DESIGN AND ANALYSIS OF MULTILEVEL CODED MODULATION FOR MULTI-NODE NETWORKS

by

Ahmed Attia Abotabl

APPROVED BY SUPERVISORY COMMITTEE:

Dr. Aria Nosratinia, Chair

Dr. John P. Fonseka

Dr. Naofal Al-Dhahir

Dr. Hlaing Minn

Copyright © 2017

Ahmed Attia Abotabl

All rights reserved

To Marwa Fahim.

DESIGN AND ANALYSIS OF MULTILEVEL CODED MODULATION FOR MULTI-NODE NETWORKS

by

AHMED ATTIA ABOTABL, BS, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

December 2017

ACKNOWLEDGMENTS

I owe my sincere gratitude to my PhD adviser Dr. Aria Nosratinia for his wisdom, endless support and teaching me the art of research. I appreciate the valuable remarks of my PhD committee members, Drs. Naofal Al-Dhahir, Hlaing Minn and John P. Fonseka. I am also indebted to Dr. Matthieu Bloch of Georgia Tech, for interesting conversations about my research.

I am thankful to my colleagues who encouraged and supported me during my studies: Ahmed Hindy, Mohamed Fadel, Ahmed Helmy, Ahmed Omar, Ahmed El-Samadouny, Mohamed Mokhtar, Noha Helal, Hussein Saad, Ahmed Hesham, Ahmed Gomaa, Fan Zhang and Hassan Zivari-Fard. My dear friends Joseph Beshay, Mohamed Hafez, Mohamed El-habbab, Ahmed Samy, Ahmed Farid and Ibrahim Ezzat. Their friendship has been a great asset during my graduate studies.

Finally, I am grateful to my family: my father, mother and brother, who have been a constant source of motivation.

June 2017

DESIGN AND ANALYSIS OF MULTILEVEL CODED MODULATION FOR MULTI-NODE NETWORKS

Ahmed Attia Abotabl, PhD
The University of Texas at Dallas, 2017

Supervising Professor: Dr. Aria Nosratinia, Chair

During the last several decades, information theory has made significant advances in the analysis of the limits of communication in multi-node networks and the methods that can approach those limits. This dissertation studies new multi-level architectures for coded modulation in multi-node networks that aim to approach, in the practical realm, the capacity limits unveiled by information theory.

For the two-user additive white Gaussian noise (AWGN) broadcast channel, a multi-level coding architecture is proposed whose performance can approach the entire capacity region, and whose attractive features include a convenient partition of the two users' data so that one and only one of the modulation bit levels (and the corresponding encoder) must contend with both users' data. Practical aspects of the problem, including allocation of levels to users and finding level-wise code rates, have been addressed.

For the full-duplex decode-forward relay channel, a pragmatic yet capacity-approaching construction is proposed that synthesizes the components of full-duplex transmission via distinct signal levels of a multi-level code at the source and at the relay. The rate penalty due to linearity of component codes is analyzed and to avoid it, a solution is proposed involving the labeling of signal constellations. Simulations show that the proposed architecture together with good point-to-point codes can achieve excellent performance.

For the full-duplex decode-compress-forward relay channel, a mutli-level coding architecture is proposed and analyzed that achieves rates very close to the best known (information theoretic) achievable rates. The performance of the proposed architecture is evaluated using a combination of low-density parity-check (LDPC) codes and polar codes.

For the design of coded modulation for the discrete-input, Gaussian noise wiretap channels, a rate splitting method is proposed to allow a convenient construction of wiretap channel codes via a combination of two separate encoders operating on the data and the dither components. This technique leads naturally to the construction of multilevel codes for the AWGN wiretap channel where the message and the dither are encoded through separate levels without compromising secrecy. The effect of maximum likelihood decoding and multistage decoding at the legitimate receiver, as well as the effect of modulation labeling, are studied.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

The last two decades have witnessed an enormous progress in network information theory. The capacity and achievable rates of many communications networks such as the broadcast, the relay and the wiretap channels have been studied [1]. In spite of this progress, practical implementation of such techniques is not very well understood to date.

Much of the existing work in coding for multi-node networks concentrates on binary-input channels. Some of the examples are for the broadcast channel [2, 3], the relay channel [4, 5, 6] and the wiretap channel [7, 8, 9, 10]. However, a more general class of coded modulations for medium and high-SNR applications has proved more challenging.

While point-to-point coded modulation is a mature topic with excellent results reported, e.g., trellis coded modulation (TCM) [11, 12, 13] and bit-interleaved coded modulation (BICM) [14, 15], it is well-known that high performance in the physical layer of multi-node networks is not always made possible by patching together independent point-to-point links. Therefore, the extension of point-to-point codes to the multi-node networks is not straightforward.

We propose a set of multilevel coding (MLC) [16] architectures for efficient communication over several multi-node networks. In doing so, we are motivated by several facts. To begin with, it is well known that a multilevel coding decomposition does not suffer from any capacity loss in the point-to-point channel, subject to certain symmetry conditions [17]. Second, recent developments in information theory for bounding the capacity of several multi-node networks have utilized a multilevel approximation of the physical channels, known as the deterministic model [18]. By approximating the physical channel into parallel, independent binary channels (levels), and furthermore approximating these levels to be either completely

free of noise or completely lost in noise, these deterministic models have been able to produce powerful insights into the calculation of the capacity of multi-node networks. The central idea of this dissertation is that a coded modulation strategy that is inspired and motivated by a similar multilevel architecture can hope to approach the capacity of multi-node networks, if the approximations are stripped away and the true effect of the noise and the interaction between levels is accounted for.

1.2 Contributions

For the two-user additive white Gaussian noise (AWGN) broadcast channel, a multi-level coding architecture is proposed whose performance can approach the entire capacity region, and whose attractive features include a convenient partition of the two users' data so that one and only one of the modulation bit levels (and the corresponding encoder) must contend with both users' data. Practical aspects of the problem, including allocation of levels to users and finding level-wise code rates, have been addressed.

For the full-duplex decode-forward relay channel, a pragmatic yet capacity-approaching construction is proposed that synthesizes the components of full-duplex transmission via distinct signal levels of a multi-level code at the source and at the relay. The rate penalty due to linearity of component codes is analyzed and to avoid it, a solution is proposed involving the labeling of signal constellations. Simulations show that the proposed architecture together with good point-to-point codes can achieve excellent performance.

For the full-duplex decode-compress-forward relay channel, a mutli-level coding architecture is proposed and analyzed that achieves rates very close to the best known (information theoretic) achievable rates. The performance of the proposed architecture is evaluated using a combination of low-density parity-check (LDPC) codes and polar codes.

For the design of coded modulation for the discrete-input, Gaussian noise wiretap channels, a rate splitting method is proposed to allow a convenient construction of wiretap chan-

nel codes via a combination of two separate encoders operating on the data and the dither components. This technique leads naturally to the construction of multilevel codes for the AWGN wiretap channel where the message and the dither are encoded through separate levels without compromising secrecy. The effect of maximum likelihood decoding and multistage decoding at the legitimate receiver, as well as the effect of modulation labeling, are studied.

1.3 Preliminaries

This section provides a brief summary of the main concepts used in this dissertation.

1.3.1 Channel Capacity

Any channel is defined by the conditional distribution $P_{Y|X}(y|x)$ of the output Y given the input X. The input-output relationship for the AWGN point-to-point channel is

$$Y = X + N \tag{1.1}$$

where N is zero-mean, Gaussian noise with variance σ^2. The capacity of this channel is given by

$$C = \log(1 + SNR) \tag{1.2}$$

where SNR is the signal-to-noise ratio. The optimal input distribution to this channel is Gaussian with zero-mean and variance equals to P where P is the transmitter power.

1.3.2 Constellation-Constrained Capacity

The constellation constrained capacity of the point-to-point AWGN channel is given by the following optimization problem:

$$C = \max_{P_X(x)} I(X;Y) \tag{1.3}$$

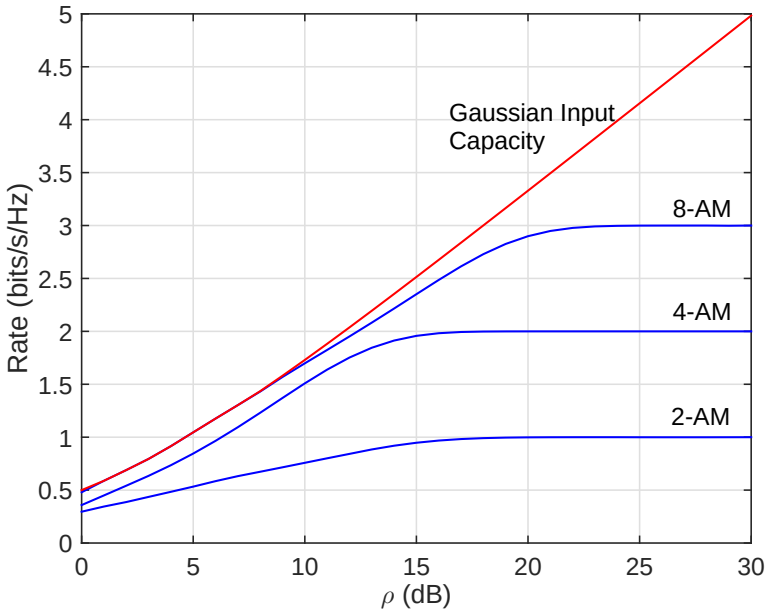

Figure 1.1. The constellation constrained capacity of several constellations in the point-to-point AWGN channel compared with the Gaussian input capacity.

However, it does not have a closed form expression. What makes it even more complicated is that the input optimal distribution depends on the channel quality which is the SNR in the case of the AWGN channel.

Exhaustive search for the optimal input distribution for the constellation constrained channel can be complicated specially for large constellations. For example, the design variable of the optimization problem in (1.3) is $m-1$ dimensional for m-ary constellation. Balhut-Arimoto algorithm [19, 20] facilitates this problem and presents an algorithm to find the optimal input distribution of the point-to-point channel. Several constellation constrained capacity are shown in Fig. 1.1.

1.3.3 Coded Modulation

Coded modulation in the point-to-point channel has a long history and has been studied in great detail [11, 12, 13]. The celebrated work of Ungerboeck [11] presented the trellis-coded

modulation (TCM) where a finite-state machine encodes the input bits and the output bits are mapped to a channel input symbol via a modulation mapper. The finite-state machine is then designed to maximize the minimum distance between every two possible sequences. However, as the data rate gets higher, the design of the trellis gets harder.

Another important coded modulation is bit-interleaved coded modulation (BICM) [15]. BICM was first introduced by Zehavi [14] and later studied extensively in [15]. In BICM, the information is encoded via a binary encoder that is followed by an infinite depth interleaver and then every group of bits is mapped to a channel input symbol. It was shown that BICM can get rates that are very close to the constellation constrained capacity however, it does not achieve the capacity. Part of the loss in BICM is that every input to the mapper is treated in the same exact way.

Multilevel coding [16] is a close relative to BICM where every input to the mapper is encoded independently with its own rate. In the point-to-point channel, binary component multilevel coding (see Fig. 1.2) is implemented by splitting the data stream into $m = \log_2(q)$ sub-streams for a q-ary constellation. Each sub-stream i is encoded independently with rate R_i. At each time instance, the outputs of the (binary) encoders are combined to construct the vector $[B_1, B_2, \ldots B_m]$ which is then mapped to a constellation point X and transmitted over. The channel is described by the conditional distribution $P_{Y|X}(y|x)$ where Y is the output of the channel. The mutual information between the input and output is given by

$$I(X;Y) = I(B_1, B_2, \ldots, B_m; Y) = \sum_{i=1}^{m} I(B_i; Y|B^{i-1}) \tag{1.4}$$

with the definition $B^{i-1} \triangleq [B_1, B_2, \ldots, B_{i-1}]$ with B^0 representing a constant, and using the chain rule for mutual information and the one-to-one relationship between X and $[B_1, B_2, \ldots, B_m]$. This equation suggests a multistage decoding where the codeword of level i is decoded using the output of the decoders of the preceding levels. A necessary and sufficient condition for multilevel coding achieving the constellation constrained capacity is that

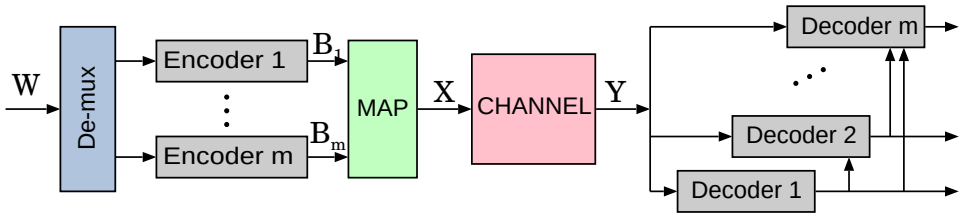

Figure 1.2. MLC and MSD in point to point channel.

the optimal distribution $P^*_{B_1,\ldots,B_m}(b_1,\ldots,b_m)$ must be independent across its components, i.e., (with a slight abuse of notation) [17]:

$$P_{B_1,\ldots,B_m}(b_1,\ldots,b_m) = \prod_{i=1}^{m} P_{B_i}(b_i) \tag{1.5}$$

A brief background survey on multilevel coding is as follows: Multilevel coding was proposed by Imai and Hirakawa in [16]. More details about the performance and the design of MLC can be found in [21, 22, 23]. Duan *et al.* [24] showed that MLC with linear mapping does not require active shaping to achieve the capacity. The MLC error exponent was analyzed by Ingber and Feder [17]. MLC was extended to the MIMO transmission [25], was used for diversity coding [26, 27, 28, 29] and in data storage [30]. Much less is known about MLC in the context of multi-node networks. A notable exception is [31] which used MLC in the context of compute and forward. But in general the optimality and efficient design of MLC for a variety of channels, including in particular the broadcast channel, relay channel and the wiretap channel has been for the most part an open problem.

CHAPTER 2

CODED MODULATION FOR THE BROADCAST CHANNEL

2.1 Introduction

The capacity of the AWGN broadcast channel is achieved via superposition coding [32, 1], but superposition of coded modulations is in general a modulation with much bigger size, and growth in the cardinality of constellation has practical costs that get progressively worse with more users. Quite aside from questions of cardinality, a superposition of coded modulations yields an irregular modulation constellation, with associated inconvenience and computational issues for the calculation of LLRs in hardware or firmware. Finally, the configuration of a superposition of constellations does not stay fixed throughout the rate region, in particular the peak-to-average power ratio (PAPR) [33], an important parameter for the efficiency of power amplifiers, becomes a variable quantity thus creating complications in the design of the transmitter.

Thus, broadcast coded modulation subject to a pre-determined transmit constellation is an important problem[1]. Coded modulation in the point-to-point channel has a long history and has been studied in great detail [11, 12, 13], but in the multi-node scenario, coded modulation introduces new and interesting phenomena and despite some progress, the design of capacity-approaching coded modulation for the broadcast channel under a *channel-input* constellation constraint has remained an essentially open problem. An outline of related work is as follows. Taubin [34] proposed the transmission of a weighted sum of two independent bit interleaved coded modulations and Sun *et al.* [35] proposed superposition Turbo TCM for the broadcast channel. Neither of these strategies obey a channel-input constellation constraint. A related area is the so-called single-user broadcasting [36], where

[1] ©[2017] IEEE. Reprint, with permission, from Ahmed Abotabl and Aria Nosratinia, Broadcast Coded Modulation: Multilevel and Bit-Interleaved Construction, IEEE Transactions on Communications, March, 2017

two streams are transmitted into a single-user channel with unequal-error protection (UEP). Earlier work in this area include Ramchandran *et al.* [37], on UEP modulation, however, the focus of their work is on providing variable error rates and not on capacity-approaching performance (see [37, Table II]).

This chapter addresses the design of multilevel coding (MLC) for the two-user AWGN broadcast channel under fixed constellation in size and shape at the channel input. In addition, a relative of MLC, the bit-interleaved coded modulation (BICM) [15] is employed for efficient implementation. For a two-user broadcast channel, we refer to the superposition code component for the weak user (experiencing lower signal-to-noise ratio) as the "outer code" and for the strong user as the "inner code." We show that for the inner code to be decomposable to multilevel code, necessary and sufficient conditions are essentially similar to the point-to-point scenario. We then show the optimality conditions for a multilevel decomposition of the outer code, and finally we highlight the optimality conditions for the (simultaneous) multilevel decomposition of the inner and outer codes. We show via numerical results that separating the two users' signals into distinct levels is in general insufficient to approach capacity. As mentioned earlier, this is the approach most commonly taken by the unequal error protection modulation schemes. Since mixing of the two users' signals is unavoidable, this chapter proposes a simple level-wise concatenation of user's codewords that closely approaches the capacity limit. The mixing of the two users' data can be limited to only one of the levels. We also propose a hybrid MLC-BICM that further simplifies the design, yet has excellent performance. Finally, we show that good point-to-point codes can be used as component codes for the multilevel encoder with excellent performance. For more than two-users, there will be more than two layers of encoders. Each layer encodes the information of a different receiver. Necessary and sufficient conditions for the decomposition of each layer into multilevel decomposition is a straight forward extensions of the results of this chapter. However, the design of bit-wise combining of more than two messages and the rate allocation per user at each level is not considered in this chapter.

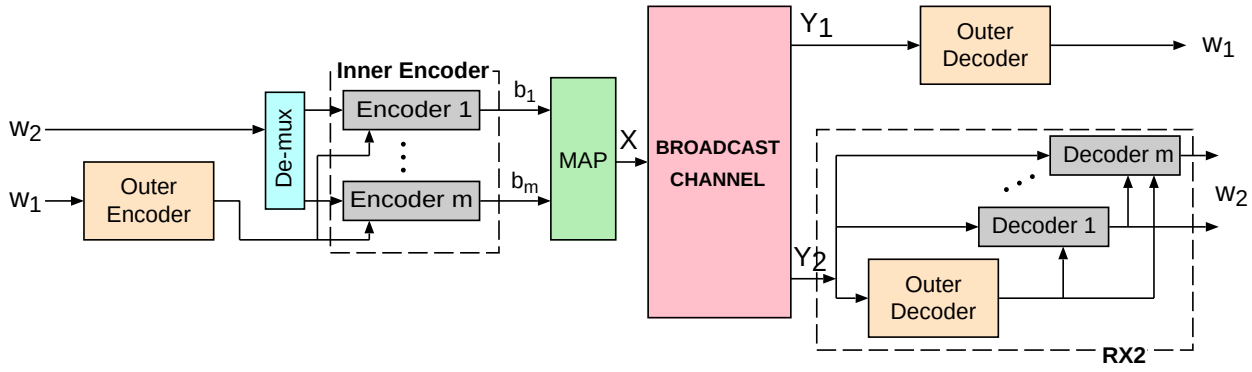

Figure 2.1. Broadcast channel with MLC for the inner code

Throughout the chapter, the SNR of a point-to-point AWGN channel is denoted by ρ and the SNR of the weak and the strong receivers of the AWGN broadcast channel are denoted by ρ_1 and ρ_2 respectively. Also, the noise variance at the weak and the strong receivers are denoted by σ_1^2 and σ_2^2.

2.2 Analysis of Multilevel Superposition Coded Modulation

2.2.1 Multilevel Inner Code

We begin by investigating multilevel decomposition of the inner code (see Fig. 2.1). The message w_1 is encoded with the outer code which is generated according to a distribution $p_U(u)$ to give the cloud centers of the superposition code (the codewords that will be decoded at both receivers). The message w_2 is split into m sub-messages. Sub-message i is encoded with inner code at level i that is generated according to a distribution $P_{B_i|U}(b_i|u)$. The inner code obeys an alphabet constraint on X as well as a multilevel coding constraint on the individual bits representing X, while the outer code in this case is unconstrained. The question is: under what conditions can such a decomposition meet the constellation constrained capacity?

The channel input X is constrained to a specific constellation via a one-to-one function $f : [B_1, \cdots, B_m] \to X$ whose domain is a vector of coded bits $[B_1, \cdots, B_m]$. The achievable

rate region of the broadcast channel subject to multilevel coding constraint on the inner code can be characterized by the following collection of weighted sum rates:

$$R = \max_{\prod_{i=1}^{m} P_{B_i|U}(b_i|u)P_U(u)} \{\theta I([B_1, \cdots, B_m]; Y_2|U) + (1-\theta) I(U; Y_1)\} \tag{2.1}$$

where $\theta \in [0, 1]$ is a parameter indicating the point achieved on the boundary of the rate region.

The modulation-constrained sum rate for the two-user degraded broadcast channel *without* any multilevel coding constraints is given by

$$R = \max_{P_{B_1,\cdots,B_m|U}(b_1,\cdots,b_m|u)P_U(u)} \{\theta I([B_1, \cdots, B_m]; Y_2|U) + (1-\theta) I(U; Y_1)\} \tag{2.2}$$

where the difference of (2.1) and (2.2) is that the former is optimized over a product conditional distribution for $B_1, \cdots, B_m$, whereas the latter is optimized over a general distribution. If the two sum-rate expressions are identical for all values of θ, it follows that the capacity regions must be identical.

Theorem 1. *A multilevel inner code achieves the constellation constrained capacity of the degraded broadcast channel if the capacity-achieving distributions on the individual bits of the modulation are conditionally independent, i.e.,*

$$P^*_{B_1,\cdots,B_m|U}(b_1, \cdots, b_m|u) = \prod_{i=1}^{m} P^*_{B_i|U}(b_i|u) \tag{2.3}$$

This optimality result is the counterpart of the point-to-point optimality result of Ingber and Feder [17]. The individual rates can be calculated using the usual peeling decoder for the strong user. When the outer decoder is implemented via multistage decoding, the achievable rates are:

$$R_1 \leq I(U; Y_1) \tag{2.4}$$

$$R_2 \leq I(X; Y_2|U) \tag{2.5}$$

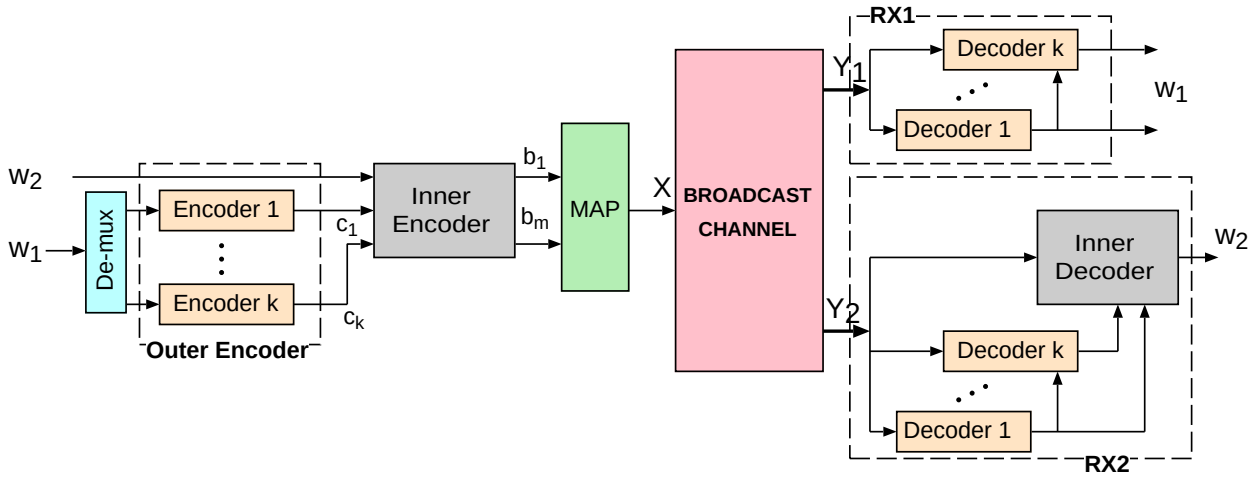

Figure 2.2. Broadcast channel with multilevel coding for the outer code

$$= \sum_{i=1}^{m} I(B_i; Y_2|U, B^{i-1}) \tag{2.6}$$

It follows that multistage decoding of the inner code is possible when

$$R_{2i} \leq I(B_i; Y_2|U, B^{i-1}) \tag{2.7}$$

where R_{2i} is the rate of the inner encoder at level i.

2.2.2 Multilevel Outer Code

We now consider the case when the inner code is unconstrained, but the outer code is a multilevel code (see Fig. 2.2). The outer code represents the cloud centers and is generated by the auxiliary random variable U, whose cardinality is enough to be bounded by the cardinality of X for optimality. The question is: when can the outer code be decomposed into *independently encoded* levels?

We now argue that it is always possible to produce a multilevel decomposition of the outer code with arbitrarily small loss, as long as it is permissible to increase the number of coding levels.

Consider a set of binary variables $C_1, \ldots, C_k$ representing the levels of the inner code, drawn independently according to Bernoulli-$\frac{1}{2}$. We now aim to find a mapping $g : [C_1, \cdots, C_k] \to U$ such that $p_U(u)$ approximates the capacity-optimizing distribution $p^*_U(u)$. Since each realization of C^k has probability 2^{-k}, the design of $g(\cdot)$ consists of crafting a many-to-one mapping from the bit vector to U so that

$$2^{-k}\left|\{[c_1, \cdots, c_k] \ : \ g(c_1, \cdots, c_k) = u_i\}\right| \approx P^*_U(u_i)$$

where $|\cdot|$ stands for the cardinality of the set it contains, and $P_{U^*}(u)$ is the optimal distribution of $P_U(u)$. It is not difficult to see that one is guaranteed to get to within 2^{-k} of approximating each $p_U(u)$.

The individual rates are therefore:

$$R_1 \leq I(U; Y_1) \tag{2.8}$$

$$= \sum_{i=1}^{k} I(C_i; Y_1 | C^{i-1}) \tag{2.9}$$

$$R_2 \leq I(X; Y_2 | C^k) \tag{2.10}$$

where $U = g([C_1, \cdots, C_k])$. Multistage decoding of the outer code at both receivers is subject to the following individual rate constraints

$$R_{1i} \leq I(C_i; Y_1 | C^{i-1}) \tag{2.11}$$

where R_{1i} is the rate of the encoder in level i of the outer encoder. Intuitively, if the weak receiver can do multistage decoding at a certain set of rates, so can the strong receiver at the same set of rates, because the strong receiver is less noisy. Formal derivation of this fact is straightforward and is relegated to Appendix 2.6.2.

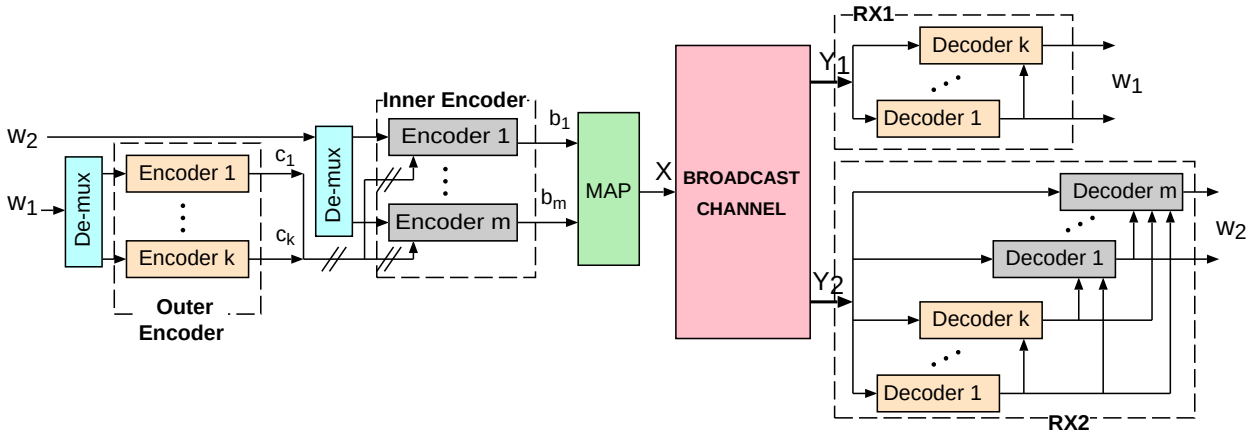

Figure 2.3. Broadcast channel with full multilevel superposition coding

2.2.3 Full Multilevel Superposition Coding

We now consider the case when the outer and the inner codes are decomposed to multilevel construction (see Fig. 2.3). Each encoder in the inner code depends on its message and the output of all the encoders of the outer code. The maximum achievable sum rate is given by

$$R = \max_{\prod_{i=1}^{m} P_{B_i|C^k}(b_i|c^k)P_{C_i}(c_i)} \theta I([B_1,\ldots,B_m];Y_2|U) + (1-\theta)I([C_1,\ldots,C_k];Y_1) \tag{2.12}$$

Denote the optimal distribution under the channel input constraint $X = f(B_1,\cdots,B_m)$ with $P^*_{X|U}(x|u)P_U(u) = P^*_{B_1,\cdots,B_m|U}(b_1,\cdots,b_m|u)P_U(u)$. A necessary and sufficient condition for the constellation-constrained optimality of a multilevel decomposition is that there exists a (potentially many-to-one) function $g(\cdot)$ so that for every u,

$$P^*_{B_1,\cdots,B_m|U}(b_1,\cdots,b_m|u)P_U(u) = \sum_{g(c^k)=u} \prod_{i=1}^{m} P_{B_i|C^k}(b_i|c^k) \prod_{j=1}^{k} P_{C_j}(c_j) \tag{2.13}$$

This means that the capacity achieving distribution on the coded bits $B_1,\cdots,B_m$ can be constructed by, firstly, cloud centers generated via independent binary variables $C_1,\ldots,C_k$ together with a mapping $g: C^k \to U$, and secondly coded bits $B_1,\ldots,B_m$ that are independent *conditioned on* $C_1,\ldots,C_k$. Using arguments similar to the ones in Section 2.2.2 and

Appendix 2.6.2, one can show that the conditions on the outer code can be satisfied to any required degree of approximation via increasing k, the number of the levels of the outer code.

Under this condition, the individual rates are:

$$R_1 \leq I(U;Y_1) \tag{2.14}$$

$$= \sum_{i=1}^{k} I(C_i;Y_1|C^{i-1}) \tag{2.15}$$

$$R_2 \leq I(X;Y_2|C^k) \tag{2.16}$$

$$= \sum_{j=1}^{m} I(B_j;Y_2|B^{j-1},C^k) \tag{2.17}$$

Multistage decoding of the outer and inner codes at both receivers is subject to the following individual rate constraints

$$R_{1i} \leq I(C_i;Y_1|C^{i-1}) \qquad 1 \leq i \leq k \tag{2.18}$$

$$R_{2j} \leq I(B_j;Y_2|B^{j-1},C^k) \qquad 1 \leq j \leq m \tag{2.19}$$

2.3 Design of Multilevel Superposition Coded Modulation

The results of the previous section show the conditions under which broadcast capacity can be achieved by multilevel coding. The remainder of this chapter shows that even in the absence of optimality conditions, MLC can still achieve rates very close to the boundary of the capacity region. This section produces a design methodology for multilevel broadcast coded modulation via a simple coding framework that greatly facilitates the design process and yet induces little or no performance penalty (allows near-optimal performance). Subsequently, we solve the problem of rate allocation between the users and layers of the multilevel code in the context of the proposed framework, thus completing the design process.

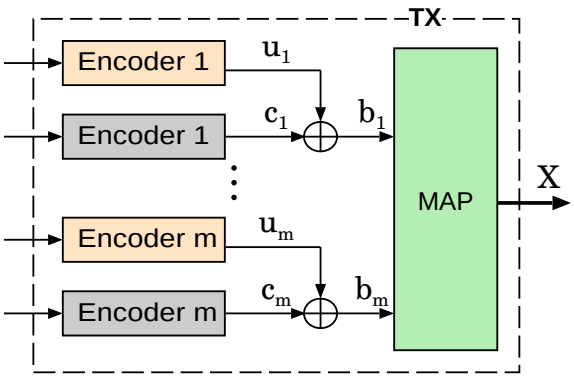

Figure 2.4. XOR implementation of multilevel bit-wise superposition coding.

2.3.1 Bit-Additive Superposition Coding

In the multilevel decomposition considered so far, each of the inner encoder levels depends on the code vector produced by *all* the outer encoders. The cross dependency of multiple codes is difficult to implement in practice, therefore it is natural to seek encoding methods whose levels are decoupled from each other *for both users*, especially considering that the notion of decoupling of levels is at the heart of motivation for the point-to-point multilevel codes [16]. This means that level-i encoder of the inner code reads only the output of level-i outer encoder, which leads to a *bit-wise superposition*. This can be optimal only if, in addition to the condition (2.13), we also have:

$$P_{B_i|C^k}(b_i|c^k) = P_{B_i|C_i}(b_i|c_i) \qquad \forall i \tag{2.20}$$

For most modulations used commonly in practice, this condition cannot be met precisely. Nevertheless, it is possible to achieve performance very close to capacity via an encoding method that decouples the bit levels from each other, and furthermore implements the superposition at each level by a simple binary additive operation. We call this simple multilevel superposition strategy the *bit-additive superposition*. We now proceed to describe this method and demonstrate its performance.

Fig. 2.4 shows the outline of the proposed method. The outer codes are generated independently according to Bernoulli-$\frac{1}{2}$ distribution, each with a prescribed rate R_{1i}, and are represented with variable C_i. The inner codes are represented by U_i, which are generated independently according to the distribution Bernoulli-α_i with $\alpha_i \in [0, 0.5]$. Bit-additive superposition is achieved via $B_i = C_i \oplus U_i$ where $\oplus$ represents the binary XOR operation. When $\alpha_i = 0$, we have $B_i = C_i$ so we have $R_{2i} = 0$. When $\alpha_i = 0.5$, B_i is independent of C_i and $R_{1i} = 0$. This method of binary superposition is mentioned, among others, in [1, Chapter 5] and [38].

The proposed bit-additive superposition can be implemented in the following manner: a binary linear code is chosen for each level of the outer code since linear codes have uniform distribution. For the encoders of the inner code, we need a code with distribution Bernoulli-α_i. Such a code can be generated from a linear code which has a uniform distribution and set the bits at randomly chosen locations with zero. For example, if the required distribution is Bernoulli-α_i, then the number of bits set to zero (regardless of their original value) should be

$$N = 2(1/2 - \alpha_i)n \tag{2.21}$$

where n is the block-length of the code.

2.3.2 Performance of Bit-Additive Superposition

We now provide numerical examples for a wide variety of modulations to demonstrate the efficacy of the proposed bit-additive superposition. The general setup for these numerical studies is as follows.

The baseline for comparisons in each case is the constellation constrained capacity, which is calculated using the modified Blahut-Arimoto algorithm [39]. In each case, the achievable rate region for the proposed bit-additive superposition is obtained in the following manner: For each level i, a uniformly distributed codeword is generated for the weak receiver and a

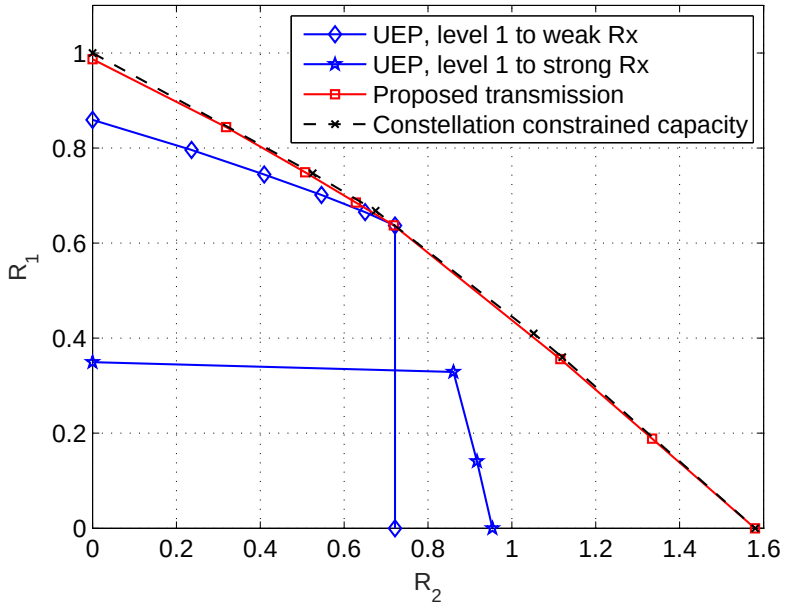

Figure 2.5. Performance of proposed technique versus UEP-type modulation that assigns levels to distinct users under 4-PAM, $\rho_1 = 5dB$, $\rho_2 = 10dB$.

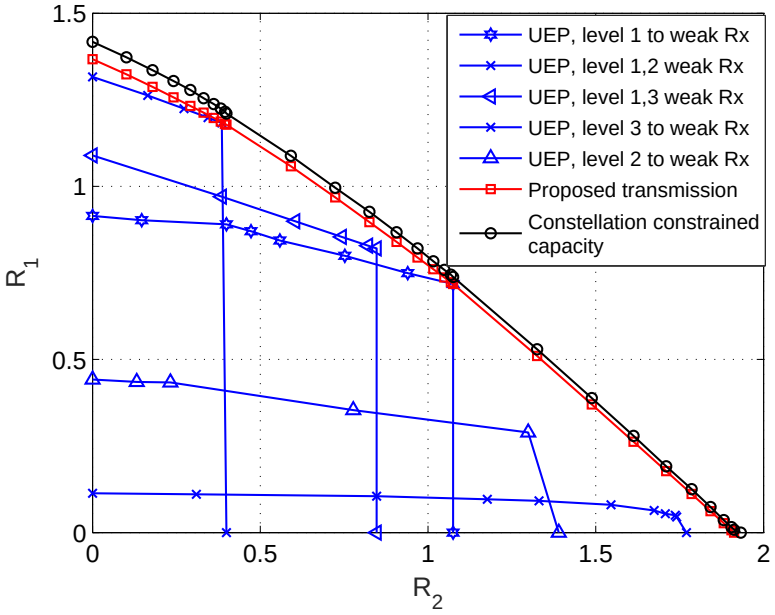

Figure 2.6. Comparison of the proposed technique with UEP-type modulation that assigns levels to distinct users under 8-PAM, $\rho_1 = 8dB$, $\rho_2 = 12dB$.

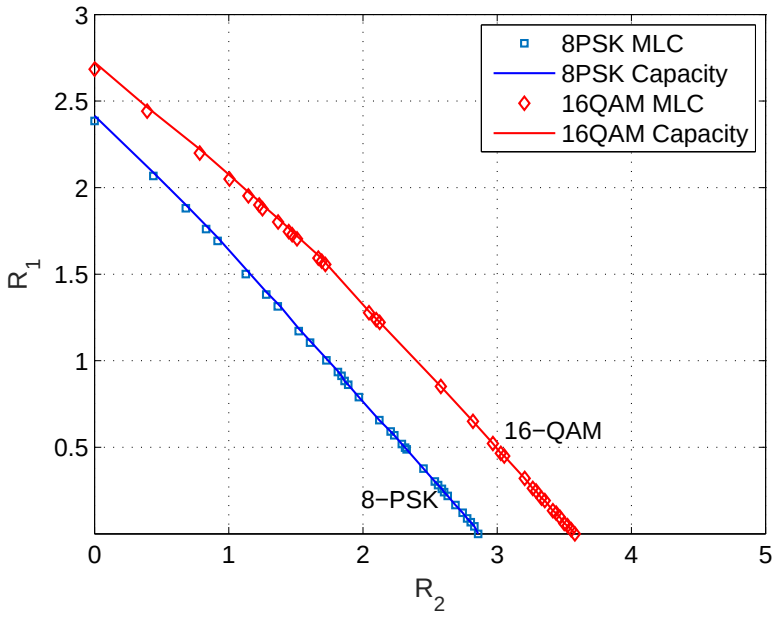

Figure 2.7. Proposed MLC transmission rates for 8-PSK and 16-QAM where $\rho_1 = 8dB$ and $\rho_2 = 12dB$.

codeword with distribution Bernoulli-α_i for the strong receiver. The input to the mapper at level i is the XOR between the weak receiver codeword at level i and the strong receiver codeword at level i. Each value of the vector $[\alpha_1, \alpha_2, \ldots, \alpha_m]$ gives a certain rate pair (R_1, R_2). For every value of the vector $[\alpha_1, \alpha_2, \ldots, \alpha_m]$, the mutual informations

$$I(C_1, \ldots, C_m; Y_1)$$

$$I(B_1, \ldots, B_m; Y_2 | C_1, \ldots, C_m)$$

are calculated. These mutual informations give an achievable rate pair R_1 and R_2 respectively.

Numerical results show a very small gap between the constellation constrained capacity and the proposed bit-additive superposition. In particular Fig. 2.5, Fig. 2.6, and Fig. 2.7 shows the performance of bit-additive superposition for 4-PAM, 8-PAM, 16-QAM and 8-PSK constellations. Simulations show the same achievable rate region via Gray and natural mapping.

Fig. 2.5 also shows comparisons to a bit-allocation strategy often used by the Unequal-Error Protection (UEP) modulations [37, 40], i.e., the higher-order bit levels are assigned to one data category and the lower-order bit levels to the other data category.

Fig. 2.5 represents 4-PAM modulation, and the UEP-type modulation curves represent the two possibilities of level-1 (respectively level-2) being assigned to weak (respectively strong) user, or vice versa. In the former case, we see that this assignment meets the capacity outer bound only at one point, otherwise it can be far from capacity. Reversing the assignment of modulation index to the users results in even worse performance.

It has been noted by [41, 42, 43] that in the UEP approach one may allocate each modulation index to one message at a time, but then allow time sharing between all such strategies. Thus one may achieve the convex hull of all points on such individual rate assignments, as well as the single-user rates. This can provide a performance closer to capacity, but requires buffering with its associated additional delay.

Remark 1. *For a fixed channel SNR and for a fixed rate pair, the larger the modulation size, the smaller is the gap-to-capacity for a static assignment of messages to modulation indices.*

Remark 2. *In Fig. 2.7 and even more so in Fig. 2.5, there is a very small gap between the modulation-constrained capacity and the multilevel coding rates, especially close to the vertical axis (when the weak user mostly occupies the channel). This can be clarified by looking at the single-user optimality condition of multilevel coding [17], finding that it is not met for PAM with natural labeling. For the single-user* 8*-PAM modulation under natural labeling, Fig. 2.8 shows the relationship of constellation constrained capacity and MLC achievable rate. 8-PAM experiences a MLC penalty that is more severe at low SNR,*[2] *therefore the slight separation*

[2]In the point-to-point channel this penalty goes away if at lower SNRs one uses a lower order modulation. Using a higher order modulation *and* requiring that all modulation points be used with equal probability (linear component codes) produces the rate penalty. In the broadcast channel this small penalty is not as easily avoidable because the same modulation is used to transmit to both users.

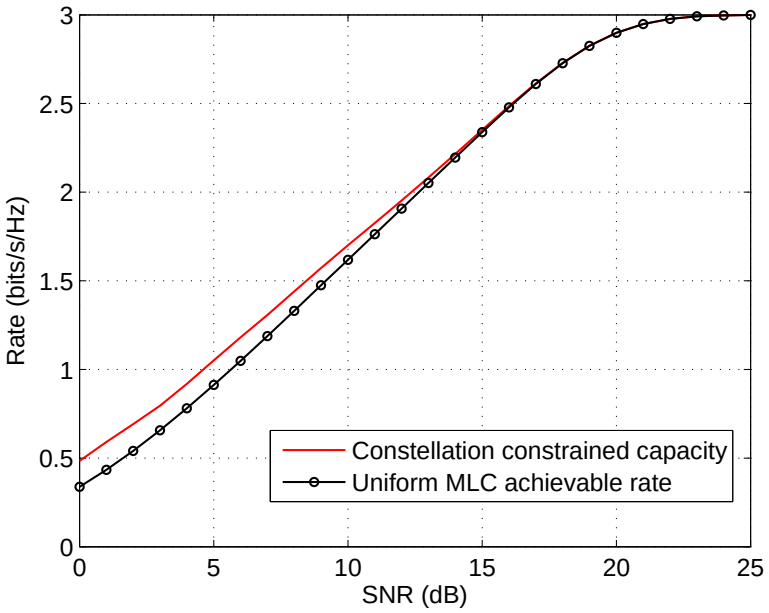

Figure 2.8. The penalty for using multilevel *linear* coding (equi-probable zeros and ones) in a single-user channel under 8-PAM with natural labeling

of rate curves in Figs. 2.5, and 2.7 is explained especially at the point where the weak user occupies the channel.

2.3.3 A Pragmatic Rate Allocation Algorithm

To achieve a desired broadcast rate pair (R_1, R_2) in the context of multilevel coding, it is necessary to identify the relevant codes at each layer, which begins by specifying the code rates R_{1i}, R_{2i} for all levels i. In this subsection, we present a pragmatic solution to this problem that in addition to its modest computational requirement, serves to reveal interactions between the rate constraints at different bit levels as well as interesting connections to the familiar single-user MLC mutual information curves. It will be demonstrated via simulations that this pragmatic method operates very close to the capacity region for most familiar modulations and mappings. Subsequently, we will discuss the rare cases where this

pragmatic method may lead to a slight departure from optimality, and propose a general (but not as computationally thrifty) algorithm for rate allocation in such cases.

We begin by casting the rate allocation problem in the form of the following optimization, where θ parametrizes the boundary of the broadcast rate region:

$$\max_{\Pi_i P_{B_i|C^k}(b_i|c^k)\Pi_j P_{C_j}(c_j)} \theta \sum_i R_{1i} + (1-\theta)\sum_j R_{2j}$$

$$\text{Subject to } R_{1i} \le I(C_i; Y_1|C^{i-1}) \qquad 1 \le i \le k$$

$$R_{2j} \le I(B_j; Y_2|B^{j-1}, C^k) \quad 1 \le j \le m$$

$$R_{1i} \ge 0 \qquad R_{2j} \ge 0 \quad \forall i, j$$

We will come back to a version of this general rate allocation problem in the sequel, but for now we concentrate on bit-additive superposition, where the rate allocation problem reduces to the following:

$$\max_{\Pi_i P_{U_i}(u_i)P_{C_i}(c_i)} \sum_i \theta R_{1i} + (1-\theta)R_{2i} \tag{2.22}$$

$$\text{Subject to } R_{1i} \le I(C_i; Y_1|C^{i-1}) \qquad 1 \le i \le m \tag{2.23}$$

$$R_{2i} \le I(U_i; Y_2|U^{i-1}, C^k) \quad 1 \le i \le m \tag{2.24}$$

$$R_{1i} \ge 0 \qquad R_{2i} \ge 0$$

The key difference is that the maximization is now over independent distributions, therefore the utility function can now be decomposed into the sum of m non-negative level-wise utility functions.

Having arrived at a simplified utility function, we now concentrate on the constraints by highlighting the shape of the feasible rate regions at each individual level, which can be thought of as cross sections of the overall feasible rate region. For insight, we look into the specific example of 8-PAM with natural labeling, where the level-wise rate constraints are shown[3] in Fig. 2.9.

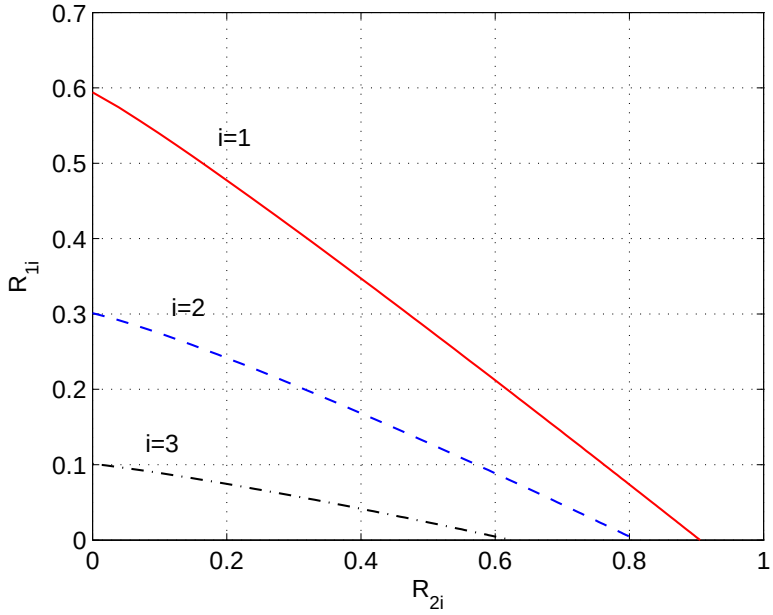

Figure 2.9. Rate constraints for the levels of 8-PAM constellation assuming natural labeling and decoding order from MSB to the LSB with $\rho_1 = 5$dB and $\rho_2 = 10$dB.

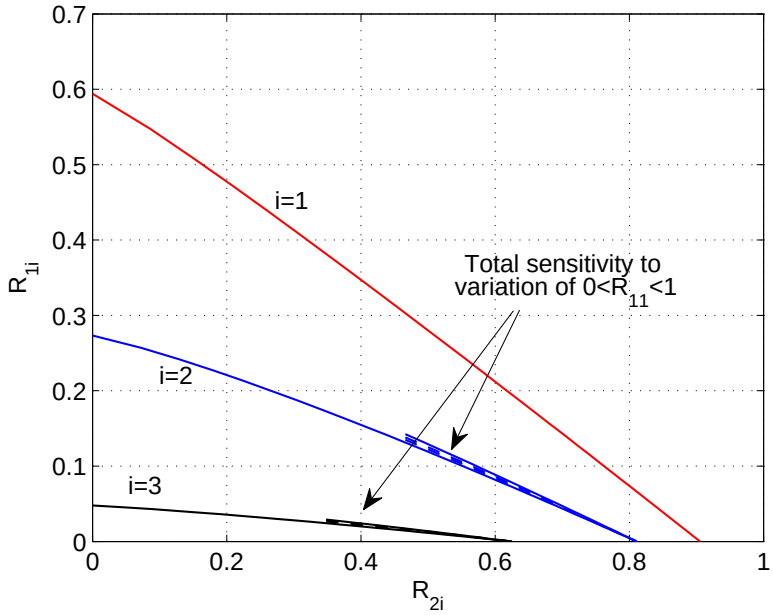

Figure 2.10. Sensitivity of each level's constraint to rates of other levels

The first interesting feature of the bit-level constraints is that, under most bit mappings including natural and Gray mapping, the binary rate constraint at each level is largely insensitive to the parameters pertaining to other levels. For example, please see Fig. 2.10, where in an 8-PAM multilevel coded modulation, the sensitivity of the rate constraints in levels $2, 3$ at the set point $R_{22} = R_{32} = 0$ is demonstrated subject to a complete sweep of the rate pair R_{11}, R_{12}. From this observation rises a pragmatic assumption: that at optimality, one may assume that the constraints at different levels are approximately independent.[4] This approximation leads to a complete decomposition of the optimization into level-wise optimizations whose only coupling is through the parameter θ, namely, for each $i = 1, \ldots, m$,

$$\max_{P_{U_i}(u_i)P_{C_i}(c_i)} \theta R_{1i} + (1-\theta)R_{2i} \tag{2.25}$$

$$\text{Subject to } g_i(R_{1i}, R_{2i}) \leq 0 \tag{2.26}$$

$$R_{1i} \geq 0 \qquad R_{2i} \geq 0 \tag{2.27}$$

where $g_i(\cdot, \cdot)$ is the rate constraint at each level whose dependence explicitly on R_{1i}, R_{2i} and omission of other variables is meant to highlight the approximate independence of the constraints at each level. Solving a typical rate allocation problem in the aforementioned example involves pushing a line with a slope determined by θ outward on the three levels mentioned above. An example is shown in Fig. 2.11, where the individual rate constraints for the three levels are shown in solid lines and the parallel dotted lines represent, for a fixed θ, the lines $\theta R_{1i} + (1-\theta)R_{2i} = \alpha_i$, and the maximization of α_i corresponds to the movement of the dotted lines as shown by arrows.

The result of this rate allocation is that Level 1 is dedicated to User 1, and levels 2 and 3 are dedicated to User 2. Note that the rate constraint curves were calculated under

[3]For each i, we have set the rates in other levels $j \neq i$ so that $R_{1j} = 0$.

[4]This approximation has been verified for all natural and Gray labeling for a variety of PAM, PSK, and QAM type modulations. There exist some irregular labeling for which this assumption fails. That case will be discussed separately in the sequel.

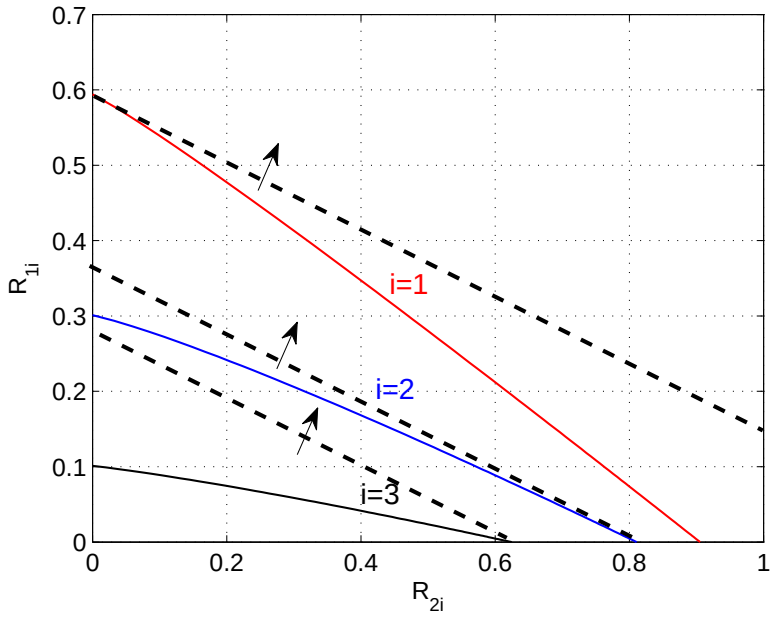

Figure 2.11. Rate allocation via optimization at each level

the operating regime that all three levels are assigned to User 2. To take into account the (small) sensitivity of the individual rate regions to the operating point of other levels, one may update the three rate curves once more and verify that optimality conditions remain satisfied at the proposed optimal point. The update may slightly adjust the intercept points.

We now consider a second empirical property of level-wise binary rate regions: that they are very nearly affine. This feature has been experimentally observed across modulations, bit level mappings, and various channel SNRs. The outcome of this second observation is that near optimal rate allocation can be achieved while allocating all the bits in each level to either one or the other user. This produces 2^m rate pairs that are close to the boundary of the rate region. Rate pairs in between can be achieved by dividing the rate in one of the levels (whose achievable rate slope is closest to $\theta R_1 + (1-\theta)R_2$ between the two users.

This approach yields results that are practically indistinguishable from optimal rate allocation, with very few exceptions that are discussed in the next subsection. The performance of this method is illustrated, for the case of a 8-PAM modulation with natural mapping,

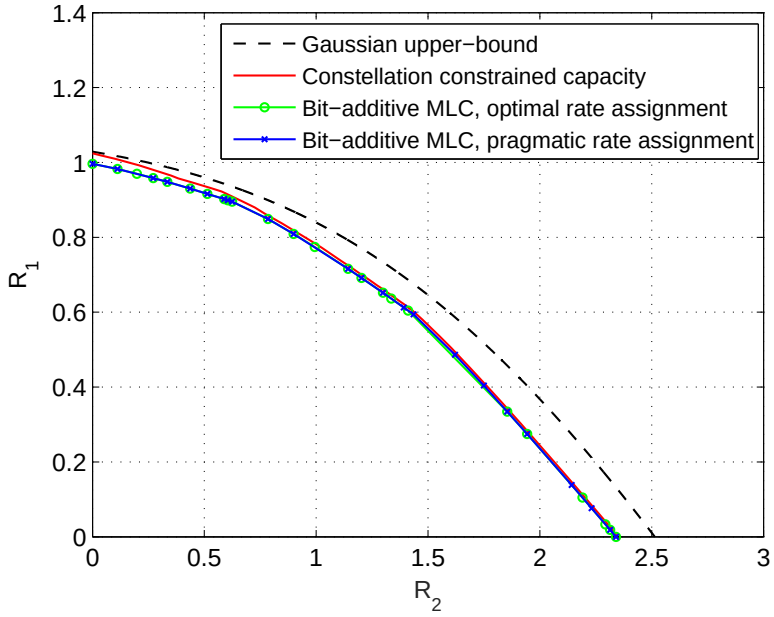

Figure 2.12. MLC rate region for 8-PAM, $\rho_1 = 5$dB, $\rho_2 = 15$dB.

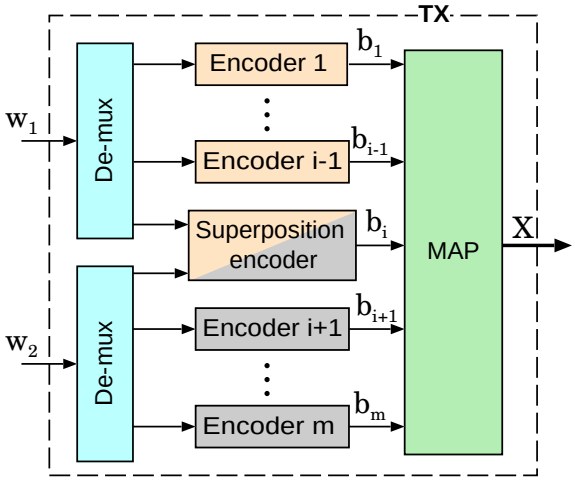

Figure 2.13. Multilevel superposition with pragmatic rate allocation.

in Fig. 2.12. In this figure, the normalized SNR of the two users are respectively 5dB and 15dB. The dotted line shows the Gaussian capacity without a modulation constraint. The red curve shows the modulation-constrained capacity that has been calculated via a variation of the Blahut-Arimoto algorithm. The achievable rate of the bit-additive multilevel coding is shown with the green plot, which is obtained by a full-search optimization for rate-allocation, potentially yielding a solution where each user's data is transmitted at all levels. The result of pragmatic rate allocation is shown with the blue plot, which is indistinguishable from the fully optimal rate allocation.

As noted earlier, the pragmatic rate allocation will result in a solution where most of the layers are allocated to one user or another, and potentially one level sees the data of both users. This will results in a solution that is shown in Fig. 2.13.

To summarize the developments so far: a pragmatic near-optimal rate allocation algorithm is being developed to allow the implementation of superposition coding in practical applications. So far, it was shown that the overall rate utility function as well as the constraints can be decomposed to level-wise utility and constraint functions that are minimally coupled (only through the shared parameter θ). The main remaining computational aspect is the calculation of the level-wise constraints. Fortunately, the affine approximation allows us to characterize the level-wise constraints via their two end-points, and the insensitivity of each constraint to other levels' parameters allows us to obtain these end points from the single-user mutual information curves of multilevel modulations. We produce in Fig.2.14 a series of such curves for PAM, PSK, and QAM type modulations. These curves may be pre-calculated and stored via lookup tables. Then the rate constraints at each level may be obtained by reading the values off these curves at the respective SNRs for the two channels.

2.3.4 Exceptions to the Decoupling of Bit-level Rate Constraints

The performance of the proposed rate allocation algorithm is virtually indistinguishable from optimal for many practical cases including many familiar modulations under natural

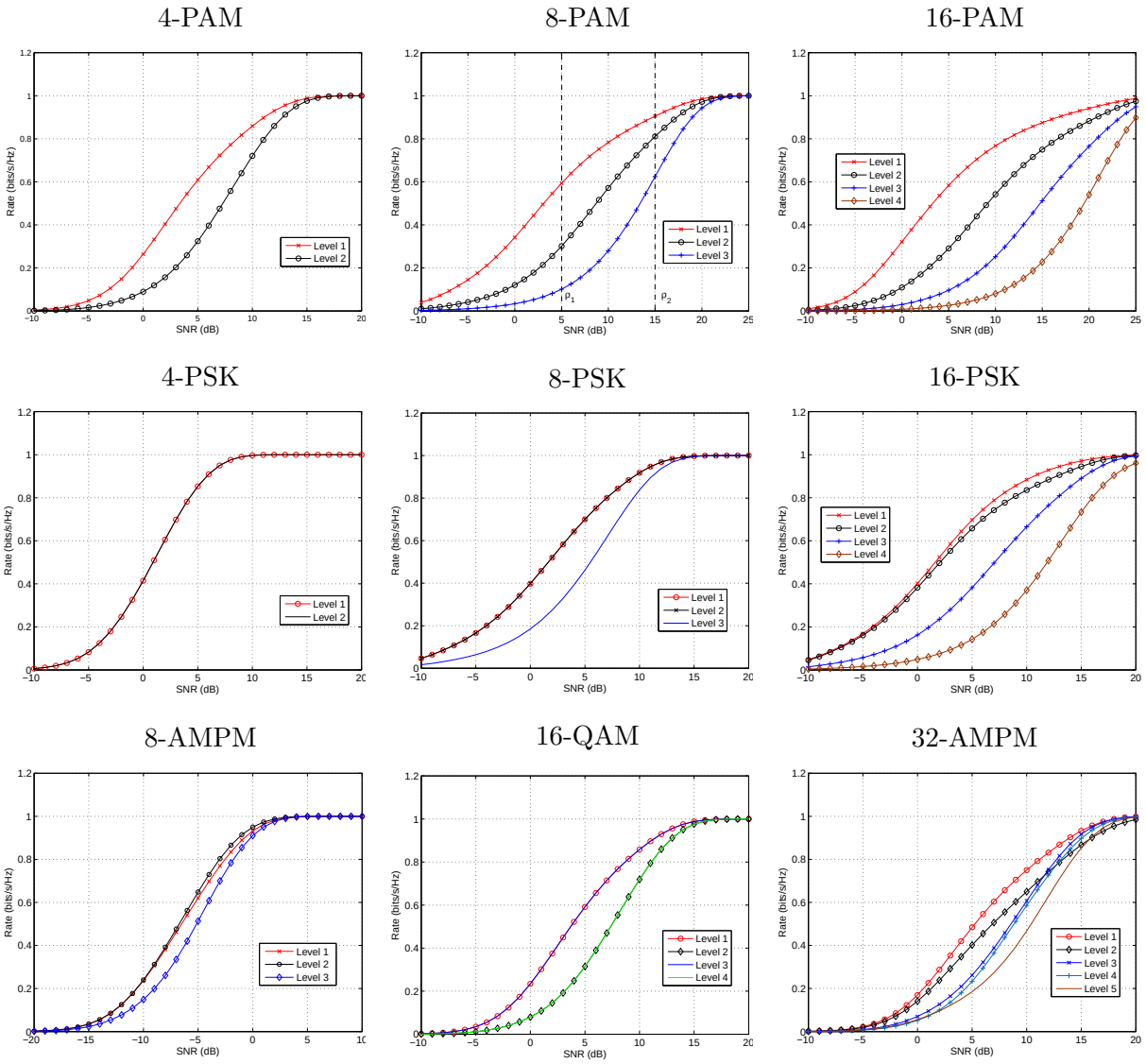

Figure 2.14. Single-user MLC mutual information curves for a variety of PAM, PSK and QAM-type constellations with natural mapping. MLC mutual information depends on decoding order, which in the case of these curves has been from the most to least significant bit of the modulation mapping. The broadcast users "see" such channels at respective operating points ρ_1 and ρ_2.

and Gray mapping. The excellent performance was explained via the insensitivity of the bit-level rate constraints to the operating point in the other bit-levels. A key remaining question is: how prevalent is this insensitivity (decoupling) condition, and what is the performance penalty of the proposed algorithm when this condition does not hold? To our experience, counter-examples to this insensitivity condition are very rare and involve irregular mappings or constellations. As an example, we offer a Gray-like mapping for 8-PAM as shown in Fig. 2.15.

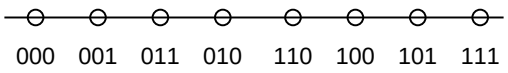

Figure 2.15. 8-PAM constellation with Gray-like mapping.

The sensitivity of the bit-level broadcast rate constraints for this modulation are demonstrated in Fig. 2.16. It is observed that unlike the previous cases, the bit-level constraint of level 3 is sensitive to the bit-level constraint in level 1. This sensitivity manifests itself in a (slight) sub-optimality of the pragmatic rate allocation technique introduced in the previous subsection. Despite the apparent sensitivity, the resulting sub-optimality is slight and is demonstrated in Fig. 2.17.

Of course an example does not make a general case, therefore in the interest of completeness, we outline in the remainder of this subsection a relaxation method can be used for allocating each level's rates to the two users, with no pre-determined constraints on the outcome of the rate allocation. Although it is our understanding that the previous subsection's pragmatic method should be sufficient for almost all practical cases.

The desired solution can be characterized in the form of two vectors $\boldsymbol{R}_1, \boldsymbol{R}_2$ whose components carry the components of the rates in individual levels dedicated to User 1 and User 2.

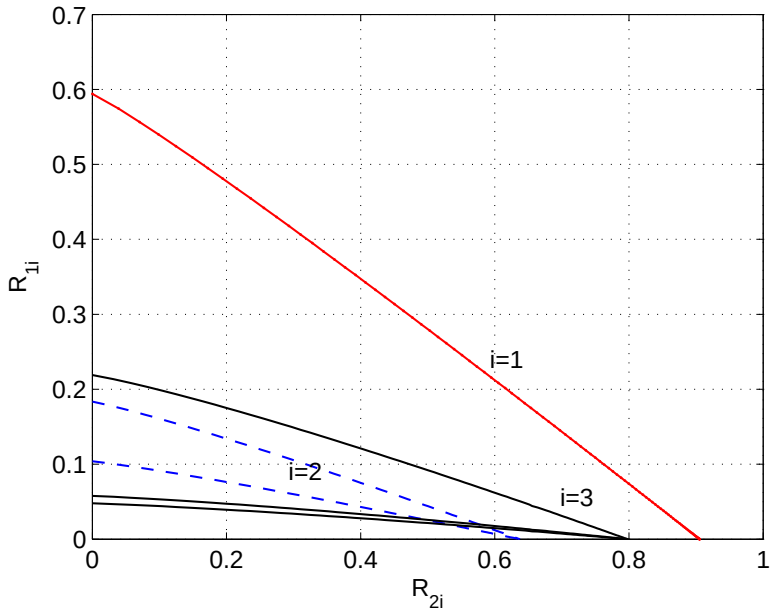

Figure 2.16. Bit-level rate constraints for the Gray-like mapping of Fig 2.15.

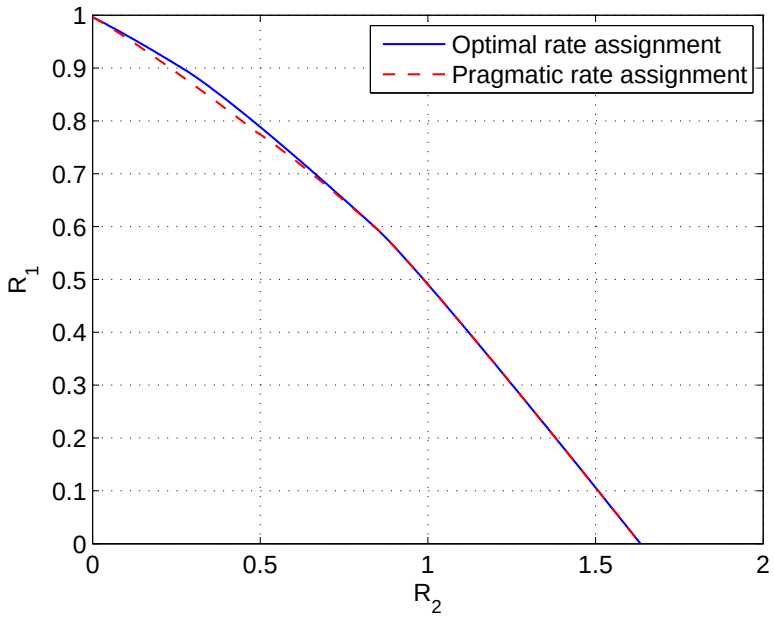

Figure 2.17. Transmission rate using the general optimization versus the efficient optimization.

One way to think about solving this optimization problem is as follows. First, we assign all the rate to one of the receivers (without loss of generality receiver 2), such that

$$\boldsymbol{R}_1 = [0 \dots 0]$$

$$\boldsymbol{R}_2 = [C_{21} \dots C_{2m}]$$

where C_{1i} and C_{2i} denote the point-to-point capacity of level-i for the weak receiver and the strong receiver respectively.

In order to move on the boundary of the capacity region so that receiver 1 is assigned a portion of the rate, each step should maximize the gain in R_1 while maintaining minimum loss to R_2.

This can be done by incrementing one of the entries of $\boldsymbol{R}_1$, i.e., increasing R_{1i} for some i. However, the corresponding loss in R_{2i} depends on the bit constraint of level i. Thus, it is reasonable to increment R_1 through level i that provides maximum gain in R_1 given a fixed loss in R_2. The remaining task is finding a plausible choice of level i as follows. First the bit-level constraint for each level i and its slope denoted by $\bar{f}_i$ are calculated at the current rate assignment. Note that $\bar{f}_i$ represents the gain in R_{1i} normalized to the loss in R_{2i}. The level i^* that results in the maximum gain in R_1 satisfies

$$|\bar{f}_{i^*}| > |\bar{f}_j| \quad \forall_j. \tag{2.28}$$

Therefore, moving close to the boundary of the capacity region can be realized by increasing R_1 through increasing R_{1i^*} and fixing R_{1j} $\forall j \neq i^*$ until either R_{1i^*}reaches its maximum value C_{1i^*} or the inequality (2.28) is violated. In either case, the same procedure is then repeated until the desired rate pair is achieved.

2.3.5 Multilevel BICM Construction

BICM is a close relative of MLC in the point-to-point channel, where the bits from multiple levels are encoded using not only the same code rate, but together as one code word. In

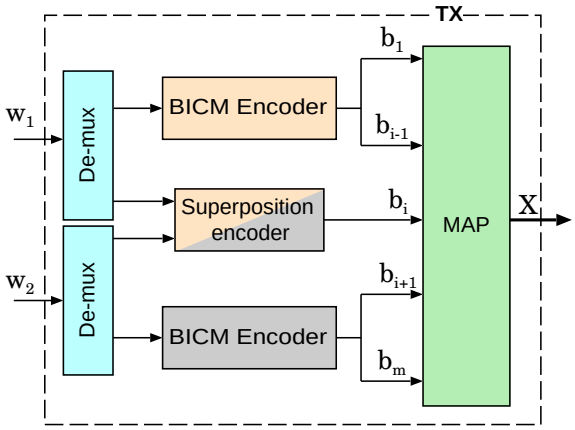

Figure 2.18. Hybrid MLC-BICM superposition

our proposed multilevel superposition coding with the efficient structure shown in Fig. 2.13, there are m encoders: some of them carry information for the weak receiver, some of them carry information for the strong receiver and at most one encoder that carries information for both receivers. We propose to combine all the encoders that carry information for a certain receiver in one BICM encoder as shown in Fig. 2.18. This way of transmission reduces the number of encoders significantly especially for big constellations. For example, for a 64-QAM constellation, the multilevel coding structure will require at least six encoders and by combining all the encoders that send to the same receiver into one BICM encoder, the number of encoders can be reduced to at most three encoders but with longer block length. We call this technique the hybrid technique since it uses multilevel coding in the sense of encoding the information independently and BICM encoder to encode the information that belong to the same receiver.

The rate of the BICM encoder and the serial to parallel conversion depends on the number of levels that the encoder feeds. The rate achieved by the hybrid transmission is shown in Fig. 2.19 for Gray and natural mappings. The achievable rate region of the hybrid transmission is in general smaller than the achievable rate region of the multilevel coding scheme since BICM is not capacity achieving. The maximum loss in rate is the point-to-point transmission since the encoding becomes completely point-to-point BICM encoding;

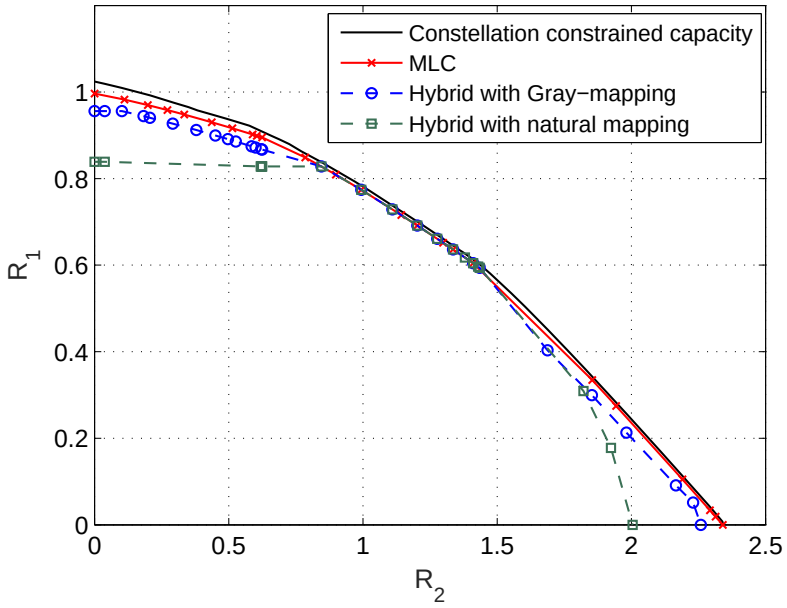

Figure 2.19. MLC and hybrid superposition achievable rates under 8-PAM, $\rho_1 = 5dB$, $\rho_2 = 15dB$.

however, when the rates of the weak and the strong receivers are not equal to zero, the transmission becomes closer to the multilevel superposition transmission. For example for the 8-PAM constellation, there is a stage in which the MLC and Hybrid schemes will be the same. This is the point when the level that carries information for both receivers is the middle level.

2.4 Simulations

Because the broadcast channel involves simultaneously two rates and two SNRs, error plots are generated for the broadcast channel by applying slight modifications to the standard methods used for plotting errors in point-to-point coding literature. For broadcasting the relative quality of the channels, indicated by the noise variances, remains fixed in the simulations, while the transmit power is allowed to increase. The rate of the two codes is chosen according to a rate pair on the boundary of the capacity region. In each plot, the value of

the transmit power corresponding to the capacity rate pair is clearly marked, a point that is the counterpart to the "capacity threshold" in the single-user error curves seen in the coding literature. A comparison between this point and the waterfall region of the error curves is an indicator of how far from optimality is the system operating.

The DVB-S2 LDPC codes are used as component codes for each of the levels to examine the performance of the proposed MLC and the hybrid (MLC-BICM) transmissions. The block length of the codes is $n = 64$k. Fig. 2.20 shows the performance of 4-PAM MLC superposition for rates $(R_2 = 0.5, R_1 = 0.6)$ with natural mapping. The information of the weak receiver is sent over level-1 and the information of the strong receiver is sent over level-2. This is considered an extreme case where each level is assigned to either the weak or the strong receiver. The bit error rate (BER) and frame error rate (FER) for each receiver are shown. The gap to capacity is approximately 0.5-dB at 10^{-5} FER, which is the similar to the gap to capacity of the DVB code in the point-to-point channel, thus suggesting that the FER gap is mostly due to the limitations of the code as opposed to the MLC.

Fig. 2.21 shows the performance of 8-PAM constellation where one bit level is shared between the weak and the strong receiver. The rates assigned are $R_1 = 0.6$ and $R_2 = 1.4$. Level-1 carries information only for the weak receiver, level-2 is shared, and level-3 carries information only for the strong receiver. In the shared level, the weak and the strong receivers messages are encoded independently using the DVB-S2 LDPC codes and combined after setting some bits of the strong receiver codeword to zeros as described in Section 2.3.1.

Fig. 2.22 shows the BER and FER of the proposed hybrid MLC-BICM (Fig. 2.18) transmission compared with the MLC transmission (Fig. 2.13) for an 8-PAM constellation with Gray mapping. Level-1 carries information for the weak receiver and the other two levels carry information for the strong receiver. The rates are $R_1 = 0.5$ and $R_2 = 1.5$. In the hybrid transmission, a BICM encoder is used with double the length of the one used in level-1 and the output of the BICM encoder is partitioned into two streams and fed to the two least

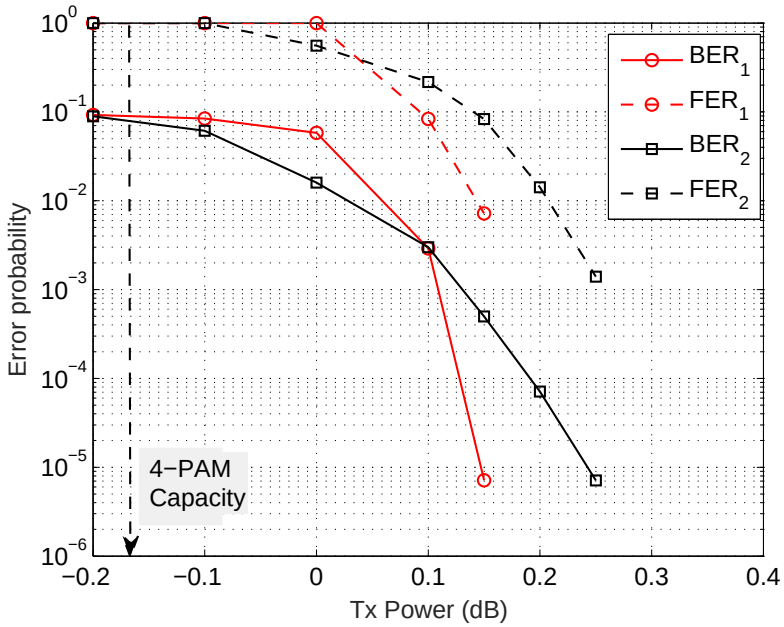

Figure 2.20. Performance of Multilevel superposition for 4-PAM constellation where $\sigma_1^2 = .48$, $\sigma_2^2 = .13$

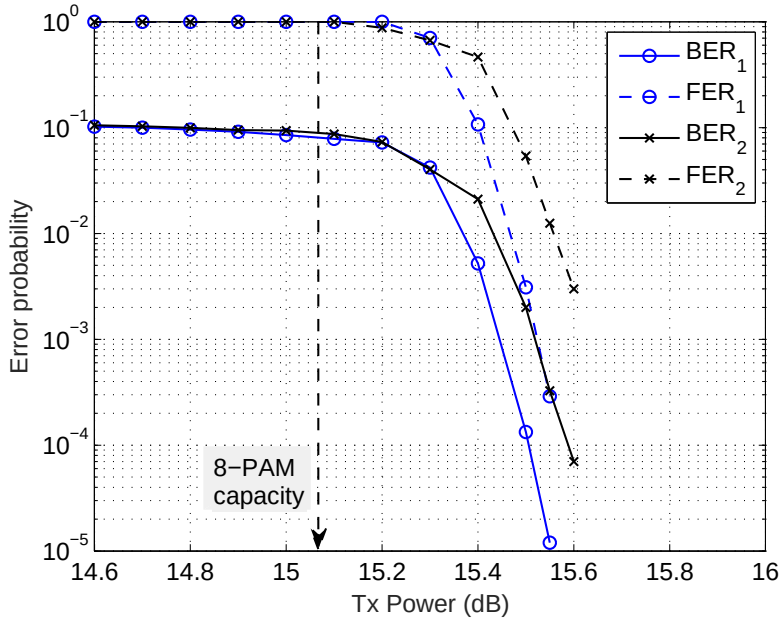

Figure 2.21. Performance of Multilevel superposition for 8-PAM constellation where $\sigma_1^2 = 8.5$, $\sigma_2^2 = 1$

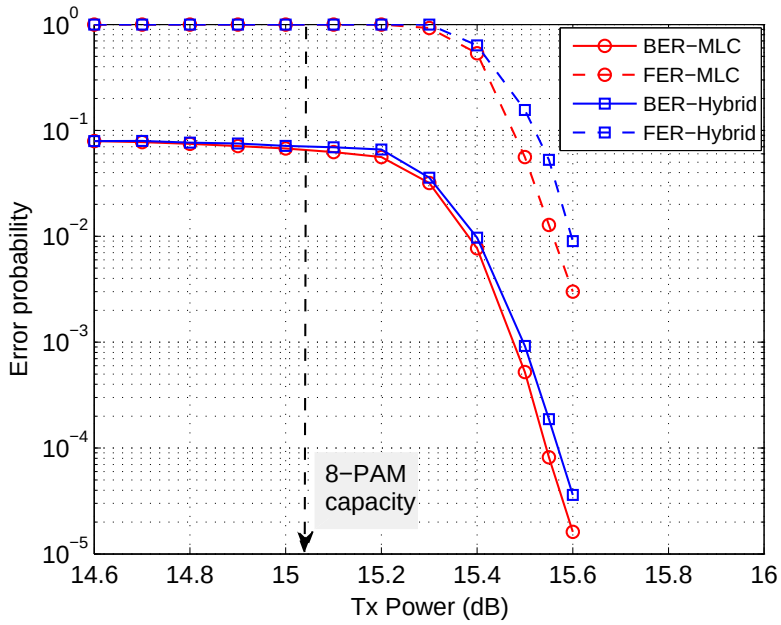

Figure 2.22. Performance of the hybrid MLC-BICM scheme for 8-PAM constellation where $\sigma_1^2 = 8.5$, $\sigma_2^2 = 1$

significant bits. Simulation show that the hybrid scheme has a performance very close to that of MLC.

Fig. 2.23 shows the error performance of 8-PSK constellation with natural mapping where level-1 carries information for the weak receiver, level-3 carries information for the strong receiver and level-2 carries information for both receivers. The rates are $R_1 = 0.4$ and $R_2 = 1.6$. The gap to capacity is around 0.5-dB at bit error probability of 10^{-5}.

Fig. 2.24 shows the performance of 16-QAM constellation with natural labeling where level-1 carries information for the weak receiver, level-2 for both receivers, and levels 3 and 4 carry information for the strong receiver. The rates are $R_1 = 1.2$ and $R_2 = 1.8$ and noise variances at the two receivers are $\sigma_1^2 = .64$ and $\sigma_2^2 = .18$. The simulations show that the proposed scheme has a gap of around 0.4-dB from the constellation constrained capacity at bit error probability of 10^{-5}. The figure also shows the performance of the Hybrid MLC-

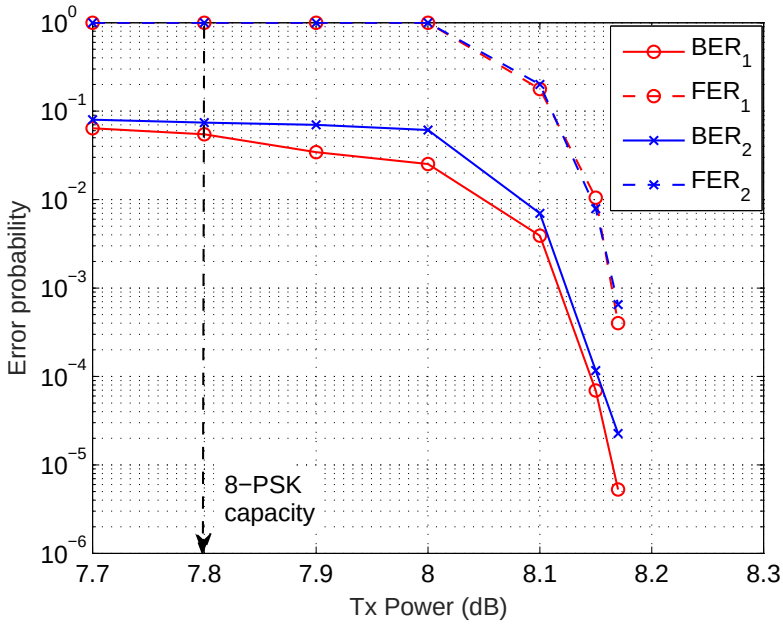

Figure 2.23. Performance of the MLC proposed transmission for 8-PSK constellation where $\sigma_1^2 = 2.2$, $\sigma_2^2 = 1$

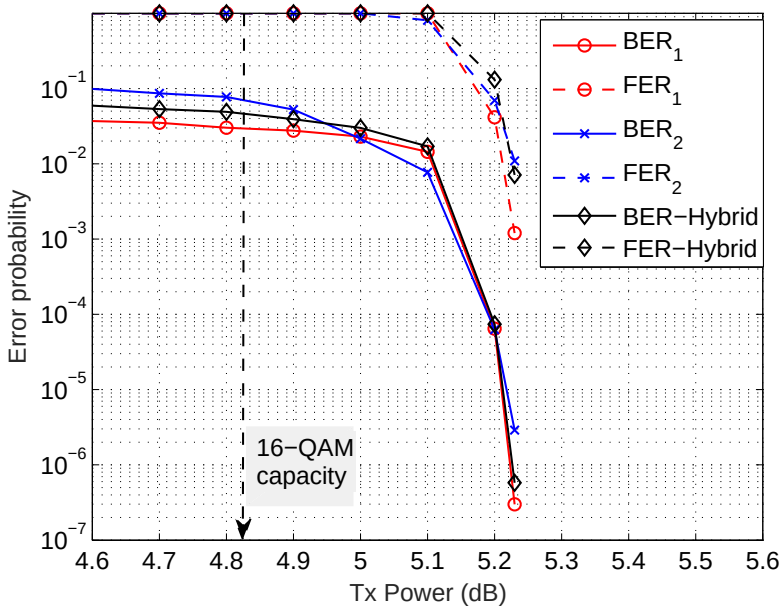

Figure 2.24. Performance of the MLC proposed transmission and the Hybrid MLC-BICM transmission for 16-QAM constellation where $\sigma_1^2 = .64$, $\sigma_2^2 = .18$

BICM transmission where the two encoders of the two least significant bits are combined in one BICM encoder while using Gray mapping.

2.5 Conclusion

This chapter studied coded modulation for the AWGN broadcast channel. multilevel coding (MLC) and bit-interleaved coded modulation (BICM) are explored under channel-input modulation constraints. It was shown that the assignment of receivers information to distinct inputs to the mapper does not approach the capacity uniformly. A bit-wise multilevel superposition transmission is proposed. Furthermore, a hybrid MLC-BICM with lower complexity is proposed. The achievable rate region of the proposed transmission is very close to the boundary of the constellation constrained capacity of the broadcast channel. Simulation results showed an excellent performance using good point-to-point codes.

2.6 Appendix

2.6.1 Degradedness of bit channels

Consider the following Markov process due to the degradedness of the channel

$$U \to X \to Y_2 \to Y_1$$

U has a multi-digit characterization $[C_1, \ldots, C_m]$.

for a specific value of $C^{i-1} = c^{i-1}$, due to the degradedness of the channel we have

$$I(C_i; Y_1|C^{i-1} = c^{i-1}) \leq I(C_i; Y_2|C^{i-1} = c^{i-1}) \tag{2.29}$$

The mutual information $I(C_i; Y_1|C^{i-1})$ and $I(C_i; Y_2|C^{i-1})$ are

$$I(C_i; Y_1|C^{i-1}) = E_{C^{i-1}}[I(C_i; Y_1|C^{i-1} = c^{i-1})] \tag{2.30}$$

$$I(C_i; Y_2|C^{i-1}) = E_{C^{i-1}}[I(C_i; Y_2|C^{i-1} = c^{i-1})] \tag{2.31}$$

where $E[.]$ is the expectation operation. The expectation operation is a convex combination for all the values that C^{i-1} can take. Since the inequality (2.29) holds for any value of C^{i-1} then it holds for any convex combination of the values of C^{i-1}, therefore:

$$I(C_i; Y_1|C^{i-1}) \leq I(C_i; Y_2|C^{i-1})$$

2.6.2 Multilevel Decomposition of the Outer Code

Consider the auxiliary random variable U representing the message to the weak user. To achieve capacity, the outer code is drawn i.i.d. according to $p_U(u)$. In the following we assume the cardinality $|U| = M$. The objective is to produce multilevel codes whose empirical distribution approaches $p_U(u)$. We now consider an m-dimensional binary vector V whose components are i.i.d. Bernoulli-$\frac{1}{2}$. Equivalently, V can be considered a random variable uniformly distributed over an alphabet size of 2^m. This is the random variable generating the m-level multilevel code. Consider the design of a mapping $U' = f(V)$ so that the random variable U', in distribution, is close to the capacity-maximizing U. We start with:

$$p_U(u) = [p_1 \cdots, p_M]$$

Rounding down each of the probabilities to a multiple of 2^{-m} via $Q(p_i) \triangleq 2^{-m}\lfloor 2^m p_i \rfloor$, and distributing the remaining probability $1 - \sum_i Q(p_i)$ over the first $K \triangleq 2^m(1 - \sum_i Q(p_i))$ components, we arrive at the following probability distribution for U':

$$p_{U'}(i) = \begin{cases} Q(p_i) + 2^{-m} & i \leq K \\ Q(p_i) & i > K \end{cases}$$

Defining $k_i \triangleq 2^m p_{U'}(i)$, the function $f(\cdot)$ given below maps the multilevel binary generator variable V to the (approximate) capacity achieving distribution U':

$$f(j) = \begin{cases} 1 & 1 \leq j < k_1 \\ 2 & k_1 \leq j < k_1 + k_2 \\ \cdots & \\ M & k_1 + \cdots + k_{M-1} \leq j < k_1 + \cdots + k_M \end{cases}$$

In the following, we assume that none of the entries of p_U are zero, and also that m is large enough so that none of the entries of $p_{U'}$ are zero. A sufficient condition is $m > -\log_2 \min_i p_U(i)$.

Now, it is straightforward to bound the divergence between p_U and $p_{U'}$:

$$\begin{aligned} D(p_U||p_{U'}) &= \sum_i P_U(i) \log \frac{P_U(i)}{P_{U'}(i)} \\ &\leq \sum_i P_U(i) \log \frac{P_{U'}(i) + 2^{-m}}{P_{U'}(i)} \\ &\overset{(a)}{\leq} \sum_i P_U(i) \frac{2^{-m}}{P_{U'}(i)} \\ &\overset{(b)}{\leq} M 2^{-m+1} \end{aligned}$$

where (a) follows from $\log(1+x) \leq x$ and (b) follows from $\frac{p_U(i)}{p_{U'}(i)} \leq \frac{p_U(i)}{Q(p_U(i))} \leq 2$.

Therefore, it follows that for a fixed M, by increasing the number of levels m one can very quickly get close to the capacity optimizing distribution.

CHAPTER 3

CODED MODULATION FOR THE FULL-DUPLEX RELAY CHANNEL

3.1 Introduction

Recent advancement in hardware design and signal processing have put full-duplex operation back on the map as a potentially viable alternative [44, 45, 46, 47], and much research is ongoing in the area of full-duplex link implementation [48, 49, 50]. The credit for this resurgence of interest goes to the new research in mitigating the so-called loop-back interference (self-interference) at the full-duplex transmitter, represented by [51, 52, 53, 54] among many others.

Focusing our attention on full-duplex *relays*, we find that while early theoretical results were on full-duplex [55], subsequent coding and signal processing results have concentrated for the most part on half-duplex scenarios, in particular low-SNR (binary) signaling [4, 5, 6]. Exceptions do exist, e.g., lattice codes for the full-duplex relay channel [56] but a nontrivial gap to capacity remains and, in the most general setting, the problem of capacity-approaching coding and modulation for the full-duplex relay channel remains open. We address this problem via multilevel coding, providing well-defined and systematic design principles that lead to near-capacity performance.

The key advantage of multilevel coding [16, 21] is that it uses binary codes whose design is by now very well understood. Moreover, the multiple binary encoders that feed the bit-levels of the modulation can operate independently, but for optimal performance some coupling between the bit-level decoders is necessary, e.g., successive decoding.

Further related results in the relay literature are as follows. Several contributions for the bandwidth limited relay channel focused on the two way relay channel. Ravindran et. al [57] studied LDPC codes with higher order modulation for the two way relay channel. Chen and Liu [58] analyzed different coded modulation transmissions for the two way relay

channel. Chen et. al [59] studied multilevel coding in the two-way relay channel. Hern and Narayanan [31] studied multilevel coding in the context of compute-and-forward. However, the two-way relay channel does not consider the direct link like in the conventional relay channel, and hence, the coded modulation techniques that are considered in the literature cannot be used for the three-nodes full-duplex relay channel.

A key contribution of this chapter is, first, to elucidate conditions under which multilevel coding for the relay channel achieves the constellation-constrained capacity. Second, to highlight the challenges involved in meeting this bound. Third, to propose solutions for these challenges, and demonstrate the performance of the proposed solutions. The bit-additive superposition used in this chapter was introduced for the broadcast channel in [60]. A preliminary version of some of the results of this chapter have appeared in a conference [61], and a related paper [62] addresses multilevel coding for the half-duplex relay channel.

We propose a simple multilevel full-duplex relay transmission. The straightforward application of multilevel coding to the relay channel would result in code specifications that require multiple inter-layer correlations between the source and relay codes. Our work produces a streamlined coding procedure where the dependencies are limited to pairwise correlation between the source/relay codes at each individual layer. Moreover, we provide a simple implementation of this idea via a binary addition between conventionally designed codes. Numerical results show that the performance of the proposed technique is almost as good as the best known decode-and-forward performance (with Gaussian codewords). We show that linearity of the source-to-relay code may impose a performance penalty. We propose a solution that minimizes this performance penalty using a proper labeling design. The error exponent of the proposed transmission is studied under sliding window decoding. Simulation results show that good point-to-point codes (DVB-S2 codes) produce performance that is very close to the fundamental limits when used in the proposed transmission. In addition, two methods are experimentally verified for directly approaching the performance

of non-linear codes in full-duplex relays: insertion of randomly located zeros into DVB-S2 codewords (using pseudo-random generators whose seed is known at source and destination), and inserting zeros at fixed locations that are determined via a puncture optimization strategy [63], resulting in a degenerate linear code. The relative performance of the two methods is discussed.

3.2 Preliminaries

Although the capacity of the full-duplex relay channel is in general unknown, we know the rates supported by several specific transmission schemes, including decode-and-forward which achieves the capacity of the degraded relay channel, partial decode-and-forward and compress-and-forward. In this chapter we consider only the decode-and-forward transmission.

The decode-and-forward transmission uses block Markov encoding. Throughout the chapter we denote the signal transmitted from the source node and the relay node in block t by $X_1^{(t)}$ and $X_2^{(t)}$. We begin by modeling the received signal at the relay, which experiences self-interference:

$$Y_2^{(t)} = H_{12}X_1^{(t)} + n_2 + n_s$$

where H_{12} is the channel from the source to the relay, n_2 is the additive Gaussian (thermal) noise at the receiver, and n_s is the sampled residual self-interference. The area of modeling and analyzing loop-back or self-interference has experienced rapid growth in the past few years. Several methods for mitigating self-interference are now in place, among them antenna design and placement (including passive components), as well as echo cancellation in the amplifier stage, as well as digital signal processing after down-conversion and sampling [51]. The collection of these methods have allowed the residual self-interference to be reduced significantly. The *residue* of self-interference, n_s, is the component that is seen by the relay decoder. Several works to date [51, 64, 65] have used a Gaussian model for

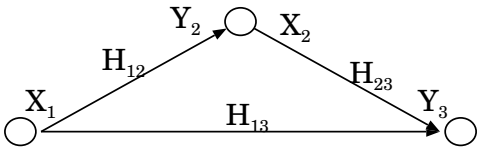

Figure 3.1. Full-Duplex relay channel.

n_s, an approximation that is confirmed by various measurements [66, 67]. Therefore, the combination $\tilde{n}_2 = n_2 + n_s$ is also Gaussian with appropriate variance.

Thus, the received signal at the relay and destination in block t are respectively given by

$$Y_2^{(t)} = H_{12}X_1^{(t)} + \tilde{n}_2 \tag{3.1}$$

$$Y_3^{(t)} = H_{13}X_1^{(t)} + H_{23}X_2^{(t)} + n_3 \tag{3.2}$$

where H_{12}, H_{13} and H_{23} are the fading channel coefficients as illustrated in Fig. 3.1.

The destination uses either backward decoding where the destination waits until the reception of the last transmission block or a sliding window decoder where the decoder uses L blocks for decoding where L is the window size.

3.3 Multilevel Decode and Forward

Subject to the channel probability distribution $P_{Y_2,Y_3|X_1,X_2}(y_2, y_3|x_1, x_2)$, the decode-and-forward achievable rate is

$$R \leq \max_{P_{X_1,X_2}(x_1,x_2)} \min\{I(X_1; Y_2|X_2), I(X_1, X_2; Y_3)\} \tag{3.3}$$

where the channel coefficients are implicitly included in the expression for any ergodic channel as follows:

$$I(X_1; Y_2|X_2) = \mathbb{E}_{H_{12},H_{23},H_{13}}[I(X_1; Y_2|X_2, H_{12}, H_{13}, H_{23})]$$

$$I(X_1, X_2; Y_3) = \mathbb{E}_{H_{12},H_{23},H_{13}}[I(X_1, X_2; Y_3|H_{12}, H_{13}, H_{23})]$$

The design variable of this optimization problem is the joint distribution $P_{X_1,X_2}(x_1, x_2)$. Assuming that $|X_i|$ is the cardinality of the variable X_i, this joint distribution can be represented by a matrix with dimensions $|X_1| \times |X_2|$. The element in row j and column k of this matrix is $P_{X_1,X_2}(x_j, x_k)$. This optimization problem is hard to solve specially when the cardinalities $|X_1|$ and $|X_2|$ are large. Moreover, it leaves open the question of a *practical* encoding with codebook that meets or approximates this distribution. In this section, we address the optimization of codebook distributions in the context of multilevel coding, and also examine its consequences on the decoder side.

3.3.1 Encoding

For ease of exposition we consider the case where the source and the relay multilevel codes have the same number of levels m, a restriction that does not lead to any loss in generality as described in Remark. 5. As shown in Fig. 3.2, the signals X_1, X_2 at the source and the relay respectively are represented by their modulation-constrained index variables $B^m = [B_1, \ldots, B_m]$ and $C^m = [C_1, \ldots, C_m]$ respectively; The relay and the source can use different sets of encoders. The source uses block-Markov superposition, therefore C^m and B^m are dependent. This dependence can be shown in Fig. 3.2 through the delay operation Z^{-nR} which is a delay of one transmission block. The two inputs of each encoder at the source are the current block message and the previous block message which is assumed to be known at the relay after successful decoding in the previous block. The two messages are encoded jointly using a generic encoder defined over a finite field. A special form of this generic encoder is shown in Fig. 3.3. The rate in (3.3) is equivalent to

$$R \leq \max_{P_{B^m,C^m}(b^m,c^m)} \min\{I(B^m; Y_2|C^m), I(B^m, C^m; Y_3)\} \tag{3.4}$$

The design variable $P_{B^m,C^m}(b^m, c^m)$ implies that the vectors B^m and C^m can be generated with any joint distribution which implies any dependency between B_is and C_is. Multilevel

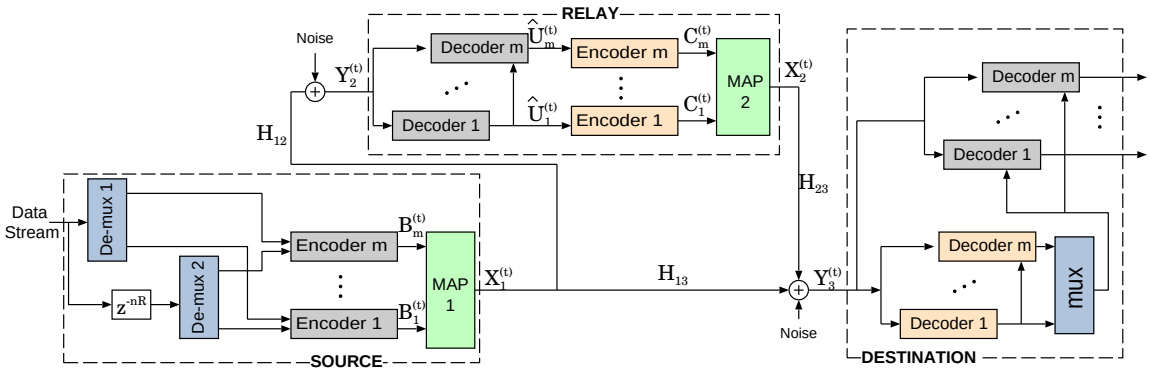

Figure 3.2. MLC and MSD in the Relay channel with regular successive decoding

coding introduces an additional constraint: that B_is should be encoded independently and C_is should be also encoded independently. However, the dependency between B^m and C^m is necessary for the superposition coding. This independence between the entries of B^m and C^m introduces a constraint on the optimization, resulting in the following rate:

$$R \leq \max_{\prod_{i=1}^{m} P_{B_i|C^m}(b_i|c^m)P_{C_i}(c_i)} \min\{I(B^m; Y_2|C^m), I(B^m, C^m; Y_3)\} \tag{3.5}$$

Multilevel coding is optimal if the new constraint is not active, i.e., if the unconstrained optimization already satisfies the constraint:

$$P^*_{B^m,C^m}(b^m, c^m) = \prod_{i=1}^{m} P^*_{B_i|C^m}(b_i|c^m)P^*_{C_i}(c_i) \tag{3.6}$$

where $P^*()$ is the optimal distribution.

So far we borrowed ideas from the point-to-point channel [68], but this is not enough to produce a multilevel scheme in the usual sense for the relay channel, because the cross-dependence of the source and relay transmissions still binds the source streams together. In other words, the source streams up to this point are only *conditionally* independent. We now proceed to address this issue via a framework allowing each level of the source signal to depend on the relay signal *only at the same level*, i.e., allowing each B_i to depend only on C_i. Then the achievable rate is

$$R \leq \max_{\prod_{i=1}^{m} P_{B_i|C_i}(b_i|c_i)P_{C_i}(c_i)} \min\{I(B^m; Y_2|C^m), I(B^m, C^m; Y_3)\} \tag{3.7}$$

A sufficient condition for this to be capacity optimal is:

$$P^*_{B_i|C^m}(b_i|c^m) = P^*_{B_i|C_i}(b_i|c_i) \qquad \forall i \tag{3.8}$$

It remains an open question exactly which channels and which modulations satisfy this sufficient condition. However, in this work we show via numerical results that this approach produces rates that are close to the constellation constrained capacity.

Remark 3. *For generality, the mutual information expressions in this section do not show explicit dependence on channel statistics. For additive Gaussian channels, Y_2 and Y_3 depend on the input variables via AWGN. In a pure line-of-sight model, the dependence is via a path loss exponent and AWGN. We consider first a path loss model with AWGN to explain the main ideas of the proposed work while a generalization of our work to the slow fading and fast fading cases are studied in the sequel. In a rich scattering (Rayleigh) model, the mutual information expressions also involve integration over fading states. Both the path loss and the Rayleigh model make an appearance in the sequel.*

Remark 4. *Coded modulation for the relay is attempting to implement a Gaussian codebook, which for the decode-and-forward consists of a superposition whose cloud centers are the relay codebook, and the satellites are the source codebook. The cloud centers are transmitted cooperatively to the destination. The satellite codewords (conditioned on the cloud center) send the relay the new information for the next transmission block. To implement this cooperative transmission, the source and the relay may use either the same modulation or two modulations from the same family (for example 16QAM and 64QAM).*

Remark 5. *The expressions above where developed for identical modulation constellation at the source and the relay. These expressions can be modified without difficulty to apply to two different modulations of the same modulation family by forcing certain B_i or C_i to be trivial random variables (constant).*

3.3.2 Multistage Decoding

Multistage decoding is simpler than joint decoding and is optimal in the point-to-point channel [68]. To investigate this issue in the relay channel, we focus on the decoding requirement at both the relay and the destination. For relay decoding, we must have at each level i:

$$R_i \leq I(B_i; Y_2|B^{i-1}, C^m) \tag{3.9}$$

So the relay is able to do multistage decoding in a straightforward matter. At the destination, the multistage decoding depends on the two possible relaying strategies [55]: in the first strategy, the relay transmits a hash at a rate supported by the relay-destination link (with partial interference from source considered as noise). The destination first decodes the hash and then the overall received signal is decoded with the help of the hash. In this case, the destination successively decodes the relay signal and then the source signal (Fig. 3.2) which requires the rates to satisfy:

$$R_i \leq I(B_i; Y_3|B^{i-1}, C^m) \tag{3.10}$$

$$R_{ri} \leq I(C_i; Y_3|C^{i-1}) \tag{3.11}$$

where R_{ri} is the rate of level i at the relay. Combining the rate constraints we obtain

$$\begin{aligned} R \leq & \max_{\prod_{i=1}^m P_{B_i|C_i}(b_i|c_i)P_{C_i}(c_i)} \min\{\sum_{i=1}^m I(B_i; Y_2|C^m, B^{i-1}), \\ & \sum_{i=1}^m I(C_i; Y_3|C^{i-1}) + I(B_i; Y_3|B^{i-1}, C^m)\} \end{aligned} \tag{3.12}$$

In the second strategy, the relay codebook has rate that may be above the capacity of the relay-destination link, but is still decodable at the destination when joined with the source signal. The multistage version of this joint decoding is shown in Fig. 3.3 and requires the individual levels to obey the following rate constraints:

$$R \leq \max_{\prod_{i=1}^m P_{B_i|C_i}(b_i|c_i)P_{C_i}(c_i)} \min\{\sum_{i=1}^m I(B_i; Y_2|C^m, B^{i-1}),$$

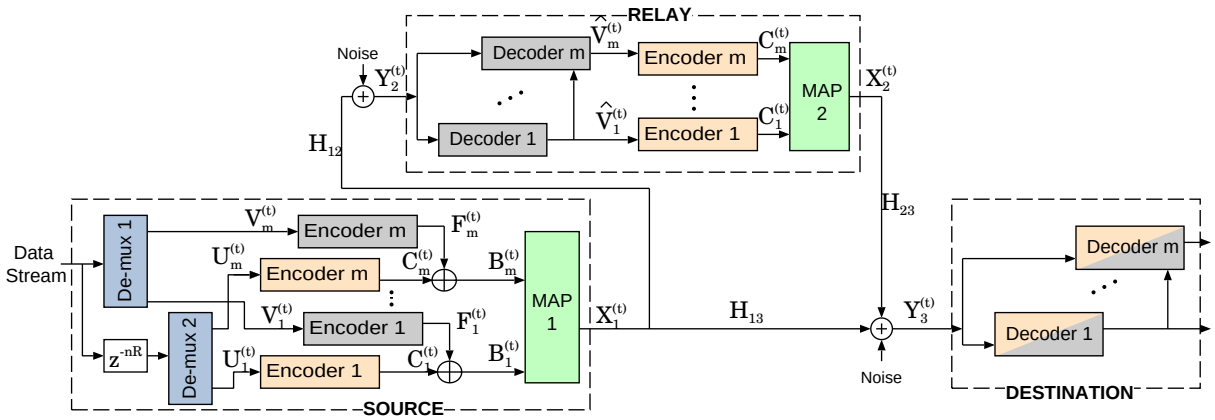

Figure 3.3. MLC and MSD in the Relay channel with level by level decoding

$$\sum_{i=1}^{m} I(B_i, C_i; Y_3 | B^{i-1}, C^{i-1})\} \tag{3.13}$$

Both (3.12) and (3.13) result in the same overall rate. However, level-wise rate allocations will be different according to the different strategies.

3.4 Code Design

Fig. 3.3 shows a block diagram of multilevel encoders and multistage decoders according to the principles outlined in the previous sections. The data is fed into the encoder in blocks of size k. Each block-Markov transmission is dependent on two successive data blocks. These two data blocks (the present and the past) are demultiplexed into levels V_i and U_i, respectively. At each level i, the two data components are encoded via superposition coding (not necessarily with XOR operation as shown in Fig. 3.3) to produce the mapping indices B_i. $\hat{V}_i$ represents the relay's estimate of V_i which is correct under decode-and-forward. C_i is the level-i relay codeword, whose data word U_i is known via relay reception at time $t-1$, i.e., $V_i^{(t)} = U_i^{(t-1)}$.

3.4.1 Bit-Additive Superposition

For superposition we propose to use a modulo-2 addition of constituent binary codes for each level, see [1, Chapter 5] and [38]. The result is shown in Fig. 3.3, where for each level i the demultiplexed data streams U_i and V_i are separately encoded into C_i and F_i, respectively, and then the input to the modulation mapper is obtained by $B_i = C_i \oplus F_i$. The achievable rates under this condition can be characterized by:

$$R \leq \max_{\prod_{i=1}^{m} P_{B_i|C_i}(b_i|c_i)P_{C_i}(c_i)} \min\{\sum_{i=1}^{m} I(B_i; Y_2|C^m, B^{i-1}), \sum_{i=1}^{m} I(B_i, C_i; Y_3|B^{i-1}, C^{i-1})\}$$
$$\text{subject to } P_{B_i|C_i}(b_i|c_i) = P_{B_i|C_i}(\bar{b}_i|\bar{c}_i) \tag{3.14}$$

The constraint $B_i = C_i \oplus F_i$ for some Bernoulli random variable F_i is equivalent to the constraint $P_{B_i|C_i}(b_i|c_i) = P_{B_i|C_i}(\bar{b}_i|\bar{c}_i)$ on the distribution of B_i, C_i, where $\bar{b}_i$ denotes the logical complement of b_i. Clearly this is a restrictive constraint as it reduces the degrees of freedom in the joint distribution of B_i, C_i. However, as will be shown in the sequel, this superposition structure does not induce a rate penalty.

Subsequently, we introduce a linearity constraint on the code with code bits C_i. Subject to this new constraint, the achievable rate will be:

$$R \leq \max_{\prod_{i=1}^{m} P_{B_i|C_i}(b_i|c_i)P_{C_i}(c_i)} \min\{\sum_{i=1}^{m} I(B_i; Y_2|C^m, B^{i-1}), \sum_{i=1}^{m} I(B_i, C_i; Y_3|B^{i-1}, C^{i-1})\}$$
$$\text{subject to } P_{B_i|C_i}(b_i|c_i) = P_{B_i|C_i}(\bar{b}_i|\bar{c}_i)$$
$$P_{C_i}(1) = \frac{1}{2} \tag{3.15}$$

Once again, numerical results show that this new constraint introduces no rate penalty. Finally, we look at the case where both codes F_i, C_i are linear and full-rank.[1] Then the achievable rates are obtained via:

$$R \leq \max_{\prod_{i=1}^{m} P_{B_i|C_i}(b_i|c_i)P_{C_i}(c_i)} \min\{\sum_{i=1}^{m} I(B_i; Y_2|C^m, B^{i-1}), \sum_{i=1}^{m} I(B_i, C_i; Y_3|B^{i-1}, C^{i-1})\}$$

[1] A code is linear when the codewords constitute a vector space.

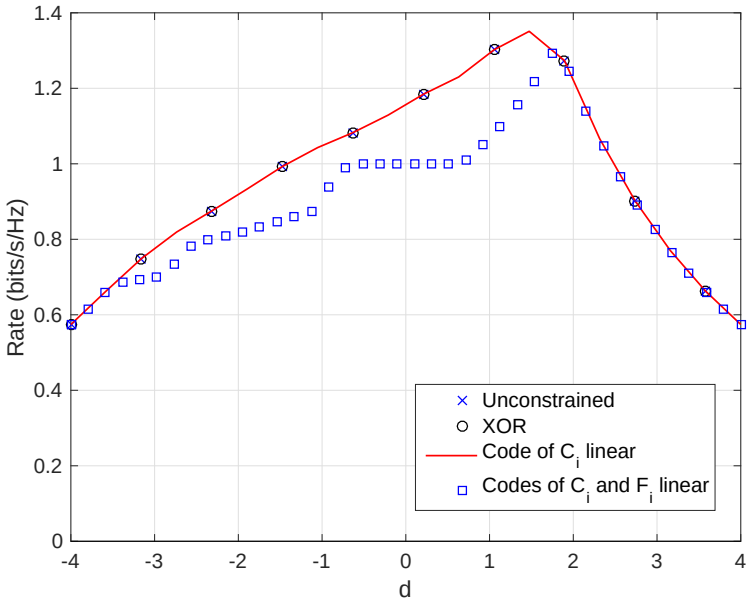

Figure 3.4. Under 4-PAM, neither XOR superposition nor linearity of C_i have a rate penalty, but linearity of F_i has a rate penalty that depends on relay location d where the source is at 0 and the destination is at 4 with a path-loss exponent of 2.

$$\begin{aligned} &\text{subject to } P_{B_i|C_i}(b_i|c_i) = P_{B_i|C_i}(\bar{b}_i|\bar{c}_i) \\ &\qquad P_{C_i}(1) = \frac{1}{2} \\ &\qquad P_{B_i=C_i} \in \{\frac{1}{2}, 1\} \end{aligned} \tag{3.16}$$

If the optimization results in a level i having $b_i = c_i$, it means that level i is only used to help the relay-destination transmission through increasing the correlation between X_1 and X_2, and carries no new information for the relay. Case studies show that Eq. (3.16) may introduce a nontrivial rate penalty compared with (3.15), especially in lower-order modulations (Fig. 3.4).

Remark 6. *We introduced constraints one-by-one to shed light on exactly which one of the practical constraints introduces rate loss. It so happens that both the XOR superposition as well as linearity of the relay code are harmless, but introducing linearity in both codes in certain cases has a cost.*

The behavior of linear codes and XOR superposition structure can be explained as follows: To begin with, assume that the distance between the source and the destination is 4 and that the distance between the source and the relay is d. When d is negative this means that the source node is between the relay and the destination and when d is positive this means that the relay is between the source and destination. In order to simply show the effect of XOR superposition and linearity, assume only a path-loss channel model with path-loss exponent $\alpha = 2$. Fig. 3.4 shows that linearity of the codes induces no rate loss when the relay is close to the destination. These are locations where source-relay link is the bottleneck and therefore the correlation between the source signal and the relay signal is not highly important. Conversely, when the relay is far from the destination and close to the source, the source should help the relay transmission to the destination, and hence, high correlation is required, and in that regime Fig. 3.4 shows linear codes can induce a rate loss, which we explain and analyze below.

The linearity of the binary code implies that the symbols are zero and one with equal probability, except for the trivial all-zero code. When F_i is always equal to zero, level i does not transmit any information to the relay. When F_i is either one or zero with a uniform distribution, B_i is independent of C_i. Therefore, under linear codes each level i can be used for only one of two purposes: either it transmits data to the relay, or it is used to help the relay transmission toward the destination via correlation (beamforming when the channel is known at the transmitter), but not both. So at each level, we must either give up the perfect allocation of rate to the relay, or give up correlation. This tension, which does not exist with nonlinear codes, gives rise to the rate loss in linear codes especially at low-order modulations. Fig. 3.5 shows this phenomenon in 4-PAM where P_1 is the transmission power of the source node and P_2 is the transmission power of the relay node.

It is observed that when the relay is far from the destination, for linear codes one of the levels transmits zero rate to the relay so that it is available for generating correlation with

the relay signal. The zero-rate assignment to some levels in this figure is due to the small constellation size of 4-PAM and because the uniform distribution constraint forces each level to either send new information or do beamforming with the source. Because of interference between the levels, the optimal strategy might abruptly change with small changes in the channel gains.

Discontinuities can be observed in the figure and this is because the 4-PAM constellation does not behave continuously due to the small discrete size. This phenomena is more obvious in the bottom figure and this is because of the added constraint of the uniform distribution. This constraint enforces each level to either send new information to the relay or provides gain to the relay-destination transmission. When the relay moves a small distance, the optimal strategy might change because the achievable rate of each level from the source to the relay change and it might be better to switch the roles of each level while the relay moves.

Remark 7. *When the modulation order is large compared with the capacity of the channel, this effect is much reduced. The reason is that the rate allocated to some layers will be small, therefore it is possible to use those layers purely for correlation without a loss of efficiency for transmission to the relay. This insight will be used subsequently to design labellings that reduce the rate loss.*

3.4.2 Labeling Design For Linear Coding

Linear codes constrain the marginal distributions that can be supported, which may not include (or be close to) the optimum.The idea of this section is to select a modulation labeling whose corresponding (optimal) input distributions is as close to uniform as possible, and therefore are suitable for use with linear codes.

This section is based on the idea that for an optimal code, changing the modulation labeling will (potentially) change the marginal distribution of the (correspondingly optimal)

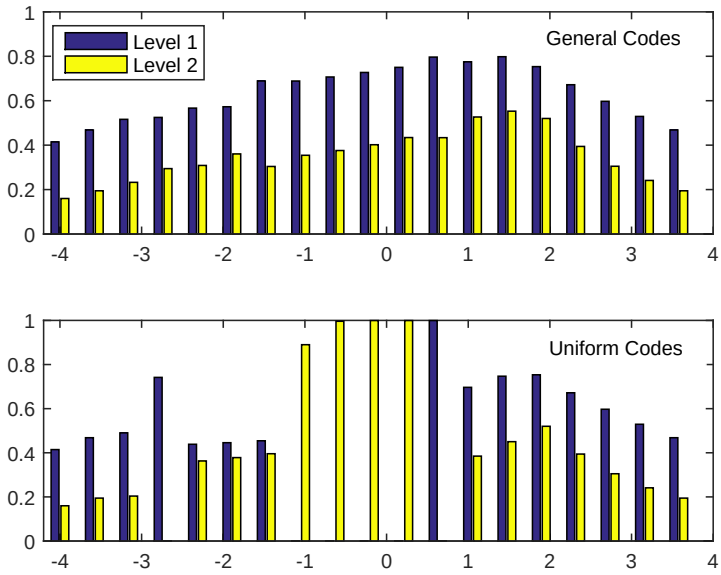

Figure 3.5. 4-PAM relay level-wise rate allocation as a function of relay location for general and uniform codes under natural labeling. Source and destination at $d = 0, 1$ respectively, $P_1 = 10$dB and $P_2 = 10$dB with respect to a reference value of 1.

binary codes that feed the modulation. Among these different modulation labelings, some of them induce optimal input distributions that are closer to those available via linear codes. In this section, we search for and study modulation labelings that are suitable for linear coding.

The bit-additive structure under linear coding admits 2^m different correlations;[2] examples for 4-PAM are shown in Table 3.1 where ρ_1 and ρ_2 are the correlations of the first level and the second level. The table shows the available source-relay correlations if the choices at each modulation are limited to $\rho_i = 1$ or $\rho_i = 0$. The corresponding source-relay rates are shown in Fig. 3.6.

The two parameters in the labeling that determine the total transmission rate are the correlation achieved by each level (if the level is used for correlation) and the source-relay

[2]Because at each level, the bit-additive linear codes can produce correlation zero or one.

Table 3.1. Correlation achieved by linear codes for different labellings

$[\rho_1, \rho_2]$	[0,0]	[0,1]	[1,0]	[1,1]
Natural {00,01,10,11}	0	0.2	0.8	1
Gray {00,01,11,10}	0	0.19	0.79	1
Custom {00,11,10,01}	0	0.41	0.51	1

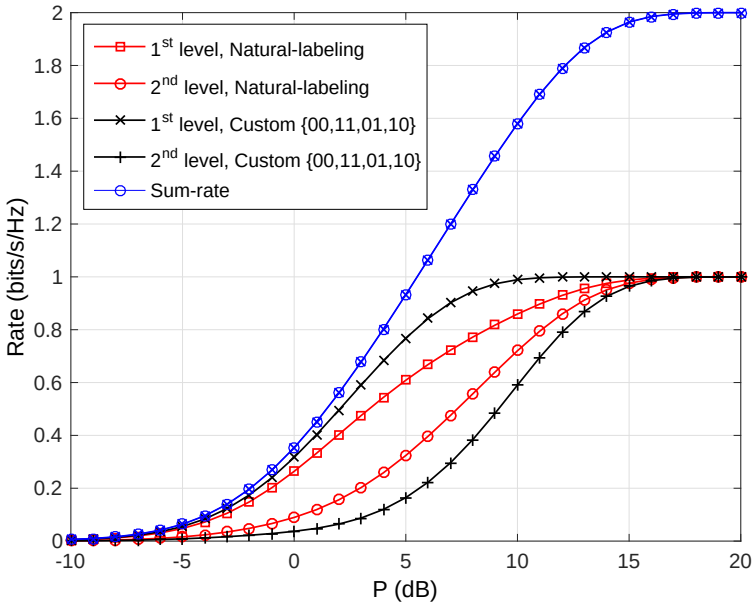

Figure 3.6. The point-to-point achievable rate for 4-PAM under different labellings

rate through each level (if the level is used for sending new information to the relay). For ease of exposition we consider a source-relay code implemented using a 4-PAM modulation, where we are free to use any modulation labeling of our choice. The two parameters discussed earlier are the available point-to-point (source-relay) rate shown in Fig. 3.6 for natural labeling and a custom labeling (00,11,01,10). This shows that different labellings correspond to different mutual information curves for each level under a fixed sum-rate for all labellings. For the custom labeling, the rate for the two levels is separated more than for the natural labeling. The second parameter that affects the rate which is the total source-relay correlations that

can be achieved by all different combinations of bit-level correlations are given in Table. 3.1 for different labellings.

Therefore, if the position of the destination requires a modest amount of beamforming, there are two cases. Firstly, if the source-relay channel is very strong, this means the SNR at the relay node is very high and the achievable rates of both levels for the two different labellings is almost the same. From the perspective of the source-relay rate, the two labellings are equivalent however, the beamforming gain will be different. From Table. 3.1, it is shown that the maximum correlation other than one[3] can be obtained by using natural labeling and assigning the most significant bit for full correlation.

Secondly, as the relay moves far from the source, the SNR value at the relay goes down (which leads to difference in the levels between the natural labeling and the custom labeling) and the required value of the correlation between the source and the relay also goes down. To accommodate source-relay rate, it is better to use the custom labeling and assign the least significant bit for beamforming because it already has a small rate penalty compared to natural labeling. From a beamforming point-of-view, Table. 3.1 shows that assigning the least significant bit of the custom labeling for beamforming will provide higher correlation than that of the natural labeling.

As explained earlier, there are also cases where beamforming is unimportant (e.g., relay very close to destination) in which case either of the labellings will perform the same.

To illustrate the effect of the choice of labeling on the performance of linear codes, we use again the 4-PAM modulation with the three labellings shown in Table 3.1. The throughput of a decode-and-forward relay is optimized subject to these labellings and under a linear coding constraint, with the results shown in Fig. 3.7, assuming the same system model with d_{13}=4.

[3]Maximum correlation between the source and the relay cannot be one because this means that zero rate will be transmitted to the relay node, leading to zero total transmission rate.

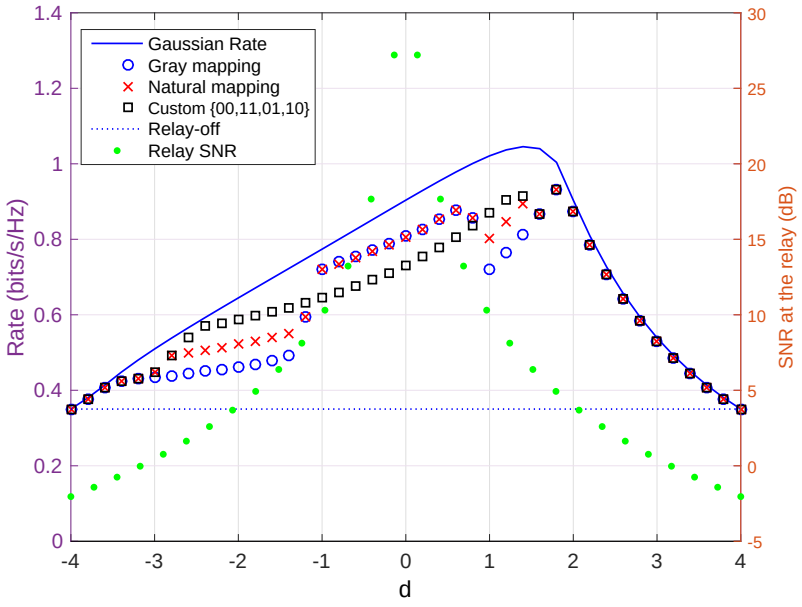

Figure 3.7. Multilevel coding transmission rate for different labellings, $P_1 = P_2 = 10$dB

Several different regions of operation clearly stand out. First, when relay is close to the destination, beamforming is not required and in fact linear coding does not incur a rate penalty. For other source-relay-destination configurations, either a natural labeling or the custom labeling performs best.

Remark 8. *We observe that the Gray labeling is never the best labeling in 4-PAM MLC in the relay channel. This is because the mutual information curves for Gray labeling are exactly the same as natural labeling, however, Gray labeling produces smaller correlation than natural labeling. We also observe that natural and Gray labeling perform very well for* $-1 < d < 1$. *This is because in this setting, the relay is so close to the source which makes the multiple-access phase to be the bottleneck of the transmission. Therefore, high correlation between the source and the relay is required. Table 3.1 shows that natural and Gray labeling can provide higher source-relay correlation.*

Remark 9. *In this Section, it was assumed that the same modulation constellation is used at the source and the relay, including the labeling. A non-identical labeling will interfere with the level-wise beamforming and does not confer any obvious advantages.*

Remark 10. *Optimization of labeling can be performed via exhaustive search for small constellations. For large constellations, as mentioned earlier, the performance penalty of linear codes is vanishingly small (due to availability of a large set of feasible correlation values), therefore any labeling (e.g., natural labeling) works well and there is no need for optimizing the labeling.*

3.4.3 Slow Fading Relay Channel

In this section, we consider that the channel coefficients are fixed over each transmission block and the channel state information is known at the receiver (CSIR). In this case, the information that can be transferred form the source node to the destination node is

$$I = \min\{I(X_1; Y_2|X_2, H_{12}), I(X_1, X_2; Y_3|H_{23}, H_{13})\} \tag{3.17}$$

and the mutual information between level i at the source and level i at the destination is

$$I_i = \min\{I(B_i; Y_2|B^{i-1}, X_2, H_{12}), I(B_i, C_i; Y_3|B^{i-1}, C^{i-1}, H_{23}, H_{13})\} \tag{3.18}$$

Assuming that the transmission rate of level i is R_i, the outage event of level i is $I_i < R_i$. The outage probability is then given by

$$P_{outage} = \bigcup_i Pr(I_i < R_i) \tag{3.19}$$

$$\leq \sum_i Pr(I_i < R_i) \tag{3.20}$$

where the last inequality is from the union bound. Each of the mutual information I_i can be calculated numerically in a similar manner to the curves in Fig. 3.6.

3.4.4 Fast Fading Relay Channel

In this section, we show the applicability of our analysis and design to the Rayleigh fading channel with channel state at the receivers. Assume that the fading coefficient between node i and node j is H_{ij}. The three channel gains are all independent and identically distributed with a normal distribution $\mathcal{N}(0,1)$. In this case, the decode-and-forward transmission rate is

$$R \leq \max_{P_{X_1,X_2}(x_1,x_2)} \min\left\{\mathbb{E}_{H_{12}}\{I(X_1;Y_2|X_2,H_{12})\}, \mathbb{E}_{H_{13},H_{23}}\{I(X_1,X_2;Y_3|H_{13},H_{23})\}\right\} \quad (3.21)$$

where $\mathbb{E}$ is the expectation operator. Therefore, the multilevel decomposition in (3.14) is still valid, given the following averaging over the channel coefficients:

$$I(B_i;Y_2|C^m,B^{i-1}) = \mathbb{E}_{H_{12}}\{I(B_i;Y_2|C^m,B^{i-1},H_{12})\} \quad (3.22)$$

$$I(B_i,C_i;Y_3|B^{i-1},C^{i-1}) = \mathbb{E}_{H_{13},H_{23}}\{I(B_i;Y_2|C^m,B^{i-1},H_{13},H_{23})\} \quad (3.23)$$

The code design criteria described earlier depends on prior knowledge of the point-to-point mutual information curves in Fig. 3.6 and the correlation supported by each level in Table 3.1. These metrics will change in a fading environment however, it can be easily obtained by averaging over the normally distributed fading coefficient. Having reached to these quantities, the design will follow directly the same steps described earlier.

3.4.5 Multi-Antenna Relay

Assume that the relay node has N receive antennas and M transmit antennas. Also, assume that the channel state is known at all nodes. The bold letters in this subsection represent the vector version of the variable. In this subsection, we show that the proposed multilevel transmission and code design follows directly in this case. We start with the source relay

transmission. The only difference in this case is that the relay receives multiple versions of the transmitted symbol and can combine them with any of the existing techniques such as maximum ratio combining. The transmission rate from the source to the relay in this case becomes

$$R_{SR} \leq I(X_1; \boldsymbol{Y}_2 | \boldsymbol{X}_2, H_{12}^{(1)}, \ldots, H_{12}^{(N)}) \tag{3.24}$$

$$= I(B^m; \boldsymbol{Y}_2 | \boldsymbol{X}_2, H_{12}^{(1)}, \ldots, H_{12}^{(N)}) \tag{3.25}$$

$$= \sum_{i=1}^{m} I(B_i; \boldsymbol{Y}_2 | \boldsymbol{X}_2, B^{i-1}, H_{12}^{(1)}, \ldots, H_{12}^{(N)}) \tag{3.26}$$

which implies that the transmission rate of level i at the source is upper bounded by

$$R_i \leq \sum_{i=1}^{m} I(B_i; \boldsymbol{Y}_2 | \boldsymbol{X}_2, B^{i-1}, H_{12}^{(1)}, \ldots, H_{12}^{(N)}) \tag{3.27}$$

Now, we show that the relay-destination transmission can be modeled as a single antenna transmission. Assume that the channel from the ith antenna at the relay node to the destination node is $H_{23}^{(i)}$. To show that the system can be modeled as a single antenna relay, we assume a Gaussian input relay channel. The relay can use the M transmit antennas to provide beamforming gain by sending the same signal X_2 from all the antennas. Assuming that the transmit power of each antenna is $P_2^{(i)}$, we have the following constraint

$$\sum_{i=1}^{M} P_2^{(i)} \leq P_2. \tag{3.28}$$

The received signal at the destination is

$$Y_3 = \sum_{i=1}^{m} H_{23}^{(i)} \sqrt{P_2^{(i)}} X_2 + H_{13} \sqrt{P_1} (X_1 + X_2) + n_3 \tag{3.29}$$

$$= (\sum_{i=1}^{m} H_{23}^{(i)} \sqrt{P_2^{(i)}} + H_{13} \sqrt{P_1}) X_2 + H_{13} \sqrt{P_1} X_1 + n_3 \tag{3.30}$$

which is equivalent to single antenna relay channel where the channel gain from the relay to the destination is

$$\sum_{i=1}^{m} H_{23}^{(i)} \sqrt{P_2^{(i)}} + H_{13} \sqrt{P_1}. \tag{3.31}$$

This requires an optimization over the powers of the transmit antenna at the relay however, once the power allocation is optimized, the problem becomes similar to the single relay antenna transmission.

3.5 Error Exponent Analysis

In this section we analyze the asymptotic error performance of the proposed transmission scheme via error exponents.The error exponents give exponential lower and upper bounds on the probability of error as a function of the blocklength and is used to understand how fast the probability of error will be vanish as the blocklength increases [69, 70]. In a point-to-point channel, the error exponent upper bound takes the form

$$P_e \leq e^{-nE(R)} \tag{3.32}$$

where n is the blocklength and $E(R)$ is the error exponent as a function of the transmission rate. The error exponent depends on the channel-input distribution and a tilting parameter both to be optimized.

In this section we derive an upper bound error exponent for the proposed transmission and compare it with the error exponent of the channel with no restrictions on the input. The error exponent of the full-duplex decode and forward relay channel was studied by Li and Georghiades [71] under backward decoding. Bradford and Laneman studied the error exponent of the full-duplex relay channel under sliding window decoding [72]. Tan [73] produced the full-duplex relay error exponent for partial decode and forward and compress and forward under backward decoding. We study the error exponent of multilevel coding in

full-duplex relay under sliding window decoding; the backward decoding analysis is similar and is omitted for brevity.

The error event in the relay channel has two components, the decoding error at the relay and the decoding error at the destination node. An error at the relay node will lead to an error at the destination with very high probability. For the sake of clarity, we need to define two error probabilities at each node, ϵ_R is the probability of error at the relay given that the previous block was decoded successfully and ϵ_D is the probability of error at the destination given that the current block is decoded successfully at the relay and the previous block is decoded successfully at the destination. These error probabilities are defined conditioned on a previously successful decoding to simplify the analysis. It was shown by Bradford and Laneman [72] that the probability of error in the full-duplex relay communications can be upper bounded by

$$Pe \leq (B-1)(\epsilon_R + \epsilon_D) \tag{3.33}$$

where B is the number of blocks.

Since each probability of error at each node has an associated error exponent that determines an upper bound on the probability of error, each error probability can be upper bounded by it's error exponent. This leads to the random coding error exponent of the entire transmission,

$$E(R) \geq \frac{1}{B}\min\{E_R(R), E_D(R)\} - \frac{\log 2(B-1)}{D} \tag{3.34}$$

where $E_R(R)$ and $E_D(R)$ are the random coding error exponents corresponding to ϵ_R and ϵ_D respectively and D is the total number of transmission symbols in the B blocks ($D = nR$) where n is the blocklength.

In the rest of this section, for the sake of completeness we state the error exponents in (3.34) for the probability of error at each node under no encoding restriction. Consequently, we present the same error exponents under multilevel coding and finally for the multilevel coding with multistage decoding. In the following, for brevity and clarity of

exposition, probability distributions are distinguished by their respective arguments. The reader is reminded that superscripted variables are vectors (e.g., $B^m = [B_1, \ldots, B_m]$ and $b^{i-1} = [b_1, \ldots, b_{i-1}]$). Summations are over the entire defined range of their subscript variable (or vector).

The error exponent for the probability of error at the relay, $E_R(R)$, is given by

$$E_R(R) = \max_{P(x),\rho_e} [E_{01}(\rho_e, P(x)) - \rho_e R] \tag{3.35}$$

where ρ_e is the random coding error exponent tilting parameter and

$$E_{01} = -\log \sum_{x_2} \int P(x_2) \Big[\sum_{x_1} P(x_1|x_2) P(y_2|x_1, x_2)^{\frac{1}{1+\rho_e}} \Big]^{1+\rho_e} dy_2 \tag{3.36}$$

In order to obtain the error exponent of the proposed multilevel encoding, we replace X_1 and X_2 by B^m and C^m and using the independence between the components of B^m and C^m, we find

$$E_{01} = -\log \sum_{c^m} \int \prod_i P(c_i) \Big[\sum_{b^m} \prod_j P(b_j|c_j) P(y_2|b^m, c^m)^{\frac{1}{1+\rho_e}} \Big]^{1+\rho_e} dy_2 \tag{3.37}$$

The error exponent under multistage decoding is more complicated. For this part, we model the multilevel encoding under multistage decoding as multiple channels with side information as follows: Multilevel coding under multistage decoding can be thought of as a decomposition of the channel from $(B_1, \ldots, B_m)$ to Y_2 into a series of channels from B_i to Y_2. Considering a successful decoding at all the decoders preceding decoder i, decoder i will have an access to B^{i-1}. Therefore, the sub-channel between B_i and Y_2 can be thought of as a channel with a "state" known at the receiver where the state of the channel is B^{i-1}. Ingber and Feder [74] derived a random coding error exponent for channels with side information at the receiver

$$E(\rho_e) = -\log \mathbb{E}[2^{-E^s(\rho_e)}], \tag{3.38}$$

where s is the state of the channel. Similarly, Calculating the error exponent under multistage decoding at level i will require averaging over B^{i-1} since at level i, the decoder knows the outputs B^{i-1} of the preceding decoders. Therefore, E_{01} of level i is given by

$$E_{01}^{i} = -\log \sum_{c^m, b^{i-1}} \int P(c^m, b^{i-1}) \Big[\sum_{b_i} P(b_i|c^m, b^{i-1}) P(y_2|b^m, c^m)^{\frac{1}{1+\rho_e}} \Big]^{1+\rho_e} dy_2 \tag{3.39}$$

Now, we are left with combining the error exponents in all the levels to obtain E_{01} under multistage decoding. In a point-to-point channel, Ingber and Feder derived a random coding error exponent of multilevel coding and multistage decoding as a function of the error exponent of the individual sub-channels "with state known at the receiver" as mentioned earlier [68, Theorem 3]. The main idea is that the total error exponent is dominated by the minimum error exponent of all the levels. Inspired by their bound, the error exponent of the decoder at the relay $E_R(R)$ under multistage decoding is

$$E_R^{MSD}(R) = \max_{R_i, P(b_i,c_i) \forall i} \min_{l} \max_{\rho} [E_{01}^{l} - \rho_e R_l] \tag{3.40}$$

The error exponent at the destination, $E_D(R)$, is more complicated as it involves sliding window decoding. Bradford and Laneman [72] decomposed this error exponent to rely on the window size L and two other metrics, namely

$$E_0(\rho_e, P(x_1, x_2)) = -\log \int \Big[\sum_{x_1, x_2} P(x_1, x_2) P(y_3|x_1, x_2)^{\frac{1}{1+\rho_e}} \Big]^{1+\rho_e} dy_3 \tag{3.41}$$

$$E_{02}(\rho_e, P(x_1, x_2)) = -\log \sum_{x_2} \int P(x_2) \Big[\sum_{x_1} P(x_1|x_2) P(y_3|x_1, x_2)^{\frac{1}{1+\rho_e}} \Big]^{1+\rho_e} dy_3 \tag{3.42}$$

Obtaining these two parameters for the proposed multilevel transmission will require replacing X_1 and X_2 with B^m and C^m respectively to give

$$E_0(\rho_e, P(b^m, c^m)) = -\log \Bigg[\int \Big[\sum_{b^m, c^m} P(b^m, c^m) P(y_3|b^m, c^m)^{\frac{1}{1+\rho_e}} \Big]^{1+\rho_e} dy_3 \Bigg] \tag{3.43}$$

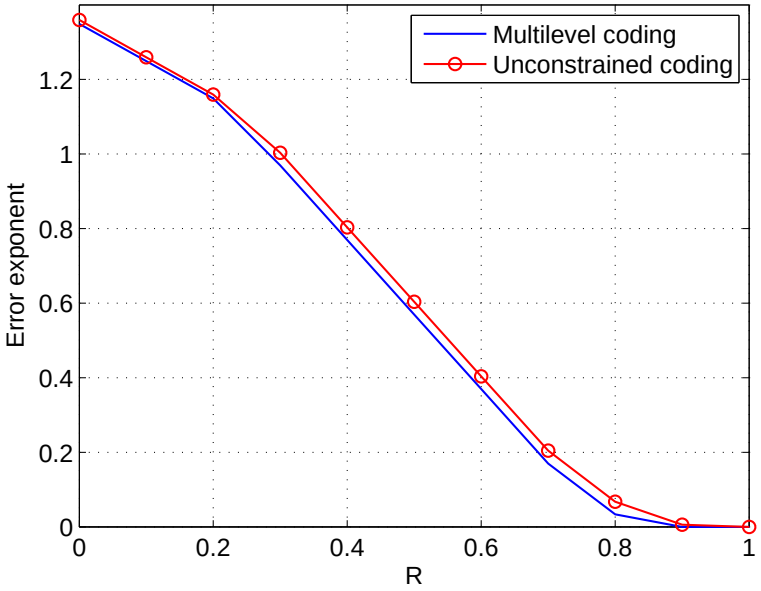

Figure 3.8. Error exponent for bit-additive MLC versus unconstrained coding for 4-PAM transmission and P=20

$$E_{02}(\rho_e,P(b^m,c^m)) = -\log\left[\sum_{c^m}\int\prod_i P(c_i)\Big[\sum_{b^m}\prod_j P(b_j|c_j)P(y_3|b^m,c^m)^{\frac{1}{1+\rho_e}}\Big]^{1+\rho_e}dy_3\right] \tag{3.44}$$

Under multistage decoding, B^{i-1} will be decoded and passed to decoder i before it starts decoding B_i. Therefore, the error exponent should be averaged over B^{i-1} in (3.43) and (3.44) to evaluate the error exponent while decoding level i. This gives

$$E_0^i(\rho_e, P(b^m,c^m)) = -\log\sum_{b^{i-1}}\int\Big[\sum_{b_i,c^m}P(b^i,c^m|b^{i-1})P(y_3|b^i,c^m)^{\frac{1}{1+\rho_e}}\Big]^{1+\rho_e}dy_3$$

$$E_{02}^i(\rho_e, P(b^m,c^m)) = -\log\sum_{c^m,b^{i-1}}\int P(c^m,b^{i-1})\Big[\sum_{b_i}P(b_i|c^m,b^{i-1})P(y_3|b^i,c^m)^{\frac{1}{1+\rho_e}}\Big]^{1+\rho_e}dy_3$$

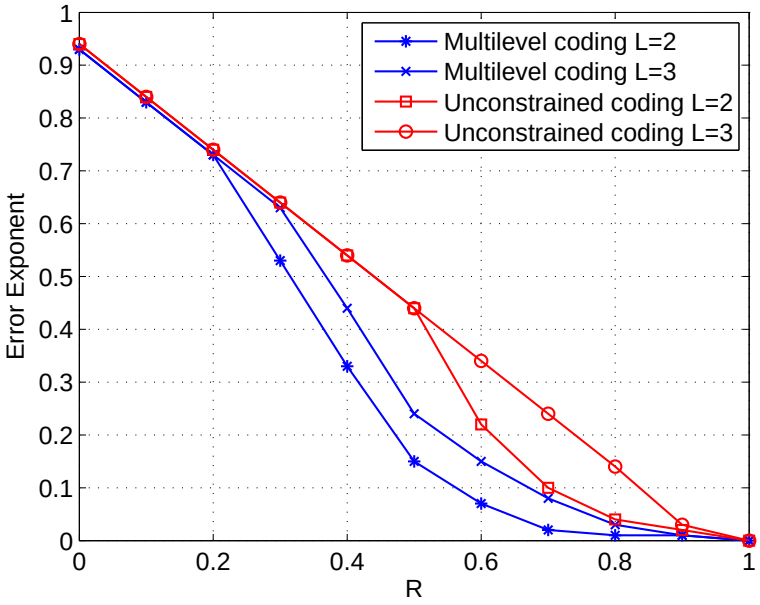

Figure 3.9. Error exponent for bit-additive MLC versus unconstrained coding for 4-PAM transmission and P=10

We now numerically compare the error exponent of the proposed transmission under multistage decoding with the general error exponent of the channel with no restrictions on the encoding or decoding. These results where obtained by exhaustive search over the input distributions $P_{B_i|C_i}(b_i|c_i)$ and $P_{C_i}(c_i)$ and the tilting parameter ρ_e to find the maximum error exponent. Please note that the random variables B_i and C_i are binary random variables, therefore, the exhaustive search includes one number that takes values between 0 and 1 for each variable. For the case with no restriction on the encoding or decoding, the error exponent was found by exhaustive search over the input constellation distribution which requires a large computational power. Fig. 3.8 and Fig. 3.9 show the error exponent of the proposed multilevel transmission under multistage decoding at the relay and destination when the window size is $L = 3$. The input channel constellation was 4-PAM. The figures show two cases, first, when the window size is not effective, Fig. 3.8, the error exponent is very close to the general encoding at the source and the relay nodes. Second, when the

window size is effective, Fig. 3.9, there is an obvious loss in the error exponent. However, as the window size increases, the error exponent of the proposed transmission gets closer to that of the general encoding at the source and relay nodes.

3.6 Simulations

3.6.1 Modulation Constellations and Achievable Rates

We assume equal transmit power and the source and the relay nodes, $P_1 = P_2 = P$, and unit variance noise at the relay and destination. The noise power at the relay node includes the thermal noise and the residual self-interference. To demonstrate the performance of the relay channel in a variety of link SNRs, we assume a path loss model following the setting of the well-known work of Kramer et al [75], with a path loss exponent $\alpha = 4$. The source, relay and destination are aligned on a line, with source-destination distance d_{13}, source-relay distance d, and relay-destination distance $d_{23} = d_{13} - d$. In our simulations we take $d_{13} = 4$. The link gains are therefore $h_{ij} = (\frac{1}{d_{ij}})^{\alpha/2}$.

The figures also include, for comparison purposes, the achievable rates for the unconstrained Gaussian relay channel:

$$R_{DF} = \max_{0 \le \rho \le 1} \min \left\{ \frac{1}{2} \log \left(1 + |H_{12}|^2 P_1 (1 - |\rho|^2) \right), \frac{1}{2} \right.$$
$$\left. \log \left(1 + |H_{13}|^2 P_1 + |H_{23}|^2 P_2 + 2\rho \sqrt{|H_{13}|^2 P_1 |H_{23}|^2 P_2} \right) \right\}$$

The transmission rates of the proposed multilevel coding are shown in Fig. 3.10 for one and two dimensional constellation at different source and relay powers. The transmission rates where obtain by exhaustive search over the input distribution to obtain the maximum achievable rate. The results show that the gap between the transmission rate of the proposed transmission and the Gaussian input transmission rate is very small and gets smaller with

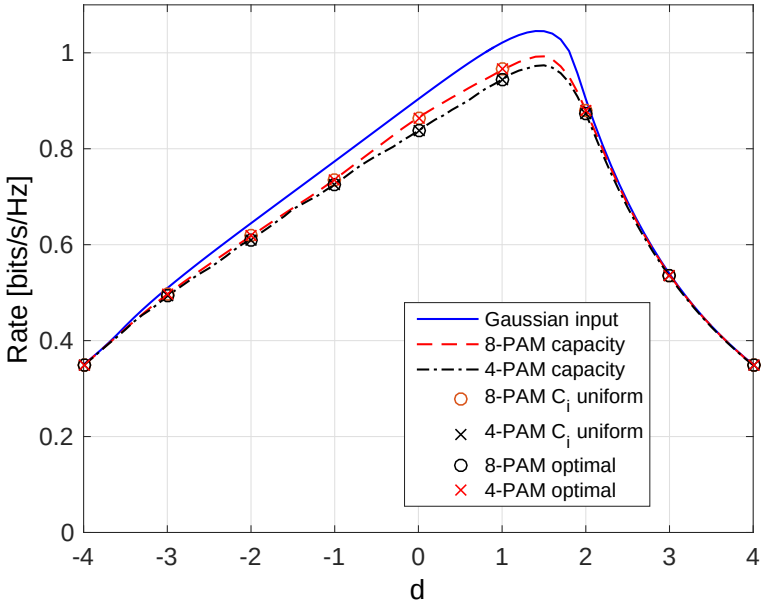

Figure 3.10. Natural labeling, PAM, $P_1 = P_2 = 10$dB

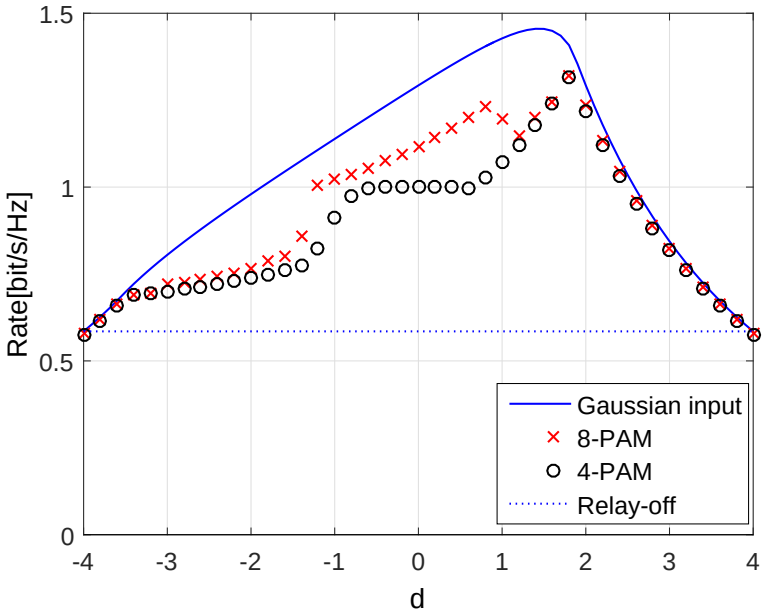

Figure 3.11. Rate of multilevel transmission when using linear codes for 4-PAM and 8-PAM constellations, $P_1 = P_2 = 13$dB

larger constellation size. The gap is smaller when the relay is far from the source and the source-relay link has smaller SNR.

Fig. 3.11 shows the degradation in the achievable rates when the source is enforced to use linear component codes. This implies that in the full-duplex relay, the achievable rates are sensitive to the correlation, which is unlike the half-duplex relay case reported in [76]. The 8-PAM constellation achieves significantly higher rates when the relay is close to the source, where the signaling calls for strong correlation, because the 8-PAM has a bigger set of feasible correlations under the linear coding constraint.

3.6.2 Error Rate Simulations

The DVB-S2 LDPC codes are used as component codes for each of the levels at the source node and the relay node to examine the performance of the proposed multilevel transmission. The rates of the LDPC codes are chosen according to the design criteria in Section 3.4. The blocklength of the component codes is $n = 64$k. Both the relay and destination nodes used belief propagation decoding at each level where the maximum number of iterations is set to 50.

The decoding at the relay node is performed as follows: While decoding level i of the signal X_1 at the relay, the relay knows two parts of X_1 already. The first is the vector U^m which is the cloud center of X_1 and the second is the vector V^{i-1} which is the output of the preceding decoders, assuming correct decoding. Therefore, the LLR of level i at the relay is

$$LLR_r = \log \frac{P(y_2|u^m, v^{i-1}, 0)}{P(y_2|u^m, v^{i-1}, 1)} \tag{3.45}$$

where

$$P(y_2|u^m, v^{i-1}, v_i) = \frac{1}{P(u^m, v^{i-1}, v_i)} \sum_{v_{i+1}^m} P(y_2|u^m, v^m)$$

The decoding at the destination node is performed as follows: Assuming that the destination node will decode the signal from the relay node and then decode the signal from the source node, the LLR of level i of the relay at the destination node is

$$LLR_{RD} = \log \frac{P(y_3|c^{i-1}, 0)}{P(y_3|c^{i-1}, 1)} \tag{3.46}$$

where

$$P(y_3|c^{i-1}, c_i) = \frac{1}{P(c^{i-1}, c_i)} \sum_{b^m, c_{i+1}^m} P(y_3|b^m, c^m)$$

The next step is to decode the signal from the source given the transmitted signal from the relay with

$$LLR_{SD} = \log \frac{P(y_3|c^m, b^{i-1}, 0)}{P(y_3|c^m, b^{i-1}, 1)} \tag{3.47}$$

where

$$P(y_3|c^m, b^{i-1}, b_i) = \frac{1}{P(c^m, b^{i-1}, b_i)} \sum_{b_{i+1}^m} P(y_3|b^m, c^m)$$

and C^m carries all the information about the cloud center of the source signal.

In each of the error plots, a capacity threshold is marked that corresponds to the relay constellation constrained capacity in each case. The source and relay powers are identical throughout all simulations, enabling the use of a single scale for power (dB) in the error curves. In each of the simulations, the rates at each level are found by exhaustive search so that the sum-rate is maximized.

Fig. 3.12 shows the bit error probability and frame error probability for 4-PAM multilevel transmission at $d_{12} = 1$ and $\alpha = 2$. The figure shows the performance of the three labellings shown in Table 3.1. The total transmission rate is $R = 0.8$. In general, for each labeling, the bit-wise correlation is not the same. However, for the current channel parameters, the bit wise correlations used in these simulations were $\rho_1 = 0$ and $\rho_2 = 1$ which means that

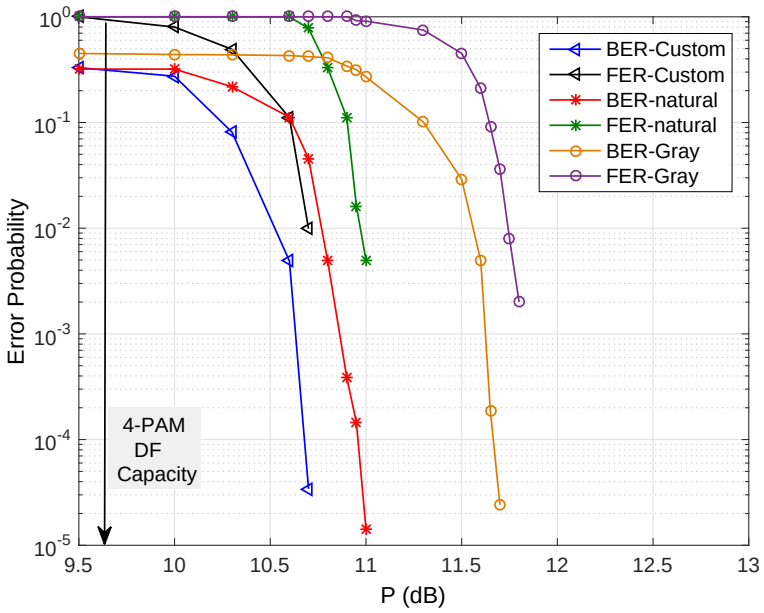

Figure 3.12. Performance of Multilevel superposition for 4-PAM constellation where $d = 1$

the least significant bit provides beamforming gain to the relay transmission and the most significant bit sends new information to the relay.

Fig. 3.13 shows the bit error probability and frame error probability of 8-PAM multilevel transmission at $d_{12} = 2.5$ with $\alpha = 4$. The total rate transmitted from the source node to the destination node is $R = 2.28$. The optimal value of the bit-wise correlations using linear codes in the current channel conditions are $\rho_1 = 0, \rho_2 = 0$ and $\rho_3 = 0$ which is the same as the general encoding case $\rho = 0$. This is because the relay-destination channel is very strong and does not need any beamforming gain from the source.

We show the performance of a 16-QAM constellation in Fig. 3.14 where $d_{12} = 1.5$ and $\alpha = 2$. The total rate transmitted from the source node to the destination node is $R = 3.5$. In this case, we used non-linear codes only at one of the least significant bits to provide the necessary gain and linear codes at the other three levels.

Fig. 3.15 shows the result for 4-PAM transmission under fast fading relay channel where $d = 1$ and the total transmission rate is 1.5 bits/transmission.

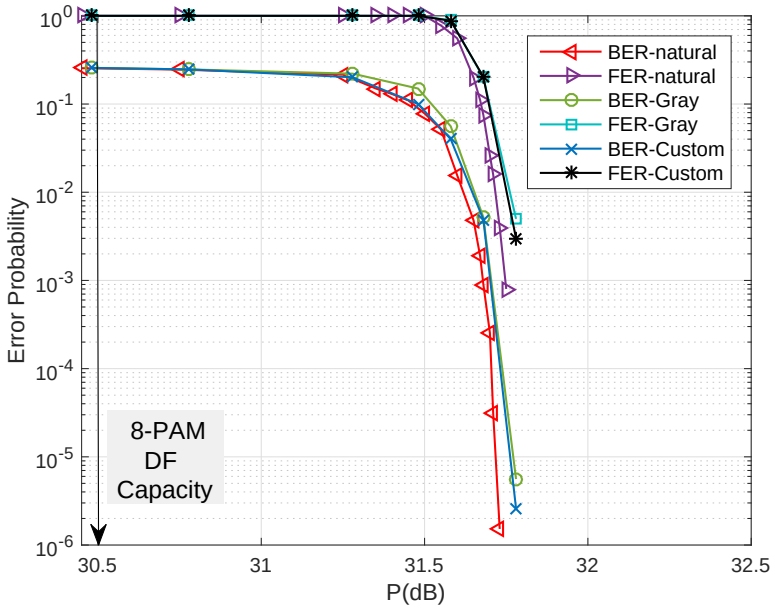

Figure 3.13. Performance of Multilevel superposition for 8-PAM constellation where $d = 2.5$

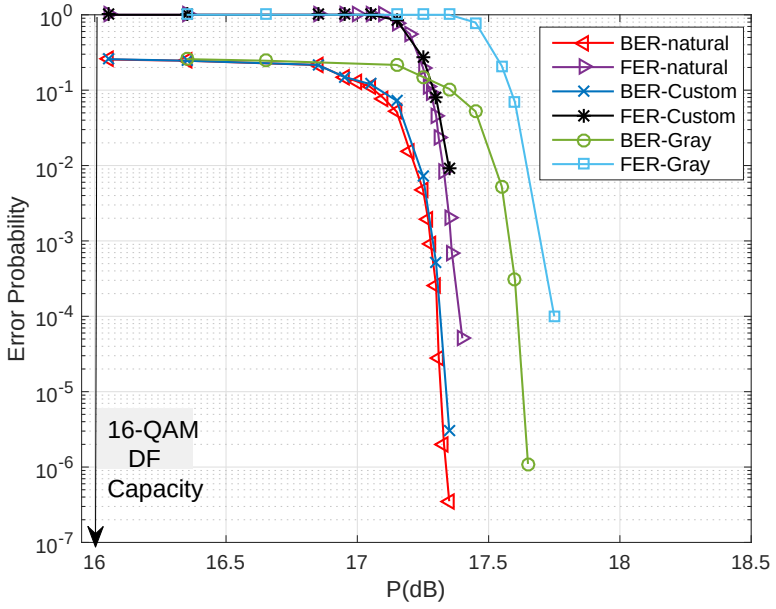

Figure 3.14. Performance of Multilevel superposition for 16-QAM constellation where $d = 1.5$

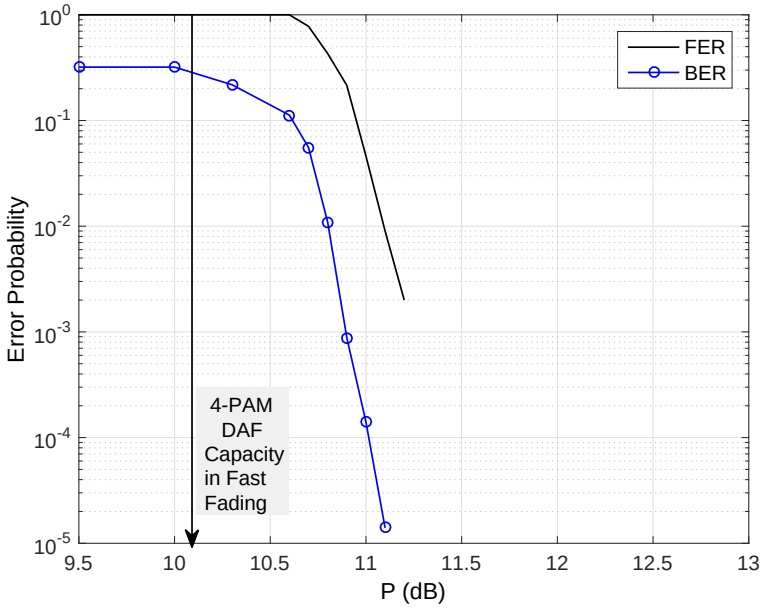

Figure 3.15. Proposed transmission under fast fading channel, 4-PAM constellation.

Remark 11. *As mentioned earlier, to avoid rate loss, the source-relay codes need a non-uniform marginal distribution, which is not available via a (full-rank) linear code. In this section, we used DVB-S2 codewords in which a prescribed number of randomly-located binary symbols were converted to zero. A practical implementation of this scheme requires a pseudo-random number generator at the transmitter and receiver and the maintenance of synchrony between them.*[4] *An alternative approach is non-random assignment of zeros using a puncture design method [63]. Fig. 3.12, Fig. 3.13, and Fig. 3.14 present simulations where superposition codes were constructed with DVB-S2 codes together with random zero assignment; parallel experiments with puncturing design resulted in roughly similar performance, i.e., within* 0.2 *to* 0.3*dB of the experiments with random zero assignment.*

[4]Decoder knowledge of location of these zeros is worth 1 to 1.5dB in performance.

3.7 Discussion and Conclusion

Multilevel coding in the decode and forward relay channel is studied. A coded modulation technique is proposed where the correlation between the source signal and the relay signal is controlled by the pairwise correlation between each level in the source and the corresponding level at the relay. Multistage decoding is studied and the necessary rates of each level for two different ways of multistage decoding are derived. A simple implementation of the proposed transmission using binary addition is presented. The labeling design is addressed and guidelines for it are presented. The error exponent of the proposed transmission is also studied, showing the loss in error exponent due to the proposed transmission is small. Numerical results show that the proposed multilevel coded modulation enjoys capacity approaching performance. From the implementation viewpoint, it is shown that a performance that is very close to the constellation constrained capacity is obtained by using standard point-to-point LDPC codes as component multi-level codes for the relay channel.

One of the main features of the present work is that it provides a systematic design process that is easily adapted to a variety of channel conditions (SNRs and rates). Furthermore, since the design of the coded modulation is reduced to the design of point-to-point binary codes, it enjoys a number of advantages including availability at a wide range of block lengths.

CHAPTER 4

CODED MODULATION FOR THE DECODE-COMPRESS-FORWARD RELAY CHANNEL

4.1 Introduction

Shortly after the introduction of the three-node relay channel by Van Der Meulen [77], Cover and El-Gamal [55] proposed and analyzed block-Markov encoded decode-forward (DF) and compress-forward (CF) for the relay channel [78, 75, 79].

Decode-forward is capacity achieving for the degraded relay channel, but due to the relay decoding constraint, it does not perform well when the source-relay link is weak. Compress-forward can also be optimal under certain conditions [80], but it also falls short under certain other conditions [78, 75]. This motivated a generalization of DF and CF into a hybrid technique by Cover and El-Gamal [55, Theorem 7], which is denoted decode-compress-forward. This technique, its performance and implementation via coded modulation, are the subjects of this chapter. Other hybrid relaying protocols include hybrid Decode-forward and Amplify-forward [81], and also a variation of DCF has appeared in the context of selective cooperation [82].

The known achievable rate of DCF in the AWGN full-duplex relay channel reduces to either DF or CF achievable rates. This is a result due to [83] that we re-derive in the following under backward decoding. In the discrete input full-duplex relay channel, DCF performance can exceed both DF and CF.

A coding implementation for the DCF in the AWGN channel is then proposed based on multilevel coding (MLC). At the source and the relay, the proposed DCF multilevel coding decomposes the overall coded modulation into a group of binary codes, each either operating via a DF or a CF protocol. The mapper combines these constituent level-wise codes into a hybrid DF-CF coded modulation. The assignment of each level to either DF or CF and the

rate allocation to each level is an optimization problem. We demonstrate the operation of this system by an implementation that employs for the DF components the DVB-S2 LDPC codes [84], and for the CF components a group of polar codes that are designed according to Blasco-Serrano [85, 86].

4.2 Decode-Compress-Forward

In this section, we re-derive Theorem 7 [55] for the discrete memoryless relay channel under backward decoding and obtain the DCF achievable rate for the AWGN relay channel as well as the constellation constrained AWGN relay channel.

4.2.1 Discrete Memoryless Full-Duplex Relay

Block Markov-encoding for the DCF is shown in Fig. 4.1 over four transmission blocks. In each transmission block, the source and the relay send a compress-forward component that is superimposed on a decode-forward component. A detailed system design and analysis is explained as follows:

In g transmission blocks or ng transmissions, a sequence of $(g-1)$ i.i.d. messages $W_j \in [1, 2^{nR}], i \in [1 : g-1]$. Each message W_j is split into two messages $W_{dj} \in [1, 2^{nR_d}]$ and $W_{cj} \in [1, 2^{nR_c}]$ for $j \in [1, g-1]$. This implies that $R = R_d + R_c$.

Codebook generation:

For each block $j \in [1 : g]$, randomly and independently generate 2^{nR_d} sequences $U_{2d}^n(w_{d(j-1)})$ according to

$$\prod_{i=1}^{n} P_{U_{2d}}(u_{2di})$$

For each $w_{d(j-1)} \in [1 : 2^{nR_d}]$, randomly and conditionally independently generate $2^{nR'}$ sequences $X_2^n(l_{j-1}|w_{d(j-1)}), l_{j-1} \in [1 : 2^{nR'}]$, each according to

$$\prod_{i=1}^{n} P_{X_2|U_{2d}}(x_{2i}|u_{2di}(w_{d(j-1)}))$$

For each $l_{j-1} \in [1 : 2^{nR'}]$, randomly and conditionally independently, generate $2^{nR''}$ sequences $\hat{Y}_2^n(k_j|l_{j-1}), k_j \in [1 : 2^{nR''}]$ each according to

$$\prod_{i=1}^{n} P_{\hat{Y}_2|X_{2i}}(\hat{y}_{2i}|x_{2i}(l_{j-1}|w_{d(j-1)})$$

Also, for each $w_{d(j-1)}$, randomly and conditionally independently generate 2^{nR_d} sequences $U_{1d}^n(w_{dj}|w_{d(j-1)}), w_{dj} \in [1 : 2^{nR_d}]$, each according to

$$\prod_{i=1}^{n} P_{U_{1d}|U_{2d}}(u_{1di}|u_{2di}(w_{d(j-1)}))$$

Finally, for each pair of messages $w_{d(j-1)}$ and $w_{d(j)}$, randomly and conditionally independent generate 2^{nR_c} sequences $X_1^n(w_{cj}|w_{d(j)}, w_{d(j-1)})$, each according to

$$\prod_{i=1}^{n} P_{X_1|U_{2d},U_{1d}}(x_{1i}|u_{2di}((w_{d(j-1)})), u_{2di}((w_{d(j-1)})))$$

This defines the codebooks

$$\mathcal{C}_j = \{x_1^n(w_{cj}|w_{dj}, w_{d(j-1)}), x_2^n(l_j|w_{d(j-1)})\}, \quad j \in [1 : g]$$

The source node:

In block j, the pair of messages w_{dj} and w_{cj} are to be transmitted. The encoder at the source node chooses $x_1(w_{cj}|w_{dj}, w_{d(j-1)})$ from codebook $\mathcal{C}_j$. The messages of the last block are considered to be $w_{cg} = w_{dg} = 1$.

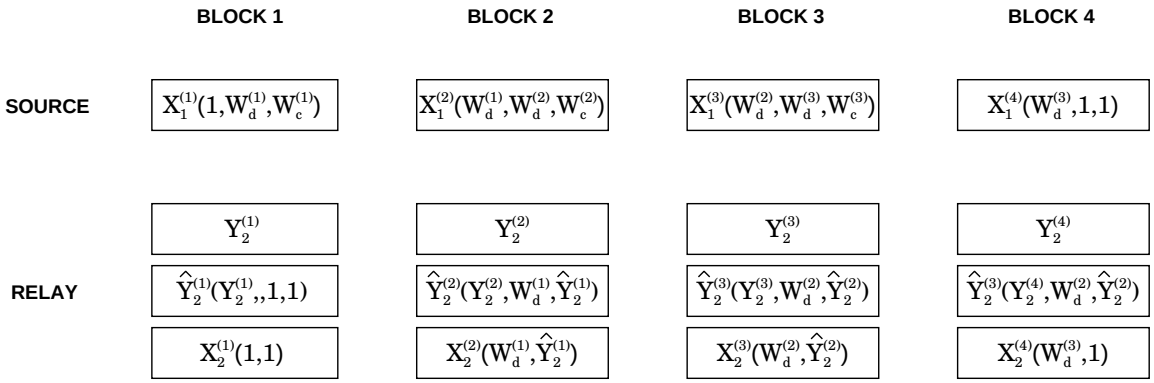

Figure 4.1. Decode-Compress-Forward transmission over four transmission blocks.

The relay node:

The decoding phase of the relay uses typicality decoding as follows: First, assume that $\tilde{w}_{d0} = 1$. Second, at the end of block j, the relay finds a unique $\tilde{w}_{dj}$ such that

$$(x_1^n(w_{cj}|\tilde{w}_{dj}, \tilde{w}_{d(j-1)}), x_2^n(l_{j-1}|\tilde{w}_{d(j-1)}), y_2^n(j)) \in \tau_\epsilon^{(n)}$$

for any w_{cj} where $\tau_\epsilon^{(n)}$ denotes the jointly typical sets of the corresponding random variables.

The relay then finds k_j such that

$$(y_2^n(j), \hat{y}_2^n(k_j|l_{j-1}), x_2^n(l_{j-1})) \in \tau_\epsilon^{(n)}$$

and if there is more than one k_j, the relay selects one at random and if the relay does not find any k_j then, it selects one uniformly at random from $[1 : 2^{nR''}]$. Based on k_j, the relay determines l_j as it is the bin index of k_j.

In the transmission phase, in block j, the relay chooses $x_2(l_{j-1}|w_{d(j-1)})$ from codebook $\mathcal{C}_j$.

The destination node:

The destination uses backward decoding so, it waits until the reception of the g blocks and then starts decoding from the last block and successively towards the first block. For $j = g-1, g-2, \ldots, 1$, The destination finds estimates $\hat{w}_{d(j)}$ and $\hat{l}_j$ such that

$$(x_1^n(\hat{w}_{c(j+1)}|\hat{w}_{d(j+1)}, \hat{w}_{d(j)}), x_2^n(\hat{l}_j|\hat{w}_{d(j)}), y_3^n(j+1)) \in \tau_\epsilon^n$$

where $w_{cg} = w_{dg} = 1$. The destination then finds an estimate $\hat{w}_{cj}$ such that

$$(x_1(\hat{w}_{c(j)}|\hat{w}_{d(j)}, \hat{w}_{d(j-1)}), x_2(\hat{l}_{j-1}|\hat{w}_{d(j-1)}), \hat{y}_2^n(\hat{k}_j|\hat{l}_{j-1})) \in \tau_\epsilon^n$$

for some $\hat{k}_j$ that belongs to the bin $\hat{l}_j$.

Probability of error analysis

Without loss of generality, we always assume that the source messages are $w_{dj} = w_{cj} = 1$ for $j \in [1, g]$. In block j, there are two error events at the relay, an error when the relay does not decode w_{dj} correctly and another when the relay makes an error in the compress-forward part. The two errors at the relay in block j are defined as follows:

$$\tilde{\mathcal{E}}_1(j) = \{\hat{W}_{dj} \neq 1\} \tag{4.1}$$

$$\tilde{\mathcal{E}}_2(j) = \{(X_2^n(L_{j-1}|W_{d(j-1)}), \hat{Y}_2^n(k_j|L_{j-1}), Y_2^n(j)) \notin \tau_\epsilon^n \text{for all } k_j \in [1 : 2^{nR''}]\} \tag{4.2}$$

while the error events at the destination are defined as follows:

$$\mathcal{E}(j+1) = \{(W_{d(j+1)} \neq 1) \cup (W_{c(j+1)} \neq 1)\} \tag{4.3}$$

$$\mathcal{E}'(j+1) = \{L_{c(j+1)} \neq 1\} \tag{4.4}$$

$$\mathcal{E}_1(j) = \{(X_1^n(\hat{W}_{c(j+1)}|\hat{W}_{d(j+1)}, \hat{W}_{dj}), X_2^n(\hat{L}_j|\hat{W}_{dj}), Y_3^n(j+1)) \notin \tau_\epsilon^n\} \tag{4.5}$$

$$\mathcal{E}_2(j) = \{(X_1^n(\hat{W}_{c(j+1)}|\hat{W}_{d(j+1)}, \hat{w}_{dj}), X_2^n(\hat{l}_j|\hat{w}_{dj}), Y_3^n(j+1)) \in \tau_\epsilon^n \text{ for some } \hat{w}_{dj} \neq 1, \hat{l}_j \neq 1\} \tag{4.6}$$

$$\mathcal{E}_3(j) = \{(X_1^n(\hat{W}_{c(j+1)}|\hat{W}_{d(j+1)}, \hat{W}_{dj}), X_2^n(\hat{l}_j|\hat{W}_{dj}), \hat{Y}_2^n(\hat{K}_j|\hat{L}_{j-1})) \notin \tau_\epsilon^n\} \tag{4.7}$$

$$\mathcal{E}_4(j) = \{(X_1^n(w_{c(j+1)}|\hat{W}_{d(j+1)}, \hat{W}_{dj}), X_2^n(\hat{l}_j|\hat{W}_{dj}), \hat{Y}_2^n(\hat{K}_j|\hat{L}_{j-1})) \in \tau_\epsilon^n \text{ for some } w_{c(j+1)} \neq 1\} \tag{4.8}$$

The probability of error is then

$$\mathcal{E}(j) = P(\tilde{\mathcal{E}}_1(j) \cup \tilde{\mathcal{E}}_2(j) \cup \mathcal{E}(j+1) \cup \mathcal{E}'(j+1) \cup \mathcal{E}_1(j) \cup \mathcal{E}_2(j) \cup \mathcal{E}_3(j) \cup \mathcal{E}_4(j)) \tag{4.9}$$

$$\leq P(\tilde{\mathcal{E}}_1(j)) + P(\tilde{\mathcal{E}}_2(j)) + P(\mathcal{E}(j+1)) \tag{4.10}$$

$$+ P((\mathcal{E}_1(j) \cup \mathcal{E}_3(j)) \cap \tilde{\mathcal{E}}_1^c(j) \cap \tilde{\mathcal{E}}_2^c(j) \cap \mathcal{E}^c(j+1) \cap \mathcal{E}'^c(j+1))$$

$$+ P(\mathcal{E}_2(j)) + P(\mathcal{E}_4(j)) \tag{4.11}$$

By the LLN and the packing lemma, $P(\tilde{\mathcal{E}}_1(j)) \to 0$ as $n \to \infty$ if

$$R_d \leq I(U_{1d}; Y_2|U_{2d}) \tag{4.12}$$

By independence of the codebooks and the covering lemma, the term $P(\tilde{\mathcal{E}}_2(j)) \to 0$ as $n \to \infty$ if

$$R'' \geq I(\hat{Y}_2; Y_2|X_2) \tag{4.13}$$

For $P(\mathcal{E}(j+1))$, since the messages of the last block is known exactly to be 1, by induction and satisfying the other constraints, $P(\mathcal{E}(j+1)) \to 0$ as $n \to \infty$.

By the independence of the codebooks and the LLN, the term $P((\mathcal{E}_1(j) \cup \mathcal{E}_3(j)) \cap \tilde{\mathcal{E}}_1^c(j) \cap \tilde{\mathcal{E}}_2^c(j) \cap \mathcal{E}^c(j+1) \cap \mathcal{E}'^c(j+1)) \to 0$ as $n \to \infty$. The term $P(\mathcal{E}_2(j)) \to 0$ as $n \to \infty$ if

$$R_d \leq I(U_{1d}, U_{2d}; Y_3) \tag{4.14}$$

$$R' \leq I(X_2; Y_3|U_{2d}) \tag{4.15}$$

$$R_c \leq I(X_1; \hat{Y}_2, Y_3|X_2, U_{1d}) \tag{4.16}$$

Eventually, the total transmission rate is $R_d + R_c$, by combining the previous rate constraints, we obtain the following theorem

Theorem 2. *The achievable rate of decode-compress-forward is given by*

$$R \leq \min\Big\{I(U_{1d}; Y_2|U_{2d}), I(U_{1d}, U_{2d}; Y_3)\Big\} + I(X_1; \hat{Y}_2, Y_3|X_2, U_{1d}) \tag{4.17}$$

subject to

$$I(Y_2; \hat{Y}_2|X_2, U_{1d}) \leq I(X_2; Y_3) - I(U_{2d}; Y_3) \tag{4.18}$$

where

$$\begin{aligned} &P_{Y_3,Y_2,\hat{Y}_2,U_{1d},U_{2d},X_1,X_2}(y_3, y_2, \hat{y}_2, u_1, u_2, x_1, x_2) \\ &\quad = P_{U_2}(u_2)P_{U_{1d}|U_{2d}}(u_{1d}|u_{2d})P_{X_1|U_{1d}}(x_1|u_1)P_{X_2|U_{2d}}(x_2|u_2)P_{Y_2|X_1}(y_2|x_1). \\ &\quad\quad P_{\hat{Y}_2|U_{1d},X_2,Y_2}(\hat{y}_2|u_1, x_2, y_2)p_{Y_3|X_1,X_2}(y_3|x_1, x_2) \end{aligned} \tag{4.19}$$

4.2.2 AWGN Full-Duplex Relay

Assume that all the variables in the Section 4.2.1 are Gaussian, [1] and the source and relay have an average power constrained by P_1 and P_2 respectively. A block-Markov encoding of

[1] Gaussian random variables are not necessarily optimal, therefore, the achievable rate is only a lower bound

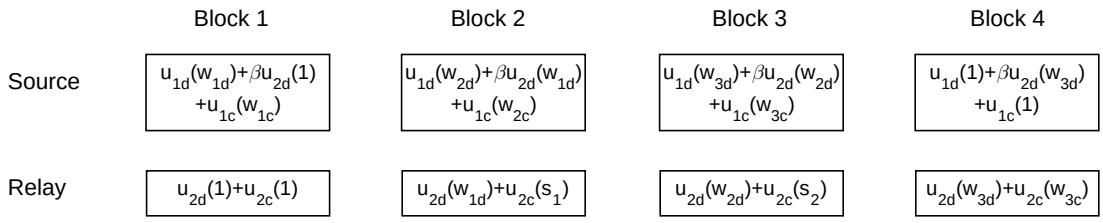

Figure 4.2. Decode-compress-forward transmission for the AWGN full-duplex relay channel over four blocks.

DCF in the AWGN channel is shown in Fig. 4.2 under four transmission blocks. The source and relay signals are given by

$$X_1 = U_{1d} + \beta U_{2d} + U_{1c} \tag{4.20}$$

$$X_2 = U_{2d} + U_{2c} \tag{4.21}$$

respectively.

Each of the codewords U_{1d}, U_{2d}, U_{1c} and U_{2c} are normally distributed. The term βU_{2d} represents the assistance that the source provides for the relay destination transmission. This assistance depends on the correlation between $U_{1d} + \beta U_{2d}$ and U_{2d} which is denoted by ρ.

Remark 12. *In DCF, normally one can optimize the power allocation of each code at the source and the relay. We fix the power of one of the DF signals U_{1d} and the CF signal U_{1c}. The power of the remaining signals can be obtained as a function of the power of U_{1d} and U_{1c} and the power constraint at the source and relay nodes. Consequently, the design variables of the rate maximization problem become the power of U_{1d}, U_{1c} and the correlation ρ.*

Assuming that the power of U_{1d} is P_{1d} and the power of U_{1c} is P_{1c}, the power of U_{2d} is then given by

$$P_{2d} = \frac{P_1 - P_{1d} - P_{1c}}{\beta^2} \tag{4.22}$$

and consequently, the power of U_{2c} is

$$\begin{aligned} P_{2c} &= P_2 - P_{2d} \\ &= P_2 - \frac{P_1 - P_{1d} - P_{1c}}{\beta^2} \end{aligned} \tag{4.23}$$

where

$$\rho = \frac{E[(U_{1d} + \beta U_{2d})U_{2d}]}{\sqrt{(P_{1d} + \beta^2 P_{2d})P_{2d}}} \tag{4.24}$$

$$\beta = \sqrt{\frac{\rho^2 P_{1d} P_{2d}}{P_{2d}^2(1-\rho^2)}} \tag{4.25}$$

The signals $Y_2, \hat{Y}_2$ and Y_3 are given by

$$Y_2 = H_{12}X_1 + n_2 \tag{4.26}$$

$$\hat{Y}_2 = Y_2 + \hat{n} \tag{4.27}$$

$$Y_3 = H_{13}X_1 + H_{23}X_2 + n_3 \tag{4.28}$$

where n_2, n_3 and $\hat{n}$ are zero mean Gaussian noise with variance σ_2^2, σ_3^2 and $\hat{N}$ respectively.

Based on the previous characterization for each of the distributions of the variables involved in calculating the transmission rate, the achievable rate for the AWGN relay channel is given by the following theorem.

Theorem 3. *The DCF achievable rate in the AWGN relay channel with all codewords normally distributed is given by*

$$R \leq \min \left\{ \frac{1}{2} \log \left(1 + \frac{|H_{12}|^2 P_{1d}}{|H_{12}|^2 P_{1c} + \sigma_2^2}\right),\right.$$

$$\frac{1}{2}\log\Big(1+\frac{(P_{1d}+\beta^2P_{2c})|H_{13}|^2}{\sigma_3^2}+\frac{P_{2d}|H_{23}|^2}{\sigma_3^2} + 2\rho\sqrt{\frac{(P_{1d}+\beta^2P_2)P_{2d}|H_{13}|^2|H_{23}|^2}{\sigma_3^4}}\Big)\Big\} + \frac{1}{2}\log\left(\frac{(|H_{12}|^2P_{1c}+\sigma_3^2+\hat{N})(|H_{13}|^2P_{1c}+\sigma_3^2)-(|H_{12}|^2|H_{13}|^2)P_{1c}^2}{(\sigma_3^2+\hat{N})\sigma_3^2}\right) \tag{4.29}$$

where

$$\hat{N}=\frac{(|H_{12}|^2P_{1c}+\sigma_2^2)(|H_{13}|^2(P_{1d}+P_{1c})+\sigma_2^2)}{|H_{23}|^2P_{2c}} \tag{4.30}$$

Proof. See Appendix 4.6.1. □

To illustrate the performance of DCF, we adopt the approach taken in [78] and to show the dependence of performance on the location of the relay, we calculate the achievable rate as a function of relay position on a line extending from the source to destination. For simplicity, a path-loss model is considered with the channel coefficient $H_{ij}=1/d_{ij}^{\alpha}$ where d_{ij} is the distance between node i and node j and α is the path-loss coefficient which is usually between 2 and 4. The distance between the source and the destination is fixed to $d_{13}=1$ while the distances d_{12} and d_{23} depend on the relay location where $d_{23}=1-d_{12}$. In Fig. 4.3, we draw the achievable rate of different transmission technique as a function of the distance between the source and the relay d_{12}. Negative values of d_{12} mean that the relay is on the side of the source that is far from the destination and positive values mean that the relay is between the source and destination.

Fig. 4.3 shows that the achievable rate of DCF in the AWGN relay channel reduces to either DF rate or CF rate. In other words, optimizing DCF results in either DF or CF.

4.2.3 Constellation-Constrained Full-Duplex relay

For the discrete input AWGN relay channel, the rate expressions in Theorem 2 can be calculated via numerical integrations. The optimizing distribution may require an exhaustive

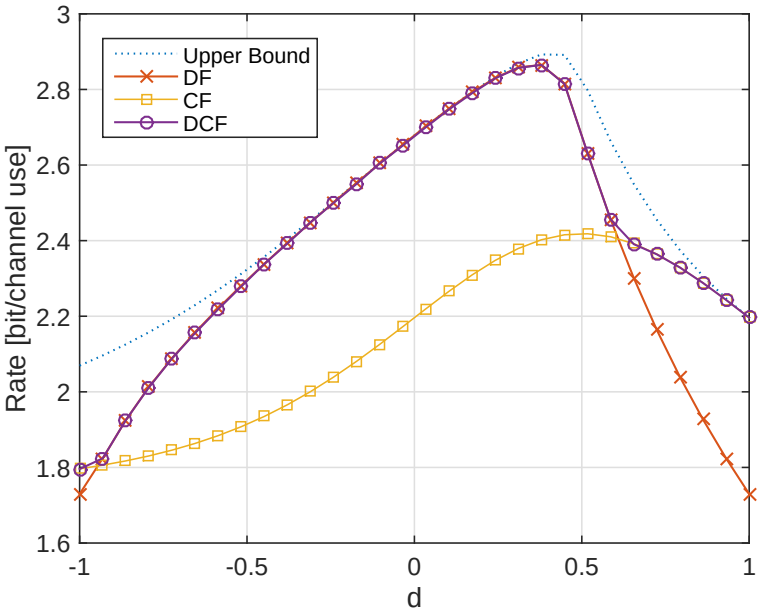

Figure 4.3. The achievable rate of different transmission techniques in the AWGN full-duplex relay channel.

search. As shown in many point-to-point and multi-user scenarios [39, 60], when the constellation becomes large enough, the achievable rate under a constrained constellation becomes very close to the Gaussian input rate.

However, the main difficulty comes from the restriction in (4.18) which is now even harder to satisfy since the mutual information $I(X_2; Y_3)$ is no longer equal to

$$\frac{1}{2}\log\left(1+\frac{|H_{23}|^2 P_2}{N_3}\right)$$

and is limited by the cardinality of the input size $|X_2|$. The exact value of the constellation constrained capacity [11], $I(X_2; Y_3)$, can be obtained from

$$I(X_2;Y_3) = \max_{P_{X_2}(x_2)} \sum_{X_2} P_{X_2}(x_2) \int_{y_3} P_{Y_3|X_2}(y_3|x_2) \log\left(\frac{P_{Y_3|X_2}(y_3|x_2)}{P_{Y_3}(y_3)}\right) dy_3 \tag{4.31}$$

An accurate approximation for $I(X_2;Y_3)$ under constrained constellation can be obtained using the Blahut-Arimoto algorithm [19, 20]. Constellation constrained point-to-point channel capacity for various constellations is shown in Fig. 1.1.

By calculating the value of $I(X_2;Y_3)$, one can find the achievable rate of compress-forward and DCF. In the following, we give an upper bound on the achievable rate of DCF under discrete relay-destination input $\mathcal{X}_2$. The upper bound is based on bounding $I(X_2;Y_3)$ by either the cardinality of X_2 or the Gaussian capacity

$$I(X_2;Y_3) \leq \min\left\{|X_2|, \log\left(1+\frac{|H_{23}|^2P_2}{\sigma_3^2}\right)\right\} \tag{4.32}$$

Therefore, the constraint in (4.18) becomes

$$I(Y_2;\hat{Y}_2|X_2,U_{1d}) \leq \min\left\{|X_2|, \log_2\left(1+\frac{|H_{23}|^2P_2}{\sigma_3^2}\right)\right\} - I(U_{2d};Y_3) \tag{4.33}$$

By using the bound in (4.33), one can find an upper bound on the DCF as follows:

$$R \leq \min\Big\{I(U_{1d};Y_2|U_{2d}), I(U_{1d},U_{2d};Y_3)\Big\} + I(X_1;\hat{Y}_2,Y_3|X_2,U_{1d}) \tag{4.34}$$

subject to

$$I(Y_2;\hat{Y}_2|X_2,U_{1d}) \leq \min\left\{\frac{1}{2}\log_2\left(1+\frac{|H_{23}|^2P_2}{\sigma_3^2}\right), |X_2|\right\} - I(U_{2d};Y_3) \tag{4.35}$$

The capacity of the decode-compress-forward technique can be obtained by evaluating these expressions using numerical integrations and exhaustive search for the optimal distribution.

In a similar manner to the Section 4.2.2, we show the achievable rate of different strategies under a constrained constellation in Fig. 4.4. Fig. 4.4 shows that when the relay is close to the

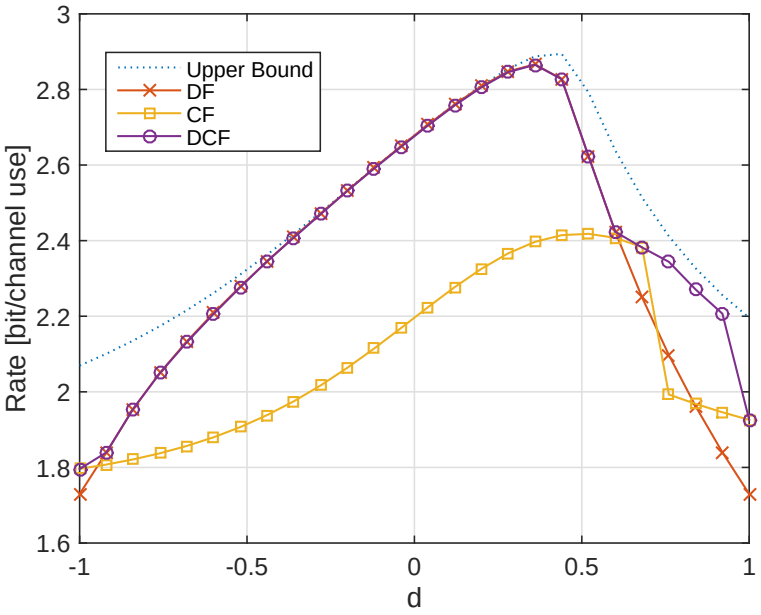

Figure 4.4. The achievable rate of different transmission techniques in the AWGN relay channel under a constrained constellation of 16-PAM at the source and 4-PAM at the relay.

destination, CF suffers a rate penalty while DCF does not suffer from this penalty. Therefore, DCF has a higher rate than DF and CF rates in this region. An intuitive explanation for this observation is as follows: Consider a 4-PAM constellation that consists of two bits with different reliability. Under certain conditions, the relay can decode the most reliable bit but not the least reliable bit. DF enforces the source to send with low rate such that the relay can decode the two bits while CF does not let the relay to decode any bit. DCF allows the relay to decode one of the bits, leading to more flexibility and higher rate. This is different from the continuous case in that the different levels have different reliabilities that are not very close to each other.

Remark 13. *compress-forward works well when the source-relay channel is weak but the relay destination channel is very strong so that it can send a precise enough estimate of* Y_2*. However, when the relay is constellation constrained, even if the quality of the relay-destination link is good, the relay cannot send a precise estimate of* Y_2*.*

Table 4.1. The achievable rate of different strategies in the relay channel under different values of channel coefficients

H_{12}	H_{23}	H_{13}	R_{No_Relay}	R_{DF}	R_{CF}	R_{DCF}	Upper bound
1	100	1	1.7	1.7	1.9	2.1	2.19
1	100	2	2.19	1.7	2.3	2.45	2.47

Clearly, the channel model of the relay channel that is considered in Fig. 4.4, does not represent every possible scenario for the relay channel. Therefore, we show two more results in Table 3.1 that cannot be observed in Fig. 4.4. The table shows different values of the channel coefficients and the achievable rates using DF, CF and DCF.

4.3 Multilevel Decode, Compress and Forward

The proposed multilevel coding is shown in Fig. 4.5, over one transmission block. The message at the source, $W^{(t)}$, is divided into two parts, $W_d^{(t)}$ which is to be transmitted using decode-forward and $W_c^{(t)}$ which is to be transmitted using compress-forward. As shown in the figures, each level at the source and the relay is responsible for either DF or CF.

Relay transmission

The levels at the relay are divided into two sets, first set sends the message $W_d^{(t-1)}$. The second set sends the quantized version of Y_2 after removing the effect of W_d, namely, $\hat{Y}_2$.

Source transmission

The levels at the source are divided into three sets. The first set sends the message $W_d^{(t-1)}$ cooperatively with the relay to the destination node. The second set sends the message $W_d^{(t)}$ to be decoded at the relay. The third set of levels sends the message $W_c^{(t)}$.

The achievable rate of the multilevel coding DCF is given by

$$R \leq \max_{\hat{B},\bar{B},\hat{C},\bar{C}} \min\{I(\hat{B};Y_2|\hat{C})\}, I(\hat{B},\hat{C};Y_3)\}$$

$$+ I(\hat{B}, \bar{B}; \hat{Y}_2, Y_3 | \hat{C}, \bar{C}, \hat{B}) \tag{4.36}$$

subject to

$$I(Y_2; \hat{Y}_2 | \hat{C}, \bar{C}) \leq I(\hat{C}, \bar{C}; Y_3) - I(\hat{C}; Y_3) \tag{4.37}$$

where $\hat{B} = [B_1, \ldots, B_j]$ are the source levels that are responsible for the decode-forward part. $\bar{B} = [B_{j+1} \ldots B_m]$ are the source levels responsible for the compress-forward part. Similarly, $\hat{C} = [C_1, \ldots C_l]$ and $\bar{C} = [C_{l+1}, \ldots C_m]$ are the levels responsible for the DF part and CF part respectively.

The question now is how to allocate the rate in each level at the source and the relay? This is determined through the following steps.

1. The optimal function of each level (being a DF or a CF level) at the source and the relay depends on the constellation and the channel conditions. Using an exhaustive search to optimize (4.36), one can find the best set of levels for decode-forward and for compress-forward at the source and the relay.

2. The number of bits of $W_d^{(t)}$ and $W_c^{(t)}$ depends on the rate of the decode-forward and compress-forward components given by

$$R_d = \min\{I(\hat{B}; Y_2 | \hat{C})\} + I(\hat{B}, \hat{C}; Y_3)\} \tag{4.38}$$

and

$$R_c = I(\hat{B}, \bar{B}; \hat{Y}_2, Y_3 | \hat{C}, \bar{C}, \hat{B}) \tag{4.39}$$

respectively.

So W_d has rate nR_d and W_c has rate nR_c.

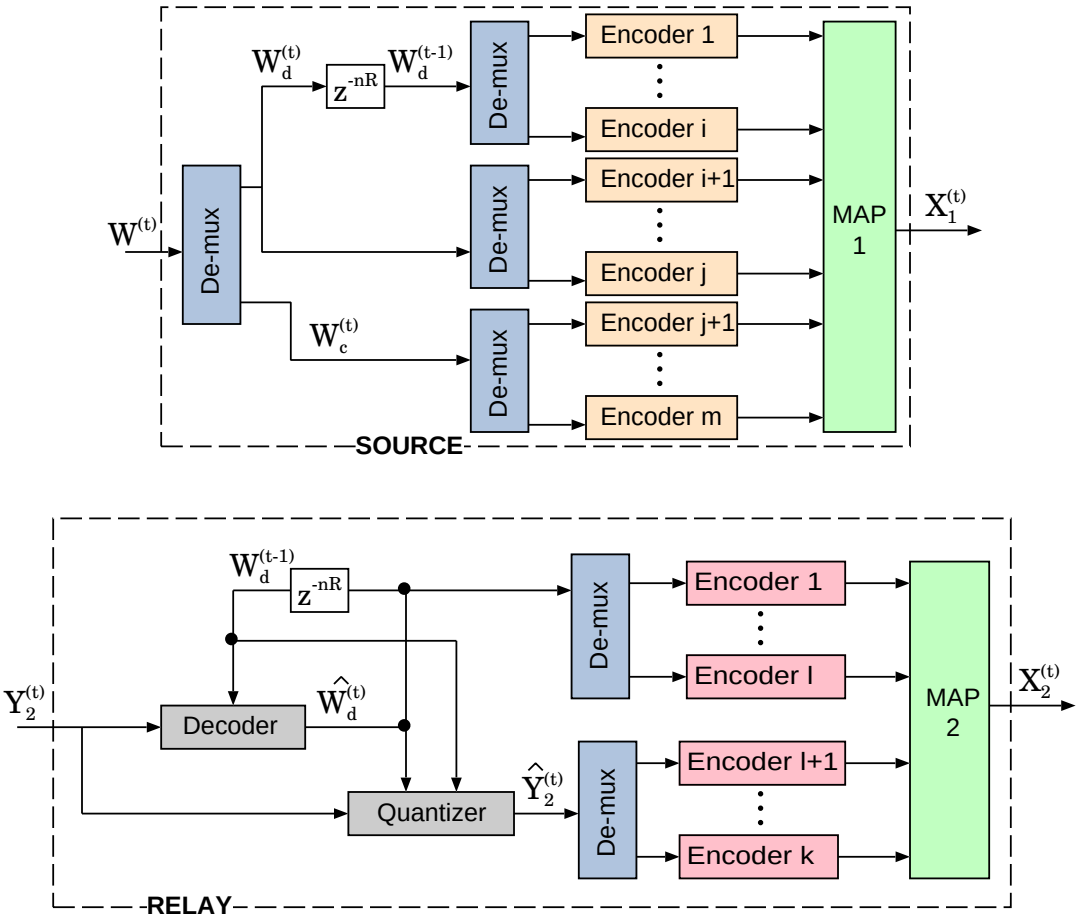

Figure 4.5. Multilevel coding of DCF in the Full-duplex relay

Remark 14. *In the proposed multilevel DCF, we assume that each level is responsible for a specific task. This is a restriction since a more general case can allow each level to perform both techniques. This restriction potentially might lead to a performance penalty. However, we have verified via extensive simulations that when DCF outperforms DF and CF, this restriction does not negatively affect the DCF rate.*

Remark 15. *In addition to level-assignment, other labeling variations are also possible, e.g., natural versus Gray labeling. Simulations in the sequel indicate that natural labeling is in general preferable to Gray labeling.*

4.4 Simulations

In this section, we produce bit error rate (BER) and frame error rate (FER) to assess the performance of the proposed multilevel coding.

In general, any kind of code can be used to implement the individual encoders at the source and the relay nodes. For the decode-forward components, we use the DVB-S2 LDPC codes. For the compress-forward components, we use the polar codes designed for compress-forward by Blasco-Serrano [85, 86]. The optimal code rate of each level is obtained via an exhaustive search to maximize the rate. The blocklength of the component codes is $n = 64$k. Both the relay and destination nodes used belief propagation decoding at each level where the maximum number of iterations is set to 50.

For simplicity of notations, in the following, we use $P(x)$ to denote $P_X(x)$. While decoding level i of the signal X_1 at the relay, the relay knows two parts of X_1 already. The first is the source assistance to the relay and the second is the output of the preceding decoders at the relay, assuming correct decoding. Therefore, the LLR of level i at the relay is

$$LLR_r = \log \frac{P(y_2|c^k, b^{i-1}, 0)}{P(y_2|c^k, b^{i-1}, 1)} \tag{4.40}$$

where

$$P(y_2|\hat{c}, b^{i-1}, b_i) = \frac{1}{P(\hat{c}, b^{i-1}, b_i)} \sum_{b_{i+1}^m} P(y_2|\hat{c}, \hat{b})$$

The relay removes the effect of $W_d^{(t-1)}$ and $W_d^{(t)}$ from Y_2 by first encoding them and removing their effect from X_1.

The decoding at the destination node is performed as follows: After the last transmission block, the destination decodes the decode-forward signal from the relay and then decode the signal from the source node, the LLR of level i of the relay at the destination node is

$$LLR_{RD} = \log \frac{P(y_3|c^{i-1}, 0)}{P(y_3|c^{i-1}, 1)} \tag{4.41}$$

where

$$P(y_3|c^{i-1}, c_i) = \frac{1}{P(c^{i-1}, c_i)} \sum_{b^m, c_{i+1}^m} P(y_3|b^m, c^m) \tag{4.42}$$

The next step is to decode the signal from the source given the transmitted signal from the relay with

$$LLR_{SD} = \log \frac{P(y_3|c^m, b^{i-1}, 0)}{P(y_3|c^m, b^{i-1}, 1)} \tag{4.43}$$

where

$$P(y_3|c^m, b^{i-1}, b_i) = \frac{1}{P(c^m, b^{i-1}, b_i)} \sum_{b_{i+1}^m} P(y_3|b^m, c^m)$$

In each of the error plots, a capacity threshold is marked that corresponds to the relay DCF constellation constrained capacity. The source and relay powers are identical throughout all simulations, enabling the use of a single scale for power (dB) in the error curves.

Fig. 4.6 shows the bit error probability and frame error probability for a 16-QAM source and QPSK relay. The figure compares the performance of DCF under different labelings. The system model is the same as the one considered in Fig. 4.4 where the source, relay and destination are all on one line. The distance between the source and relay is 0.8, the distance between the relay and destination is 0.2 and the distance between the source and destination is 1. The path-loss coefficient is $\alpha = 2$.

The total transmission rate is 2.2 bits/s/Hz. The four levels of the source operate as follows: The most significant bit transmits the same information of the most significant

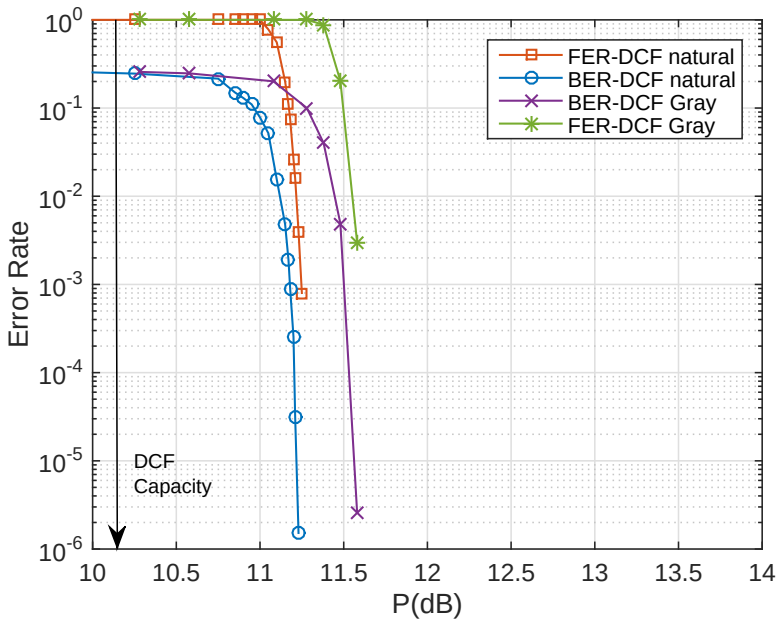

Figure 4.6. Performance of Multilevel superposition for 8-PAM constellation where $d = 2.5$

bit of the relay to provide a beamforming gain to the relay-destination transmission. The following two levels at the source transmit new information to be decodable at the relay while the last level transmits a compress-forward component.

Fig. 4.7 shows the bit error probability and frame error probability for a 16-QAM source and QPSK relay. The model that is considered this time is given by the channel gains, more specifically, $H_{12} = 1$, $H_{23} = 100$ and $H_{13} = 1$. This figure compares the proposed DCF with DF and CF. The rate is 2 bits/s/Hz. The figure shows that DCF outperforms DF and CF.

4.5 Discussion

The present work studied the hybrid decode-forward compress-forward strategy. This work showed certain cases where this combination can actually exceed the rates of decode-forward and compress-forward leading to higher rates in the constellation constrained full-duplex relay channel.

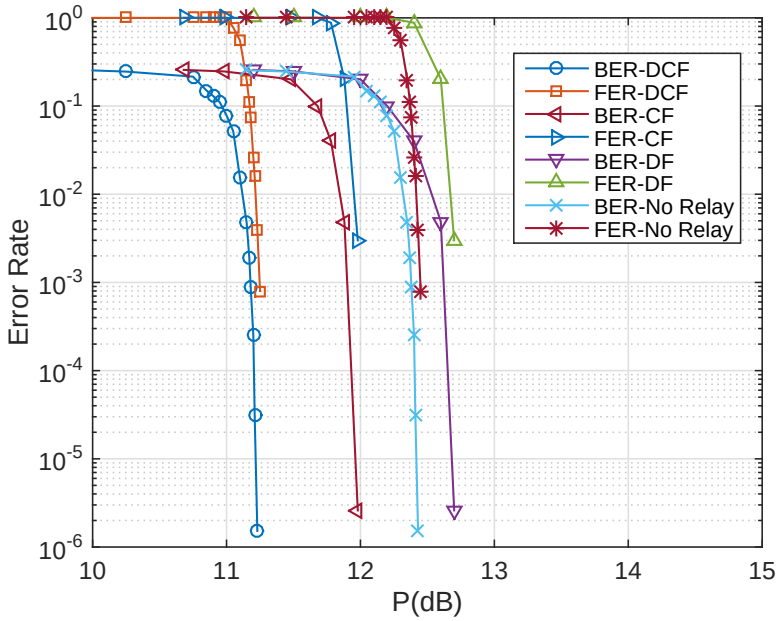

Figure 4.7. Performance of multilevel DCF vs. DF, CF, and no-relay.

This work presented a multilevel coding that systematically implements decode-compress-forward. The proposed multilevel coding was shown to approach the rates achieved by decode-compress-forward in the constellation constrained AWGN full-duplex relay channel.

The main points of this chapter can be summarized as follows: Each of the two widely known transmission techniques, decode-forward and compress-forward impose their own constraints that limit the rate under certain channel conditions. The decode-forward technique enforces the source to transmit the message with a rate that allows the relay to decode its message. Compress-forward enforces the relay not to decode the received signal even if the relay can actually decode it or partially remove the noise. Decode-compress-forward let the source split the message so that part of it to be decoded at the relay and the other part to be compressed.

This chapter studied one case where decode-compress-forward can improve the cooperation however, it is interesting to explore other cases of network cooperation and see if this technique can actually provide higher rates.

4.6 Appendix

4.6.1 Achievable Rate of DCF in the AWGN Relay Channel

First we start by obtaining the optimal value of $\hat{N}$. In the AWGN relay channel with all codewords normally distributed, the compress-forward constraint in (4.18) becomes

$$\frac{1}{2}\log\left(1+\frac{|H_{12}|^2P_{1c}+\sigma_2^2}{\hat{N}}\right)\leq\frac{1}{2}\log\left(1+\frac{|H_{23}|^2P_{2c}}{|H_{13}|^2(P_{1d}+P_{1c})+\sigma_3^2}\right) \tag{4.44}$$

The following value $\hat{N}$ is the value that satisfies (4.44) with equality [87, 88], and hence,

$$\hat{N}=\frac{(|H_{12}|^2P_{1c}+\sigma_2^2)(|H_{13}|^2(P_{1d}+P_{1c})+\sigma_2^2)}{|H_{23}|^2P_{2c}} \tag{4.45}$$

Now, we calculate the two terms that represent the decode-forward bound. Given that $Y_2=H_{12}X_1+n_2$,

$$\begin{aligned}I(U_{1d};Y_2|U_{2d})&=h(Y_2|U_{2d})-h(Y_2|U_{1d},U_{2d})\\&=h(U_{1d}+U_{1c})-h(U_{1c})\\&=\frac{1}{2}\log\left(1+\frac{|H_{12}|^2P_{1d}}{|H_{12}|^2P_{1c}+\sigma_2^2}\right)\end{aligned} \tag{4.46}$$

The other term in the decode-forward bound is $I(U_{1d}+\beta u_{2d},U_{2d};Y_3)$. Given that $Y_3=H_{13}X_1+H_{23}X_2+n_3$, this term can be obtain evaluated as follows:

$$\begin{aligned}I(U_{1d}+\beta U_{2d},U_{2d};Y_3)&=h(Y_3)-h(Y_3|U_{1d},U_{2d})\\&=\frac{1}{2}\log\left(\frac{|H_{12}|^2P_1+|H_{23}|^2P_2+\sigma_3^2}{|H_{13}|^2P_{1c}+|H_{23}|^2P_{2c}+\sigma_3^2}\right)\end{aligned} \tag{4.47}$$

after some mathematical manipulations, we can write

$$I(U_{1d} + \beta U_{2d}, U_{2d}; y_3) =$$
$$\frac{1}{2}\log\Big(1 + \frac{(P_{1d} + \beta^2 P_{2c})|H_{13}|^2}{\sigma_3^2} + \frac{P_{2d}|H_{23}|^2}{\sigma_3^2} + 2\rho\sqrt{\frac{(P_{1d} + \beta^2 P_2)P_{2d}|H_{13}|^2|H_{23}|^2}{\sigma_3^4}}\Big)\Bigg\} \quad (4.48)$$

The rate of the compress-forward part is given by

$$I(X_1; \hat{Y}_2, Y_3|X_2, U_{1d}) = h(\hat{Y}_2, Y_3|X_2, U_{1d}) - h(\hat{Y}_2, Y_3|X_2, U_{1d}, X_1) \quad (4.49)$$

where U_{1d} is in the given expression because the assumption of decoding the decode-forward part first. By treating $[\hat{Y}_2 Y_3]$ as a random vector, from the covariance matrix, the entropies in (4.49) can be calculating and give

$$I(X_1; \hat{Y}_2, Y_3|X_2, U_{1d}) =$$
$$\frac{1}{2}\log\Big(\frac{(|H_{12}|^2 P_{1c} + \sigma_3^2 + \hat{N})(|H_{13}|^2 P_{1c} + \sigma_3^2) + (|H_{12}|^2|H_{13}|^2)P_{1c}^2}{(\sigma_3^2 + \hat{N})\sigma_3^2}\Big) \quad (4.50)$$

By combining these mutual informations and substituting in the rate described in 4.17, we obtain the rate in Theorem 3.

CHAPTER 5

CODED MODULATION FOR THE WIRETAP CHANNEL

5.1 Introduction

Perfect secrecy can be achieved by sharing a secret key k with number of bits equal to the number of bits in the transmitted message W [89]. However, if the transmitter can securely share with the receiver a key that has the same quantity of information in the message, then it is better to send the message instead of the key. For that reason, practical secrecy involved variations of this technique that maximizes the effort and time required by the eavesdropper in order to decode the message. Wyner [90] has introduced the wiretap channel, Fig. 5.1, and showed that it is possible that Alice sends a message to Bob and Eve get zero information without sharing a secret key.

Similar to Shannon's random coding argument, Wyner showed that secrecy can be achieved by random binning. Guidelines for designing practical codes has been a research interest. The design criteria is based on two metrics:

- **Reliability**

  Bob should be able to recover the message with arbitrarily low error probability

$$\lim_{n\to\infty} P(\hat{W} \neq W) \to 0 \tag{5.1}$$

- **Secrecy**

  Eve should not reveal any information about the transmitted message

$$H(W|Z^n) = H(W) \tag{5.2}$$

The reliability of practical codes has been studied in depth for different classes of codes. For example, density evolution and EXIT charts can predict the performance of the LDPC

codes. However, the secrecy metric is more complicated since the calculation of the entropy of the transmitted message is not easy to calculate specially that the secrecy condition assumes infinite computational power at the eavesdropper. Therefore, this conditional entropy should be calculated under maximum likelihood decoding.

Thangaraj et al. [7] showed that nested codes can be used to achieve the secrecy capacity and gave examples for constructing such codes from LDPC ensambles. Rathi et al. [8] presented a method for constructing nested LDPC codes through multi-edge type LDPC codes. The main idea in these techniques is that the reliability can be understood from the existing techniques in the literature and the secrecy can be understood from the conditional entropy in (5.2). For the binary erasure channel, the conditional entropy can be calculated by the so called Maxwell construction [91]. LDPC codes have been also used to achieve the strong secrecy capacity of the binary erasure channel [9].

For the AWGN wiretap channel, the bit-error rate was used as a metric to design the codes [10] and analyzing the security gap that was first introduced in [92]. The security gap is defined as the difference between the minimum SNR required by the legitimate receiver to decode the message reliably and the maximum SNR at the eavesdropper that derives the bit error rate to 0.5. Examples of the techniques that were designed according to this metric with the aim of minimizing the security gap are puncturing [10], scrambling of systematic codes [93] and a combination of scrambling, concatenation and hybrid automatic retransmission request (HARQ) [94]. Other techniques such as polar codes [95] have used the stronger metric of the weak secrecy, namely, the leakage in the mutual information to build codes for the wiretap channel [96, 97]. However, polar codes have an error exponent that scales with $\sqrt{n}$ where n is the block length. For the non-binary input, lattice codes were studied for the AWGN wiretap channel [98, 99] and they were also considered for the cooperative jamming [100].

Coded modulation in the wiretap channel remains an open problem that we try to solve in this chapter. Different techniques of coded modulation can be used. BICM [15] and

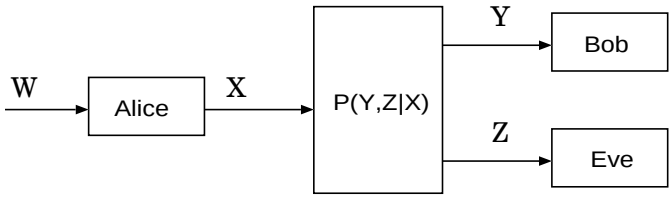

Figure 5.1. General wiretap channel

multilevel coding [16, 21]. In BICM the binary data is first encoded and then the codeword is randomly interleaved before it is mapped to symbols from the constellation. This implies that if BICM is to be used for the wiretap channel, binary encoding is necessary and it can follow any of the previously mentioned approaches.

In this chapter we use multilevel coding in the wiretap channel to achieve the constellation constrained capacity of the wiretap channel. We show that under multilevel coding, it is possible to achieve perfect secrecy of each binary level even if all the other levels are known to the wiretapper. We present an explicit transmission that uses only point-to-point codes in each level without nesting and achieves the entire rate-equivocation region. Simulation results show that good point-to-point LDPC codes can be used as component codes in multilevel coding and achieves very good performance.

The chapter is organized as follows: In Section 5.2, we present a systematic way of splitting the message from the randomness during the encoding and show that it achieves the rate-equivocation region of the general wiretap channel. In Section 5.3, we present a multilevel coded modulation structure that achieves the rate-equivocation region by encoding the message and the randomness through distinct levels. In Section 5.4 we present simulation results using point-to-point LDPC codes.

5.2 Independent Encoding of the Message and Randomness

The rate splitting technique is shown in Fig. 5.2, briefly, the message W and the randomness D are encoded independently with rates R and R_e producing two independent codewords

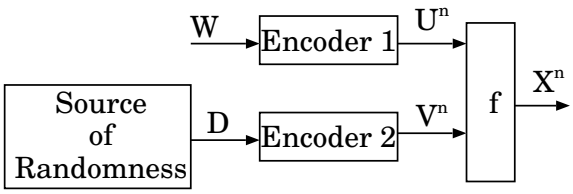

Figure 5.2. Secrecy Splitting in the Wiretap Channel

U^n and V^n respectively. The outputs of the encoders at each time instant are combined with the function $f : \mathcal{U} \times \mathcal{V} \to \mathcal{X}$ to produce the vector X^n where $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{X}$ are the alphabets of U, V and X respectively. The goal of this Section is to show that encoding the message and the randomness does not incur any loss in the achievable rate-equivocation region of the wiretap channel. The following Theorem presents the main result of this Section.

Theorem 4. *The rate-equivocation region of the independently encoded message and randomness is given by*

$$\mathcal{R} = \bigcup_{P_U P_{X|U} P_{YZ|X}} \left\{ \begin{array}{l} (R, R_e) \\ R \leq I(U;Y) \\ R_e \leq R \\ R_e \leq I(U;Y) - I(U;Z) \end{array} \right\} \tag{5.3}$$

Proof. Since independent encoding of W and D is one special case of encoding in the wiretap channel, the best upper bound on the rate-equivocation region is also an upper bound on the rate-equivocation region under independent encoding. Therefore, we consider the same converse proof of the general wiretap channel which is omitted for brevity.

The Achievability proof is more involved and follows the following codebook construction, reliability analysis and equivocation analysis.

Codebook Generation:

Construct the codebooks

$$\mathcal{C}_u = \{U_i^n, \quad i = 1, \ldots, 2^{n(I(X;Y)-I(X;Z))}\} \tag{5.4}$$

$$\mathcal{C}_v = \{V_j^n, \quad j = 1, \ldots, 2^{nI(X;Z)}\} \tag{5.5}$$

According to the distribution

$$P_{U_i^n}(u^n) = \prod_j P_{U_i j}(u_j) \tag{5.6}$$

$$P_{V_i^n}(v^n) = \prod_j P_{V_i j}(v_j) \tag{5.7}$$

respectively. Therefore, $u^n \in T_\epsilon^n(P_{U^n})$ and $v^n \in T_\epsilon^n(P_{V^n})$ mean that u^n is strongly typical according to P_{U^n} and v^n is strongly typical according to P_{V^n}. The function f maps each vector U^n and V^n into a vector X^n. This mapping leads to a codebook $\mathcal{C}_x$ that contains all the possible codewords X^n.

Encoding:

For a given message $w \in [1 : 2^{nR}]$, a codeword $U^n(w)$ is selected. In each transmission, a uniformly distributed random variable $d \in [1 : 2^{nR_e}]$ is generated and mapped to the codeword $V^n(d)$. The random variable D represents the randomness in the transmission. A function $f : (u_i, v_i) \to x_i$ combines the vectors U^n and V^n to generate X^n based on a sample by sample mapping.

Decoding:

The legitimate receiver obtains the vector Y^n and find an estimate to the message $\hat{w}$ and the random variable $\hat{d}$ such that

$$(U^n(\hat{w}), V^n(\hat{d}), Y^n) \tag{5.8}$$

are jointly typical for some $\hat{w}$ and $\hat{d}$.

Probability of Error Analysis:

According to the LLN and the covering lemma, the probability of decoding error at the legitimate receiver $P_e \to 0$ as $n \to \infty$ as long as

$$R + R_e \leq I(X;Y) \tag{5.9}$$

which is satisfied in this case.

Equivocation Calculations:

The equivocation at the eavesdropper is

$$H(W|Z^n) = H(W, Z^n) - H(Z^n) \tag{5.10}$$

$$= H(W, Z^n, X^n) - H(X^n|W, Z^n) - H(Z^n) \tag{5.11}$$

$$= H(W, X^n) + H(Z^n|W, X^n) - H(X^n|W, Z^n) - H(Z^n) \tag{5.12}$$

$$\geq H(X^n) + H(Z^n|X^n) - H(X^n|W, Z^n) - H(Z^n) \tag{5.13}$$

Now, we study the four terms in (5.13).

For the first term in (5.13), we need the following Lemma from [101].

Lemma 1. *Consider a discrete random variable X taking on the mass points $x_1, \ldots, x_m$ with probability mass function satisfying*

$$\frac{Pr\{X = x_i\}}{Pr\{X = x_j\}} \leq 2 \times 2^{\delta}, \quad \forall i, j \in [1, \ldots, k]. \tag{5.14}$$

Then,

$$H(X) \geq \log k - \delta - 1 \tag{5.15}$$

Assuming that the function f leads to a one-to-one mapping from $U^n \times V^n \to X^n$, then all codewords in X^n are equi-probable if $\mathcal{C}_u$ and $\mathcal{C}_v$ have equi-probable codewords. Under this assumption,

$$\frac{\Pr\{X^n = x^n\}}{\Pr\{X^n = x'^n\}} \leq 2 \quad \forall x^n, x'^n \in \mathcal{C}_x \tag{5.16}$$

therefore,

$$\frac{1}{n}H(X^n) \geq \frac{1}{n}\log|\mathcal{C}_u| + \frac{1}{n}\log|\mathcal{C}_v| - \frac{1}{n} \tag{5.17}$$

$$= I(X;Y) - \frac{1}{n} \tag{5.18}$$

Remark 16. *Please note that the bound* (5.18) *depends only on the distribution of the codewords* X^n *in the codebook* $\mathcal{C}_x$*. The one-to-one mapping* $U^n \times V^n \to X^n$ *is only a sufficient condition for* (5.16) *but not necessary. One can find a function* f *that does not follow the one-to-one mapping property and still obtain* (5.16) *but the appropriate number of vectors* U^n *and* V^n*.*

For the second term,

$$\frac{1}{n}H(Z^n|X^n) = \frac{1}{n}\sum_{x^n} \Pr\{X^n = x^n\} H(Z^n|X^n = x^n) \tag{5.19}$$

$$= \frac{1}{n}\sum_{x^n} \Pr\{X^n = x^n\} \sum_{a\in\mathcal{X}} N(a|x^n) \sum_{z} -P(z|a)\log p(z|a) \tag{5.20}$$

$$\geq \sum_{x^n} \Pr\{X^n = x^n\} \sum_{a\in\mathcal{X}} (P(X=a) - \epsilon) \tag{5.21}$$

$$\times \sum_{z} -P(z|a)\log P(z|a) \tag{5.22}$$

$$= \sum_{x^n} \Pr\{X^n = x^n\} (H(Z|X) - O(\epsilon)) \tag{5.23}$$

$$= H(Z|X) - O(\epsilon) \tag{5.24}$$

Please note that (5.24) does not depend on the function f.

For the third term, in a manner similar to [102] define

$$\rho(W, Z^n) = \begin{cases} x^n_{w,d} & \text{if } \exists d \quad \text{s.t.} \quad (x^n_{w,d}, z^n) \in T^n_\epsilon(P_{XZ}) \\ \text{arbitrary}, & \text{otherwise} \end{cases} \tag{5.25}$$

Then

$$Pr\{X^n \neq \rho(W, Z^n)\} = \sum_{w,b} Pr\{x^n_{w,d}\} Pr\{x^n_{w,d} \neq \rho(W, z^n)|W, D\} \tag{5.26}$$

$$= \lambda_1 \leq \epsilon \tag{5.27}$$

and by Fano's inequality, we obtain

$$\frac{1}{n}H(X^n|W, Z^n) \leq \frac{1}{n}(1 + \lambda_1 I(X;Y)) < \epsilon_2 \tag{5.28}$$

The last inequality means that if the wiretapper have an access to the message W, then no more ambiguity is left in X^n.

For the fourth term,

$$\frac{1}{n}H(Z^n) \leq \frac{1}{n}\log|T_\epsilon^n(P_Z)| \tag{5.29}$$

$$\leq H(Z) + \epsilon \tag{5.30}$$

By substitution in (5.13), we have

$$\frac{1}{n}H(W|Z^n) \geq I(X;Y) + H(Z|X) - H(Z) - \epsilon_4 \tag{5.31}$$

$$= I(X;Y) - I(X;Z) - \epsilon_4 \tag{5.32}$$

Finally, we get

$$R_e \leq I(X;Y) - I(X;Z) \tag{5.33}$$

By now, we can show that the following rate-equivocation region is achievable

$$\mathcal{R} = \bigcup_{P_X P_{YZ|X}} \left\{ \begin{array}{l} (R, R_e) \\ R \leq I(X;Y) \\ R_e \leq R \\ R_e \leq I(X;Y) - I(X;Z) \end{array} \right\} \tag{5.34}$$

Consider a DMC from U to X with the conditional distribution $P_{U|X}$, one can show that the rate-equivocation region is now

$$\mathcal{R} = \bigcup_{P_U P_{X|U} P_{YZ|X}} \left\{ \begin{array}{l} (R, R_e) \\ R \leq I(U;Y) \\ R_e \leq R \\ R_e \leq I(U;Y) - I(U;Z) \end{array} \right\} \tag{5.35}$$

Please note that this is the same expression for the rate-equivocation region for the general wiretap channel however, the input distribution $P_U P_{X|U}$ is restricted due to the proposed structure. Therefore, if the optimizing input distribution for the general wiretap channel can be achieved under this restriction, then independent encoding of the message and randomness will achieve the same rate-equivocation region of the general wiretap channel.

□

Another way to think about the condition on the function f to obtain weak secrecy can be described in the following argument. The weak secrecy condition is equivalent to having $\frac{1}{n}I(U^n; Z^n) = 0$ since $H(W) = H(U^n)$. Therefore,

$$I(X^n; Z^n) = I(U^n; Z^n|V^n) \tag{5.36}$$

$$H(X^n) - H(X^n|Z^n) = H(U^n|V^n) - H(U^n|Z^n, V^n) \tag{5.37}$$

$$H(X^n) - H(X^n|Z^n) = H(U^n) - H(U^n|Z^n, V^n) \tag{5.38}$$

$$H(X^n) - H(X^n|Z^n) = H(U^n) + H(V^n) - H(X^n|Z^n) \tag{5.39}$$

and hence,

$$H(X^n) = H(U^n) + H(V^n) \tag{5.40}$$

where the condition (5.40) implies that the function f should lead to a one-to-one mapping from U^n and V^n to X^n.

The following is an example that shows the existence of such a function f that satisfies the conditions in Theorem 4.

Example 1. *Assume that*

$$\mathcal{U} = \mathcal{V} = \mathcal{X} = \{0, 1\}$$

and that the function f =XOR(U, V). The codebook $\mathcal{C}_x$ should be optimal for the legitimate receiver to be able to decode W and D. In addition, according to the enhanced decoder

concept, the codebook $\mathcal{C}_u$ should be optimal for the eavesdropper given D. Now, assume that the optimal input distribution to both the legitimate receiver channel and the eavesdropper channel is uniform then, a proper choice for P_U is to be uniform which will lead to uniform input distribution. Now, it remains to show that the mapping $U^n \times V^n \to X^n$ is a one-to-one mapping. This will depend on the distribution of the bits in V^n because V^n should be sufficiently sparse.

The previous example shows that encoding for the binary input wiretap channel can take the form of a superposition code where the cloud centers of the code represent the source of randomness while the satellite codewords represent the message.

Remark 17. *Another perspective at the proposed independent encoding of the message and randomness can be looked at as a rate-splitting approach. The splitting of the message and randomness creates two multiple-access channels. The first MAC is given by $P_{Y|U,V}$ and the second MAC is given by $P_{Z|U,V}$. The capacity region of each multiple access channel is shown in Fig. 5.3. Every point on the boundary of the capacity region of the legitimate receiver represents a point on the rate-equivocation region. For example, point (a) represents the secrecy capacity rate.*

$$R_M = I(X;Y) - I(X;Z) \tag{5.41}$$

$$R_D = I(X;Z) \tag{5.42}$$

Any point on the boundary of the capacity region of the legitimate receiver to the right of point (a) represents more transmission rate and less security while any point to the 45 degrees line in the equivocation region. This different perspective shows that the proper rate allocation between the message and the randomness determines which point on the rate-equivocation region the system is operating. For example, to operate at the secrecy capacity This shows that the independent encoding of the message and the randomness provides a systematic way

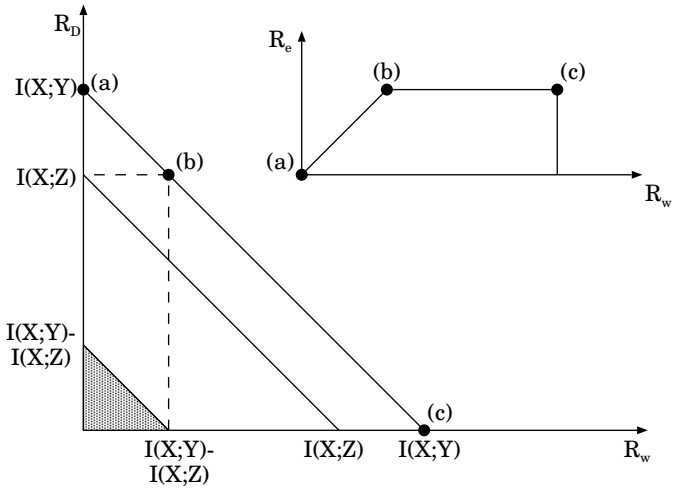

Figure 5.3. Message, Randomness rate-region with the corresponding equivocation region

of operating at any point on the rate-equivocation region by using the appropriate code rates at each encoder.

Independent encoding of the message and the randomness can have several benefits such as using point-to-point codes which are easier to design than nested codes. It can also have benefits in encoding for multi-node networks with a wiretapper.

5.3 Multilevel Coding in the Wiretap Channel

In a general sense, multilevel coding in the wiretap channel can split the actual message W and the randomness $\bar{W}$ into m sub-streams. Each pair of streams W_i and $\bar{W}_i$ are encoded by encoder i whose output drives the ith input to the mapper where $i \in \{1, \ldots, m\}$. Each encoder i should take into account the reliability and secrecy conditions. As mentioned in the introduction, it is challenging to design such codes that require mixing of information and randomness in the same codeword with a guarantee on the equivocation in the AWGN channel. Even if this difficulty is ignored, there is another important task which is to find the optimal rate of W_i and $\bar{W}_i$, for every level i. The most optimistic solution to this optimization

problem is to isolate the message sub-streams in certain levels and the randomness sub-streams in the other levels so that we can avoid the joint design of binary codes that satisfy reliability and secrecy conditions. Formally speaking,

$$R_i = 0 \quad \text{for} \quad i \in \bar{S} \tag{5.43}$$

$$\bar{R}_i = 0 \quad \text{for} \quad i \in S \tag{5.44}$$

for some set of levels S and its complement $\bar{S}$ where R_i is the transmission rate of W_i and $\bar{R}_i$ is the transmission rate of $\bar{W}_i$.

Due to the very attractive features of the construction resulting from the optimistic solution described earlier, we adopt this construction and show that it can be designed such that the possible rate loss is negligible.

The optimality of such construction can be understood by modeling the multilevel coding as a multiple-access channel and highlighting the difference. The multiple-access wiretap channel was introduced by Tekin and Yener [103, 104]. The strong secrecy capacity region of the multiple access wiretap channel was studied by Yassaee and Aref [105]. In these results, the secrecy conditions for a two-user multiple-access channel with messages W_1 and W_2 are

$$H(W_1, W_2|Z^n) = H(W_1, W_2) \tag{5.45}$$

$$H(W_1|Z^n) = H(W_1) \tag{5.46}$$

$$H(W_2|Z^n) = H(W_2) \tag{5.47}$$

The intuitive idea that makes the proposed solution achieves the secrecy rate can be explained as follows: assume first a rate pair in the secrecy capacity region of the multiple-access wiretap channel, this rate pair can be achieved via sending the necessary randomness through each transmitter. However, if the sum-rate is all what matters then, the necessary randomness can be sent from any set of transmitters as long as the channel from these transmitters to the legitimate receiver can support the encoding rate of the randomness.

Since the levels of a multilevel encoder act in an independent manner, the multilevel coding transmitter can be modeled as the set of transmitters in a multiple-access channel with the number of transmitters equal to the number of levels. The corner points of the capacity region of this multiple access channel represent different decoding orders of the levels. Unlike the multiple-access channel, only the sum-rate matters. This is what allows the simple construction described earlier, in other words, it does not matter which level is sending the information and which level is sending randomness therefore, the necessary randomness, with rate $I(X;Z)$, can be sent through a subset of levels while the other levels send only information.

The question now is under what conditions there exists a set of levels that can support the necessary rate of the randomness? We begin by showing, in the following theorem, that if such condition exists, secrecy capacity is achieved and then present the cases under which these conditions are satisfied.

Theorem 5. *Multilevel coding can achieve the secrecy capacity via encoding the information and the randomness through distinct levels as long as:*

$$I(X;Z) = I(B_{\bar{S}};Y|B_K) \tag{5.48}$$

for some set of levels S and $K \subset S$ where $\bar{S}$ is the complement of S.

Proof. The converse proof has two components. The first component is to show that any rate higher than the secrecy rate will not result in a vanishing error probability. This component follows directly from the standard converse proof of the wiretap channel and is omitted for brevity.

The second component is to show that if the condition in (5.48) is not satisfied, then encoding the message and the randomness through distinct levels will not be possible to achieve the secrecy capacity. In order to prove this part, we first explain the roles of the sets

S and K. Assuming that the message and the randomness will be encoded through distinct levels S and $\bar{S}$ respectively. Then, the capacity between the levels $\bar{S}$ and the legitimate receiver should be equal to $I(X;Z)$ in order to accommodate the rate of the randomness. However, the capacity between the set of levels $\bar{S}$ and the legitimate receiver depends on the decoding order at the receiver. Assuming that the legitimate receiver will decode the levels in the set K first and conditionally decode $\bar{S}$, then the capacity between the set of levels $\bar{S}$ and the legitimate receiver is given by

$$C_{\bar{S}} = I(B_{\bar{S}}; Y|B_K) \tag{5.49}$$

for some set $K \subset S$.

The achievability proof on the other hand is as follows: Assume that the input levels to the mapper are divided into two sets, S and its compliment $\bar{S}$, assume also for simplicity that the set $S = \{1, 2, \ldots, |S|\}$ and $\bar{S} = \{|S|+1, \ldots, m\}$. This last assumption will simplify the notation in the rest of the proof and does not restrict the construction since the indexes in S and $\bar{S}$ can represent any level.

Codebook Generation:

Construct the following codebooks,

$$\mathcal{C}_g^{(i)} = \{B_g^{(i)n}, g = 1, \ldots, G_i\} \quad \text{for} \quad i \in S$$
$$\mathcal{C}_h^{(i)} = \{B_h^{(i)n}, h = 1, \ldots, H_i\} \quad \text{for} \quad i \in \bar{S}$$

Each codeword in every codebook is generated independently according to

$$P_{B_g^{(i)n}} = \prod_{j=1}^{n} P_{B_{gj}^{(i)}}(b_{gj}^{(i)}) \tag{5.50}$$

where $B_{gj}^{(i)}$ is the symbol in location j in the vector $B_g^{(i)n}$. Please note that the subscripts g and h in the codebooks are to distinguish between the codebooks that will be used for the

transmission of the message M, sub-scripted with g and the codebooks that will be used for the transmission of the randomness D, sub-scripted with h.

The size of each codebook is restricted by the following:

$$\frac{1}{n}\log H_i \leq I(B^{(i)}; Y|B^{(1)}, \ldots, B^{(i-1)}, B^{(i+1)}, B^{(k)}) \tag{5.51}$$

$$\begin{aligned}\frac{1}{n}\log G_i \leq & I(B^{(|S|+i)}; Y| \\ & B^{(1)}, \ldots, B^{(|S|+i-1)}, B^{(|S|+i+1)}, B^{(k)})\end{aligned} \tag{5.52}$$

$$\frac{1}{n}\sum_i \log G_i = I(X;Y) - I(X;Z) \tag{5.53}$$

$$\frac{1}{n}\sum_i \log H_i = I(X;Z) \tag{5.54}$$

The last two inequalities guarantee that no level will transmit above its capacity.

Encoding:

Split the message $W \in [1 : 2^{nR}]$ into $|S|$ sub-message $W_i \in [1 : 2^{nR_i}]$ where $i \in S$ and encode W_i through level i into a codeword $B_g^{(i)n}(W_i)$. Split the random sequence $D \in [1 : 2^{nR_2}]$ into $|\bar{S}|$ sub-sequences $D_j \in [1 : 2^{nR_j}]$ where $j \in \bar{S}$ and encode D_j through level j into a codeword $B_g^{(j)n}(D_j)$. The mapper then generates X^n via one-to-one mapping from $B_h^{(i)n}(W_i)$ and $B_g^{(j)n}(D_j)$.

Decoding:

The legitimate decoder decodes the transmitted message by finding jointly typical sequences

$$\begin{aligned}\big(&B^{(1)n}(\hat{w}_1), \ldots, B^{(|S|)n}(\hat{w}_{|S|}), \\ &B^{(|S|+1)n}(\hat{d}_{|S|+1}), \ldots, B^{(m)n}(\hat{d}_m), Y^n\big)\end{aligned}$$

for some estimates $\hat{m}_i$ and $\hat{d}_i$

Reliability Analysis: According to the LLN and the covering Lemma, the probability of decoding error at the legitimate receiver $Pe \to 0$ as $n \to \infty$ as long as

$$R + R_2 \leq I(X;Y) \tag{5.55}$$

which is satisfied by the codebook construction.

Equivocation Analysis:

The equivocation at the eavesdropper is

$$\begin{aligned} H(W|Z^n) &= H(W, Z^n) - H(Z^n) && (5.56)\\ &= H(W, Z^n, X^n) - H(X^n|W, Z^n) - H(Z^n) && (5.57)\\ &= H(W, X^n) + H(Z^n|W, X^n) - H(X^n|W, Z^n) \\ &\quad - H(Z^n) && (5.58)\\ &\geq H(X^n) + H(Z^n|X^n) - H(X^n|W, Z^n) \\ &\quad - H(Z^n) && (5.59) \end{aligned}$$

Now, we study the four terms in (5.59).

To upper bound the first term in (5.59), we use lemma 1 from [101]. Since the messages W_is and the random sequences D_is are uniformly distributed and the mapping function is one-to-one, the codebook X^n has equally probable codewords and hence,

$$\frac{\Pr\{X^n = x^n\}}{\Pr\{X^n = x'^n\}} \leq 2 \quad \forall x^n, x'^n \in \mathcal{X} \tag{5.60}$$

therefore,

$$\begin{aligned} \frac{1}{n}H(X^n) &\geq \frac{1}{n}\sum_i \log G_i + \frac{1}{n}\sum_i \log H_i - \frac{1}{n} && (5.61)\\ &= I(X;Y) - \frac{1}{n} && (5.62) \end{aligned}$$

Please note that the restriction on the mapping function requires a one-to-one mapping between the vectors $B_g^{(i)n}$ and $B_h^{(i)n}$ to the vector X^n and not a one-to-one mapping from

the individual samples. This means that the mapping function might not be a one-to-one function on a sample by sample basis and still satisfies the conditions necessary for the previous upper bound. An example of such function is an XOR function used for binary superposition in the broadcast channel [1].

For the second term,

$$\frac{1}{n}H(Z^n|X^n) = \frac{1}{n}\sum_{x^n}\Pr\{X^n = x^n\}H(Z^n|X^n = x^n) \tag{5.63}$$

$$= \frac{1}{n}\sum_{x^n}\Pr\{X^n = x^n\}$$

$$\times \sum_{a\in\mathcal{X}} N(a|x^n)\sum_z -P(z|a)\log p(z|a) \tag{5.64}$$

$$\geq \sum_{x^n}\Pr\{X^n = x^n\}\sum_{a\in\mathcal{X}}(P(X=a)-\epsilon)$$

$$\times \sum_z -P(z|a)\log P(z|a) \tag{5.65}$$

$$= \sum_{x^n}\Pr\{X^n = x^n\}(H(Z|X) - O(\epsilon)) \tag{5.66}$$

$$= H(Z|X) - O(\epsilon) \tag{5.67}$$

The previous bound does not depend on the mapping function but depends only on the distribution of X^n and the channel. The distribution of X^n is restricted by the distribution of the inputs to the mapper as well as the mapping function. However, as will be shown in the sequel, in most of practical cases, the mapping function does not impose any penalty on the optimal distribution of X^n.

For the third term in (5.59), in a manner similar to [102] define

$$\rho(W, z^n) = \begin{cases} z^n, & \text{if } \exists b \quad \text{s.t.} \quad (x^n, z^n) \in T_\epsilon^n(P_{XZ}) \\ \text{arbitrary}, & \text{otherwise} \end{cases} \tag{5.68}$$

Then

$$Pr\{X^n \neq \rho(W, Z^n)\} = \sum_{w,d} Pr\{x^n\} \tag{5.69}$$

$$\times Pr\{x^n \neq \rho(w, z^n)|w, d\} \tag{5.70}$$

$$= \lambda_1 \leq \epsilon \tag{5.71}$$

and by Fano's inequality, we obtain

$$\frac{1}{n}H(X^n|W, Z^n) \leq \frac{1}{n}(1 + \lambda_1 \log(AB)) < \epsilon_2 \tag{5.72}$$

The last inequality means that if the wiretapper have an access to the message W, then no more ambiguity is left in X^n.

For the fourth term,

$$\frac{1}{n}H(Z^n) \leq \frac{1}{n}\log|T_\epsilon^n(P_Z)| \tag{5.73}$$

$$\leq H(Z) + \epsilon \tag{5.74}$$

By substitution in (5.59), we have

$$\frac{1}{n}H(W|Z^n) \geq I(X;Y) + H(Z|X) - H(Z) - \epsilon_4 \tag{5.75}$$

$$= I(X;Y) - I(X;Z) - \epsilon_4 \tag{5.76}$$

$$= H(W) - \epsilon_4 \tag{5.77}$$

□

The condition in (5.48) implies that there is a set of levels $\bar{S}$ that can support the necessary rate of the randomness. This implies that the complement of this set, S, can support a message with a rate of $I(X;Y) - I(X;Z)$.

Theorem 5 shows that it is possible to encode the messages W_is and the randomness D_is through distinct levels while the eavesdropper is asymptotically ignorant about the transmitted messages. The following Corollary shows that Theorem 5 is also true under a stronger secrecy condition.

Corollary 1. *Theorem 5 is also true under the following stronger secrecy condition:*

$$H(W_i|Z^n, W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_m) = H(W_i). \tag{5.78}$$

Proof.

$$\begin{aligned}
&\frac{1}{n} I(W_i; Z^n | W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_m) \\
&= \frac{1}{n} I(W; Z^n) - \frac{1}{n} I(W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_m; Z^n) && (5.79) \\
&= -\frac{1}{n} I(W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_m; Z^n) && (5.80) \\
&= -I(B_2; Z) && (5.81) \\
&= 0 && (5.82)
\end{aligned}$$

where the last equality is due to the positivity of the mutual information. □

The stronger secrecy definition in Corollary 1 goes back to the original secrecy definition by Shannon [89] which implies that having an access to part of the message does not reveal any information about the rest of the message at the Eavesdropper.

In the following, we study the proposed transmission and the conditions under which (5.48) is satisfied under joint decoding and multistage decoding at the legitimate receiver while we always consider maximum likelihood decoding at the wiretapper.

What makes a difference in the achievable secrecy rates between the decoding strategies is the rate that can be supported by each level under different decoding strategies. For example, under joint decoding, each level can support any rate as long as it belongs to the capacity region of the equivalent MAC channel. Under multistage decoding on the other hand, the individual levels can only support the rates on the corner points of the capacity region. In the following, we consider only two levels at the transmitter for simplicity while the generalization to any number of levels is straightforward.

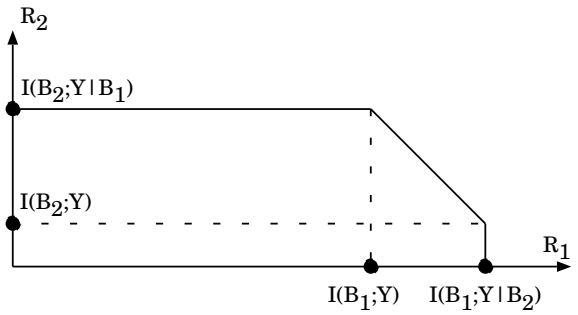

Figure 5.4. The capacity region for the channel between two-levels MLC to the legitimate receiver.

5.3.1 Proposed Transmission Under Joint Decoding

Assuming only two levels with outputs B_1 and B_2, the capacity region from these two levels to the legitimate receiver is shown in Fig. 5.4. For optimality of multilevel coding, the transmission rate is equal to what the channel of the legitimate receiver can support, in other words, the sum-rate constraint should be active. This means that the transmission rate pair should be any point on the line connecting the corner points in Fig. 5.4. The rates of the individual levels under joint decoding while the sum-rate constraint is active are

$$R_1 \in [I(B_1;Y), I(B_1;Y|B_2)] \tag{5.83}$$

$$R_2 = I(X;Y) - R_1 \tag{5.84}$$

or

$$R_2 \in [I(B_2;Y), I(B_2;Y|B_1)] \tag{5.85}$$

$$R_1 = I(X;Y) - R_2 \tag{5.86}$$

This means that if $I(X;Z)$ belongs to the set in (5.83) or (5.85), then a randomness with rate $I(X;Z)$ can occupy the first or second level respectively and the other level can send real information with rate $I(X;Y) - I(X;Z)$. The problem arises when $I(X;Z)$ does not belong to any of the sets, approximations should be made at this point.

There are two cases when this can happen. First, when

$$I(X;Z) < \min_i I(B_i;Y) \tag{5.87}$$

which means that the amount of necessary randmness is less than the rate that can be carried by the smallest-rate level. This means that this level will be wasted in sending randomness. The rate loss in this case is:

$$\min_i\{I(B_i;Y)\} - I(X;Z) \leq 1 \tag{5.88}$$

Second, when

$$I(X;Z) > \max\{I(B_1;Y|B_2), I(B_2;Y|B_1)\} \tag{5.89}$$

which means that the necessary randomness is more than the rate that the largest rate level can support. In both cases, the loss is upper bounded by 1 since the rate-loss is bounded by the largest rate that a level can support.

5.3.2 Proposed Transmission Under Multistage Decoding

Unlike joint decoding, multistage decoding can only achieve the corner points on the capacity region. Therefore, the rate pairs that can be supported are:

$$R_1 = I(B_1;Y) \tag{5.90}$$

$$R_2 = I(B_2;Y|B_1) \tag{5.91}$$

or

$$R_1 = I(B_1;Y|B_2) \tag{5.92}$$

$$R_2 = I(B_2;Y) \tag{5.93}$$

depending on the decoding order. Clearly, this is a huge restriction over the joint decoding case. Assuming that each level is either going to carry actual information or randomness

but not both, the necessary randomness rate $I(X;Z)$ should be equal to any of the rates in (5.90) to (5.93). If this equality is not satisfied, some approximations should be made as the joint decoding case and secrecy capacity will not be achieved.

The decoding order at the legitimate receiver depends on the value of $I(X;Z)$. For example, assume that

$$I(X;Z) \approx I(B_2;Y) \tag{5.94}$$

then, the corner point defined by the rate pair $(I(B_1;Y|B_2), I(B_2;Y))$ will be active where the first level carries the message and the second level carries the randomness. The legitimate receiver then decodes the second level and then the first level. However, if

$$I(X;Z) \approx I(B_2;Y|B_1) \tag{5.95}$$

then, the corner point defined by the rate pair $(I(B_1;Y), I(B_2;Y|B_1))$ will be active where the first level carries the message and the second level carries the randomness. The legitimate receiver then decodes level-1 first. It is worth noting here that the legitimate receiver does not need to decode the randomness which makes the proposed transmission even simpler.

Remark 18. *Iterative decoding of multilevel coding can approximate the joint decoding and achieve any point on the sum-rate constraint of the capacity region. Therefore, iterative decoding of the proposed transmission has the same treatment as joint decoding.*

Remark 19. *Increasing the constellation size leads to finer distribution of the total rate on the levels which allows more flexibility that helps the existence of the condition in* (5.48)*. Therefore, as the constellation size increases, more available secrecy rates can be obtained.*

The question now is whether or not we can increase the set of achievable secrecy rates for a given constellation size. In the following, we show that the labeling design can play a very important role in increasing the set of achievable secrecy rates.

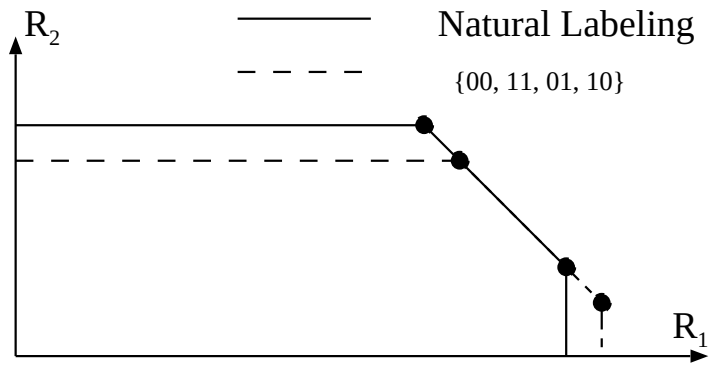

Figure 5.5. The capacity region for the channel between two-levels MLC to the legitimate receiver under two different labelings.

In order to explain this point, we need to shed the light on the fact that different labelings for the constellation might have different distribution of the total rate across the levels while maintaining the sum-rate. This observation can be shown in Fig. 3.6 where the rate of each level in a 4-PAM constellation in a point-to-point channel is shown for two different labelings. This leads to more degrees of freedom that can consequently reduce the rate loss.

Formally speaking, the two labelings described in Fig. 3.6 lead to different equivalent MAC channels as shown in Fig. 5.5. The capacity region of the two MAC channels have the sum-rate but different corner points. This leads to more possible corner points and hence, more possible values for the rates (5.90) through 5.93.

Remark 20. *Increasing the constellation size does not only result in a finer distribution of the total rate across the levels but also results in more possible labelings that lead to different equivalent MAC channels with different corner points. Therefore, increasing the constellation size leads to increasing the set of available secrecy rates.*

In Fig. 5.6, we find the achievable secrecy rate of the proposed multilevel coding for an 8-PAM transmission versus the power of the transmitter in dB while fixing the noise power at the legitimate receiver and the eavesdropper at 1 and 2 respectively. Please note that the constellation constrained secrecy capacity goes does after a certain value of the transmission

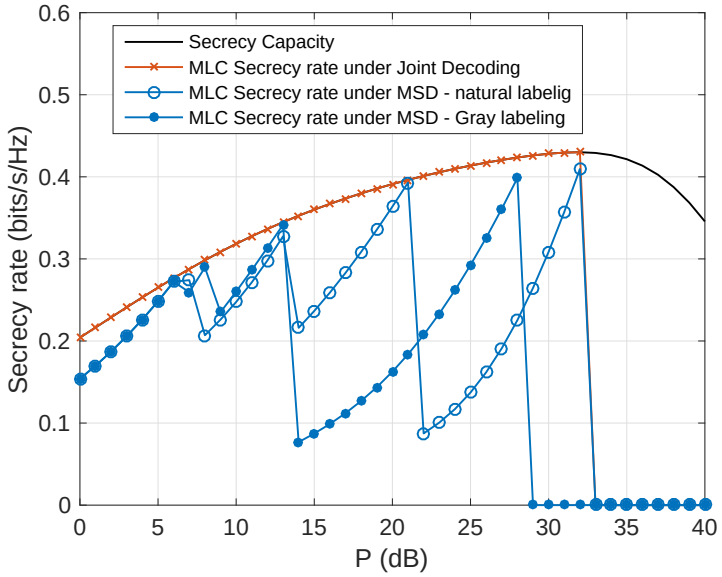

Figure 5.6. Achievable secrecy rate of the proposed multilevel coding under joint decoding and multi-stage decoding

power since the transmitted sequence becomes clear to both the legitimate receiver and the eavesdropper as noted in [106]. As the figure shows, the proposed multilevel coding achieves the constellation constrained secrecy capacity under joint decoding until the power, $P = 25$dBw. This is because after this point, the value of $I(X;Z)$ needs all the levels to be carried over the channel. The figure also shows that depending on the channel conditions, different labelings will provide different secrecy rates under multi-stage decoding as explained earlier.

5.3.3 The Rate-Equivocation Region

In this Section, we present how the proposed multilevel coding can extend to the entire rate-equivocation region. This needs to prove the achievable rate-equivocation region however, the proof goes along the same lines of the proof of Theorem 5 and we only present a sketch of the proof in this Section. The rate-equivocation region (see Fig. 5.7) consists of two parts: rates that are less than the secrecy capacity represented in the 45 degrees line and

rates that are larger than the secrecy capacity represented by the horizontal line. To extend the proposed multilevel coding to the rate-equivocation region, we start from the secrecy capacity point and move in both directions.

First, assume that the maximum secrecy rate achieved by the proposed multilevel coding is C_{MLC} which is, as explained earlier, is within 1 bit from the secrecy capacity. Clearly, any rate below C_{MLC} can be achieved by reducing the rate that is transmitted from the set S, the levels that are responsible for transmitting the message, while keeping the rate that is transmitted from $\bar{S}$, the levels that are responsible for transmitting the randomness, the same.

Second, in order to show that the proposed multilevel coding can achieve rates that are higher than C_{MLC}, consider for simplicity that we have two levels while the generalization is straightforward. Assume that at the achievable secrecy rate, level 1 is assigned to encode the message with rate $R_1 = C_{MLC}$ while level 2 is assigned to encode the randomness with rate $I(X;Y) - I(X;Z)$. Under joint decoding, R_1 is upper bounded by

$$R_1 \leq I(B_1; Y|B_2) \tag{5.96}$$

therefore, R_1 can increase up to the bound $I(B_1;Y|B_2)$ while decreases until it becomes equal to $I(B_2;Y)$. In order to increase the rate beyond this point, level 2 also should be used to send message bits as well. However, if level 2 is also used to send the message, $H(D) = 0$ which means that an infinitesimally small increase in the message rate beyond $I(B_1;Y|B_2)$ will lead to a loss in the equivocation. This loss will linearly be reduced as the information rate increases till $I(X;Y)$.

The number of alternations that we see in Fig. 5.7 is equal to the number of levels in $\bar{S}$. Each alternation spans the rate of each level in $\bar{S}$. As the number of levels increases, the capacity of each level goes down which means that every alternation will span a very small range. Therefore, as the number of levels increases, the entire rate-equivocation region is achieved.

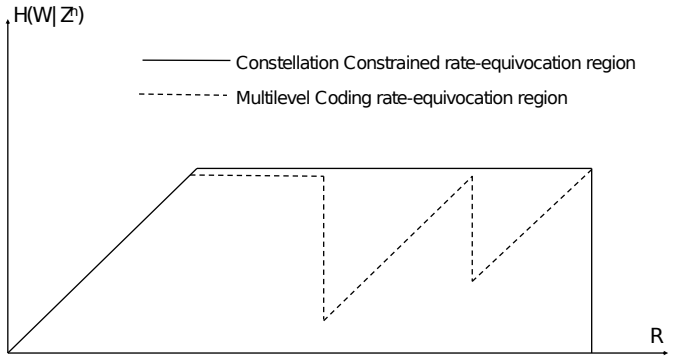

Figure 5.7. The constellation constrained rate-equivocation region versus the multilevel coding rate-equivocation region.

Under multi-stage decoding at the legitimate receiver, increasing the rate beyond C_{MLC} requires adding a new level in the set S directly since R_1 can take only one of two values. This makes the alternation that we see in Fig. 5.7 starts from the secrecy rate C_{MLC}.

5.4 Simulations

In the simulations, we use the DVB-S2 LDPC codes with length of 64800. The bit-error rate and the frame-error rate are obtain for different transmission power P under fixed noise variance at the legitimate receiver σ_1^2 and the Eavesdropper σ_2^2. Each level is either driven by a message or by randomness. We show the error-rate curves at both the legitimate receiver and the eavesdropper. However, observing the error-rate curve at the eavesdropper is not the best indication about the quality of the codes. We only show the error-rate curve at the eavesdropper as an available indication while the best indication is simulating the equivocation which is not easy to simulate.

Fig. 5.8 shows the performance of the proposed transmission under 16-QAM. The transmission power is changing while the noise variance at the legitimate receiver and the eavesdropper are $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$ respectively. The first level encodes the message with rate 0.6 while the second level encodes the dither with rate 0.4. The figure shows a separation of

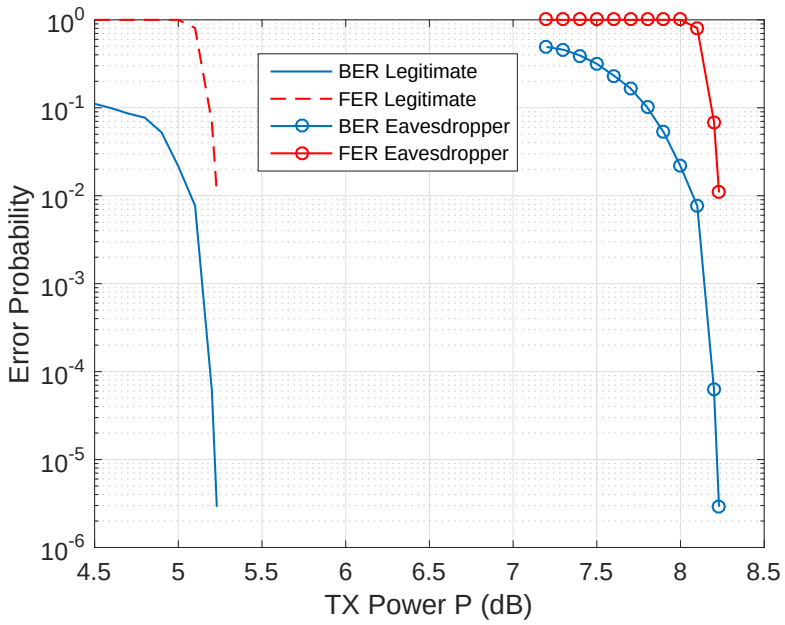

Figure 5.8. Error-rate at the legitimate receiver and the wiretapper under 16-QAM constellation.

3dB between the two frame-error rate curves which means that the eavesdropper will have high probability of error until it has the minimum SNR that the legitimate receiver needs to be able to decode.

5.5 Conclusion

In this chapter, we showed that multilevel coding can achieve the secrecy capacity of the constellation constrained wiretap channel. An explicit construction of multilevel coding where the message and the randomness are encoded through distinct levels is shown to achieve the secrecy capacity under certain conditions. The proposed construction was studied under joint decoding and multistage decoding at the legitimate receiver. The labeling design was shown to play an important role in the design if multistage decoding is to be used at the legitimate receiver. Simulation results show good codes that are designed for the point-

to-point channel can be used in the proposed construction and achieve both reliability and secrecy.

CHAPTER 6

CONCLUSION

This dissertation studied multilevel coded modulation for multi-node networks. Multilevel coding for the AWGN broadcast channel was studied. Necessary and sufficient conditions for optimality of multilevel coding in the AWGN broadcast channel were presented. A pragmatic rate allocation that achieves rate pairs that are very close to the constellation constrained capacity were presented. Surprisingly, it was shown that to achieve any rate pair that is very close to the constellation constrained capacity, mixing of the user information is not necessary except in one level.

Multilevel coding was also studied for the full-duplex relay channel. First, necessary and sufficient conditions for optimality of multilevel coding to achieve the decode-and-forward rate were studied. The effect of linear codes and the labeling design were investigated. The error exponent of the system was studied. Furthermore, multilevel coding for the hybrid decode-compress-forward was proposed. It is shown that the proposed multilevel coding can achieve rates that are very close to the achievable rates under Gaussian input.

Finally, coding for the wiretap channel was studied. Independent encoding of the message and randomness was studied for the discrete memory-less as well as the AWGN wiretap channel showing that it can still achieve the rate-equivocation region. A multilevel coding for the wiretap channel where the message and the randomness are encoded through distinct levels is proposed. The proposed multilevel coding was shown to achieve the secrecy capacity of the constellation constrained AWGN wiretap channel.

REFERENCES

[1] A. E. Gamal and Y. Kim, *Network Information Theory*. Cambridge University Press, 2012.

[2] P. Berlin and D. Tuninetti, "LDPC codes for gaussian broadcast channels," in *Signal Processing Advances in Wireless Communications*, 2004, pp. 444–448.

[3] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," *IEEE Trans. Inform. Theory*, vol. 61, no. 2, pp. 758–782, Feb. 2015.

[4] A. Chakrabarti, A. de Baynast, A. Sabharwal, and B. Aazhang, "Low density parity check codes for the relay channel," *IEEE J. Select. Areas Commun.*, vol. 25, no. 2, pp. 280–291, Feb. 2007.

[5] T. V. Nguyen, A. Nosratinia, and D. Divsalar, "Bilayer protograph codes for half-duplex relay channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1969–1977, May 2013.

[6] P. Razaghi and W. Yu, "Bilayer low-density parity-check codes for decode-and-forward in relay channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3723–3739, Oct. 2007.

[7] A. Thangaraj, S. Dihidar, A. R. Calderbank, S. W. McLaughlin, and J. M. Merolla, "Applications of LDPC codes to the wiretap channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 8, pp. 2933–2945, Aug. 2007.

[8] V. Rathi, M. Andersson, R. Thobaben, J. Kliewer, and M. Skoglund, "Performance analysis and design of two edge-type LDPC codes for the BEC wiretap channel," *IEEE Trans. Inform. Theory*, vol. 59, no. 2, pp. 1048–1064, Feb. 2013.

[9] A. Subramanian, A. Thangaraj, M. Bloch, and S. W. McLaughlin, "Strong secrecy on the binary erasure wiretap channel using large-girth LDPC codes," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 585–594, Sep. 2011.

[10] D. Klinc, J. Ha, S. W. McLaughlin, J. Barros, and B. J. Kwak, "LDPC codes for the Gaussian wiretap channel," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 532–540, Sep. 2011.

[11] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. 28, no. 1, pp. 55–67, 1982.

[12] G. Ungerboeck and I. Csajka, "On improving data-link performance by increasing the channel alphabet and introducing sequence coding," in *Proc. of IEEE International Symposium on Information Theory ISIT*, 1976.

[13] G. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.

[14] E. Zehavi, "8-PSK trellis codes for a rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, 1992.

[15] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[16] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inform. Theory*, vol. 23, no. 3, pp. 371–377, 1977.

[17] A. Ingber and M. Feder, "On the optimality of multilevel coding and multistage decoding," in *IEEE 25th Convention of Electrical and Electronics Engineers in Israel.*, Dec. 2008, pp. 731–735.

[18] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.

[19] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[20] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 14–20, 1972.

[21] U. Wachsmann, R. F. H. Fischer, and J. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1361–1391, 1999.

[22] J. Huber and U. Wachsmann, "Capacities of equivalent channels in multilevel coding schemes," *Electronics Letters*, vol. 30, no. 7, pp. 557–558, 1994.

[23] ——, "Design of multilevel codes," in *Proc. of IEEE Information Theory Workshop (ITW), Rydzyna, Poland*, 1995.

[24] L. Duan, B. Rimoldi, and R. Urbanke, "Approaching the AWGN channel capacity without active shaping," in *Proc. of IEEE International Symposium on Information Theory ISIT97*, Jun. 1997, pp. 374–.

[25] L. H.-J. Lampe, R. F. H. Fischer, and R. Schober, "Multilevel coding for multiple-antenna transmission," in *Proc. of IEEE International Symposium on Information Theory*, Jun. 2002.

[26] R. Yeung, "Multilevel diversity coding with distortion," *IEEE Trans. Inform. Theory*, vol. 41, no. 2, pp. 412–422, Mar. 1995.

[27] J. Roche, R. Yeung, and K. P. Hau, "Symmetrical multilevel diversity coding," *IEEE Trans. Inform. Theory*, vol. 43, no. 3, pp. 1059–1064, May 1997.

[28] S. Mohajer, C. Tian, and S. Diggavi, "Asymmetric Multilevel Diversity Coding and Asymmetric Gaussian Multiple Descriptions," *IEEE Trans. Inform. Theory*, vol. 56, no. 9, pp. 4367–4387, Sep. 2010.

[29] J. Jiang, N. Marukala, and T. Liu, "Symmetrical multilevel diversity coding and subset entropy inequalities," *IEEE Trans. Inform. Theory*, vol. 60, no. 1, pp. 84–103, Jan 2014.

[30] K. Abdel-Ghaffar and M. Hassner, "Multilevel error-control codes for data storage channels," *IEEE Trans. Inform. Theory*, vol. 37, no. 3, pp. 735–741, May 1991.

[31] B. Hern and K. Narayanan, "Multilevel coding schemes for compute-and-forward with flexible decoding," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7613–7631, Nov. 2013.

[32] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.

[33] E. McCune, *Dynamic Power Supply Transmitters*. Academic Press, 2015.

[34] F. Taubin, "Performance of BICM transmission over Gaussian broadcast channels," in *Proc. of IEEE International Conference on Telecommunications (ICT)*, Jun. 2008, pp. 1–4.

[35] T. Sun, R. Wesel, M. Shane, and K. Jarett, "Superposition turbo TCM for multi-rate broadcast," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 368–371, Mar. 2004.

[36] S. Shamai and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.

[37] K. Ramchandran, A. Ortega, K. Uz, and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source/channel coding," *IEEE J. Select. Areas Commun.*, vol. 11, no. 1, pp. 6–23, Jan. 1993.

[38] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, "Superposition coding for side-information channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1872–1889, May 2006.

[39] Z. Mheich, F. Alberge, and P. Duhamel, "Achievable rates optimization for broadcast channels using finite size constellations under transmission constraints," *Journal on Wireless Communications and Networking EURASIP*, no. 1, p. 254, 2013.

[40] A. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1234–1248, Jul. 1993.

[41] L.-F. Wei, "Coded modulation with unequal error protection," *IEEE Trans. Commun.*, vol. 41, no. 10, pp. 1439–1449, Oct. 1993.

[42] R. Morelos-Zaragoza, O. Takeshita, H. Imai, M. Fossorier, and S. Lin, "Coded modulation for satellite broadcasting," in *Proc. of IEEE Global Telecommunication Conference (GLOBECOM)*, Nov. 1996, pp. 31–35.

[43] R. Morelos-Zaragoza and S. Lin, "QPSK block-modulation codes for unequal error protection," *IEEE Trans. Inform. Theory*, vol. 41, no. 2, pp. 576–581, Mar. 1995.

[44] A. Khandani, "Two-way (true full-duplex) wireless," in *13th Canadian Workshop on Information Theory (CWIT)*, Jun. 2013, pp. 33–38.

[45] A. Sabharwal, P. Schniter, D. Guo, D. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Select. Areas Commun.*, vol. 32, no. 9, Sep. 2014.

[46] E. Ahmed, A. Eltawil, and A. Sabharwal, "Rate gain region and design tradeoffs for full-duplex wireless communications," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3556–3565, Jul. 2013.

[47] T. Riihonen, S. Werner, and R. Wichman, "Hybrid Full-Duplex/Half-Duplex relaying with transmit power adaptation," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 3074–3085, Sep. 2011.

[48] O. Agazzi, D. Messerschmitt, and D. Hodges, "Nonlinear echo cancellation of data signals," *IEEE Trans. Commun.*, vol. 30, no. 11, pp. 2421–2433, Nov. 1982.

[49] W. Afifi and M. Krunz, "Incorporating self-interference suppression for full-duplex operation in opportunistic spectrum access systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2180–2191, Apr. 2015.

[50] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Processing*, vol. 59, no. 12, pp. 5983–5993, Dec. 2011.

[51] M. Duarte, "Full-duplex wireless: Design, implementation and characterization," Ph.D. dissertation, Rice University, 2012.

[52] S. Hong, J. Brand, J. I. Choi, M. Jain, J. Mehlman, S. Katti, and P. Levis, "Applications of self-interference cancellation in 5G and beyond," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 114–121, 2014.

[53] D. Korpi, T. Riihonen, V. Syrjälä, L. Anttila, M. Valkama, and R. Wichman, "Full-duplex transceiver system calculations: Analysis of ADC and linearity challenges," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3821–3836, 2014.

[54] A. Balatsoukas-Stimming, A. C. Austin, P. Belanovic, and A. Burg, "Baseband and RF hardware impairments in full-duplex wireless systems: experimental characterisation and suppression," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 1, 2015.

[55] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[56] N. Ferdinand, M. Nokleby, and B. Aazhang, "Low-density lattice codes for full-duplex relay channels," *IEEE Trans. Wireless Commun.*, vol. PP, no. 99, pp. 1–1, 2014.

[57] K. Ravindran, A. Thangaraj, and S. Bhashyam, "LDPC codes for network-coded bidirectional relaying with higher order modulation," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 1975–1987, Jun. 2015.

[58] Z. Chen and H. Liu, "Spectrum-efficient coded modulation design for two-way relay channels," *IEEE J. Select. Areas Commun.*, vol. 32, no. 2, pp. 251–263, Feb. 2014.

[59] Z. Chen, B. Xia, Z. Hu, and H. Liu, "Design and analysis of multi-level physical-layer network coding for gaussian two-way relay channels," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 1803–1817, Jun. 2014.

[60] A. A. Abotabl and A. Nosratinia, "Broadcast coded modulation: Multilevel and bit-interleaved construction," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 969–980, Mar. 2017.

[61] A. Abotabl and A. Nosratinia, "Multilevel coding for the full-duplex relay channel," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2015, pp. 1–6.

[62] ——, "Multi-level coding and multi-stage decoding in MAC, broadcast, and relay channel," in *Proc. of IEEE International Symposium on Information Theory*, Jun. 2014, pp. 96–100.

[63] M. Smolnikar, T. Javornik, M. Mohorcic, S. Papaharalabos, and P. T. Mathiopoulos, "Rate-compatible punctured DVB-S2 LDPC codes for DVB-SH applications," in *International Workshop on Satellite and Space Communications*, Sep. 2009, pp. 13–17.

[64] N. Shende, O. Gurbuz, and E. Erkip, "Half-duplex or full-duplex relaying: A capacity analysis under self-interference," in *47th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2013, pp. 1–6.

[65] E. Ahmed and A. M. Eltawil, "All-digital self-interference cancellation technique for full-duplex systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3519–3532, Jul. 2015.

[66] K. Alexandris, A. Balatsoukas-Stimming, and A. Burg, "Measurement-based characterization of residual self-interference on a full-duplex MIMO testbed," in *IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jun. 2014, pp. 329–332.

[67] N. H. Mahmood, I. S. Ansari, G. Berardinelli, P. Mogensen, and K. A. Qaraqe, "Analysing self interference cancellation in full duplex radios," in *IEEE Wireless Communications and Networking Conference*, Apr. 2016, pp. 1–6.

[68] A. Ingber and M. Feder, "Capacity and error exponent analysis of multilevel coding with multistage decoding," in *IEEE International Symposium on Information Theory*, Jun. 2009, pp. 1799–1803.

[69] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, no. 1, pp. 3–18, 1965.

[70] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[71] Q. Li and C. Georghiades, "On the error exponent of the wideband relay channel," in *Signal Processing Conference, 2006 14th European*, Sep. 2006, pp. 1–5.

[72] G. Bradford and J. Laneman, "Error exponents for block markov superposition encoding with varying decoding latency," in *IEEE Information Theory Workshop ITW*, Sep. 2012, pp. 237–241.

[73] V. Tan, "On the reliability function of the discrete memoryless relay channel," *IEEE Trans. Inform. Theory*, vol. 61, no. 4, pp. 1550–1573, Apr. 2015.

[74] A. Ingber and M. Feder, "Finite blocklength coding for channels with side information at the receiver," in *IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, Nov. 2010, pp. 000 798–000 802.

[75] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[76] A. Chakrabarti, A. Sabharwal, and B. Aazhang, "Sensitivity of achievable rates for half-duplex relay channel," in *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications*, Jun. 2005, pp. 970–974.

[77] E. C. Van Der Meulen, "Three-terminal communication channels," *Advances in applied Probability*, pp. 120–154, 1971.

[78] G. Kramer, I. Marić, and R. D. Yates, "Cooperative communications," *Foundations and Trends® in Networking*, vol. 1, no. 3–4, pp. 271–425, 2007.

[79] J. N. Laneman, "Cooperative diversity in wireless networks: Algorithms and architectures," Ph.D. dissertation, Massachusetts Institute of Technology, 2002.

[80] S. H. Lee and S. Y. Chung, "When is compress-and-forward optimal?" *Information Theory and Applications Workshop (ITA)*, pp. 1–3, Jan. 2010.

[81] X. Bao and J. Li, "Efficient message relaying for wireless user cooperation: Decode-amplify-forward (DAF) and hybrid DAF and coded-cooperation," *IEEE Trans. Wireless Commun.*, vol. 6, no. 11, pp. 3975–3984, Nov. 2007.

[82] J. Haghighat and W. Hamouda, "Decode-compress-and-forward with selective-cooperation for relay networks," *IEEE Communications Letters*, vol. 16, no. 3, pp. 378–381, Mar. 2012.

[83] H. F. Chong, M. Motani, and H. K. Garg, "Generalized backward decoding strategies for the relay channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 394–401, Jan. 2007.

[84] A. Morello and V. Mignone, "DVB-S2: The second generation standard for satellite broad-band services," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 210–227, 2006.

[85] R. Blasco-Serrano, "Coding strategies for compress-and-forward relaying," Ph.D. dissertation, KTH, 2010.

[86] R. Blasco-Serrano, R. Thobaben, V. Rathi, and M. Skoglund, "Polar codes for compress-and-forward in binary relay channels," in *the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, Nov. 2010, pp. 1743–1747.

[87] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.

[88] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *IEEE International Symposium on Information Theory. Proceedings.* IEEE, 2002, p. 136.

[89] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, Oct. 1949.

[90] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, no. 8, Oct. 1975.

[91] C. Measson, A. Montanari, and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5277–5307, Dec. 2008.

[92] D. Kline, J. Ha, S. W. McLaughlin, J. Barros, and B. J. Kwak, "LDPC codes for the Gaussian wiretap channel," in *IEEE Information Theory Workshop*, Oct. 2009, pp. 95–99.

[93] M. Baldi, M. Bianchi, and F. Chiaraluce, "Non-systematic codes for physical layer security," in *IEEE Information Theory Workshop*, Aug. 2010, pp. 1–5.

[94] ——, "Coding with scrambling, concatenation, and HARQ for the AWGN wire-tap channel: A security gap analysis," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 883–894, Jun. 2012.

[95] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[96] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.

[97] E. Hof and S. Shamai, "Secrecy-achieving polar-coding," in *IEEE Information Theory Workshop*, Aug. 2010, pp. 1–5.

[98] F. Oggier, P. Sol, and J. C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," *IEEE Trans. Inform. Theory*, vol. 62, no. 10, pp. 5690–5708, Oct. 2016.

[99] C. Ling, L. Luzzi, J. C. Belfiore, and D. Stehl, "Semantically secure lattice codes for the gaussian wiretap channel," *IEEE Trans. Inform. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.

[100] X. He and A. Yener, "Providing secrecy with structured codes: Two-user gaussian channels," *IEEE Trans. Inform. Theory*, vol. 60, no. 4, pp. 2121–2138, Apr. 2014.

[101] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[102] Y. Liang, H. V. Poor, S. Shamai *et al.*, "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4–5, pp. 355–580, 2009.

[103] E. Tekin and A. Yener, "The Gaussian multiple access wire-tap channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5747–5755, Dec. 2008.

[104] ——, "The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2735–2751, 2008.

[105] M. H. Yassaee and M. R. Aref, "Multiple access wiretap channels with strong secrecy," in *IEEE Information Theory Workshop (ITW)*, Aug. 2010, pp. 1–5.

[106] Z. Mheich, F. Alberge, and P. Duhamel, "Achievable secrecy rates for the broadcast channel with confidential message and finite constellation inputs," *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 195–205, Jan. 2015.

BIOGRAPHICAL SKETCH

Ahmed Abotabl received his B.Sc degree in Electrical Engineering from Alexandria University, Alexandria, Egypt in 2010 and his M.Sc. degree in Electrical Engineering from Nile University, Cairo, Egypt in 2012. From October 2010 to August 2012 he was a research assistant in the Wireless Intelligent Networks Center, Nile University where he worked on Vehicular communications problems in collaboration with General Motors USA. From September 2012 to August 2017 has has been working as a research assistant in the Multimedia Communications Laboratory, The University of Texas at Dallas where he was pursuing his Ph.D. degree under the supervision of Prof. Aria Nosratinia. His research interests are information theory, Coding theory, Estimation and Detection and its applications in machine learning and security.

CURRICULUM VITAE

Ahmed Attia Abotabl

August 14, 2017

Educational History:

B.S., Electrical Engineering, Alexandria University, 2010
M.S., Electrical Engineering, Nile University, 2012
Ph.D., Electrical Engineering, The University of Texas at Dallas, 2017

Employment History:

Research assistant, The University of Texas at Dallas, September 2012 – August 2017
Research assistant, Nile University, October 2010 – August 2012
Optical communications engineer, Egypt Telecom, June 2009 – August 2009

Professional Recognitions and Honors:

The Industrial Advisory Board Graduate Fellowship award, UTD, 2016
Louis-Beecherl, Jr. Graduate Fellowship award, UTD, 2015
Johnson Distinguished scholarship from the Erik Jonsson school of Engineering UTD, 2012

Professional Memberships:

Institute of Electrical and Electronics Engineers (IEEE), 2015–present