Coded Bidirectional Relaying in Wireless Networks

Petar Popovski and Toshiaki Koike - Akino

Abstract The communication strategies for coded bidirectional (two–way) relaying emerge as a result of successful amalgamation of the recent ideas from network coding and the broadcast/interfering nature of the wireless communication medium. In this chapter we consider the basic scenario in which two terminal nodes carry out a two–way communication, helped by a third, relay node. Insights are provided into several important ideas that arise in this simple communication scenario. The communication mechanisms can be divided into two groups, depending whether the relay node is or is not required to decode the messages from the terminals. For the class of techniques in which the relay *denoises* rather than decodes the messages from the terminals, we discuss the role of structured codes in achieving the highest aggregate rates in the system. The treatment of the topics is generally in an information–theoretic setting, but we also present the design of communication mechanisms when finite–length packets and practical modulation constellations are used.

1 Introduction

The techniques for coded bidirectional or two–way relaying have received significant attention in the recent years [1–15]. The mechanisms for two–way relaying or, more general, multi–way relaying, leverage on two conceptual blocks. The *first conceptual block* is the shared nature of the wireless communication medium. On one hand, this implies that there is interference when multiple transmissions are oc-

---

Petar Popovski
Department of Electronic Systems, Aalborg University, Niels Jernes Vej 12, DK-9220 Aalborg, Denmark and Oticon A/S, Kongebakken 9, Smørum, Denmark. e-mail: petarp@es.aau.dk

Toshiaki Koike - Akino
School of Engineering and Applied Sciences, Harvard University, 33 Oxford Street, Cambridge, MA 02138, U.S.A. e-mail: koike@seas.harvard.edu

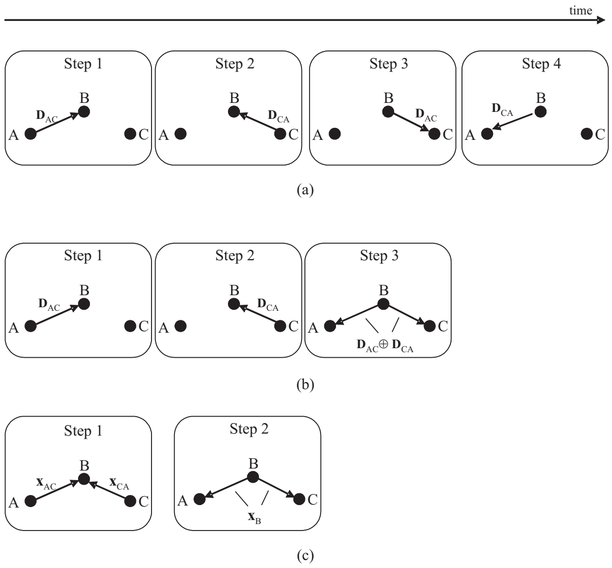


Fig. 1 (a) Uncoded bidirectional relaying (b) Example of coded bidirectional relaying with three steps (phases) (c) Example of coded bidirectional relaying with two steps.

curring simultaneously. On the other hand, the *wireless broadcasting is "cheap"* in a sense that a single transmission can be received by multiple nodes. The *second conceptual block* is the idea of network coding. In short, the traditional design of communication networks observes the data flows as conventional commodity flows. Therefore, a routing node in the network essentially replicates the data packets from an incoming link to an outgoing link (or multiple links, in case of multicast). The network coding recognizes that a data flow is different from a physical commodity flow and generalizes the routing such that the data on a given outgoing link is a function of the data from two or more incoming links.

To see how these building blocks accrue into novel modes of wireless communication, consider the example on Fig. 1 [1]. The communication scenario is that the node A has packets destined to C and vice versa. However, in this example we assume that the capacity of the direct channel between A and C is zero and the communication between them must be done by using B as a relay node. The packet from the source node i to the destination j is denoted by $\mathbf{D}_{ij}$. For this example, all the packets have identical sizes. The node B is neither source nor destination of the data traffic. The conventional (uncoded) relaying regards the transmission of $\mathbf{D}_{AC}$ and $\mathbf{D}_{CA}$ as two separate problems. Fig. 1(a) shows that the uncoded relaying consumes four steps, each step having a duration of a time slot. On the other hand, the coded bidirectional relaying Fig. 1(b) considers the two transmissions as a single problem. In the first two slots the relay gathers the packets $\mathbf{D}_{AC}$ and $\mathbf{D}_{CA}$ from the respective source node. In Step 3 B broadcasts the packet $\mathbf{D}_B = \mathbf{D}_{AC} \oplus \mathbf{D}_{CA}$, where $\oplus$ is XOR operation. After receiving $\mathbf{D}_B$, the node A decodes $\mathbf{D}_{CA} = \mathbf{D}_B \oplus \mathbf{D}_{AC}$ by using its a priori knowledge of $\mathbf{D}_{AC}$. In analogous way, C decodes $\mathbf{D}_{AC}$. Thus, to transfer the same amount of data, the coded two–way relaying requires only three time slots, which is an improvement of 33 % over the uncoded relaying. This simple example clearly illustrates the two building blocks, mentioned above: (1) it utilizes the feature of "cheap wireless broadcast" to save one transmission slot; and (2) the packets of different incoming communication flows are combined before being sent over the outgoing link (broadcast) of the relay node B.

In the example on Fig. 1(b) the relay node gathers the packets from from A and C in a time–division manner. Alternatively, A and C can simultaneously transmit over the multiple access channel [3], as shown in Fig. 1(c), where we use the notation $\mathbf{x}_i$ to denote the baseband signal transmitted by the node i. In Step 2, B transmits the signal $\mathbf{x}_B$, which is created as a function of the signal that B received in Step 1. This function does not necessarily assume that B is able to decode the individual signals. In the simplest case, B needs only to amplify the signal that it has received in Step 1 and broadcast it back to A and C in Step 2. Since A (C) a priori knows its contribution to the interfered signal received at B in Step 1, then it can utilize this information to reliably extract $\mathbf{x}_C$ ($\mathbf{x}_A$) from the signal $\mathbf{x}_B$. Ideally, if both A and C decode each other's signals correctly, the time to transmit the data is only two steps, which means throughput improvement over the uncoded relaying of 100 %.

These illustrations of the coded two–way relaying suggest that the underlying ideas have a great potential to improve the performance of wireless networks. The chapter does not aim to cover all the aspects of the two–way relaying, which for example include multiple–antenna techniques, scenario–specific channel estimation, scheduling, etc. Instead, the objective is to elaborate on several important ideas and insights brought by the two–way relaying scenario. We will see that already this simple three–node scenario gives rise to many novel techniques, such that the cases with more than three nodes are outside the scope.

The text is organized as follows. After the preliminaries in the next section, we first describe the bidirectional relaying techniques that require decoding at the relay. Section 4 is dedicated to the discussion of the techniques used when the relay does not decode the messages from the terminals. A class of such techniques is based on denoising (rather than decoding) and Section 5 outlines the special significance that the structured codes have for such denoising techniques. The next section departs form the information–theoretic setting to provide insight into the more practical aspects of the coded bidirectional relaying, by considering finite–length packets and practical modulation setting. The last section concludes the chapter.

2 Preliminaries

We assume that there are only two communication flows, $A \to C$ and $C \to A$, respectively. The relay B is neither a source nor a sink of any data in the system. All the nodes are half–duplex, such that a node can either transmit or receive at a given time. The rate at which node $i \in \{A,B,C\}$ transmits is denoted by R_i. During the transmission, the node i has n_i channel uses, such that the message that it sends has n_iR_i bits. Let us denote by $N > n_i$ the total number of channel uses during the whole round of two–way relaying. If after N channel uses the message $W_{AC}(W_{CA})$ is received at $C(A)$ successfully, then the rate R_{AC} achieved from A to C and the rate R_{CA} achieved from C to A are given as:

$$R_{AC} = \frac{n_A R_A}{N} \qquad R_{CA} = \frac{n_C R_C}{N} \tag{1}$$

We will be interested in determining the rate pair (R_{AC}, R_{CA}) and the sum (two–way) rate $R_{AC} + R_{CA}$. We will assume that in each round A and C transmit only fresh data, independent of any information exchange from the previous rounds [1].

The message sent from node i and destined for node j is denoted by W_{ij} and the corresponding binary representation is the vector $\mathbf{w}_{ij}$. The size (in bits) of the message is denoted by $|\mathbf{w}_{ij}|$. The codeword transmitted by node i is denoted by $\mathbf{x}_i$ and is a vector of dimension n_i, whose $m-$th element is denoted by $x_i[m]$. The random variable that stands for a symbol sent (received) by node i is denoted by $X_i(Y_i)$. The received vector at node j is $\mathbf{y}_j$. We mainly consider Gaussian channels, explicitly stating if the channel is discrete.

For Gaussian channels X_i, Y_j are complex numbers, unless stated otherwise (Section 5.2). If only one node $U \in \{A, B, C\}$ is transmitting, then the received symbol at the node $V \in \{A, B, C\} \setminus U$ is given by:

$$Y_V = h_{UV} X_U + Z_V \tag{2}$$

or if we want to emphasize that it is the $m-$th symbol:

$$y_V[m] = h_{UV} x_U[m] + z_V[m] \tag{3}$$

where h_{UV} is the complex channel coefficient between U and V. $z_V[m]$ is the complex additive white Gaussian noise $\mathscr{CN}(0, N_0)$. The transmitted symbols have $E\{x_U[m]\} = 0$ and a normalized power $E\{|x_U[m]|^2\} = 1$. If A and C transmit simultaneously, then B receives:

$$Y_B = h_1 X_A + h_2 X_B + Z_B \tag{4}$$

Each node uses the same transmission power, which makes the links symmetric:

$$h_{AC} = h_{CA} = h_0; \quad h_{AB} = h_{BA} = h_1; \quad h_{CB} = h_{BC} = h_2 \tag{5}$$

This assumption is certainly restrictive, as one can optimize the two–way transmissions by allocating appropriate power to the nodes, while keeping some global power constraint for all the transmitters; however, that discussion is outside the scope of this text. The bandwidth is normalized to 1 Hz, such that the time is measured in number of symbols (channel uses) and the signal–to–noise ratios (SNR) is

[1] This is analogous to the what Shannon in [16] refers as a scheme for capacity for the *restricted* two–way channel

given as:

$$\gamma_i = \frac{|h_i|^2}{N_0} \qquad i = 0, 1, 2 \tag{6}$$

and a point–to–point link with SNR of γ can reliably transfer up to:

$$C(\gamma) = \log_2(1+\gamma) \text{ [bit/s]} \tag{7}$$

Without loss of generality, we can assume that

$$\gamma_2 \geq \gamma_1 \tag{8}$$

and we also assume that the direct link is worse than both links, i. e. $\gamma_0 < \gamma_1$. We will see that there are schemes in which the signal received over the direct link can be used as a side information to improve the end–to–end rates.

3 Two–way Relaying with Decoding at the Relay

Here we discuss the methods for two–way relaying in which the relay node decodes the messages W_{AC} and W_{CA}. We consider three–step scheme called *Decode–and–Forward (DF)* and two–step scheme, termed *Joint–Decode–and–Forward (JDF)*. Both schemes consist of an *uplink* phase and a *broadcast* phase. In the uplink phase B gathers data from A and C, while in the broadcast phase B transmits to A, C.

3.1 The Uplink Phase

The uplink phase of three–step scheme consists of two steps. In Step 1 the node A transmits W_{AC} using n_A symbols and in Step 2 the node C transmits W_{CA} using n_C symbols. The rates R_A and R_C should be chosen:

$$R_A \leq C(\gamma_1) \qquad R_C \leq C(\gamma_2) \tag{9}$$

where $C(\cdot)$ is defined in(7). After the uplink phase, B successfully decodes the $n_A R_A$ bits sent by A and the $n_C R_C$ bits sent by C. The message that B needs to relay to C and A is denoted W_{BC} and W_{BA}, respectively. In the simplest case, the direct link between A and C is considered to have zero capacity, such that after the uplink phase, A has still not any information about the packet sent by C, and vice versa. Hence, when $\gamma_0 = 0$, B needs to completely relay the messages W_{AC} and W_{CA}. However, in general $\gamma_0 > 0$, such that the direct link between A and C carries non–zero information. In this case, at each step of the 3–step scheme there is broadcast transmission, since one node is transmitting and two are receiving. In Step 1, A broadcasts to B and C at a rate R_A, where $R_0 \leq R_A \leq C(\gamma_1)$. The node C receives some partial information from the broadcast of A and vice versa, which decreases the amount of

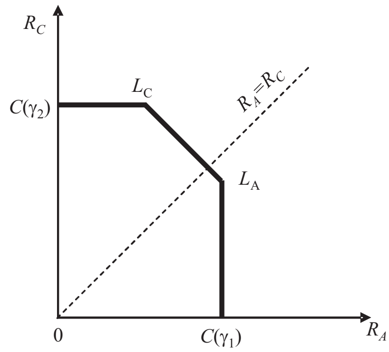


Fig. 2 Achievable rate region for the multiple–access channel used in the uplink phase.

data that needs to be broadcasted by B. In this case, W_{BC} contains $n_A[R_A - R_0]$ bits, while W_{BA} has $n_C[R_C - R_0]$ bits, where $R_0 \leq C(\gamma_0)$. The relay can use random binning [17] to create the messages that need to be relayed. With such approach, W_{BC} is uniquely determined from W_{AC} and W_{BA} is uniquely determined from W_{CA}. The uplink transmission of the three–step scheme is illustrated on Fig. 3(a).

While the uplink phase of the three–step scheme consists of two broadcast transmissions, the uplink phase of the JDF scheme consists of two simultaneous transmissions over a multiple access (MA) channel. In this case the number of channel uses by A and C is equal $n_A = n_C$, and the rates R_A and R_C should be selected within the capacity region of the MA channel [17] with B as a receiver (see Fig. 2):

$$\begin{aligned} R_A \leq C(\gamma_1) \qquad R_C \leq C(\gamma_2) \\ R_A + R_C \leq C(\gamma_1 + \gamma_2) \end{aligned} \tag{10}$$

Due to the half–duplex restriction, in Step 1 both A and C cannot receive information over the direct link, such that the relayed messages cannot be reduced and $W_{BC} = W_{AC}, W_{BA} = W_{CA}$. The uplink transmission of the JDF scheme is shown on Fig. 3(b).

3.2 The Broadcast Phase

A feature of the two schemes described above is that the relay node B knows the messages W_{AC}, W_{CA} and has a complete freedom in combining them for the broadcast phase. We first describe a simple broadcast strategy that combines the data by using XOR. In relation to the quantity of data that B should relay to A, C, there are two different cases:

- $|\mathbf{w}_{BA}| \geq |\mathbf{w}_{BC}|$. The shorter packet $\mathbf{w}_{BC}$ is padded with $|\mathbf{w}_{BA}| - |\mathbf{w}_{BC}|$ zeros and the padded packet is denoted by $\mathbf{w}_{BC}^P$. Then B broadcasts the packet $\mathbf{w}_{BA} \oplus \mathbf{w}_{BC}^P$ at the transmission rate $R_B \leq C(\gamma_1)$, limited by the SNR of the weaker link. This is

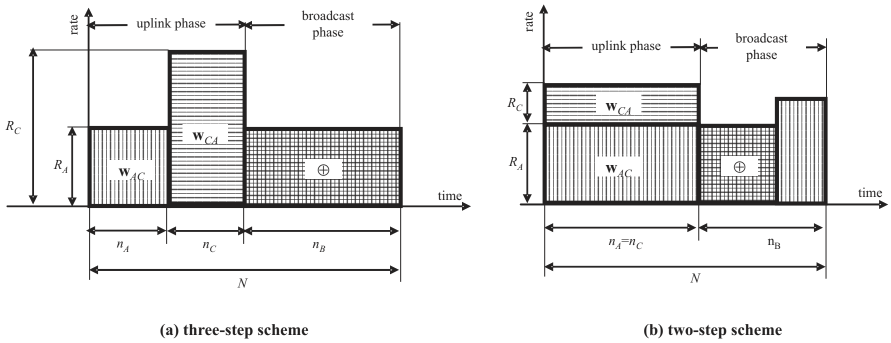


Fig. 3 Illustration of the uplink and the broadcast phase for the schemes with decoding at the relay.

required since both A and C should correctly receive the XOR packet. This case of the broadcast phase is illustrated on Fig. 3(a).

- $|\mathbf{w}_{BA}| < |\mathbf{w}_{BC}|$. In this case we take the first $|\mathbf{w}_{BA}|$ bits of $\mathbf{w}_{BC}$ to create the packet $\mathbf{w}_{BC}^{(1)}$. The vector of the last $|\mathbf{w}_{BC}| - |\mathbf{w}_{BA}|$ bits of $\mathbf{w}_{BC}$ is denoted $\mathbf{w}_{BC}^{(2)}$. The relay node transmits as follows:
  – The packet $\mathbf{w}_{BC}^{(1)} \oplus \mathbf{w}_{BA}$ at rate $R_B^{(1)} \leq C(\gamma_1)$
  – The packet $\mathbf{w}_{BC}^{(2)}$ at rate $R_B^{(2)} \leq C(\gamma_2)$, as only C needs to receive it.

  This simple broadcast is depicted on Fig. 3(b). Such a case can occur if during the uplink phase the rate pair (R_A, R_C) is selected to be e. g. at the point L_A.

Fig. 3 shows only two of the four possible combinations of the different options for the uplink phases and the broadcast phases. For example, also in the three–step scheme it can happen that $|\mathbf{w}_{BA}| < |\mathbf{w}_{BC}|$, such that the broadcast phase uses transmissions at two different rates.

3.3 Improved Broadcast Strategies

The simple broadcast strategy does not efficiently use the available degrees of freedom. In this section we describe two strategies that can enlarge the achievable rate region: (1) broadcast strategy based on superposition coding and (2) broadcast strategy with side information.

3.3.1 Broadcast with Superposition Coding

Superposition coding has been introduced as coding strategy that achieves the capacity region of degraded broadcast channels [17]. The Gaussian broadcast channels belong are degraded and here we explain the superposition strategy for the Gaussian case. We reuse the notations from the scenario above and consider the case in which B broadcasts to A and C. The codeword transmitted by B is:

$$\mathbf{x}_B = \sqrt{1-\theta}\mathbf{x}_{B,1} + \sqrt{\theta}\mathbf{x}_{B,2} \tag{11}$$

where θ is the power division coefficient. The codeword $\mathbf{x}_{B,1}$ should be decoded by both A and C, while the codeword $\mathbf{x}_{B,2}$ needs to be decoded only by C. When decoding $\mathbf{x}_{B,1}$, the codeword $\mathbf{x}_{B,2}$ should be treated as noise with power θ, such that its rate should satisfy:

$$R_{B,1} \leq C\left(\frac{(1-\theta)|h_{BA}|^2}{N_0+\theta|h_{BA}|^2}\right) = C\left(\frac{(1-\theta)\gamma_1}{1+\theta\gamma_1}\right) \tag{12}$$

C uses successive interference cancelation: after decoding $\mathbf{x}_{B,1}$, its contribution from the received signal is removed and $\mathbf{x}_{B,2}$ is decoded, such that its rate should satisfy:

$$R_{B,2} \leq C(\theta\gamma_2) \tag{13}$$

If B broadcasts during n_B channel uses, then A receives $n_B R_{B,1}$ bits, while C receives $n_B R_{B,1} + n_B R_{B,2}$ bits. This implies that this strategy is effective when the relay node B has more data destined for C than A, i. e. $|\mathbf{w}_{BC}| \geq |\mathbf{w}_{BA}|$. Recall the corresponding case from the simple broadcast strategy, in which we have split the broadcast data of B into two messages:

- The packet $\mathbf{w}_{BC}^{(1)} \oplus \mathbf{w}_{BA}$ is common for both A and C and is sent using the codeword $\mathbf{x}_{B,1}$, at a rate $R_{B,1}$.
- The packet $\mathbf{w}_{BC}^{(2)}$ is sent using the codeword $\mathbf{x}_{B,2}$, at a rate $R_{B,2}$.

Once the parameters of the uplink phase are fixed, one can pose the question: which θ will result in most efficient broadcast? We can answer that by setting equalities in (12-13) and solving the the balance equations for the data in the uplink and the broadcasts phase. For the three–step DF scheme the balance equations are given as:

$$n_A(R_A - R_0) = n_B(R_{B,1} + R_{B,2}) \tag{14}$$

$$n_C(R_C - R_0) = n_B R_{B,1} \tag{15}$$

The first and second balance equation is for $\mathbf{w}_{BC}$ and $\mathbf{w}_{BA}$, respectively. The only unknowns are n_B and θ, which can be found by solving the system of equations. For the JDF scheme the same equations are applied by setting $n_A = n_C$ and $R_0 = 0$.

3.3.2 Broadcast Strategy Optimized for Two–Way Relaying

In the previous broadcast strategies, the relay node "digitally" combines the data destined for the two different users by using the XOR operation. Therefore B can use codebooks and strategies that are used in the conventional broadcast scenario, where B is the source of information. On the other hand, B *is not the source* of information and each of the nodes A,C knows part of the data that B broadcasts, since it has itself sent the data in the uplink phase. This motivates to consider a different broadcast strategy, in which the codebooks at B are designed in order to account for the side information present at the nodes A,C, as introduced in [12].

Intuitive description

Let us consider a very simplified scenario in which B has two bits $\mathbf{w}_{BA} = [c_1\ c_2]$ to be transmitted to A and four bits $\mathbf{w}_{BC} = [a_1\ a_2\ a_3\ a_4]$ to be transmitted to C. Furthermore, assume that A knows $\mathbf{w}_{BC}$ and C knows $\mathbf{w}_{BA}$ and this knowledge is used as a side information in the decoding process. We also make the following simplifying assumptions. The link $B-C$ has SNR γ_2 that is sufficient for C to reliably decode 16–QAM symbols sent by B. The link $B-A$ has a lower SNR and can reliably decode only QPSK symbols, but not constellations of a higher level. The term *reliable* as used here is not precise and does not reflect the probabilistic nature of the errors, but serves well for illustrating purposes. We pose the following question:

Can B use a single 16–QAM symbol to send both $\mathbf{w}_{BA}$ *and* $\mathbf{w}_{BC}$*?*

The trick is to observe that the 16–QAM constellation consists of four shifted QPSK cosets or subconstellations, see Fig. 4. A cannot reliably decode a 16–QAM symbol, but if it has a side information to which QPSK coset does the symbol belong, it can reliably decode it. Note that a "coset QPSK" has approximately 1.0dB loss compared to the pure QPSK signal, but for the discussion here we assume that both can be received at A equally reliably. The mentioned decoding of cosets at A can be implemented as follows. The node B creates the following two–bit messages:

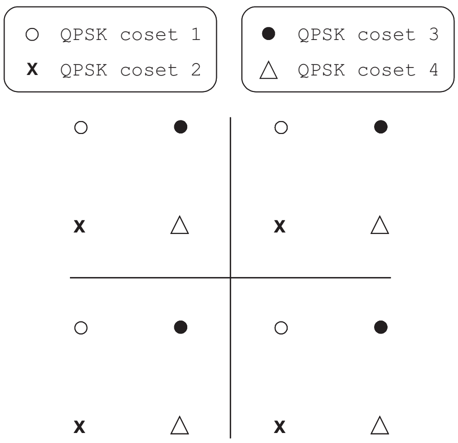


Fig. 4 Representation of 16–QAM with four QPSK subconstellations.

- The message $\mathbf{w}_{B,1} = [\,a_1\ a_2\,] \oplus [\,c_1\ c_2\,]$ and needs to be decoded both by A and C.
- The message $\mathbf{w}_{B,2} = [\,a_3\ a_4\,]$ needs to be decoded by C only.

To see how the transmitted 16–QAM symbol is selected for given $\mathbf{w}_{B,1}, \mathbf{w}_{B,2}$, we refer to Fig. 4. The node B uses $\mathbf{w}_{B,2}$ to determine which QPSK coset is being sent, while it uses $\mathbf{w}_{B,1}$ to determine the symbol within the coset. Now, since A knows $\mathbf{w}_{B,2}$ and thereby the coset a priori, it decodes the received signal by defining decision regions only for the coset of interest and ignoring the other 12 constellation points. After decoding $\mathbf{w}_{B,1}$, it retrieves the bits $[c_1 \quad c_2]$ by using XOR. On the other hand, C decodes the transmission of B by using the full 16–QAM constellation, retrieves the 2–bit packets $\mathbf{w}_{B,1}$ and $\mathbf{w}_{B,2}$ and uses XOR to extract the bits $[a_1 \quad a_2]$ from $\mathbf{w}_{B,1}$.

One may argue that we have again used digital combining with XOR. Nevertheless, here the decoding at A uses the side information in the "analog" part of the decoding, which is not the case with the usual XOR packet combining.

The result from [12] shows that a related broadcast strategy can be devised in information–theoretic sense by creating appropriate codebooks and decoding rules. Thus, the message from B can carry data to $A(C)$ at a rate $R_{BA}(R_{BC})$ where:

$$R_{BA} \leq C(\gamma_1) \qquad R_{BC} \leq C(\gamma_2) \tag{16}$$

With given data $\mathbf{w}_{BA}$ and $\mathbf{w}_{BC}$, this broadcast strategy is carried out as follows. Let us assume that B uses the maximal possible rates by putting the equality in (16). We first consider the case when

$$\frac{|\mathbf{w}_{BA}|}{C(\gamma_1)} \geq \frac{|\mathbf{w}_{BC}|}{C(\gamma_2)} \tag{17}$$

B divides $\mathbf{w}_{BA}$ into two messages $\mathbf{w}_{BA}^{(1)}, \mathbf{w}_{BA}^{(2)}$ such that:

- With $n_{B,1}$ channel uses and using the broadcast codebooks with side information it sends $\mathbf{w}_{BA}^{(1)}$ to A and $\mathbf{w}_{BC}$ to C, such that the following should be satisfied:

$$\frac{|\mathbf{w}_{BA}^{(1)}|}{C(\gamma_1)} = \frac{|\mathbf{w}_{BC}|}{C(\gamma_2)} \tag{18}$$

  form which the size of $\mathbf{w}_{BA}^{(1)}$ can be determined.
- With $n_{B,2}$ channel uses ordinary single–user codebook with rate $C(\gamma_1)$ to transmit $\mathbf{w}_{BA}^{(2)}$ to A.

From the above conditions the duration of the broadcast phase is found to be:

$$n_B = n_{B,1} + n_{B,2} = \frac{|\mathbf{w}_{BA}|}{C(\gamma_1)} \tag{19}$$

For the other case, when (17) is not satisfied, with a similar analysis it can be found that the duration of the broadcast phase is:

$$n_B = \frac{|\mathbf{w}_{BC}|}{C(\gamma_2)} \tag{20}$$

3.4 Numerical Illustration

Figures 5 and 6 compare the regions of achievable rate pairs (R_{AC}, R_{CA}) for different uplink/broadcast strategies of the schemes with decoding at the relay. Fig. 5 compares the three–step DF scheme with the two–step JDF scheme when the simple broadcast strategy is used. The SNRs of the links are chosen such as to illustrate that none of the two achievable regions is completely contained in the other. As shown in [7], the maximal two–way sum rate is achieved when $|\mathbf{w}_{BA}| = |\mathbf{w}_{BC}|$. The two strategies can be combined by time–sharing [17] in order to achieve rate pairs that are not achievable by any single strategy DF or JDF. The grey line shows the convex region of achievable rates when time–sharing is used. For example, the rate pair at point M can be achieved by using the JDF strategy with rates at the point J for a fraction δ of time, while for the remaining fraction $(1-\delta)$ of time using one–way transmission from A to C with B as helper (the rates at the point D).

Fig. 6 compares the achievable rate region for the three–step DF scheme for the three different broadcast strategies. Note that the plot starts at a value $R_{AC} =$ 1.5 in order to emphasize the differences between the strategies. As expected, the largest region is achieved by the broadcast strategy with side information. Similar

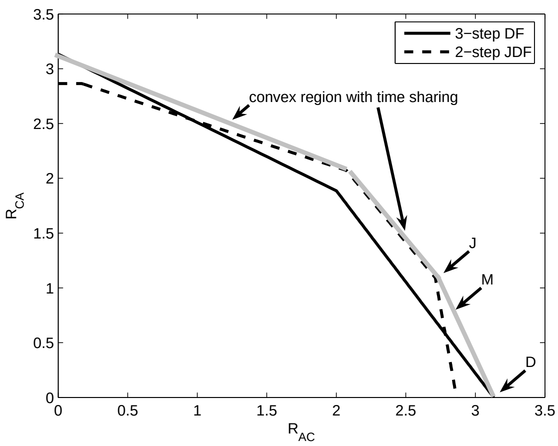


Fig. 5 Achievable rate regions with the three–step DF scheme and the two–step JDF scheme when the simple broadcast strategy is used. The parameters are $\gamma_0 = 0$ dB, $\gamma_1 = 15$ dB, $\gamma_2 = 20$ dB.

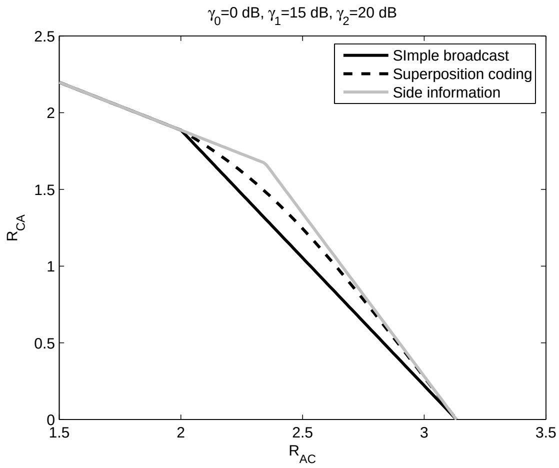


Fig. 6 Achievable rate regions with the three–step DF scheme when different broadcast strategies are used. The parameters are $\gamma_0 = 0$ dB, $\gamma_1 = 15$ dB, $\gamma_2 = 20$ dB.

results are obtained when JDF strategy is considered. In general, the benefits of the improved broadcast strategy are observable when the amount of data $|\mathbf{w}_{BC}|$ that needs to be send over the stronger link, is larger than $|\mathbf{w}_{BA}|$.

4 Two–way Relaying without Decoding at the Relay

The relay node B is not the intended destination of the data from the source and hence it is not necessary that it decodes the data. By leveraging on that observation, we can markedly increase the space of available communication strategies. In this section we discuss several techniques that do not require decoding at the relay. Note that the strategies without decoding at the relay are not novelty brought by the two–way relaying, and they have already been applied to one–way relaying, see e. g. [18]. Nevertheless, the two–way relaying scenario has distinctive features that give rise to some completely novel strategies, such as Denoise–and Forward (DNF), described further in the text.

We describe three strategies that have two–steps, as in the JDF scheme: an uplink step over multiple–access channel, such that $n_A = n_C$, and a step of n_B channel uses for broadcast. As B does not obtain the messages W_{AC}, W_{CA}, it cannot combine them digitally by using the XOR operation. Instead, the signals from A and C are inherently combined through the multiple access channel, such that the network–coding is *analog* and done at the physical layer [4] [7] [11].

4.1 Amplify–and–Forward (AF)

The simplest strategy is Amplify–and–Forward (AF) [3, 6], in which the relay B amplifies and broadcasts the symbols of the received noisy signal $\mathbf{y}_B$ from the uplink phase, such that the vector transmitted in the broadcast phase is:

$$\mathbf{x}_B = \beta \mathbf{y}_B = \beta (h_1 \mathbf{x}_A + h_2 \mathbf{x}_C + \mathbf{z}_B) \tag{21}$$

The amplification factor β is chosen as:

$$\beta = \sqrt{\frac{1}{|h_1|^2 + |h_2|^2 + N_0}} \tag{22}$$

to make the average per–symbol transmitted energy at B equal to 1. For the AF scheme $n_B = n_A = n_C$. The $m-$th symbol received by A is the amplified and additionally noised version of the $m-$th symbol received by B in step 1:

$$\begin{aligned} y_A[m] &= \beta h_1 y_B[m] + z_A[m] = \\ &= \beta h_1^2 x_A[m] + \beta h_1 h_2 x_C[m] + \beta h_1 z_B[m] + z_A[m] \end{aligned}$$

Assuming that A knows $x_A[m], h_1, h_2$ and β, then it can remove the contribution of $\beta h_1^2 x_A[m]$ from $y_A[m]$ and obtain:

$$r_A[m] = \beta h_1 h_2 x_C[m] + \beta h_1 z_B[m] + z_A[m] \tag{23}$$

which is a Gaussian channel for receiving $x_C[m]$ with SNR:

$$\gamma_{C\to A}^{(AF)} = \frac{\beta^2 |h_1|^2 |h_2|^2}{(\beta^2 |h_1|^2 + 1) N_0} = \frac{\gamma_1 \gamma_2}{2\gamma_1 + \gamma_2 + 1} \tag{24}$$

This notation denotes that $\gamma_{C\to A}^{(AF)}$ is the SNR that determines the transmission rate R_C, such that $\mathbf{x}_C$ can be successfully decoded by A. Similarly, we can find the SNR which determines the rate R_A:

$$\gamma_{A\to C}^{(AF)} = \frac{\gamma_1 \gamma_2}{\gamma_1 + 2\gamma_2 + 1} \tag{25}$$

The rate pair (R_A, R_C) used in Step 1 should satisfy:

$$R_A \leq C\left(\gamma_{A\to C}^{(AF)}\right) \qquad R_C \leq C\left(\gamma_{C\to A}^{(AF)}\right) \tag{26}$$

and the achievable rate region is determined as:

$$R_{AC} \leq \frac{1}{2} C\left(\gamma_{A\to C}^{(AF)}\right) \qquad R_C \leq \frac{1}{2} C\left(\gamma_{C\to A}^{(AF)}\right) \tag{27}$$

since the end–to–end rates are calculated over $2n_B$ channel uses.

4.2 Denoise–and–Forward (DNF)

Even if the relay does not decode the messages from A and C, it can still process the received signal $\mathbf{y}_B$ beyond mere amplification, while still not decoding the messages W_{AC}, W_{CA}. In general, we term those strategies Denoise–and–Forward (DNF) in order to emphasize that the noise is mitigated, but the signal is not decoded.

Motivating example for Denoise–and–Forward (DNF).

Assume that the multiple access channel at B is specified with $h_{AB} = h_{CB} = 1$:

$$Y_B = X_A + X_C + Z_B \tag{28}$$

Let A, C use BPSK modulation $X_A, X_C \in \{-1, 1\}$. Assume at first that there is no noise $Z_B = 0$. Then the possible signals that B can observe are $\{-2, 0, 2\}$. If the received symbol is either 2 or -2 then B can infer that the signals sent by A and C are $(X_A, X_C) = (1, 1)$ and $(X_A, X_C) = (-1, -1)$, respectively. If B receives 0, then it has ambiguity whether the signals sent are $(X_A, X_C) = (1, -1)$ or $(X_A, X_C) = (-1, 1)$. However e. g. if A sends 1 and learns that B has observed 0, then A can infer that $X_C = -1$. How much information does B need to send so that A and C can retrieve each other's symbols? One bit of information from B is sufficient: If B also uses BPSK modulation, then it can broadcast $X_B = -1$ when it observes $Y_B = -2$ or $Y_B = 2$ and it can broadcast $X_B = 1$ when $Y_B = 0$. One can easily check that, with the knowledge of X_A and X_B, A can infer X_C and, vice versa, knowing X_C and X_B, C can infer X_A.

If the channel at B is noisy, then B needs to set appropriate decision regions for the symbols, see Fig. 7. The output of the decision process is a symbol from the set $\{-2, 0, 2\}$ and in the next step B applies *denoise mapping* in order to compress the ternary symbol from the decision process to a binary symbol $\{-1, 1\}$ that needs to be sent in the broadcast phase.

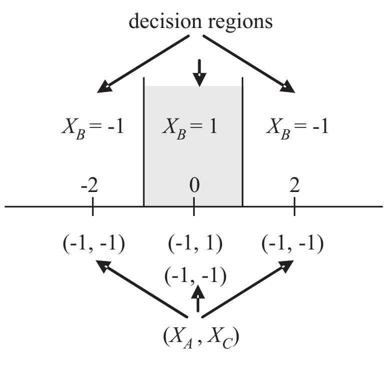


Fig. 7 Example of the decision regions and the denoise mapping when A, C use BPSK signaling and $h_{AB} = h_{CB} = 1$.

The example above outlines the main operations that need to be done by B to implement a DNF scheme: (1) decision process that quantizes the received signal and (2) mapping of the quantized signal to a message that is sent in the broadcast phase. In the example we have used *per–symbol denoising*, where the quantization and the mapping are done for each individual symbol. Such an approach is interesting when we consider non–information–theoretic analysis by considering finite–length packets and constellations, see Section 6.

For information–theoretic analysis, we need to consider *per–codeword denoising*. In that case, the decision/mapping at B is based on the received vector $\mathbf{y}_B$ during the $n_A = n_C$ channel uses of the multiple–access phase. The denoise mapping at B produces the message W_B and should be designed in such a way that by knowing W_{AC} and W_B, the node A can uniquely determine W_{CA} (analogous for C). More generally, the noisy versions of the broadcasted signal $\mathbf{x}_B$ are $\mathbf{y}_A, \mathbf{y}_C$, each being a vector of dimension n_B. After the broadcast phase, the node A should be able to decode W_{CA} from the observation $\mathbf{y}_A$ and using W_{AC} as a side information.

It is interesting to find the rate pair (R_{AC}, R_{CA}) that is on the outer bound of the achievable rate region. Assume that in the uplink phase, the transmission rates are $R_A = C(\gamma_1)$ and $R_C = C(\gamma_2)$. The relay B cannot decode W_{AC}, W_{CA} because the point $(R_A, R_C) = (C(\gamma_1), C(\gamma_2))$ is outside the capacity region of the multiple access channel with B - this is shown by simply observing that $C(\gamma_1 + \gamma_2) < C(\gamma_1) + C(\gamma_2)$. Now assume that in the broadcast phase B uses $n_B = n_A = n_C$ to transmit $\mathbf{y}_B$ noiselessly, such that A, C can observe the exact output vector $\mathbf{y}_B$. Clearly, no DNF scheme can do better than this, since from $\mathbf{y}_A$ and $\mathbf{y}_C$ can be represented as noisy copies of $\mathbf{y}_B$. If $\mathbf{x}_A$ is sent at rate R_A, then to be decodable at C it has to satisfy

$$R_A \leq I(\mathbf{x}_A; \mathbf{y}_A | W_{CA}) \stackrel{(a)}{\leq} I(\mathbf{x}_A; \mathbf{y}_B | W_{CA}) = n_A C(\gamma_1) \tag{29}$$

where (a) follows from the data processing inequality [17]. However, C decodes the signal from A after $2n_A$ channel uses, such that the end–to–end rate from A to C is $\frac{1}{2}R_A = \frac{1}{2}C(\gamma_1)$. Making analogous observations for C, we can write that a point in the outer bound to the achievable region and the two–way sum rate is given by:

$$(R_{AC}, R_{CA}) = \left(\frac{1}{2}C(\gamma_1), \frac{1}{2}C(\gamma_2) \right) \tag{30}$$

The achievability of the outer bound with structured codes is discussed in Section 5.

4.3 Compress–and–Forward (CF)

An interesting variant of the DNF schemes is using the quantization/compression framework to determine the operations carried out at the relay node B and is termed Compress–and–Forward (CF) [15]. Fig. 8 shows the data dependencies in the CF scheme.

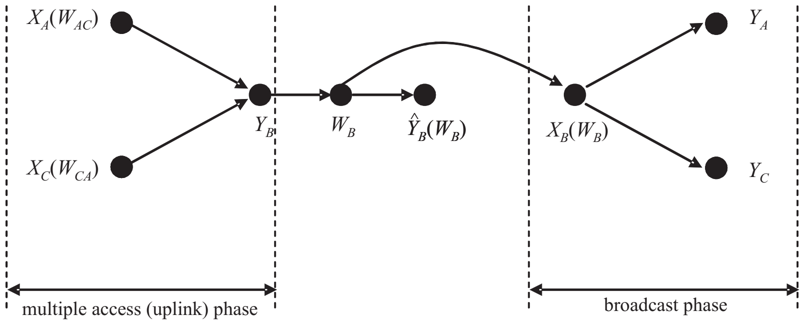


Fig. 8 Conceptual description of the Compress–and–Forward (CF) scheme.

After observing $\mathbf{y}_B$, B can obtain its quantized version $\hat{\mathbf{y}}_B$ as well as the message W_B, consisting of $n_B R_B$ bits, which is uniquely associated with $\hat{\mathbf{y}}_B$. We consider the decoding at C. The relay broadcasts $\mathbf{x}_B(W_B)$ and C receives $\mathbf{y}_C$. From the knowledge of $\mathbf{x}_C$ sent in the MA phase, C can create the set $\mathscr{W}^{MA}$ defined as:

$$\mathscr{W}^{MA} = \{W_B | (\mathbf{x}_C, \hat{\mathbf{y}}_B(W_B)) \text{isjointlytypical}\} \tag{31}$$

For formal definition of joint typicality, see [17]. Informally, we can say that from the knowledge of $\mathbf{x}_C$, the node C can create the set of candidate messages W_B that correspond to set of quantized vectors $\{\hat{\mathbf{y}}_B\}$, such that the observed signal $\mathbf{y}_B$ is likely to be "close" to one of the vectors in that set $\{\hat{\mathbf{y}}_B\}$.

Next, after observing $\mathbf{y}_C$, C creates the set $\mathscr{W}^{BC}$ defined as:

$$\mathscr{W}^{BC} = \{W_B | (\mathbf{x}_B(W_B), \mathbf{y}_C) \text{ is jointly typical}\} \tag{32}$$

which is the set of candidate messages that are likely to be sent by B in the broadcast phase. In a more conventional approach, one would select the transmission rate R_B so that W_B is decodable at C only from the observation $\mathbf{y}_B$. In that case $\mathscr{W}^{BC}$ should contain unique message W_B. The trick used here is W_B does not need to be decodable only from $\mathbf{y}_C$. Instead, W_B can be found as the unique message that lies in the intersection $\mathscr{W}^{MA} \cap \mathscr{W}^{BC}$. Having found W_B, C can find the unique message W_{AC}.

An important element of the CF scheme is the quantization performed at B. For Gaussian channels, the parameters quantization in [15] is determined by using Gaussian test channel and the achievable rate regions are obtained by numerical optimization of the parameters of the test channel.

4.4 Numerical Illustration and Variations

In this section we illustrate the achievable rate regions for the schemes that do not require decoding at the relay. Fig. 9 compares the achievable rate region for the

AF scheme with the achievable rate region of the two–step JDF scheme (where the relay decodes). The point from the outer bound is also plotted as a reference. It is interesting to see that, for the chosen values of γ_1 and γ_2, the achievable region for AF contains points that are not achievable by decoding at the relay. From the results in [15] it can be seen that the CF scheme can bring a larger region as compared to AF, but not for all configurations of γ_1, γ_2.

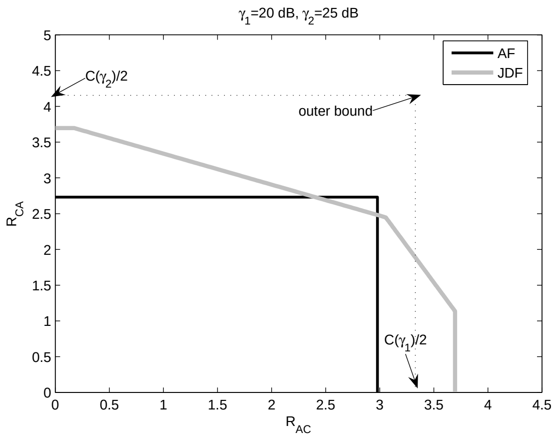


Fig. 9 Comparison of the achievable rate regions for the two–step schemes Amplify–and–Forward (AF) and Joint Decode–And–Forward (JDF). The point from the outer bound is also plotted. The parameters are $\gamma_1 = 20$ dB and $\gamma_2 = 25$ dB.

In discussing the schemes without decoding at the relay, we have focused on the two–step schemes in which the uplink phase consists of transmissions over the multiple–access channel. The three–step schemes, which leverage on the direct link between *A* and *C*, can also utilize relaying which does not require decoding at the relay. For example, consider the possible way to use AF in the three–step scheme. We can constrain $n_A = n_C = n_B$. *B* sums the n_A received symbols from the first step with the n_C symbols from the second step and amplifies the sum by respecting the transmit power constraint. Then *C* decodes by (A) combining the n_A symbols from step 1 with the n_B symbols from step 3 and (b) using its knowledge of the signal it has sent in step 2. Another generalization of the schemes described here is the Mixed Forward [15], where the relay decodes the data in one direction (e. g. from *A* to *C*), while uses compression in the other direction (*C* to *A*).

5 Achieving the Two–Way Rates with Structured Codes

Using the arguments of random coding, it has been seen that it is hard, if possible at all, to achieve the maximal two–way sum rates i .e. the outer bound point from

Fig. 9. In this section we discuss the usage of structured codes in the DNF schemes, by which B can do reliable decisions about the combination of the codewords sent by A and C without actually decoding the messages W_{AC}, W_{CA}. We first discuss the case of discrete channel, where the structured codes are parity–check codes. For the Gaussian case, the structured codes are based on lattices.

5.1 Parity–Check Codes for Binary Symmetric Channels

In this section the input/output variables X_i, Y_j, as well as the corresponding vectors $\mathbf{x}_i, \mathbf{y}_j$ are binary. The output of the multiple access channel at the relay node B is defined as:

$$Y_B = X_A \oplus X_C \oplus Z_B \tag{33}$$

In the broadcast phase, the outputs at A and C are given as:

$$Y_A = X_B \oplus Z_A \tag{34}$$

$$Y_C = X_B \oplus Z_C \tag{35}$$

The noise variables $z_i, i \in \{A, B, C\}$ have Bernoulli distribution, with error probability $P(Z_i = 1) = p$. If B knows X_C a priori, then the capacity of the channel from A to B is:

$$C_{AB|C} = 1 - H(p) = C_{CB|A} \tag{36}$$

where, analogously, $C_{CB|A}$ is the capacity from C to B, if B knows X_A a priori. Let the source nodes select the transmission rates as:

$$R_A = R_C = 1 - H(p) \tag{37}$$

We now use the coding theorem for the binary symmetric channel (BSC) for parity–check codes, see [19], Sec. 6.2. This theorem states that the channel capacity of the binary symmetric channel can be approached arbitrarily closely by using the so–called coset codes. A (n,k) coset code is defined as a code with 2^k codewords of block length $n > k$ and each message has k bits. Each message is represented by a $1 \times k$ vector $\mathbf{w}$ and the corresponding codeword $\mathbf{x} \in \{0,1\}^{[1\times n]}$ is obtained as:

$$\mathbf{x} = \mathbf{wG} \oplus \mathbf{c} \tag{38}$$

where $\mathbf{G}$ is a $k \times n$ binary matrix and c is $1 \times n$ binary vector. The theorem states that a code that is constructed with random selection of the bit entries in $\mathbf{G}$ and $\mathbf{c}$ can achieve the capacity of BSC.

Let us assume that, for a given codeword length n, the rates are selected

$$R_A = R_C = \frac{k}{n} < 1 - H(p) \tag{39}$$

and $\mathbf{G}$ and $\mathbf{c}$ are chosen according to the discussion above and are known to all the nodes. When the messages $\mathbf{w}_A$ and $\mathbf{w}_C$ are transmitted, A and C select their codewords, respectively, as follows:

$$\mathbf{x}_A = \mathbf{w}_A\mathbf{G} \oplus \mathbf{c} \qquad \mathbf{x}_C = \mathbf{w}_C\mathbf{G} \tag{40}$$

The received vector at B is:

$$\mathbf{y}_B = \mathbf{x}_A \oplus \mathbf{x}_C \oplus \mathbf{z}_B = (\mathbf{w}_A \oplus \mathbf{w}_C)\mathbf{G} \oplus \mathbf{c} \tag{41}$$

Conceptually, B can treat the received vector as an output of a BSC, over which a *virtual* transmitter sends the message $\mathbf{w}' = \mathbf{w}_A \oplus \mathbf{w}_C$ by using the same coset code as used by A. Then B tries to reliably decode the message $\mathbf{w}'$, rather than the individual messages $\mathbf{w}_A$ and $\mathbf{w}_C$. $\mathbf{m}'$ is a message of k bits that, by assumption, can be reliably decoded over the BSC. In the broadcast phase, B sends the codeword $\mathbf{x}' = \mathbf{w}'\mathbf{G} \oplus \mathbf{c}$, which can be reliably decoded by both A and C.

We can conclude that when the two–way relaying system features binary symmetric channels, the rate pair $(R_{AC}, R_{CA}) = (1 - H(p), 1 - H(p))$ of the outer bound point is achievable.

5.2 Gaussian Channel

At the moment we do not have a rigorous proof that the outer bound point is achievable in Gaussian channels and the discussion in this section should be taken as a possible way forward toward proving such achievability. Note that several works [10] have considered the usage of lattice codes for two–way relaying in Gaussian channels, but there is no known technique that certainly achieves the outer bound point.

In order to simplify the discussion, in this section we assume that the wireless channels are *real* Gaussian channels, such that the received signals are:

$$Y_B = X_A + X_C + Z_B \tag{42}$$

$$Y_A = X_B + Z_A \tag{43}$$

$$Y_C = X_B + Z_C \tag{44}$$

where for the transmitter i the average power is $E[X_i^2] \leq 1$ and the noise power is $E[Z_i^2] = \sigma^2$. We slightly abuse the notation used so far by defining:

$$C(\gamma_1) = \frac{1}{2}\log_2(1 + \gamma_1) \tag{45}$$

to correspond to the capacity of a real Gaussian channel with SNR γ_1. We assume that the transmission rates of A and C are chosen:

$$R_A = R(C) = C(\gamma_1) = C\left(\frac{1}{\sigma^2}\right) \tag{46}$$

A lattice Λ is discrete subgroup of the Euclidean space $\mathbb{R}^n$ with the ordinary vector addition operation. If $\lambda_1, \lambda_2 \in \Lambda$, then $(\lambda_1 - \lambda_2) \in \Lambda$ and $(\lambda_1 + \lambda_2) \in \Lambda$. The lattices introduce algebraic structure for the codewords in the continuous channel. For detailed discussion on the lattices and their usage in achieving the capacity of Gaussian channel, the reader is directed to [20] and the references therein.

For our discussion, we use $n = n_A = n_B$, such that the transmitted codewords belong to $\mathbb{R}^n$. We use Construction A to generate the lattice, as described in [20]. These are the steps to generate the lattices for A and C:

1. Pick a prime number p, such that the rate R and the number of channel uses n satisfy:
$$p = 2^{nR} \tag{47}$$
2. Draw a generating vector $\mathbf{g} = (g_1, g_2, \dots g_n)$, where g_i is an integer uniformly picked from the set $\{0, 1, 2, \dots, p-1\}$.
3. Define the discrete codebook
$$\mathscr{C} = \{\mathbf{c} \in \mathbb{Z}_p^n : \mathbf{c} = (\mathbf{g} \cdot q)\text{mod}p \quad q = 0, 1, \dots, p-1\} \tag{48}$$
4. The node A selects a random vector $\mathbf{u}_A = \{u_{A1}, u_{A2}, \dots u_{An}\}$, where each u_{Ai} is independently generated as a real number uniformly distributed in $[0, p)$. The codewords transmitted by A are obtained as follows: For each $c \in \mathscr{C}$, the corresponding codeword $\mathbf{x}_A(\mathbf{c})$ is obtained as:
$$\mathbf{x}_A(\mathbf{c}) = \beta(p^{-1}[\mathbf{c} + \mathbf{u}_A]_p - 0.5) \tag{49}$$
where the notation $[v]_p$ denotes the per–component modulo$-p$ operation:
$$[v]_p = \begin{cases} v & \text{if } v < p \\ v - p & \text{otherwise} \end{cases} \tag{50}$$
and the coefficient β is a normalization coefficient that ensures that the average power of the transmitted codewords is P. Using the crypto–lemma [21](or see Lemma 1 in [20]), one can show that the $i-$th component of each vector is a uniformly distributed real number in the interval $[-\frac{\beta}{2}, \frac{\beta}{2})$. Hence, in this case β should be chosen to be $\frac{\beta^2}{12} = P$ in order to match the power constraint.
5. The node C uses identical procedure as A, except that it generates an independent vector $\mathbf{u}_C$.

Let us consider the properties of the sum of such generated codewords. Each pair of codewords $(\mathbf{x}_A, \mathbf{x}_C)$ can be uniquely associated with the pair (q_A, q_C) where $q_A, q_C \in \{0, 1, \dots, p-1\}$ are the integers used in generating the correspondent $(\mathbf{c}_A, \mathbf{c}_C)$ in (48), such that we can use the notation $(\mathbf{x}_A(q_A), \mathbf{x}_C(q_C))$ or simply (q_A, q_C) to denote a pair of codewords. The sum of the codeword pair (q_A, q_C) is:

$$\begin{aligned} s_i(q_A, q_C) &= \beta(p^{-1}[[g_i q_A]_p + u_{Ai}]_p - 0.5 + p^{-1}[[g_i q_C]_p + u_{Ci}]_p - 0.5) \\ &\overset{(a)}{=} \beta p^{-1}(([g_i q_A + u_{Ai}]_p + [g_i q_C + u_{Ci}]_p) - p) \end{aligned} \tag{51}$$

where (a) follows from the distributive property of the modulo operation. The important component of $s_i(q_A, q_C)$ is:

$$\tau_i(q_A, q_C) = [g_i q_A + u_{Ai}]_p + [g_i q_C + u_{Ci}]_p \tag{52}$$

which can be represented as follows:

$$\tau_i(q_A, q_C) = [g_i[q_A + q_C]_p + u_{Ai} + u_{Ci}]_p + \delta_i p \tag{53}$$
$$= t_i + \delta_i p \tag{54}$$

where

$$\delta_i = \begin{cases} 0 & \text{if} \quad \tau_i(q_A, q_C) < p \\ 1 & \text{otherwise} \end{cases} \tag{55}$$

On the other hand, using the crypto–lemma, one can infer that t_i is a real number that is uniformly distributed in $[0, p)$. Hence, the $i-$th component of sum of the pair of codewords (q_A, q_C) can be represented by:

$$s_i(q_A, q_C) = \beta p^{-1}(t_i + \delta_i p - p) = \beta(p^{-1}t_i + \delta_i - 1) = \beta(p^{-1}t_i - 0.5 + \delta_i - 0.5) \tag{56}$$

where $p^{-1}t_i - 0.5$ is a random variable that is uniformly distributed in $[-0.5, 0.5)$ and $\delta_i - 0.5$ is a binary random variable that can have a value of either -0.5 or 0.5. The distribution of δ_i is dependent on t_i and can be determined by observing that the distribution of $p^{-1}t_i + \delta_i - 1$ is equal to the distribution obtained from a sum of two random numbers uniformly distributed in $[-0.5, 0.5]$, which implies:

$$P(\delta_i = -1|t_i) = \frac{t_i}{p} \qquad P(\delta_i = 1|t_i) = 1 - \frac{t_i}{p} \tag{57}$$

In the sequel we will need the following definition:

Definition 1. The codeword pair $(\mathbf{x}_A(q_A), \mathbf{x}_C(q_C))$ belongs to the class m if:

$$[q_A + q_C]_p = (q_A + q_C)\text{mod}p = m \tag{58}$$

There are p different classes, each containing p codeword pairs. If two codeword pairs (q_{A1}, q_{C1}) and (q_{A2}, q_{C2}) belong to the same class, then one can easily show that for each component i:

$$t_i(q_{A1}, q_{C1}) = t_i(q_{A2}, q_{C2}) \tag{59}$$

such that $s_i(q_{A1}, q_{C1})$ *and* $s_i(q_{A2}, q_{C2})$ *are either equal or their difference is exactly* $\frac{\beta}{2}$.

The relay B receives:

$$\mathbf{y}_B = \mathbf{s}(q_A, q_C) + \mathbf{z}_B \tag{60}$$

If from $\mathbf{y}_B$ B can determine to which of the $p = 2^{nR}$ classes of codewords does (q_A, q_C) belong, then in the broadcast phase B can use “conventional” random codebooks to send the nR bits that describe the class of the pair of observed codewords.

If A knows the class, then it can uniquely determine q_C and thereby W_{AC}. The same is valid for C and in such case we would have achieved the point of the outer bound.

Having introduced this framework, we can think that the communication from A,C to B is done between a virtual transmitter V and B. The virtual transmitter generates p codewords $\{\mathbf{x}_V\}$ where the $i-$th component of each codeword is uniformly and independently distributed in $\left(-\frac{\beta}{2}, \frac{\beta}{2}\right)$. Before transmitting the signal to B, V generates *binary self–noise* at each component of $\{\mathbf{x}_V\}$: for the $i-$th component, V looks at x_{Vi} and uses the appropriate distribution for δ_i, see (56), to determine whether to add or subtract $\frac{\beta}{2}$.

In the absence of the binary self–noise, the codewords of $\{\mathbf{x}_V\}$ can be reliably decoded at B, which follows from the manner in which the codebook is generated and the noisy channel coding theorem for Gaussian channels. *We conjecture that there are codebooks $\{\mathbf{x}_V\}$ such that B can reliably decode $\mathbf{x}_V$ even in the presence of the binary self-noise, but the rigorous proof for this claim is not available at this time and is a subject of ongoing work.*

6 Signalling Constellations for Finite Packet Lengths

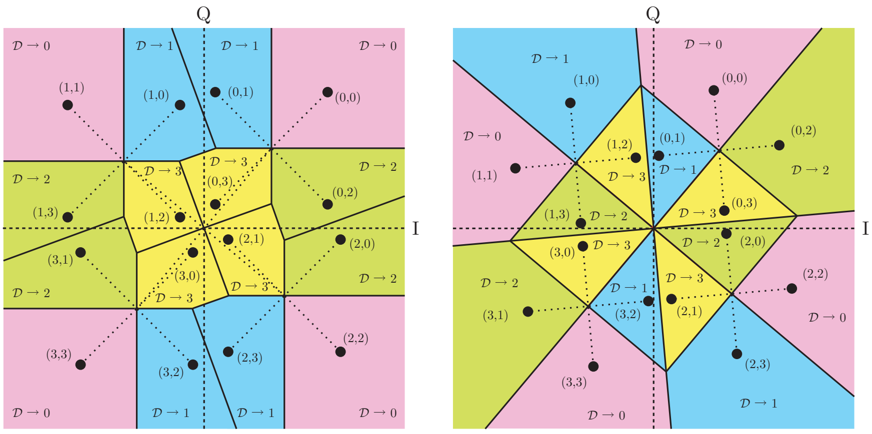
The discussion hitherto has been mostly focused on information–theoretic aspects of the bidirectional relaying, which assumed arbitrarily long codewords/packets and error probability that is asymptotically zero. In this section, we focus on the practical aspects of the two–way relaying by considering finite–length packets and finite constellations. We restrict ourselves to the most interesting case, namely the two–step DNF schemes that do not require decoding at the relay. The schemes use with per–symbol denoising, rather than per–codeword one, and we address the following questions: How should the relaying node B generate a denoised signal for the broadcast phase by observing the received signal at the uplink phase? How many constellation points are required to reliably forward two distinct messages to the destination nodes A and C? Considering the case in which both terminals use QPSK constellations in the uplink phase, we present one example of signalling strategies for reliable bidirectional relaying. The ensuing discussions tell us two interesting results: (1) we should use multiple network coding rules adaptively optimized according to the channel state information at the relay B and (2) some specific channel conditions necessitate the use of unconventional 5–ary signal constellations, rather than the QPSK.

6.1 XOR Denoising

When A,C are using QPSK modulation, the combined signal $h_1 X_A + h_2 X_C$ can have up to 16 possible values, depending on the channel coefficients h_1, h_2. As it has been noted in Section 4.2, the relay needs to compress the observed signal Y_B before forwarding it in the broadcast phase. Letting $\mathscr{M}(\cdot)$ be the QPSK signal mapper, the 2–bit digital data of the ML estimate $\hat{X}_i$ can be written as $\hat{\mathbf{w}}_i = \mathscr{M}^{-1}(\hat{X}_i) \in \{0,1,2,3\}$. In the DNF scheme that uses XOR–based network coding, the relay B creates the QPSK signal for the broadcast phase in the following way. It maps the received signal within the ML region for $(\hat{\mathbf{w}}_A, \hat{\mathbf{w}}_C)$ into the QPSK constellation as $X_B = \mathscr{M}(\mathbf{w}_B)$ where $\mathbf{w}_B = \mathscr{D}(\hat{\mathbf{w}}_A, \hat{\mathbf{w}}_C) = \hat{\mathbf{w}}_A \oplus \hat{\mathbf{w}}_C$. The mapper $\mathscr{D}(\cdot)$ denotes the digital denoising function. Since a terminal node knows its own information, it can detect the desired data from the denoising signal. The XOR denoising works very well for some channel conditions, e. g., $\phi \simeq 0$ where we define

$$h_2/h_1 = \rho \exp(\mathrm{j}\phi) \tag{61}$$

Fig. 10(a) shows an example of the received signal constellation Y_B at the uplink phase for $\rho \simeq 1$ and $\phi \simeq 0$. The four signal points at the center may be unreliable in the ML estimation because the distance between the points are short. This produces unreliable relaying when we adopt the JDF scheme at the relay. Meanwhile, as the DNF scheme maps such closest neighboring points into the same denoising signal, it offers a significant improvement in achievable throughput performance.



(a) $h_2/h_1 \simeq 1$ (b) $h_2/h_1 \simeq (1+j)/2$

Fig. 10 Examples of received signal constellations at the relay B and XOR denoising.

6.2 *Adaptive Denoising with Quintary Cardinality*

The denoising function $\mathscr{D}(\cdot)$ should be adaptively changed as a function of the channel state information h_1 and h_2 because the ML region highly depends on the channel condition. The XOR denoising function $\mathscr{D}(\hat{w}_A, \hat{w}_C) = \hat{w}_A \oplus \hat{w}_C$ does not work well for some specific channel conditions as illustrated in Fig. 10(b). This figure shows an example of the received signal y_B for $\rho \simeq 1/\sqrt{2}$ and $\phi = \pi/4$. As this figure shows, all the closest neighbor points are mapped into different denoising points. In particular, the distance between the ML points $(\hat{w}_A, \hat{w}_C) = (0,1)$ and $(1,2)$ is very short, which leads to an unreliable relaying performance.

It is very interesting to observe that for this case there does not exist a denoising function with 4–ary cardinality which can map all the closest neighbors into the same denoising point. The minimum achievable cardinality of the denoising function to avoid the distance shortening is 5, which implies that we need to use 5–ary modulation in the broadcast phase. For these channel conditions, the coding rule is given as:

$$\mathscr{D}(\hat{w}_A, \hat{w}_C) = \begin{cases} 0 & \text{for } (\hat{W}_A, \hat{W}_C) \in \{(0,0),(1,1),(2,2),(3,3)\} \\ 1 & \text{for } (\hat{W}_A, \hat{W}_C) \in \{(0,3),(2,0),(3,1)\} \\ 2 & \text{for } (\hat{W}_A, \hat{W}_C) \in \{(0,1),(1,2),(2,3)\} \\ 3 & \text{for } (\hat{W}_A, \hat{W}_C) \in \{(3,2),(2,1),(1,0)\} \\ 4 & \text{for } (\hat{W}_A, \hat{W}_C) \in \{(1,3),(3,0),(0,2)\} \end{cases} \tag{62}$$

The above denoising function performs well for $h_2/h_1 \simeq (1+\mathrm{j})/2$ or $(1-\mathrm{j})$. For the other channel conditions, we require three more 5–ary denoising functions and one more 4–ary denoising function to avoid all the possible distance shortening; in total there are six coding rules. For the 5–ary denoising, we should use some kind of 5–QAM signalling for broadcasting. Although it exhibits a slight loss in the Euclidean distances as compared to the QPSK modulation, the adaptive use of six denoising functions can bring a substantial benefit, as the results in the following section show.

6.3 End–to–End Throughput Performance

In Fig. 11, we show the performance comparisons in end–to–end throughput as a function of average SNR under Nakagami–Rice fading channels for a Ricean factor of 10 dB. We assume $E[|h_1|^2] = E[|h_2|^2]$ for simplicity. In this figure, we plot the curves for the conventional 4–step relaying scheme, the 3–step DF scheme with XOR network coding, and the 2–step DNF scheme with XOR denoising and adaptive denoising. The direct link between A and C is assumed to have SNR of 0. Due to the time efficiency, the 3–step network coding is superior to the 4–step protocol with an improvement of 33%, as discussed before. The 2–step DNF scheme further improves the throughput by a maximum of 100%. The XOR denoising can offer an excellent throughput performance if we can use the precoding technique for phase–synchronization to achieve $\phi = 0$. However, it seems technically infeasible to obtain an accurate phase synchronization among the distributed terminals. Without such a precoding, the XOR denoising suffers from a serious performance degradation because of the distance shortening occurred in several channel condi-

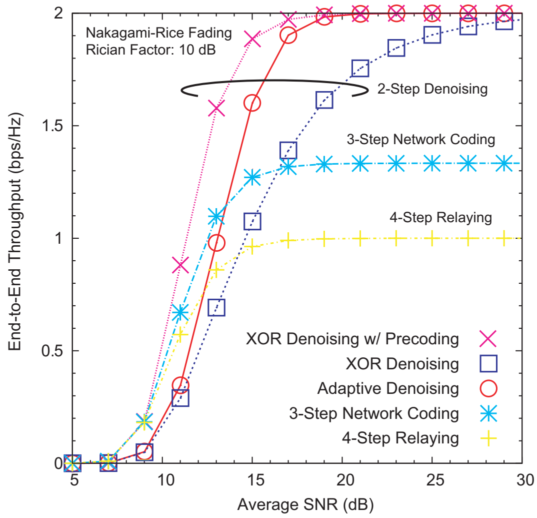


Fig. 11 End–to–end throughput performance as a function of average SNR in Nakagami–Rice fading channels for 10 dB Ricean factor.

tions. This degradation can be significantly compensated by using adaptive network coding which allows the use of 5–ary modulations as well as 4–ary one.

7 Conclusions and the Way Forward

The shared wireless medium fosters cooperative (relay–aided) communication modes among the proximate nodes. The two–way relaying scenarios lift such cooperative techniques to the next level, by leveraging on the recent ideas of network coding. We have shown that even the simple 3–node scenario with two–way relaying is a fertile ground for devising novel communication strategies. For a class of two–way relaying techniques, the relaying node decodes the messages before broadcasting them back to the terminals. For that class of techniques, the novel ideas are featured in the broadcast methods, since the relay has a large freedom in combining the two received messages. We have also elaborated on the techniques that do not require decoding at the relay, where the innovative techniques are combining several operations at the relay: detection, quantization, mapping and recoding. In the information–theoretic setting, rather than using solely random codebooks, we have shown that the structured codes, such as the lattices, can have great utility if the does not decode but only denoises the received signals. Besides the information–theoretic discussion, we have provided insights into the design of the two–way relaying schemes with practical modulation constellations.

The techniques for coded bidirectional relaying represent a prime example of the large freedom that a wireless network designer has in devising novel communication modes. This is in particular visible if we compare them with the protocol designs that are aligned to the layered protocol structure. The presented communication schemes somehow violate the layering, since: (a) the intermediate nodes (network layer) do not decode the information before forwarding it further, and (b) the adopted network–coding approach observes more than one flow simultaneously, while the layered architecture is single–flow–oriented. The approaches for coded bidirectional relaying leave largely open venues for future work: code designs, channel estimation, scheduling, retransmission protocols, etc. To name one more general topic - despite the apparent performance advantages in the simple 3–node scenario, it is not straightforward to see how the techniques without decoding at the relay can scale to networks with may nodes.

References

1. Larsson, P., Johansson, N., and Sunell, K.-E.: Coded Bi–directional Relaying, *In Proc. 5th Scandinavian Workshop on Ad Hoc Networks (ADHOC'05)*, Stockholm, Sweden, May 2005.
2. Katti, S., Katabi, D., Hu, W., Rahul, H., and Médard, M.: The Importance of Being Opportunistic: Practical Network Coding for Wireless Environments, *In Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing)*, Sep. 2005.
3. Popovski, P. and Yomo, H.: Bi-directional Amplification of Throughput in a Wireless Multi-Hop Network, *In Proc. IEEE 63rd Vehicular Technology Conference (VTC)*, Melbourne, Australia, May 2006.
4. Popovski, P. and Yomo, H.: The Anti–Packets can Increase the Achievable Throughput of a Wireless Multi–Hop Network, *In Proc. IEEE International Conference on Communications (ICC)*, Istanbul, Turkey, Jun. 2006.
5. Hausl, C. and Hagenauer, J.: Iterative Network and Channel Decoding for the Two-Way Relay Channel, *In Proc. IEEE International Conference on Communications (ICC)*, Istanbul, Turkey, Jun. 2006.
6. Rankov, B. and Wittneben, A.: Spectral efficient protocols for half-duplex relay channels, *IEEE J. Select. Areas Commun.*, vol. 25, pp. 379389, Feb. 2007.
7. Popovski, P. and Yomo, H.: Physical Network Coding in TwoWay Wireless Relay Channels, *In Proc. IEEE International Conference on Communications (ICC)*, Glasgow, Scotland, Jun. 2007.
8. Yomo, H. and Popovski, P.: Opportunistic Scheduling for Wireless Network Coding, *In Proc. IEEE International Conference on Communications (ICC)*, Glasgow, Scotland, Jun. 2007.
9. Kim, S. J., Mitran, P. and Tarokh, V.: Performance bounds for bi-directional coded cooperation protocols, *To appear in IEEE Trans. Inform. Theory*, http://arxiv.org/abs/cs/0703017.
10. Narayanan, K., Wilson, M. P., and Sprintson, A.: Joint Physical Layer Coding and Network Coding for Bi-Directional Relaying, *In Proc. 2007 Allerton Conference on Communication, Control and Computing*, Monticello, IL, USA, 2007.
11. Katti, S., Gollakota, S., and Katabi, D.: Embracing Wireless Interference: Analog Network Coding, *In Proc. ACM SIGCOMM*, 2007.
12. Oechtering, T. J., Schnurr, C., Bjelakovic, I., and Boche, H.: Broadcast Capacity Region of Two-Phase Bidirectional Relaying, *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 454458, Jan. 2008.
13. Mitran, P.: The Diversity-Multiplexing Tradeoff for Independent Parallel MIMO Channels, *In Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Toronto, Canada, Jul. 2008.
14. Koike-Akino, T., Popovski, P., and Tarokh, V.: Denoising Maps and Constellations for Wireless Network Coding in TwoWay Relaying Systems, *In Proc. IEEE Globecom*, 2008, to appear.

15. Kim, S. J., Devroye, N., Mitran, P. and Tarokh, V.: Achievable Rate Regions for Bi–directional Relaying, http://arxiv.org/abs/0808.0954.
16. Shannon, C. E: Two–Way Communications Channels,*In Proc. 4th Berkeley Symp. Math. Stat. Prob.*, Chicago, IL, Jun. 1961.
17. Cover, T. M., and Thomas, J.: Elements of Information Theory, John Wiley & Sons Inc., 1991.
18. Laneman, J. N., Tse, D. N. C., and Wornell, G. W.: Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior, *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.
19. Gallager, R.: Information Theory and Reliable Communication, New York: Wiley, 1968.
20. Erez, U. and Zamir, R.: Achieving $(1/2)\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding, *IEEE Transactions on Information Theory*, vol. 50, pp. 2293-2314, Oct. 2004.
21. Forney Jr., G. D.: On the Role of MMSE Estimation in Approaching the Information-Theoretic Limits of Linear Gaussian Channels: Shannon meets Wiener, *In Proc. 2003 Allerton Conference on Communication, Control and Computing*, Monticello, IL, USA, 2003.