Practical wireless transceivers are commonly HD constrained [1] and cannot transmit and receive at the same time. Accounting for this constraint is usually done by assuming a particular transmission protocol (or equivalently a sequence of network states [19]) which predefines the (non-adaptive) temporal sequence of transmission and reception phases for each node. For the sTRC with HD device nodes a general two-phase protocol is depicted in Figure 2.2 (a). Here the channel is used $2n$ times in total and the channel uses are divided into an uplink and downlink phase of relative durations Δ_1 and Δ_2 respectively, where



(a) HD, general two-phase protocol

(b) HD, restricted two-phase protocol

(c) FD, transmission of M blocks of dimension n

Figure 2.2.: Relationship between the protocol strategies for HD and FD nodes. In this thesis the restricted two-phase protocol (b) is always used to describe the relaying strategies and generalization to FD nodes is done with the block transmission strategy (c).

$0 \leq \Delta_1 \leq 1$, $0 \leq \Delta_2 \leq 1$, and $\Delta_1 + \Delta_2 = 1$. In the uplink, both users a and b transmit signals and the relay receives. Consequently the relay remains silent and the users don't receive any information. In the downlink the relay transmits while both users receive. For HD nodes the power constraint (2.2) is assumed to hold for each phase separately. In this thesis a restricted two-phase protocol is considered where $\Delta_1 = \Delta_2 = 1/2$, i.e. the relative durations of the uplink and downlink are equal[2]. The following lemma shows, that for this particular transmission protocol the relationship between achievable rates for HD and FD nodes is simple.

Lemma 1. *If a particular rate pair $(R_a/2, R_b/2)$ is shown to be achievable for the Gaussian sTRC with HD nodes under the assumption of the restricted two-phase protocol, then the rate pair (R_a, R_b) is achievable for the Gaussian sTRC with FD nodes.*

Proof. The strategy showing the achievability for HD nodes is used M times as shown in Figure 2.2 (c). Then MnR_a bits are transmitted from user a to b in $(M+1)n$ channel uses and thus the rate approaches R_a for large M (and similarly for user b). Achievability follows from the fact that the error probability for each uplink/downlink pair can be made arbitrarily small for $n \to \infty$. □

Because of this lemma, the following convention will be used when describing the relaying

[2]For the general two-phase protocol it would be possible to find optimal Δ_1 and Δ_2 with respect to, say, the overall throughput for the network based on a particular relaying strategy, see [20].

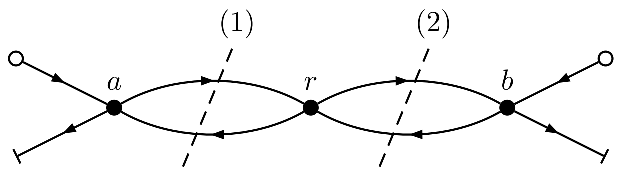


Figure 2.3.: Illustration of the two possible cuts (1) and (2) for the sTRC.

strategies in the following sections. All relaying strategies are described with respect to HD nodes assuming the restricted two-phase protocol (with other words one uplink/downlink phase is described and depicted in block diagrams). The achievable rates are then stated for FD nodes, i.e. multiplied by a factor of two, mainly in order to avoid the prelog factors that would occur otherwise.

2.2. Cut-set Bound

The cut-set bound [9] provides an upper bound on the achievable rate tuples for general multi-terminal networks. In this section this bound is used to obtain an upper bound on the achievable rate pairs for the Gaussian sTRC with FD nodes. The network topology illustrating the possible cuts is depicted in Figure 2.3. A cut divides all device nodes in a network into two disjoint subsets. The flow of information from one subset to another is limited by the mutual information between the inputs of one subset and the outputs of the other subset, conditioned by the outputs of the other subset [9]. Applying the cut-set bound to the Gaussian sTRC yields the following cut-set region[3]:

$$\mathcal{R}_{cut} = \left\{ (R_a, R_b) \ : \ \begin{array}{l} 0 < R_a < \min\left(\mathrm{C}\left(h_{ar}^2 P\right), \mathrm{C}\left(h_{rb}^2 P\right)\right) \\ 0 < R_b < \min\left(\mathrm{C}\left(h_{br}^2 P\right), \mathrm{C}\left(h_{ra}^2 P\right)\right) \end{array} \right\}, \tag{2.5}$$

where $\mathrm{C}(x) \stackrel{\text{def.}}{=} \log_2(1+x)/2$. The cut-set region upper bounds the capacity region, in the sense that $\mathcal{R}_c \subseteq \mathcal{R}_{cut}$. The cut-set bound in [9] is derived for FD nodes and consequently (2.5) also only applies to FD nodes. An upper bound for the (not necessarily separated) Gaussian TRC with HD nodes under the assumption of the general two-phase protocol as described in the previous section is derived in [22] and given by:

$$\mathcal{R}_{cut}^{\mathrm{HD}} = \left\{ (R_a, R_b) \ : \ \begin{array}{l} 0 < R_a < \min\left(\Delta_1 \mathrm{C}\left(h_{ar}^2 P\right), \Delta_2 \mathrm{C}\left(h_{rb}^2 P\right)\right) \\ 0 < R_b < \min\left(\Delta_1 \mathrm{C}\left(h_{br}^2 P\right), \Delta_2 \mathrm{C}\left(h_{ra}^2 P\right)\right) \end{array} \right\}. \tag{2.6}$$

[3] This result first appeared in [14] and a complete proof for both discrete memoryless channels and the Gaussian case is given in [21].

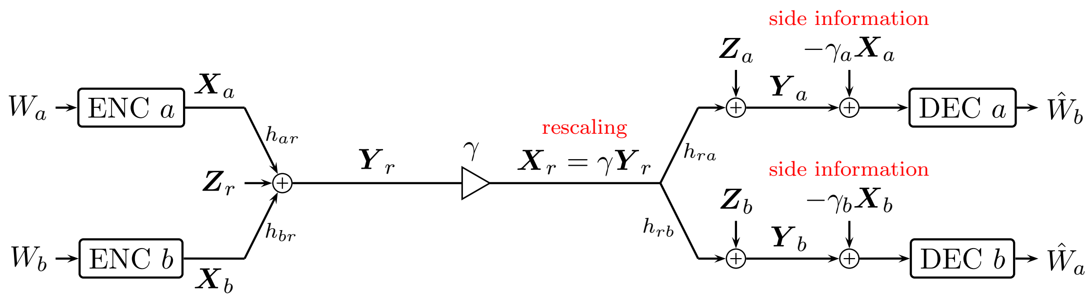


Figure 2.4.: Block diagram for the amplify-and-forward (AF) strategy. See the text for an explanation of the scaling factors γ, γ_a and γ_b.

Note that the assumption of separation (absence of a direct link) is not necessary here in order to state the upper bound because the nature of the two-phase protocol in combination with the HD assumption renders any existing direct link unexploitable. Also note that for HD nodes the assumed transmission protocol is key in order to state an upper bound and different protocols lead to different upper bounds (see [18]). It can be seen that for the restricted two-phase protocol, i.e. for $\Delta_1 = \Delta_2 = 1/2$, the cut-set region for the Gaussian sTRC with FD nodes differs from the cut-set region for the Gaussian TRC with HD nodes only by factor of 2.

2.3. Amplify-and-forward

One of the simplest relaying strategies for the Gaussian case of the sTRC is called amplify-and-forward (AF) and is described in [3], [4] and [23] for HD nodes. In the following the derivation of the achievable rates for AF is briefly reviewed. See Figure 2.4 for a block diagram of this scheme. In the uplink both users simultaneously transmit their respective signals $\boldsymbol{X}_a$ and $\boldsymbol{X}_b$ which are assumed to be independent to have average power P. The basic idea is that the relay broadcasts a *rescaled* version of the received signal $\boldsymbol{X}_r = \gamma \boldsymbol{Y}_r$, where the scaling factor γ is chosen such that the relay power constraint is met, i.e.

$$\frac{1}{n}\mathrm{E}\left[||\boldsymbol{X}_r||^2\right] = \frac{1}{n}\mathrm{E}\left[||\gamma \boldsymbol{Y}_r||^2\right] = \gamma^2(h_{ar}^2 P + h_{br}^2 P + 1) \stackrel{!}{=} P \tag{2.7}$$

$$\Rightarrow \gamma^2 = \frac{P}{h_{ar}^2 P + h_{br}^2 P + 1}. \tag{2.8}$$

Upon reception both users can subtract a scaled version of their own transmit signal from the received signal prior to decoding, i.e. they employ “analog” SI [24]. For example user

a receives

$$\begin{aligned}\boldsymbol{Y}_a &= h_{ra}\boldsymbol{X}_r + \boldsymbol{Z}_a = h_{ra}\gamma\boldsymbol{Y}_r + \boldsymbol{Z}_a \\ &= h_{ra}\gamma(h_{ar}\boldsymbol{X}_a + h_{br}\boldsymbol{X}_b + \boldsymbol{Z}_r) + \boldsymbol{Z}_a \end{aligned} \tag{2.9}$$

and then subtracts $\gamma_a\boldsymbol{X}_a = h_{ra}\gamma h_{ar}\boldsymbol{X}_a$ in order to obtain

$$\tilde{\boldsymbol{Y}}_a = \boldsymbol{Y}_a \underbrace{-\gamma_a\boldsymbol{X}_a}_{\substack{\text{analog side}\\ \text{information}}} = \gamma h_{br}h_{ra}\boldsymbol{X}_b + \gamma h_{ra}\boldsymbol{Z}_r + \boldsymbol{Z}_a. \tag{2.10}$$

Obviously the assumption of full CSI for all nodes is necessary in order to calculate the scaling factors correctly. Equation (2.10) shows that user a “sees” an additive Gaussian noise channel for which an equivalent signal-to-noise ratio SNR_a can be computed given by

$$\mathrm{SNR}_a = \frac{\gamma^2 h_{br}^2 h_{ra}^2 P}{\gamma^2 h_{ra}^2 + 1} = \frac{h_{br}^2 h_{ra}^2 P^2}{h_{ra}^2 P + h_{ar}^2 P + h_{br}^2 P + 1}. \tag{2.11}$$

The same steps can be repeated for user b to obtain a similar expression for SNR_b. Effectively, this scheme converts the sTRC into a Gaussian two-way channel [4] (see also [9, p. 519]).

Therefore, the achievable rates for AF are given by all positive rate pairs satisfying

$$\begin{aligned} R_a &< \mathrm{C}\left(\mathrm{SNR}_b\right) = \mathrm{C}\left(\frac{h_{ar}^2 h_{rb}^2 P^2}{h_{rb}^2 P + h_{ar}^2 P + h_{br}^2 P + 1}\right) \\ R_b &< \mathrm{C}\left(\mathrm{SNR}_b\right) = \mathrm{C}\left(\frac{h_{br}^2 h_{ra}^2 P^2}{h_{ra}^2 P + h_{ar}^2 P + h_{br}^2 P + 1}\right). \end{aligned} \tag{2.12}$$

Note that this strategy can be applied to FD nodes by assuming that forwarding is done on a *per-symbol* basis (rather than on a *per-block* basis), i.e. the relay immediately transmits the rescaled version of each received symbol $Y_{r,i}$. The relay then transmits nothing in the first channel use and receives and transmits simultaneously in the following $n-1$ uses according to the previous description. In the last channel use only the relay transmits because the users are finished with transmitting their own signals which have length n. Therefore the whole message exchange takes $n+1$ channel uses and the effective rate is reduced by a factor of $n/(n+1)$ which can be neglected for $n \to \infty$.

2.4. Decode-and-forward

A decode-and-forward (DF) strategy for the sTRC is proposed in [3] for the Gaussian case and in [4] for the Gaussian case as well as for general discrete memoryless channel models.

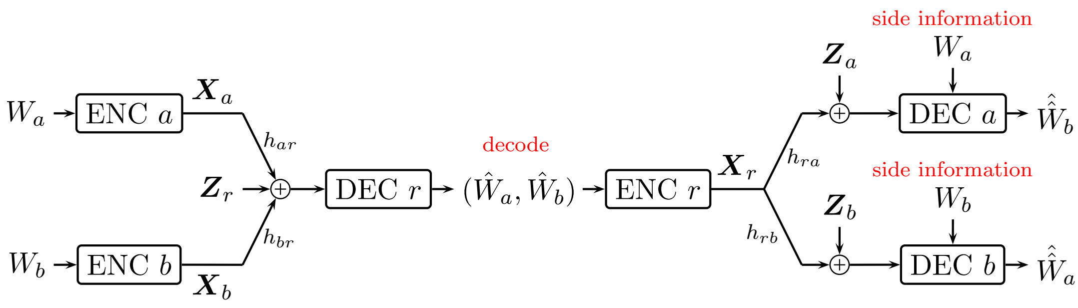


Figure 2.5.: Block diagram for the decode-and-forward (DF) strategy. Here the relay tries to completely decode both messages in the uplink phase and re-encodes them for the downlink.

The key idea for this strategy is that the relay attempts to fully decode both messages in the first phase and then re-encodes them for transmission in the second phase. A block diagram for this strategy is depicted in Figure 2.5.
Interestingly the DF strategy decomposes the sTRC setup into two separate problems. In the uplink the setup is identical to the classical Gaussian multi-access channel. The capacity region for the uplink is therefore given by the set of rate pairs satisfying [9]

$$\begin{aligned} R_a &< \mathrm{C}\left(h_{ar}^2 P\right) \\ R_b &< \mathrm{C}\left(h_{br}^2 P\right) \\ R_a + R_b &< \mathrm{C}\left(h_{ar}^2 P + h_{br}^2 P\right). \end{aligned} \tag{2.13}$$

It is tempting to regard the second phase as the classical Gaussian broadcast channel, but this would not account for the fact that each user knows its own message and therefore has SI to exploit. It is not obvious however to see what the best use of this SI is. In [3] "analog" SI is considered, in the sense that superposition coding is used at the relay. The relay re-encodes the decoded messages by using the same codebooks as the users but with adjustable power allocation. Then, each user can subtract its own codeword prior to decoding, provided that the power allocation is known. Another approach is to use NC. Here one can achieve all rate pairs satisfying

$$R_a, R_b < \min\left(\mathrm{C}\left(h_{ra}^2 P\right), \mathrm{C}\left(h_{rb}^2 P\right)\right). \tag{2.14}$$

This may be achieved by zero-padding the shorter message vector such that both message vectors have equal length and computing $\boldsymbol{W} = \boldsymbol{W}_a \oplus \boldsymbol{W}_b$ at the relay where $\oplus$ denotes the componentwise XOR of two binary vectors. Then, a random Gaussian codebook is used for transmitting the "common message" $\boldsymbol{W}$ and decoding at both users will be successful with large probability if the restriction (2.14) is fulfilled, i.e. the cardinalities of the message sets are adjusted to the *worse* SNR in the downlink. After successful decoding, each user can XOR the decoded common message with its own and thereby extract the other message. However, both approaches described above turn out to be suboptimal in general[4]. It has been shown by several authors [4, 26–28] that in the downlink phase the following rates are achievable[5]:

$$\begin{aligned} R_a &< \mathrm{C}\left(h_{rb}^2 P\right) \\ R_b &< \mathrm{C}\left(h_{ra}^2 P\right). \end{aligned} \tag{2.15}$$

Now one can combine the two results for the uplink and downlink phase and argue that for a successful message exchange with vanishing error probability the rates have to be chosen such that they lie in both achievability regions simultaneously, i.e. the intersection of (2.13) and (2.15).

Then, in summary the following rates are achievable with DF:

$$\begin{aligned} R_a &< \min\left(\mathrm{C}\left(h_{ar}^2 P\right), \mathrm{C}\left(h_{rb}^2 P\right)\right) \\ R_b &< \min\left(\mathrm{C}\left(h_{br}^2 P\right), \mathrm{C}\left(h_{ra}^2 P\right)\right) \\ R_a + R_b &< \mathrm{C}\left(h_{ar}^2 P + h_{br}^2 P\right). \end{aligned} \tag{2.16}$$

Comparing the achievable rates given by (2.16) to those implied by the upper bound in (2.5), it can be seen that both are identical with the exception that the last inequality in (2.16) puts an additional constraint on the sum rate of both users. In the context of the sTRC this additional constraint leads to a so-called multiplexing loss [4]. The explanation for this loss is that in the DF scheme the relay attempts to decode both messages individually even though this is not necessarily required. The strategies that are described in the next chapter are based on lattices and designed such that this multiplexing loss is avoided. Roughly, the relay does not attempt to decode both messages, but rather “protects” the sum of two lattice points (which can be seen as a function of the two messages) instead.

[4]The NC approach is optimal if and only if the downlink channel gains are equal [25].

[5]Note that these rates are also the point-to-point capacities and therefore an upper bound on the capacity region [28] for each link. Thus equation (2.15) gives the capacity region for the second phase of the DF strategy [25].

CHAPTER 3

Relaying Strategies based on Lattices

The reason to study lattices in the context of bidirectional relaying is motivated by the shortcoming of the decode-and-forward (DF) strategy described in the last chapter, i.e. the multiplexing loss. In this chapter it is shown that the *group-property* of a lattice in combination with the linearity of the wireless channel allows the relay to decode a function of the users messages instead of decoding each message individually and thereby avoiding this loss. The chapter starts with a general introduction to lattices and nested lattice codes while focusing on the application to the bidirectional relaying problem. The first section is largely based on a few introductory texts about lattices[1], in particular [30] and [31]. Then, the modulo-and-forward (MF) strategy [15] is reviewed and slightly generalized. In [15] the communication links from the relay to the two users are assumed to be symmetric and to have equal SNR. Here the MF scheme is applied with nested lattice codes to the general Gaussian sTRC as defined in the previous chapter and the achievable rates are given. After that, it is shown that the capacity region of this network can be characterized to within a constant gap by using a nested lattice code and superposition coding. This result is similar to the one that was recently published in [7], but the proof presented here uses a different technique for the uplink, following closely the analysis presented in [32] and [33]. Finally, two other relaying approaches that are based on lattices are briefly mentioned for completeness: Even though very tight "gap-to-capacity" results are available, the capacity region of the Gaussian sTRC is still not known in general.

[1] A recent tutorial paper on lattices for various problems of communication theory including communication networks can be found in [29].

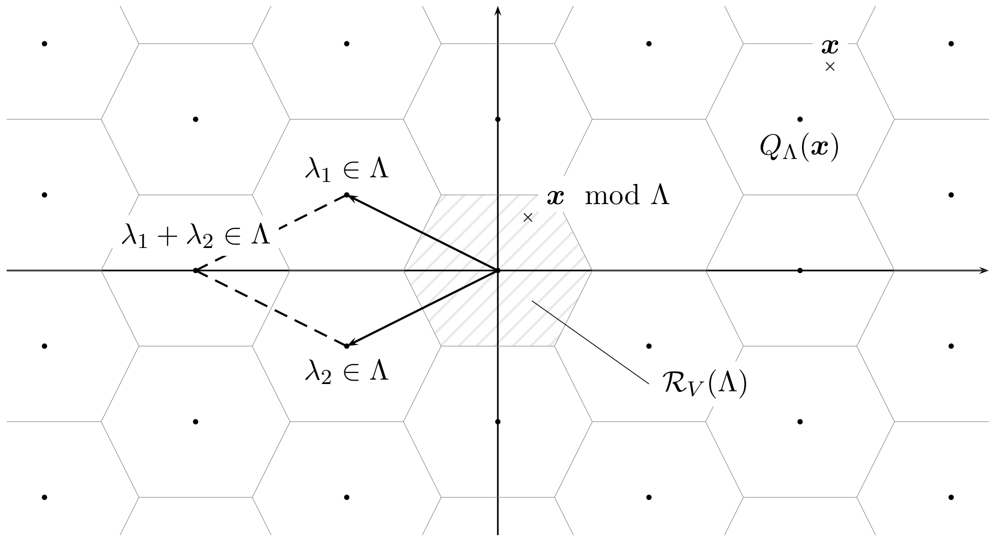


Figure 3.1.: Visualization of important definitions and concepts regarding lattices in $\mathbb{R}^2$.

3.1. Introduction

3.1.1. Lattices

A lattice Λ is a discrete subgroup of n-dimensional space $\mathbb{R}^n$ and an element of Λ is called a lattice point. In particular, this means that the all-zero vector $\mathbf{0} = (0, \dots, 0)$ is a lattice point and for any two lattice points $\lambda_1, \lambda_2 \in \Lambda$ the linear superposition $\lambda_1 + \lambda_2$ is also a lattice point. Moreover if $\lambda \in \Lambda$ then $-\lambda \in \Lambda$. The *nearest neighbor lattice quantizer* of a lattice Λ is defined as

$$Q_\Lambda(\boldsymbol{x}) = \arg\min_{\lambda \in \Lambda} ||\boldsymbol{x} - \lambda|| \tag{3.1}$$

and returns the closest lattice point to any given point $\boldsymbol{x} \in \mathbb{R}^n$, where $||\cdot||$ specifies the Euclidean norm and ties are broken in a systematic fashion. Two points are said to be "equivalent modulo Λ" if their difference is a lattice point. The modulo operation is defined as

$$\boldsymbol{x} \mod \Lambda = \boldsymbol{x} - Q_\Lambda(\boldsymbol{x}) \tag{3.2}$$

and satisfies the distributive property

$$(\boldsymbol{x} \mod \Lambda + \boldsymbol{y}) \mod \Lambda = (\boldsymbol{x} + \boldsymbol{y}) \mod \Lambda. \tag{3.3}$$

Associated with each lattice Λ is the *fundamental Voronoi region* or *fundamental region* $\mathcal{R}_V(\Lambda)$, which contains all points that are closer to the origin than to all other lattice points, i.e.

$$\mathcal{R}_V(\Lambda) = \{\boldsymbol{x} \mid Q_\Lambda(\boldsymbol{x}) = \boldsymbol{0}\}. \tag{3.4}$$

Each lattice point in Λ is surrounded by a translation of the fundamental region (which is then called the *Voronoi region* or *Voronoi cell*) and the union of all such regions is equal to n-dimensional space $\mathbb{R}^n$. A visualization of these concepts in two-dimensional space $\mathbb{R}^2$ is provided in Figure 3.1.
Important properties of a compact *region* $\mathcal{R} \subset \mathbb{R}^n$ in n-dimensional space include (see [34])

- the *volume*

$$\mathrm{Vol}(\mathcal{R}) = \int_{\mathcal{R}} \mathrm{d}\boldsymbol{x}, \tag{3.5}$$

- the *second moment per dimension* or average energy of a uniform distribution $\boldsymbol{U} \sim \mathrm{Unif}(\mathcal{R})$ over the region

$$\sigma^2(\mathcal{R}) = \frac{1}{n}\mathrm{E}\left[||\boldsymbol{U}||^2\right] = \frac{1}{n\mathrm{Vol}(\mathcal{R})} \int_{\mathcal{R}} ||\boldsymbol{x}||^2 \,\mathrm{d}\boldsymbol{x}, \tag{3.6}$$

- and the *normalized second moment*

$$\mathrm{G}(\mathcal{R}) = \frac{\sigma^2(\mathcal{R})}{\mathrm{Vol}(\mathcal{R})^{2/n}}. \tag{3.7}$$

  Note that for any region $\mathrm{G}(\mathcal{R}) > 1/(2\pi e)$, where $1/(2\pi e)$ is the normalized second moment of an n-ball for $n \to \infty$.

The above definitions apply to any compact bounding region in n-space (such as an n-ball or n-cube), and in particular to the fundamental region of a lattice Λ. In this context an abbreviated notation is used, i.e. instead of writing $\mathrm{Vol}(\mathcal{R}_V(\Lambda))$ and so on, the volume, second moment and normalized second moment of the fundamental region of a lattice Λ are denoted by $\mathrm{Vol}(\Lambda)$, $\sigma^2(\Lambda)$, and $\mathrm{G}(\Lambda)$ respectively.
It is possible to characterize a lattice Λ with the help of a generator matrix $G \in \mathbb{R}^{n\times n}$. Assuming that G has full rank then an n-dimensional Lattice is defined as

$$\Lambda = \{\lambda = \boldsymbol{i}G \mid \boldsymbol{i} \in \mathbb{Z}^n\}. \tag{3.8}$$

With other words, Λ is the image of the integer lattice $\mathbb{Z}^n$ under a linear transformation of $\mathbb{R}^n$ [35].
The following lemma is of fundamental importance with regard to the relaying strategies described in this chapter.

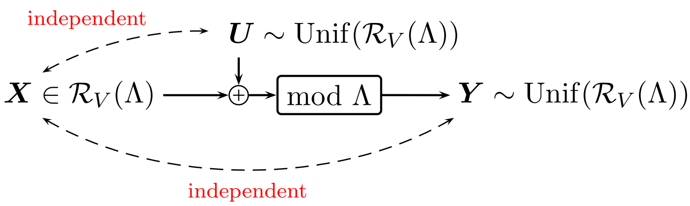


Figure 3.2.: Graphical illustration of the crypto lemma as a basic building block for the lattice based relaying strategies.

Lemma 2 (Crypto lemma [31, 34])**.** *Let $\boldsymbol{U} \sim \mathrm{Unif}(\mathcal{R}_V(\Lambda))$ and $\boldsymbol{X} \in \mathcal{R}_V(\Lambda)$ be any random vector statistically independent of $\boldsymbol{U}$. Then the sum $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{U} \mod \Lambda$ is uniformly distributed over $\mathcal{R}_V(\Lambda)$ and statistically independent of $\boldsymbol{X}$.*

A graphical illustration of this lemma[2] is shown in Figure 3.2. The crypto lemma is a basic building block for the lattice based strategies described in this chapter. $\boldsymbol{U}$ is referred to as a *dither variable* and statistically "decouples" $\boldsymbol{Y}$ from $\boldsymbol{X}$.

3.1.2. Nested Lattice Codes

A *coset* of a lattice, denoted by $\Lambda + \boldsymbol{x}$, is a translation of the original lattice by $\boldsymbol{x}$, where $\boldsymbol{x} \in \mathbb{R}^n$. Note that a lattice is geometrically uniform, in the sense that if $\boldsymbol{x} \in \Lambda$ then the coset will be identical to the lattice. Now consider two lattices Λ_c and Λ where $\Lambda_c \supset \Lambda$, i.e. Λ is a *sublattice* of Λ_c. Algebraically this induces a so called *partition*, denoted by Λ_c/Λ, of Λ_c into *cosets* of Λ, where one coset is Λ itself. A nested lattice code, also known as a Voronoi code[3], which will be denoted by $\mathcal{C}(\Lambda_c/\Lambda)$, is defined as the set of all coset leaders (or minimum-energy coset representatives) of the cosets that are introduced by this partition. Geometrically this can be seen as an intersection of Λ_c with the fundamental region of Λ, and algebraically as a reduction of Λ_c modulo Λ, i.e.

$$\mathcal{C}(\Lambda_c/\Lambda) = \{\Lambda_c \cap \mathcal{R}_V(\Lambda)\} = \{\Lambda_c \mod \Lambda\}. \tag{3.9}$$

In this context Λ_c is referred to as the *fine* or *coding lattice* and Λ as the *coarse* or *shaping lattice*. Note that any point in the fine lattice can be represented by a point in the coarse lattice plus the corresponding coset leader such that $\Lambda_c = \Lambda + \mathcal{C}(\Lambda_c/\Lambda)$ which is also called the coset leader decomposition. The number of codewords, i.e. the number of elements in

[2] The name of the lemma stems from the fact that it is impossible to infer any information about $\boldsymbol{X}$ ("plaintext") by observing $\boldsymbol{Y}$ ("cyphertext") without knowing $\boldsymbol{U}$ ("key") [34].

[3] According to [34] these codes were first called "Voronoi codes" or "Voronoi constellations" and later "nested lattice codes". We will use both terms throughout this thesis.

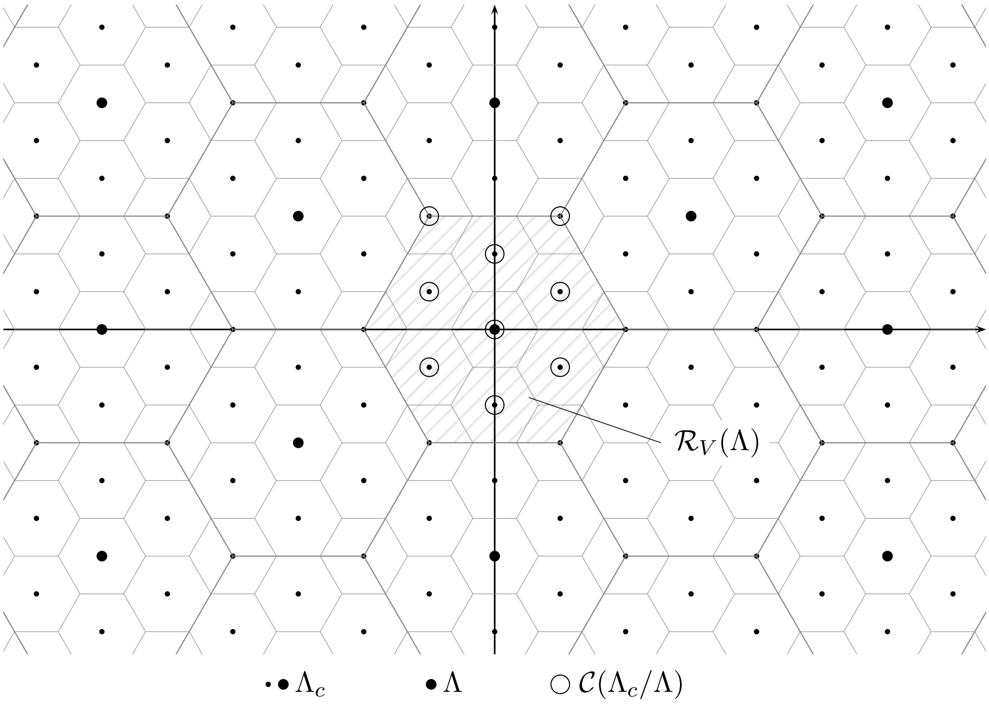


Figure 3.3.: Visualization of nested lattices $\Lambda_c \supset \Lambda$ and the corresponding nested lattice code $\mathcal{C}(\Lambda_c/\Lambda)$ in two dimensions.

the code is given by

$$|\mathcal{C}(\Lambda_c/\Lambda)| = \frac{\mathrm{Vol}(\Lambda)}{\mathrm{Vol}(\Lambda_c)}, \tag{3.10}$$

and the *code rate* of $\mathcal{C}(\Lambda_c/\Lambda)$ (in bits per dimension) is defined as

$$R_{\mathrm{Lattice}} = \frac{1}{n}\log_2|\mathcal{C}(\Lambda_c/\Lambda)|. \tag{3.11}$$

A visualization of the concepts which concern the nesting of lattices as well as nested lattice codes can be found in Figure 3.3.

3.1.3. Goodness of a Lattice

The results presented in this thesis rely on the existence of certain types of lattices[4] (and nested lattices) which are *good* in the following senses:

Good shaping lattices. A lattice Λ is said to be Rogers-good if the covering efficiency goes to 1 as $n \to \infty$ [37]. This implies that the normalized second moment of Λ approaches that of an n-ball, i.e. $G(\Lambda) \to 1/(2\pi e)$ as $n \to \infty$.

[4] More precisely, the existence of sequences of lattices $\Lambda^{(n)}$, where a sequence is indexed by the dimension n [36]. For convenience the superscript as well as the phrase "sequence of" are dropped.

Good coding lattices. Let $\boldsymbol{Z}$ be a random vector distributed according to $\boldsymbol{Z} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$. A lattice Λ is said to be Poltyrev-good if $\Pr(\{\boldsymbol{Z} \notin \mathcal{R}_V(\Lambda)\})$, i.e. the probability of $\boldsymbol{Z}$ falling outside of the fundamental region of Λ, goes to zero exponentially as $n \to \infty$ for any $\sigma^2 < \sigma^2(\Lambda)$.

Simultaneously good lattices. A lattice Λ is said to be simultaneously good, if Λ is both Rogers-good and Poltyrev-good.

The main results about the existence of good lattices are summarized in the following list:

- There exist lattices Λ that are simultaneously good [37].
- There exist nested lattices $\Lambda_c \supset \Lambda$ such that Λ is simultaneously good and Λ_c is Poltyrev-good [31].
- There exist nested lattices $\Lambda_c \supset \Lambda$ such that Λ and Λ_c are simultaneously good [36].

In the following example, it is shown how these results are used in order to derive achievable rates for the relaying strategies. Moreover, the decoding method – *lattice decoding* – is introduced.

Consider an additive white Gaussian noise (AWGN) channel model $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}$ with $\boldsymbol{Z} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ and a nested lattice code $\mathcal{C}(\Lambda_c/\Lambda)$, where both lattices are assumed to be simultaneously good. The elements of the code, i.e. the lattice points, are used as channel inputs and the receiver uses lattice decoding, which finds the closest lattice point to $\boldsymbol{Y}$ *in the coding lattice* Λ_c:

$$\hat{\boldsymbol{X}} = Q_{\Lambda_c}(\boldsymbol{Y}), \tag{3.12}$$

where $\hat{\boldsymbol{X}}$ is the estimated codeword. Note that lattice decoding – as opposed maximum likelihood (ML) decoding, which finds the closest lattice point to $\boldsymbol{Y}$ *in the code* $\mathcal{C}(\Lambda_c/\Lambda)$ – does not take into account any boundary of the code and therefore preserves the symmetry of the lattice [38]. In particular, this means that the decoding error at the receiver is independent of the transmitted codeword and simply given by:

$$p_e = \Pr(\{\boldsymbol{X} \neq Q_{\Lambda_c}(\boldsymbol{Y})\}) = \Pr(\{\boldsymbol{Z} \notin \mathcal{R}_V(\Lambda_c)\}). \tag{3.13}$$

Because Λ_c is Poltyrev-good, $p_e \to 0$ exponentially for $n \to \infty$ as long as $\sigma^2 < \sigma^2(\Lambda_c)$. Note that the coding rate of $\mathcal{C}(\Lambda_c/\Lambda)$ can be written as

$$R_{\text{Lattice}} = \frac{1}{n} \log_2 \left(\frac{\text{Vol}(\Lambda)}{\text{Vol}(\Lambda_c)} \right) = \frac{1}{2} \log_2 \left(\frac{\sigma^2(\Lambda)}{\sigma^2(\Lambda_c)} \right) + o_n(1), \tag{3.14}$$

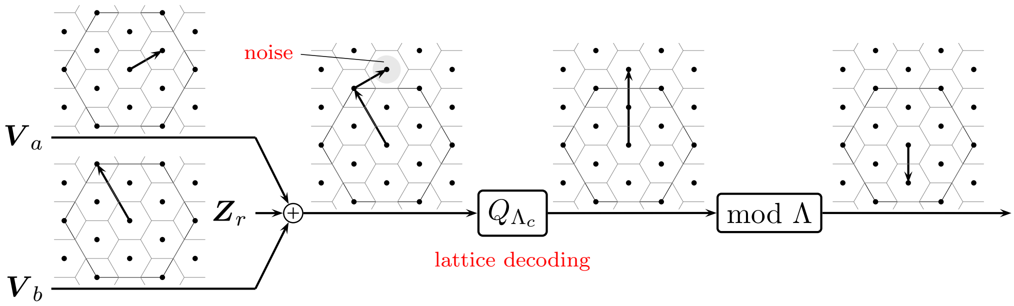


Figure 3.4.: Illustration of the basic idea behind PNC with nested lattice codes. The noise is visualized as a small uncertainty region around the received lattice point.

where $o_n(1) \to 0$ as $n \to \infty$, because both lattices are Rogers-good (cp. (3.7) and the definition of Rogers-goodness). If we assume $\sigma^2(\Lambda) = P$ and insert the condition for reliable decoding into (3.14), it can be seen that this transmission scheme can reliably approach any rate up to

$$R_{\text{Lattice}} < \frac{1}{2}\log_2\left(\frac{P}{\sigma^2}\right). \tag{3.15}$$

3.2. PNC with Nested Lattice Codes

In this section the basic idea behind the use of nested lattice codes for PNC is illustrated. Assume that the channel gains from the two users to the relay are given by $h_{ar} = h_{br} = 1$. Both users map their messages $\boldsymbol{W}_a$ and $\boldsymbol{W}_b$ one-to-one to codewords $\boldsymbol{V}_a$ and $\boldsymbol{V}_b$ from a nested lattice code $\mathcal{C}(\Lambda_c/\Lambda)$. The users then simply transmit lattice points corresponding to their messages. Because the channel gains are equal, these lattice points will "add up" such that the relay receives a (noisy) lattice point, due to the group property and the linearity of the wireless channel. The received lattice point is however not necessarily in the code $\mathcal{C}(\Lambda_c/\Lambda)$. Lattice decoding with respect to the coding lattice Λ_c and applying the modulo-operation with respect to the shaping lattice Λ allows the relay to recover a point in $\mathcal{C}(\Lambda_c/\Lambda)$, provided that the noise is not too strong. See Figure 3.4 for a visualization of this concept.

The relaying strategies described in the following sections are in a sense just extensions and modifications of this general idea. Roughly, they differ from the above description in the following aspects:

1. The nodes do not transmit lattice points, but *dithered* lattice points in accordance with the crypto lemma. This dithering serves two purposes. First, it ensures that the transmit power constraint is met because the dithered signals are uniformly

distributed over the basic Voronoi region of Λ. Secondly, the transmitted signals are statistically independent of the actual codewords.

2. At a receiving node linear minimum mean square error (MMSE) estimation is employed. The received signal is therefore multiplied by a linear scaling factor prior to subtracting the dither variables. It is shown in [31] that for an AWGN channel this linear scaling is key in order to achieve capacity. In [34] it is pointed out that this linear MMSE estimation is not only sufficient but necessary.

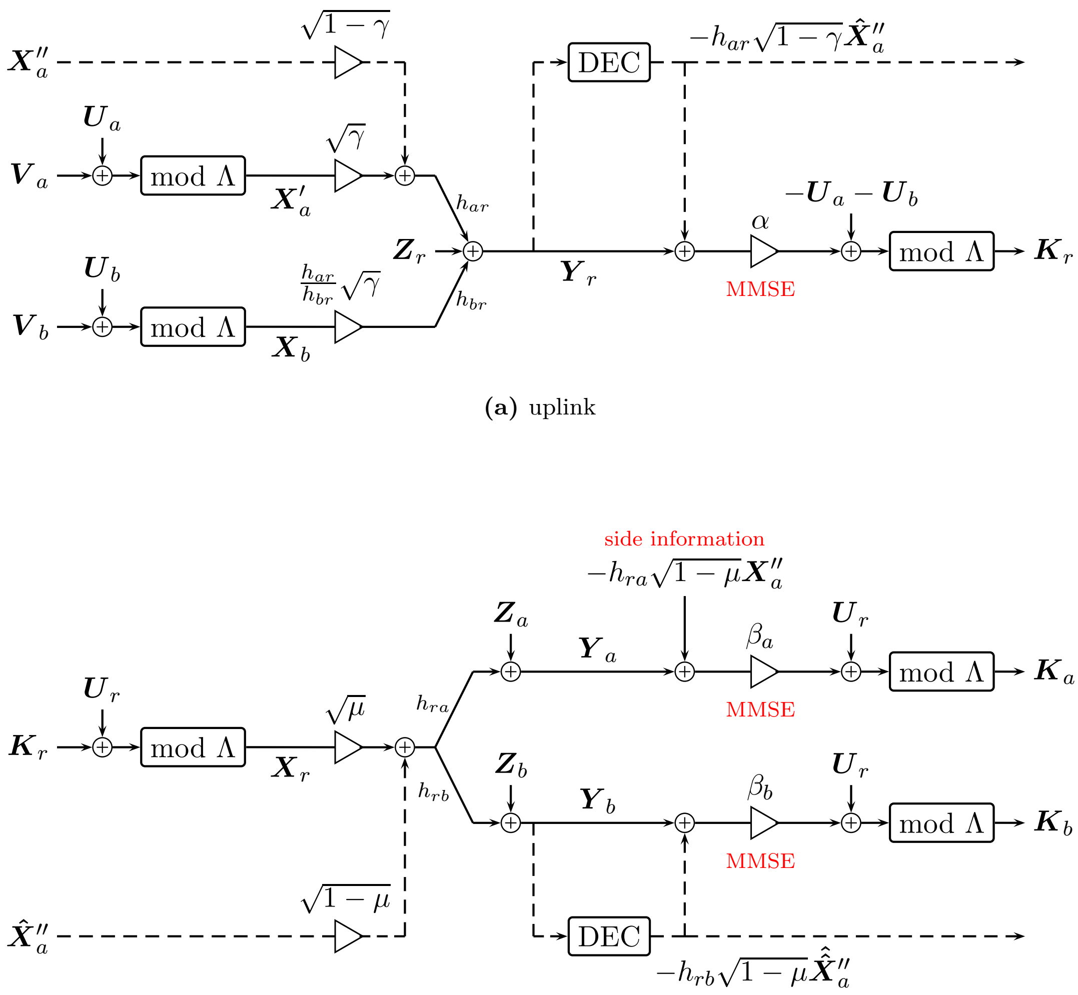
3. For the *general* Gaussian sTRC the assumption $h_{ar} = h_{br} = 1$ may not hold and the lattice points can be "mismatched" after passing through the channel. In order to circumvent this problem, superposition coding can be used for the user with the better uplink channel gain. This user splits the message into two parts and encodes the first part with a nested lattice code and the second part with a random Gaussian codebook. The power allocation is chosen such that the lattice codewords by both users are scaled correctly after passing through the channel. Note that this technique requires the assumption of full CSI at the transmitter.

3.3. Modulo-and-forward

The MF strategy is proposed in [15] and described as a power-efficient alternative to the AF scheme with less complexity compared to the DF scheme. In this section MF is applied to the *general* Gaussian sTRC as defined in chapter 2, i.e. all channel gains can be different, and the achievable rates are given. The description of the strategy and the derivation of the achievable rates presented here differs from [15] in two aspects. First, we use nested lattice codes instead of random codes. The use of structured codes is not necessarily required as will become clear later, but it makes the connection of MF to the PNC idea outlined in the previous section more apparent. Secondly, we introduce a power allocation factor for the user who employs superposition coding. A complete block diagram of the MF scheme is provided in Figure 3.5.

Encoding at the Users

It is assumed without loss of generality that $h_{ar} \geq h_{br}$ and user a splits the message vector $\boldsymbol{W}_a$ into two parts $(\boldsymbol{W}'_a, \boldsymbol{W}''_a)$, the lattice-encoded message and an extra message, where $\boldsymbol{W}'_a$ has length nR'_a and $\boldsymbol{W}''_a$ has length nR''_a. Both users use a nested lattice code for encoding the messages $\boldsymbol{W}'_a$ and $\boldsymbol{W}_b$ respectively. The code for user a is denoted by $\mathcal{C}(\Lambda_a/\Lambda)$ and the code for user b by $\mathcal{C}(\Lambda_b/\Lambda)$, where Λ is the common shaping lattice with



(a) uplink

(b) downlink

Figure 3.5.: Complete block diagram for the MF strategy. All paths corresponding to superposition coding are depicted with dashed lines while all paths corresponding to lattice coding are depicted with solid lines.

$\sigma^2(\Lambda) = P$, and $\Lambda_a \supset \Lambda$ and $\Lambda_b \supset \Lambda$ are coding lattices. All lattices are assumed to be simultaneously good. The messages are one-to-one mapped to codewords $\boldsymbol{W}'_a \leftrightarrow \boldsymbol{V}_a$ and $\boldsymbol{W}_b \leftrightarrow \boldsymbol{V}_b$, where $\boldsymbol{V}_a \in \mathcal{C}(\Lambda_a/\Lambda)$ and $\boldsymbol{V}_b \in \mathcal{C}(\Lambda_b/\Lambda)$. These codewords are dithered prior to transmission in accordance with the crypto lemma, i.e.

$$\boldsymbol{X}'_a = (\boldsymbol{V}_a + \boldsymbol{U}_a) \mod \Lambda \tag{3.16}$$

$$\boldsymbol{X}_b = (\boldsymbol{V}_b + \boldsymbol{U}_b) \mod \Lambda, \tag{3.17}$$

where $\boldsymbol{U}_a$ and $\boldsymbol{U}_b$ are independent dither variables, uniformly distributed over the fundamental Voronoi region of Λ, but known to all nodes. The signals $\boldsymbol{X}'_a$ and $\boldsymbol{X}_b$ are therefore statistically independent of the chosen codewords and it is ensured that they have average power $\sigma^2(\Lambda) = P$. The extra message $\boldsymbol{W}''_a$ of user a is encoded with a random Gaussian codebook with power P and the corresponding transmitted codeword is denoted by $\boldsymbol{X}''_a$. User a then uses superposition coding to transmit both the lattice-encoded message and the extra message according to

$$\boldsymbol{X}_a = \sqrt{\theta}\boldsymbol{X}'_a + \sqrt{1-\theta}\boldsymbol{X}''_a, \tag{3.18}$$

where θ is a power allocation factor and $0 \leq \theta \leq 1$ in order to satisfy the power constraint. For the transmitted signal of user b it is required that the (dithered) lattice codewords of both users are scaled correctly at the relay. This may be achieved by multiplying $\boldsymbol{X}_b$ at user b by $\sqrt{\theta}h_{ar}/h_{br}$ prior to transmission. Note that the power constraint at user b is met only if $\theta \leq h_{br}^2/h_{ar}^2$ and consequently it is required that $0 \leq \theta \leq \min(1, h_{br}^2/h_{ar}^2)$. One can check that by choosing θ smaller than $\min(1, h_{br}^2/h_{ar}^2)$ transmit power at user b is actually wasted. In [15] it is assumed that $\theta = \min(1, h_{br}^2/h_{ar}^2)$, i.e. the lattice codewords always get the maximal available power and only the rest is used for the Gaussian codebook. However, reducing the power that is used for transmitting the lattice codewords in favor of increasing power that is used for transmitting the codeword corresponding to the extra message of user a might be beneficial in terms of overall sum-rate throughput.

Uplink

Both users transmit the signals $\boldsymbol{X}_a$ and $\boldsymbol{X}_b$ and the relay receives

$$\begin{aligned}\boldsymbol{Y}_r &= h_{ar}(\sqrt{1-\theta}\boldsymbol{X}''_a + \sqrt{\theta}\boldsymbol{X}'_a) + h_{br}\frac{h_{ar}}{h_{br}}\sqrt{\theta}\boldsymbol{X}_b + \boldsymbol{Z}_r \\ &= h_{ar}\sqrt{1-\theta}\boldsymbol{X}''_a + h_{ar}\sqrt{\theta}(\boldsymbol{X}'_a + \boldsymbol{X}_b) + \boldsymbol{Z}_r.\end{aligned} \tag{3.19}$$

where (3.19) shows that the signals corresponding to the lattice points add up as desired. The relay first tries to decode the extra message of user a by treating all the other terms in (3.19) as noise. Note that $\boldsymbol{X}'_a$ and $\boldsymbol{X}_b$ are statistically independent and uniformly distributed over the fundamental Voronoi region of Λ according to the crypto lemma. Because Λ is Rogers-good, we can apply the so-called Gaussian approximation principle and assume that the distribution of $\boldsymbol{X}'_a$ and $\boldsymbol{X}_b$ is Gaussian and that this approximation becomes exact as $n \to \infty$, i.e. $\boldsymbol{X}'_a$ and $\boldsymbol{X}_b$ are *Gaussian in the limit* [34]. Decoding will therefore be successful with high probability if the rate of the extra message satisfies

$$R''_a < \mathrm{C}\left(\frac{(1-\theta)h_{ar}^2 P}{2\theta h_{ar}^2 P + 1}\right). \tag{3.20}$$

The relay can then subtract the codeword corresponding to the extra message from the received signal[5]. After that, the resulting signal is scaled by α, the random dithers $\boldsymbol{U}_a$ and $\boldsymbol{U}_b$ are subtracted and the modulo-operation is applied:

$$\begin{aligned}
\boldsymbol{K}_r &= (\alpha(\boldsymbol{Y}_r - h_{ar}\sqrt{1-\theta}\hat{\boldsymbol{X}}''_a) - \boldsymbol{U}_a - \boldsymbol{U}_b) \bmod \Lambda \\
&= (\alpha(h_{ar}\sqrt{\theta}(\boldsymbol{X}'_a + \boldsymbol{X}_b) + \boldsymbol{Z}_r) - \boldsymbol{U}_a - \boldsymbol{U}_b) \bmod \Lambda \\
&\overset{(a)}{=} ((\boldsymbol{V}_a + \boldsymbol{U}_a) \bmod \Lambda - \boldsymbol{X}'_a \\
&\quad + (\boldsymbol{V}_b + \boldsymbol{U}_b) \bmod \Lambda - \boldsymbol{X}_b \\
&\quad + \alpha(h_{ar}\sqrt{\theta}(\boldsymbol{X}'_a + \boldsymbol{X}_b) + \boldsymbol{Z}_r) - \boldsymbol{U}_a - \boldsymbol{U}_b) \bmod \Lambda \\
&\overset{(b)}{=} (\boldsymbol{V}_a + \boldsymbol{V}_b + (\alpha h_{ar}\sqrt{\theta} - 1)(\boldsymbol{X}'_a + \boldsymbol{X}_b) + \alpha \boldsymbol{Z}_r) \bmod \Lambda \\
&= (\boldsymbol{V}_a + \boldsymbol{V}_b + \tilde{\boldsymbol{Z}}_r) \bmod \Lambda,
\end{aligned} \tag{3.21}$$

where the so-called effective noise $\tilde{\boldsymbol{Z}}_r$ is defined as

$$\tilde{\boldsymbol{Z}}_r = \underbrace{(\alpha h_{ar}\sqrt{\theta} - 1)(\boldsymbol{X}'_a + \boldsymbol{X}_b)}_{\text{self-noise}} + \alpha \boldsymbol{Z}_r. \tag{3.22}$$

Note that for step (a) the identities (3.16) and (3.17) are used and for step (b) the distributive property of the modulo-operation is applied. The above calculation reveals that the relay receives the superposition of two lattice codewords[6] plus an effective noise term (modulo Λ). The effect of scaling the received signal prior to subtracting the dithers can be seen in equation (3.22). Suppose that $\alpha = 1/(h_{ar}\sqrt{\theta})$ for which the term marked as self-noise would be zero. This choice of α would correspond to a usual linear equalizer, which inverts the effect of the channel. The SNR in this case would amount to $\theta h_{ar}^2 P$. For now, the choice of α is left open and is described and explained further below.

[5] In the following it will be assumed that all intermediate decoding steps are successful and the estimates are equal to actually transmitted signals.

[6] However, because the coding lattices are different, this superposition is not necessarily a lattice point.

Relaying Strategy

An important characteristic of the MF strategy is that the relay simply forwards $\boldsymbol{K}_r$, i.e. the "noisy" superposition of the two lattice points[7]. The relay uses superposition coding in order to transmit

$$\boldsymbol{X}_r' = (\boldsymbol{K}_r + \boldsymbol{U}_r) \mod \Lambda, \tag{3.23}$$

which is the dithered version of $\boldsymbol{K}_r$, and the decoded extra message of user a according to

$$\boldsymbol{X}_r = \sqrt{\mu}\boldsymbol{X}_r' + \sqrt{1-\mu}\hat{\boldsymbol{X}}_a'', \tag{3.24}$$

where μ is a power allocation and $0 \leq \mu \leq 1$ in order to satisfy the power constraint at the relay.

Downlink

The two users receive the following signals:

$$\boldsymbol{Y}_a = h_{ra}\sqrt{\mu}\boldsymbol{X}_r' + h_{ra}\sqrt{1-\mu}\hat{\boldsymbol{X}}_a'' + \boldsymbol{Z}_a \tag{3.25}$$

$$\boldsymbol{Y}_b = h_{rb}\sqrt{\mu}\boldsymbol{X}_r' + h_{rb}\sqrt{1-\mu}\hat{\boldsymbol{X}}_a'' + \boldsymbol{Z}_b. \tag{3.26}$$

User a can use SI in order to remove the second term in the sum of (3.25), because $\boldsymbol{X}_a''$ is known and decoding is assumed to be successful at the relay. User b attempts to decode $\hat{\boldsymbol{X}}_a''$ by treating the other parts of the signal as noise. By the same argument as before, the signal $\boldsymbol{X}_r'$ is Gaussian in the limit and decoding will be successful if

$$R_a'' < \mathrm{C}\left(\frac{(1-\mu)h_{rb}^2 P}{\mu h_{rb}^2 P + 1}\right). \tag{3.27}$$

The codeword corresponding to the decoded extra message can then be stripped off. After that, both users scale the resulting signals by β_a and β_b, and subtract the random dither that is added at the relay according to

$$\boldsymbol{K}_a = (\beta_a(\boldsymbol{Y}_a - h_{ra}\sqrt{1-\mu}\boldsymbol{X}_a'') - \boldsymbol{U}_r) \mod \Lambda \tag{3.28}$$

$$\boldsymbol{K}_b = (\beta_b(\boldsymbol{Y}_b - h_{rb}\sqrt{1-\mu}\hat{\hat{\boldsymbol{X}}}_a'') - \boldsymbol{U}_r) \mod \Lambda, \tag{3.29}$$

where $\hat{\hat{\boldsymbol{X}}}_a''$ denotes the estimate of the relay's estimate of the Gaussian codeword of user a. In the following, the calculation of $\boldsymbol{K}_a$ is described while the steps are similar to the

[7] Note that, because the relay does not perform lattice decoding, there is no necessity to use the same coding lattices at the users. In fact, in [15] the codewords are assumed to be uniformly distributed over the fundamental Voronoi region of the shaping lattice Λ.

calculation for $\boldsymbol{K}_r$ at the relay above, i.e.

$$\begin{aligned}
\boldsymbol{K}_a &= (\beta_a(\boldsymbol{Y}_a - h_{ra}\sqrt{1-\mu}\boldsymbol{X}''_a) - \boldsymbol{U}_r) \mod \Lambda \\
&\overset{(a)}{=} ((\boldsymbol{K}_r + \boldsymbol{U}_r) \mod \Lambda - \boldsymbol{X}'_r + \beta_a h_{ra}\sqrt{\mu}\boldsymbol{X}'_r + \beta_a \boldsymbol{Z}_a - \boldsymbol{U}_r) \mod \Lambda \\
&\overset{(b)}{=} (\boldsymbol{K}_r + (\beta_a h_{ra}\sqrt{\mu} - 1)\boldsymbol{X}'_r + \beta_a \boldsymbol{Z}_a) \mod \Lambda, \qquad (3.30)
\end{aligned}$$

where for step (a) the identity (3.23) and for step (b) the distributive property of the modulo-operation is used. Recall that $\boldsymbol{K}_r$ is the noisy superposition of two lattice codewords defined by equation (3.21). Inserting this into (3.30) gives

$$\begin{aligned}
\boldsymbol{K}_a = (\underbrace{\boldsymbol{V}_a}_{\text{known}} &+ \boldsymbol{V}_b \qquad (3.31)\\
&+ (\alpha h_{ar}\sqrt{\theta} - 1)(\underbrace{\boldsymbol{X}'_a}_{\text{known}} + \boldsymbol{X}_b) + \alpha \boldsymbol{Z}_r \qquad (3.32)\\
&+ (\beta_a h_{ra}\sqrt{\mu} - 1)\boldsymbol{X}'_r + \beta_a \boldsymbol{Z}_a) \mod \Lambda. \qquad (3.33)
\end{aligned}$$

User a can cancel out the parts of the signal that are marked as known, i.e. the transmitted lattice point $\boldsymbol{V}_a$ and also some of the self-noise introduced at the relay. This yields

$$\begin{aligned}
\tilde{\boldsymbol{K}}_a &= \left(\boldsymbol{V}_b + (\alpha h_{ar}\sqrt{\theta} - 1)\boldsymbol{X}_b + \alpha \boldsymbol{Z}_r + (\beta_a h_{ra}\sqrt{\mu} - 1)\boldsymbol{X}'_r + \beta_a \boldsymbol{Z}_a\right) \mod \Lambda \qquad (3.34)\\
&= (\boldsymbol{V}_b + \tilde{\boldsymbol{Z}}_a) \mod \Lambda, \qquad (3.35)
\end{aligned}$$

where the effective noise $\tilde{\boldsymbol{Z}}_a$ is defined as

$$\tilde{\boldsymbol{Z}}_a = (\alpha h_{ar}\sqrt{\theta} - 1)\boldsymbol{X}_b + \alpha \boldsymbol{Z}_r + (\beta_a h_{ra}\sqrt{\mu} - 1)\boldsymbol{X}'_r + \beta_a \boldsymbol{Z}_a. \qquad (3.36)$$

Similar steps can be applied for the derivation of $\tilde{\boldsymbol{K}}_b$ which is given by

$$\begin{aligned}
\tilde{\boldsymbol{K}}_b &= \left(\boldsymbol{V}_a + (\alpha h_{ar}\sqrt{\theta} - 1)\boldsymbol{X}'_a + \alpha \boldsymbol{Z}_r + (\beta_b h_{rb}\sqrt{\mu} - 1)\boldsymbol{X}_r + \beta_b \boldsymbol{Z}_b\right) \mod \Lambda \qquad (3.37)\\
&= (\boldsymbol{V}_a + \tilde{\boldsymbol{Z}}_b) \mod \Lambda, \qquad (3.38)
\end{aligned}$$

where the effective noise $\tilde{\boldsymbol{Z}}_b$ is defined as

$$\tilde{\boldsymbol{Z}}_b = (\alpha h_{ar}\sqrt{\theta} - 1)\boldsymbol{X}'_a + \alpha \boldsymbol{Z}_r + (\beta_b h_{rb}\sqrt{\mu} - 1)\boldsymbol{X}'_r + \beta_b \boldsymbol{Z}_b. \qquad (3.39)$$

In summary, by applying the MF scheme both users “see” the transmitted lattice point of the respective other user plus some effective noise which is given by (3.36) for user a and by (3.39) for user b.

Lattice Decoding at the Users

It is assumed that both users employ lattice decoding in order estimate the lattice codeword, i.e. user a computes $\hat{\boldsymbol{V}}_b = Q_{\Lambda_b}(\tilde{\boldsymbol{K}}_a)$ while user b computes $\hat{\boldsymbol{V}}_a = Q_{\Lambda_a}(\tilde{\boldsymbol{K}}_b)$. The scaling factors α, β_a and β_b are chosen such that the effective noise power of $\tilde{\boldsymbol{Z}}_a$ and $\tilde{\boldsymbol{Z}}_b$ is minimized. The derivation can be found in Appendix A and with this choice for the scaling factors the effective noise power of $\tilde{\boldsymbol{Z}}_a$ and $\tilde{\boldsymbol{Z}}_b$ are then given by:

$$\tilde{\sigma}_a^2 = \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_a||^2\right] = \frac{P}{\theta h_{ar}^2 P + 1} + \frac{P}{\mu h_{ra}^2 P + 1} \tag{3.40}$$

$$\tilde{\sigma}_b^2 = \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_b||^2\right] = \frac{P}{\theta h_{ar}^2 P + 1} + \frac{P}{\mu h_{rb}^2 P + 1}. \tag{3.41}$$

It can be seen from (3.40) and (3.41) that the noise power is greater for the user with the weaker downlink channel gain. In the following, we describe the condition for reliable decoding at user a. The probability of decoding error is given by

$$p_e = \Pr(\{\boldsymbol{V}_b \neq Q_{\Lambda_b}(\tilde{\boldsymbol{K}}_a)\}) \tag{3.42}$$

$$= \Pr(\{\tilde{\boldsymbol{Z}}_a \mod \Lambda \notin \mathcal{R}_V(\Lambda_c)\}) \tag{3.43}$$

$$\leq \Pr(\{\tilde{\boldsymbol{Z}}_a \notin \mathcal{R}_V(\Lambda_c)\}), \tag{3.44}$$

where the second equality follows from the symmetry of lattice decoding and the inequality follows from the fact that the modulo operation is a many-to-one mapping [34]. Note that $\tilde{\boldsymbol{Z}}_a$ is Gaussian in the limit and therefore approaches a Gaussian distribution as $n \to \infty$. With this Gaussian approximation for $\tilde{\boldsymbol{Z}}_a$, the error probability (3.44) goes to zero exponentially[8] for $n \to \infty$ as long as $\tilde{\sigma}_a^2 < \sigma^2(\Lambda_b)$ because the coding lattice Λ_b is Poltyrev-good. Recall that the rate of $\mathcal{C}(\Lambda_b/\Lambda)$ can be written as

$$R_b = \frac{1}{n}\log_2\left(\frac{\mathrm{Vol}(\Lambda)}{\mathrm{Vol}(\Lambda_b)}\right) \tag{3.45}$$

$$= \frac{1}{2}\log_2\left(\frac{\sigma^2(\Lambda)}{\sigma^2(\Lambda_b)}\right) + o_n(1) \tag{3.46}$$

where $o_n(1) \to 0$ as $n \to \infty$, because both lattices are Rogers-good. It can be seen that the message of user b can reliably approach any rate up to

$$R_b < \frac{1}{2}\log_2\left(\frac{\sigma^2(\Lambda)}{\tilde{\sigma}_a^2}\right) = \frac{1}{2}\log_2\left(\frac{P}{\tilde{\sigma}_a^2}\right) \tag{3.47}$$

and inserting (3.40) into (3.47) yields:

$$R_b < \frac{1}{2}\log_2\left(\frac{P}{\frac{P}{\theta h_{ar}^2 P+1} + \frac{P}{\mu h_{ra}^2 P+1}}\right) = \frac{1}{2}\log_2\left(1 + \frac{\theta h_{ar}^2 P \mu h_{ra}^2 P - 1}{\theta h_{ar}^2 P + \mu h_{ra}^2 P + 2}\right). \tag{3.48}$$

[8] For a detailed analysis including error exponents see [31].

For user b the same calculation can be done which results in the following rate constraint for the lattice-encoded message of user a:

$$R_a'' < \frac{1}{2}\log_2\left(1 + \frac{\theta h_{ar}^2 P \mu h_{rb}^2 P - 1}{\theta h_{ar}^2 P + \mu h_{rb}^2 P + 2}\right). \tag{3.49}$$

In summary, the achievable rates for MF are given by

$$R_a < \mathrm{C}\left(\frac{\theta h_{ar}^2 P \mu h_{rb}^2 P - 1}{\theta h_{ar}^2 P + \mu h_{rb}^2 P + 2}\right) + \min\left(\mathrm{C}\left(\frac{(1-\mu)h_{rb}^2 P}{\mu h_{rb}^2 P + 1}\right), \mathrm{C}\left(\frac{(1-\theta)h_{ar}^2 P}{2\theta h_{ar}^2 P + 1}\right)\right) \tag{3.50}$$

$$R_b < \mathrm{C}\left(\frac{\theta h_{ar}^2 P \mu h_{ra}^2 P - 1}{\theta h_{ar}^2 P + \mu h_{ra}^2 P + 2}\right) \tag{3.51}$$

for any choice of $0 \le \theta \le \min(1, h_{br}^2/h_{ar}^2)$ and $0 \le \mu \le 1$.

3.4. Achievability of the Capacity Region to Within a Constant Gap

In this section it is shown that the capacity region of the Gaussian sTRC can be characterized to within a constant gap. This is made precise with the following theorem.

Theorem 1. *The cut-set region $\mathcal{R}_{cut}$ of the Gaussian sTRC is achievable to within* $1/2$ *bits per dimension for each user, i.e. if $(R_a, R_b) \in \mathcal{R}_{cut}$, then $([R_a - 1/2]^+, [R_b - 1/2]^+)$ is achievable, where $[x]^+ \overset{def.}{=} \max(x, 0)$.*

Theorem 1 is proved with a similar superposition coding technique as used in the MF strategy. However, the relay performs lattice decoding instead of broadcasting a noisy superposition of two lattice points. Moreover, for the downlink a nested, random Gaussian codebook is used. This is similar to the downlink strategies that are described in [7] and [39]. Note that in [7] a complete proof of theorem 1 is also provided. A particular strategy based different shaping lattices is considered for the uplink phase which is different from the approach described in the following.

Proof Outline

The proof involves the following main steps:

- First it is shown that it is sufficient to proof theorem 1 for the channel configuration where $h_{ar} = h_{rb} = h$ and $h_{br} = h_{ra} = \bar{h}$ with $\bar{h} \le h$. Roughly, all other channel configurations can be converted to this one without changing the cut-set bound. The case $\bar{h} > h$ follows by symmetry.

- The uplink phase of the achievability strategy is based on the uplink phase of the MF strategy described in the previous section (cp. Figure 3.5 (a)). However, the relay uses lattice decoding to decode the sum of the users lattice codewords (modulo Λ). Then, after the uplink phase the relay is aware of the extra message of user a as well as a function of the lattice codewords of the users. The latter can be regarded as a structured binning of the common-rate message of user a and the message of user b [7].

- For the downlink phase, a two-dimensional (or nested) random Gaussian codebook is used to convey the structured binning reliably to both users as well as the extra message from user a to user b. From the binning, each user can extract the corresponding message of the other user.

In the following subsection the proof is described in detail.

3.4.1. Proof

It is assumed without loss of generality that the weakest of the four channel gains is either h_{br} or h_{rb} (i.e. a gain in the path $b \to a$). The following lemma, which is a modified version of lemma 1 in [32] and lemma 2 in [33], then simplifies the number of channel gain orderings that must be considered in order to proof theorem 1.

Lemma 3. *For the case where the weakest channel gain is either h_{br} or h_{rb}, it is possible to convert the Gaussian sTRC with parameters parameters $(P, h_{ar}, h_{br}, h_{ra}, h_{rb})$ to the case with parameters $(\bar{P}, h, h, h', h')$ satisfying $h' \leq h$ without changing the cut-set bound.*

Proof. See Figure 3.6 for an illustration, where $\mathrm{C}_{ij} \overset{\text{def.}}{=} \mathrm{C}(h_{ij}^2 P)$. We have two (independent) orderings for $\mathrm{C}_{ar}, \mathrm{C}_{rb}$ and $\mathrm{C}_{br}, \mathrm{C}_{ra}$ respectively which are depicted in either black or red color. The operations that are involved in order to convert the channel to the case described in the lemma are depicted in the picture with the corresponding color. □

It can now be assumed that $h_{ar} = h_{rb}$, $h_{br} = h_{ra}$ and $h_{ar} \geq h_{br}$ without loss of generality. The encoding process at the users is similar to the encoding process for the MF strategy described in the previous section and only the differences are pointed here. In particular user a splits its message into two parts $\boldsymbol{W}_a = (\boldsymbol{W}'_a, \boldsymbol{W}''_a)$, where the lattice-encoded message $\boldsymbol{W}'_a$ has the same length nR_b as $\boldsymbol{W}_b$, and the extra message has length nR''_a. $\boldsymbol{W}'_a$ and $\boldsymbol{W}_b$ are both encoded using *the same* nested lattice code $\mathcal{C}(\Lambda_c/\Lambda)$, where both lattices

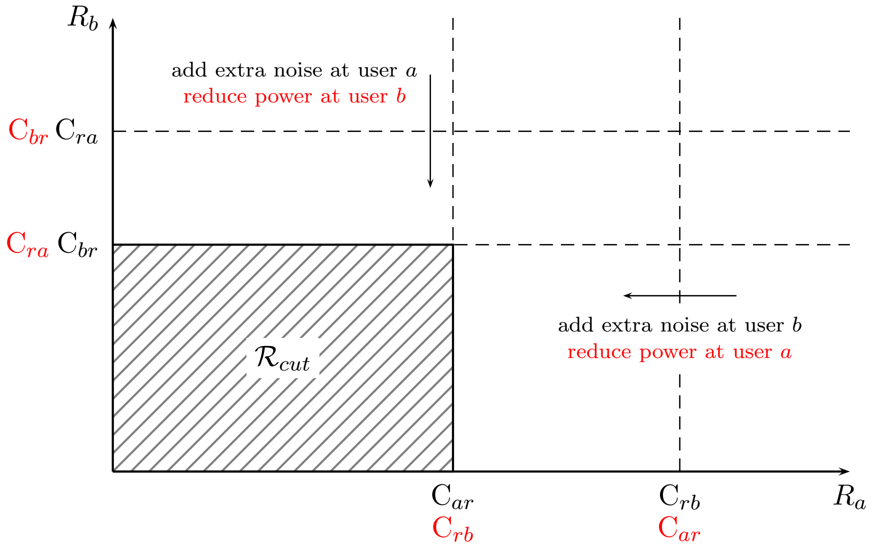


Figure 3.6.: Picture for the proof of lemma 3. Note that the smallest channel gain is in the path $b \to a$.

are simultaneously good. The codewords are dithered prior to transmission as before. The extra message of user a is encoded with a random Gaussian codebook. In contrast to the MF scheme, the power allocation factor for the superposition coding at user a is given by $\theta = h_{br}^2/h_{ar}^2$ and consequently no scaling of the transmitted signal of user b occurs (cp. Figure 3.5 (a)). This means that user b always transmits with the maximal available power.

Uplink

The two users transmit their signals and the relay first decodes the extra message of user a. The condition for reliable decoding of $\boldsymbol{W}_a''$ at the relay is given by

$$R_a'' < \mathrm{C}\left(\frac{(h_{ar}^2 - h_{br}^2)P}{2h_{br}^2 P + 1}\right), \tag{3.52}$$

which follows from equation (3.20) if θ is chosen as described. The codeword corresponding to the extra message of user a can then be stripped off the received signal. After scaling the resulting signal by α and subtracting the dither variables, the relay obtains (cp. (3.21) and (3.22))

$$\boldsymbol{K}_r = (\boldsymbol{V}_a + \boldsymbol{V}_b + \tilde{\boldsymbol{Z}}_r) \bmod \Lambda, \tag{3.53}$$

where the effective noise $\tilde{\boldsymbol{Z}}_r$ is defined as

$$\tilde{\boldsymbol{Z}}_r = (\alpha h_{br} - 1)(\boldsymbol{X}_a' + \boldsymbol{X}_b) + \alpha \boldsymbol{Z}_r. \tag{3.54}$$

The scaling factor α is chosen such that the effective noise power is minimized and the calculation is included in Appendix A. For this choice of α the effective noise power of $\tilde{\boldsymbol{Z}}_r$ is given by

$$\tilde{\sigma}_r^2 = \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_r||^2\right] = \frac{2P}{2h_{br}^2 P + 1}. \tag{3.55}$$

The relay then uses lattice decoding according to $\hat{\boldsymbol{V}} = Q_{\Lambda_c}(\boldsymbol{K}_r)$ in order to estimate the superposition of the two transmitted lattice points (modulo Λ), i.e.

$$\boldsymbol{V} = (\boldsymbol{V}_a + \boldsymbol{V}_b) \mod \Lambda. \tag{3.56}$$

Note that $\boldsymbol{V}_a$ and $\boldsymbol{V}_b$ are both lattice points in the same coding lattice. Therefore, the superposition described by (3.56) is also a lattice point. The condition for reliable decoding with this method is given by

$$R_b < \frac{1}{2}\log_2\left(\frac{P}{\tilde{\sigma}_r^2}\right) = \frac{1}{2}\log_2\left(\frac{1}{2} + h_{br}^2 P\right). \tag{3.57}$$

The error probability of lattice decoding (i.e. the probability that the effective noise $\tilde{\boldsymbol{Z}}_r$ falls outside the fundamental region of Λ_c) goes to zero exponentially as $n \to \infty$ by the virtue of Λ_c being Poltyrev-good [6].

Downlink

It is assumed that decoding at the relay is successful. Then the relay knows both the extra message $\boldsymbol{W}_a''$ of user a as well as $\boldsymbol{V} = (\boldsymbol{V}_a + \boldsymbol{V}_b) \mod \Lambda$, i.e. a function of the lattice codewords of the users. If the codewords of the nested lattice code $\mathcal{C}(\Lambda_c/\Lambda)$ are interpreted as bins, then the latter equation can be regarded as a structured binning of message pairs [7]: Each of the 2^{nR_b} bins (one for each lattice codeword) “contains” 2^{nR_b} message pairs.

The codebook construction for the downlink is visualized in Figure 3.7. The 2^{nR_a} codewords $\boldsymbol{X}_r$ are assumed to be randomly generated according to a Gaussian distribution. The codewords are arranged in such a way that a two-dimensional codebook is formed. The first dimension is indexed by the extra message of user a (i.e. the codebook has $M = 2^{n(R_a - R_b)}$ columns) and the second dimension is indexed by the bin number corresponding to the received lattice codeword $\boldsymbol{V}$ (i.e. the codebook has $N = 2^{nR_b}$ rows). This can also be seen as a nested code, in the sense that each column of the two-dimensional code forms a codebook of its own and is a sub-code “nested” in the original code, e.g. the codewords that are shaded in Figure 3.7. The relay then simply broadcasts the codeword $\boldsymbol{X}_r(i,j)$, that corresponds to the received extra message j and the bin number i.

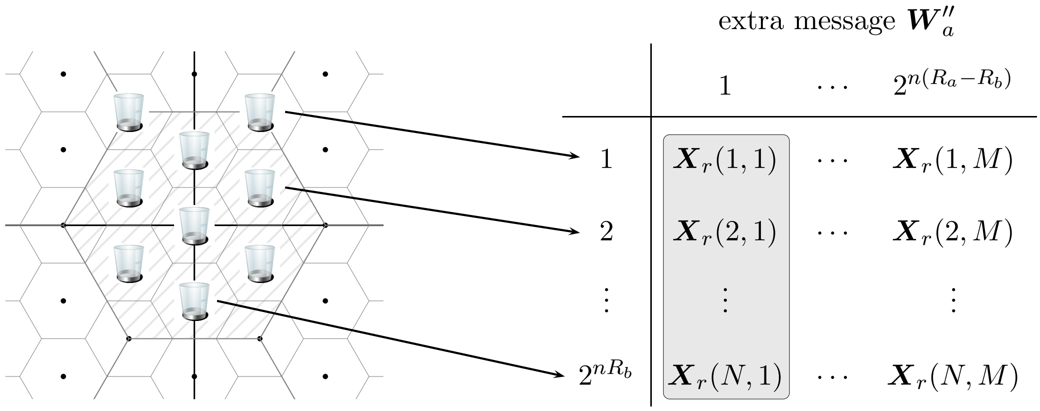


Figure 3.7.: Two-dimensional codebook generation at the relay. The first dimension is indexed by the extra message of user a and the second dimension by the bin index corresponding to the received lattice codeword. Each column forms a sub-code.

Recall that user a knows the extra message $\boldsymbol{X}_a''$ as SI and therefore only needs to decode with respect to a particular sub-code. It can be shown with joint typicality arguments that the condition for reliable decoding at user a is given by [26]

$$R_b < \frac{1}{2}\log_2(1 + h_{ra}^2 P) = \mathrm{C}\left(h_{ra}^2 P\right). \tag{3.58}$$

User a then knows the bin index that is received at the relay. User b cannot exploit any SI for decoding and therefore performs decoding with respect to all codewords of the two-dimensional codebook, which has cardinality 2^{nR_a}. The condition for reliable decoding is therefore given by

$$R_a < \frac{1}{2}\log_2(1 + h_{rb}^2 P) = \mathrm{C}\left(h_{rb}^2 P\right). \tag{3.59}$$

User b then knows both the extra message of user a as well as the bin index that is received at the relay. Note that if each user knows the correct bin index, the message of the other user can be obtained by calculating

$$(\boldsymbol{V} - \boldsymbol{V}_a) \bmod \Lambda = ((\boldsymbol{V}_a + \boldsymbol{V}_b) \bmod \Lambda - \boldsymbol{V}_a) \bmod \Lambda = \boldsymbol{V}_b \tag{3.60}$$

for user a and

$$(\boldsymbol{V} - \boldsymbol{V}_b) \bmod \Lambda = ((\boldsymbol{V}_a + \boldsymbol{V}_b) \bmod \Lambda - \boldsymbol{V}_b) \bmod \Lambda = \boldsymbol{V}_a \tag{3.61}$$

for user b. Also note that (3.58) and (3.59) correspond to the capacity region of the downlink phase revealing that the described strategy for the downlink is optimal.

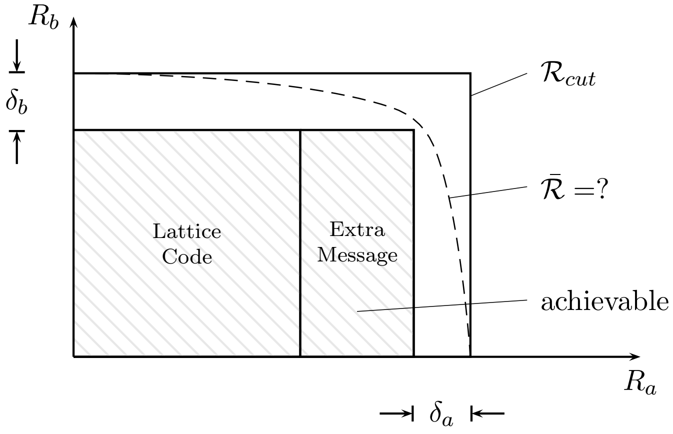


Figure 3.8.: Visualization of the achievable rate region in comparison to the cut-set region and the (unknown) capacity region .

Gap Analysis

Figure 3.8 shows a conceptual visualization of the achievable rate region for the described strategy in comparison to the cut-set region $\mathcal{R}_{cut}$. For user b the gap δ_b to the cut-set can be upper-bounded by

$$\delta_b = \frac{1}{2}\log_2(1+h_{br}^2P) - \frac{1}{2}\log_2\left(\frac{1}{2}+h_{br}^2P\right) \tag{3.62}$$

$$= \frac{1}{2}\log_2(1+h_{br}^2P) - \frac{1}{2}\log_2(1+2h_{br}^2P) + \frac{1}{2} \tag{3.63}$$

$$= \frac{1}{2}\log_2\left(\frac{1+h_{br}^2P}{1+2h_{br}^2P}\right) + \frac{1}{2} \tag{3.64}$$

$$< \frac{1}{2}. \tag{3.65}$$

For user a the gap δ_a can be upper-bounded by

$$\delta_a = \frac{1}{2}\log_2(1+h_{ar}^2P) - \frac{1}{2}\log_2\left(\frac{1}{2}+h_{br}^2P\right) - \frac{1}{2}\log_2\left(1+\frac{(h_{ar}^2-h_{br}^2)P}{2h_{br}^2P+1}\right) \tag{3.66}$$

$$= \frac{1}{2}\log_2(1+h_{ar}^2P) - \frac{1}{2}\log_2(1+2h_{br}^2P) - \frac{1}{2}\log_2\left(1+\frac{(h_{ar}^2-h_{br}^2)P}{2h_{br}^2P+1}\right) + \frac{1}{2} \tag{3.67}$$

$$= \frac{1}{2}\log_2(1+h_{ar}^2P) - \frac{1}{2}\log_2(1+h_{br}^2P+h_{ar}^2P) + \frac{1}{2} \tag{3.68}$$

$$= \frac{1}{2}\log_2\left(\frac{1+h_{ar}^2P}{1+h_{br}^2P+h_{ar}^2P}\right) + \frac{1}{2} \tag{3.69}$$

$$< \frac{1}{2}. \tag{3.70}$$

This proofs the theorem. Moreover, from (3.64) it can be seen that the gap vanishes for user b for fixed channel gains and increasing node power, i.e. $\delta_b \to 0$ as $P \to \infty$. Also from (3.69) it can be seen that the gap for user a is close to zero for fixed channel gains and increasing node power if the uplink channel gains are almost equal. However, the gap is close to 1/2 if the uplink channel gain for user b is small compared to the uplink channel gain of user a, i.e. if $h_{br} \lesssim h_{ar}$, $\delta_a \approx 0$ as $P \to \infty$ and if $h_{br} \ll h_{ar}$, $\delta_a \approx 1/2$ as $P \to \infty$.

3.5. Other Approaches

In this section two other approaches for the Gaussian sTRC that are based on lattices are briefly mentioned for completeness.

In [5] the use of lattice chains is considered. In particular user a uses $\mathcal{C}(\Lambda_c/\Lambda_a^S)$ and user b uses $\mathcal{C}(\Lambda_c/\Lambda_b^S)$ where $\Lambda_c \supset \Lambda_b^S \supset \Lambda_a^S$ forms a lattice chain, i.e. the nested lattice codes of the users are generated with the same coding lattice but different shaping lattices in order to account for asymmetric channel gains. It is shown that the achievable rate pairs for the Gaussian sTRC using this approach are given by

$$\begin{aligned} R_a &< \min\left(\left[\frac{1}{2}\log_2\left(\frac{h_{ar}^2}{h_{ar}^2+h_{br}^2}+h_{ar}^2P\right)\right]^+, \frac{1}{2}\log_2(1+h_{rb}^2P)\right) \\ R_b &< \min\left(\left[\frac{1}{2}\log_2\left(\frac{h_{br}^2}{h_{br}^2+h_{ar}^2}+h_{br}^2P\right)\right]^+, \frac{1}{2}\log_2(1+h_{ra}^2P)\right), \end{aligned} \tag{3.71}$$

where $[x]^+ \overset{\text{def.}}{=} \max(x,0)$. This result also implies a gap of 1/2 bit per dimension to capacity for each user. Moreover, the gap to the cut-set bound vanishes for increasing uplink SNR for *both* users, i.e. this scheme is essentially optimal when the uplink SNR is high.

Next to this, the compute-and-forward (CF) strategy [12] provides another approach for this network model which is also based on nested lattice codes. It should be noted however, that CF is much more general and provides a whole framework for computing functions of messages over noisy multi-access channels. Also, this approach is very relevant from a practical implementation point-of-view because no CSI is required at the users in the uplink phase. The main idea can be described as follows. The users transmit (dithered) codewords using the same nested lattice code. If fading occurs, i.e. if the uplink channel gains are unequal, the relay tries to recover an *integer combination* of the codewords, which is itself a codeword again. In [16] a tutorial introduction for CF is included.

CHAPTER 4

Sum Rate Comparison

In this chapter the relaying strategies that have been described in the previous chapters, namely amplify-and-forward (AF), decode-and-forward (DF), and modulo-and-forward (MF), are compared in terms of achievable sum rate. The sum rate is denoted by $R_{\text{sum}} = R_a + R_b$ and is a measure of total throughput for the network. We compare the maximal sum rate that is achievable with the aforementioned relaying strategies for different channel parameters P, h_{ar}, h_{br}, h_{ra} and h_{rb}. At first, the symmetric case is considered, where all channel gains are set to 1 and the sum rate is plotted for varying node power P. After that, three cases are considered where the channel gains have different values. In particular we show the sum rate performance of the strategies for the case where the uplink channel gains (h_{ar}, h_{br}) are weak compared to the downlink channel gains (h_{ra}, h_{rb}), then for the case where the uplink gains are strong compared to the downlink gains, and finally for the case where the channel gains are reciprocal, i.e. where $h_{ar} = h_{ra}$ and $h_{br} = h_{rb}$.
For MF the power allocation factor at user a is fixed at $\theta = \min(1, h_{br}^2/h_{ar}^2)$, i.e. the codewords of the nested lattice codes of both users get the maximum power and only the rest that is available at user a is allocated to the extra message. The power allocation factor μ at the relay is optimized numerically. This means that MF is effectively applied and compared as it is proposed in [15]. However, we also illustrate the benefit of optimizing θ numerically with a short example.

4.1. Symmetric Case

The first case that is considered is the symmetric case, where all channel gains are equal and given by $h_{ar} = 1$, $h_{br} = 1$, $h_{rb} = 1$, and $h_{ra} = 1$. The maximal sum rates for the different strategies are:

$$\text{AF:} \qquad R_{\text{sum}} < \log_2\left(1 + P\,\frac{P}{3P+1}\right) \tag{4.1}$$

$$\text{DF:} \qquad R_{\text{sum}} < \frac{1}{2}\log_2(1 + 2\,P) \tag{4.2}$$

$$\text{MF:} \qquad R_{\text{sum}} < \log_2\left(\frac{1}{2} + \frac{P}{2}\right) \tag{4.3}$$

An upper bound follows from the cut-set region and is given by

$$R_{\text{sum}} < \log_2(1 + P). \tag{4.4}$$

In addition to the relaying strategies AF, DF, and MF, for this channel configuration we also plot the maximal sum rate of the achievability strategy that is used in the proof of section 3.4 in the previous chapter. For the symmetric case the maximal sum rate for this strategy is given by

$$R_{\text{sum}} < \log_2\left(\frac{1}{2} + P\right). \tag{4.5}$$

Note that this is identical to the maximal sum rate that can be achieved with CF [16] as well as the lattice chain scheme [7] (cp. equation (3.71)). With other words, these approaches are identical in terms of sum rate for the symmetric case and for simplicity we refer to these schemes as CF in the following.

The achievable sum rates are plotted in Figure 4.1. It can be seen that in the low SNR regime ($P < 0$ dB) the DF strategy dominates the other strategies and appears to be optimal for decreasing node power P. If we compare the achievable sum rate for DF to the upper bound, i.e.

$$\frac{1}{2}\log_2(1 + 2P) \le \log_2(1 + P) = \frac{1}{2}\log_2(1 + 2P + P^2), \tag{4.6}$$

it can be seen that at low SNR, the term P^2 becomes small compared to the other terms and therefore the DF strategy in fact approaches the upper bound for decreasing node power. The AF scheme has very poor performance at low SNR for this channel configuration because the relay does not remove the noise that is introduced in the uplink but rather amplifies the received signal and therefore also the noise. It can be seen from (4.1) that for low node power P the maximal sum rate for AF can be approximated by $\approx \log_2(1 + P^2)$, i.e. the sum rate approaches zero much more quickly than the sum rate for the DF scheme

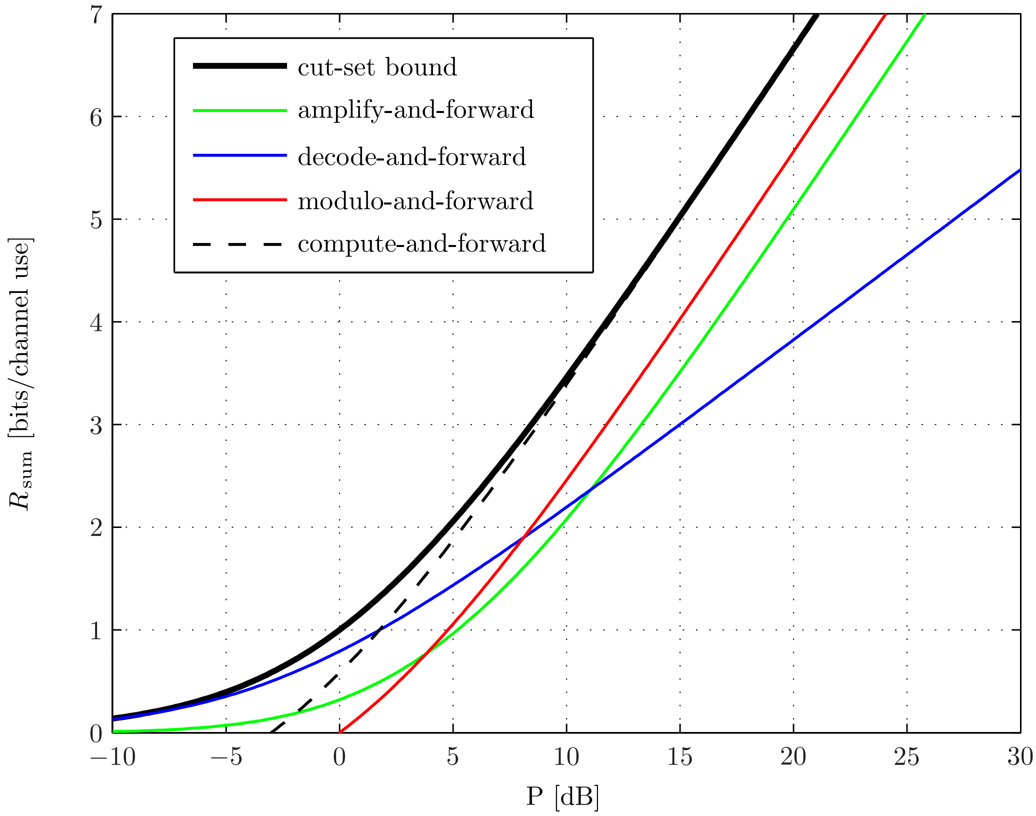


Figure 4.1.: Achievable sum rates for the symmetric case, i.e. $h_{ar} = 1$, $h_{br} = 1$, $h_{ra} = 1$, and $h_{rb} = 1$.

for $P \to 0$. Both MF and CF don't achieve a positive sum rate below a certain threshold, which is given by 0 dB and −3 dB, respectively. For high node power ($P > 0$ dB), the situation is very different and almost reverses. First, it can be seen that all strategies except DF have a positive slope of 1 bit per 3 dB power increase. For DF however, every 3 dB increase in node power only yields 1/2 bit increase in sum rate and therefore this strategy gets outperformed by the other strategies at high SNR. The maximal sum rate for AF can be approximated by $\approx \log_2(P/3) \approx \log_2(P) - 1.6$ and for MF by $\approx \log_2(P) - 1$ at high SNR. Recall that both strategies don't decode and therefore noise accumulates. However, MF is superior to AF at high SNR for this channel configuration. Finally, it can be observed that the CF approach is essentially optimal at high SNR and the gap to the upper bound approaches zero for increasing node power which can be seen from (4.5) and (4.4).

In the following, we briefly illustrate the effect of the power allocation factor θ in the MF scheme on the sum rate. In Figure 4.2 the maximal sum rate that is achievable with

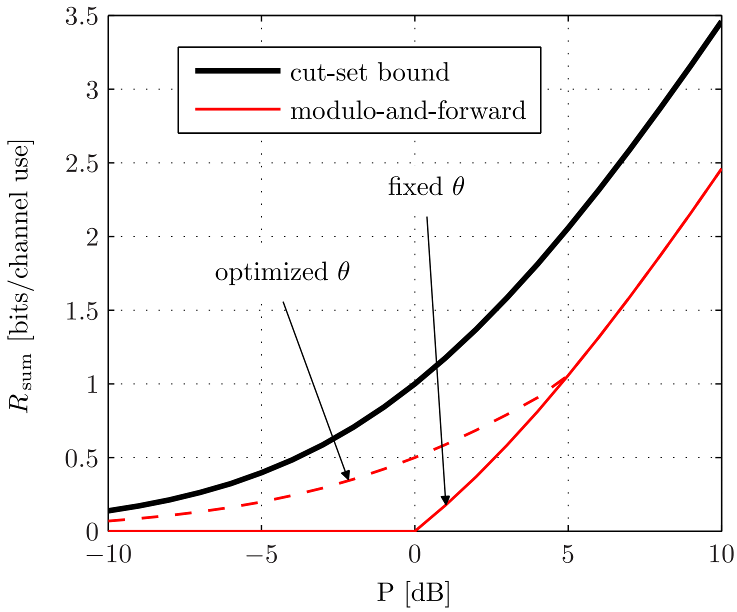


Figure 4.2.: Effect of the power allocation factor θ on the sum rate for the MF scheme.

MF is plotted for two different scenarios. In the first scenario (solid line), the power allocation factor θ is set to the maximal value, i.e. $\theta = \min(1, h_{br}^2/h_{ar}^2)$, and consequently the resulting curve is identical to the red curve plotted in Figure 4.1. In the second scenario (dashed line), the power allocation factor θ is chosen such that the overall sum rate R_{sum} is maximized. It can be seen that for low node power P, it is beneficial to reduce the power that is used for the lattice in favor of increasing the power for the extra message at user a.

4.2. Asymmetric Cases

In the following, three cases are considered that demonstrate the sum rate performance of the strategies for asymmetric channel configurations.

4.2.1. Weak Uplink

First, we consider the case where the channel coefficients are given by $h_{ar} = -20$ dB, $h_{br} = -30$ dB, $h_{ra} = 0$ dB, and $h_{rb} = 0$ dB. This implies that the uplink SNR is weak compared to the downlink SNR. The sum rates that can be achieved with the different strategies are plotted in Figure 4.3. It can be seen that AF performs very close to the upper bound for the depicted range of P. MF is slightly outperformed by AF because

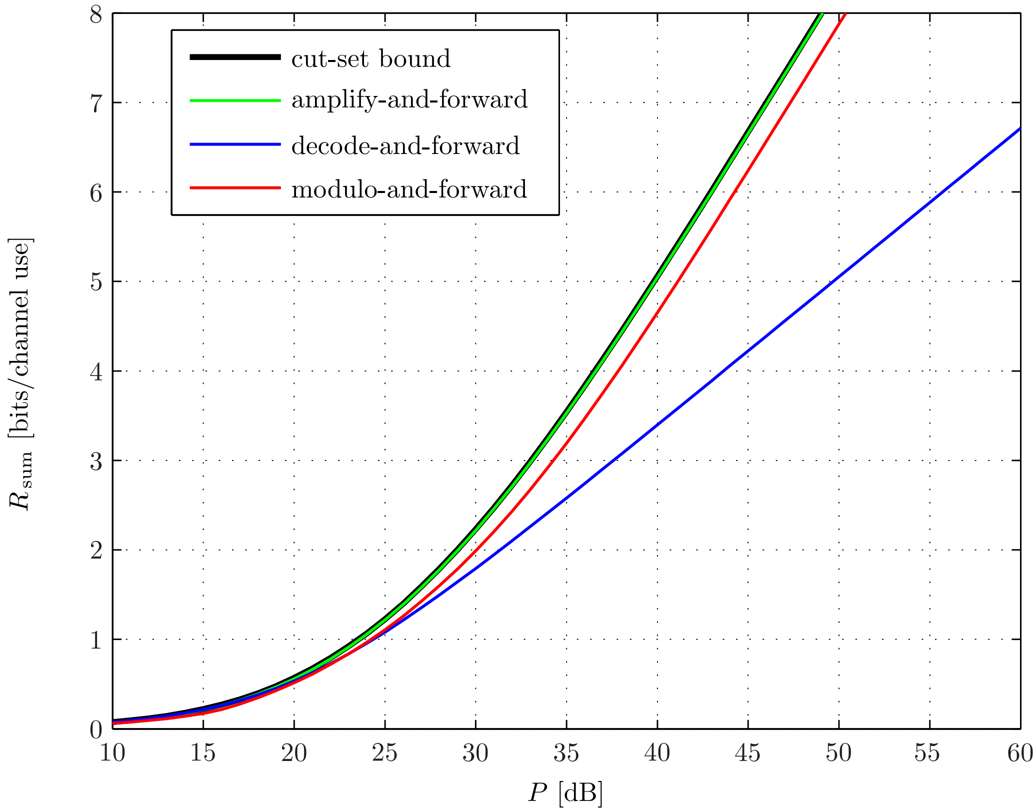


Figure 4.3.: Achievable sum rates for $h_{ar} = -20$ dB, $h_{br} = -30$ dB, $h_{ra} = 0$ dB, and $h_{rb} = 0$ dB.

the uplink channel gains are chosen to be unequal and superposition coding is used. For the case where the uplink gains are equal (e.g. $h_{ar} = -20$ dB and $h_{br} = -20$ dB) the performance of AF and MF becomes almost identical (not shown here). It can also be seen that both schemes outperform DF at high SNR. The gap to the upper bound for DF becomes more pronounced when the node power P is increased.

4.2.2. Strong Uplink

Now we consider the case where the uplink gains are strong compared to the downlink gains, e.g. $h_{ar} = 0$ dB, $h_{br} = 0$ dB, $h_{ra} = -15$ dB, and $h_{rb} = -25$ dB. The achievable sum rates for this channel configuration are plotted in Figure 4.4. It can be seen that for this case, DF performs very close to the upper bound is in fact optimal for $P < 43$ dB. This can be explained by the fact that in this scenario the sum rate constraint for the DF scheme (cp. equation (2.16)) is not active and therefore does not limit the achievable sum rates. Then, it is in fact optimal to fully decode both messages at the relay and no

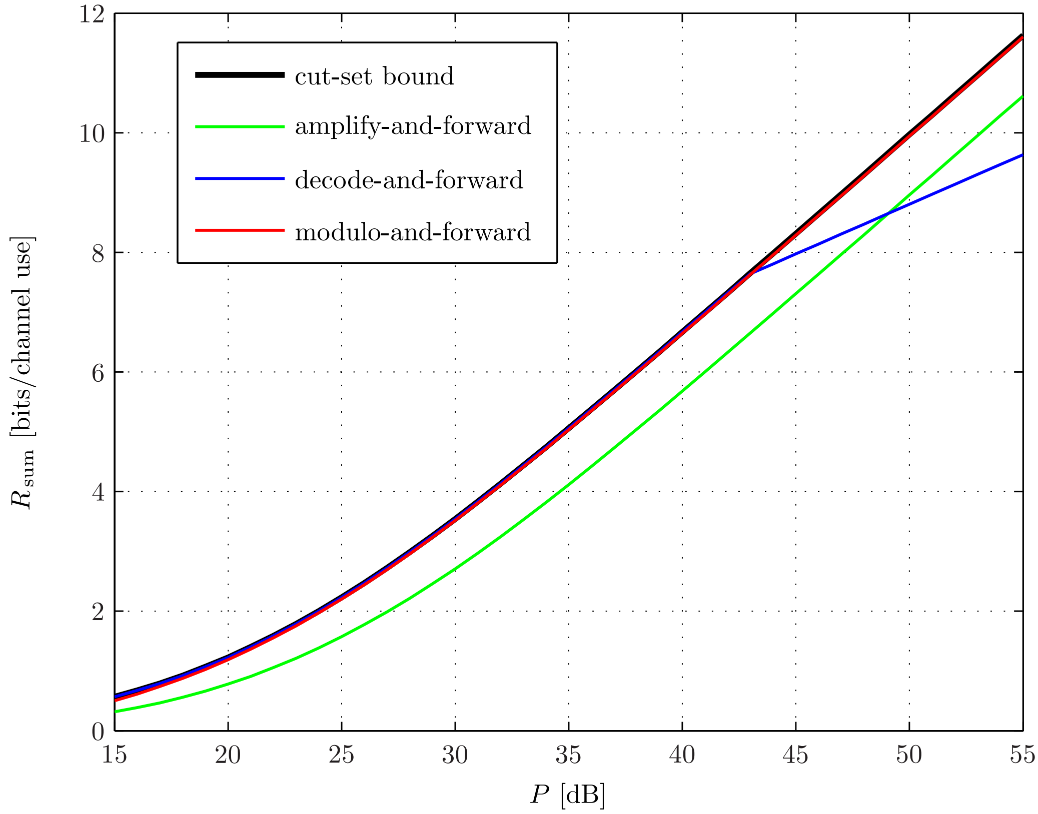


Figure 4.4.: Achievable sum rates for $h_{ar} = 0$ dB, $h_{br} = 0$ dB, $h_{ra} = -15$ dB, and $h_{rb} = -25$ dB.

multiplexing loss occurs. However, the slope changes for $P > 43$ dB and the DF strategy gets outperformed by the other two strategies for increasing node power. It can also be seen that MF also performs very close to the upper bound and outperforms AF for the depicted range of P in this scenario. Note that the channel gains in the uplink are equal and therefore no superposition coding is used here (i.e. $\theta = 1$) for MF, i.e. both users only use a nested lattice code.

4.2.3. Reciprocal Gains

Finally, we consider the case where the channel gains are given by $h_{ar} = 0$ dB, $h_{br} = -10$ dB, $h_{ra} = 0$ dB, and $h_{rb} = -10$ dB, i.e. the channel gains are reciprocal. The achievable sum rates are plotted in Figure 4.5. It can be seen that for $P < 20$ dB the DF scheme outperforms the two other schemes. However, for $P > 20$ dB the slope of the DF scheme changes because the sum rate constraint becomes active. Therefore the performance gap of the DF to the cut-set bound increases as P increases. For $P > 22$ dB

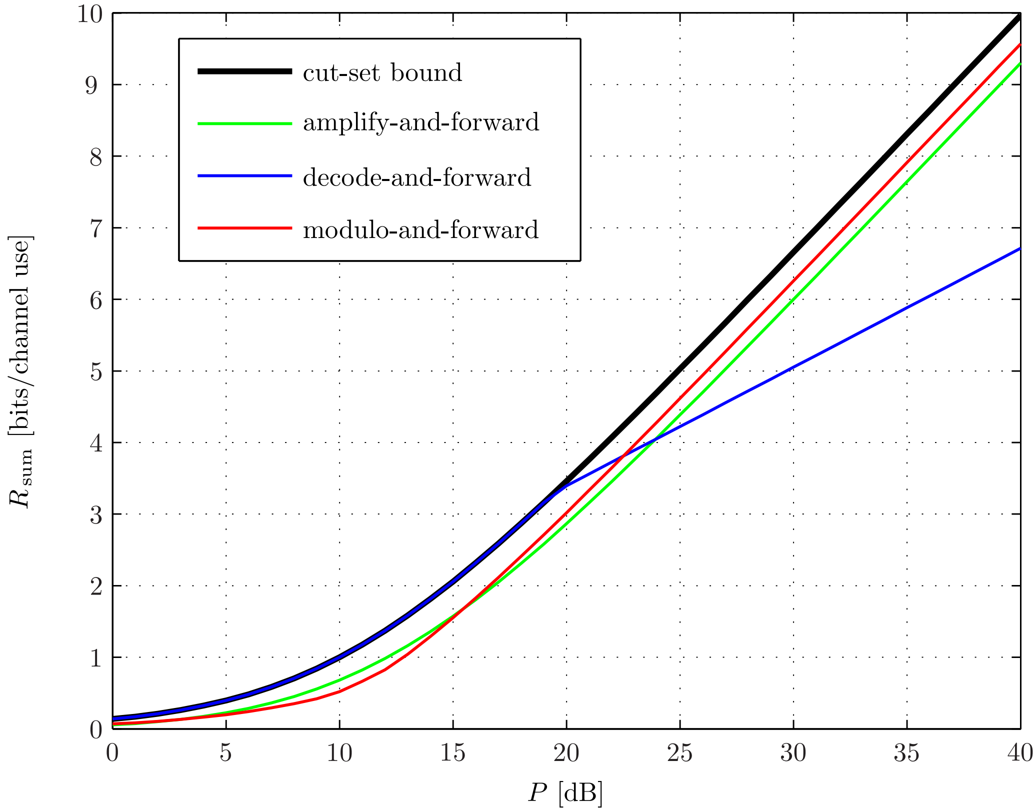


Figure 4.5.: Achievable sum rates for $h_{ar} = 0$ dB, $h_{br} = -10$ dB, $h_{ra} = 0$ dB, and $h_{rb} = -10$ dB.

the MF scheme achieves the highest sum rate among the considered strategies, followed by the AF scheme. In this case, none of the considered strategies is optimal (or approaches optimality) for $P > 20$ dB.

4.3. Discussion

The above comparison reveals that it is rather difficult to make generalized statements about the achievable sum rate performance of the different strategies. In general, the performance is highly dependent on the realization of the channel gains and also on the node power P. In fact, for fixed channel gains the sum rate performance can change significantly with increasing node power P and different schemes perform better for different SNR regimes. It is however possible to identify a few situations where some schemes are strictly better than others. For example the DF scheme is always optimal if the sum rate constraint is already implied by the individual rate constraints for each user. This may be the case when the uplink is very strong compared to the downlink. For this scenario, MF also performs very close to the upper bound. The AF strategy shows good performance for the case when the uplink is weak compared to the downlink.

A more exhaustive comparison of different strategies for the Gaussian TRC (including AF and DF, but not including MF) can be found in [40]. In addition to sum rates the authors also plot the rate regions that are achievable with the different strategies for various channel conditions. However, the authors also come to the conclusion that in general different schemes are optimal (or approach optimality) for different channel configurations and that “the rate regions that are achievable are not subsets of one another” [40].

CHAPTER 5

The Separated Two-Way Two-Relay Channel

In this chapter the network is modified by adding a second relay, i.e. two relays r_1 and r_2 assist in the message exchange between the two users a and b. It is assumed that there exists no direct communication link between the users. Similarly, there is no link between user a and the second relay r_2, nor is there a link between the first relay r_1 and user b. Hence, both relays are necessary in order to enable the message exchange. The resulting setup can be seen as a multi-hop extension of the sTRC and consequently this network is termed the (fully) separated two-way two-relay channel (sTTRC). See Figure 5.1 for an illustration of the network topology, where W_a and W_b are the user messages, $\hat{W}_a$ and $\hat{W}_b$ are the estimates, and (X_j, Y_j) are the input and output variables associated with each device node $j \in \{a, r_1, r_2, b\}$. Note that both relays r_1 and r_2 do not have any message to transmit and are not necessarily required to decode any of the user messages (i.e. no source or sink nodes are attached to the relay device nodes). In the following, FD nodes are considered and we begin by assuming a noiseless, finite field physical layer

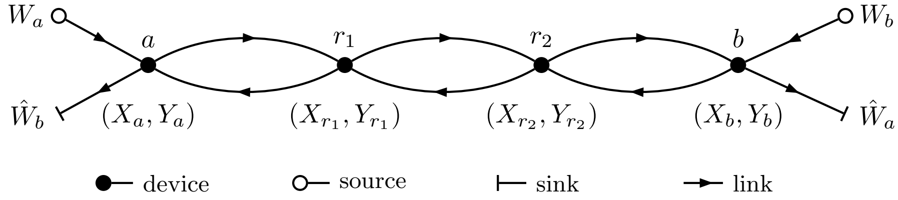


Figure 5.1.: Illustration of the network topology for the sTTRC.

in order to introduce the proposed block transmission strategy. All channel input and output variables are binary and interference at the relay nodes is modeled as the modulo-2 sum of two binary signals. For this model, two strategies are described where the users can simultaneously transmit their messages error-free at any rate up to 1 bit per channel use. After that, (binary) noise is included in the finite field model. It is shown that the cut-set region of the resulting network is achievable by using nested linear codes. Thus, for binary symmetric channels (BSCs) the capacity region of the sTTRC is known and coincides with the cut-set region. This is similar to the result that has been shown for the binary symmetric case of the sTRC [5, 39]. Finally, an achievable rate region for the Gaussian case under certain channel conditions is derived. In particular, we assume that all transmitted signals are average power constrained to P, the channel gains are 1, but all noise variances can be different. This channel configuration in combination with the proposed block transmission protocol leads to a new variant of the "broadcasting with user side-information" problem. In particular, in each transmission block the relay r_1 broadcasts two messages[1] – one low-rate message destined for user a and one high-rate message destined for the second relay r_2. The new element is that, at the same time, user b also transmits to r_2. Therefore it is required that the broadcast signal of r_1 is suitable for PNC, such that r_2 can decode a function of the messages transmitted by r_1 and b. The employed coding scheme is based on a lattice partition chain. It is shown that the achievable rate region derived with this strategy is within 1/2 bit per dimension of the capacity region for each user. Moreover, the gap vanishes for fixed noise variances and increasing node power, i.e. the scheme closely approaches the capacity region at high SNR.

5.1. Finite Field Physical Layer

Inspired by [16], at first a finite field physical layer is considered in order to become familiar with the network as well as to introduce the proposed block transmission strategy. In the following, it is assumed that all nodes have FD capability, i.e. a device node can receive and transmit at the same time. The definitions of messages, rates, encoding and decoding functions, error probability, and the capacity region for the sTRC in chapter 2 directly apply to this network as well. It is assumed that all input and output variables are binary and take on values in $\{0, 1\}$. Interference at the relay nodes is modeled as the modulo-2 sum of two incoming signals. First, the noiseless case is described and then binary noise is added (modulo 2) to each output variable, which can be seen as a binary symmetric channel model.

[1] More precisely, functions of messages.

5.1.1. Noiseless Case

In this subsection we assume that the channel is noiseless and represented by the following set of equations:

$$Y_a = X_{r_1} \tag{5.1a}$$
$$Y_{r_1} = X_a \oplus X_{r_2} \tag{5.1b}$$
$$Y_{r_2} = X_{r_1} \oplus X_b \tag{5.1c}$$
$$Y_b = X_{r_2}. \tag{5.1d}$$

The channel is memoryless and the ith channel use or transmission block[2] is denoted by $X_a^{(i)}$ for the input variable of user a and similarly for the other variables. User a wants to transmit the message vector $\boldsymbol{W}_a = (u_a^{(1)}, u_a^{(2)}, \ldots, u_a^{(M)}) \in \{0,1\}^M$ to user b, where we call the individual bits u "information bits" or "packets". Similarly user b wants to transmit the message vector $\boldsymbol{W}_b = (u_b^{(1)}, u_b^{(2)}, \ldots, u_b^{(M)}) \in \{0,1\}^M$ to user a. As we are going to use the channel $M+2$ times in order to exchange these bits, the transmission rate for both users is given by $R_a = R_b = M/(M+2)$ in bits per channel use[3].
One possible strategy for exchanging these bits would be as follows. The users transmit

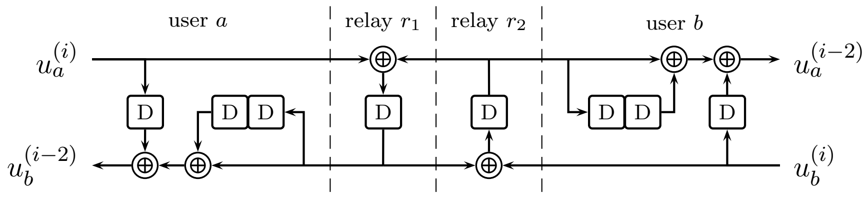
$$X_a^{(i)} = u_a^{(i)} \tag{5.2}$$
$$X_b^{(i)} = u_b^{(i)} \tag{5.3}$$

at time i. The relays broadcast the received signal given by (5.1b) and (5.1c) in the next channel use according to

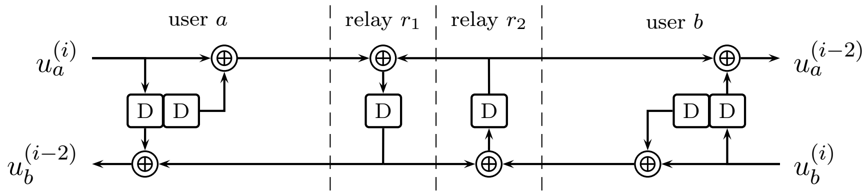
$$X_{r_1}^{(i)} = Y_{r_1}^{(i-1)} \tag{5.4}$$
$$X_{r_2}^{(i)} = Y_{r_2}^{(i-1)}, \tag{5.5}$$

where $X_{r_1}^{(1)} \stackrel{\text{def.}}{=} 0$ and $X_{r_2}^{(1)} \stackrel{\text{def.}}{=} 0$. What must be shown is that this is a solution, in the sense that each user is able to recover the bits of the other user based on the received channel outputs and the own transmitted bits. The received signal of user a in the ith

[2]For the noiseless case each channel use can be seen as a transmission block because no coding is required.
[3]Smaller or unequal user rates may be achieved through zero-padding the message vector.



(a) strategy 1



(b) strategy 2

Figure 5.2.: Block diagrams for the noiseless finite field physical layer model of the sTTRC.

channel use is given by

$$\begin{aligned} Y_a^{(i)} &= X_{r_1}^{(i)} = Y_{r_1}^{(i-1)} = X_a^{(i-1)} \oplus X_{r_2}^{(i-1)} \\ &= u_a^{(i-1)} \oplus Y_{r_2}^{(i-2)} \\ &= u_a^{(i-1)} \oplus X_{r_1}^{(i-2)} \oplus X_b^{(i-2)} \\ &= u_a^{(i-1)} \oplus X_{r_1}^{(i-2)} \oplus u_b^{(i-2)} \\ &= \underbrace{u_a^{(i-1)} \oplus Y_a^{(i-2)}}_{\text{known}} \oplus u_b^{(i-2)}. \end{aligned} \tag{5.6}$$

It can be seen that all parts of the received signal are known to user a as SI except the packet $u_b^{(i-2)}$ which can therefore be extracted in the ith channel use. The same follows for user b. A block diagram visualizing the strategy is depicted in Figure 5.2 (a). Note that the operators are modulo-2 additions and a delay element D delays the corresponding input signal by one block.

However, the above strategy is not unique and in the following we describe another block transmission strategy, where the users transmit the modulo-2 sum of two message bits. This is similar to the concept of transmitting "anti-packets" [23] in order to cancel packets at the physical layer that "flow" in the wrong direction due to the broadcast nature of the network. However, in our case these anti-packets do not increase the throughput of the network, but lead to simple closed form expressions of the received signals at the relay nodes and therefore simplify the description. In particular the users transmit

$$X_a^{(i)} = u_a^{(i)} \oplus u_a^{(i-2)} \tag{5.7}$$

$$X_b^{(i)} = u_b^{(i)} \oplus u_b^{(i-2)}, \tag{5.8}$$

where $u_a^{(i)} = 0$ and $u_b^{(i)} = 0$ for $i < 1$, and the relays broadcast the received signal in the next channel use according to

$$X_{r_1}^{(i)} = Y_{r_1}^{(i-1)} \tag{5.9}$$

$$X_{r_2}^{(i)} = Y_{r_2}^{(i-1)}, \tag{5.10}$$

where $X_{r_1}^{(1)} \overset{\text{def.}}{=} 0$ and $X_{r_2}^{(1)} \overset{\text{def.}}{=} 0$ as before. For this strategy, the received signals at the two relays are given by

$$Y_{r_1}^{(i)} = u_a^{(i)} \oplus u_b^{(i-1)} \tag{5.11}$$

$$Y_{r_2}^{(i)} = u_a^{(i-1)} \oplus u_b^{(i)}, \tag{5.12}$$

which can be shown by induction over i and will be described in the following for the first relay r_1. For $i = 1$ we have $Y_{r_1}^{(1)} = X_a^{(1)} \oplus X_{r_2}^{(1)} = u_a^{(1)}$, because $X_{r_2}^{(1)} = 0$ and $u_a^{(-1)} = 0$. For $i = 2$ we have $Y_{r_1}^{(2)} = X_a^{(2)} \oplus X_{r_2}^{(2)} = u_a^{(2)} \oplus u_b^{(1)}$, because $X_{r_1}^{(1)} = 0$, $u_a^{(0)} = 0$, and $u_b^{(-1)} = 0$. Finally, for $i = l + 1$ we have

$$\begin{aligned} Y_{r_1}^{(l+1)} &= X_a^{(l+1)} \oplus X_{r_2}^{(l+1)} \\ &= u_a^{(l+1)} \oplus u_a^{(l-1)} \oplus Y_{r_2}^{(l)} \\ &= u_a^{(l+1)} \oplus u_a^{(l-1)} \oplus X_{r_1}^{(l)} \oplus X_b^{(l)} \\ &= u_a^{(l+1)} \oplus u_a^{(l-1)} \oplus Y_{r_1}^{(l-1)} \oplus u_b^{(l)} \oplus u_b^{(l-2)} \\ &= u_a^{(l+1)} \oplus u_b^{(l)}, \end{aligned} \tag{5.13}$$

where the last equation follows by inductive assumption (cp. equation (5.11)) and thus the claim is established for r_1. The steps for the second relay r_2 are similar. With this strategy, a user can extract the information bit of the other user by computing $Y_a^{(i)} \oplus u_a^{(i-1)}$ and $Y_b^{(i)} \oplus u_b^{(i-1)}$ respectively. The strategy is visualized as a block diagram in Figure 5.2 (b). It can be seen that both strategies allow the users to successfully extract the information bits of the other user and thus a rate of up to 1 bit per channel use can be approached simultaneously for each user for $M \to \infty$. Moreover, from the block diagrams we obtain the following picture: After two initial blocks of transmission the network reaches a *steady state* where both information flows pass through the relays without influencing each other. Effectively, it looks as though user a directly sees the packet stream of user b (delayed by two channel uses) and vice versa.

5.1.2. Binary Symmetric Channel

Now we consider the case where binary noise is added (modulo 2) at each receiving node. Thus the channel is modeled by the following equations

$$Y_a = X_{r_1} \oplus Z_a \tag{5.14a}$$

$$Y_{r_1} = X_a \oplus X_{r_2} \oplus Z_{r_1} \tag{5.14b}$$

$$Y_{r_2} = X_{r_1} \oplus X_b \oplus Z_{r_2} \tag{5.14c}$$

$$Y_b = X_{r_2} \oplus Z_a, \tag{5.14d}$$

where $Z_a, Z_{r_1}, Z_{r_2}, Z_a$ are independent binary random variables and the crossover probabilities $\Pr(\{Z_a = 1\})$, $\Pr(\{Z_{r_1} = 1\})$, $\Pr(\{Z_{r_2} = 1\})$, and $\Pr(\{Z_b = 1\})$ are denoted by $\varepsilon_a, \varepsilon_{r_1}, \varepsilon_{r_2}$, and ε_b respectively. The cut-set region for this (binary symmetric) channel is given by

$$\mathcal{R}_{cut}^{\text{BSC}} = \left\{ (R_a, R_b) \ : \ \begin{array}{l} 0 < R_a < \min\left(1 - \text{H}(\varepsilon_{r_1}), 1 - \text{H}(\varepsilon_{r_2}), 1 - \text{H}(\varepsilon_b)\right) \\ 0 < R_b < \min\left(1 - \text{H}(\varepsilon_{r_2}), 1 - \text{H}(\varepsilon_{r_1}), 1 - \text{H}(\varepsilon_a)\right) \end{array} \right\}, \tag{5.15}$$

where $\text{H}(\,\cdot\,)$ is the binary entropy function [9]. In order to show achievability of (5.15), consider the following two linear binary codes $\mathcal{C}_a$ and $\mathcal{C}_b$ of dimension n:

$$\mathcal{C}_a = \left\{ \boldsymbol{c} = \boldsymbol{u}\boldsymbol{G} = (\boldsymbol{u}', \boldsymbol{u}'') \begin{pmatrix} \boldsymbol{G}' \\ \boldsymbol{G}'' \end{pmatrix} = \boldsymbol{u}'\boldsymbol{G}' \oplus \boldsymbol{u}''\boldsymbol{G}'' \ : \ \boldsymbol{u} \in \{0,1\}^{nR_a} \right\} \tag{5.16}$$

and

$$\mathcal{C}_b = \left\{ \boldsymbol{c} = \boldsymbol{u}\boldsymbol{G} = (\boldsymbol{u}', \boldsymbol{0}) \begin{pmatrix} \boldsymbol{G}' \\ \boldsymbol{G}'' \end{pmatrix} = \boldsymbol{u}'\boldsymbol{G}' \ : \ \boldsymbol{u}' \in \{0,1\}^{nR_b} \right\}, \tag{5.17}$$

where $\boldsymbol{u}''$ has length $n(R_a - R_b)$ and $\boldsymbol{G}$ is a generator matrix[4] of dimension $nR_a \times n$ which can also be written as a "stacked" generator matrix using $\boldsymbol{G}'$ and $\boldsymbol{G}''$ whose dimensions are $nR_b \times n$ and $n(R_a - R_b) \times n$. We use the same variables in (5.16) and (5.17) to make the connection between the codes more apparent. From the code construction the following observations can be made:

- $(\mathcal{C}_a, \mathcal{C}_b)$ are nested codes $\mathcal{C}_b \subseteq \mathcal{C}_a$, i.e. each codeword in $\mathcal{C}_b$ is also in $\mathcal{C}_a$. This means that $\mathcal{C}_a$ and $\mathcal{C}_b$ share the same generating subspace [5]. Moreover, it follows that the (componentwise) modulo-2 sum of any codeword in $\mathcal{C}_a$ and any codeword in $\mathcal{C}_b$ is itself a valid codeword in $\mathcal{C}_a$.

[4] It is also possible to characterize these codes with parity-check matrices, see [41].

- In the context of communication networks, $\mathcal{C}_a$ is a code with "multiple interpretations" [42]. The idea here is that the code $\mathcal{C}_a$ can have different effective rates (in this case either R_a or R_b) based on the SI that is available at a decoding receiver. In particular, a receiver that knows the information sequence corresponding to $\boldsymbol{u}'$ as SI is able to decode with respect to $\mathcal{C}_b$ *and* $\mathcal{C}_a$. A receiver with no SI can only decode with respect to $\mathcal{C}_a$.

We now divide the total number of channel uses into $M+2$ successive blocks of n channel uses and use an upper index to refer to each block, i.e. user a receives the signal $\boldsymbol{Y}_a^{(i)} = \boldsymbol{X}_{r_1}^{(i)} \oplus \boldsymbol{Z}_a^{(i)}$ of length n in the ith block and so on.

Encoding at the Users

User a transmits M binary information vectors $\boldsymbol{u}_a^{(1)}, \ldots, \boldsymbol{u}_a^{(M)}$, where each vector has length nR_a, i.e. in total MnR_a bits are transmitted to user b. Similarly user b transmits M binary information vectors $\boldsymbol{u}_b'^{(1)}, \ldots, \boldsymbol{u}_b'^{(M)}$, where each vector has length nR_b, i.e. in total MnR_b bits are transmitted to user a. It is assumed without loss of generality that $R_b \leq R_a$ and the zero-padded information vectors of user b (to length nR_a) are denoted by $\boldsymbol{u}_b^{(1)}, \ldots, \boldsymbol{u}_b^{(M)}$. The effective rates are slightly lower than R_a and R_b because the whole message exchange takes $n(M+2)$ channel uses. However, by choosing M large enough, the rates R_a and R_b can be approached as closely as desired.
User a maps the information vectors in each transmission block to codewords in $\mathcal{C}_a$ and transmits

$$\boldsymbol{X}_a^{(i)} = \boldsymbol{c}_a^{(i)} \oplus \boldsymbol{c}_a^{(i-2)} = \boldsymbol{u}_a^{(i)}\boldsymbol{G} \oplus \boldsymbol{u}_a^{(i-2)}\boldsymbol{G} = \left(\boldsymbol{u}_a^{(i)} \oplus \boldsymbol{u}_a^{(i-2)}\right)\boldsymbol{G}, \tag{5.18}$$

where $\boldsymbol{u}_a^{(i)} = \boldsymbol{0}$ for $i < 1$. It can be seen that $\boldsymbol{X}_a^{(i)} \in \mathcal{C}_a$, i.e. the transmitted signal is itself a codeword. Moreover, it can be seen that from the distributive property of matrix multiplication it follows that it is equivalent to compute the componentwise XOR of two codewords or to perform the componentwise XOR of the corresponding information vectors [42], i.e. user a effectively performs NC.
Similarly, user b maps the information vectors in each block to codewords by using the code $\mathcal{C}_b$ and transmits

$$\boldsymbol{X}_b^{(i)} = \boldsymbol{c}_b^{(i)} \oplus \boldsymbol{c}_b^{(i-2)} = \boldsymbol{u}_b'^{(i)}\boldsymbol{G}' \oplus \boldsymbol{u}_b'^{(i-2)}\boldsymbol{G}' = \left(\boldsymbol{u}_b'^{(i)} \oplus \boldsymbol{u}_b'^{(i-2)}\right)\boldsymbol{G}', \tag{5.19}$$

where $\boldsymbol{u}_b'^{(i)} = \boldsymbol{0}$ for $i < 1$. Note that $\boldsymbol{X}_b^{(i)} \in \mathcal{C}_b$ and moreover, because $\mathcal{C}_b$ is nested in $\mathcal{C}_a$, the transmitted signal of user b is also a valid codeword in $\mathcal{C}_a$ and can also be written as

$$\boldsymbol{X}_b^{(i)} = \left(\boldsymbol{u}_b{}^{(i)} \oplus \boldsymbol{u}_b{}^{(i-2)}\right)\boldsymbol{G}. \tag{5.20}$$

From this encoding process it follows that the componentwise XOR of two transmitted signals (independent of any particular transmission block) is a valid codeword in $\mathcal{C}_a$.

Relaying Strategy

Both relays r_1 and r_2 perform ML decoding in each block with respect to the code $\mathcal{C}_a$. For a binary symmetric channel, ML decoding amounts to finding the codeword that is closest to the received vector in terms of Hamming-distance [43]. Also note that for a linear code, the decoding decision only depends on the noise and not on the codeword. Therefore, ML decoding is equivalent to finding the ML estimate of the noise vector [43]. Since the relays are not interested in the particular messages of the users (or a linear combination thereof), the relays don't reverse the encoding operation, but rather compute the componentwise XOR of the ML estimate of the noise and the received signal. This can also be regarded as a quantization operation with respect to $\mathcal{C}_a$ and will be denoted by $Q_{\mathcal{C}_a}(\,\cdot\,)$, analogous to the nearest neighbor lattice quantizer. The relays then simply broadcast the "quantized" signal in the next block, i.e.

- the relay r_1 broadcasts $\boldsymbol{X}_{r_1}^{(i)} = Q_{\mathcal{C}_a}(\boldsymbol{Y}_{r_1}^{(i-1)}) = \boldsymbol{Y}_{r_1}^{(i-1)} \oplus \hat{\boldsymbol{Z}}_{r_1}^{(i-1)}$ and
- the relay r_2 broadcasts $\boldsymbol{X}_{r_2}^{(i)} = Q_{\mathcal{C}_a}(\boldsymbol{Y}_{r_2}^{(i-1)}) = \boldsymbol{Y}_{r_2}^{(i-1)} \oplus \hat{\boldsymbol{Z}}_{r_2}^{(i-1)}$,

where $\hat{\boldsymbol{Z}}_{r_1}$ and $\hat{\boldsymbol{Z}}_{r_2}$ denote the (ML) estimate of the binary noise vector for r_1 and r_2, respectively. Note that if all intermediate decoding steps in each block are successful, i.e. $\hat{\boldsymbol{Z}}_{r_1}^{(i)} = \boldsymbol{Z}_{r_1}^{(i)}$ and $\hat{\boldsymbol{Z}}_{r_2}^{(i)} = \boldsymbol{Z}_{r_2}^{(i)}$ holds for all i, we have

$$Q_{\mathcal{C}_a}(\boldsymbol{Y}_{r_1}^{(i)}) = \boldsymbol{c}_a^{(i)} \oplus \boldsymbol{c}_b^{(i-1)} \tag{5.21}$$
$$Q_{\mathcal{C}_a}(\boldsymbol{Y}_{r_2}^{(i)}) = \boldsymbol{c}_a^{(i-1)} \oplus \boldsymbol{c}_b^{(i)}, \tag{5.22}$$

which can be shown by induction over i in a similar way as was done for the noiseless case. This means that each relay broadcasts the componentwise XOR of two user codewords, which is itself a codeword in $\mathcal{C}_a$, provided that all preceding decoding steps are successful.

Decoding at the Users

Assuming that the relays perform error-free decoding, in the ith transmission block the users receive

$$\boldsymbol{Y}_a^{(i)} = \boldsymbol{X}_{r_1}^{(i)} \oplus \boldsymbol{Z}_a^{(i)} = Q_{\mathcal{C}_a}(\boldsymbol{Y}_{r_1}^{(i-1)}) \oplus \boldsymbol{Z}_a^{(i)} = \boldsymbol{c}_a^{(i-1)} \oplus \boldsymbol{c}_b^{(i-2)} \oplus \boldsymbol{Z}_a^{(i)} \tag{5.23}$$
$$\boldsymbol{Y}_b^{(i)} = \boldsymbol{X}_{r_2}^{(i)} \oplus \boldsymbol{Z}_b^{(i)} = Q_{\mathcal{C}_a}(\boldsymbol{Y}_{r_2}^{(i-1)}) \oplus \boldsymbol{Z}_b^{(i)} = \boldsymbol{c}_a^{(i-2)} \oplus \boldsymbol{c}_b^{(i-1)} \oplus \boldsymbol{Z}_b^{(i)} \tag{5.24}$$

and compute

$$Q_{\mathcal{C}_b}(\boldsymbol{Y}_a^{(i)} \oplus \boldsymbol{c}_a^{(i-1)}) = Q_{\mathcal{C}_b}(\boldsymbol{c}_b^{(i-2)} \oplus \boldsymbol{Z}_a^{(i)}) \tag{5.25}$$

$$Q_{\mathcal{C}_a}(\boldsymbol{Y}_b^{(i)} \oplus \boldsymbol{c}_b^{(i-1)}) = Q_{\mathcal{C}_a}(\boldsymbol{c}_a^{(i-2)} \oplus \boldsymbol{Z}_a^{(i)}) \tag{5.26}$$

in order to recover the codeword (and therefore the information vector) of the respective other user. Note that user a is able to decode with respect to the code $\mathcal{C}_b$ in equation (5.25) because there is sufficient SI available. It is crucial to observe that user b could compute $Q_{\mathcal{C}_a}(\boldsymbol{Y}_b^{(i)})$ and extract the codeword $\boldsymbol{c}_a^{(i-2)}$ from the decoded linear superposition of two codewords, i.e. from $\boldsymbol{c}_a^{(i-2)} \oplus \boldsymbol{c}_b^{(i-1)}$, while this is not the case for user a. Computing $Q_{\mathcal{C}_b}(\boldsymbol{Y}_a^{(i)})$ would fail because $\boldsymbol{Y}_a^{(i)}$ is a "noisy" codeword in $\mathcal{C}_a$ but not necessarily $\mathcal{C}_b$. However, computing $Q_{\mathcal{C}_a}(\boldsymbol{Y}_a^{(i)})$ would work, but does not lead to the desired rate region, because this would put a rate constraint on R_a and not on R_b.

Achievability of the Cut-set Region

The overall probability of error p_e (for the message exchange) can be bounded by the sum over the decoding error probabilities in each block for all devices. Each relay decodes (or quantizes) $M + 2$ times and each user M times. Therefore

$$p_e \leq (4M + 2) \Pr(\{\hat{\boldsymbol{Z}} \neq \boldsymbol{Z}\}) \tag{5.27}$$

where $\Pr(\{\hat{\boldsymbol{Z}} \neq \boldsymbol{Z}\})$ denotes the probability that any of the decoding steps fails[5]. Note that because of the symmetry of linear codes, the error probability does not depend on the particular codeword that is sent. Achievability of the cut-set region in the sense of vanishing p_e follows from the fact that there exist nested linear codes $(\mathcal{C}_a, \mathcal{C}_b)$ such that each code is a good channel code for a BSC [41], i.e. $\Pr(\{\hat{\boldsymbol{Z}} \neq \boldsymbol{Z}\})$ can be made as small as desired by increasing n. Note that both relays and user b decode with respect to $\mathcal{C}_a$ and therefore the probability of decoding error in each block for these nodes can be made arbitrarily small as long as

$$R_a < \min\left(1 - \mathrm{H}(\varepsilon_{r_1}), 1 - \mathrm{H}(\varepsilon_{r_2}), 1 - \mathrm{H}(\varepsilon_b)\right). \tag{5.28}$$

User a decodes with respect to $\mathcal{C}_b$ and therefore the probability of decoding error in each block for user a can be made arbitrarily small as long as

$$R_b < 1 - \mathrm{H}(\varepsilon_a). \tag{5.29}$$

[5]If the decoding error probability is different for the two codes, then $\Pr(\{\hat{\boldsymbol{Z}} \neq \boldsymbol{Z}\})$ is the probability of decoding error for the worse code and serves as an upper bound on the decoding error probability for both codes.

Recall that $R_b \leq R_a$ and therefore the inequality (5.28) also holds for R_b if it holds for R_a. It can then be seen that the achievable rate region implied by (5.29) and (5.28) is identical to the cut-set region (5.15) if $1 - \mathrm{H}(\varepsilon_b) \geq 1 - \mathrm{H}(\varepsilon_a)$, and for $1 - \mathrm{H}(\varepsilon_b) < 1 - \mathrm{H}(\varepsilon_a)$ we can just relabel the users.

5.2. Gaussian Channel

In the preceding section we described a block transmission strategy and derived the capacity region of the sTTRC assuming a (binary symmetric) finite field physical layer. In this section the methods and insights gained from this analysis are applied to the Gaussian case of the sTTRC. The notation and most of the definitions regarding the sTRC introduced in chapter 2 are also valid for this network and only the differences are pointed out here. The channel is modeled by the following equations

$$Y_a = X_{r_1} + Z_a \tag{5.30a}$$

$$Y_{r_1} = X_a + X_{r_2} + Z_{r_1} \tag{5.30b}$$

$$Y_{r_2} = X_{r_1} + X_b + Z_{r_2} \tag{5.30c}$$

$$Y_b = X_{r_2} + Z_a, \tag{5.30d}$$

where Z_a, Z_{r_1}, Z_{r_2} and Z_b are assumed to be i.i.d. Gaussian random variables with zero mean and variance σ_a^2, $\sigma_{r_1}^2$, $\sigma_{r_2}^2$ and σ_b^2 respectively. All transmit signals are average power constrained to P (cp. equation (2.2)). Note that this is equivalent to the case where all noise variables have unit variance and the input signals are scaled by channel coefficients h_{r_1a}, h_{ar_1}, $h_{r_2r_1}$, $h_{r_1r_2}$, h_{br_2}, and h_{r_2b} with the additional constraint that $h_{ar_1} = h_{r_2r_1}$ and $h_{r_1r_2} = h_{br_2}$. However, in this section the notation without channel coefficients[6] is used for convenience. Also note that this is not the general case of the Gaussian sTTRC where unequal power constraints at the device nodes have to be taken into account (or equivalently the equality constraints are removed when channel coefficients are used). In the last section of this chapter we propose two ideas that might be useful in order to fully extend the derived results to the general case.

5.2.1. Cut-set Bound

The cut-set region of the Gaussian sTTRC is given by

$$\mathcal{R}_{cut} = \left\{ (R_a, R_b) \ : \ \begin{array}{l} 0 < R_a < \min\left(\mathrm{C}\left(P/\sigma_{r_1}^2\right), \mathrm{C}\left(P/\sigma_{r_2}^2\right), \mathrm{C}\left(P/\sigma_b^2\right)\right) \\ 0 < R_b < \min\left(\mathrm{C}\left(P/\sigma_{r_2}^2\right), \mathrm{C}\left(P/\sigma_{r_1}^2\right), \mathrm{C}\left(P/\sigma_a^2\right)\right) \end{array} \right\}, \tag{5.31}$$

[6] More precisely, all channel coefficients are assumed to be 1.

where $\mathrm{C}(x) \stackrel{\text{def.}}{=} \log_2(1+x)/2$. This region follows from the cut-set bound [9] and applying similar arguments as presented in [14] for the sTRC.

5.2.2. An Achievable Rate Region

The main result of this section is given in form of the following theorem.

Theorem 2. *For the Gaussian sTTRC defined above, one can achieve all positive rate pairs* (R_a, R_b) *satisfying:*

$$0 < R_a < \min\left(\tilde{\mathrm{C}}\left(P/\sigma_{r_1}^2\right), \tilde{\mathrm{C}}\left(P/\sigma_{r_2}^2\right), \mathrm{C}\left(P/\sigma_b^2\right)\right) \tag{5.32}$$

$$0 < R_b < \min\left(\tilde{\mathrm{C}}\left(P/\sigma_{r_2}^2\right), \tilde{\mathrm{C}}\left(P/\sigma_{r_1}^2\right), \mathrm{C}\left(P/\sigma_a^2\right)\right), \tag{5.33}$$

where $\mathrm{C}(x) \stackrel{\text{def.}}{=} \log_2(1+x)/2$ *and* $\tilde{\mathrm{C}}(x) \stackrel{\text{def.}}{=} \log_2(1/2+x)/2$.

Note that the achievable rate region implied by (5.32) and (5.33) is within 1/2 bit per dimension of the cut-set region and thus the capacity region for each user. Moreover, the gap to the cut-set region approaches zero for fixed noise variances and increasing node power P.
In the remainder of this section we describe the transmission strategy that is used in order to prove the theorem. In particular, we will show that the Gaussian sTTRC defined above is conceptually very similar to the binary symmetric case by applying a so-called MLAN conversion [31]. The continuous counterpart of (binary) nested linear codes are nested Voronoi codes which are constructed using a lattice partition chain.

MLAN Conversion of the sTTRC

In the following, we describe how the Gaussian sTTRC can be converted to an MLAN channel. This conversion makes the similarities to the (binary symmetric) finite field physical layer apparent and also simplifies the analysis. The conversion is based on the description contained in [31] and [44].
We assume that the channel is used in $M+2$ consecutive blocks, where each block contains n channel uses. Λ is an n-dimensional lattice which is simultaneously good and the fundamental region of Λ has second moment $\sigma^2(\Lambda) = P$. The transmitted signals of the

nodes in the ith block are given by

$$\boldsymbol{X}_a^{(i)} = (\tilde{\boldsymbol{X}}_a^{(i)} + \boldsymbol{U}_a^{(i)}) \mod \Lambda \tag{5.34}$$

$$\boldsymbol{X}_{r_1}^{(i)} = (\tilde{\boldsymbol{X}}_{r_1}^{(i)} + \boldsymbol{U}_{r_1}^{(i)}) \mod \Lambda \tag{5.35}$$

$$\boldsymbol{X}_{r_2}^{(i)} = (\tilde{\boldsymbol{X}}_{r_2}^{(i)} + \boldsymbol{U}_{r_2}^{(i)}) \mod \Lambda \tag{5.36}$$

$$\boldsymbol{X}_b^{(i)} = (\tilde{\boldsymbol{X}}_b^{(i)} + \boldsymbol{U}_b^{(i)}) \mod \Lambda, \tag{5.37}$$

where $\tilde{\boldsymbol{X}}_a^{(i)}$, $\tilde{\boldsymbol{X}}_{r_1}^{(i)}$, $\tilde{\boldsymbol{X}}_{r_2}^{(i)}$, $\tilde{\boldsymbol{X}}_b^{(i)} \in \mathcal{R}_V(\Lambda)$ are the inputs signals to the MLAN channel and $\boldsymbol{U}_a^{(i)}$, $\boldsymbol{U}_{r_1}^{(i)}$, $\boldsymbol{U}_{r_2}^{(i)}$, $\boldsymbol{U}_b^{(i)}$ are random dither vectors. All dither vectors in each block are uniformly distributed over $\mathcal{R}_V(\Lambda)$ and known to all nodes. The crypto lemma (cp. section 3.1) ensures that the input signals to the MLAN channel and the transmitted signals of the nodes are statistically independent. Moreover, the transmitted signals are uniformly distributed over $\mathcal{R}_V(\Lambda)$ and therefore the power constraint at all nodes is met.
Upon reception the nodes compute

$$\tilde{\boldsymbol{Y}}_a^{(i)} = (\beta \boldsymbol{Y}_a^{(i)} - \boldsymbol{U}_{r_1}^{(i)}) \mod \Lambda \tag{5.38}$$

$$\tilde{\boldsymbol{Y}}_{r_1}^{(i)} = (\alpha \boldsymbol{Y}_{r_1}^{(i)} - \boldsymbol{U}_a^{(i)} - \boldsymbol{U}_{r_2}^{(i)}) \mod \Lambda \tag{5.39}$$

$$\tilde{\boldsymbol{Y}}_{r_2}^{(i)} = (\alpha \boldsymbol{Y}_{r_2}^{(i)} - \boldsymbol{U}_{r_1}^{(i)} - \boldsymbol{U}_b^{(i)}) \mod \Lambda \tag{5.40}$$

$$\tilde{\boldsymbol{Y}}_b^{(i)} = (\beta \boldsymbol{Y}_b^{(i)} - \boldsymbol{U}_{r_2}^{(i)}) \mod \Lambda, \tag{5.41}$$

where $\tilde{\boldsymbol{Y}}_a^{(i)}$, $\tilde{\boldsymbol{Y}}_{r_1}^{(i)}$, $\tilde{\boldsymbol{Y}}_{r_2}^{(i)}$, $\tilde{\boldsymbol{Y}}_b^{(i)}$ are the output signals of the MLAN channel in block i and β and α are scaling factors for the MMSE estimation. For user a the output signal of the MLAN channel can be written as

$$\begin{aligned}
\tilde{\boldsymbol{Y}}_a^{(i)} &= (\beta \boldsymbol{Y}_a^{(i)} - \boldsymbol{U}_{r_1}^{(i)}) \mod \Lambda \\
&= (\beta(\boldsymbol{X}_{r_1}^{(i)} + \boldsymbol{Z}_a^{(i)}) - \boldsymbol{U}_{r_1}^{(i)}) \mod \Lambda \\
&\overset{(5.35)}{=} ((\tilde{\boldsymbol{X}}_{r_1}^{(i)} + \boldsymbol{U}_{r_1}^{(i)}) \mod \Lambda - \boldsymbol{X}_{r_1}^{(i)} \\
&\quad + \beta(\boldsymbol{X}_{r_1}^{(i)} + \boldsymbol{Z}_a^{(i)}) - \boldsymbol{U}_{r_1}^{(i)}) \mod \Lambda \\
&= (\tilde{\boldsymbol{X}}_{r_1}^{(i)} + (\beta - 1)\boldsymbol{X}_{r_1}^{(i)} + \beta \boldsymbol{Z}_a^{(i)}) \mod \Lambda \\
&= (\tilde{\boldsymbol{X}}_{r_1}^{(i)} + \tilde{\boldsymbol{Z}}_a^{(i)}) \mod \Lambda,
\end{aligned} \tag{5.42}$$

where $\tilde{\boldsymbol{Z}}_a^{(i)}$ is the effective noise given by

$$\tilde{\boldsymbol{Z}}_a^{(i)} = (\beta - 1)\boldsymbol{X}_{r_1}^{(i)} + \beta \boldsymbol{Z}_a^{(i)}. \tag{5.43}$$

Recall that from the crypto lemma we have that $\boldsymbol{X}_{r_1}^{(i)}$ is independent of $\tilde{\boldsymbol{X}}_{r_1}^{(i)}$. Therefore $\tilde{\boldsymbol{Z}}_a^{(i)}$ is also independent of $\tilde{\boldsymbol{X}}_{r_1}^{(i)}$. Moreover, from the fact that $\boldsymbol{X}_{r_1}^{(i)}$ is uniform over

$\mathcal{R}_V(\Lambda)$ and Λ is Poltyrev-good we have that $\tilde{\boldsymbol{Z}}_a^{(i)}$ is Gaussian in the limit [34]. We choose $\beta = P/(P+\sigma_a^2)$ in order to minimize the effective noise power (see also Appendix A) and obtain

$$\tilde{\sigma}_a^2 = \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_a^{(i)}||^2\right] = \frac{P\sigma_a^2}{P+\sigma_a^2}. \tag{5.44}$$

The steps for user b are similar, yielding

$$\tilde{\boldsymbol{Y}}_b^{(i)} = (\tilde{\boldsymbol{X}}_{r_2}^{(i)} + \tilde{\boldsymbol{Z}}_b^{(i)}) \bmod \Lambda, \tag{5.45}$$

where $\boldsymbol{Z}_b^{(i)}$ is the effective noise at user b is independent of $\tilde{\boldsymbol{X}}_{r_2}^{(i)}$, Gaussian in the limit, and given by

$$\tilde{\boldsymbol{Z}}_b^{(i)} = (\beta-1)\boldsymbol{X}_{r_2}^{(i)} + \beta\boldsymbol{Z}_b^{(i)}. \tag{5.46}$$

The effective noise power is

$$\tilde{\sigma}_b^2 = \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_b^{(i)}||^2\right] = \frac{P\sigma_b^2}{P+\sigma_b^2}. \tag{5.47}$$

for the same choice of β as for user a. For the first relay r_1 the output vector of the MLAN channel $\tilde{\boldsymbol{Y}}_{r_1}^{(i)}$ can be written as

$$\begin{aligned}
\tilde{\boldsymbol{Y}}_{r_1}^{(i)} &= (\alpha\boldsymbol{Y}_{r_1}^{(i)} - \boldsymbol{U}_a^{(i)} - \boldsymbol{U}_{r_2}^{(i)}) \bmod \Lambda \\
&= (\alpha(\boldsymbol{X}_a^{(i)} + \boldsymbol{X}_{r_2}^{(i)} + \boldsymbol{Z}_{r_1}^{(i)}) - \boldsymbol{U}_a^{(i)} - \boldsymbol{U}_{r_2}^{(i)}) \bmod \Lambda \\
&\overset{(5.34)}{\underset{(5.36)}{=}} ((\tilde{\boldsymbol{X}}_a^{(i)} + \boldsymbol{U}_a^{(i)}) \bmod \Lambda - \boldsymbol{X}_a^{(i)} \\
&\quad + (\tilde{\boldsymbol{X}}_{r_2}^{(i)} + \boldsymbol{U}_{r_2}^{(i)}) \bmod \Lambda - \boldsymbol{X}_{r_2}^{(i)} \\
&\quad + \alpha(\boldsymbol{X}_a^{(i)} + \boldsymbol{X}_{r_2}^{(i)} + \boldsymbol{Z}_{r_1}^{(i)}) - \boldsymbol{U}_a^{(i)} - \boldsymbol{U}_{r_2}^{(i)}) \bmod \Lambda \\
&= (\tilde{\boldsymbol{X}}_a^{(i)} + \tilde{\boldsymbol{X}}_{r_2}^{(i)} + (\alpha-1)(\boldsymbol{X}_a^{(i)} + \boldsymbol{X}_{r_2}^{(i)}) + \alpha\boldsymbol{Z}_{r_1}^{(i)}) \bmod \Lambda \\
&= (\tilde{\boldsymbol{X}}_a^{(i)} + \tilde{\boldsymbol{X}}_{r_2}^{(i)} + \tilde{\boldsymbol{Z}}_{r_1}^{(i)}) \bmod \Lambda,
\end{aligned} \tag{5.48}$$

where $\tilde{\boldsymbol{Z}}_{r_1}$ is the effective noise given by

$$\tilde{\boldsymbol{Z}}_{r_1}^{(i)} = (\alpha-1)(\boldsymbol{X}_a^{(i)} + \boldsymbol{X}_{r_2}^{(i)}) + \alpha\boldsymbol{Z}_{r_1}^{(i)}. \tag{5.49}$$

Note that (5.48) reflects the interference at the relay r_1. It can be seen that because of this interference the effective noise term is different from the effective noise term at the users – in particular, we have two self-noise terms whereas at the users only one self-noise term is present. However, by the crypto lemma both self-noise terms are statistically independent of the input signals and therefore $\tilde{\boldsymbol{Z}}_{r_1}^{(i)}$ is also independent of $\tilde{\boldsymbol{X}}_a^{(i)}$ and $\tilde{\boldsymbol{X}}_{r_2}^{(i)}$. Here we choose $\alpha = 2P/(2P + \sigma_{r_1}^2)$ in order to minimize the effective noise power, which is then given by

$$\tilde{\sigma}_{r_1}^2 = \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_{r_1}||^2\right] = \frac{2P\sigma_{r_1}^2}{2P + \sigma_{r_1}^2}. \tag{5.50}$$

The steps for the second relay r_2 are similar, yielding

$$\tilde{\boldsymbol{Y}}_{r_2}^{(i)} = (\tilde{\boldsymbol{X}}_{r_1}^{(i)} + \tilde{\boldsymbol{X}}_b^{(i)} + \tilde{\boldsymbol{Z}}_{r_2}^{(i)}) \bmod \Lambda, \tag{5.51}$$

where the effective noise $\tilde{\boldsymbol{Z}}_{r_2}^{(i)}$ is independent of $\tilde{\boldsymbol{X}}_{r_1}^{(i)}$ and $\tilde{\boldsymbol{X}}_b^{(i)}$ and given by

$$\tilde{\boldsymbol{Z}}_{r_2}^{(i)} = (\alpha - 1)(\boldsymbol{X}_{r_1}^{(i)} + \boldsymbol{X}_b^{(i)}) + \alpha \boldsymbol{Z}_{r_2}^{(i)}. \tag{5.52}$$

For the same choice of α as for r_1 the second moment of the effective noise is

$$\tilde{\sigma}_{r_2}^2 = \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_{r_2}||^2\right] = \frac{2P\sigma_{r_2}^2}{2P + \sigma_{r_2}^2}. \tag{5.53}$$

Again, $\tilde{\boldsymbol{Z}}_{r_1}^{(i)}$ and $\tilde{\boldsymbol{Z}}_{r_2}^{(i)}$ can be well approximated by a Gaussian random vector with the same variance and the approximation becomes exact as $n \to \infty$.

In summary, the equivalent MLAN channel of the Gaussian sTTRC is given by

$$\tilde{\boldsymbol{Y}}_a^{(i)} = (\tilde{\boldsymbol{X}}_{r_1}^{(i)} + \tilde{\boldsymbol{Z}}_a^{(i)}) \bmod \Lambda \tag{5.54}$$

$$\tilde{\boldsymbol{Y}}_{r_1}^{(i)} = (\tilde{\boldsymbol{X}}_a^{(i)} + \tilde{\boldsymbol{X}}_{r_2}^{(i)} + \tilde{\boldsymbol{Z}}_{r_1}^{(i)}) \bmod \Lambda \tag{5.55}$$

$$\tilde{\boldsymbol{Y}}_{r_2}^{(i)} = (\tilde{\boldsymbol{X}}_{r_1}^{(i)} + \tilde{\boldsymbol{X}}_b^{(i)} + \tilde{\boldsymbol{Z}}_{r_2}^{(i)}) \bmod \Lambda \tag{5.56}$$

$$\tilde{\boldsymbol{Y}}_b^{(i)} = (\tilde{\boldsymbol{X}}_{r_2}^{(i)} + \tilde{\boldsymbol{Z}}_b^{(i)}) \bmod \Lambda \tag{5.57}$$

where $\tilde{\boldsymbol{X}}_a^{(i)}, \tilde{\boldsymbol{X}}_b^{(i)}, \tilde{\boldsymbol{X}}_{r_1}^{(i)}, \tilde{\boldsymbol{X}}_{r_2}^{(i)} \in \mathcal{R}_V(\Lambda)$ are the input signals to the channel in block i and the second moment of the effective noise $\tilde{\boldsymbol{Z}}_a^{(i)}$, $\tilde{\boldsymbol{Z}}_{r_1}^{(i)}$, $\tilde{\boldsymbol{Z}}_{r_2}^{(i)}$, and $\tilde{\boldsymbol{Z}}_b^{(i)}$ is denoted by $\tilde{\sigma}_a^2$, $\tilde{\sigma}_{r_1}^2$, $\tilde{\sigma}_{r_2}^2$, and $\tilde{\sigma}_b^2$ respectively. The noise is statistically independent of the input signals and approaches a Gaussian distribution for $n \to \infty$, i.e. the noise vectors are Gaussian in the limit.

We now describe the proposed coding scheme that is applied to this MLAN channel.

Encoding at the Users

A conceptual visualization of our encoding scheme (i.e. the involved lattices and codes) is depicted in Figure 5.3, which will be explained in the following. We consider a lattice

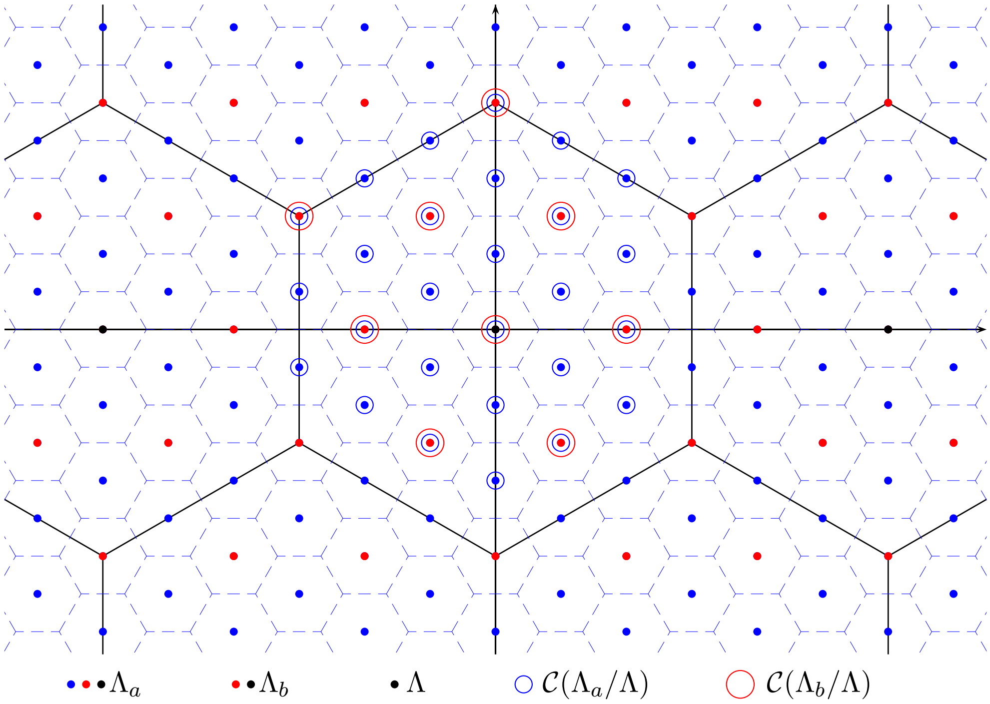


Figure 5.3.: Conceptual visualization of a lattice partition chain $\Lambda_a/\Lambda_b/\Lambda$ and the corresponding Voronoi codes. The Voronoi regions of Λ_b are not shown.

partition chain $\Lambda_a/\Lambda_b/\Lambda$, i.e. $\Lambda_a \supseteq \Lambda_b \supset \Lambda$, where Λ_a and Λ_b serve as coding lattices and Λ is the shaping lattice. In the following, we assume that all lattices are simultaneously good. The existence of lattice partition chains where all lattices are simultaneously good is proved in [36]. From the partition chain we obtain the following two Voronoi codes[7]

$$\mathcal{C}_a \overset{\text{def.}}{=} \mathcal{C}(\Lambda_a/\Lambda) \tag{5.58}$$

$$\mathcal{C}_b \overset{\text{def.}}{=} \mathcal{C}(\Lambda_b/\Lambda), \tag{5.59}$$

which are nested, i.e. $\mathcal{C}_b \subseteq \mathcal{C}_a$. The number of codewords in $\mathcal{C}_a$ is given by

$$|\mathcal{C}_a| = \frac{\mathrm{Vol}(\Lambda)}{\mathrm{Vol}(\Lambda_a)} = 2^{nR_a} \tag{5.60}$$

and we assume a one-to-one mapping between a particular message vector $\boldsymbol{W}_a^{(i)} \in \{0,1\}^{nR_a}$ and a lattice codeword in each block which is denoted by $\boldsymbol{V}_a^{(i)}$ for user a. Similarly the

[7] In the following we will use the original term "Voronoi codes" instead of "nested lattice codes" because in our case the codes are nested as well.

number of codewords in $\mathcal{C}_b$ is given by

$$|\mathcal{C}_b| = \frac{\mathrm{Vol}(\Lambda)}{\mathrm{Vol}(\Lambda_b)} = 2^{nR_b} \tag{5.61}$$

and user b maps the message $\boldsymbol{W}_b^{(i)} \in \{0,1\}^{nR_b}$ to a codeword $\boldsymbol{V}_b^{(i)}$ in each block.
Based on the intuition gained from the finite field model, the input signals to the MLAN channel in block i for both users are given by

$$\tilde{\boldsymbol{X}}_a^{(i)} = (\boldsymbol{V}_a^{(i)} - \boldsymbol{V}_a^{(i-2)}) \mod \Lambda \tag{5.62}$$

$$\tilde{\boldsymbol{X}}_b^{(i)} = (\boldsymbol{V}_b^{(i)} - \boldsymbol{V}_b^{(i-2)}) \mod \Lambda, \tag{5.63}$$

where $\boldsymbol{V}_a^{(i)} \stackrel{\text{def.}}{=} 0$ and $\boldsymbol{V}_b^{(i)} \stackrel{\text{def.}}{=} 0$ for $i < 1$. Note that $\tilde{\boldsymbol{X}}_a^{(i)} \in \mathcal{C}_a$ and $\tilde{\boldsymbol{X}}_b^{(i)} \in \mathcal{C}_b \subseteq \mathcal{C}_a$, i.e. both users transmit valid codeword in $\mathcal{C}_a$.

Relaying Strategy

The relays perform lattice decoding with respect to the coding lattice Λ_a, i.e. the relays find the closest lattice point to the received signal in Λ_a. This can also be regarded as a quantization operation because the relays don't recover individual user messages. The quantized signals (modulo Λ) are simply forwarded in the next block according to

$$\tilde{\boldsymbol{X}}_{r_1}^{(i)} = Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_1}^{(i-1)}) \mod \Lambda = Q_{\Lambda_a}(\tilde{\boldsymbol{X}}_a^{(i-1)} + \tilde{\boldsymbol{X}}_{r_2}^{(i-1)} + \tilde{\boldsymbol{Z}}_{r_1}^{(i-1)}) \mod \Lambda \tag{5.64}$$

$$\tilde{\boldsymbol{X}}_{r_2}^{(i)} = Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_2}^{(i-1)}) \mod \Lambda = Q_{\Lambda_a}(\tilde{\boldsymbol{X}}_{r_1}^{(i-1)} + \tilde{\boldsymbol{X}}_b^{(i-1)} + \tilde{\boldsymbol{Z}}_{r_2}^{(i-1)}) \mod \Lambda, \tag{5.65}$$

where we used the fact that

$$Q_{\Lambda_a}(\boldsymbol{x}) \mod \Lambda = Q_{\Lambda_a}(\boldsymbol{x} \mod \Lambda) \mod \Lambda \tag{5.66}$$

for any $\boldsymbol{x} \in \mathbb{R}^n$ if $\Lambda_a \supset \Lambda$ [45]. By convention, in the first block the input signals to the MLAN channel[8] are given by $\tilde{\boldsymbol{X}}_{r_1}^{(1)} = \tilde{\boldsymbol{X}}_{r_2}^{(1)} \stackrel{\text{def.}}{=} 0$.
From the above definitions and the encoding process at the users, it follows that $\tilde{\boldsymbol{X}}_a^{(i)} + \tilde{\boldsymbol{X}}_{r_2}^{(i)}$ as well as $\tilde{\boldsymbol{X}}_{r_1}^{(i)} + \tilde{\boldsymbol{X}}_b^{(i)}$ are lattice points in Λ_a for each block. Therefore, decoding at the relays is successful in block i if

$$Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_1}^{(i)}) \mod \Lambda = (\tilde{\boldsymbol{X}}_a^{(i)} + \tilde{\boldsymbol{X}}_{r_2}^{(i)}) \mod \Lambda \tag{5.67}$$

$$Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_2}^{(i)}) \mod \Lambda = (\tilde{\boldsymbol{X}}_{r_1}^{(i)} + \tilde{\boldsymbol{X}}_b^{(i)}) \mod \Lambda \tag{5.68}$$

[8]Note that this does not imply that the relays remain silent in the first block, but rather only transmit random dithers.

holds, which is the case when the effective noise remains inside the fundamental Voronoi region of Λ_a. The probability of decoding error at the relays vanishes for $n \to \infty$ if $\sigma^2(\Lambda_a) < \tilde{\sigma}_{r_1}^2$ and $\sigma^2(\Lambda_a) < \tilde{\sigma}_{r_2}^2$ by the virtue of Λ_a being Poltyrev-good. Then, for the first relay we have that for successful decoding the rate R_a must satisfy

$$R_a < \frac{1}{2}\log_2\left(\frac{\sigma^2(\Lambda)}{\tilde{\sigma}_{r_1}^2}\right) = \frac{1}{2}\log_2\left(\frac{P}{\frac{2P\sigma_{r_1}^2}{2P+\sigma_{r_1}^2}}\right) = \frac{1}{2}\log_2\left(\frac{1}{2}+\frac{P}{\sigma_{r_1}^2}\right). \tag{5.69}$$

Similarly for the second relay we obtain

$$R_a < \frac{1}{2}\log_2\left(\frac{\sigma^2(\Lambda)}{\tilde{\sigma}_{r_2}^2}\right) = \frac{1}{2}\log_2\left(\frac{1}{2}+\frac{P}{\sigma_{r_2}^2}\right). \tag{5.70}$$

By code construction we assumed that $R_a \leq R_b$ and in summary we have the following conditions for vanishing error probabilities at the relays:

$$R_b \leq R_a < \min\left(\tilde{\mathrm{C}}\left(P/\sigma_{r_1}^2\right), \tilde{\mathrm{C}}\left(P/\sigma_{r_2}^2\right)\right). \tag{5.71}$$

Decoding at the Users

Before describing the decoding steps at the users, first we state the following lemma.

Lemma 4. *Assuming that all decoding steps at the relays are successful up to (and including) block i, the quantized signals at the relays in block i can be written as*

$$Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_1}^{(i)}) \bmod \Lambda = (\boldsymbol{V}_a^{(i)} + \boldsymbol{V}_b^{(i-1)}) \bmod \Lambda \tag{5.72}$$

$$Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_2}^{(i)}) \bmod \Lambda = (\boldsymbol{V}_a^{(i-1)} + \boldsymbol{V}_b^{(i)}) \bmod \Lambda \tag{5.73}$$

Proof. The lemma is proved by induction over i. Only the steps for user a are shown. For $i = 1$ we have

$$Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_1}^{(1)}) \bmod \Lambda = (\tilde{\boldsymbol{X}}_a^{(1)} + \tilde{\boldsymbol{X}}_{r_2}^{(1)}) \bmod \Lambda = (\boldsymbol{V}_a^{(1)} - \boldsymbol{V}_a^{(-1)}) \bmod \Lambda = \boldsymbol{V}_a^{(1)}, \tag{5.74}$$

where we used the assumption that $\tilde{\boldsymbol{X}}_{r_2}^{(1)} \overset{\text{def.}}{=} 0$ and $\boldsymbol{V}_a^{(i)} \overset{\text{def.}}{=} 0$ for $i < 1$. For $i = 2$ we have

$$\begin{aligned} Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_1}^{(2)}) \bmod \Lambda &= (\tilde{\boldsymbol{X}}_a^{(2)} + \tilde{\boldsymbol{X}}_{r_2}^{(2)}) \bmod \Lambda \\ &= ((\boldsymbol{V}_a^{(2)} - \boldsymbol{V}_a^{(0)}) \bmod \Lambda + Q_{\Lambda_a}(\boldsymbol{Y}_{r_2}^{(1)}) \bmod \Lambda) \bmod \Lambda \\ &= (\boldsymbol{V}_a^{(2)} + (\tilde{\boldsymbol{X}}_{r_1}^{(1)} + \tilde{\boldsymbol{X}}_b^{(1)}) \bmod \Lambda) \bmod \Lambda \\ &= (\boldsymbol{V}_a^{(2)} + (\boldsymbol{V}_b^{(1)} - \boldsymbol{V}_b^{(-1)}) \bmod \Lambda) \bmod \Lambda \\ &= (\boldsymbol{V}_a^{(2)} + \boldsymbol{V}_b^{(1)}) \bmod \Lambda, \end{aligned} \tag{5.75}$$

where we used the assumption that $\tilde{\boldsymbol{X}}_{r_1}^{(1)} \stackrel{\text{def.}}{=} 0$ and $\boldsymbol{V}_b^{(i)} \stackrel{\text{def.}}{=} 0$ for $i < 1$. Finally for $i = l+1$ we have

$$\begin{aligned}
Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_1}^{(l+1)}) \bmod \Lambda &= (\tilde{\boldsymbol{X}}_a^{(l+1)} + \tilde{\boldsymbol{X}}_{r_2}^{(l+1)}) \bmod \Lambda \\
&= (\boldsymbol{V}_a^{(l+1)} - \boldsymbol{V}_a^{(l-1)} + Q_{\Lambda_a}(\boldsymbol{Y}_{r_2}^{(l)})) \bmod \Lambda \\
&= (\boldsymbol{V}_a^{(l+1)} - \boldsymbol{V}_a^{(l-1)} + \tilde{\boldsymbol{X}}_{r_1}^{(l)} + \tilde{\boldsymbol{X}}_b^{(l)}) \bmod \Lambda \\
&= (\boldsymbol{V}_a^{(l+1)} - \boldsymbol{V}_a^{(l-1)} + Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_{r_1}^{(l-1)}) + \boldsymbol{V}_b^{(l)} - \boldsymbol{V}_b^{(l-2)}) \bmod \Lambda \\
&= (\boldsymbol{V}_a^{(l+1)} + \boldsymbol{V}_b^{(l)}) \bmod \Lambda \qquad (5.76)
\end{aligned}$$

where the last step follows by inductive assumption (cp. equation (5.72)). The steps for the second relay are similar. □

From (5.72) and (5.73) it follows that in each block the relays broadcast a linear combination of two user codewords according to

$$\tilde{\boldsymbol{X}}_{r_1}^{(i)} = (\boldsymbol{V}_a^{(i-1)} + \boldsymbol{V}_b^{(i-2)}) \bmod \Lambda \qquad (5.77)$$

$$\tilde{\boldsymbol{X}}_{r_2}^{(i)} = (\boldsymbol{V}_a^{(i-2)} + \boldsymbol{V}_b^{(i-1)}) \bmod \Lambda, \qquad (5.78)$$

assuming that the relays decode successfully. In block i user b tries to recover the codeword of user a, $\boldsymbol{V}_a^{(i-2)}$, by first subtracting $\boldsymbol{V}_b^{(i-1)}$, which is the transmitted codeword in the preceding block, and then quantizing the resulting signal with respect to Λ_a. Hence for user b the decoding process is similar to the decoding process at the relays. If we assume that all decoding steps at the relays are successful, then the estimate is given by

$$\begin{aligned}
\hat{\boldsymbol{V}}_a^{(i-2)} &= Q_{\Lambda_a}(\tilde{\boldsymbol{Y}}_b^{(i)} - \boldsymbol{V}_b^{(i-1)}) \bmod \Lambda \qquad (5.79)\\
&= Q_{\Lambda_a}(\tilde{\boldsymbol{X}}_{r_2}^{(i)} + \tilde{\boldsymbol{Z}}_b^{(i)} - \boldsymbol{V}_b^{(i-1)}) \bmod \Lambda \qquad (5.80)\\
&= Q_{\Lambda_a}(\boldsymbol{V}_a^{(i-2)} + \boldsymbol{V}_b^{(i-1)} + \tilde{\boldsymbol{Z}}_b^{(i)} - \boldsymbol{V}_b^{(i-1)}) \bmod \Lambda \qquad (5.81)\\
&= Q_{\Lambda_a}(\boldsymbol{V}_a^{(i-2)} + \tilde{\boldsymbol{Z}}_b^{(i)}) \bmod \Lambda. \qquad (5.82)
\end{aligned}$$

The error probability at user b vanishes for $n \to \infty$ if $\sigma^2(\Lambda_a) < \tilde{\sigma}_b^2$. Therefore the code rate of $\mathcal{C}_a$ has to satisfy

$$R_a < \frac{1}{2}\log_2\left(\frac{\sigma^2(\Lambda)}{\tilde{\sigma}_b^2}\right) = \frac{1}{2}\log_2\left(\frac{P}{\frac{P\sigma_b^2}{P+\sigma_b^2}}\right) = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma_b^2}\right). \qquad (5.83)$$

From $\boldsymbol{V}_a^{(i-2)}$ the second user can obtain the message by inverting the one-to-one mapping between message vectors and codewords. For user a the estimate of the codeword is given by

$$\hat{\boldsymbol{V}}_b^{(i-2)} = Q_{\Lambda_b}(\tilde{\boldsymbol{Y}}_a^{(i)} - \boldsymbol{V}_a^{(i-1)}) \bmod \Lambda \tag{5.84}$$
$$= Q_{\Lambda_b}(\tilde{\boldsymbol{X}}_{r_1}^{(i)} + \tilde{\boldsymbol{Z}}_a^{(i)} - \boldsymbol{V}_a^{(i-1)}) \bmod \Lambda \tag{5.85}$$
$$= Q_{\Lambda_b}(\boldsymbol{V}_a^{(i-1)} + \boldsymbol{V}_b^{(i-2)} + \tilde{\boldsymbol{Z}}_a^{(i)} - \boldsymbol{V}_a^{(i-1)}) \bmod \Lambda \tag{5.86}$$
$$= Q_{\Lambda_b}(\boldsymbol{V}_b^{(i-2)} + \tilde{\boldsymbol{Z}}_a^{(i)}) \bmod \Lambda. \tag{5.87}$$

Note that lattice decoding at user a is with respect to Λ_b. The error probability vanishes for $n \to \infty$ if $\sigma^2(\Lambda_b) < \tilde{\sigma}_b^2$, because Λ_b is also assumed to be Poltyrev-good. Therefore, the message of user b can approach any rate up to

$$R_b < \frac{1}{2}\log_2\left(\frac{\sigma^2(\Lambda)}{\tilde{\sigma}_a^2}\right) = \frac{1}{2}\log_2\left(\frac{P}{\frac{P\sigma_a^2}{P+\sigma_a^2}}\right) = \frac{1}{2}\log_2\left(1+\frac{P}{\sigma_a^2}\right). \tag{5.88}$$

Note that user b could just quantize $\tilde{\boldsymbol{Y}}_b^{(i)}$ without subtracting $\boldsymbol{V}_b^{(i-1)}$. From the quantized signal the codeword of the other user could be extracted by employing SI *after* lattice decoding. User a could, in principle, proceed in a similar way as user b and first quantize the received signal with respect to Λ_a and then apply SI to extract the message. This would however put an additional constraint on the rate R_a. Therefore, for user a it is crucial to employ SI *prior to* lattice decoding which is then performed with respect to Λ_b. A complete block diagram visualizing the relationships between the input and output signals of the MLAN channel is depicted in Figure 5.4.
In summary, this transmission strategy can approach any rate pairs (R_a, R_b) satisfying

$$R_a < \min\left(\tilde{\mathrm{C}}\left(P/\sigma_{r_1}^2\right), \tilde{\mathrm{C}}\left(P/\sigma_{r_2}^2\right), \mathrm{C}\left(P/\sigma_b^2\right)\right) \tag{5.89}$$
$$R_b < \mathrm{C}\left(P/\sigma_a^2\right) \tag{5.90}$$

while making the error probability as small as desired. These constraints are identical to the constraints stated in the theorem if $\mathrm{C}\left(P/\sigma_b^2\right) \geq \mathrm{C}\left(P/\sigma_a^2\right)$. For the case where $\mathrm{C}\left(P/\sigma_b^2\right) < \mathrm{C}\left(P/\sigma_a^2\right)$ we relabel the users. This proves the theorem.

5.3. Discussion

The preceding analysis shows that there are striking similarities between the finite field physical layer and the Gaussian model assuming equal power constraints at the nodes. The key in order to make these similarities apparent is the MLAN conversion – a technique that was originally proposed in [31] for point-to-point channels and extended in [44] to

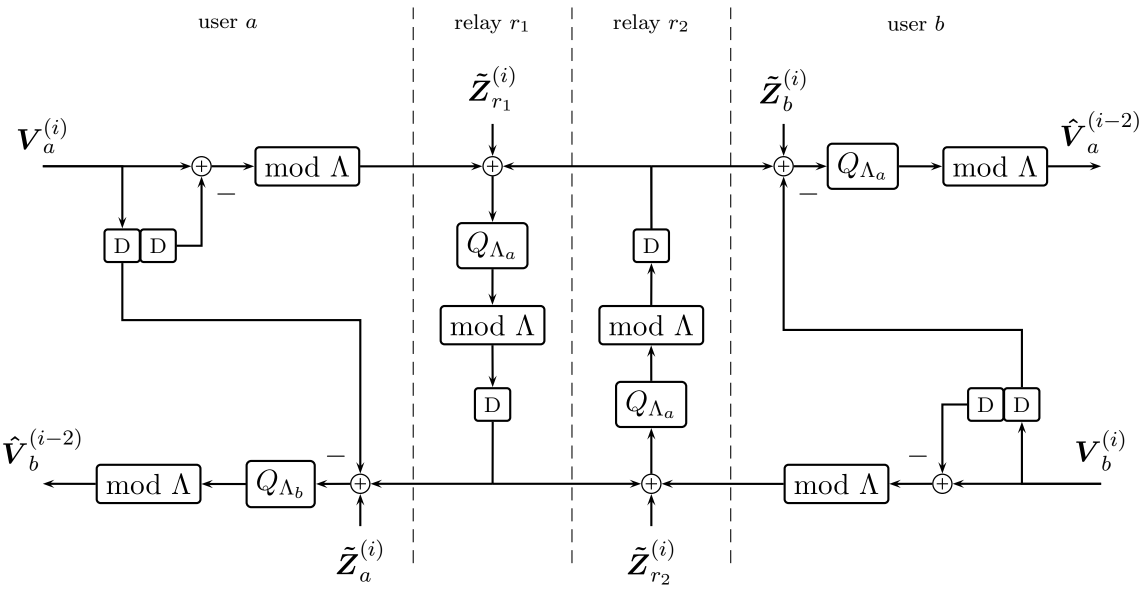


Figure 5.4.: Block diagram of the transmission strategy for the MLAN converted Gaussian sTTRC.

multi-access channels. The use of structured codes for the sTTRC – nested linear codes for the binary symmetric case and nested Voronoi codes for the Gaussian case – is necessary in order to exploit the linearity of the channel models. Note that for a point-to-point MLAN channel a capacity-achieving code can be constructed simply by choosing the codewords at random according to a uniform distribution over the fundamental Voronoi region of Λ [31]. A random code for the sTTRC, however, is unsuited for PNC because it lacks structure and therefore would not allow the relays to "protect" linear combinations of codewords.

In this chapter we assumed that the nodes have FD capability, mainly in order to simplify the analysis. This assumption is commonly not fulfilled in practice and therefore we briefly illustrate how the techniques developed here may be applied to the case with HD constrained device nodes and we also point out an important caveat that should be kept in mind regarding practical applications. In Figure 5.5 two modes (or network states) are depicted for the sTTRC. During mode 1 only user a and the second relay r_2 transmit while the other nodes receive and therefore remain silent. During mode 2 the situation reverses and user a and the first relay r_1 transmit. For a complete message exchange one would then alternate between these two modes and each "mode cycle" corresponds to a transmission block. An important point to observe here is the fact that one would have to make sure that the users are sufficiently separated, even for HD nodes. Recall that for the sTRC with only one relay and HD nodes the assumption of separation between

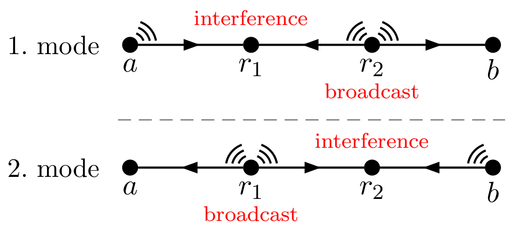


Figure 5.5.: Possible transmission modes for HD constrained device nodes.

the users is not necessary when using a two-phase protocol, i.e. the separation is implied (and therefore perfectly fulfilled in practice) due to the protocol. For a practical setup corresponding to the sTTRC however, the separation is not implied by the protocol (i.e. the transmit signal of a user may interfere with the transmit signal of a relay at the other user) and it is obvious that physical separation for wireless nodes may only be fulfilled *approximately* in practice.

Finally, we would like to address two ideas that might be useful in order to extend the derived results to the general Gaussian sTTRC with unequal power constraints at the relay nodes. One possible strategy is to use superposition coding, similar to the strategy that was used in the proof in section 3.4. However, we conjecture that it is not possible to achieve a constant gap to the upper bound based on a straightforward application of superposition coding in combination with the block transmission protocol. As an example, consider the case where the power constraints for the device nodes are given by P_a, $P_{r_1} = cP$, P_{r_2}, $P_b = P$, and the noise variances by $\sigma_a^2 = c$, $\sigma_{r_1}^2 = 0$, $\sigma_b^2 = 1$, $\sigma_b^2 = 0$ for $c > 1$. One can check that for these configuration the cut-set region is given by all positive rate pairs (R_a, R_b) for which $R_a < \mathrm{C}(cP)$ and $R_b < \mathrm{C}(P)$. The parameters are chosen such that it can be assumed that any transmission to r_1 and to user b happens instantaneously and without error because zero noise implies unlimited capacity independent of the transmit power (i.e. the available power at user a and the second relay r_2 is irrelevant here). Assume now that in the first transmission block r_1 transmits the message of user a to r_2 while at the same time b also transmits its message. For unequal rates we can use the rate splitting approach, i.e. r_1 splits the message of user a into two parts, one common-rate message with rate R_b which is encoded with a nested lattice code and one extra message with rate R_a'' which is encoded with a Gaussian codebook. User b encodes its message with the same nested lattice code. If we require the lattice codewords to be aligned, then we allocate a fraction of the available power of at r_1, in this case P, to the lattice, and the rest $(c-1)P$ to the Gaussian codeword. The relay r_2 uses successive cancellation and lattice decoding and therefore receives the extra message of user a and a combination of the common-rate message of user a and the message of user b. This information is instantaneously available also at r_1 and b and consequently (because sufficient SI is available) b has the message of user a and r_1 has the message of user b. However, in the second block r_1 has to convey the message of user b to user a while at the same time transmitting a new message from user a to r_2 as before. If we proceed in a straightforward manner, the relay r_1 encodes the message of user b also with the nested lattice code and adds the resulting codeword to the codeword corresponding to the new common-rate message of user a (modulo the coarse lattice). The problem is that, because only a fraction of the available power at r_1 is allocated to the lattice, lattice decoding at user a is only successful (the Gaussian codeword can be stripped off, because the message is known) if $R_b < \mathrm{C}\left(P/c\right)$ which can be arbitrarily far away from the cut-set bound $\mathrm{C}\left(P\right)$ by increasing c. Therefore this approach is unsuited to show achievability of the upper bound to within a constant gap. It may still be possible to use different coding approaches that are partially based on superposition coding for different channel parameters but the analysis for this approach may be infeasible because the parameter space is quite large. Another strategy would be to use different shaping lattices similar to the approach outlined in [7] for the sTRC. Further investigation of this approach for the sTTRC might be an interesting topic for future work.

CHAPTER 6

Conclusion

In this thesis we studied the bidirectional message exchange of two users via relay nodes. In particular we considered the separated two-way relay channel (sTRC) and the separated two-way two-relay channel (sTTRC).
For the Gaussian case of the sTRC, we described and compared three relaying strategies: amplify-and-forward (AF), decode-and-forward (DF), and modulo-and-forward (MF). AF and DF are standard relaying approaches that find application in a variety of relay network setups [46]. The sum rate comparison revealed that both strategies are capable of providing near-optimal (in the case of AF) and optimal performance (in the case of DF) for some channel configurations, but are suboptimal in general. The particular shortcoming of DF, namely the multiplexing loss, stems from the fact that the relay attempts to decode individual user messages, while this is not necessarily required. The main idea to avoid this loss is to exploit the additive nature of the wireless channel with the help of structured codes, in particular Voronoi codes or nested lattice codes. We provided a detailed description of the MF strategy which is partially based on AF, because the relay broadcasts a noisy version of the received signal, and partially on the structured code approach, because the fundamental region of a coarse lattice is used to shape the signal. The scheme was also generalized by introducing a second power allocation factor at user a and we applied MF to the *general* case of the Gaussian sTRC with no constraint on the downlink SNR. A brief example revealed that by optimizing the introduced power allocation factor improvements in terms of sum rate may be achieved. It was also pointed out that the comparison presented here was not meant to be exhaustive. The list of po-

tential relaying strategies that have been proposed for this network model is quite long (see e.g. [18] and [16]) and in general, different techniques are superior for different channel conditions. From an information-theoretic viewpoint it is desirable to find strategies that are good enough to show achievability of upper bounds. Therefore, we also provided a proof based on nested lattice codes showing that the capacity region of the Gaussian sTRC is achievable to within 1/2 bit per dimension for each user for arbitrary channel conditions – a result that was also recently published in [7]. However, for the uplink we used superposition coding instead of two different shaping lattices. Even though our approach is good enough to show achievability of the capacity region to within a constant gap, we acknowledge that it is inferior to the approach presented in [7], in the sense that the gap to capacity vanishes only for *one* user for increasing uplink SNR, while in [7] the gap vanishes for *both* users.

After that, we studied the sTTRC. This network can be seen as a multi-hop extension of the sTRC where the users exchange their messages via *two* relays. First, a finite field physical layer was considered and it was shown that for both the noiseless case and for the binary symmetric case the cut-set bound is achievable. For the binary symmetric case the capacity-achieving coding strategy is based on nested linear codes. The linearity of the code is used to exploit the linearity of the assumed channel model such that the relays can "protect" linear combinations of messages rather than decode individual messages. The insights gained from the finite field model were then applied to the Gaussian case of the sTTRC where we considered a slightly restricted channel model. In particular no unequal power constraints were allowed at the device nodes. It was shown that with this restriction the Gaussian case is very similar to the binary symmetric case by converting the continuous channel to an MLAN channel. The continuous counterpart of nested linear codes are nested Voronoi codes. With these two ingredients we derived an achievable rate region which is within 1/2 bit per dimension of the capacity region for each user. The gap to capacity vanishes for fixed noise variances and increasing node power.

As a last point, we would like to address a few topics that might be interesting for future work. One possible extension would be to study the bidirectional message exchange via L relays. It seems reasonable to assume that, for the case with equal power constraints at the nodes and full separation, it should possible to generalize the results by using a similar block transmission strategy and nested codes. However, it might be difficult to obtain closed form expressions for the relay signals and therefore a detailed analysis may be cumbersome. Another interesting topic for investigation is the performance of the lattice based strategies at low SNR. In the comparison chapter it was illustrated that for the symmetric case of the Gaussian sTRC no positive rate can be achieved using lattices below a certain SNR threshold and it remains an open question why this is the case. In [16] the authors state that “several groups have unsuccessfully tried to find a lattice scheme that can attain the upper bound” (with respect to the symmetric case). It is not clear whether the shortcoming of the lattice based schemes, i.e. the poor performance at low SNR, stems from the particular decoding scheme that is used – in this case lattice decoding – or from other factors. In order to partially resolve this question the authors in [6] considered minimum angle decoding as an alternative to lattice decoding. The analysis revealed that the achievable rates assuming this decoding method are identical to the rates for lattice decoding, suggesting that the underlying problem is of a more general nature.

APPENDIX A

Linear MMSE Estimation

Let $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}$ where $\boldsymbol{Z} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ and $\boldsymbol{X}$ is independent of $\boldsymbol{Z}$ with average power P. Let $\alpha \boldsymbol{Y}$ denote the linear MMSE estimator of $\boldsymbol{X}$ given $\boldsymbol{Y}$. Note that the estimation error (or effective noise) is given by $\tilde{\boldsymbol{Z}} = \boldsymbol{X} - \alpha \boldsymbol{Y} = (\alpha - 1)\boldsymbol{X} + \alpha \boldsymbol{Z}$ which has average power

$$\frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}||^2\right] = (\alpha - 1)^2 P + \alpha^2 \sigma^2. \tag{A.1}$$

Minimizing (A.1) with respect to α results in

$$\alpha^* = \frac{P}{P + \sigma^2}. \tag{A.2}$$

The effective noise power for this choice of α is given by

$$\frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}||^2\right]_{\alpha=\alpha^*} = \frac{P\sigma^2}{P + \sigma^2}. \tag{A.3}$$

In the following the scaling factors for the linear MMSE estimators that are used for the relaying strategies are calculated using (A.2) and (A.3).

Modulo-and-forward (Section 3.3)

User a

The effective noise power is given by (cp. (3.36)):

$$\begin{aligned}\tilde{\boldsymbol{Z}}_a &= (\alpha h_{ar}\sqrt{\theta} - 1)\boldsymbol{X}_b + \alpha \boldsymbol{Z}_r + (\beta_a h_{ra}\sqrt{\mu} - 1)\boldsymbol{X}'_r + \beta_a \boldsymbol{Z}_a \\ &= \boldsymbol{Z}'_a + \boldsymbol{Z}''_a.\end{aligned} \tag{A.4}$$

We have

$$\alpha^* = \frac{P}{P + \frac{1}{\theta h_{ar}^2}} = \frac{\theta h_{ar}^2 P}{\theta h_{ar}^2 P + 1} \tag{A.5}$$

and

$$\beta_a^* = \frac{P}{P + \frac{1}{\mu h_{ra}^2}} = \frac{\mu h_{ra}^2 P}{\mu h_{ra}^2 P + 1} \tag{A.6}$$

and therefore

$$\frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_a||^2\right] = \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}'_a||^2\right] + \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}''_a||^2\right] = \frac{P}{\gamma h_{ar}^2 P + 1} + \frac{P}{\mu h_{ra}^2 P + 1} \tag{A.7}$$

User b

The effective noise power is given by (cp. (3.39)):

$$\begin{aligned}\tilde{\boldsymbol{Z}}_b &= (\alpha h_{ar}\sqrt{\theta} - 1)\boldsymbol{X}'_a + \alpha \boldsymbol{Z}_r + (\beta_b h_{rb}\sqrt{\mu} - 1)\boldsymbol{X}'_r + \beta_b \boldsymbol{Z}_b. \\ &= \boldsymbol{Z}'_b + \boldsymbol{Z}''_b.\end{aligned} \tag{A.8}$$

We have the same α^* as for user a and

$$\beta_b^* = \frac{P}{P + \frac{1}{\mu h_{rb}^2}} = \frac{\mu h_{rb}^2 P}{\mu h_{rb}^2 P + 1} \tag{A.9}$$

and therefore

$$\frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_b||^2\right] = \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}'_b||^2\right] + \frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}''_b||^2\right] = \frac{P}{\gamma h_{ar}^2 P + 1} + \frac{P}{\mu h_{rb}^2 P + 1} \tag{A.10}$$

Proof (Section 3.4)

Relay r

The effective noise power is given by (cp. (3.54)):

$$\tilde{\boldsymbol{Z}}_r = (\alpha h_{br} - 1)(\boldsymbol{X}'_a + \boldsymbol{X}_b) + \alpha \boldsymbol{Z}_r. \tag{A.11}$$

We have

$$\alpha^* = \frac{2P}{2P + \frac{1}{h_{br}^2}} = \frac{2h_{br}^2 P}{2h_{br}^2 P + 1} \tag{A.12}$$

and therefore

$$\frac{1}{n}\mathrm{E}\left[||\tilde{\boldsymbol{Z}}_r||^2\right] = \frac{2P}{2h_{br}^2 P + 1}. \tag{A.13}$$

APPENDIX B

Notation

B.1. Acronyms

Abbreviation	Meaning
AF	amplify-and-forward
AWGN	additive white Gaussian noise
BPSK	binary phase-shift keying
BSC	binary symmetric channel
CF	compute-and-forward
CSI	cannnel state information
DF	decode-and-forward
FD	full-duplex
HD	half-duplex
MF	modulo-and-forward
ML	maximum likelihood
MLAN	modulo-lattice additive noise
MMSE	minimum mean square error
NC	network coding
PNC	physical-layer network coding
SI	side information
SNR	signal-to-noise ratio

Abbreviation	Meaning
sTRC	separated two-way relay channel
sTTRC	separated two-way two-relay channel
TRC	two-way relay channel

B.2. List of Symbols

Symbol	Description
$\|\|\cdot\|\|$	Euclidean norm
$\oplus$	exclusive or
$\mathbf{0}$	all-zero vector
a	device node, first user
α	MMSE scaling factor for the relay
b	device node, second user
β_a	MMSE scaling factor for user a
β_b	MMSE scaling factor for user b
$\boldsymbol{c}$	codeword
$\mathcal{C}$	code
$\mathcal{C}(\Lambda_c/\Lambda)$	nested lattice code
δ_a	gap to capacity for user a
δ_b	gap to capacity for user b
$\mathrm{E}[\cdot]$	expectancy operator
ε	crossover probability
G	normalized second moment of a region
$\boldsymbol{G}$	generator matrix
γ	scaling factor, used in the AF strategy
γ_a	scaling factor, used in the AF strategy
γ_b	scaling factor, used in the AF strategy
h	channel coefficient
$\mathrm{H}(\cdot)$	binary entropy function
$\boldsymbol{I}$	identity matrix
$\boldsymbol{K}$	noisy lattice point
$\tilde{\boldsymbol{K}}$	noisy lattice point
λ	arbitrary lattice point
Λ	shaping (or coarse) lattice
Λ_a	coding (or fine) lattice

Symbol	Description
Λ_b	coding (or fine) lattice
Λ_c	coding (or fine) lattice
M	number of transmission blocks
mod	modulo operation
μ	power allocation factor for superposition coding
n	number of channel uses / dimensions
$\mathcal{N}$	normal distribution
P	transmit power of a device node
$\Pr(\,\cdot\,)$	probability of an event
$Q_\Lambda(\,\cdot\,)$	nearest neighbor lattice quantizer
$Q_\mathcal{C}(\,\cdot\,)$	binary quantizer with respect to a linear code $\mathcal{C}$
r	device node, relay
r_1	device node, first relay in the two-relay setup
r_2	device node, second relay in the two-relay setup
R_a	information rate from user a to b, in bits per channel use
R_b	information rate from user b to a, in bits per channel use
R_{Lattice}	rate of the nested lattice code
R_{sum}	sum rate
$\mathcal{R}$	compact bounding region
$\mathcal{R}_c$	capacity region
$\mathcal{R}_{cut}$	cut-set region
$\mathcal{R}_V$	fundamental Voronoi region of a lattice
$\mathbb{R}$	set of real numbers
$\mathbb{R}^n$	real Euclidean n-dimensional space
σ^2	variance / second moment of a region
$\tilde{\sigma}^2$	effective noise power
θ	power allocation factor for superposition coding
u	information bit, packet
$\boldsymbol{u}$	information vector
$\boldsymbol{U}$	random dither vector
Unif	uniform distribution
$\boldsymbol{V}$	(lattice) codeword
$\hat{\boldsymbol{V}}$	estimated (lattice) codeword
Vol	volume of a region
W_a	message of user a

Symbol	Description
W_b	message of user b
$\boldsymbol{W}$	message vector
$\mathcal{W}$	message set
X	channel input variable
$\boldsymbol{X}$	transmitted signal
$\hat{\boldsymbol{X}}$	estimated signal
$\tilde{\boldsymbol{X}}$	input signal for the MLAN channel
Y	channel output variable
$\boldsymbol{Y}$	received signal
$\tilde{\boldsymbol{Y}}$	output signal of the MLAN channel
Z	random noise variable
$\boldsymbol{Z}$	random noise vector
$\tilde{\boldsymbol{Z}}$	effective noise vector
$\hat{\boldsymbol{Z}}$	estimated noise vector
$\mathbb{Z}$	set of integer numbers

Bibliography

[1] G. Kramer, I. Marić, and R. D. Yates, "Cooperative communications," *Foundations and Trends in Networking*, vol. 1, p. 271–425, 2006, ACM ID: 1295179.

[2] D. Gündüz, E. Tuncel, and J. Nayak, "Rate regions for the separated two-way relay channel," in *Proc. 46th Annual Allerton Conf. on Comm., Control, and Computing*, 2008.

[3] B. Rankov and A. Wittneben, "Spectral efficient signaling for half-duplex relay channels," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, 2005, p. 1066–1071.

[4] R. Knopp, "Two-Way radio networks with a star topology," in *2006 International Zurich Seminar on Communications*, Zurich, Switzerland, 2006, pp. 154–157. [Online]. Available: http://www.eurecom.fr/util/popuppubli.en.htm?page=detail&id=1894

[5] W. Nam, S. Chung, and Y. H. Lee, "Capacity bounds for Two-Way relay channels," in *2008 IEEE International Zurich Seminar on Communications*, Zurich, Switzerland, 2008, pp. 144–147.

[6] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5641–5654, 2010, asdf.

[7] W. Nam, S. Chung, and Y. Lee, "Capacity of the Gaussian Two-Way relay channel to within 1/2 bit," *Information Theory, IEEE Transactions on*, vol. 56, no. 11, pp. 5488–5494, 2010.

[8] R. Ahlswede, N. Cai, S. Li, and R. Yeung, “Network information flow,” *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*, 2nd ed. Wiley-Interscience, Jul. 2006.

[10] Y. Wu, P. A. Chou, and S. Y. Kung, “Information exchange in wireless networks with network coding and physical-layer broadcast,” Tech. Rep., 2004.

[11] S. Zhang, S. C. Liew, and P. P. Lam, “Hot topic: physical-layer network coding,” in *Proceedings of the 12th annual international conference on Mobile computing and networking*, ser. MobiCom ’06. New York, NY, USA: ACM, 2006, p. 358–365, ACM ID: 1161129.

[12] B. Nazer and M. Gastpar, “Compute-and-Forward: harnessing interference through structured codes,” *0908.2119*, Aug. 2009. [Online]. Available: http://arxiv.org/abs/0908.2119

[13] ——, “Computation over multiple-access channels,” *Information Theory, IEEE Transactions on*, vol. 53, no. 10, p. 3498–3516, 2007.

[14] R. Knopp, “Two-way wireless communication via a relay station,” 2007.

[15] I. J. Baik and S. Y. Chung, “Network coding for two-way relay channels using lattices,” in *Communications, 2008. ICC’08. IEEE International Conference on*, 2008, p. 3898–3902.

[16] B. Nazer and M. Gastpar, “Reliable physical layer network coding,” *Proceedings of the IEEE*, no. 99, p. 1–23, 2011.

[17] P. Larsson, N. Johansson, and K. Sunell, “Coded bi-directional relaying,” in *2006 IEEE 63rd Vehicular Technology Conference*, Melbourne, Australia, 2006, pp. 851–855.

[18] S. J. Kim, N. Devroye, P. Mitran, and V. Tarokh, “Comparison of bi-directional relaying protocols,” in *Sarnoff Symposium, 2008 IEEE*, 2008, p. 1–5.

[19] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, “Bounds on achievable rates for general multi-terminal networks with practical constraints,” in *Proceedings of the 2nd international conference on Information processing in sensor networks*, 2003, p. 146–161.

[20] T. J. Oechtering, R. F. Wyrembelski, and H. Boche, "Optimal time-division of two-phase decode-and-forward bidirectional relaying," in *2008 International Symposium on Information Theory and Its Applications*, Auckland, New Zealand, 2008, pp. 1–6.

[21] S. L. Fong and R. Yeung, "Capacity bounds for full-duplex two-way relay channel with feedback," in *Information Theory and Applications Workshop (ITA), 2011*, 2011, p. 1–10.

[22] S. J. Kim, P. Mitran, and V. Tarokh, "Performance bounds for bidirectional coded cooperation protocols," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5235–5241, 2008.

[23] P. Popovski and H. Yomo, "The Anti-Packets can increase the achievable throughput of a wireless Multi-Hop network," in *2006 IEEE International Conference on Communications*, Istanbul, 2006, pp. 3885–3890.

[24] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: analog network coding," in *ACM SIGCOMM Computer Communication Review*, vol. 37. New York, NY, USA: ACM, 2007, p. 397–408, ACM ID: 1282425.

[25] T. J. Oechtering, C. Schnurr, I. Bjelakovic, and H. Boche, "Broadcast capacity region of two-phase bidirectional relaying," *Information Theory, IEEE Transactions on*, vol. 54, no. 1, p. 454–458, 2008.

[26] L. L. Xie, "Network coding and random binning for multi-user channels," in *Information Theory, 2007. CWIT'07. 10th Canadian Workshop on*, 2007, p. 85–88.

[27] Y. Wu, "Broadcasting when receivers know some messages a priori," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, 2008, p. 1141–1145.

[28] G. Kramer and S. Shamai, "Capacity for classes of broadcast channels with receiver side information," in *Information Theory Workshop, 2007. ITW'07. IEEE*, 2007, p. 313–318.

[29] R. Zamir, "Lattices are everywhere," in *Information Theory and Applications Workshop, 2009*, 2009, p. 392–421.

[30] G. D. Forney, Jr., "Coset codes. I. introduction and geometrical classification," *Information Theory, IEEE Transactions on*, vol. 34, no. 5, p. 1123–1151, 1988.

[31] U. Erez and R. Zamir, "Achieving 1/2 log (1+ SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, 2004.

[32] A. S. Avestimehr, A. Sezgin, and D. N. Tse, "Approximate capacity of the two-way relay channel: A deterministic approach," in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, 2008, pp. 1582–1589.

[33] A. Sezgin, A. S. Avestimehr, M. A. Khajehnejad, and B. Hassibi, "Divide-and-conquer: Approaching the capacity of the two-pair bidirectional gaussian relay network," *Arxiv preprint arXiv:1001.4271*, 2010.

[34] G. D. Forney, Jr., "On the role of MMSE estimation in approaching the information-theoretic limits of linear gaussian channels: Shannon meets wiener," in *Proceedings of the annual Allerton conference on communication control and computing*, vol. 41, 2003, p. 430–439.

[35] ——, "Principles of digital communication II, lecture notes," 2005.

[36] D. Krithivasan and E. Pradhan, "A proof of the existence of good nested lattices," 2007. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.152.1198

[37] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *Information Theory, IEEE Transactions on*, vol. 51, no. 10, p. 3401–3416, 2005.

[38] H. A. Loeliger, "Averaging bounds for lattices and linear codes," *Information Theory, IEEE Transactions on*, vol. 43, no. 6, p. 1767–1773, 1997.

[39] E. Yilmaz and R. Knopp, "Hash-and-Forward relaying for Two-Way relay channel," *1102.3082*, 2011. [Online]. Available: http://arxiv.org/abs/1102.3082

[40] S. J. Kim, N. Devroye, P. Mitran, and V. Tarokh, "Achievable rate regions for bi-directional relaying," *0808.0954*, 2008. [Online]. Available: http://arxiv.org/abs/0808.0954

[41] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *Information Theory, IEEE Transactions on*, vol. 48, no. 6, p. 1250–1276, 2002.

[42] L. Xiao, T. E. Fuja, J. Kliewer, and D. J. C. Jr, "Nested codes with multiple interpretations," in *Information Sciences and Systems, 2006 40th Annual Conference on*, 2006, p. 851–856.

[43] M. Bossert, *Kanalcodierung*, 2nd ed. Teubner Verlag, Jul. 1998.

[44] U. Erez and R. Zamir, "A modulo-lattice transformation for multiple-access channels," in *Electrical and Electronics Engineers in Israel, 2008. IEEEI 2008. IEEE 25th Convention of*, 2009, p. 836–840.

[45] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference with structured codes," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, 2008, p. 772–776.

[46] A. E. Gamal and Y. Kim, "Lecture notes on network information theory," *1001.3404*, Jan. 2010. [Online]. Available: http://arxiv.org/abs/1001.3404