



KTH Electrical Engineering

# Polar Codes for Cooperative Relaying

©2012 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

RICARDO BLASCO-SERRANO, RAGNAR THOBABEN, MATTIAS ANDERSSON,  
VISHWAMBHAR RATHI, AND MIKAEL SKOGLUND

Stockholm 2012

---

Communication Theory Department  
School of Electrical Engineering  
KTH Royal Institute of Technology

# Polar Codes for Cooperative Relaying

Ricardo Blasco-Serrano, *Student Member, IEEE*, Ragnar Thobaben, *Member, IEEE*,  
Mattias Andersson, *Student Member, IEEE*, Vishwambhar Rathi, and Mikael Skoglund, *Senior Member, IEEE*

**Abstract**—We consider the symmetric discrete memoryless relay channel with orthogonal receiver components and show that polar codes are suitable for decode-and-forward and compress-and-forward relaying. In the first case we prove that polar codes are capacity achieving for the physically degraded relay channel; for stochastically degraded relay channels our construction provides an achievable rate. In the second case we construct sequences of polar codes that achieve the compress-and-forward rate by nesting polar codes for source compression into polar codes for channel coding. In both cases our constructions inherit most of the properties of polar codes. In particular, the encoding and decoding algorithms and the bound on the block error probability  $O(2^{-N^\beta})$  which holds for any  $0 < \beta < \frac{1}{2}$ .

## I. INTRODUCTION

The relay channel, introduced by van der Meulen in [1], is an information theoretical model for cooperative communication in which a source wants to convey reliably a message to a destination with the help of a third node known as the relay. To determine the capacity of the relay channel in general is still an open problem. In [2] Cover and El Gamal established an outer bound on the capacity which is known as the cut-set bound and two coding strategies based on two different philosophies of information processing at the relay: decode-and-forward (DF) and compress-and-forward (CF). In DF the relay recovers the message transmitted by the source and forwards some information about it to the destination that complements the observation obtained through the source-destination link. In contrast, in CF the relay describes its raw channel observation. Since the destination has some side information (i.e. its own observation) this approach is connected to the Slepian-Wolf [3] and Wyner-Ziv [4] problems. Neither of these strategies (nor the combination of both [2]) is capacity achieving in general. Moreover, neither DF nor CF outperform the other in all the scenarios; as a rule of thumb DF performs better when the source-relay channel is good while CF is better when the quality of this channel is low [5].

In the last decade there have been many research efforts to implement DF relaying in practice. The work has mainly focused on adapting capacity-approaching/achieving codes from

the point-to-point channel to the relay channel, e.g. distributed turbo codes [6], distributed serially concatenated codes [7], or low-density parity-check codes [8]. Although in general their optimality cannot be proved, many of these solutions have been shown empirically to perform remarkably well. In contrast, the amount of practical implementations of CF relaying is much smaller. They are mostly based on realizing Slepian-Wolf and Wyner-Ziv coding [9], [10]. Interestingly, the achievability of the Wyner-Ziv bound with codes defined on sparse graphs under optimal encoding and decoding was shown in [11].

Recently, Arıkan introduced the phenomenon of channel polarization and its application to construct capacity-achieving codes which are known as polar codes (PCs) [12], [13]. This has led to a breakthrough in terms of achievability results in information theory with structured codes (as opposed to random coding). For example, the optimality of PCs for source compression was established by Korada and Urbanke in [14], [15] for the case of binary reproduction alphabets (later extended in [16]). These results were obtained by considering the duality between channel and source coding. A different approach was taken by Arıkan in [17] to show that sources also polarize and that this phenomenon is useful for designing codes for lossless compression. PCs were first applied to multi-terminal problems in [14], where it was shown that they are optimal for binary Slepian-Wolf (see also [17]) and Wyner-Ziv coding, among others. To our knowledge, the first application of PCs to the relay channel was reported in [18]. There it was shown that PCs achieve the capacity of symmetric binary-input physically degraded relay channels. Construction and performance aspects of PCs were considered in [19], [20] and a design method based on density evolution [21] was proposed in [22].

The contributions of this paper are the following: first, we extend the achievability results from [18] on PCs for DF relaying in stochastically degraded binary symmetric relay channels with orthogonal receivers to arbitrary discrete alphabets. Second, we show that PCs are also suitable for CF relaying. Then we specialize this result to two cases of interest: CF relaying based on Slepian-Wolf coding which is capacity-achieving in special cases, and to channels with a special structure where the performance is independent of the choice of frozen symbols at the relay. Finally, we present the first numerical results on the performance of PCs with finite block lengths for both DF and CF relaying.

This paper is organized as follows: In Section II we review the background on PCs and the relay channel and introduce the scenario along with the notation. In Section III we state our main contributions in the form of two theorems. In Section IV we review a few properties of PCs for degraded

This work was supported in part by the European Community's Seventh Framework Programme under grant agreement no 216076 FP7 (SENDORA), the Swedish Research Council, and VINNOVA.

Parts of the material in this paper were presented at the 44th Asilomar Conference on Signals, Systems, and Computers, 2010.

R. Blasco-Serrano, R. Thobaben, M. Andersson, and M. Skoglund are with Communication Theory Laboratory, School of Electrical Engineering and ACCESS Linnaeus Centre, KTH Royal Institute of Technology, SE-10044, Stockholm, Sweden (e-mail: ricardo.blasco@ee.kth.se; ragnar.thobaben@ee.kth.se; mattias.andersson@ee.kth.se; mikael.skoglund@ee.kth.se).

V. Rathi was with Communication Theory Laboratory, School of Electrical Engineering and ACCESS Linnaeus Centre, KTH Royal Institute of Technology, SE-10044, Stockholm, Sweden. He is now with Nvidia Corporation, Bristol, United Kingdom (e-mail: vrathi@gmail.com).

channels. These properties are used to establish the proofs of the two theorems in Sections V and VI. We evaluate the performance of our constructions for DF and CF relaying for finite block lengths in Section VII using simulation results. Finally, Section VIII concludes the paper.

## II. NOTATION, BACKGROUND, AND SCENARIO

### A. Notation

Random variables and their realizations are represented using upper case and lower case letters  $X$  and  $x$ , respectively. Vectors are represented using bold face  $\mathbf{x}$  and the  $i^{\text{th}}$  component of  $\mathbf{x}$  is denoted by  $x_i$ . For a vector  $\mathbf{x}$  we write  $\mathbf{x}_i^j$  as shorthand for  $(x_i, \dots, x_j)$  (void if  $j < i$ ). More generally, for a set  $F = \{f_0, \dots, f_{|F|-1}\}$  with cardinality  $|F|$ ,  $\mathbf{x}_F$  denotes the sub-vector  $(x_{f_0}, \dots, x_{f_{|F|-1}})$ . An alphabet is represented with a calligraphic letter  $\mathcal{X}$ . We shall assume that together with the addition  $\oplus$ , the alphabet forms an Abelian group  $(\mathcal{X}, \oplus)$ . Without loss of generality we label the elements in  $\mathcal{X}$  as  $\{0, 1, \dots, |\mathcal{X}| - 1\}$ . We denote the inverse (with respect to  $\oplus$ ) of an element  $x \in \mathcal{X}$  by  $-x$ ; that is,  $x \oplus (-x) = 0$ , where 0 is the identity element. For vectors  $\oplus$  works element-wise.

We follow the standard notation to denote entropy  $H(U)$ , mutual information  $I(U; V)$ , and conditional mutual information  $I(U; V|T)$ . In addition, for a given set of conditional probabilities  $W(v|u)$ ,  $I(W)$  is shorthand for the mutual information  $I(U; V)$  when  $U$  is uniformly distributed. Information is measured in  $q$ -ary units. We use the Landau notation  $O(N)$  to denote the asymptotic behavior of functions.

### B. Polar Codes for Channel and Source Coding

*Channel coding:* Let  $W(y|x)$  be a  $q$ -ary input discrete memoryless channel<sup>1</sup>(DMC) with *symmetric capacity*  $I(W)$  (i.e.  $I(X; Y)$  when  $X$  is uniformly distributed). Channel polarization is a method for constructing codes based on the recursive application of a simple linear transformation (expressed by the invertible  $N \times N$  matrix  $\mathbf{G}_N$ ) to  $N$  independent copies of  $W$  to synthesize a set of  $N$   $q$ -ary channels with extreme properties: as  $N$  grows the synthetic channels become, except for a vanishing fraction, either noiseless or pure noise. Moreover, the fraction of noiseless channels tends to  $I(W)$ . PCs are based on this phenomenon: information is sent at full rate through the noiseless channels (known as the *information set of channels*,  $F_C^c$ ) and a known sequence of symbols is sent through the noisy ones (*frozen set*,  $F_C$ ). A PC is simply defined by its frozen set  $F_C$  and the values of the fixed (frozen) symbols  $\mathbf{u}_{F_C}$ .

We consider  $N = 2^n$  ( $n \in \mathbb{N}$ ) independent uses (copies) of  $W$ . That is, the transmission channel is<sup>2</sup>

$$W^N(\mathbf{y}|\mathbf{u}\mathbf{G}_N) = W^N(\mathbf{y}|\mathbf{x}) = \prod_{i=0}^{N-1} W(y_i|x_i) \quad (1)$$

<sup>1</sup>Here we consider only  $q$  prime. In addition, we disregard the uninteresting cases where  $I(W) \in \{0, 1\}$ .

<sup>2</sup>We follow the standard notation for PCs and use  $W^N(\mathbf{y}|\mathbf{x})$  to denote  $N$  uses of the channel  $W : \mathcal{X} \rightarrow \mathcal{Y}$ . Similarly  $W_N(\mathbf{y}|\mathbf{u})$  denotes  $N$  uses of the channel induced by  $\mathbf{x} = \mathbf{u}\mathbf{G}_N$ . Finally, the  $i^{\text{th}}$  synthetic channel ( $i \in \{0, 1, \dots, N-1\}$ ) is denoted by  $W_N^{(i)}$ .

where  $\mathbf{u}$  contains the  $q$ -ary frozen (i.e. fixed and known) and information symbols and  $\mathbf{x} = \mathbf{u}\mathbf{G}_N$  is the codeword put into the channel. This induces the conditional probability mass function (pmf)

$$W_N(\mathbf{y}|\mathbf{u}) = \prod_{i=0}^{N-1} W(y_i|(\mathbf{u}\mathbf{G}_N)_i) \quad (2)$$

and the distribution of the  $i^{\text{th}}$  synthetic channel<sup>3</sup> ( $i \in \{0, 1, \dots, N-1\}$ ), obtained from (2) as

$$W_N^{(i)}(\mathbf{y}, \mathbf{u}_0^{i-1}|u_i) = \frac{1}{q^{N-1}} \sum_{\mathbf{u}_{i+1}^{N-1}} W_N(\mathbf{y}|\mathbf{u}).$$

In order to decode PCs Arkan introduced a simple Successive Cancellation (SC) algorithm that generates hard estimates  $\hat{u}_i$  of the information symbols  $u_i$  sequentially (with increasing  $i$ ) using the previous estimates  $\hat{\mathbf{u}}_0^{i-1}$  by considering  $W_N^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|u_i)$ . PCs decoded with the SC algorithm can be used to transmit at any rate  $R < I(W)$  with error probability  $\Pr(\hat{\mathbf{U}} \neq \mathbf{U})$  bounded as  $O(2^{-N^\beta})$  for any  $0 < \beta < \frac{1}{2}$ . This result on the rate of polarization was proved in [23].

*Source coding:* Korada and Urbanke showed that it is possible to achieve the (symmetric) rate-distortion function  $R_s(D)$  of a discrete memoryless source (DMS)  $Y$  using PCs. The PCs are designed similarly as if they were to be used for transmission over the test channel  $W_Q(y|\hat{y})$  associated with  $R_s(D)$ . In this case the frozen set  $F_Q$  is given by the very noisy synthetic channels and the frozen values  $\mathbf{u}_{F_Q}$  are assumed for their symbols. Compression at any rate  $R_Q > R_s(D)$  of a length- $N$  source vector  $\mathbf{y}$  into a vector  $\hat{\mathbf{u}}$  with PCs is performed using the SC algorithm as follows. Let the design probability distribution be

$$P_{\mathbf{U}, \mathbf{Y}, \hat{\mathbf{Y}}}(\mathbf{u}, \mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{q^N} \mathbb{1}_{\{\hat{\mathbf{y}}=\mathbf{u}\mathbf{G}_N\}} \prod_{i=0}^{N-1} W_Q(y_i|\hat{y}_i), \quad (3)$$

where  $\mathbb{1}_{\{\cdot\}}$  is an indicator function that takes value 1 if the argument is true and 0 otherwise. For each  $i \in \{0, 1, \dots, N-1\}$  if  $i \in F_Q$  then set  $\hat{u}_i = u_i$  (i.e.  $\hat{\mathbf{u}}_{F_Q} = \mathbf{u}_{F_Q}$ ), otherwise set  $\hat{u}_i = \tilde{u}$  ( $\tilde{u} \in \mathcal{U}$ ) with probability<sup>4</sup>

$$P_{U_i|\mathbf{U}_0^{i-1}, \mathbf{Y}}(u_i = \tilde{u}|\hat{\mathbf{u}}_0^{i-1}, \mathbf{y}) \quad (4)$$

where (4) is obtained from (3) by conditioning and marginalizing. If the values of the frozen symbols  $\mathbf{u}_{F_Q}$  are sampled i.i.d. from a uniform distribution then source compression with PCs is described by the pmf  $Q(\mathbf{u}, \mathbf{y}, \hat{\mathbf{y}}) = Q(\mathbf{u}, \mathbf{y}) \mathbb{1}_{\{\hat{\mathbf{y}}=\mathbf{u}\mathbf{G}_N\}}$  where  $Q(\mathbf{u}, \mathbf{y})$  has marginal  $Q(\mathbf{y}) = P(\mathbf{y})$  and satisfies [14]:

$$Q(u_i|\mathbf{u}_0^{i-1}, \mathbf{y}) = \begin{cases} \frac{1}{q} & \text{if } i \in F_Q, \\ P(u_i|\mathbf{u}_0^{i-1}, \mathbf{y}) & \text{if } i \in F_Q^c. \end{cases} \quad (5)$$

The encoder only needs to communicate the symbols  $\mathbf{u}_{F_Q^c}$  to the decoder to allow it to reconstruct  $\mathbf{y}$  within distortion  $D$  as  $\hat{\mathbf{y}} = \hat{\mathbf{u}}\mathbf{G}_N$ . The compression rate is therefore  $R_Q = |F_Q^c|/N > I(W_Q)$ . Throughout this paper we will denote the

<sup>3</sup>Whenever it is clear from the context we shall drop the subindex  $N$ .

<sup>4</sup>This encoding operation is random. This implies that encoding twice the same source output may not produce the same vector  $\hat{\mathbf{u}}$ . This can be avoided if both encodings are performed using a source of common randomness [14].

source encoding of a vector  $\mathbf{y}$  using frozen symbols  $\mathbf{u}_{F_Q}$  by  $\hat{U}(\mathbf{y}, \mathbf{u}_{F_Q})$  (resulting in the vector  $\hat{\mathbf{u}}$ ) and the reconstruction simply by  $\hat{\mathbf{y}} = \hat{\mathbf{u}}\mathbf{G}_N$ , where it is assumed that the frozen symbols are set to the same values  $\mathbf{u}_{F_Q}$ .

The previous development is still valid if a different set of transition probabilities is chosen. That is, if we design the PC following the same principles but using some arbitrary  $\hat{W}(y|\hat{y})$  instead of the test channel  $W_Q(y|\hat{y})$  then, for sufficiently large block length  $N$ , it performs arbitrarily close to the rate-distortion pair  $(I(\hat{W}), \hat{D})$  defined by  $\hat{W}$ . This observation plays a fundamental role in our proofs.<sup>5</sup>

The following bound on the variational distance between the marginals of the design distribution  $P_{\mathbf{Y}, \hat{\mathbf{Y}}}$  and the distribution  $Q(\mathbf{y}, \hat{\mathbf{y}})$  induced by the PC was established in [14], [16]:

**Lemma 1.** *Let  $0 < \beta < \frac{1}{2}$  and  $\delta_N \triangleq \frac{1}{N}2^{-N^\beta}$ . For any compression rate  $R > I(W_Q)$ , if the frozen set  $F$  is defined as*

$$F \triangleq \left\{ i \in \{0, 1, \dots, N-1\} : I(W_Q^{(i)}) \leq \delta_N \right\},$$

we have that

$$\frac{1}{2} \sum_{\mathbf{y}, \hat{\mathbf{y}}} |P(\mathbf{y}, \hat{\mathbf{y}}) - Q(\mathbf{y}, \hat{\mathbf{y}})| \leq O(2^{-N^\beta}). \quad (6)$$

### C. The Relay Channel

In this paper we consider a particular instance of the discrete memoryless relay channel that has orthogonal receiver components [24]. That is, the pmf governing its behavior factorizes as

$$p(y_d, y_{sr} | x, x_r) = p(y_{sd}, y_{sr} | x) p(y_{rd} | x_r) \quad (7)$$

with  $Y_D = (Y_{SD}, Y_{RD})$ . Fig. 1 illustrates this channel and includes the following length- $N$  vectors:  $\mathbf{M}$  contains the information and the frozen symbols (the latter are also available at the decoders) transmitted by the source, and  $\hat{\mathbf{M}}$  and  $\hat{\mathbf{M}}_R$  are the corresponding estimates at the destination and the relay (only in DF relaying), respectively.  $\mathbf{X}$  is the vector of symbols put into the channel by the source, and  $\mathbf{Y}_{SR}$  and  $\mathbf{Y}_{SD}$  are the channel outputs at the relay and at the destination, respectively. Similarly,  $\mathbf{X}_R$  is the vector put into the channel by the relay and  $\mathbf{Y}_{RD}$  is the observation at the destination. Finally, in the case of CF we will consider the vectors  $\mathbf{Y}_Q$ , which is a compressed version of  $\mathbf{Y}_{SR}$  generated by the relay, and  $\hat{\mathbf{Y}}_Q$  which is the corresponding estimate generated by the destination. We will refer to the marginal pmfs of  $y_{sr}$ ,  $y_{sd}$ , and  $y_{rd}$  from (7) as the source-relay, source-destination, and relay-destination channels, respectively and denote them by  $W_{SR}$ ,  $W_{SD}$ , and  $W_{RD}$ , respectively.

Before introducing several bounds on the capacity of the relay channel we review the concept of degradation, which is a precise statement of the notion that some channels are better than other ones.

**Definition 1** (Stochastic degradation). *Let  $W_1 : \mathcal{X} \rightarrow \mathcal{Y}_1$  and  $W_2 : \mathcal{X} \rightarrow \mathcal{Y}_2$  be two arbitrary DMCs with  $q$ -ary input*

<sup>5</sup>As opposed to the work on rate-distortion theory with PCs the distortion does not play an explicit role in our work.

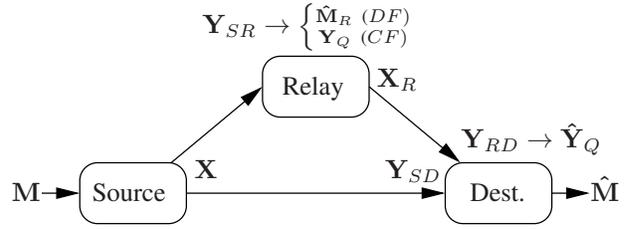


Fig. 1. Relay channel with orthogonal receiver components.

alphabet  $\mathcal{X}$ .  $W_1$  is stochastically degraded with respect to  $W_2$  if there exists a DMC  $\tilde{W} : \mathcal{Y}_2 \rightarrow \mathcal{Y}_1$  such that for every  $y_1 \in \mathcal{Y}_1$  and  $x \in \mathcal{X}$

$$W_1(y_1|x) = \sum_{y_2 \in \mathcal{Y}_2} W_2(y_2|x) \tilde{W}(y_1|y_2).$$

A relay channel is said to be stochastically degraded whenever the source-destination channel  $W_{SD}(y_{sd}|x)$  is stochastically degraded with respect to the source-relay channel  $W_{SR}(y_{sr}|x)$ . Similarly, the relay channel with orthogonal receivers is said to be physically degraded whenever

$$p(y_{sd}, y_{sr} | x) = W_{SR}(y_{sr}|x) p(y_{sd} | y_{sr}). \quad (8)$$

Note that physical degradation implies stochastic degradation.

We now introduce three bounds on the capacity of the relay channel that were established in [2] for the general case. We write them adapted to the instance of the relay channel considered here (cf. [24]) and with the following additional restrictions that take into account the nature of PCs:

*Additional constraints:* We consider the following additional constraints (as compared to [2], [24]) when evaluating the rates achievable with our code constructions (i.e. (9) and (10)):

- $X$  and  $X_R$  must follow uniform distributions with  $|\mathcal{X}|$  and  $|\mathcal{X}_R|$  prime numbers.
- The admissible conditional probabilities  $p(y_q | y_{sr})$  in the characterization of the CF achievable rate must induce a uniform distribution on  $\mathcal{Y}_Q$  with  $|\mathcal{Y}_Q|$  a prime number.

They are natural consequences of the special properties of PCs. We have decided to omit the existing workarounds to most of these issues (see e.g., [12]–[16]) since they bring no insight to the understanding of the problem discussed in this paper.

Taking into account these constraints the bounds on the capacity  $C$  of the relay channel are:

**Definition 2** (Cut-set upper bound).

$$C \leq \max_{p(x)p(x_r)} \min \{I(X; Y_{SR}Y_{SD}), I(X; Y_{SD}) + I(X_R; Y_{RD})\}.$$

**Definition 3** (Symmetric DF rate for relay channels with orthogonal receivers).

$$R_s^{DF} = \min \{I(W_{SR}), I(W_{SD}) + I(W_{RD})\}. \quad (9)$$

It is well known that DF relaying achieves the capacity of the physically degraded relay channel [2]. When such a channel is symmetric we shall denote its capacity by  $C_s^{PD}$ .

**Definition 4** (Symmetric CF rate for relay channels with orthogonal receivers).

$$R_s^{CF} = \sup I(X; Y_Q Y_{SD}) \quad (10)$$

The supremum is over all conditional distributions  $p(y_q|y_{sr})$  that induce a uniform distribution on  $\mathcal{Y}_Q$  and with  $|\mathcal{Y}_Q|$  a prime number, and such that  $I(W_{RD}) \geq I(Y_Q; Y_{SR}|Y_{SD})$ .

### III. MAIN RESULTS

The main results of this paper are given in the following two theorems:

**Theorem 1** (Symmetric decode-and-forward relaying with polar codes). Consider a stochastically degraded relay channel with orthogonal receiver components. For any transmission rate  $R < R_s^{DF}$  there exists a sequence of polar codes (indexed by the block length  $N$ ) with block error probability  $P_e = \Pr(\hat{\mathbf{M}} \neq \mathbf{M})$  under SC decoding bounded as  $P_e \leq O(2^{-N^\beta})$  for any  $0 < \beta < \frac{1}{2}$ .

**Theorem 2** (Symmetric compress-and-forward relaying using polar codes). Consider a relay channel with orthogonal receiver components. For any fixed rate  $R < R_s^{CF}$  there exists a sequence of polar codes (indexed by the block length  $N$ ) with block error probability at the destination  $P_e = \Pr(\hat{\mathbf{M}} \neq \mathbf{M})$  under SC decoding bounded as  $P_e \leq O(2^{-N^\beta})$  for any  $0 < \beta < \frac{1}{2}$ .

We prove the two theorems in Sections V and VI but first we review some results on PCs for degraded channels.

### IV. POLAR CODES FOR DEGRADED $Q$ -ARY CHANNELS

In this section we review some results that relate PC constructions for stochastically degraded channels introduced in [14] for binary input alphabets and extend them to general discrete input alphabets. In order to do this we first define the Bhattacharyya parameter of a  $q$ -ary DMC.

**Definition 5** (Average Bhattacharyya parameter of a  $q$ -ary DMC  $W(y|x)$ ).

$$Z(W) = \frac{1}{q(q-1)} \sum_{x,x' \in \mathcal{X}, x \neq x'} \sum_{y \in \mathcal{Y}} \sqrt{W(y|x)W(y|x')}.$$

The Bhattacharyya parameter is an upper bound on the error probability of uncoded transmission over the channel  $W$  [13].

The following result, introduced in [14] for binary DMCs and easily extended to  $q$ -ary DMCs, states a fundamental property of PCs constructed based on stochastically degraded channels.

**Proposition 1.** Let  $W_1(y|x)$  and  $W_2(y|x)$  be two  $q$ -ary DMCs such that  $W_1$  is stochastically degraded with respect to  $W_2$ . Let  $W_1^{(i)}$  and  $W_2^{(i)}$  be the  $i^{\text{th}}$  synthetic channels in PCs generated from  $W_1$  and  $W_2$ , respectively ( $i \in \{0, 1, \dots, N-1\}$ ). For each PC (labeled by  $j$ ,  $j = \{1, 2\}$ ) let the frozen set be

$$F_j = \left\{ i \in \{0, 1, \dots, N-1\} : Z(W_j^{(i)}) \geq \delta_N \right\}.$$

for some  $\delta_N > 0$ . Then for all  $i \in \{0, 1, \dots, N-1\}$  the following properties hold:

- 1) The Bhattacharyya parameters of the synthetic channels satisfy:  $Z(W_1^{(i)}) \geq Z(W_2^{(i)})$ .
- 2) If  $i \in F_2$  then  $i \in F_1$  as well. That is,  $F_2 \subseteq F_1$ .

*Proof:* The proof of the first claim follows the same lines as in the binary case [14]. The second claim follows directly from the first one. ■

The frozen set of a PC is usually defined in terms of either  $I(W^{(i)})$  or  $Z(W^{(i)})$  of the individual synthetic channels. We will use the following result that connects these two (different) definitions for  $q$ -ary DMCs:

**Lemma 2.** Consider an arbitrary  $q$ -ary DMC  $W$ . Let  $F$  and  $F'$  be the sets defined by

$$F \triangleq \left\{ i \in \{0, 1, \dots, N-1\} : I(W_N^{(i)}) \leq \delta_N \right\},$$

$$F' \triangleq \left\{ i \in \{0, 1, \dots, N-1\} : Z(W_N^{(i)}) \geq 1 - \delta'_N \right\}$$

with  $\delta'_N = \frac{q-q^{1-\delta_N}}{q-1}$ . Then  $F \subseteq F'$ . Moreover, for  $\delta_N \in [0, 1]$  and given  $0 < \beta < \frac{1}{2}$ , if  $\delta_N \leq O(2^{-N^\beta})$  then we have that  $\delta_N \leq \delta'_N \leq O(2^{-N^\beta})$ .

*Proof:* The fact that  $F \subseteq F'$  follows from [13, Proposition 3]. To prove that  $\delta_N \leq \delta'_N$  note that  $\delta'_N = \delta_N$  for  $\delta_N \in \{0, 1\}$  and that  $\delta'_N$  is an increasing concave function of  $\delta_N$  in  $[0, 1]$ . Therefore  $\delta_N \leq \delta'_N$  in  $[0, 1]$ . Now note that  $\delta_N \leq O(2^{-N^\beta})$  means that there exists  $K > 0$  such that  $\delta_N \leq K2^{-N^\beta}$  for all sufficiently large  $N$ . Consider the function  $\delta'_N(\delta_N)$ . We have that

$$\lim_{N \rightarrow \infty} \frac{\delta'_N(K2^{-N^\beta})}{2^{-N^\beta}} = \frac{Kq \log_e q}{q-1}$$

so that  $\delta'_N(K2^{-N^\beta}) \leq O(2^{-N^\beta})$ . Since for all sufficiently large  $N$  we have that  $\delta'_N(\delta_N) \leq \delta'_N(K2^{-N^\beta})$ , we conclude that  $\delta'_N(\delta_N) \leq O(2^{-N^\beta})$ . ■

### V. DECODE-AND-FORWARD RELAYING USING POLAR CODES

The construction of PCs for DF relaying in binary-input channels in [18] exploited the nested structure of PCs for stochastically degraded channels introduced in the previous section. We extend this construction here to relay channels with larger discrete input alphabets.

*Proof of Theorem 1*

*Encoding at the source node:* The source node chooses a rate  $R < I(W_{SR})$  and uses a (sequence of) PC that is capacity achieving for the channel  $W_{SR}$ . Let  $0 < \beta < \frac{1}{2}$  and  $\delta_N \triangleq \frac{1}{N}2^{-N^\beta}$ . The frozen set  $F_{SR}$  is defined as

$$F_{SR} \triangleq \left\{ i \in \{0, 1, \dots, N-1\} : Z(W_{SR}^{(i)}) \geq \delta_N \right\}$$

and its complement is denoted by  $F_{SR}^c$ . This (sequence of) PC is used to encode the information symbols  $\mathbf{M}_{F_{SR}^c}$ . The values of the frozen symbols  $\mathbf{M}_{F_{SR}}$  used for encoding are generated using common randomness and hence they are also available at the relay and the destination. Note that this code is not directly

decodable by the destination. For direct communication over  $W_{SD}$  the frozen set would need to be defined as

$$F_{SD} \triangleq \left\{ i \in \{0, 1, \dots, N-1\} : Z(W_{SD}^{(i)}) \geq \delta_N \right\}$$

and the decoder would need to know the symbols  $\mathbf{M}_{F_{SD}}$  to recover the message. Note that  $F_{SR} \subseteq F_{SD}$  by Proposition 1 so the destination decoder already knows the values of some of these symbols (i.e. those  $M_i$  with  $i \in F_{SR}$ ). The rest of them contain information and will be provided by the relay.

*Processing at the relay:* The relay decodes the message from the source, extracts the information symbols with indices in  $F_{SD} \cap F_{SR}^c$  (see Fig. 2), and re-encodes them using a (sequence of) PC that is capacity achieving for  $W_{RD}$ . In addition, if  $|\mathcal{X}| \neq |\mathcal{X}_R|$ , the relay changes the representation of these information symbols from  $|\mathcal{X}|$ -ary to  $|\mathcal{X}_R|$ -ary.

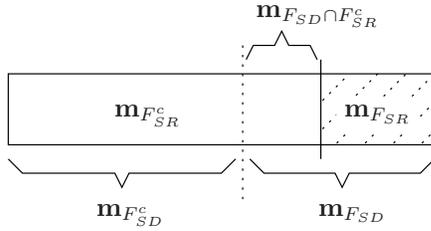


Fig. 2. Nested structure of PCs for decode-and-forward relaying. The symbols  $\mathbf{m}_{F_{SD}^c}$  constitute the message transmitted by the source while the frozen symbols  $\mathbf{m}_{F_{SR}^c}$  are known by source, relay, and destination. The relay obtains the message symbols  $\mathbf{m}_{F_{SR}^c}$  from its observation (obtained from the channel  $W_{SR}$ ) using the SC algorithm, extracts the part  $\mathbf{m}_{F_{SD} \cap F_{SR}^c}$  and forwards it to the destination. The destination uses these symbols  $\mathbf{m}_{F_{SD} \cap F_{SR}^c}$  in addition to  $\mathbf{m}_{F_{SR}^c}$  to recover the message symbols  $\mathbf{m}_{F_{SD}^c}$  from its observation (obtained from the channel  $W_{SD}$ ) using the SC algorithm.

*Decoding at the destination:* The destination decodes the message from the relay that contains part of the message transmitted by the source. These symbols together with  $\mathbf{M}_{F_{SR}^c}$  are all the symbols in the frozen set  $F_{SD}$ . Therefore, the destination can decode  $\mathbf{M}$  from  $\mathbf{Y}_{SD}$  using the SC algorithm.

*Analysis of the error probability:* Let  $\mathcal{E}$ ,  $\mathcal{E}_{SR}$ , and  $\mathcal{E}_{RD}$  denote the events  $\{\hat{\mathbf{M}} \neq \mathbf{M}\}$ ,  $\{\hat{\mathbf{M}}_R \neq \mathbf{M}\}$ , and an erroneous relay-destination transmission, respectively. Let  $\mathcal{E}^c$ ,  $\mathcal{E}_{SR}^c$ , and  $\mathcal{E}_{RD}^c$  denote their respective complementary events. Using this notation we write:

$$\begin{aligned} \Pr(\mathcal{E}) &= \Pr(\mathcal{E}|\mathcal{E}_{SR}) \Pr(\mathcal{E}_{SR}) + \Pr(\mathcal{E}|\mathcal{E}_{SR}^c) \Pr(\mathcal{E}_{SR}^c) \\ &\leq \Pr(\mathcal{E}_{SR}) + \Pr(\mathcal{E}|\mathcal{E}_{SR}^c). \end{aligned}$$

Since the source uses a PC for channel transmission at rate  $R < I(W_{SR})$  the first term can be bounded as  $O(2^{-N^\beta})$ . For the second term we have:

$$\begin{aligned} \Pr(\mathcal{E}|\mathcal{E}_{SR}^c) &= \Pr(\mathcal{E}|\mathcal{E}_{SR}^c, \mathcal{E}_{RD}) \Pr(\mathcal{E}_{RD}|\mathcal{E}_{SR}^c) \\ &\quad + \Pr(\mathcal{E}|\mathcal{E}_{SR}^c, \mathcal{E}_{RD}^c) \Pr(\mathcal{E}_{RD}^c|\mathcal{E}_{SR}^c) \\ &= \Pr(\mathcal{E}|\mathcal{E}_{SR}^c, \mathcal{E}_{RD}) \Pr(\mathcal{E}_{RD}) \\ &\quad + \Pr(\mathcal{E}|\mathcal{E}_{SR}^c, \mathcal{E}_{RD}^c) \Pr(\mathcal{E}_{RD}^c) \\ &\leq \Pr(\mathcal{E}_{RD}) + \Pr(\mathcal{E}|\mathcal{E}_{SR}^c, \mathcal{E}_{RD}^c) \end{aligned} \quad (11)$$

where we have used the fact that  $\mathcal{E}_{SR}$  and  $\mathcal{E}_{RD}$  are independent. Both terms in (11) describe the operation of the PC under the conditions for which it was designed and can therefore be

bounded as  $O(2^{-N^\beta})$ . The bound on the first term induces a constraint on the transmission rate from relay to destination:  $R_{RD} < I(W_{RD})$ . Similarly, the bound on the second term requires

$$R < \frac{|F_{SD}^c|}{N} + R_{RD} < I(W_{SD}) + I(W_{RD}).$$

Collecting all the terms we obtain the desired bound on  $P_e$  and the constraints on  $R$ . ■

This achievable rate coincides with the capacity of the symmetric physically degraded relay channel  $C_s^{PDD}$  [2], [24]. However, the preceding proof only requires the source-destination channel  $W_{SD}$  to be stochastically degraded with respect to the source-relay channel  $W_{SR}$ . Unfortunately, it is well known that, in general, DF relaying does not achieve the capacity of the stochastically degraded relay channel [2].

## VI. COMPRESS-AND-FORWARD RELAYING USING POLAR CODES

In this section we prove Theorem 2 by combining PCs for source and channel coding in a nested way. Then we specialize our result to the case where the relay describes its observation perfectly to the destination and observe that in some cases this achieves the capacity of the relay channel with orthogonal receiver components. Finally, we introduce a type of channels with a special structure for which any sequence of frozen symbols used by the relay during source compression is equally good.

### A. Proof of Theorem 2

Select a rate  $R < R_s^{CF}$  by choosing a valid  $p(y_q|y_{sr})$  in (10) and let  $W_Q(y_{sr}|y_q) = p(y_q|y_{sr})p(y_{sr})|\mathcal{Y}_Q|$  where  $p(y_{sr}) = \sum_{y_{sd}, x} p(y_{sd}, y_{sr}|x) \frac{1}{|\mathcal{X}|}$ .

*Encoding at the source node:* The source node encodes the information and frozen symbols  $\mathbf{M}$  using a (sequence of) PC that is capacity achieving for the channel  $W : \mathcal{X} \rightarrow \mathcal{Y}_{SD} \times \mathcal{Y}_Q$ :

$$W(y_{sd}, y_q|x) = \sum_{y_{sr} \in \mathcal{Y}_{SR}} p(y_q|y_{sr})p(y_{sd}, y_{sr}|x)$$

where  $p(y_{sd}, y_{sr}|x)$  comes from the channel pmf (7). Assume that the information symbols are chosen i.i.d. randomly from a uniform distribution. Similarly, let the frozen symbols be chosen i.i.d. from a uniform distribution (using common randomness, so that they are also available at the destination node). Then the observation at the relay  $\mathbf{Y}_{SR}$  has the distribution of a DMS:

$$\begin{aligned} P_{\mathbf{Y}_{SR}}(y_{sr}) &= \sum_{\mathbf{x}} \frac{1}{q^N} \prod_{i=0}^{N-1} W_{SR}(y_{sr,i}|x_i) \\ &= \prod_{i=0}^{N-1} \frac{1}{q} \sum_{x_i} W_{SR}(y_{sr,i}|x_i) \\ &= \prod_{i=0}^{N-1} p(y_{sr,i}). \end{aligned}$$

*Processing at the relay:* Let  $0 < \beta < \frac{1}{2}$  and  $\delta_N \triangleq \frac{1}{N}2^{-N^\beta}$ . The relay compresses  $\mathbf{Y}_{SR}$  into  $\mathbf{Y}_Q$  using a (sequence of) PC for source coding (see Section II-B) designed based on  $W_Q$  with compression rate  $R_Q > I(W_Q)$ , frozen set

$$F_Q \triangleq \left\{ i \in \{0, 1, \dots, N-1\} : I(W_Q^{(i)}) \leq \delta_N \right\},$$

and frozen symbols  $\mathbf{U}_{F_Q}$ , generated using common randomness so that they are also available at the destination. Let  $F_Q^c$  be the complement of  $F_Q$ . The compressed signal can be reconstructed as  $\mathbf{Y}_Q = \hat{U}(\mathbf{Y}_{SR}, \mathbf{U}_{F_Q})\mathbf{G}_N = \mathbf{U}\mathbf{G}_N$ . Hence it would be possible to reconstruct  $\mathbf{Y}_Q$  at the destination if  $\mathbf{U}_{F_Q^c}$  were sent over the relay-destination channel. However, this would require communicating over  $W_{RD}$  at rate  $R_Q$ , which is more than it is allowed by the constraint in Definition 4. To lower the rate required while still allowing the destination to reconstruct  $\mathbf{Y}_Q$  we exploit the correlation between  $\mathbf{Y}_Q$  and the direct-link observation at the destination  $\mathbf{Y}_{SD}$ . This correlation has some distribution that is determined by the channel distribution and the PC used for source compression. Rather than considering the true distribution we simply model this correlation as if  $\mathbf{Y}_Q$  were transmitted over the *virtual* DMC

$$W_V(y_{sd}|y_q) = \sum_{y_{sr}} \left( W_Q(y_{sr}|y_q) \frac{1}{p(y_{sr})} \sum_{x \in \mathcal{X}} \frac{1}{|\mathcal{X}|} p(y_{sd}, y_{sr}|x) \right) \quad (12)$$

giving rise to  $\mathbf{Y}_{SD}$  as channel output and then show that this approximation is good enough for large  $N$ . Note that  $W_V$  defines a Markov chain:  $Y_Q \rightarrow Y_{SR} \rightarrow Y_{SD}$ . This simplified correlation model allows for nesting PCs for source coding into PCs for channel coding. This was introduced in [14] for binary Wyner-Ziv coding. We extend it here in the context of CF relaying to larger alphabets using the results from Section IV, [13], and [16].

Note that the input  $\mathbf{Y}_Q = \mathbf{U}\mathbf{G}_N$  to the virtual channel is a valid codeword from a PC. Assuming that our model for the correlation in (12) is correct we can use the SC algorithm to decode the (virtual) channel input  $\mathbf{Y}_Q$  from the (virtual) channel output  $\mathbf{Y}_{SD}$ , if the (virtual) communication rate satisfies  $R_V < I(W_V)$ . For this to be possible (see Section II-B) we need to define a frozen set

$$F_V \triangleq \left\{ i \in \{0, 1, \dots, N-1\} : Z(W_V^{(i)}) \geq \delta'_N \right\},$$

(with complement  $F_V^c$ ) and ensure that the decoder knows the values of the symbols in this frozen set  $\mathbf{U}_{F_V}$ . Note that, unlike in regular channel coding scenarios (here the virtual channel acts as a model), we cannot choose the values of these symbols; they are set by the compression algorithm. Hence we need to communicate them to the destination. Fortunately, we can proceed as we did in DF relaying to reduce the information to be conveyed to the destination. Consider the set<sup>6</sup>

$$F'_Q \triangleq \left\{ i \in \{0, 1, \dots, N-1\} : Z(W_Q^{(i)}) \geq 1 - \delta'_N \right\}.$$

<sup>6</sup> $F'_Q$  is an auxiliary set that is only used to show that  $F_Q \subseteq F_V$ . No PC is explicitly built using this set.

If we choose  $\delta'_N = \frac{q-q^{1-\delta_N}}{q-1}$  then Lemma 2 ensures that  $F_Q \subseteq F'_Q$ . Moreover, since  $W_V$  is stochastically degraded with respect to  $W_Q$ , for  $\delta_N$  such that  $\delta'_N < 1 - \delta'_N$ , we have that  $F'_Q \subseteq F_V$  by Proposition 1 and hence  $F_Q \subseteq F_V$ . Since the values  $\mathbf{U}_{F_Q}$  are already known at the destination, only the symbols in  $F_V \cap F_Q^c$  need to be transmitted over the relay-destination channel (see Fig. 3). Note that it is by choosing  $F_V$  that we set the (virtual) communication rate, i.e.  $R_V \leq |F_V^c|/N$ . Recall that we have that  $\delta_N \leq \delta'_N \leq O(2^{-N^\beta})$ . Due to the optimality of PCs for (symmetrical) channel and source coding, for any rates  $R_V < I(W_V)$  and  $R_Q > I(W_Q)$ , we have that  $R_V \leq |F_V^c|/N < I(W_V)$  and  $R_Q \geq |F_Q^c|/N > I(W_Q)$  if  $N$  is sufficiently large. Hence, reliable transmission over the relay-destination channel  $W_{RD}$  at any rate  $R_{RD}$  that satisfies

$$R_{RD} = \frac{|F_Q^c| - |F_V^c|}{N} > I(W_Q) - I(W_V) \quad (13)$$

is sufficient to be able to recover  $\mathbf{Y}_Q$  at the destination. In addition, if  $|\mathcal{Y}_Q| \neq |\mathcal{X}_R|$ , the relay changes the representation of the symbols from  $|\mathcal{Y}_Q|$ -ary to  $|\mathcal{X}_R|$ -ary. This allows the relay to use a (sequence of) PC for channel coding for transmission over  $W_{RD}$ , with the frozen symbols generated i.i.d. using common randomness so that they are also known at the destination.

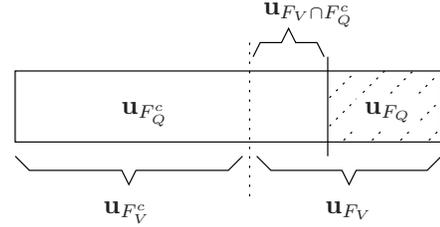


Fig. 3. Nested structure of PCs for compress-and-forward relaying. The symbols  $\mathbf{u}_{F_Q}$  are fixed when performing source compression with PCs at the relay and therefore known by both the relay and the destination. The destination needs to obtain the remaining symbols  $\mathbf{u}_{F_Q^c}$ . If the relay communicates only the symbols  $\mathbf{u}_{F_V \cap F_Q^c}$ , the destination can use its own channel observation to recover the remaining symbols  $\mathbf{u}_{F_V^c}$  using the SC algorithm.

Finally, we rewrite the bound on  $R_{RD}$  by considering the Markov chain  $Y_Q \rightarrow Y_{SR} \rightarrow Y_{SD}$ :

$$\begin{aligned} I(Y_Q; Y_{SR}) &= I(Y_Q; Y_{SR}Y_{SD}) \\ &= I(Y_Q; Y_{SD}) + I(Y_Q; Y_{SR}|Y_{SD}). \end{aligned}$$

This holds for any distribution on  $Y_Q$  as long as the Markov chain relationship is satisfied. In particular, when  $Y_Q$  is uniformly distributed we have that  $I(Y_Q; Y_{SR}|Y_{SD}) = I(W_Q) - I(W_V)$ .

*Decoding at the destination:* The destination performs three decoding steps. First it decodes the message transmitted by the relay, i.e. it obtains  $\mathbf{U}_{F_V \cap F_Q^c}$ . These symbols together with  $\mathbf{U}_{F_Q}$  are all the symbols in the frozen set for the virtual channel  $F_V$ . Knowing them allows the destination node to decode the compressed vector  $\mathbf{Y}_Q = \mathbf{U}\mathbf{G}_N$  from  $\mathbf{Y}_{SD}$  using the SC algorithm. Finally, the destination decodes the message  $\mathbf{M}$  from the estimate  $\hat{\mathbf{Y}}_Q$  and  $\mathbf{Y}_{SD}$  using the SC algorithm.

*Analysis of the error probability:* We want to evaluate the probability of the event  $\mathcal{E} = \{\hat{\mathbf{M}} \neq \mathbf{M}\}$  over the distribution induced by the channel and the different codes in our scheme. We denote this distribution as  $P_{\mathcal{S}}(s)$  where  $\mathcal{S}$  is the set of all the random vectors present in the scenario. To make the dependency of  $P_{\mathcal{S}}(s)$  on the distribution  $Q(\mathbf{y}_{sr}, \mathbf{y}_q)$  induced by the PC used for source coding at the relay explicit we define  $\mathcal{S}_s = \mathcal{S} \setminus \{\mathbf{Y}_{SR}, \mathbf{Y}_Q\}$ , the subset of  $\mathcal{S}$  that excludes  $\mathbf{Y}_{SR}, \mathbf{Y}_Q$ , and write

$$P_{\mathcal{S}}(s) = Q(\mathbf{y}_{sr}, \mathbf{y}_q) P_{\mathcal{S}_s | \mathbf{Y}_{SR}, \mathbf{Y}_Q}(s_s | \mathbf{y}_{sr}, \mathbf{y}_q). \quad (14)$$

To compute  $\Pr(\mathcal{E})$  we can replace the term  $Q(\mathbf{y}_{sr}, \mathbf{y}_q)$  by any distribution  $p(\mathbf{y}_{sr}, \mathbf{y}_q, \tilde{\mathbf{y}}_{sr}, \tilde{\mathbf{y}}_q)$  as long as its marginal coincides with  $Q(\mathbf{y}_{sr}, \mathbf{y}_q)$  because the probability of our event of interest only depends on the marginal  $Q(\mathbf{y}_{sr}, \mathbf{y}_q)$ . In particular, we consider  $\tilde{\mathcal{Y}}_{SR} = \mathcal{Y}_{SR}$  and  $\tilde{\mathcal{Y}}_Q = \mathcal{Y}_Q$  and replace  $Q(\mathbf{y}_{sr}, \mathbf{y}_q)$  by the optimal coupling (see [25])  $P_E(\mathbf{y}_{sr}, \mathbf{y}_q, \tilde{\mathbf{y}}_{sr}, \tilde{\mathbf{y}}_q)$  between the distribution  $P(\mathbf{y}_{sr}, \mathbf{y}_q)$  used for designing the PC for source compression and the distribution  $Q(\mathbf{y}_{sr}, \mathbf{y}_q)$  induced by the PC (cf. (3) and (5), respectively). This optimal coupling  $P_E$  has marginals equal to  $P(\tilde{\mathbf{y}}_{sr}, \tilde{\mathbf{y}}_q)$  and  $Q(\mathbf{y}_{sr}, \mathbf{y}_q)$ , and for  $(\mathbf{Y}_{SR}, \mathbf{Y}_Q, \tilde{\mathbf{Y}}_{SR}, \tilde{\mathbf{Y}}_Q)$  distributed according to  $P_E$  the probability of the event  $\mathcal{E}_E \triangleq \{(\mathbf{Y}_{SR}, \mathbf{Y}_Q) \neq (\tilde{\mathbf{Y}}_{SR}, \tilde{\mathbf{Y}}_Q)\}$  is:

$$\Pr(\mathcal{E}_E) = \frac{1}{2} \sum_{\mathbf{y}_{sr}, \mathbf{y}_q} |P(\mathbf{y}_{sr}, \mathbf{y}_q) - Q(\mathbf{y}_{sr}, \mathbf{y}_q)|.$$

In addition to the already defined events  $\mathcal{E}$  and  $\mathcal{E}_E$ , consider also  $\mathcal{E}_{Y_Q}$  and  $\mathcal{E}_{RD}$  which denote the events  $\{\hat{\mathbf{Y}}_Q \neq \mathbf{Y}_Q\}$ , and an erroneous relay-destination transmission, respectively. Let  $\mathcal{E}^c, \mathcal{E}_E^c, \mathcal{E}_{Y_Q}^c$ , and  $\mathcal{E}_{RD}^c$  denote the respective complementary events. Using this notation and the new probability distribution obtained by replacing  $Q(\mathbf{y}_{sr}, \mathbf{y}_q)$  by the optimal coupling  $P_E$  in (14) we write

$$\begin{aligned} \Pr(\mathcal{E}) &= \Pr(\mathcal{E} | \mathcal{E}_{RD}) \Pr(\mathcal{E}_{RD}) + \Pr(\mathcal{E} | \mathcal{E}_{RD}^c) \Pr(\mathcal{E}_{RD}^c) \\ &\leq \Pr(\mathcal{E}_{RD}) + \Pr(\mathcal{E} | \mathcal{E}_{RD}^c). \end{aligned} \quad (15)$$

If a (sequence of) PC is used for transmission over  $W_{RD}$  then  $\Pr(\mathcal{E}_{RD}) \leq O(2^{-N^\beta})$  by using the SC decoder provided that  $R_{RD} < I(W_{RD})$  [13]. We rewrite the second term in (15) as

$$\begin{aligned} \Pr(\mathcal{E} | \mathcal{E}_{RD}^c) &= \Pr(\mathcal{E} | \mathcal{E}_{RD}^c, \mathcal{E}_E) \Pr(\mathcal{E}_E | \mathcal{E}_{RD}^c) \\ &\quad + \Pr(\mathcal{E} | \mathcal{E}_{RD}^c, \mathcal{E}_E^c) \Pr(\mathcal{E}_E^c | \mathcal{E}_{RD}^c) \\ &\leq \Pr(\mathcal{E}_E | \mathcal{E}_{RD}^c) + \Pr(\mathcal{E} | \mathcal{E}_{RD}^c, \mathcal{E}_E^c) \\ &= \Pr(\mathcal{E}_E) + \Pr(\mathcal{E} | \mathcal{E}_{RD}^c, \mathcal{E}_E^c), \end{aligned} \quad (16)$$

where the last step is due to the independence of  $\mathcal{E}_E$  and  $\mathcal{E}_{RD}$ . We know from Lemma 1 that for our choice of  $R_Q$  and  $F_Q$  we have that  $\Pr(\mathcal{E}_E) \leq O(2^{-N^\beta})$ . We bound the second term in (16) as

$$\begin{aligned} \Pr(\mathcal{E} | \mathcal{E}_{RD}^c, \mathcal{E}_E^c) &= \Pr(\mathcal{E} | \mathcal{E}_{RD}^c, \mathcal{E}_E^c, \mathcal{E}_{Y_Q}) \Pr(\mathcal{E}_{Y_Q} | \mathcal{E}_{RD}^c, \mathcal{E}_E^c) \\ &\quad + \Pr(\mathcal{E} | \mathcal{E}_{RD}^c, \mathcal{E}_E^c, \mathcal{E}_{Y_Q}^c) \Pr(\mathcal{E}_{Y_Q}^c | \mathcal{E}_{RD}^c, \mathcal{E}_E^c) \\ &\leq \Pr(\mathcal{E}_{Y_Q} | \mathcal{E}_{RD}^c, \mathcal{E}_E^c) + \Pr(\mathcal{E} | \mathcal{E}_{RD}^c, \mathcal{E}_E^c, \mathcal{E}_{Y_Q}^c). \end{aligned} \quad (17)$$

The first term in (17) is upper bounded by  $\Pr(\mathcal{E}_{Y_Q} | \mathcal{E}_{RD}^c)$  when  $\mathbf{Y}_Q$  is generated according to the design distribution

$P(\tilde{\mathbf{y}}_{sr}, \tilde{\mathbf{y}}_q)$  rather than the distribution induced by the code  $Q(\mathbf{y}_{sr}, \mathbf{y}_q)$  because in this case we are decoding  $\mathbf{Y}_Q$  from  $\mathbf{Y}_{SD}$  using the SC algorithm under the design conditions. Hence, as long as (13) is satisfied, we know from Lemma 2 that for our choice of  $\delta_N^l$  the first term in (17) can be bounded as  $O(2^{-N^\beta})$ . The same bound holds for the second term in (17) since the PC used by the source node is designed precisely under the hypothesis that  $\mathcal{E}_E^c$  and  $\mathcal{E}_{Y_Q}^c$  happen.

Collecting the different terms we obtain the bound on  $P_e$  and the constraint in Definition 4.  $\blacksquare$

### B. Compress-and-Forward Using Slepian-Wolf Coding

The following theorem follows as a special case of Theorem 2 for the case when the relay does not perform any lossy source compression<sup>7</sup>, i.e.  $\mathbf{Y}_Q = \mathbf{Y}_{SR}$ :

**Theorem 3** (CF relaying with PCs based on Slepian-Wolf coding). *Consider a relay channel with orthogonal receiver components. For any fixed rate  $R < I(X; Y_{SR} Y_{SD})$  there exists a sequence of polar codes (indexed by the block length  $N$ ) with block error probability at the destination  $P_e = \Pr(\hat{\mathbf{M}} \neq \mathbf{M})$  under SC decoding bounded as  $P_e \leq O(2^{-N^\beta})$  for any  $0 < \beta < \frac{1}{2}$ , as long as  $I(W_{RD}) \geq H(Y_{SR} | Y_{SD})$ .*

Note that the construction of PCs is greatly simplified due to the absence of a lossy compression PC. Since  $Y_Q = Y_{SR}$  the constraint on  $R_{RD}$  in (13) reduces to the Slepian-Wolf result [3]:

$$\begin{aligned} R_{RD} &> I(Y_Q; Y_{SR}) - I(Y_Q; Y_{SD}) \\ &= H(Y_Q) - H(Y_Q | Y_{SR}) - H(Y_Q) + H(Y_Q | Y_{SD}) \\ &= H(Y_{SR} | Y_{SD}). \end{aligned}$$

The importance of this particular case lies on the fact that in some special circumstances it coincides with the cut-set bound and hence with the capacity of the relay channel:

**Corollary 1.** *If all the channels are symmetric and  $I(W_{RD}) \geq H(Y_{SR} | Y_{SD})$  then CF relaying based on Slepian-Wolf coding achieves the cut-set bound, given in this case by the term  $I(X; Y_{SR} Y_{SD})$ .*

*Proof:* In this case the two terms of the cut-set bound satisfy:

$$\begin{aligned} I(X; Y_{SR} Y_{SD}) &\leq I(X; Y_{SD}) + H(Y_{SR} | Y_{SD}) \\ &\leq I(W_{SD}) + I(W_{RD}). \end{aligned}$$

Therefore, the rate  $R_s^{CF} = I(X; Y_{SR} Y_{SD})$  coincides with the cut-set bound.  $\blacksquare$

### C. Shift-Invariant Test Channels

The proof of Theorem 2 relied on bounding the probability of the event  $\mathcal{E}_E$  using Lemma 1. In (6) we are summing over all possible codewords. Since  $\mathbf{G}_N$  is a bijective mapping this means summing not only over all possible messages (i.e. information symbols) but also over all choices of frozen

<sup>7</sup>Note that in this case  $Y_{SR}$  must follow a uniform distribution and  $|\mathcal{Y}_{SR}|$  needs to be a prime number.

symbols. In this section we concentrate on a specific type of  $q$ -ary DMC that allows us to derive results that are independent of the choice of frozen symbols. This section parallels similar results on channel and source coding with PCs for binary alphabets [12], [14]. As it will become clear from the definition of a shift-invariant channel, the results on compress-and-forward reported in this section assume that  $|\mathcal{Y}_{SR}| = |\mathcal{Y}_Q|$ .

**Definition 6** (Shift-invariant DMC, SI-DMC). *A DMC  $W : \mathcal{X} \rightarrow \mathcal{Y}$ , with  $\mathcal{Y} = \mathcal{X}$ , is shift-invariant if for every  $x, a \in \mathcal{X}$  and  $y \in \mathcal{Y}$  we have that  $W(y|x) = W(y \oplus a|x \oplus a)$ .*

This is a particular case of the class of binary symmetric DMCs from [12] extended to  $q$ -ary alphabets. A properly designed (sequence of) PC for transmission over a SI-DMC will yield arbitrarily low error probability regardless of the choice of frozen symbols. This can easily be seen by considering the proof for binary symmetric DMCs in [12, Section VI].

The effect of transmitting through a SI-DMC can be expressed in terms of an additive error term. That is,  $\mathbf{y} = \mathbf{x} \oplus \mathbf{e}$  where the error  $\mathbf{e}$  is independent of the input and has the following distribution:

$$P_{SI}(\mathbf{e}) \triangleq \prod_{i=0}^{N-1} W(e_i|0).$$

For SI-DMCs and any  $\mathbf{u}, \mathbf{x}, \mathbf{y}$ , and  $\mathbf{a}$  we have that

$$\begin{aligned} W^N(\mathbf{y}|\mathbf{x}) &= W^N(\mathbf{y} \oplus \mathbf{a}|\mathbf{x} \oplus \mathbf{a}), \\ W_N(\mathbf{y}|\mathbf{u}) &= W_N(\mathbf{y} \oplus \mathbf{a}|\mathbf{u} \oplus \mathbf{a}\mathbf{G}_N^{-1}). \end{aligned}$$

It is also easy to see that the matrix of transition probabilities for a SI-DMC is doubly stochastic (i.e. each of its rows and columns sums up to 1). Using these properties we prove the following lemmata. Let  $W$  be a  $q$ -ary SI-DMC used for constructing a PC for source coding. The first lemma establishes the relationship that two vectors of source realizations and two choices of frozen symbols need to satisfy so that their compression rules (i.e. (4)) yield the same result.

**Lemma 3.** *Let  $\mathbf{y}, \mathbf{y}' \in \mathcal{Y}^N$  with  $\mathbf{y}' = \mathbf{y} \oplus \mathbf{a}$ . Let  $\mathbf{u}, \mathbf{u}' \in \mathcal{X}^N$  with  $\mathbf{u}'_0^{i-1} = \mathbf{u}_0^{i-1} \oplus \mathbf{v}_0^{i-1}$ , where  $\mathbf{v} = \mathbf{a}\mathbf{G}_N^{-1}$ . Consider the encoding rule  $P_{U_i|\mathbf{U}_0^{i-1}, \mathbf{Y}}$  in (4). We have that  $P(u_i|\mathbf{u}_0^{i-1}, \mathbf{y}) = P(u_i \oplus v_i|\mathbf{u}'_0^{i-1}, \mathbf{y}')$ .*

*Proof:* From the design distribution in (3) and using the definition of  $W^N$  in (1) we obtain:

$$\begin{aligned} P(u_i|\mathbf{u}_0^{i-1}, \mathbf{y}) &= \sum_{\mathbf{u}_{i+1}^{N-1}} \frac{P(\mathbf{u}_0^{i-1}, u_i, \mathbf{u}_{i+1}^{N-1}, \mathbf{y})}{P(\mathbf{u}_0^{i-1}, \mathbf{y})} \\ &= \frac{1}{q^N} \frac{1}{P(\mathbf{u}_0^{i-1}, \mathbf{y})} \sum_{\mathbf{u}_{i+1}^{N-1}} W^N(\mathbf{y}|\mathbf{u}\mathbf{G}_N). \end{aligned} \quad (18)$$

Similarly,  $P(u_i \oplus v_i|\mathbf{u}'_0^{i-1}, \mathbf{y}')$  can be written as

$$\frac{1}{q^N} \frac{1}{P(\mathbf{u}'_0^{i-1}, \mathbf{y}')} \sum_{\mathbf{u}_{i+1}^{N-1}} W^N(\mathbf{y}'|(\mathbf{u}'_0^{i-1}, u_i \oplus v_i, \mathbf{u}_{i+1}^{N-1})\mathbf{G}_N). \quad (19)$$

In deriving (19) we have used that the index of the summation is a dummy vector. The summations in (18) and (19) are equal

since  $W$  is a SI-DMC. The terms  $P(\mathbf{u}_0^{i-1}, \mathbf{y})$  and  $P(\mathbf{u}'_0^{i-1}, \mathbf{y}')$  are only scaling factors independent of  $u_i$  and  $v_i$ . However, because for any  $v_i$

$$\sum_{u_i} P(u_i|\mathbf{u}_0^{i-1}, \mathbf{y}) = \sum_{u_i} P(u_i \oplus v_i|\mathbf{u}'_0^{i-1}, \mathbf{y}') = 1$$

these scaling factors must be equal. Hence we conclude that  $P(u_i|\mathbf{u}_0^{i-1}, \mathbf{y}) = P(u_i \oplus v_i|\mathbf{u}'_0^{i-1}, \mathbf{y}')$ . ■

Using this result we now show that, under common randomness, the outputs of the SC compression algorithm for source realizations that satisfy the aforementioned relationship are also related:

**Lemma 4.** *Let  $\mathbf{y}, \mathbf{y}', \mathbf{a} \in \mathcal{Y}^N$ , with  $\mathbf{y}' = \mathbf{y} \oplus \mathbf{a}$ . Consider any set  $F \subseteq \{0, 1, \dots, N-1\}$  and  $\mathbf{u}_F, \mathbf{u}'_F$  with  $\mathbf{u}'_F = \mathbf{u}_F \oplus (\mathbf{a}\mathbf{G}_N^{-1})_F$  then, under common randomness,  $\hat{U}(\mathbf{y}', \mathbf{u}'_F) = \hat{U}(\mathbf{y}, \mathbf{u}_F) \oplus \mathbf{a}\mathbf{G}_N^{-1}$ .*

*Proof:* By induction, identical to that of [14, Lemma 9]. ■

Since PCs designed based on SI-DMCs use the source alphabet as reconstruction alphabet (i.e.  $\mathcal{X} = \mathcal{Y}$ ) we can talk about the error incurred by quantizing a source output  $\mathbf{y}$  into  $\hat{\mathbf{y}}$ . We define this quantization error as  $\mathbf{e} = \hat{\mathbf{y}} \oplus (-\mathbf{y})$ . We have the following property regarding this error.

**Corollary 2.** *For fixed frozen symbols  $\mathbf{u}_F$ , under common randomness, all source vectors that belong to the same coset  $\{\mathbf{y} : (\mathbf{y}\mathbf{G}_N^{-1})_F = \mathbf{v}_F\}$  have quantization error equal to  $\hat{U}(\mathbf{0}, \mathbf{u}_F \oplus (-\mathbf{v}_F))\mathbf{G}_N$ .*

*Proof:* Similar to that of its binary counterpart in [14]. ■

Moreover, for a PC for source compression designed based on a SI-DMC the distribution of the quantization error is close to the effect of transmission over the SI-DMC in the following sense:

**Lemma 5.** *Let  $Y$  be a  $q$ -ary DMS that puts out letters according to a uniform distribution. Let  $W(y|\hat{y})$  be a SI-DMC used in the design of a sequence of PCs for compression of  $Y$  into  $\hat{Y}$ . Let  $0 < \beta < \frac{1}{2}$  and  $\delta_N \triangleq \frac{1}{N}2^{-N^\beta}$  and let the frozen set be*

$$F = \left\{ i \in \{0, 1, \dots, N-1\} : I(W_N^{(i)}) \leq \delta_N \right\}.$$

*Let  $P_E^{\mathbf{u}_F}(\mathbf{e})$  denote the distribution of the quantization error when the symbols in the frozen set are fixed to  $\mathbf{u}_F$ . Then for any  $\mathbf{u}_F \in \mathcal{X}^{|F|}$  the variational distance between  $P_E^{\mathbf{u}_F}(\mathbf{e})$  and  $P_{SI}(\mathbf{e})$  satisfies*

$$\sum_{\mathbf{e}} \left| P_E^{\mathbf{u}_F}(\mathbf{e}) - \prod_{i=0}^{N-1} P_{SI}(e_i) \right| \leq O(2^{-N^\beta}).$$

*Proof:* Let  $P_{\mathbf{U}, \mathbf{Y}, \hat{\mathbf{Y}}}(\mathbf{u}, \mathbf{y}, \hat{\mathbf{y}})$  and  $Q(\mathbf{u}, \mathbf{y})$  be the design and induced distributions associated with the PC, respectively. Consider  $P_{\mathbf{U}, \mathbf{Y}}$ ,  $P_{\hat{\mathbf{Y}}, \hat{\mathbf{Y}}}$ , and  $P_{\mathbf{Y}, \hat{\mathbf{Y}}}$  obtained from  $P_{\mathbf{U}, \mathbf{Y}, \hat{\mathbf{Y}}}$ . We have that

$$\begin{aligned} P_E^{\mathbf{u}_F}(\mathbf{e}) &\triangleq \Pr \left\{ \hat{U}(\mathbf{y}, \mathbf{u}_F)\mathbf{G}_N \oplus (-\mathbf{y}) = \mathbf{e} \right\} \\ &= \Pr \left\{ \hat{U}(\mathbf{0}, \mathbf{u}_F \oplus (-\mathbf{y}\mathbf{G}_N^{-1})_F)\mathbf{G}_N = \mathbf{e} \right\} \quad (20) \\ &= Q(\mathbf{u}|\mathbf{0})\mathbf{1}_{\{\mathbf{e}=\mathbf{u}\mathbf{G}_N\}}. \quad (21) \end{aligned}$$

To write (20) we have used Corollary 2. To obtain (21) we have used the fact that  $Y$  is a uniform DMS and hence it induces a uniform distribution over  $\mathcal{U}^{|\mathcal{F}|}$  on  $\mathbf{u}_F \oplus (-\mathbf{y}\mathbf{G}_N^{-1})_F$ , independent of the actual value of  $\mathbf{u}_F$ . Since  $W$  is described by a doubly stochastic matrix we have that

$$\begin{aligned} \prod_{i=0}^{N-1} P_{SI}(e_i) &= P_{\mathbf{Y}|\hat{\mathbf{Y}}}(\mathbf{e}|\mathbf{0}) \\ &= P_{\hat{\mathbf{Y}}|\mathbf{Y}}(\mathbf{e}|\mathbf{0}) \\ &= P_{\mathbf{U}|\mathbf{Y}}(\mathbf{u}|\mathbf{0})\mathbb{1}_{\{\mathbf{e}=\mathbf{u}\mathbf{G}_N\}}. \end{aligned} \quad (22)$$

Combining (21) and (22) and using the fact that  $\mathbf{G}_N$  is a one-to-one mapping we obtain

$$\begin{aligned} \sum_{\mathbf{e}} \left| P_E^{\mathbf{u}_F}(\mathbf{e}) - \prod_{i=0}^{N-1} P_{SI}(e_i) \right| &= \sum_{\mathbf{u}} |Q(\mathbf{u}|\mathbf{0}) - P_{\mathbf{U}|\mathbf{Y}}(\mathbf{u}|\mathbf{0})| \\ &= q^N \sum_{\mathbf{u}} |Q(\mathbf{u}, \mathbf{y} = \mathbf{0}) - P_{\mathbf{U}, \mathbf{Y}}(\mathbf{u}, \mathbf{y} = \mathbf{0})|. \end{aligned}$$

Due to the shift-invariant property of the channel the summation takes the same value regardless of which  $\mathbf{y} \in \mathcal{Y}^N$  we consider. Using this property and Lemma 1 we obtain the desired result:

$$\begin{aligned} \sum_{\mathbf{e}} \left| P_E^{\mathbf{u}_F}(\mathbf{e}) - \prod_{i=0}^{N-1} P_{SI}(e_i) \right| &= \sum_{\mathbf{y}, \mathbf{u}} |Q(\mathbf{u}, \mathbf{y}) - P_{\mathbf{U}, \mathbf{Y}}(\mathbf{u}, \mathbf{y})| \\ &\leq O(2^{-N^\beta}). \end{aligned} \quad \blacksquare$$

Note that although we have again invoked Lemma 1 our result is valid for any choice of frozen symbols  $\mathbf{u}_F$ . This Lemma allows us to build a coupling that states that with high probability the effect of source coding using PCs with fixed frozen symbols will be the same as that of a transmission through the test channel. One important point is that by fixing the frozen symbols used for source coding at the relay we are also fixing the frozen symbols in the PC used over the virtual channel  $W_V$ . Using this Lemma we obtain the following corollary to Theorem 2:

**Corollary 3.** *If the supremum in (10) is taken over SI-DMC distributions  $p(y_q|y_{sr})$  that induce virtual SI-DMCs  $W_V(y_{sd}|y_q)$ , Theorem 2 holds for any choice of frozen symbols for source coding at the relay.*

*Proof:* Source, relay, and destination operate in the same way as in the proof of Theorem 2 with the difference that the frozen bits used for source compression at the relay are fixed to some  $\tilde{\mathbf{u}}_{F_Q}$ . The analysis of the error probability also follows similar lines. However, in this case it is convenient to explicitly include the compression error in the distribution  $P_S(s)$  over which we evaluate the probability of the event  $\mathcal{E}$ . We can then write  $P_S(s) = P_E^{\mathbf{u}_F}(\mathbf{e})P_{S_s|\mathbf{E}}(s_s|\mathbf{e})$  with  $S_s = S \setminus \mathbf{E}$  and use Lemma 5 to establish a coupling. The rest of the proof is identical to the one for Theorem 2.  $\blacksquare$

Note that the above corollary does not include the choice of frozen symbols for channel coding used by the source and by the relay for transmission over  $W_{RD}$ . In order to extend the result for any choice of these frozen symbols as well it is necessary to take into account the results for channel

coding [12] and verify that the vectors follow the appropriate distributions.

## VII. SIMULATIONS

We first consider PCs for DF relaying in a physically degraded relay channel (see (8)) with the following characteristics:  $W_{SR}(y_{sr}|x)$  and  $p(y_{sd}|y_{sr})$  are independent BSCs with crossover probabilities of 0.05 and 0.15, respectively. We study two different realizations of the relay-destination channel: an error-free channel with fixed capacity, and an independent BSC with crossover probability 0.1. We will only consider cases where the capacity of the relay channel equals that of the source-relay channel, i.e.,  $C_s^{PD} = I(W_{SR}) \approx 0.71$ . The relay-destination BSC has  $I(W_{RD}) \approx 0.53$ . Direct transmission without cooperation is limited by  $I(W_{SD}) \approx 0.31$ .

In Fig. 4 we show the BER ( $\Pr(\hat{U} \neq U)$ ) of our construction from Section V for different values of the transmission rates used by the source ( $R$ , coordinate axis) and the relay ( $R_{RD}$ , line face), and block lengths ( $N = 2^n$ , line marker) for the case of an error-free relay-destination channel (with capacity equal to  $R_{RD}$ ). As one would expect it is possible to lower the BER by increasing  $n$  (and hence the complexity and delay) and also by reducing  $R$  (i.e., the efficiency). Moreover, increasing  $R_{RD}$  also yields a lower BER. This is due to the fact that we are reducing the amount of information that has to be decoded from the direct link observation at no cost since relay-destination channel is error-free.

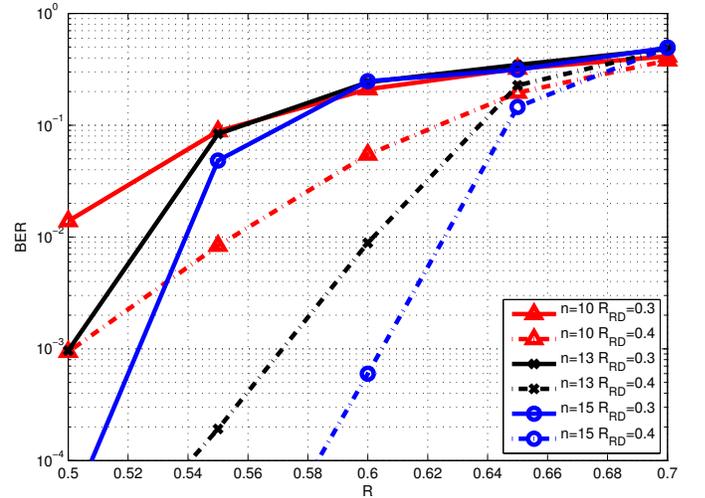


Fig. 4. Performance of DF relaying with PCs: error-free relay-destination channel.

In Fig. 5 we show the BER when the relay-destination channel is a BSC. In this case we observe that increasing  $R_{RD}$  does not always improve the performance. The reason for this is that now transmission from relay to destination takes place over a channel that introduces errors. At some point, the effect of these errors becomes the bottleneck of the system since the destination cannot recover the additional information conveyed by the relay that is necessary to decode the direct link observation. That is,  $\mathcal{E}_{RD}$  becomes the dominant error event in (11).

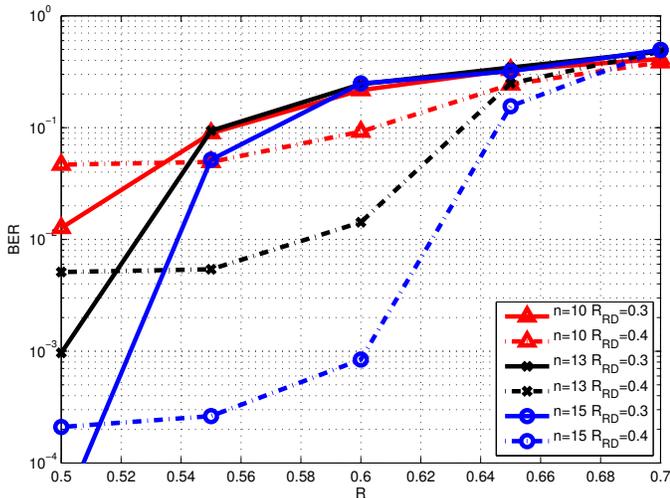


Fig. 5. Performance of DF relaying with PCs: binary symmetric relay-destination channel.

We now consider CF relaying. We model the source-relay and source-destination channels as independent BSCs with crossover probabilities of 0.1 and 0.05, respectively. The relay compresses its observation at a rate of  $R_Q = 0.8$  bits per sample using a PC designed using a BSC as the test channel. As before, we will first model the relay-destination channel as an error-free link with limited capacity equal to  $R_{RD}$  and then as an independent BSC, in this case with capacity equal to 0.85. Numerical evaluation of the limits in Theorem 2 results in  $I(X; Y_{SD} Y_Q) \approx 0.81$  and  $I(Y_{SR}; Y_Q | Y_{SD}) \approx 0.44$ . Without cooperation the scenario is limited by  $I(W_{SD}) \approx 0.71$ .

In Fig. 6 we show the BER of our construction from Section VI for the case of an error-free relay-destination channel. Again, performance improves with larger  $n$  and  $R_{RD}$  or with lower  $R$ . However, in this case we observe a saturation effect if only one of the rates is changed. For example, for  $R < 0.7$  the BER curves flatten out. This is due to the fact that in this region the errors in decoding the channel code used over the virtual channel start to dominate the error probability (i.e., the first term in (17)). A similar effect is observed if only  $R_{RD}$  is increased. For example, for  $R_{RD} > 0.65$  the reduction in BER is negligible. In this case the virtual channel becomes nearly error free and the error probability is dominated by the weakness of the PC used by the source (i.e., the second term in (17)).

In Fig. 7 we show the BER when the relay-destination channel is a BSC. As for DF relaying we observe that increasing  $R_{RD}$  may degrade the performance (in this case for  $R_{RD} > 0.65$ ). The reason is similar: a larger value of  $R_{RD}$  brings the PC used for transmission from relay to destination close to the capacity of the channel. At some point the destination cannot decode the relay contribution without errors. Using the wrong estimate  $\hat{Y}_Q$  implies error propagation when decoding the message transmitted by the source. This means a larger BER.

As a final remark we note that the constructions of PCs for relaying introduced here in spite of being optimal for  $N \rightarrow \infty$

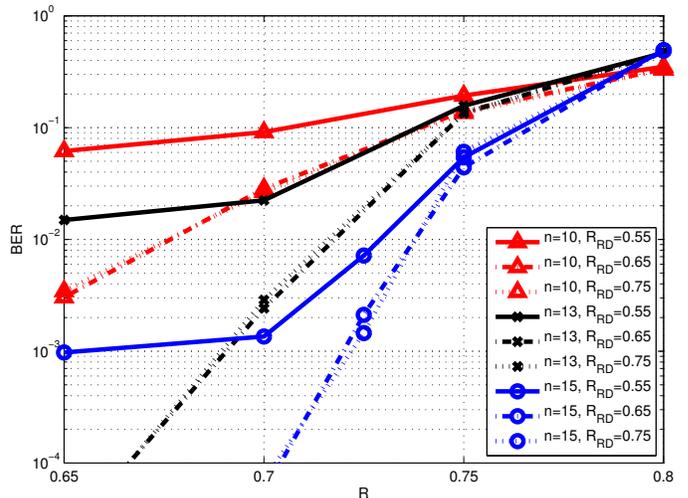


Fig. 6. Performance of CF relaying with PCs: error-free relay-destination channel.

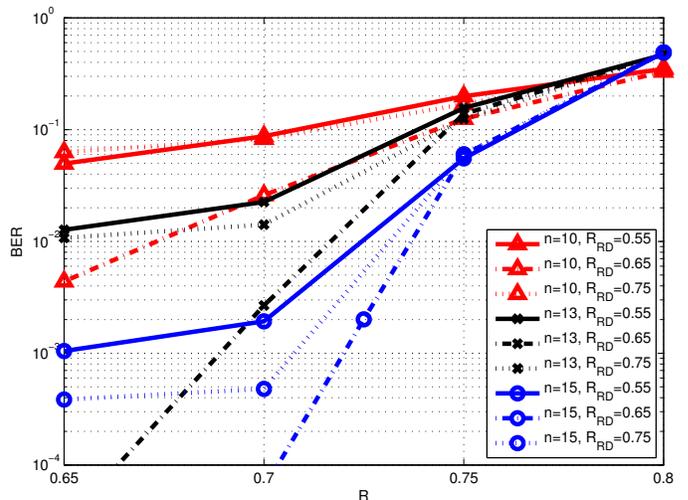


Fig. 7. Performance of CF relaying with PCs: binary symmetric relay-destination channel.

operate quite far away from these asymptotical limits when the block length is moderate. This behavior is common to all constructions of PCs for finite block lengths.

## VIII. CONCLUSION

In this paper we have shown that polar codes are suitable for decode-and-forward and compress-and-forward relaying in relay channels with orthogonal receivers. In the first case we have exploited the natural nesting of polar codes in stochastically degraded channels to extract the information present at the relay that complements the direct-link observation at the destination. In the second case we have employed a nested construction of polar codes for channel and source coding to adapt the quality of the description conveyed from the relay to the destination to the conditions of the channel while exploiting the side information available at the destination in the form of the direct-link observation. In both cases we have found scenarios where the strategies achieve the full capacity of the relay channel.

Simulation results validate our constructions of polar codes for relaying. Although these constructions are not directly applicable for moderate block lengths, they provide insights on how to implement the two relaying protocols with structured codes.

## REFERENCES

- [1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. in Applied Probability*, no. 3, pp. 120–154, 1971.
- [2] T. Cover and A.A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, pp. 572–584, Sep. 1979.
- [3] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [4] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [5] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
- [6] M. Valenti and B. Zhao, "Distributed Turbo codes: towards the capacity of the relay channel," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, vol. 1, Oct. 2003, pp. 322–326.
- [7] Z. Si, R. Thobaben, and M. Skoglund, "On distributed serially concatenated codes," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2009, pp. 653–657.
- [8] A. Chakrabarti, A. D. Baynast, A. Sabharwal, and B. Aazhang, "Low density parity check codes for the relay channel," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 280–291, Feb. 2007.
- [9] M. Uppal, Z. Liu, V. Stankovic, and Z. Xiong, "Compress-forward coding with BPSK modulation for the half-duplex Gaussian relay channel," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4467–4481, Nov. 2009.
- [10] R. Blasco-Serrano, R. Thobaben, and M. Skoglund, "Bandwidth efficient compress-and-forward relaying based on joint source-channel coding," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, Mar. 2011.
- [11] M. Wainwright and E. Martinian, "Low-density graph codes that are optimal for binning and coding with side information," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1061–1079, Mar. 2009.
- [12] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.
- [13] E. Şaşıoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Proc. IEEE Information Theory Workshop (ITW)*, Oct. 2009, pp. 144–148.
- [14] S. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Transactions on Information Theory*, pp. 1751–1768, Apr. 2010.
- [15] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Switzerland, 2009.
- [16] M. Karzand and E. Telatar, "Polar codes for  $Q$ -ary source coding," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2010, pp. 909–912.
- [17] E. Arıkan, "Source polarization," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2010, pp. 899–903.
- [18] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested polar codes for wiretap and relay channels," *IEEE Communications Letters*, vol. 14, no. 8, pp. 752–754, Aug. 2010.
- [19] E. Arıkan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Communications Letters*, vol. 12, no. 6, pp. 447–449, Jun. 2008.
- [20] N. Hussami, S. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2009, pp. 1488–1492.
- [21] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York: Cambridge University Press, 2008.
- [22] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Communications Letters*, vol. 13, no. 7, pp. 519–521, Jul. 2009.
- [23] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2009, pp. 1493–1495.
- [24] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Sep. 2007, pp. 129–135.
- [25] D. Aldous, "Random walks on finite groups and rapidly mixing Markov chains," *Séminaire de probabilités (Strasbourg)*. Springer-Verlag, vol. 17, pp. 243–297, 1983.



**Ricardo Blasco-Serrano** (S'05) received the M.Sc. degree on Telecommunication Engineering from the Technical University of Catalonia (UPC), Spain in 2007. Since 2007 he is with the Communication Theory laboratory of the School of Electrical Engineering at the KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include digital communications and information theory with focus on code design for cooperation and coordination in networks. He served as the Finance Vice Chair for the 2011 IEEE Swedish Communication

Technologies Workshop.



**Ragnar Thobaben** (M'07) received the Dipl.-Ing. degree (M.Sc.) in Electrical Engineering in 2001 and the Dr.-Ing. degree (Ph.D.) in Electrical Engineering in 2007 from the University of Kiel, Germany. In December 2006, Dr. Thobaben joined the Communication Theory laboratory at the KTH Royal Institute of Technology, Stockholm, Sweden, as a post-doctoral researcher, where he serves since July 2008 as an Assistant Professor. Dr. Thobaben's current research activities are dedicated to the design and analysis of coding and transmission schemes for

communication networks with a special focus on cognitive radio, cooperative communication, coordination, as well as physical-layer and network security. Dr. Thobaben serves currently on the technical program committees for the 2012 IEEE PIMRC and the 2012 ISWCS and as publicity chair for the 2012 International Symposium on Turbo Codes & Iterative Information Processing. He served as well as publicity chair for the 2011 IEEE Swedish Communication Technologies Workshop.



**Mattias Andersson** (S'07) received the M.Sc. degree in Engineering Physics from the KTH Royal Institute of Technology, Stockholm, Sweden in 2007. In 2007 he joined the Communication Theory laboratory of the School of Electrical Engineering at the KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include digital communications and information theory with focus on code design for physical layer security.



**Vishwamhar Rathi** obtained the Bachelor of Technology (B. Tech) in Electrical Engineering from the Indian Institute of Technology, Bombay (IITB), India. He obtained his PhD degree in the domain of sparse graph codes from Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland. In 2009-2010, he was a post doctoral researcher at KTH Royal Institute of Technology, Stockholm, Sweden. In 2010, he joined a wireless communication start up Icera Semiconductors which was acquired by Nvidia Corporation in 2011. He is currently a modem software engineer at Nvidia. His main research interests are information theory, coding theory, wireless communication, and graphical models.



**Mikael Skoglund** (S'93-M'97-SM'04) received the Ph.D. degree in 1997 from Chalmers University of Technology, Sweden. In 1997, he joined the KTH Royal Institute of Technology, Stockholm, Sweden, where he was appointed to the Chair in Communication Theory in 2003. At KTH, he heads the Communication Theory laboratory and he is the Assistant Dean for Electrical Engineering. Dr. Skoglund's research interests are in the theoretical aspects of wireless communications. He has worked on problems in source-channel coding, coding and transmission for wireless communications, Shannon theory and statistical signal processing. He has authored and co-authored more than 300 scientific papers in these areas, and he holds six patents. Dr. Skoglund has served on numerous technical program committees for IEEE conferences. During 2003–08 he was an associate editor with the IEEE Transactions on Communications and he is presently on the editorial board for IEEE Transactions on Information Theory.