

Polar Codes for Coordination in Cascade Networks

Ricardo Blasco-Serrano, Ragnar Thobaben, and Mikael Skoglund

KTH Royal Institute of Technology and ACCESS Linnaeus Centre
SE-100 44, Stockholm, Sweden

E-mail: {ricardo.blasco, ragnar.thobaben, mikael.skoglund}@ee.kth.se

Abstract—We consider coordination in cascade networks and construct sequences of polar codes that achieve any point in a special region of the empirical coordination capacity region. Our design combines elements of source coding to generate actions with the desired type with elements of channel coding to minimize the communication rate. Moreover, we bound the probability of malfunction of a polar code for empirical coordination. Possible generalizations and open problems are discussed.

I. INTRODUCTION

The limits of coordination in networks have been recently studied from a mathematical point of view (e.g. [1], [2]). Simple questions like how to measure coordination or how much communication is necessary to achieve a desired level of coordination have been posed and to some extent answered. The authors of [1] developed elements of a fundamental theory with two notions of coordination. In the first one, empirical coordination, the sequences of actions generated in the network must have a type that is close to a desired probability distribution. Empirical coordination is closely related to rate-distortion theory [3]. In fact any good code for rate-distortion is useful for empirical coordination and vice versa [1]. The second notion is that of strong coordination. Here the sequences of actions generated by the nodes must be statistically indistinguishable from those obtained by sampling a certain distribution.

In this work we construct sequences of polar codes (PCs) that achieve a special region of the empirical coordination capacity region for two-node and three-node cascade networks. PCs were introduced by Arkan as a method to achieve the capacity of any symmetric binary-input discrete memoryless channel (BI-DMC) [4]. Since then they have emerged as a powerful technique to develop achievability results in information theory with structured codes (as opposed to random coding). Korada and Urbanke established in [5] the optimality of PCs for lossy source coding of (symmetric) discrete memoryless sources (DMS) with binary reproduction alphabets. These results were later extended to non-binary channels and reproduction alphabets in [6] and [7], respectively.

Our constructions combine elements of PCs for source compression with PCs for channel coding. In addition, we show that, in line with [1], common randomness is not necessary for implementing PCs for empirical coordination, although it is useful in the proofs. We use the properties of PCs to extend the application of results on the rate of polarization [8] to PCs designed for empirical coordination. Finally, we discuss possible generalizations of our constructions as well as some open problems for future research.

This paper is organized as follows. In Section II we summarize basic results on empirical coordination and polar codes along with the notation. We analyze a simple two-node network in Section III. This serves as a building block for the cascade network, which is addressed in Section IV. We conclude our work in Section V with a discussion on the results of the paper as well as on some open problems.

II. PRELIMINARIES

A. Notation

Scalars are written using normal face x and vectors using bold face \mathbf{x} . The i^{th} element of a vector \mathbf{x} is denoted by x_i . For a given set of natural numbers F with size $|F|$, \mathbf{x}_F is shorthand for the subvector with elements whose positions belong to F . We use upper case letters for random variables (RVs) Y and lower case letters for their realizations y . The joint probability distribution on (X, Y) is denoted by $P_{X,Y}(x, y)$. For convenience we shall alternatively drop the subindices or the arguments whenever they are clear from the context. We follow the standard information-theoretic notation from [3].

B. Empirical Coordination

Consider the three-node network in Fig. 1. Node X observes a sequence of N external actions \mathbf{X} chosen independently and identically distributed (i.i.d.) according to P_X . Communication from Node X to Node Y and from Node Y to Node Z is possible at rates R_1 and R_2 (bits per action), respectively. We use these resources to have Node Y and Node Z generate sequences of actions \mathbf{Y} and \mathbf{Z} , respectively, with length N and joint type close to a desired probability function $P_{Y,Z|X}P_X$.

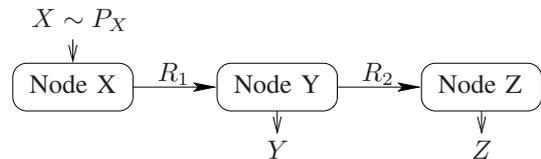


Fig. 1. Cascade network.

A $(2^{NR_1}, 2^{NR_2}, N)$ coordination code for the network in Fig. 1 consists of an encoding, a recoding, and two decoding functions (see [1]). All of them may have access to a source of common randomness (CR) independent of the external actions. Each coordination code induces a joint distribution on the actions $Q_{\mathbf{X},\mathbf{Y},\mathbf{Z}}$.

In this paper we are interested in the joint type of a tuple of action sequences $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ which is defined as

$$T_{\mathbf{x},\mathbf{y},\mathbf{z}}(x, y, z) = \frac{1}{N} \sum_{i=1}^N \mathbb{1} \{ (x_i, y_i, z_i) = (x, y, z) \} \quad (1)$$

for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, where $\mathbf{1}\{\cdot\}$ is the indicator function. In order to measure the distance between two probability distributions $P_{X,Y,Z}$ and $Q_{X,Y,Z}$ we consider their total variation, which is defined as

$$\|P_{X,Y,Z} - Q_{X,Y,Z}\| \triangleq \frac{1}{2} \sum_{x,y,z} |P(x,y,z) - Q(x,y,z)|.$$

We say that a triple $(R_1, R_2, P_{X,Y,Z})$ is achievable for empirical coordination if for any $\epsilon > 0$ there exists a sequence of $(2^{NR_1}, 2^{NR_2}, N)$ coordination codes and a choice of CR such that

$$\Pr(\|P_{X,Y,Z} - T_{\mathbf{X},\mathbf{Y},\mathbf{Z}}\| > \epsilon) < \epsilon$$

for sufficiently large N under the distribution induced by the codes. The empirical coordination capacity region, denoted by \mathcal{C}_{P_X} , is the closure of the set of achievable triples $(R_1, R_2, P_{X,Y,Z})$.

Due to the nature of PCs we restrict our attention to binary actions Y and Z (although no restriction is placed on X) and to choices of $P_{Y,Z|X}$ that induce uniform distributions on Y and Z for the given P_X . All the results in this paper are restricted to this subset of \mathcal{C}_{P_X} that we shall refer to as the symmetrical empirical coordination capacity region and denote by $\mathcal{C}_{P_X}^s$. Possible generalizations are discussed in Section V.

C. Polar Codes

Channel polarization is a method for transforming¹ N identical copies of a BI-DMC $P_{Y|X}$ into N distinct BI-DMCs $P^{(i)}(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i)$ ($i \in (1, \dots, N)$) with extremal properties in the sense that, for sufficiently large N , a fraction $I(X; Y)$ (for X uniformly distributed) of these synthetic channels is noise-free while the rest is virtually useless.

1) *Channel Coding*: Channel polarization leads naturally to a code construction that achieves the capacity of any symmetric BI-DMC. These codes are known as PCs and consist of two elements: an encoding matrix \mathbf{G}_N and a Successive Cancellation (SC) decoding algorithm. \mathbf{G}_N synthesizes the channels with extremal properties. Fixed (i.e. *frozen*) bits are put into the “bad” channels, which are those in the *frozen set* F defined as

$$F = \{i : Z(P^{(i)}) \geq \delta_N\} \quad (2)$$

for some $\delta_N > 0$, where $Z(P^{(i)})$ denotes the Bhattacharyya parameter (an upper bound on the error probability for uncoded transmission [4]) of the BI-DMC $P^{(i)}$. Information is transmitted at full rate through the rest of channels, i.e. those in the complement of the frozen set F^c . The SC decoding algorithm generates sequentially estimates \hat{u}_i for the information bits using the channel distribution $P^{(i)}(\mathbf{y}, \hat{\mathbf{u}}_0^{i-1}|u_i)$ and the previous estimates.

The following observation about PCs designed for degraded channels (as defined in [3]) will turn out to be important in the sequel.

¹We restrict our attention to PCs as introduced in [4] and therefore consider vectors of $N = 2^n$ actions (with $n \in \mathbb{N}$).

Lemma 1 (Lemma 21 in [5]). *Let $P_{Y|X}$ and $\tilde{P}_{Y|X}$ be two BI-DMCs with $\tilde{P}_{Y|X}$ degraded with respect to $P_{Y|X}$ (denoted by $\tilde{P}_{Y|X} \preceq P_{Y|X}$). Then for all $i \in \{1, \dots, N\}$ the synthesized channels satisfy $\tilde{P}^{(i)} \preceq P^{(i)}$ and $Z(\tilde{P}^{(i)}) \geq Z(P^{(i)})$. \square*

2) *Source Coding*: Korada and Urbanke showed in [5] that PCs are also optimal for lossy source compression of a DMS $X \sim P_X$ into Y . In this context the SC algorithm is used for compression while reconstruction is performed with the matrix \mathbf{G}_N . PCs for source compression are designed upon the test channel $P_{X|Y}$ that solves the minimization defining the (symmetric) rate-distortion function. Their fundamental property is that the distribution they induce $Q_{\mathbf{X},\mathbf{Y}}$ is arbitrarily close to the N -product design distribution $P_{\mathbf{X},\mathbf{Y}} = \prod P_{Y|X} P_X$, as summarized in the following Lemma:

Lemma 2 (Lemmas 2, 5, and Theorem 16 in [5]). *For any compression rate $R > I(Y; X)$ (Y uniformly distributed) we can choose $\delta_N > 0$ such that if the frozen set F is defined as*

$$F = \{i : Z(P^{(i)}) \geq 1 - \delta_N^2\} \quad (3)$$

then for any $0 < \beta < \frac{1}{2}$ we have

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

We will combine this result with the optimal coupling. For any two RVs $X \sim P_X$ and $Y \sim Q_Y$ defined on the same space their *optimal coupling* $C_{PQ}(x, y)$ is a probability distribution with marginals P_X and Q_Y and that satisfies [5], [9]

$$\Pr(X \neq Y) = \frac{1}{2} \sum_{x,y} |P_X(x) - Q_Y(y)|. \quad (4)$$

III. TWO-NODE NETWORK

In this Section we restrict our attention to the two-node network in Fig. 2. A $(2^{NR}, P_{Y|X})$ coordination code for this scenario consists only of an encoding and a decoding function.

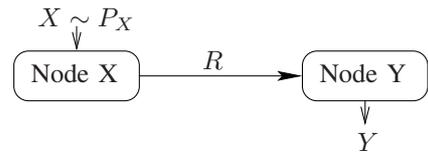


Fig. 2. Two-node network.

The empirical coordination capacity region for the network in Fig. 2 is given by [1]

$$\mathcal{C}_{P_X} = \{(R, P_{Y|X}) : R \geq I(Y; X)\}.$$

Theorem 1. *For any pair $(R, P_{Y|X}) \in \mathcal{C}_{P_X}^s$ (i.e. considering the aforementioned additional restrictions with respect to \mathcal{C}_{P_X}) there exists a sequence of PCs that achieves empirical coordination in the two-node network in Fig. 2. \square*

Proof: Let $P_{Y|X}$ be the desired distribution and let

$$P_{X|Y} = \frac{P_{X,Y}}{P_Y} = \frac{P_{Y|X} P_X}{P_Y}$$

where P_Y is the uniform distribution. Construct a PC for source compression upon $P_{X|Y}$ with the frozen set defined

as in (3). For any compression rate $R > I(Y; X)$ we know from the results on PCs for source coding in [5] that $R > \frac{|F^c|}{N}$ if N is sufficiently large. Using the CR both nodes generate $|F|$ i.i.d. random bits \mathbf{U}_F according to a uniform distribution. Node X compresses its observation \mathbf{X} using the SC encoding algorithm with frozen bits set to \mathbf{U}_F . The output of the algorithm is the vector \mathbf{U}_{F^c} which contains $|F^c|$ information bits. Therefore a rate of R bits per action suffices to communicate them to Node Y . This allows Node Y to generate the sequence $\mathbf{Y} = \mathbf{U}\mathbf{G}_N$. From Lemma 2 we know that the induced distribution is close to the desired N -product distribution $P_{\mathbf{X}, \mathbf{Y}} = \prod P_{Y|X} P_X$:

$$\|P_{\mathbf{X}, \mathbf{Y}} - Q_{\mathbf{X}, \mathbf{Y}}\| = \frac{1}{2} \sum_{\mathbf{x}, \mathbf{y}} |P(\mathbf{x}, \mathbf{y}) - Q(\mathbf{x}, \mathbf{y})| \leq O(2^{-N^\beta})$$

for any $0 < \beta < \frac{1}{2}$.

We now verify that this implies empirical coordination. Let $C_{PQ}(\mathbf{x}_P, \mathbf{y}_P, \mathbf{x}_Q, \mathbf{y}_Q)$ denote the optimal coupling between $P_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_P, \mathbf{y}_P)$ and $Q_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_Q, \mathbf{y}_Q)$. Consider the events

$$\begin{aligned} \mathcal{E} &= \{\|P_{X, Y} - T_{\mathbf{X}_Q, \mathbf{Y}_Q}\| > \epsilon\}, \\ \mathcal{E}_{XY} &= \{(\mathbf{X}_P, \mathbf{Y}_P) \neq (\mathbf{X}_Q, \mathbf{Y}_Q)\}, \end{aligned}$$

and their complements \mathcal{E}^c and \mathcal{E}_{XY}^c . Since C_{PQ} has marginal $Q_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_Q, \mathbf{y}_Q)$, \mathcal{E} evaluated over C_{PQ} is our event of interest. We upper bound its probability using the properties of the optimal coupling (see (4)). For any $0 < \beta < \frac{1}{2}$ we have that

$$\begin{aligned} \Pr(\mathcal{E}) &= \Pr(\mathcal{E}|\mathcal{E}_{XY}) \Pr(\mathcal{E}_{XY}) + \Pr(\mathcal{E}|\mathcal{E}_{XY}^c) \Pr(\mathcal{E}_{XY}^c) \\ &\leq \Pr(\mathcal{E}_{XY}) + \Pr(\mathcal{E}|\mathcal{E}_{XY}^c) \\ &\leq O(2^{-N^\beta}). \end{aligned} \quad (5)$$

The second term in (5) is upper bounded by the probability that a tuple $(\mathbf{X}, \mathbf{Y}) \sim P_{\mathbf{X}, \mathbf{Y}}$ has a type with total variation with respect to $P_{X, Y}$ larger than ϵ . From the properties of the Asymptotic Equipartition Property (AEP) (see [1], [3]) we know that this probability goes to zero exponentially fast with N . Therefore the bound for the first term in (5), which is due to the properties of the optimal coupling, dominates. Hence we conclude that the sequence of PCs achieves empirical coordination. ■

The above proof requires both nodes to generate the frozen bits using CR. The following corollary shows that there exists a fixed choice of the values of these bits such that the sequence of PCs achieves empirical coordination.

Corollary 1. *For any pair $(R, P_{Y|X}) \in \mathcal{C}_{P_X}^s$ there exists a sequence of PCs and a choice of frozen bits $\bar{\mathbf{u}}_F$ that achieves empirical coordination in the two-node network in Fig. 2. □*

Proof: The proof is given in the Appendix. ■

IV. CASCADE NETWORK

Consider again the network given in Fig. 1. Its empirical coordination capacity region is given by [1]

$$\mathcal{C}_{P_X} \triangleq \begin{cases} (R_1, R_2, P_{Y, Z|X}) : \\ R_1 \geq I(Y, Z; X) \\ R_2 \geq I(Z; X) \end{cases}.$$

Theorem 2. *For any triple $(R_1, R_2, P_{Y, Z|X}) \in \mathcal{C}_{P_X}^s$ (i.e. considering the aforementioned additional restrictions with respect to \mathcal{C}_{P_X}) there exists a sequence of PCs that achieves empirical coordination in the cascade network in Fig. 1. □*

Proof: Let $P_{Y, Z|X}$ be the desired distribution. Design a PC for source compression based on $P_{X|Z}$ (obtained from $P_{Y, Z|X} P_X$ by conditioning and marginalizing). For any compression rate $R > I(Z; X)$ (for Z uniformly distributed) and any $0 < \beta < \frac{1}{2}$ we know that

$$\|P_{\mathbf{X}, \mathbf{Z}} - Q_{\mathbf{X}, \mathbf{Z}}\| \leq O(2^{-N^\beta}) \quad (6)$$

where $P_{\mathbf{X}, \mathbf{Z}} = \prod P_{Z|X} P_X$ and $Q_{\mathbf{X}, \mathbf{Z}}$ is the distribution induced by the code. The values of the frozen bits are chosen i.i.d. according to a uniform distribution using CR so that they are also available at Nodes Y and Z . The vector resulting from source compression is sent to Node Y and then forwarded to Node Z . This allows both nodes to generate \mathbf{Z} . Therefore

$$\begin{aligned} R_1 &\geq R'_1 + I(Z; X) \\ R_2 &\geq I(Z; X) \end{aligned}$$

where R'_1 is the fraction of R_1 which is not yet used. Clearly we need R'_1 to be arbitrarily close to $I(Y, Z; X) - I(Z; X)$.

Now assume that $(\mathbf{X}, \mathbf{Z}) \sim P_{\mathbf{X}, \mathbf{Z}}$ (instead of $Q_{\mathbf{X}, \mathbf{Z}}$). Design a new PC for compression of (\mathbf{X}, \mathbf{Z}) into $\tilde{\mathbf{Y}}$ using the conditional probability $P_{X, Z|Y}$. Let $\tilde{Q}_{\tilde{\mathbf{Y}}|\mathbf{X}, \mathbf{Z}}$ be the distribution induced by this new PC. We know from Section III that for any compression rate $R > I(Y; X, Z)$ (for Y uniformly distributed) and any $0 < \beta < \frac{1}{2}$ we can choose the frozen set F_Q as in (3) so that for $P_{\mathbf{Y}|\mathbf{X}, \mathbf{Z}} = \prod P_{Y|X, Z}$

$$\|P_{\mathbf{Y}|\mathbf{X}, \mathbf{Z}} P_{\mathbf{X}, \mathbf{Z}} - \tilde{Q}_{\tilde{\mathbf{Y}}|\mathbf{X}, \mathbf{Z}} P_{\mathbf{X}, \mathbf{Z}}\| \leq O(2^{-N^\beta}). \quad (7)$$

Using the triangle inequality it is easy to show that (6) and (7) imply that for the true distribution on (\mathbf{X}, \mathbf{Z}) , i.e. $Q_{\mathbf{X}, \mathbf{Z}}$,

$$\|P_{\mathbf{Y}|\mathbf{X}, \mathbf{Z}} P_{\mathbf{X}, \mathbf{Z}} - \tilde{Q}_{\tilde{\mathbf{Y}}|\mathbf{X}, \mathbf{Z}} Q_{\mathbf{X}, \mathbf{Z}}\| \leq O(2^{-N^\beta}) \quad (8)$$

for any $0 < \beta < \frac{1}{2}$. As in Section III we use this result to build the optimal coupling $C_{P\tilde{Q}}$ between $P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}$ and $\tilde{Q}_{\mathbf{X}, \tilde{\mathbf{Y}}}$ which satisfies, for any $0 < \beta < \frac{1}{2}$,

$$\Pr(\mathcal{E}_{XYZ}) \triangleq \Pr((\mathbf{X}_P, \mathbf{Y}_P, \mathbf{Z}_P) \neq (\mathbf{X}_{\tilde{Q}}, \tilde{\mathbf{Y}}_{\tilde{Q}}, \mathbf{Z}_{\tilde{Q}})) \leq O(2^{-N^\beta}).$$

This allows us to evaluate the probability of the event

$$\mathcal{E}_1 \triangleq \left\{ \|P_{X, Y, Z} - T_{\mathbf{X}, \tilde{\mathbf{Y}}, \mathbf{Z}}\| > \epsilon \right\}$$

and bound it, for any $0 < \beta < \frac{1}{2}$, as

$$\Pr(\mathcal{E}_1) \leq \Pr(\mathcal{E}_{XYZ}) + \Pr(\mathcal{E}_1 | \mathcal{E}_{XYZ}^c) \leq O(2^{-N^\beta}).$$

This means that our construction will yield sequences of actions with the desired type with high probability. Now we only need to make $\tilde{\mathbf{Y}}$ available at Node Y . To accomplish this at the desired rate we take advantage of the fact that \mathbf{Z} is already available at Node Y and that it is indeed correlated with $\tilde{\mathbf{Y}}$. We model this correlation as transmitting $\tilde{\mathbf{Y}} = \mathbf{U}\mathbf{G}_N$ (i.e. a codeword from a PC) through the DMC $P_{Z|Y}$ and use the SC decoding algorithm to obtain \mathbf{Y} , an estimate of $\tilde{\mathbf{Y}}$.

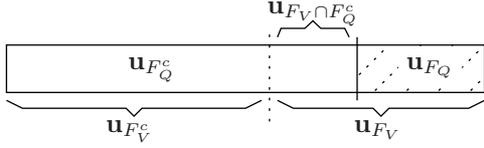


Fig. 3. Nesting of the frozen sets

Such a PC for transmission over $P_{Z|Y}$ would have a frozen set F_V as defined in (2). To obtain \mathbf{Y} the decoder needs to have access to the frozen bits \mathbf{U}_{F_V} . Moreover, since due to Lemma 1 we know that $F_Q \subseteq F_V$ for $\delta_N^2 < \delta_N$ and sufficiently large N , only the bits in the set $F_V \cap F_Q^c$ need to be conveyed (see Figure 3). This can be done at any rate

$$R'_1 \geq \frac{|F_V| - |F_Q|}{N} \geq I(Y; X, Z) - I(Y; Z),$$

or equivalently $R'_1 \geq I(Y, Z; X) - I(Z; X)$, as desired.

It is easy to show that this design will also work fine when $(\tilde{\mathbf{Y}}, \mathbf{Z}) \sim \tilde{Q}_{\tilde{\mathbf{Y}}, \mathbf{Z}}$ by considering the optimal coupling between $P_{\mathbf{Y}, \mathbf{Z}}$ and $\tilde{Q}_{\tilde{\mathbf{Y}}, \mathbf{Z}}$ and the fact that, for any $0 < \beta < \frac{1}{2}$,

$$\frac{1}{2} \sum_{\mathbf{y}, \mathbf{z}} \left| P_{\mathbf{Y}, \mathbf{Z}}(\mathbf{y}, \mathbf{z}) - \tilde{Q}_{\tilde{\mathbf{Y}}, \mathbf{Z}}(\mathbf{y}, \mathbf{z}) \right| \leq O(2^{-N^\beta})$$

which is derived from (8) with the triangle inequality. That is,

$$\Pr(\mathcal{E}_Y) \triangleq \Pr(\mathbf{Y} \neq \tilde{\mathbf{Y}}) \leq O(2^{-N^\beta})$$

for any $0 < \beta < \frac{1}{2}$.

Finally, we see that the probability of the event

$$\mathcal{E} = \{ \|P_{X, Y, Z} - T_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}\| > \epsilon \},$$

when evaluated over the distribution induced by the code

$$Q_{\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{Z}} = P_{\mathbf{X}} Q_{\mathbf{Z}|\mathbf{X}} \tilde{Q}_{\tilde{\mathbf{Y}}|\mathbf{X}, \mathbf{Z}} Q_{\mathbf{Y}|\tilde{\mathbf{Y}}},$$

where $Q_{\mathbf{Y}|\tilde{\mathbf{Y}}}$ accounts for the possible errors in reproducing $\tilde{\mathbf{Y}}$ at Node Y, is arbitrarily low. That is, for any $0 < \beta < \frac{1}{2}$

$$\begin{aligned} \Pr(\mathcal{E}) &\leq \Pr(\mathcal{E}_Y) + \Pr(\mathcal{E}|\mathcal{E}_Y^c) \\ &= \Pr(\mathcal{E}_Y) + \Pr(\mathcal{E}_1) \leq O(2^{-N^\beta}). \end{aligned}$$

Hence the sequence of PC achieves empirical coordination. ■

As for the two-node network, there exists a choice of the values of the frozen bits such that the sequence of PCs achieves empirical coordination.

Corollary 2. For any triple $(R_1, R_2, P_{Y, Z|X}) \in \mathcal{C}_{P_X}^s$ there exists a sequence of PCs and a choice of frozen bits that achieves empirical coordination in the network in Fig. 1. □

Proof: The proof is similar to that of Corollary 1. ■

V. DISCUSSION

The previous sections show that PCs are suitable for empirical coordination. However, a few questions remain open. First of all, our discussion has been restricted to the symmetrical empirical coordination capacity region, a portion of the region \mathcal{C}_{P_X} . Preliminary results show that the usual techniques for

extending the applicability of PCs (see [4]–[7]) may also be useful here to make more general statements as one would expect from [1, Section VII]. Moreover, we have only considered empirical coordination. No practical codes for strong coordination have been designed so far. The problem gets more involved due to the need for exploiting common randomness.

ACKNOWLEDGMENT

The authors would like to thank one anonymous reviewer for helpful comments.

APPENDIX

Proof of Corollary 1: Since the total variation is a bounded measure, Theorem 1 implies that there is a sequence of coordination codes such that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\mathbf{X}, \mathbf{Y}} \{ \|T_{\mathbf{X}, \mathbf{Y}} - P_{X, Y}\| \} = 0.$$

The expectation, which is taken with respect to the induced distribution, can be written as

$$\begin{aligned} \mathbb{E}_{\mathbf{X}, \mathbf{Y}} \{ \|T_{\mathbf{X}, \mathbf{Y}} - P_{X, Y}\| \} &= \mathbb{E}_{\mathbf{X}, \mathbf{U}} \{ \|T_{\mathbf{X}, \mathbf{U}} - P_{X, Y}\| \} \quad (9) \\ &= \sum_{\mathbf{x}, \mathbf{u}} Q(\mathbf{x}, \mathbf{u}) \|T_{\mathbf{x}, \mathbf{u}} - P_{X, Y}\| \quad (10) \end{aligned}$$

$$= \mathbb{E}_{\mathbf{U}_F} \{ \mathbb{E}_{\mathbf{X}, \mathbf{U}_{F^c}} \{ \|T_{\mathbf{X}, \mathbf{U}_{F^c}, \mathbf{U}_F} - P_{X, Y}\| \} \}. \quad (11)$$

In (9) and (10) we have used the fact that \mathbf{G}_N is a one-to-one mapping to modify slightly the definitions of $Q_{\mathbf{X}, \mathbf{Y}}$ and of the type function to depend on \mathbf{u} by setting $\mathbf{y} = \mathbf{u}\mathbf{G}_N$. The notation $T_{\mathbf{x}, \mathbf{u}_{F^c}, \mathbf{u}_F} = T_{\mathbf{x}, \mathbf{u}}$ in (11) simply makes explicit the fact that $\mathbf{u} = (\mathbf{u}_{F^c}, \mathbf{u}_F)$. For the outer expectation in (11) to go to zero with the block length there must be a choice of the values of the frozen bits $\tilde{\mathbf{u}}_F$ such that

$$\mathbb{E}_{\mathbf{X}, \mathbf{U}_{F^c}} \{ \|T_{\mathbf{X}, \mathbf{U}_{F^c}, \tilde{\mathbf{u}}_F} - P_{X, Y}\| \} \leq \mathbb{E}_{\mathbf{X}, \mathbf{U}} \{ \|T_{\mathbf{X}, \mathbf{U}} - P_{X, Y}\| \}.$$

Moreover, since convergence in the first mean implies convergence in probability, we conclude that there exists a sequence of PCs for coordination and a choice of the frozen bits that achieves empirical coordination. ■

REFERENCES

- [1] P. Cuff, H. Permuter, and T. Cover, “Coordination capacity,” *IEEE Transactions on Information Theory*, vol. 56, no. 9, Sep. 2010.
- [2] G. Kramer and S. A. Savari, “Communicating probability distributions,” *IEEE Transactions on Information Theory*, vol. 53, no. 2, Feb. 2007.
- [3] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: John Wiley, 2006.
- [4] 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, Jul. 2009.
- [5] S. Korada and R. Urbanke, “Polar codes are optimal for lossy source coding,” *IEEE Transactions on Information Theory*, Apr. 2010.
- [6] E. Şaşıoğlu, E. Telatar, and E. Arıkan, “Polarization for arbitrary discrete memoryless channels,” in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2009.
- [7] M. Karzand and E. Telatar, “Polar codes for Q -ary source coding,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2010.
- [8] E. Arıkan and E. Telatar, “On the rate of channel polarization,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2009.
- [9] D. Aldous, “Random walks on finite groups and rapidly mixing Markov chains,” *Séminaire de probabilités (Strasbourg)*. Springer-Verlag, 1983.