

Article

Capacity Bounds and Mapping Design for Binary Symmetric Relay Channels

Majid Nasiri Khormuji * and Mikael Skoglund *

School of Electrical Engineering and ACCESS Linnaeus Center, Royal Institute of Technology (KTH), Stockholm, 100 44, Sweden

* Authors to whom correspondence should be addressed; E-Mails: khormuji@ee.kth.se (M.N.K.); skoglund@ee.kth.se (M.S.).

Received: 1 September 2012; in revised form: 4 December 2012 / Accepted: 6 December 2012 / Published: 17 December 2012

Abstract: Capacity bounds for a three-node binary symmetric relay channel with orthogonal components at the destination are studied. The cut-set upper bound and the rates achievable using decode-and-forward (DF), partial DF and compress-and-forward (CF) relaying are first evaluated. Then relaying strategies with finite memory-length are considered. An efficient algorithm for optimizing the relay functions is presented. The Boolean Fourier transform is then employed to unveil the structure of the *optimized* mappings. Interestingly, the optimized relay functions exhibit a simple structure. Numerical results illustrate that the rates achieved using the optimized low-dimensional functions are either comparable to those achieved by CF or superior to those achieved by DF relaying. In particular, the optimized low-dimensional relaying scheme can improve on DF relaying when the quality of the source-relay link is worse than or comparable to that of other links.

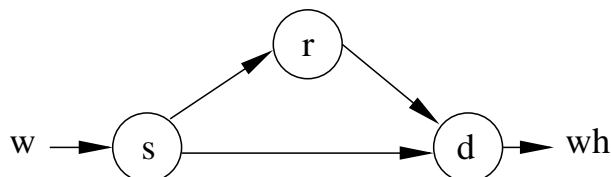
Keywords: binary symmetric relay channel; decode-and-forward; compress-and-forward; linear relaying; capacity bounds; binary Fourier transform

1. Introduction

In this paper we consider a relay network consisting of a source, a relay, and a destination, as illustrated in Figure 1. This channel is considered in [1,2]. The communication task is to reproduce the transmitted message M , uniformly chosen from the set $\mathcal{M} = \{1, 2, \dots, 2^{nR}\}$, at the destination such that the probability of error is arbitrarily small. The transmission of a message consumes n channel

uses. We desire to quantify the supremum of the set of rates, R , for which the average message error probability at the destination can be made to approach zero as the number of channel uses n goes to infinity. This number is the capacity, C , in the communication between the source and destination. The channel capacity for the general relay channel is still an open problem.

Figure 1. Three-node relay network.



One special case of the general relay channel is the Gaussian relay channel, which is studied, e.g., in [2–9]. In this paper, however, we focus on a relay network in which each link is a binary symmetric channel. Some instances of the binary symmetric relay channel (BSRC) are considered in [10–14]. The work of [10,11] considers BSRCs with correlated noises and [12,13] focuses on the detect-and-forward protocol for multi-hop relaying and the relay channel, respectively. In this paper, we consider a model that is also studied in an independent work [14]. We study decode-and-forward (DF), partial decode-and-forward (PDF), compress-and-forward (CF) and general finite-memory relay mappings. One of our main contributions is to propose a general structure for *finite dimensional* relaying. We show that one can improve on the result presented in [14] and illustrate that it is possible to approach the capacity upper bound for some cases by using the proposed low-dimensional mappings. Interestingly we recover the lower bound in [14] by a simple one-dimensional mapping at the relay. We also illustrate that one can obtain higher reliable rates by employing optimized relaying functions with larger memories. The rate obtained via optimized finite-length mappings can be superior to those achieved by DF protocol in some cases and are comparable to those achieved by CF protocol in general.

The remaining part of the paper is organized as follows. In Section 2, we present the three-node BSRC with orthogonal components at the destination. Section 3 derives capacity bounds for the BSRC when the relay is assumed to have infinite memory. Section 4 investigates capacity bounds for the BSRC when the relay is assumed to have a finite memory. We present an algorithm to optimize the relay mapping in Section 4. In Section 5 we compare the achievable rates as a function of memory length. Finally, Section 6 concludes the paper.

Notation: We use the following notation for brevity.

- $\mathbb{GF}_b$ denotes the binary Galois field, *i.e.*, $\{0, 1\}$.
- $\mathbb{GF}_b^k$ denotes the k -dimensional binary Galois field, *i.e.*, $\{0, 1\}^k$.
- $\mathbb{H}_b(\epsilon) := -\epsilon \log(\epsilon) - (1 - \epsilon) \log(1 - \epsilon)$ denotes binary entropy function where $\epsilon \in [0, 1]$.
- $\mathbb{D}_H(x_1^k, y_1^k)$ denotes the Hamming distance between the two binary sequences of length k .
- The operation $*$ is defined as $\epsilon_1 * \epsilon_2 := \epsilon_1 + \epsilon_2 - 2\epsilon_1\epsilon_2$.
- We denote that the binary random variable Z has a Bernoulli distribution by $Z \sim \text{Ber}(\epsilon)$ where $\Pr(Z = 1) = \epsilon$ and $\Pr(Z = 0) = 1 - \epsilon$.
- We denote the binary Kronecker delta function by $\delta_b(x)$, where $\delta_b(0) = 1$ and $\delta_b(1) = 0$ for $x \in \mathbb{GF}_b$.

2. Binary Symmetric Relay Channel

In this section, we introduce the channel model that we consider in the paper. Figure 2 shows a BSRC consisting of three nodes: a source, a relay and a destination. In this model, we however assume that received signals at the destination are *orthogonal*. That is, the signals transmitted from the source and the relay do not interfere with each other. We further assume that all links are corrupted with modulo-sum noises distributed according to the Bernoulli distribution and all quantities are binary; *i.e.*, $X, X_r, Y_1, Y_2, Y_r \in \{0, 1\}$.

The received signal at the relay Y_r is given by

$$Y_r = X \oplus Z_r \quad (1)$$

where X is the transmitted symbol from the source and $Z_r \sim \text{Ber}(\epsilon_r)$ is the additive Bernoulli noise. (The Bernoulli noise models for example the conventional communication set-up using BPSK modulation followed by a hard decision (see also [14]).) The received signal from the source at the destination Y_1 is given by

$$Y_1 = X \oplus Z_1 \quad (2)$$

where $Z_1 \sim \text{Ber}(\epsilon_1)$ is the additive Bernoulli noise. Similarly, the received signal from the relay at the destination Y_2 is given by

$$Y_2 = X_r \oplus Z_2 \quad (3)$$

where X_r is the transmitted symbol from the relay and $Z_2 \sim \text{Ber}(\epsilon_2)$ is the additive Bernoulli noise. We assume that the random variables Z_r, Z_1 , and Z_2 are mutually independent. Note that the addition in Equations (1)–(3) is done in $\mathbb{GF}(2)$.

Remark 1. Figure 3 shows a BSRC with non-orthogonal reception at the destination. The received signal at the destination is given by

$$Y = X \oplus X_r \oplus Z \quad (4)$$

where X and X_r respectively denote the symbols transmitted by the source and the relay. The random variable $Z \sim \text{Ber}(\epsilon)$ is the additive Bernoulli noise and is independent of Z_r . The capacity of this channel is

$$C_{no} = 1 - \mathbb{H}_b(\epsilon) \quad (5)$$

By setting $X_r = 0$, we have $Y = X \oplus Z$. Then the achievability follows since $\max_{p(x)} I(X; Y | X_r = 0) = C_{no}$. The converse follows from the cut-set bound. Invoking the multiple access bound, we have

$$\begin{aligned} C &\leq I(X, X_r; Y) \\ &= H(Y) - H(Y | X, X_r) \\ &= H(Y) - H(Z) \\ &\leq 1 - \mathbb{H}_b(\epsilon) \end{aligned} \quad (6)$$

In the sequel, we present various capacity bounds for the orthogonal BSRC.

Figure 2. Orthogonal Binary Symmetric Relay Channel.

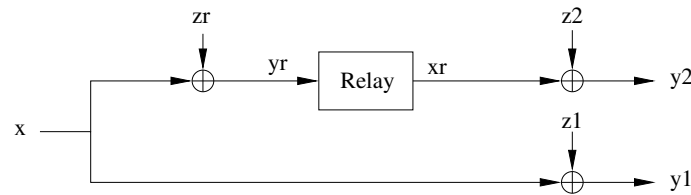
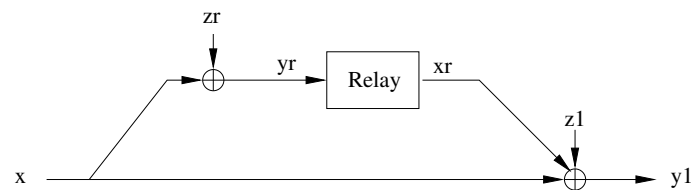


Figure 3. Non-Orthogonal Binary Symmetric Relay Channel.



3. Capacity Bounds for the Orthogonal BSRC: Infinite Memory Relay Case

In this section, we consider the cut-set upper bound and three lower bounds on the capacity: DF, PDF and CF. These three bounds are evaluated based on the results in [2] for the general relay channel where the relay has infinite memory and unlimited processing capability.

Proposition 1 (cutset bound). *For the relay channel in Figure 2, the capacity is upper bounded by*

$$R_{UB} = \min \{1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_r), 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2)\}. \quad (7)$$

Proof. See Appendix A and also [14]. □

Proposition 2 (DF lower bound). *For the relay channel in Figure 2, the capacity is lower bounded by*

$$R_{DF} = \max \{1 - \mathbb{H}_b(\epsilon_1), \min \{1 - \mathbb{H}_b(\epsilon_r), 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2)\}\} \quad (8)$$

where the bound is achieved using the block Markov encoding scheme in [2] (known as DF relaying) and the only-direct-transmission (i.e., when the relay is off).

Proof. See Appendix B. □

Note that in Equation (8), we take the maximum of the rates achieved using the conventional DF [2] and the only-direct-link transmission. In some cases it is possible to improve on DF by using PDF at the relay. That is, the relay only decodes a part of the transmitted message. The achievable rate of PDF is given by

$$R_{PDF} = \max_{p(u,x,x_r)} \min \{I(X, X_r; Y_1, Y_2), I(U; Y_r | X_r) + I(X; Y_1, Y_2 | X_r, U)\} \quad (9)$$

where U denotes the part of the transmitted message that the relay decodes. (See Theorem 7 in [2] and also [15].)

Proposition 3 (PDF lower bound). *For the relay channel in Figure 2, generalized block Markov encoding attains the same rate as the modified DF given in Equation (8).*

Proof. See Appendix C. □

Corollary 1. *DF is an optimal relaying strategy if $\mathbb{H}_b(\epsilon_1) + \mathbb{H}_b(\epsilon_2) - \mathbb{H}_b(\epsilon_r) \geq 1$.*

Proof. The proof follows from Propositions 2 and 1. □

Proposition 4 (CF lower bound). *For the relay channel in Figure 2, the capacity is lower bounded by*

$$R_{CF} = \begin{cases} 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_1), & \text{if } 1 < \mathbb{H}_b(\epsilon_2) + \mathbb{H}_b(\epsilon_1 * \epsilon_r) \\ 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r) - \mathbb{H}_b(\epsilon_r) - \mathbb{H}_b(\epsilon_1), & \text{if } 1 \geq \mathbb{H}_b(\epsilon_2) + \mathbb{H}_b(\epsilon_1 * \epsilon_r) \end{cases} \quad (10)$$

where ϵ_q satisfies

$$\mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_q) + \mathbb{H}_b(\epsilon_2) = 1 \quad (11)$$

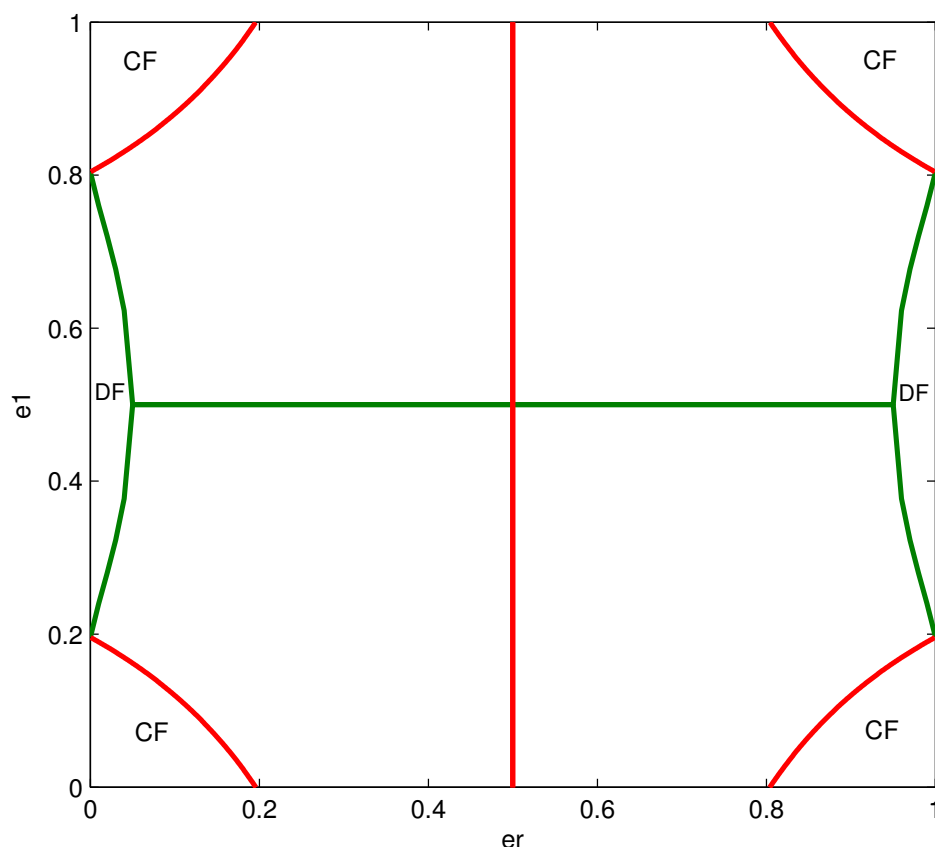
Proof. See Appendix D. □

The bound in Equation (10) is achieved using the side-information encoding scheme in [2], known as CF relaying.

Corollary 2. *CF is an optimal relaying strategy if $\mathbb{H}_b(\epsilon_1 * \epsilon_r) + \mathbb{H}_b(\epsilon_2) \leq 1$.*

Proof. The proof follows from Propositions 4 and 1. □

Figure 4. Optimal regions for $\epsilon_2 = 0.05$.

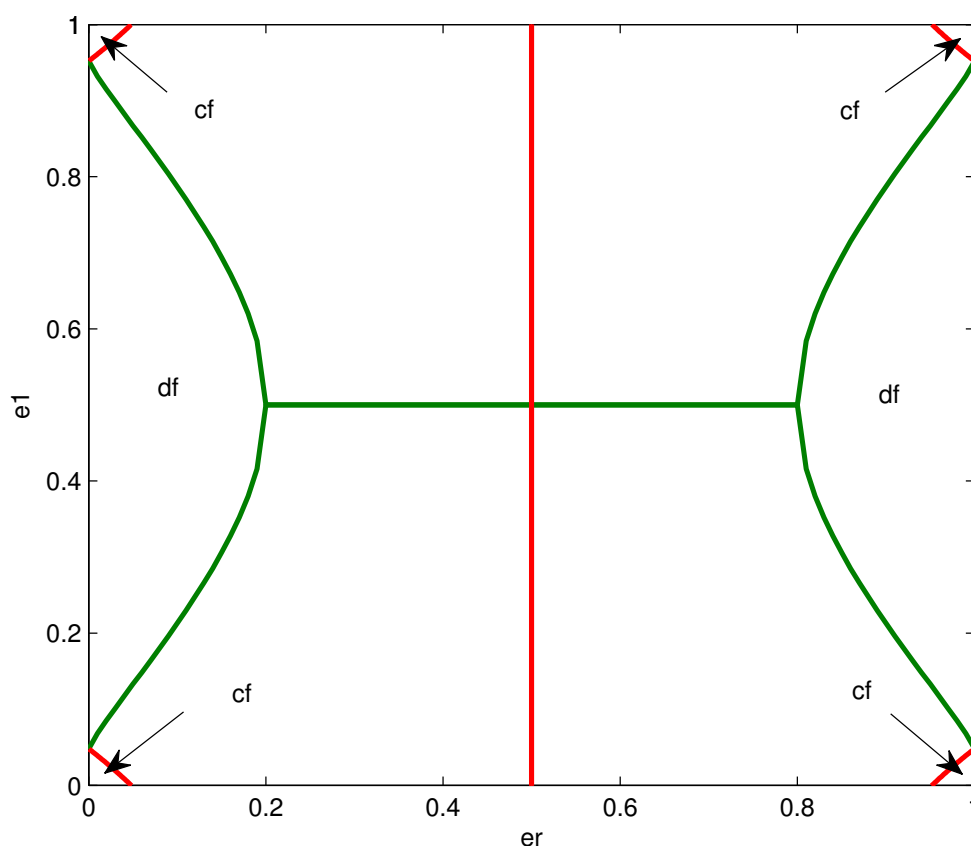


Figures 4 and 5 show optimal regions in the plane $[0, 1] \times [0, 1]$ for ϵ_2 set to 0.05 and 0.2, respectively. One can see that for some parameters of ϵ_r and ϵ_1 , CF or DF is optimal.

From Figures 4 and 5, we see that DF is optimal when ϵ_r is small and ϵ_1 is close to 0.5. When $\epsilon_1 = 0.5$, the BSRC simplifies to a two-hop channel for which DF is capacity achieving. However, DF is also optimal when the relay is relatively strong compared with the direct link.

From Figures 4 and 5, we similarly see that CF is optimal at the corner of the plane, *i.e.*, when ϵ_1 and ϵ_r are small. As ϵ_2 increases, the area of the region for which CF is optimal shrinks. This is due to the fact that small value of ϵ_2 allows the relay to use higher rate to describe the received signal to destination. Small value of ϵ_1 means the destination has better side-information in order to decode the reproduction sequence generated at the relay, and small value of ϵ_r means that the relay receives a less-noisy sequence on average that makes compression of the “true” signal easier.

Figure 5. Optimal regions for $\epsilon_2 = 0.2$.



4. Capacity Bounds for the Orthogonal BSRC: Finite Memory Relay Case

We next consider the case when the relay has a finite memory length. For the Gaussian relay channel, optimized *memoryless* relaying is investigated in [16–19]. Here we consider higher-dimension mappings for the BSRC.

If the relay has a storage memory of k bits, it can process the last $k - 1$ and the presently received symbol to generate k new symbols using k possibly different k -dimensional functions. This results

in a low-complexity relaying protocol suitable for delay-sensitive or inexpensive applications. In the following, we denote the relay functions by

$$\begin{aligned} f_i &: \mathbb{GF}_b^k \mapsto \mathbb{GF}_b \\ x_{r,i} &= f_i(y_{r,1}, \dots, y_{r,k}) \end{aligned} \quad (12)$$

for $i \in \{1, 2, \dots, k\}$. Note that we here consider the whole sub-block of k symbols to generate k new symbols to be transmitted to the destination. This is different from the classical definition with strictly causal relaying. We are allowed to do this, without any particular condition in signal reception at the relay, since the relay has an orthogonal link to the destination.

4.1. Achievable Rate

For a given set of relay functions $\{f_i\}_{i=1}^k$, the channel is parameterized by the pmf $p(\mathbf{y}|\mathbf{x})$, by defining $\mathbf{y} \equiv (y_{1,1}^k, y_{2,1}^k)$ and $\mathbf{x} \equiv x_1^k$, where

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &:= p(y_{1,1}^k, y_{2,1}^k | x_1^k) \\ &= \sum_{y_{r,1}^k \in \mathbb{GF}_b^k} p(y_{1,1}^k, y_{2,1}^k | y_{r,1}^k, x_1^k) p(y_{r,1}^k | x_1^k) \\ &= \sum_{y_{r,1}^k \in \mathbb{GF}_b^k} p(y_{1,1}^k | x_1^k) p(y_{2,1}^k | y_{r,1}^k) p(y_{r,1}^k | x_1^k) \\ &= \sum_{y_{r,1}^k \in \mathbb{GF}_b^k} \prod_{i=1}^k p(y_{1,i} | x_i) p(y_{r,i} | x_i) p(y_{2,i} | f_i(y_{r,1}^k)) \\ &= \sum_{y_{r,1}^k \in \mathbb{GF}_b^k} \prod_{i=1}^k \left[(1 - \epsilon_1) \delta_b(y_{1,i} \oplus x_i) + \epsilon_1 \delta_b(y_{1,i} \oplus x_i \oplus 1) \right] \times \\ &\quad \left[(1 - \epsilon_r) \delta_b(y_{r,i} \oplus x_i) + \epsilon_r \delta_b(y_{r,i} \oplus x_i \oplus 1) \right] \times \\ &\quad \left[(1 - \epsilon_2) \delta_b(y_{2,i} \oplus f_i(y_{r,1}^k)) + \epsilon_2 \delta_b(y_{2,i} \oplus f_i(y_{r,1}^k) \oplus 1) \right] \end{aligned} \quad (13)$$

Now one can apply the standard random coding argument for the equivalent discrete memoryless point-to-point communication link with the input $\mathbf{x}$ and the output $\mathbf{y}$ whose relation is governed by the pmf $p(\mathbf{y}|\mathbf{x})$ as follows. Generate $2^{(nk)C_k}$ i.i.d. codewords where each has length kn and each k subsequent symbols in every codeword is distributed according to $p(x_1^k)$ (i.e., $p(x_1^{kn}) = \prod_{i=1}^n p(x_1^k)$). Thus, the achievable rate using the finite memory relay is given by

$$C_k = \sup_{\{f_i\}_{i=1}^k, p(x_1^k)} \frac{1}{k} I(X_1^k; Y_{1,1}^k, Y_{2,1}^k) \quad (14)$$

where the supremum is taken over the set of Boolean functions $\{f_i\}_{i=1}^k$ and the joint pmf $p(x_1^k)$ of k symbols at the source. Since the channel is used k times, the mutual information in Equation (14) is divided by k (see also [1,4]).

Achievable Rate for $k = 1$: The simplest case is the memoryless relay in which the relay just transmits the received *noisy* bit to the destination without any further processing. That is $x_{r,i} = y_{r,i}$ for $1 \leq i \leq n$. For this relay function, the optimal input distribution is $X \sim \text{Ber}(\frac{1}{2})$.

Proposition 5. For the relay channel in Figure 2, the rate

$$C_1 = 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_2) - \mathbb{H}_b(\epsilon_r * \epsilon_2) - \mathbb{H}_b(\epsilon_1) \quad (15)$$

is achievable.

Proof. See Appendix E. □

We note that the rate C_1 is also derived in [14] via a suboptimal evaluation of the CF lower bound. However, here we arrive at this rate using a one-dimensional mapping without any need for a compression codebook at the relay.

Computation of C_k for $k \geq 2$ is a cumbersome task. Nevertheless, there is no unique set of relay functions that is optimal for all channel parameters $\epsilon_r, \epsilon_1, \epsilon_2$. To see this, consider the case with $\epsilon_2 = 0$. For this case, the simple strategy with $k = 1$ used in Proposition 5 is optimal. However, this relay function is not necessarily optimal for cases with $\epsilon_2 \neq 0$ since one can potentially provide error protection on the relay-destination link by utilizing functions with higher dimensions.

4.2. Mapping Optimization for an Arbitrary k

In the following, we confine the pmf at the source to be $p(x_1^k) = \prod_{i=1}^k p(x)$, and $p(x) = \frac{1}{2}\delta_b(x) + \frac{1}{2}\delta_b(x \oplus 1)$, i.e., $X \sim \text{Ber}(\frac{1}{2})$.

Lemma 1. The mutual information given in Equation (14) can be written as

$$\frac{1}{k} I(X_1^k; Y_{1,1}^k, Y_{2,1}^k) = 1 - \mathbb{H}_b(\epsilon_1) - \frac{1}{k} \mathbb{E} [\log(p(y_{2,1}^k | y_{1,1}^k))] + \frac{1}{k} \mathbb{E} [\log(p(y_{2,1}^k | x_1^k))] \quad (16)$$

where

$$p(y_{2,1}^k | x_1^k) = (1 - \epsilon_2)^k (1 - \epsilon_r)^k \sum_{y_{r,1}^k \in \mathbb{GF}_b^k} \left(\frac{\epsilon_2}{1 - \epsilon_2} \right)^{\mathbb{D}_H(y_{2,1}^k, f(y_{r,1}^k))} \left(\frac{\epsilon_r}{1 - \epsilon_r} \right)^{\mathbb{D}_H(y_{r,1}^k, x_1^k)} \quad (17)$$

and

$$p(y_{2,1}^k | y_{1,1}^k) = (1 - \epsilon_1)^k (1 - \epsilon_2)^k (1 - \epsilon_r)^k \sum_{y_{r,1}^k \in \mathbb{GF}_b^k} \left(\frac{\epsilon_2}{1 - \epsilon_2} \right)^{\mathbb{D}_H(y_{2,1}^k, f(y_{r,1}^k))} \sum_{x_1^k \in \mathbb{GF}_b^k} \left(\frac{\epsilon_r}{1 - \epsilon_r} \right)^{\mathbb{D}_H(y_{r,1}^k, x_1^k)} \left(\frac{\epsilon_1}{1 - \epsilon_1} \right)^{\mathbb{D}_H(y_{1,1}^k, x_1^k)}. \quad (18)$$

Proof. See Appendix F. □

To compute the rate in Equation (14), one needs to select the best functions among 2^{k2^k} possible choices, which has an exponential complexity. (For $k = 4$, there are approximately 1.8×10^{19} possible functions.) In order to cope with the complexity, we implement an efficient hill-climbing search algorithm as follows. For a given k , we first initialize the relay functions with a random mapping and compute the rate using Lemma 1. Then we randomly select one function and one corresponding dimension, and flip the mapping and recompute the rate. If the new mapping provides a higher

rate, we accept the change. Otherwise we repeat the process until the mapping converges. Since the algorithm by construction may terminate in a local optimum, we repeat the whole algorithm with different initializations and pick the mapping that attains the highest rate.

One example of the optimized mapping for $k = 4$ when $\epsilon_1 = \epsilon_2 = \epsilon_r = 0.01$ is

$$F = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

Here F_{ij} denotes the output of the relay along the i th dimension for the j th input configuration (we have $1 \leq i \leq 4$ and $1 \leq j \leq 2^4$). The relay for a given combination of the received bits finds the decimal representation and transmits the bits in the column given by the decimal representation plus one. For example, for the received string 0000, the relay transmits the bits in the first column, *i.e.*, 0111. By studying the matrix F , we can get some insight into the underlying structure of the mapping. Finding an efficient structure at the relay simplifies the design and the implementation of relaying. We employ the Fourier transform to accomplish this task. Our use of the binary Fourier (or Hadamard) transform is related to how it was used in, e.g., [20,21], to analyze the performance of quantizers over noisy channels.

4.3. Fourier Spectrum of the Optimized Mappings

In order to define the Fourier transform, we need an orthonormal basis [22]. Consider the following set of functions

$$\begin{aligned} \mathcal{X}_S(\mathbf{x}) &: \{-1, +1\}^k \mapsto \{-1, +1\} \\ \mathcal{X}_S(\mathbf{x}) &= \prod_{i \in S} x_i \end{aligned} \quad (19)$$

where $S \subseteq \{1, 2, \dots, k\}$. Then, any function $f : \{0, 1\}^k \rightarrow \{0, 1\}$ can uniquely be represented as (We use the one-to-one mapping $0 \iff +1$ and $1 \iff -1$.)

$$f = \sum_{S \subseteq \{1, 2, \dots, k\}} \hat{f}(S) \mathcal{X}_S \quad (20)$$

where $\hat{f}(S)$ is the Fourier coefficient of f and is given by

$$\begin{aligned} \hat{f}(S) &= \langle f, \mathcal{X}_S \rangle = \frac{1}{2^k} \sum_{\mathbf{x} \in \{+1, -1\}^k} f(\mathbf{x}) \mathcal{X}_S(\mathbf{x}) \\ &= \mathbb{E}[f \cdot \mathcal{X}_S] \end{aligned} \quad (21)$$

The expectation in Equation (21) is taken uniformly over $\mathbf{x} \in \{+1, -1\}^k$. Note that the Fourier expansion of f can potentially have up to 2^k terms. As an example consider the following randomly chosen function.

x_3	x_2	x_1	f
+1	+1	+1	+1
+1	+1	-1	+1
+1	-1	+1	-1
+1	-1	-1	-1
-1	+1	+1	-1
-1	+1	-1	+1
-1	-1	+1	+1
-1	-1	-1	+1

This function can be expanded as

$$f(x_1, x_2, x_3) = \frac{1}{4} - \frac{1}{4}x_1 + \frac{1}{4}x_2 - \frac{1}{4}x_3 - \frac{1}{4}x_1x_2 + \frac{1}{4}x_1x_3 + \frac{3}{4}x_2x_3 + \frac{1}{4}x_1x_2x_3 \quad (22)$$

That is, the Fourier spectrum has eight terms and the function is not linear. In general, we would like a sparse Fourier spectrum for efficient implementation. (Sparsity of mappings allows to realize them with much fewer multiplications and additions. This can be also compared to codes with low-density generator/parity matrices that allow simpler encoding and decoding in general.)

Table 1 presents the Fourier expansion of optimized relay functions for $k = 2, 3, 4, 6$ when $\epsilon_1 = \epsilon_2 = \epsilon_r = 0.01$. Interestingly, we see that the Fourier expansion of the optimized functions is indeed sparse. Using the results in Table 1, we can rewrite the functions in the following form

$$f : \{0, 1\}^k \mapsto \{0, 1\}^k$$

$$\mathbf{x}_r = A_k \mathbf{y}_r + b_k \quad (23)$$

where $\mathbf{x}_r = [x_{r,1}, \dots, x_{r,k}]^T$ and $\mathbf{y}_r = [y_{r,1}, \dots, y_{r,k}]^T$, and $A_k \in \{0, 1\}^{[k \times k]}$ and $b_k \in \{0, 1\}^{[k \times 1]}$. For example for $k = 6$, we have

$$A_6 = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}, \quad b_6 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Note that the mapping given in Equation (23) is not linear in the binary field, when $b_k \neq 0$. However, the linear mapping $\mathbf{x}_r = A_k \mathbf{y}_r$ gives the same performance as $\mathbf{x}_r = A_k \mathbf{y}_r + b_k$. In other words the bias term b_k does not improve the rate. This essentially follows from the data processing inequality. Therefore, the underlying relay functions define a linear code of rate one on the noisy received bits at the relay. Additionally, the code used at the relay performs joint source–channel coding, it therefore should be good for both source and channel coding.

Table 1. Fourier expansion of the optimized relay functions for the orthogonal BSRC with $\epsilon_1 = \epsilon_2 = \epsilon_r = 0.01$.

$k = 2$	$f_1(y_{r1}^k) = y_{r1}$
	$f_2(y_{r1}^k) = y_{r2}y_{r1}$
$k = 3$	$f_1(y_{r1}^k) = y_{r3}y_{r2}y_{r1}$
	$f_2(y_{r1}^k) = -y_{r3}y_{r1}$
	$f_3(y_{r1}^k) = y_{r3}y_{r2}$
$k = 4$	$f_1(y_{r1}^k) = -y_{r4}y_{r3}y_{r1}$
	$f_2(y_{r1}^k) = -y_{r4}y_{r2}y_{r1}$
	$f_3(y_{r1}^k) = -y_{r3}y_{r2}y_{r1}$
	$f_4(y_{r1}^k) = y_{r4}y_{r3}y_{r2}$
$k = 6$	$f_1(y_{r1}^k) = y_{r5}y_{r3}y_{r2}$
	$f_2(y_{r1}^k) = -y_{r6}y_{r4}y_{r3}y_{r1}$
	$f_3(y_{r1}^k) = -y_{r5}y_{r4}y_{r2}$
	$f_4(y_{r1}^k) = -y_{r4}y_{r3}y_{r2}y_{r1}$
	$f_5(y_{r1}^k) = y_{r6}y_{r5}y_{r1}$
	$f_6(y_{r1}^k) = -y_{r6}y_{r4}y_{r3}y_{r2}$

4.4. Effect of Channel Parameters on the Structure of the Optimized Mappings

In this section, we investigate the structure of the optimized mappings for different channel parameters. Our numerical search indicates that the linear mapping $\mathbf{x}_r = A_k \mathbf{y}_r$ is an efficient strategy among all classes of mappings for low-dimensional relaying. That is, the relay employs the binary matrix A_k to generate the relay outputs using k received bits.

Table 2 shows the optimized generator matrices A_k for various values of ϵ_r when $k = 6$, $\epsilon_1 = \epsilon_2 = 0.05$. In particular, for $\epsilon_r = 0.25$ the optimized generator matrix is the identity matrix, *i.e.*, $A_6 = I_6$. For this case, the relay is better off transmitting the received noisy bits without any further processing. However, as ϵ_r decreases, the relay starts combining the received bits at the relay before transmitting. The number of ones in a row of the generator matrix indicates the number of inputs that the relay combines. The density of ones in the generator matrices $\rho := \frac{\# \text{ of ones}}{k^2}$ are also shown in Table 2. We see that as ϵ_r decreases, ρ increases. That is, the relay starts to transmit combinations of more bits in one single channel use. This occurs because of two main reasons: firstly when ϵ_r decreases the relay receives *less* noisy bits on average and secondly the destination has some partial knowledge of individual bits via the received signal from the source.

Table 3 shows the optimized generator matrices for various values of ϵ_1 when $k = 6$, $\epsilon_r = 0.01$ and $\epsilon_2 = 0.1$. We similarly see that as ϵ_1 decreases, ρ increases. This is due to the fact that when ϵ_1 decreases, the destination receives better descriptions of the transmitted bits via the source-destination link. The relay then forwards combinations of several incoming bits when the destination has access to more reliable side information.

Table 2. Optimized generator matrices as a function of ϵ_r for $k = 6$ and $\epsilon_1 = \epsilon_2 = 0.05$.

ϵ_r	0.25	0.1	0.05	0.01	0.001	0.0001
A_6	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$
ρ	0.1667	0.3611	0.4444	0.5556	0.6111	0.6389

Table 3. Optimized generator matrices as a function of ϵ_1 for $k = 6$ and $\epsilon_r = 0.01$, $\epsilon_2 = 0.1$.

ϵ_1	0.4	0.25	0.1	0.05	0.01	0.001
A_6	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$
ρ	0.1667	0.3333	0.5556	0.6389	0.6667	0.7500

Table 4. Optimized generator matrices as a function of ϵ_2 for $k = 6$ and $\epsilon_r = 0.01$, $\epsilon_1 = 0.1$.

ϵ_2	0.4	0.25	0.05	0.01	0.001	0.0001
A_6	$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
ρ	0.8333	0.6667	0.6389	0.5833	0.4167	0.1667

Table 4 shows the optimized generator matrices for various values of ϵ_2 when $k = 6$, $\epsilon_r = 0.01$ and $\epsilon_1 = 0.1$. We can see that as ϵ_2 decreases, the density of ones ρ decreases as well. For $\epsilon_2 = 0.0001$, we have $A_6 = I_6$. This is due to the fact that for small value of ϵ_2 , the relay can reliably transmits the received bits to the destination. (Note that this strategy is optimal if $\epsilon_2 = 0$). However, for high values of ϵ_2 , the relay combines several bits prior to the transmission. This helps the destination combat the noise on the relay-destination link.

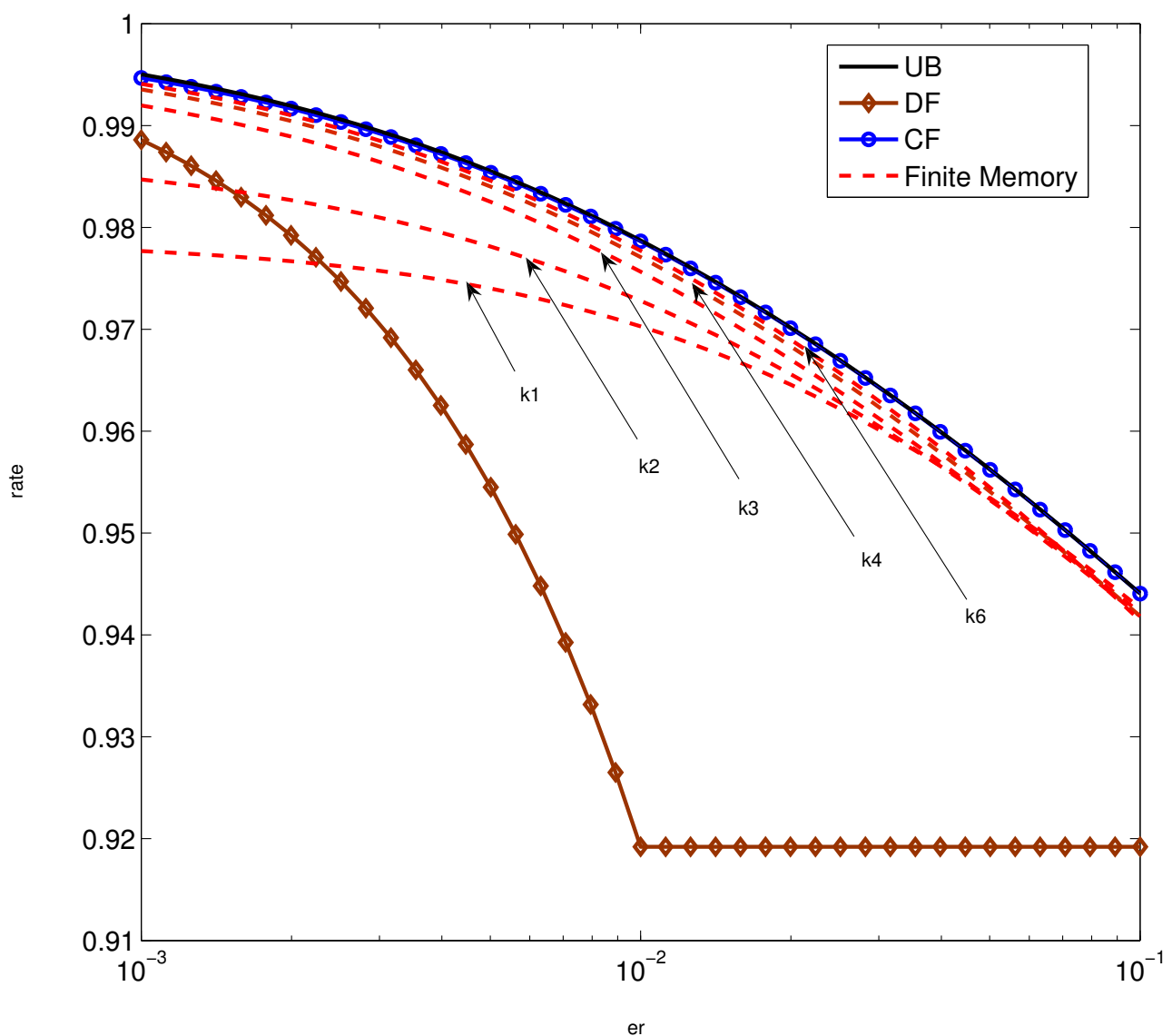
5. Numerical Examples

Figure 6 shows the capacity results for the orthogonal BSRC shown in Figure 2 as a function of ϵ_r when $\epsilon_1 = \epsilon_2 = 0.01$. In this figure, we have plotted the cut-set upper bound (UB) (Equation (7)), rates

achieved using decode-and-forward (DF) (Equation (8)), compress-and-forward (CF) (Equation (10)), and optimized finite memory relay (Equation (14)) for different memory size. The relay functions are optimized for the channel parameters $\epsilon_1 = \epsilon_2 = \epsilon_r = 0.01$ and are given in Table 1.

From Figure 6, we see that the achievable rate of DF decreases as ϵ_r increases to 0.01. For $\epsilon_r \geq 0.01$, the achievable rate of DF coincides with that achievable with direct transmission (*i.e.*, the relay is off). On the other hand, the rates achieved by CF coincides with the upper bound for the chosen channel parameters since the condition in Corollary 2 is satisfied. More interestingly, optimized low-dimensional relaying with $k = 6$ achieves rates close to those achieved by CF and operates close to the capacity.

Figure 6. Capacity results for the binary symmetric relay channel as a function of ϵ_r when $\epsilon_1 = \epsilon_2 = 0.01$.



6. Summary and Concluding Remarks

We introduced a binary symmetric relay channel with orthogonal components at the destination, and investigated three main relaying strategies: decode-and-forward (DF) relaying, compress-and-forward

(CF) relaying, and optimized low-dimensional relaying. We used a bit-switching numerical algorithm to find *optimized* mappings. We initialized our algorithm with arbitrary random *nonlinear* mappings, and after optimization based on Fourier analysis, we observed that all optimized mappings that we found were *linear*. We also illustrated that one can obtain rates very close to the upper bound by using the proposed optimized low-dimensional relaying scheme. It is worth noting that DF and CF require codebooks with *infinite* block length codewords at the relay. This stands in a sharp contrast to the proposed low-dimensional relaying scheme. Additionally, the suggested relaying protocol has low-delay processing and paves the way for implementation of inexpensive relaying protocols. We finally note that the sufficiency of linear mappings for the problem of optimal relaying remains open.

Acknowledgements

The authors wish to acknowledge Abbas El Gamal for helpful input, and the Swedish Research Council for funding the work in part.

Claims

The material in this paper in part was presented in proceedings of the IEEE ITW, Cairo, Egypt, 6–8 January 2010 [23].

Appendix

A. Proof of Proposition 1

In order to proceed, we first recite a lemma given in [13] (We thank one of the anonymous reviewers for bringing this result to our attention.) that we occasionally use in the sequel.

Lemma 2. Consider a binary symmetric channel with input X and outputs Y_1 and Y_2 , where

$$\begin{aligned} Y_1 &= X \oplus Z_1 \\ Y_2 &= X \oplus Z_2 \end{aligned}$$

The random variable $Z_1 \sim \text{Ber}(\epsilon_1)$ and is independent of $Z_2 \sim \text{Ber}(\epsilon_2)$. The capacity of this channel is given by

$$C = 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_2) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) \quad (24)$$

Proof. A proof of this lemma can be found in [24], but for completeness we present a slightly different proof in the following. The channel is a standard 1×2 SIMO (single-input multiple-output) link and its capacity is given by

$$C = \max_{p(x)} I(X; Y_1, Y_2) \quad (25)$$

Next consider

$$\begin{aligned}
 C &= H(Y_1, Y_2) - H(Y_1, Y_2|X) \\
 &= H(Y_1, Y_2) - H(Z_1, Z_2|X) \\
 &= H(Y_1, Y_2) - H(Z_1) - H(Z_2) \\
 &= H(Y_1) + H(Y_2|Y_1) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) \\
 &= H(Y_1) + H(Y_2 \oplus Y_1|Y_1) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) \\
 &\leq H(Y_1) + H(Y_2 \oplus Y_1) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) \\
 &= H(Y_1) + H(Z_2 \oplus Z_1) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) \\
 &\leq 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_2) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2)
 \end{aligned} \tag{26}$$

We finally note that the upper bound can be achieved by choosing $X \sim \text{Ber}(0.5)$. □

Proof of Proposition 1: Using the cut-set bound [2], we have

$$C \leq \max_{p(x, x_r)} \min\{I(X, X_r; Y_1, Y_2), I(X; Y_1, Y_2, Y_r|X_r)\}$$

Now we bound each term.

$$\begin{aligned}
 I(X, X_r; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X, X_r) \\
 &= H(Y_1, Y_2) - H(Z_1, Z_2|X, X_r) \\
 &= H(Y_1, Y_2) - H(Z_1) - H(Z_2) \\
 &\leq H(Y_1) + H(Y_2) - H(Z_1) - H(Z_2) \\
 &\leq 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_1)
 \end{aligned} \tag{27}$$

Similarly, we have

$$\begin{aligned}
 I(X; Y_r, Y_1, Y_2|X_r) &= I(X; Y_1, Y_r|X_r) + I(X; Y_2|Y_1, Y_r, X_r) \\
 &= I(X; Y_1, Y_r|X_r) + H(Y_2|Y_1, Y_r, X_r) \\
 &\quad - H(Y_2|Y_1, Y_r, X_r, X) \\
 &= I(X; Y_1, Y_r|X_r) + H(Z_2|Y_1, Y_r, X_r) \\
 &\quad - H(Z_2|Y_1, Y_r, X_r, X) \\
 &= I(X; Y_1, Y_r|X_r) \\
 &= H(Y_1, Y_r|X_r) - H(Y_1, Y_r|X, X_r) \\
 &= H(Y_1, Y_r|X_r) - H(Z_1, Z_r|X, X_r) \\
 &= H(Y_1, Y_r|X_r) - H(Z_1, Z_r) \\
 &\leq H(Y_1, Y_r) - H(Z_1, Z_r) \\
 &= I(X; Y_1, Y_r) \\
 &\leq 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r) - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_r)
 \end{aligned} \tag{28}$$

where the last inequality follows from Lemma 2. Combining (27) and (28) and noting that (27) and (28) are maximized when the input distribution is chosen as

$$p(x, x_r) = p(x)p(x_r) = [0.5\delta_b(x) + 0.5\delta_b(x \oplus 1)] [0.5\delta_b(x_r) + 0.5\delta_b(x_r \oplus 1)]$$

yield the result.

B. Proof of Proposition 2

Using Theorem 1 in [2], the rate

$$R_{DF} = \max_{p(x, x_r)} \min\{I(X, X_r; Y_1, Y_2), I(X; Y_r|X_r)\}$$

is achievable. The first term is evaluated in Appendix A and is maximized when X and X_r are independent and have uniform distribution. One can show the same distribution maximizes the second term. This yields

$$R_{DF} = \min\{1 - \mathbb{H}_b(\epsilon_r), 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2)\}. \quad (29)$$

C. Proof of Proposition 3

Using Theorem 7 in [2], the rate

$$R_{PDF} = \max_{p(u, x, x_r)} \min\{I(X, X_r; Y_1, Y_2), I(U; Y_r|X_r) + I(X; Y_1, Y_2|X_r, U)\}$$

is achievable. (See also [15].) The first term is evaluated in Appendix A and is bounded as

$$I(X, X_r; Y_1, Y_2) \leq 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) \quad (30)$$

Next consider

$$\begin{aligned} & I(U; Y_r|X_r) + I(X; Y_1, Y_2|X_r, U) \\ &= I(U; Y_r|X_r) + H(Y_1, Y_2|X_r, U) - H(Y_1, Y_2|X, X_r, U) \\ &= I(U; Y_r|X_r) + H(Y_1, Y_2|X_r, U) - H(Z_1, Z_2|X, X_r, U) \\ &= I(U; Y_r|X_r) + H(Y_1, Y_2|X_r, U) - H(Z_1) - H(Z_2) \\ &= I(U; Y_r|X_r) + H(Y_1|X_r, U) + H(Y_2|X_r, U, Y_1) - H(Z_1) - H(Z_2) \\ &= I(U; Y_r|X_r) + H(Y_1|X_r, U) + H(Z_2|X_r, U, Y_1) - H(Z_1) - H(Z_2) \\ &= H(Y_r|X_r) - H(Y_r|X_r, U) + H(Y_1|X_r, U) - H(Z_1) \\ &\leq H(Y_r) - H(Y_r|X_r, U) + H(Y_1|X_r, U) - H(Z_1) \\ &\leq 1 - \mathbb{H}_b(\epsilon_1) + H(Y_1|X_r, U) - H(Y_r|X_r, U) \end{aligned} \quad (31)$$

Now we bound $H(Y_1|X_r, U) - H(Y_r|X_r, U)$. First, define $V := (X_r, U)$ where $p(V = v_i) = p_i$ and $\sum_i p_i = 1$. Further assume that $p(x = 0|V = v_i) = \delta_i$, $p(x = 1|V = v_i) = 1 - \delta_i$. Next consider

$$\begin{aligned}
H(Y_1|X_r, U) - H(Y_r|X_r, U) &= H(X + Z_1|V) - H(X + Z_r|V) \\
&= \sum_i (H(X + Z_1|V = v_i) - H(X + Z_r|V = v_i)) p_i \\
&\leq \sum_i \max_{\delta_i} (H(X + Z_1|V = v_i) - H(X + Z_r|V = v_i)) p_i \\
&= \max_{\delta_i} (H(X + Z_1|V = v_i) - H(X + Z_r|V = v_i)) \sum_i p_i \\
&= \max_{\delta_i} (H(X + Z_1|V = v_i) - H(X + Z_r|V = v_i)) \\
&= \max_{\delta_i} (\mathbb{H}_b(\delta_i * \epsilon_1) - \mathbb{H}_b(\delta_i * \epsilon_r))
\end{aligned} \tag{32}$$

In the following, let $\mathfrak{h}(\delta) := \mathbb{H}_b(\delta * \epsilon_1) - \mathbb{H}_b(\delta * \epsilon_r)$ and $\max\{\epsilon_1, \epsilon_r\} \leq 0.5$. We then obtain

$$\begin{aligned}
\frac{\partial \mathfrak{h}}{\partial \delta} &= (1 - 2\epsilon_1) \log \left(\frac{1 - (\delta * \epsilon_1)}{\delta * \epsilon_1} \right) - (1 - 2\epsilon_r) \log \left(\frac{1 - (\delta * \epsilon_r)}{\delta * \epsilon_r} \right) \\
\frac{\partial^2 \mathfrak{h}}{\partial \delta^2} &= \left(\frac{(1 - 2\epsilon_r)^2}{(\delta * \epsilon_r)(1 - (\delta * \epsilon_r))} - \frac{(1 - 2\epsilon_1)^2}{(\delta * \epsilon_1)(1 - (\delta * \epsilon_1))} \right)
\end{aligned} \tag{33}$$

Now let $g(\epsilon) := \frac{(1-2\epsilon)^2}{(\delta * \epsilon)(1 - (\delta * \epsilon))}$. One can show that $\frac{\partial g}{\partial \epsilon} \leq 0$ if $\epsilon \leq 0.5$ and hence $g(\epsilon)$ is a non-increasing function. Thus we conclude that

$$\begin{cases} \epsilon_r \leq \epsilon_1 \implies g(\epsilon_r) \geq g(\epsilon_1) \implies \frac{\partial^2 \mathfrak{h}}{\partial \delta^2} \geq 0 \implies \mathfrak{h}(\delta) \text{ is convex} \\ \epsilon_r \geq \epsilon_1 \implies g(\epsilon_r) \leq g(\epsilon_1) \implies \frac{\partial^2 \mathfrak{h}}{\partial \delta^2} \leq 0 \implies \mathfrak{h}(\delta) \text{ is concave} \end{cases} \tag{34}$$

Therefore

$$\begin{cases} \mathfrak{h}(\delta) \leq \max\{\mathfrak{h}(0), \mathfrak{h}(1)\}, & \text{if } \epsilon_r \leq \epsilon_1 \\ \mathfrak{h}(\delta) \leq \mathfrak{h}(\delta^*), & \text{if } \epsilon_r \geq \epsilon_1 \end{cases} \tag{35}$$

where δ^* is the solution of $\frac{\partial \mathfrak{h}}{\partial \delta} = 0$. Finally we obtain the following bound

$$H(Y_1|X_r, U) - H(Y_r|X_r, U) \leq \begin{cases} \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_r), & \text{if } \epsilon_r \leq \epsilon_1 \\ 0, & \text{otherwise} \end{cases} \tag{36}$$

Combining Equations (30), (31) and (36) proves that partial DF does not improve on DF for the BSRC.

D. Proof of Proposition 4

We use the equivalent formulation of the original CF (Theorem 6 in [2]) given in [4]. CF achieves the rate

$$R_{CF} = \max_{p(q)p(x|q)p(x_r|q)p(y_r|x_r, y_r, q)} \min \left\{ I(X, X_r; Y_1, Y_2|Q) - I(Y_r; \hat{Y}_r|X, X_r, Y_1, Y_2, Q), \right. \\ \left. I(X; Y_1, Y_2, \hat{Y}_r|X_r, Q) \right\} \tag{37}$$

First consider

$$\begin{aligned}
 I(X, X_r; Y_1, Y_2 | Q) &= I(X; Y_1, Y_2 | Q) + I(X_r; Y_1, Y_2 | X, Q) \\
 &= I(X; Y_1 | Q) + I(X; Y_2 | Y_1, Q) + I(X_r; Y_2 | X, Q) + I(X_r; Y_1 | Y_2, X, Q) \\
 &= I(X; Y_1 | Q) + I(X_r; Y_2 | X, Q) \\
 &= I(X; Y_1 | Q) + I(X_r; Y_2 | Q) \\
 &\leq 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2)
 \end{aligned} \tag{38}$$

where the upper bound can be achieved by choosing $X \sim \text{Ber}(0.5)$ and $X_r \sim \text{Ber}(0.5)$.

In order to proceed we choose the following binary test channel:

$$\hat{Y}_r = Y_r \oplus Z_q \tag{39}$$

where $Z_q \sim \text{Ber}(\epsilon_q)$ and is independent of other random variables.

This yields

$$\begin{aligned}
 I(Y_r; \hat{Y}_r | X, X_r, Y_1, Y_2, Q) &= H(\hat{Y}_r | X, X_r, Y_1, Y_2, Q) - H(\hat{Y}_r | Y_r, X, X_r, Y_1, Y_2, Q) \\
 &= H(Z_r \oplus Z_q) - H(Z_q) = \mathbb{H}_b(\epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_q)
 \end{aligned} \tag{40}$$

$$\begin{aligned}
 I(X; Y_1, Y_2, \hat{Y}_r | X_r, Q) &= I(X; Y_1, \hat{Y}_r | X_r, Q) + I(X; Y_2 | Y_1, \hat{Y}_r, X_r, Q) \\
 &= I(X; Y_1, \hat{Y}_r | Q) \\
 &\leq 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_1)
 \end{aligned} \tag{41}$$

where the last inequality follows from Lemma 2 and it is achieved by choosing $X \sim \text{Ber}(0.5)$.

Putting all together, the following rate is achievable

$$\begin{aligned}
 R_{CF} = \max_{\epsilon_q \in [0,1]} \min \left\{ 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) - \mathbb{H}_b(\epsilon_r * \epsilon_q) + \mathbb{H}_b(\epsilon_q), \right. \\
 \left. 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_1) \right\}
 \end{aligned} \tag{42}$$

Now define

$$\begin{aligned}
 R_1(\epsilon_q) &:= 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) - \mathbb{H}_b(\epsilon_r * \epsilon_q) + \mathbb{H}_b(\epsilon_q) \\
 R_2(\epsilon_q) &:= 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_1)
 \end{aligned} \tag{43}$$

Using the fact that $f(\epsilon) := \mathbb{H}_b(\epsilon * \delta) - \mathbb{H}_b(\epsilon)$ is convex $\forall \delta \in [0, 1]$, we conclude that $R_1(\epsilon_q)$ is concave and $R_2(\epsilon_q)$ is convex in ϵ_q . We next note that

$$\begin{aligned}
 \max_{\epsilon_q} R_1 &= R_1(0.5) = 2 - \mathbb{H}_b(\epsilon_1) - \mathbb{H}_b(\epsilon_2) \\
 \min_{\epsilon_q} R_2 &= R_2(0.5) = 1 - \mathbb{H}_b(\epsilon_1)
 \end{aligned} \tag{44}$$

Since $R_2(0.5) \leq R_1(0.5)$ we only need to consider two following cases:

- Case 1: $R_1(0) < R_2(0)$

If $R_1(0) < R_2(0)$ we have $1 - \mathbb{H}_b(\epsilon_2) < \mathbb{H}_b(\epsilon_1 * \epsilon_r)$ and there exists ϵ_q such that $R_1(\epsilon_q) = R_2(\epsilon_q)$. Thus

$$R_{CF} = 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_1) \quad (45)$$

is achievable where ϵ_q satisfies

$$\mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_q) - \mathbb{H}_b(\epsilon_q) + \mathbb{H}_b(\epsilon_2) = 1 \quad (46)$$

- Case 2: $R_1(0) \geq R_2(0)$

If $R_1(0) \geq R_2(0)$ we have $1 - \mathbb{H}_b(\epsilon_2) \geq \mathbb{H}_b(\epsilon_1 * \epsilon_r)$. Thus

$$R_{CF} = \max_{\epsilon_q} \min\{R_1(\epsilon_q), R_2(\epsilon_q)\} = \max_{\epsilon_q} R_2(\epsilon_q) = R_2(0) = 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r) - \mathbb{H}_b(\epsilon_r) - \mathbb{H}_b(\epsilon_1) \quad (47)$$

is achievable.

E. Proof of Proposition 5

For the memoryless relay, we have $x_r = y_r$ and hence

$$\begin{aligned} Y_1 &= X \oplus Z_1 \\ Y_2 &= X_r \oplus Z_2 = X \oplus Z_r \oplus Z_2 = X \oplus Z_{eq} \end{aligned} \quad (48)$$

where $Z_{eq} := Z_r \oplus Z_2 \sim \text{Ber}(\epsilon_r * \epsilon_2)$. Then the achievable rate is given by $C_1 = \max_{p(x)} I(X; Y_1, Y_2)$. Now using Lemma 2, we obtain

$$C_1 = 1 + \mathbb{H}_b(\epsilon_1 * \epsilon_r * \epsilon_2) - \mathbb{H}_b(\epsilon_r * \epsilon_2) - \mathbb{H}_b(\epsilon_1) \quad (49)$$

F. Proof of Lemma 1

Consider

$$\begin{aligned} I(X_1^k; Y_{1,1}^k, Y_{2,1}^k) &= I(X_1^k; Y_{1,1}^k) + I(X_1^k; Y_{2,1}^k | Y_{1,1}^k) \\ &\stackrel{(a)}{=} kI(X_1; Y_1) + I(X_1^k; Y_{2,1}^k | Y_{1,1}^k) \\ &= kI(X_1; Y_1) + H(Y_{2,1}^k | Y_{1,1}^k) - H(Y_{2,1}^k | Y_{1,1}^k, X_1^k) \\ &\stackrel{(b)}{=} kI(X_1; Y_1) + H(Y_{2,1}^k | Y_{1,1}^k) - H(Y_{2,1}^k | X_1^k) \\ &= kI(X_1; Y_1) + H(Y_{2,1}^k | Y_{1,1}^k) - H(Y_{2,1}^k | X_1^k) \\ &= k(1 - \mathbb{H}_b(\epsilon_1)) - \mathbb{E} [\log_2(p(y_{2,1}^k | y_{1,1}^k))] + \mathbb{E} [\log_2(p(y_{2,1}^k | x_1^k))] \end{aligned} \quad (50)$$

where (a) holds since $\{X_i\}_{i=1}^k$ are i.i.d. and the channel is memoryless and (b) holds since $Y_{1,1}^k = X_1^k \oplus Z_{1,1}^k$ and $Z_{1,1}^k$ is independent of other random sequences. The conditional probabilities can be computed as follows

$$\begin{aligned}
 p(y_{2,1}^k | x_1^k) &= \sum_{y_{r,1}^k} p(y_{2,1}^k | x_1^k, y_{r,1}^k) p(y_{r,1}^k | x_1^k) \\
 &= \sum_{y_{r,1}^k} p(y_{2,1}^k | y_{r,1}^k) p(y_{r,1}^k | x_1^k) \\
 &= \sum_{y_{r,1}^k} p(y_{2,1}^k | y_{r,1}^k) \prod_{i=1}^k p(y_{r,i} | x_i) \\
 &= \sum_{y_{r,1}^k} \prod_{i=1}^k p(y_{2,i} | f_i(y_{r,1}^k)) \prod_{i=1}^k p(y_{r,i} | x_i) \\
 &= \sum_{y_{r,1}^k} \epsilon_2^{\mathbb{D}_H(y_{2,1}^k, f(y_{r,1}^k))} (1 - \epsilon_2)^{k - \mathbb{D}_H(y_{2,1}^k, f(y_{r,1}^k))} \\
 &\quad \times \epsilon_r^{\mathbb{D}_H(y_{r,1}^k, x_1^k)} (1 - \epsilon_r)^{k - \mathbb{D}_H(y_{r,1}^k, x_1^k)} \\
 &= (1 - \epsilon_2)^k (1 - \epsilon_r)^k \times \\
 &\quad \sum_{y_{r,1}^k} \left(\frac{\epsilon_2}{1 - \epsilon_2} \right)^{\mathbb{D}_H(y_{2,1}^k, f(y_{r,1}^k))} \left(\frac{\epsilon_r}{1 - \epsilon_r} \right)^{\mathbb{D}_H(y_{r,1}^k, x_1^k)} \tag{51}
 \end{aligned}$$

We similarly obtain

$$\begin{aligned}
 p(y_{2,1}^k | y_{1,1}^k) &= \sum_{y_{r,1}^k} p(y_{2,1}^k | y_{1,1}^k, y_{r,1}^k) p(y_{r,1}^k | y_{1,1}^k) \\
 &= \sum_{y_{r,1}^k} p(y_{2,1}^k | y_{r,1}^k) p(y_{r,1}^k | y_{1,1}^k) \\
 &= \sum_{y_{r,1}^k} p(y_{2,1}^k | y_{r,1}^k) \sum_{x_1^k} p(y_{r,1}^k | y_{1,1}^k, x_1^k) p(x_1^k | y_{1,1}^k) \\
 &= \sum_{y_{r,1}^k} p(y_{2,1}^k | y_{r,1}^k) \sum_{x_1^k} p(y_{r,1}^k | x_1^k) \frac{p(y_{1,1}^k | x_1^k) p(x_1^k)}{p(y_{1,1}^k)} \\
 &= \sum_{y_{r,1}^k} \prod_{i=1}^k p(y_{2,i} | f_i(y_{r,1}^k)) \sum_{x_1^k} \prod_{i=1}^k p(y_{r,i} | x_i) \prod_{i=1}^k p(y_{1,i} | x_i) \\
 &= (1 - \epsilon_1)^k (1 - \epsilon_2)^k (1 - \epsilon_r)^k \times \sum_{y_{r,1}^k} \left(\frac{\epsilon_2}{1 - \epsilon_2} \right)^{\mathbb{D}_H(y_{2,1}^k, f(y_{r,1}^k))} \times \\
 &\quad \sum_{x_1^k} \left(\frac{\epsilon_r}{1 - \epsilon_r} \right)^{\mathbb{D}_H(y_{r,1}^k, x_1^k)} \left(\frac{\epsilon_1}{1 - \epsilon_1} \right)^{\mathbb{D}_H(y_{1,1}^k, x_1^k)} \tag{52}
 \end{aligned}$$

References

1. van der Meulen, E.C. Three-terminal communication channels. *Adv. Appl. Probab.* **1971**, *3*, 120–154.
2. Cover, T.M.; El Gamal, A. Capacity theorems for the relay channel. *IEEE Trans. Inform. Theory* **1979**, *25*, 572–584.
3. Høst-Madsen, A.; Zhang, J. Capacity bounds and power allocation for wireless relay channels. *IEEE Trans. Inform. Theory* **2006**, *51*, 2020–2040.
4. El Gamal, A.; Mohseni, M.; Zahedi, S. Bounds on capacity and minimum energy-per-bit for AWGN relay channels. *IEEE Trans. Inform. Theory* **2006**, *52*, 1545–1561.
5. Kramer, G.; Gastpar, M.; Gupta, P. Cooperative strategies and capacity theorems for relay networks. *IEEE Trans. Inform. Theory* **2005**, *51*, 3037–3063.
6. Dabora, R.; Servetto, S.D. On the role of estimate-and-forward with time sharing in cooperative communication. *IEEE Trans. Inform. Theory* **2008**, *54*, 4409–4431.
7. Laneman, J.N.; Wornell, G.W.; Tse, D.N.C. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. Inform. Theory* **2004**, *50*, 3062–3080.
8. Sendonaris, A.; Erkip, E.; Aazhang, B. User cooperation diversity-Part I: System description. *IEEE Trans. Commun.* **2003**, *51*, 1927–1938.
9. Sendonaris, A.; Erkip, E.; Aazhang, B. User cooperation diversity-Part II: Implementation aspects and performance analysis. *IEEE Trans. Commun.* **2003**, *51*, 1939–1948.
10. Aleksic, M.; Razaghi, P.; Yu, W. Capacity of a class of modulo-sum relay channels. *IEEE Trans. Inform. Theory* **2009**, *55*, 921–930.
11. Kim, Y.H. Coding Techniques for Primitive Relay Channels. In Proceedings of the Forty-Fifth Annual Allerton Conference, Allerton House, UIUC, IL, USA, 26–28 September 2007.
12. Lau, A.P.T.; Cui, S. Joint power minimization in wireless relay channels. *IEEE Trans. Wirel. Commun.* **2007**, *6*, 2820–2824.
13. Karystinos, G.N.; Liavas, A.P. Outage Capacity of a Cooperative Scheme with Binary Input and a Simple Relay. In Proceedings of the IEEE ICASSP 2008-International Conference Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008.
14. Sagar, Y.; Kwon, H.M.; Ding, Y. Capacity of Modulo-Sum Simple Relay Network. In Proceedings International Zurich Seminar on Communications, Zurich, Switzerland, 3–5 March 2010.
15. El Gamal, A.; Aref, M. The capacity of the semi-deterministic relay channel. *IEEE Trans. Inform. Theory* **1982**, *28*, 536.
16. Khormuji, M.N.; Larsson, E.G. Rate-optimized constellation rearrangement for the relay channel. *IEEE Commun. Lett.* **2008**, *12*, 618–620.
17. Khormuji, M.N.; Skoglund, M. Piecewise linear relaying: Low complexity parametric relaying. In Proceedings of the IEEE SPAWC, Perugia, Italy, 21–24 June 2009.
18. Khormuji, M.N.; Skoglund, M. On instantaneous relaying. *IEEE Trans. Inform. Theory* **2010**, *56*, 3378–3394.
19. Zaidi, A.; Khormuji, M.N.; Yao, S.; Skoglund, M. Rate-maximizing Mappings for Memoryless Relaying. In Proceedings of the IEEE ISIT, Seoul, Korea, 28 June–3 July, 2009.

20. Skoglund, M. On channel-constrained vector quantization and index assignment for discrete memoryless channels. *IEEE Trans Inform. Theory* **1999**, *45*, 2615–2622.
21. Mehes, A.; Zeger, K. Performance of quantizers on noisy channels using structured families of codes. *IEEE Trans. Inform. Theory* **2000**, *46*, 2468–2476.
22. Rudin, W. *Fourier Analysis on Groups*; John Wiley And Sons Ltd: Hoboken, NJ, USA, 1990.
23. Khormuji, M.N.; Skoglund, M. On the Capacity of the Binary Symmetric Relay Channel with a Finite Memory Relay. In Proceedings of the IEEE ITW, Cairo, Egypt, 6–8 January 2010.
24. Liavas, A. Outage Capacity of a Cooperative Scheme with Binary Input and a Simple Relay. Year-1 Work-Package-1 report of European Commission FET Project FP6-033533-COOPCOM, November 2007.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).