Joint Coordination-Channel Coding for Strong Coordination over Noisy Channels Based on Polar Codes

Sarah A. Obead, Jörg Kliewer
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, New Jersey 07102
Email:sao23@njit.edu, jkliewer@njit.edu

Badri N. Vellambi
Research School of Computer Science
Australian National University
Acton, Australia 2601
Email: badri.n.vellambi@ieee.org

Abstract—We construct a joint coordination-channel polar coding scheme for strong coordination of actions between two agents X and Y, which communicate over a discrete memoryless channel (DMC) such that the joint distribution of actions follows a prescribed probability distribution. We show that polar codes are able to achieve our previously established inner bound to the strong noisy coordination capacity region and thus provide a constructive alternative to a random coding proof. Our polar coding scheme also offers a constructive solution to a channel simulation problem where a DMC and shared randomness are together employed to simulate another DMC. In particular, our proposed solution is able to utilize the randomness of the DMC to reduce the amount of local randomness required to generate the sequence of actions at agent Y. By leveraging our earlier random coding results for this problem, we conclude that the proposed joint coordination-channel coding scheme strictly outperforms a separate scheme in terms of achievable communication rate for the same amount of injected randomness into both systems.

I. INTRODUCTION

A fundamental problem in decentralized networks is to coordinate activities of different agents with the goal of reaching a state of agreement. Such a problem arises in a multitude of applications, including networks of autonomous robots, smart traffic control, and distributed computing problems. For such applications, coordination is understood to be the ability to arrive at a prescribed joint distribution of actions at all agents in the network. In information theory, two different notions of coordination are explored: (i) empirical coordination, which only requires the normalized histogram of induced joint actions to approach a desired target distribution, and (ii) strong coordination, where the sequence of induced joint actions must be statistically close (i.e., nearly indistinguishable) from a given target probability mass function (pmf).

A significant amount of work has been devoted to finding the capacity regions of various coordination problems based on both empirical and strong coordination [1]–[10], where [4], [6]–[8], [10] focus on small to moderate network settings.

While all these works address the noiseless case, coordination over noisy channels has received only little attention in the literature so far. However, notable exceptions are [11]–[13]. For example, in [11] joint empirical coordination of the channel inputs/outputs of a noisy communication channel with source and reproduction sequences is considered. Also, in [12] the notion of strong coordination is used to simulate a discrete memoryless channel via another channel. Recently, [13] explored the strong coordination variant of the problem investigated in [11].

As an alternative to the impracticalities of random coding, solutions for empirical and strong coordination problems have been proposed based on low-complexity polar-codes introduced by Arikan [14], [15]. For example, polar coding for strong point-to-point coordination is addressed in [16], [17], and empirical coordination for cascade networks in [18], respectively. The only existing design of polar codes for the noisy empirical coordination case [19] is based on the joint source-channel coordination approach in [11]. However, to the best of our knowledge, polar code designs for noisy *strong* coordination have not been proposed in the literature.

In this work we consider the point-to-point coordination setup depicted in Fig. 1 where only source and reproduction sequences are coordinated via a suitable polar coding scheme over DMCs. In particular, we design an explicit low-complexity nested polar coding scheme for strong coordination over noisy channels that achieves the inner bound of the two-node network capacity region of our earlier work [20]. In this work, we show that a joint coordination-channel coding scheme is able to strictly outperform a separation-based scheme in terms of achievable communication rate if the same amount of randomness is injected into the system. Note that our proposed joint coordination-channel polar coding scheme employs nested codebooks similar to the polar codes for the broadcast channel [21]. Further, our polar coding scheme also offers a constructive solution to a channel simulation problem where a DMC is employed to simulate another DMC in the presence of shared randomness [12].

The remainder of the paper is organized as follows. Sec-

tion II introduces the notation, the model under investigation, and a random coding construction. Section III provides our proposed joint coordination-channel coding design and a proof to show that this design achieves the random coding inner bound.

II. PROBLEM STATEMENT

A. Notation

Let $N \triangleq 2^n, n \in \mathbb{N}$. We denote the source polarization transform as $G_n = RF^{\otimes n}$, where R is the bit-reversal mapping defined in [14], $F = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, and $F^{\otimes n}$ denotes the n-th Kronecker power of F. Given $X^{1:N} \triangleq (X^1, X^2, \ldots, X^N)$ and $\mathcal{A} \subset [\![1, N]\!]$, we let $X^N[\mathcal{A}]$ denote the components X^i such that $i \in \mathcal{A}$. Given two distributions $P_X(x)$ and $Q_X(x)$ defined over an alphabet $\mathcal{X}$, we let $\mathbb{D}(P_X(x)\|Q_X(x))$ and $\|P_X(x) - Q_X(x)\|_{TV}$ denote the Kullback-Leibler (KL) divergence and the total variation, respectively. Given a pmf $P_X(x)$ we let $\min^*(P_X) = \min\{P_X(x) : P_X(x) > 0\}$.

B. System Model

The point-to-point coordination setup considered in this work is depicted in Fig. 1. Node X receives a sequence of actions $X^N \in \mathcal{X}^N$ specified by nature where X^N is i.i.d. according to a pmf p_X. Both nodes have access to shared randomness J at rate R_o bits/action from a common source, and each node possesses local randomness M_ℓ at rate $\rho_\ell, \ell = 1, 2$.
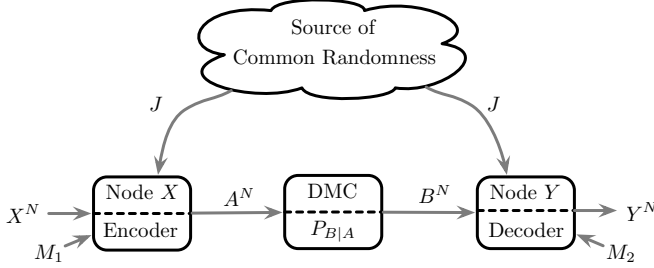

Fig. 1. Point-to-point strong coordination over a DMC.

We wish to communicate a codeword A^N corresponding to the coordination message over the rate-limited DMC $P_{B|A}$ to Node Y. The *codeword* A^N is constructed based on the input action sequence X^N, the local randomness M_1 at Node X, and the common randomness J. Node Y generates a sequence of actions $Y^N \in \mathcal{Y}^N$ based on the received codeword B^N, common randomness J, and local randomness M_2.

By assumption, the common randomness is independent of the action specified at Node X. A strong coordination coding scheme with rates $(R_c, R_o, \rho_1, \rho_2)$ is deemed *achievable* if for each $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that the joint pmf of actions $\tilde{P}_{X^N, Y^N}$ induced by this scheme and the N i.i.d. copies of desired joint pmf $(X, Y) \sim q_{XY}, Q_{X^N Y^N}$, are *close* in total variation, i.e.,

$$\|\tilde{P}_{X^N Y^N} - Q_{X^N Y^N}\|_{TV} < \epsilon. \tag{1}$$

C. Random Coding Construction

Consider auxiliary random variables $A \in \mathcal{A}$ and $C \in \mathcal{C}$ with $(A, C) \sim P_{AC}$ be jointly correlated with (X, Y) as $P_{XYABC} = P_{AC}P_{X|AC}P_{B|A}P_{Y|BC}$. The joint

strong coordination-channel random code with parameters (R_c, R_o, R_a, N) [20], where $\mathcal{I} \triangleq [\![1, 2^{NR_c}]\!]$, $\mathcal{J} \triangleq [\![1, 2^{NR_o}]\!]$, and $\mathcal{K} \triangleq [\![1, 2^{NR_a}]\!]$, consists of

1) Nested codebooks: A codebook $\mathcal{C}$ of size $2^{N(R_o+R_c)}$ is generated i.i.d. according to a pmf P_C, i.e., $C_{ij}^N \sim \prod_{l=1}^N P_C(\cdot)$ for all $(i, j) \in \mathcal{I} \times \mathcal{J}$. A codebook $\mathcal{A}$ is generated by randomly selecting $A_{ijk}^N \sim \prod_{l=1}^N P_{A|C}(\cdot|C_{ij}^N)$ for all $(i, j, k) \in \mathcal{I} \times \mathcal{J} \times \mathcal{K}$.
2) Encoding functions:
$C^N : [\![1, 2^{NR_c}]\!] \times [\![1, 2^{NR_o}]\!] \to \mathcal{C}^N$,
$A^N : [\![1, 2^{NR_c}]\!] \times [\![1, 2^{NR_o}]\!] \times [\![1, 2^{NR_a}]\!] \to \mathcal{A}^N$.
3) The indices I, J, K are independent and uniformly distributed over $\mathcal{I}$, $\mathcal{J}$, and $\mathcal{K}$, respectively. These indices select the pair of codewords C_{IJ}^N and A_{IJK}^N from codebooks $\mathcal{C}$ and $\mathcal{A}$.
4) The selected codeword A_{IJK}^N is sent through the communication DMC $P_{B|A}$, whose output B^N is used to decode codeword $C_{\hat{I}J}^N$, and both are then passed through a DMC $P_{Y|BC}$ to obtain Y^N.

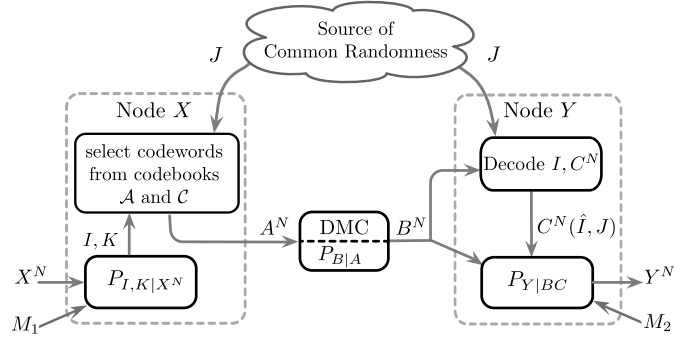The corresponding scheme is displayed in Fig. 2.

Fig. 2. Joint strong coordination-channel coding scheme.

The following theorem provides the inner bound for strong coordination region achieved by such joint coordination-channel code.

Theorem 1. *(Strong coordination inner bound [20]) A tuple (R_o, ρ_1, ρ_2) is achievable for the strong noisy communication setup in Fig. 1 if for some $R_a, R_c \geq 0$,*

$$R_a + R_o + R_c > I(XY; AC), \tag{2a}$$
$$R_o + R_c > I(XY; C), \tag{2b}$$
$$R_a + R_c > I(X; AC), \tag{2c}$$
$$R_c > I(X; C), \tag{2d}$$
$$R_c < I(B; C), \tag{2e}$$
$$\rho_1 > R_a + R_c - I(X; AC), \tag{2f}$$
$$\rho_2 > H(Y|BC). \tag{2g}$$

The underlying proofs and details of the coding mechanism for this joint coordination-channel coding scheme for noisy strong coordination are based on a complex channel resolvability framework [20]. Channel resolvability has been successfully used to study different strong coordination problems due to its ability to approximate channel output statistics with random codebooks [22]. We now propose a scheme

based on polar codes that achieves the inner bound stated by Theorem 1 for the strong coordination region as follows.

III. NESTED POLAR CODE FOR STRONG COORDINATION OVER NOISY CHANNELS

Since the proposed joint coordination-channel coding scheme is based on a channel resolvability framework, we adopt the channel resolvability-based polar construction for noise-free strong coordination [17] in combination with polar coding for the degraded broadcast channel [21].

A. Coding Scheme

Consider the random variables $X, Y, A, B, C, \widehat{C}$ distributed according to $Q_{XYABC\widehat{C}}$ over $\mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B} \times \mathcal{C}$ such that $X - (A,C) - (B,\widehat{C}) - Y$. Assume that $|A| = 2$ and the distribution Q_{XY} is achievable with $|C| = 2$ [†]. Let $N \triangleq 2^n$. We describe the polar coding scheme as follows:
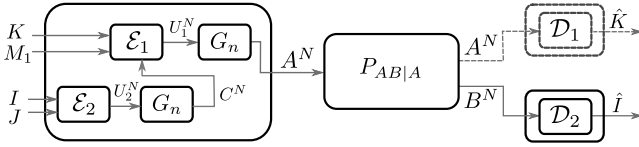

Fig. 3. Block diagram of the superposition polar code.

Consider a 2-user *physically* degraded discrete memoryless broadcast channel (DM-BC) $P_{AB|A}$ in Fig. 3 where A denotes the channel input and A, B denote the output to the first and second receiver, respectively. In particular, the channel DMC $P_{B|A}$ is physically degraded with respect to the *perfect* channel $P_{A|A}$ (i.e., $P_{A|A} \succ P_{B|A}$), leading to the Markov chain $A - A - B$. We construct the nested polar coding scheme in a similar fashion as in [21] as this mimics the nesting of the codebooks $\mathcal{C}$ and $\mathcal{A}$ in Step 1) of the random coding construction in Section II-C. Here, the second (weaker) user is able to recover its intended message I, while the first (stronger) user is able to recover both messages K and I. Let C be the auxiliary random variable (cloud center) required for superposition coding over the DM-BC leading to the Markov chain $C - A - (A,B)$. As a result, the channel $P_{B|C}$ is also degraded with respect to $P_{A|C}$ (i.e., $P_{A|C} \succ P_{B|C}$) [21, Lemma 3]. Let V be a matrix of the selected codewords A^N and C^N as

$$V \triangleq \begin{bmatrix} A^N \\ C^N \end{bmatrix}. \tag{3}$$

Now, apply the polar linear transformation G_n as

$$U \triangleq \begin{bmatrix} U_1^N \\ U_2^N \end{bmatrix} = VG_n. \tag{4}$$

First, consider $C^N \triangleq U_2^N G_n$ from (3) and (4) where U_2^N is generated by the second encoder $\mathcal{E}_2$ in Fig. 3. For $\beta < \frac{1}{2}$

[†]For the sake of exposition, we only focus on the set of joint distributions over $\mathcal{X} \times \mathcal{Y}$ that are achievable with binary auxiliary random variables C, A, and over a binary-input DMC. The scheme can be generalized to non-binary C, A with non-binary polar codes in a straightforward way [23].

and $\delta_N \triangleq 2^{-N^\beta}$ we define the very high and high entropy sets

$$\mathcal{V}_C \triangleq \{i \in [\![1, N]\!] : H(U_2^i|U_2^{1:i-1}) > 1 - \delta_N\},$$
$$\mathcal{V}_{C|X} \triangleq \{i \in [\![1, N]\!] : H(U_2^i|U_2^{1:i-1} X^N) > 1 - \delta_N\},$$
$$\mathcal{V}_{C|XY} \triangleq \{i \in [\![1, N]\!] : H(U_2^i|U_2^{1:i-1} X^N Y^N) > 1 - \delta_N\},$$
$$\mathcal{H}_{C|B} \triangleq \{i \in [\![1, N]\!] : H(U_2^i|U_2^{1:i-1} B^N) > \delta_N\},$$
$$\mathcal{H}_{C|A} \triangleq \{i \in [\![1, N]\!] : H(U_2^i|U_2^{1:i-1} A^N) > \delta_N\}, \tag{5}$$

which by [24, Lemma 7] satisfy

$$\lim_{N\to\infty} \frac{|\mathcal{V}_C|}{N} = H(C), \qquad \lim_{N\to\infty} \frac{|\mathcal{V}_{C|X}|}{N} = H(C|X),$$
$$\lim_{N\to\infty} \frac{|\mathcal{V}_{C|XY}|}{N} = H(C|XY), \qquad \lim_{N\to\infty} \frac{|\mathcal{H}_{C|B}|}{N} = H(C|B),$$
$$\lim_{N\to\infty} \frac{|\mathcal{H}_{C|A}|}{N} = H(C|A).$$

These sets are illustrated in Fig. 4. Note that the set $\mathcal{H}_{C|B}$ indicates the noisy bits of the DMC $P_{B|C}$ (i.e., the unrecoverable bits of the codeword C^N intended for the weaker user in the DM-BC setup in Fig. 3) and is in general not aligned with other sets. Let

$$\mathcal{L}_1 \triangleq \mathcal{V}_C \setminus \mathcal{H}_{C|A}, \qquad \mathcal{L}_2 \triangleq \mathcal{V}_C \setminus \mathcal{H}_{C|B},$$

where the set $\mathcal{H}_{C|A}$ indicates the noisy bits of the DMC $P_{A|C}$ (i.e., the unrecoverable bits of the codeword C^N intended for the stronger user). From the relation $P_{A|C} \succ P_{B|C}$ we obtain $\mathcal{H}^c_{C|B} \subseteq \mathcal{H}^c_{C|A}$. This ensures that the polarization indices are guaranteed to be aligned (i.e., $\mathcal{L}_2 \subseteq \mathcal{L}_1$) [25], [21, Lemma 4]. As a consequence, the bits decodable by the weaker user are also decodable by the stronger user.

Now, consider $A^N \triangleq U_1^N G_n$ (see (3) and (4)), where U_1^N is generated by the first encoder $\mathcal{E}_1$ with C^N as a side information as seen in Fig. 3. We define the very high entropy sets illustrated in Fig. 5 as
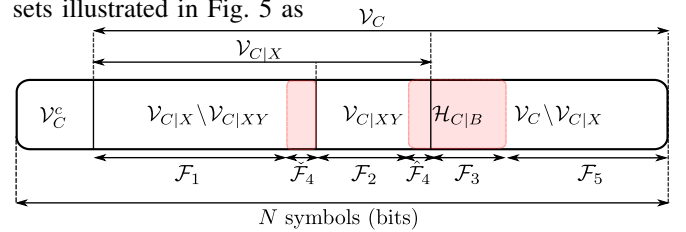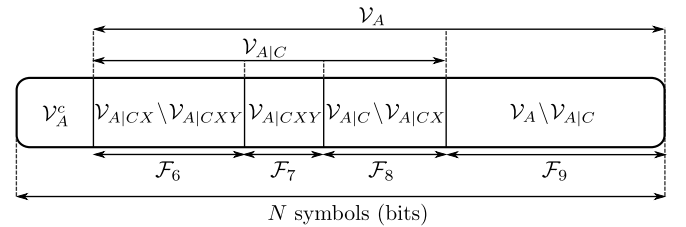

Fig. 4. Index sets for codeword C.

Fig. 5. Index sets for codeword A.

$$\mathcal{V}_A \triangleq \{i \in [\![1, N]\!] : H(U_1^i|U_1^{1:i-1}) > 1 - \delta_N\},$$
$$\mathcal{V}_{A|C} \triangleq \{i \in [\![1, N]\!] : H(U_1^i|U_1^{1:i-1} C^N) > 1 - \delta_N\},$$
$$\mathcal{V}_{A|CX} \triangleq \{i \in [\![1, N]\!] : H(U_1^i|U_1^{1:i-1} C^N X^N) > 1 - \delta_N\},$$
$$\mathcal{V}_{A|CXY} \triangleq \{i \in [\![1, N]\!] : H(U_1^i|U_1^{1:i-1} C^N X^N Y^N) > 1 - \delta_N\} \tag{6}$$

satisfying

$$\lim_{N\to\infty}\frac{|\mathcal{V}_A|}{N}=H(A), \qquad \lim_{N\to\infty}\frac{|\mathcal{V}_{A|CX}|}{N}=H(A|CX),$$

$$\lim_{N\to\infty}\frac{|\mathcal{V}_{A|C}|}{N}=H(A|C), \quad \lim_{N\to\infty}\frac{|\mathcal{V}_{A|CXY}|}{N}=H(A|CXY).$$

Note that, in contrast to Fig. 4, here there is no channel dependent set overlapping with all other sets as $P_{A|A}$ is a noiseless channel with rate $H(A)$ and hence $\mathcal{H}_{A|A}=\emptyset$.

Accordingly, in terms of the polarization sets in (5) and (6) we define the sets combining channel resolvability for strong coordination and broadcast channel construction

$$\mathcal{F}_1 \triangleq (\mathcal{V}_{C|X}\setminus\mathcal{V}_{C|XY})\cap\mathcal{H}_{C|B}^c,$$
$$\mathcal{F}_2 \triangleq \mathcal{V}_{C|XY}\cap\mathcal{H}_{C|B}^c,$$
$$\mathcal{F}_3 \triangleq \mathcal{V}_{C|X}^c\cap\mathcal{H}_{C|B}=\mathcal{H}_{C|B}\setminus\mathcal{H}_{C|BX},$$
$$\mathcal{F}_4 \triangleq \mathcal{V}_{C|X}\cap\mathcal{H}_{C|B}=\mathcal{H}_{C|BX},$$
$$\hat{\mathcal{F}}_4 \triangleq \mathcal{H}_{C|BXY},$$
$$\check{\mathcal{F}}_4 \triangleq \mathcal{H}_{C|BX}\setminus\mathcal{H}_{C|BXY},$$
$$\mathcal{F}_5 \triangleq (\mathcal{V}_C\setminus\mathcal{V}_{C|X})\cap\mathcal{H}_{C|B}^c,$$
$$\mathcal{F}_6 \triangleq \mathcal{V}_{A|CX}\setminus\mathcal{V}_{A|CXY},$$
$$\mathcal{F}_7 \triangleq \mathcal{V}_{A|CXY},$$
$$\mathcal{F}_8 \triangleq \mathcal{V}_{A|C}\setminus\mathcal{V}_{A|CX},$$
$$\mathcal{F}_9 \triangleq \mathcal{V}_A\setminus\mathcal{V}_{A|C}.$$

Finally, with $Y^N\triangleq T^N G_n$, we define the very high entropy set:

$$\mathcal{V}_{Y|BC}\triangleq\{i\in[\![1,N]\!]:H(T^i|T^{1:i-1}B^N C^N)>\log|\mathcal{Y}|-\delta_N\}, \quad (7)$$

satisfying

$$\lim_{N\to\infty}\frac{|\mathcal{V}_{Y|BC}|}{N}=H(Y|BC).$$

This set is useful for expressing the randomized generation of Y^N via simulating the channel $P_{Y|BC}$ in Fig. 2 as a source polarization operation [17]. We now proceed to describe the encoding and decoding algorithms.

1) Encoding: The encoding protocol described in Algorithm 1 is performed over $k\in\mathbb{N}$ blocks of length N. Since for strong coordination the goal is to approximate a target joint distribution with a minimum amount of randomness, the encoding scheme performs channel resolvability while reusing a fraction of the common randomness over several blocks (i.e., randomness recycling) as in [17]. However, since the communication is over a noisy channel, the encoding scheme also considers a block chaining construction to mitigate the channel noise influence as in [19], [24]–[26].

More precisely, as demonstrated in Fig. 2, we are interested in successfully recovering the message I that is intended for the weak user channel given by $P_{B|A}$ in Fig. 3. However, the challenge is to communicate the set $\mathcal{F}_3$ that includes bits of the message I that are corrupted by the channel noise. This suggests that we apply a variation of block chaining only at encoder $\mathcal{E}_2$ generating the codeword C^N as follows (see Fig. 6). At encoder $\mathcal{E}_2$, the set $\mathcal{F}_3$ of block $i\in[\![1,k]\!]$ is embedded in the reliably decodable bits of $\mathcal{F}_1\cup\mathcal{F}_2$ of
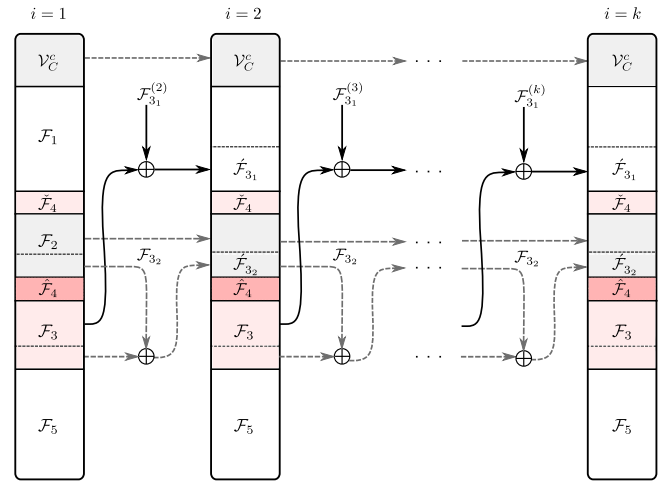

Fig. 6. Chaining construction for block encoding.

the following block $i+1$. This is possible by following the decodability constraint (see (2d), (2e) of Theorem 1) that ensures that the size of the set $\mathcal{F}_3$ is smaller than the combined size of the sets $\mathcal{F}_1$ and $\mathcal{F}_2$ [19]. However, since these sets originally contain uniformly distributed common randomness J [17], the bits of $\mathcal{F}_3$ can be embedded while maintaining the uniformity of the randomness by taking advantage of the Crypto Lemma [27, Lemma 2]. Then, to ensure that $\mathcal{F}_3$ is equally distributed over $\mathcal{F}_1\cup\mathcal{F}_2$, $\mathcal{F}_3$ is partitioned according to the ratio between $|\mathcal{F}_1|$ and $|\mathcal{F}_2|$. To utilize the Crypto Lemma, we introduce $\mathcal{F}_{3_2}$ and $\mathcal{F}_{3_1}^{(i)}$, which represent uniformly distributed common randomness used to randomize the information bits of $\mathcal{F}_3$. The difference is that $\mathcal{F}_{3_2}$, as $\mathcal{F}_2$, represents a fraction of common randomness that can be reused over k blocks whereas a realization of the randomness in $\mathcal{F}_{3_1}^{(i)}$ needs to be provided in each new block. Note that, as visualized in Fig. 6, both the subsets $\acute{\mathcal{F}}_{3_1}\subset\mathcal{F}_1$ and $\acute{\mathcal{F}}_{3_2}\subset\mathcal{F}_2$ represent the resulting uniformly distributed bits of $\mathcal{F}_3$ of the previous block, where $|\acute{\mathcal{F}}_{3_1}|=|\mathcal{F}_{3_1}|$ and $|\acute{\mathcal{F}}_{3_2}|=|\mathcal{F}_{3_2}|$. Finally, in an additional block $k+1$ we use a good channel code to reliably transmit the set $\mathcal{F}_3$ of the last block k.

2) Decoding: The decoder is described in Algorithm 2. Recall that we are only interested in the message $\hat{I}$ intended for the weak user channel given by $P_{B|A}$ in Figure 3. As a result, we only state the decoding protocol at $\mathcal{D}_2$ that recovers the codeword $\widehat{C}^N$. Note that the decoding is done in reverse order after receiving the extra $k+1$ block containing the bits of set $\mathcal{F}_3$ of the last block k. In particular, in each block $i\in[1,k-1]$ the bits in $\mathcal{F}_3$ are obtained by successfully recovering the bits in both $\mathcal{F}_1$ and $\mathcal{F}_2$ in block $i+1$.

B. Scheme Analysis

We now provide an analysis of the coding scheme of Section III. The analysis is based on KL divergence which upper bounds the total variation in (1) by Pinsker's inequality. We start the analysis with a set of sequential lemmas. In particular, Lemma 1 is useful to show in Lemma 2 that the strong coordination scheme based on channel resolvability

Algorithm 1: Encoding algorithm at Node X for strong coordination

---

Input: $X_{1:k}^N$, uniformly distributed local randomness bits $M_{1_{1:k}}$ of the size $k|\mathcal{F}_6|$, common randomness bits $\bar{J} = (\bar{J}_1, \bar{J}_2)$ of sizes $|\mathcal{F}_2 \cup \hat{\mathcal{F}}_4|$, and $|\mathcal{F}_7|$, respectively, and $J_{1:k}$ of size $k|\check{\mathcal{F}}_4 \cup \mathcal{F}_1|$ shared with Node Y.
Output: $\widetilde{A}_{1:k}^N$
1. **for** $i = 2, \ldots, k$ **do**
2. $\mathcal{E}_2$ in Fig. 3 constructs $\widetilde{U}_{2_i}^N$ bit-by-bit as follows:
   if $i = 1$ **then**
   - $\widetilde{U}_{2_i}^N[\mathcal{F}_1 \cup \check{\mathcal{F}}_4] \leftarrow J_i$
   - $\widecheck{U}_{2_i}^N[\mathcal{F}_2 \cup \hat{\mathcal{F}}_4] \leftarrow \bar{J}_1$

   else
   - Let $\mathcal{F}_{3_1}^{(i)}$, $\mathcal{F}_{3_2}$ be sets of the size $(|\mathcal{F}_m| \times |\mathcal{F}_3|)/(|\mathcal{F}_1| + |\mathcal{F}_2|)$ for $m \in \{1, 2\}$.
   - $\left(\widetilde{U}_{2_i}^N[(\mathcal{F}_1 \setminus \acute{\mathcal{F}}_{3_1}) \cup \check{\mathcal{F}}_4], \mathcal{F}_{3_1}^{(i)}\right) \leftarrow J_i$
   - $\left(\widetilde{U}_{2_i}^N[(\mathcal{F}_2 \setminus \acute{\mathcal{F}}_{3_2}) \cup \hat{\mathcal{F}}_4], \mathcal{F}_{3_2}\right) \leftarrow \bar{J}_1$
   - $\widetilde{U}_{2_i}^N[\acute{\mathcal{F}}_{3_1}] \leftarrow \widetilde{U}_{2_{i-1}}^N[\mathcal{F}_3 \setminus \mathcal{F}_{3_2}] \oplus \mathcal{F}_{3_1}^{(i)}$
   - $\widetilde{U}_{2_i}^N[\acute{\mathcal{F}}_{3_2}] \leftarrow \widetilde{U}_{2_{i-1}}^N[\mathcal{F}_3 \setminus \mathcal{F}_{3_1}] \oplus \mathcal{F}_{3_2}$

   end
   - Given X_i^N, successively draw the remaining components of $\widetilde{U}_{2_i}^N$ according to $\tilde{P}_{U_{2_i}^j|U_{2_i}^{1:j-1}X_i^N}$ defined by

$$\tilde{P}_{U_{2_i}^j|U_{2_i}^{1:j-1}X_i^N} \triangleq \begin{cases} Q_{U_2^j|U_2^{1:j-1}} & j \in \mathcal{V}_C^c, \\ Q_{U_2^j|U_2^{1:j-1}X^N} & j \in \mathcal{F}_3 \cup \mathcal{F}_5. \end{cases} \tag{8}$$

3. $\widetilde{C}_i^N \leftarrow \widetilde{U}_{2_i}^N G_n$
4. $\mathcal{E}_1$ in Fig. 3 constructs $\widetilde{U}_{1_i}^N$ bit-by-bit as follows:
   - $\widetilde{U}_{1_i}^N[\mathcal{F}_6] \leftarrow M_{1_i}$
   - $\widehat{U}_{1_i}^N[\mathcal{F}_7] \leftarrow \bar{J}_2$
   - Given X_i^N and $\widetilde{C}_i^N$, successively draw the remaining components of $\widehat{U}_{1_i}^N$ according to $\tilde{P}_{U_{1_i}^j|U_{1_i}^{1:j-1}C_i^N X_i^N}$ defined by

$$\tilde{P}_{U_{1_i}^j|U_{1_i}^{1:j-1}C_i^N X_i^N} \triangleq \begin{cases} Q_{U_1^j|U_1^{1:j-1}} & j \in \mathcal{V}_A^c, \\ Q_{U_1^j|U_1^{1:j-1}C^N} & j \in \mathcal{F}_9, \\ Q_{U_1^j|U_1^{1:j-1}C^N X^N} & j \in \mathcal{F}_8. \end{cases} \tag{9}$$

5. $\widetilde{A}_i^N \leftarrow \widetilde{U}_{1_i}^N G_n$
6. Transmit $\widetilde{A}_i^N$
7. **end for**

---

holds for each block individually regardless of the randomness recycling.

Lemma 1. *For block $i \in [\![1, k]\!]$, we have*

$$\mathbb{D}(Q_{A^N C^N X^N} || \tilde{P}_{A_i^N C_i^N X_i^N}) \leq 2N\delta_N.$$

Algorithm 2: Decoding algorithm at Node Y for strong coordination

---

Input: $B_{1:k}^N$, uniformly distributed common randomness, $\bar{J}_1$, and $J_{1:k}$ shared with Node X.
Output: $\widetilde{Y}_{1:k}^N$
1. **For** block $i = k, \ldots, 1$ **do**
2. $\mathcal{D}_2$ in Fig. 3 constructs $\widehat{U}_{2_i}^N$ bit-by-bit as follows:
   - $\left(\widehat{U}_{2_i}^N[(\mathcal{F}_1 \setminus \acute{\mathcal{F}}_{3_1}) \cup \check{\mathcal{F}}_4], \mathcal{F}_{3_1}^{(i)}\right) \leftarrow J_i$
   - $\left(\widehat{U}_{2_i}^N[(\mathcal{F}_2 \setminus \acute{\mathcal{F}}_{3_2}) \cup \hat{\mathcal{F}}_4], \mathcal{F}_{3_2}\right) \leftarrow \bar{J}_1$
   - Given B_i^N successively draw the components of $\widehat{U}_{2_i}^N$ according to $\tilde{P}_{U_{2_i}^j|U_{2_i}^{j-1},B_i^N}$ defined by

$$\tilde{P}_{U_{2_i}^j|U_{2_i}^{j-1}B_i^N} \triangleq \begin{cases} Q_{U_2^j|U_2^{1:j-1}} & j \in \mathcal{V}_C^c, \\ Q_{U_2^j|U_2^{1:j-1}B^N} & j \in \acute{\mathcal{F}}_{3_2} \cup \acute{\mathcal{F}}_{3_1} \cup \mathcal{F}_5. \end{cases} \tag{10}$$

3. **if** $i = k$ **then**
   - $\widehat{U}_{2_i}^N[\mathcal{F}_3] \leftarrow B_{k+1}^N$

   else
   - $\widehat{U}_{2_i}^N[\mathcal{F}_3 \setminus \mathcal{F}_{3_2}] \leftarrow \widehat{U}_{2_{i+1}}^N[\acute{\mathcal{F}}_{3_1}] \oplus \mathcal{F}_{3_1}^{(i+1)}$
   - $\widehat{U}_{2_i}^N[\mathcal{F}_3 \setminus \mathcal{F}_{3_1}] \leftarrow \widehat{U}_{2_{i+1}}^N[\acute{\mathcal{F}}_{3_2}] \oplus \mathcal{F}_{3_2}$
4. Let
   - $\widehat{U}_{2_i}^N[\acute{\mathcal{F}}_{3_1}] \leftarrow \mathcal{F}_{3_1}^{(i)}$
   - $\widehat{U}_{2_i}^N[\acute{\mathcal{F}}_{3_2}] \leftarrow \mathcal{F}_{3_2}$
5. $\widehat{C}_i^N \leftarrow \widehat{U}_{2_i}^N G_n$
6. Channel simulation: given $\widehat{C}_i^N$ and B_i^N, successively draw the components of $\widetilde{T}_i^N$ according to

$$\tilde{P}_{T_i^j|T_i^{1:j-1}B_i^N C_i^N} \triangleq \begin{cases} 1/|\mathcal{Y}| & j \in \mathcal{V}_{Y|BC}, \\ Q_{T^j|T^{1:j-1}B^N C^N} & j \in \mathcal{V}_{Y|BC}^c. \end{cases} \tag{11}$$

5. $\widetilde{Y}_i^N \leftarrow \widetilde{T}_i^N G_n$
6. **end for**

Proof. We have

$$\mathbb{D}(Q_{A^N C^N X^N} || \tilde{P}_{A_i^N C_i^N X_i^N})$$
$$\overset{(a)}{=} \mathbb{D}(Q_{U_1^N U_2^N X^N} || \tilde{P}_{U_{1_i}^N U_{2_i}^N X_i^N})$$
$$\overset{(b)}{=} \mathbb{E}_{Q_{X^N}}\left[\mathbb{D}(Q_{U_1^N U_2^N|X^N} || \tilde{P}_{U_{1_i}^N U_{2_i}^N|X_i^N})\right]$$
$$= \mathbb{E}_{Q_{X^N}}\left[\mathbb{D}(Q_{U_2^N|X^N}Q_{U_1^N|U_2^N X^N} || \tilde{P}_{U_{2_i}^N|X_i^N}\tilde{P}_{U_{1_i}^N|U_{2_i}^N X_i^N})\right]$$
$$\overset{(c)}{=} \mathbb{E}_{Q_{X^N}}\Big[\mathbb{D}(Q_{U_2^N|X^N} || \tilde{P}_{U_{2_i}^N|X_i^N})$$
$$\qquad + \mathbb{D}(Q_{U_1^N|U_2^N X^N} || \tilde{P}_{U_{1_i}^N|U_{2_i}^N X_i^N})\Big]$$
$$\overset{(d)}{=} \sum_{j=1}^N \mathbb{E}_{Q_{U_2^{1:j-1}X^N}}\left[\mathbb{D}(Q_{U_2^j|U_2^{1:j-1}X^N} || \tilde{P}_{U_{2_i}^j|U_{2_i}^{1:j-1}X_i^N})\right]$$
$$+ \sum_{j=1}^N \mathbb{E}_{Q_{U_1^{1:j-1}U_2^N X^N}}\left[\mathbb{D}(Q_{U_1^j|U_1^{1:j-1}U_2^N X^N} || \tilde{P}_{U_{1_i}^j|U_{1_i}^{1:j-1}U_{2_i}^N X_i^N})\right]$$

$$\stackrel{(e)}{=} \sum_{j \notin \mathcal{F}_3 \cup \mathcal{F}_5} \mathbb{E}_{Q_{U_2^{1:j-1}X^N}} \left[\mathbb{D}(Q_{U_2^j|U_2^{1:j-1}X^N} || \tilde{P}_{U_{2_i}^j|U_{2_i}^{1:j-1}X_i^N}) \right]$$

$$+ \sum_{j \notin \mathcal{F}_8} \mathbb{E}_{Q_{U_1^{1:j-1}U_2^N X^N}} \left[\mathbb{D}(Q_{U_1^j|U_1^{1:j-1}U_2^N X^N} || \tilde{P}_{U_{1_i}^j|U_{1_i}^{1:j-1}U_{2_i}^N X_i^N}) \right]$$

$$\stackrel{(f)}{=} \sum_{j \in \mathcal{V}_C^c \cup \mathcal{V}_{C|X}} \mathbb{E}_{Q_{U_2^{1:j-1}X^N}} \left[\mathbb{D}(Q_{U_2^j|U_2^{1:j-1}X^N} || \tilde{P}_{U_{2_i}^j|U_{2_i}^{1:j-1}X_i^N}) \right]$$

$$+ \sum_{j \in \mathcal{V}_A^c \cup \mathcal{V}_{A|CX} \cup \mathcal{V}_A \backslash \mathcal{V}_{A|C}} \mathbb{E}_{Q_{U_1^{1:j-1}U_2^N X^N}} \Big[$$

$$\mathbb{D}(Q_{U_1^j|U_1^{1:j-1}U_2^N X^N} || \tilde{P}_{U_{1_i}^j|U_{1_i}^{1:j-1}U_{2_i}^N X_i^N}) \Big]$$

$$\stackrel{(g)}{=} \sum_{j \in \mathcal{V}_C^c} \left(H(U_2^j|U_2^{1:j-1}) - H(U_2^j|U_2^{1:j-1}X^N) \right)$$

$$+ \sum_{j \in \mathcal{V}_{C|X}} \left(1 - H(U_2^j|U_2^{1:j-1}X^N) \right)$$

$$+ \sum_{j \in \mathcal{V}_A^c} \left(H(U_1^j|U_1^{1:j-1}) - H(U_1^j|U_1^{1:j-1}U_2^N X^N) \right)$$

$$+ \sum_{j \in \mathcal{V}_{A|CX}} \left(1 - H(U_1^j|U_1^{1:j-1}U_2^N X^N) \right)$$

$$+ \sum_{j \in \mathcal{V}_{A|C}^c \backslash \mathcal{V}_A^c} \left(H(U_1^j|U_1^{1:j-1}U_2^N) - H(U_1^j|U_1^{1:j-1}U_2^N X^N) \right)$$

$$\stackrel{(h)}{=} \sum_{j \in \mathcal{V}_C^c} \left(H(U_2^j|U_2^{1:j-1}) - H(U_2^j|U_2^{1:j-1}X^N) \right)$$

$$+ \sum_{j \in \mathcal{V}_{C|X}} \left(1 - H(U_2^j|U_2^{1:j-1}X^N) \right)$$

$$+ \sum_{j \in \mathcal{V}_A^c} \left(H(U_1^j|U_1^{1:j-1}) - H(U_1^j|U_1^{1:j-1}C^N X^N) \right)$$

$$+ \sum_{j \in \mathcal{V}_{A|CX}} \left(1 - H(U_1^j|U_1^{1:j-1}C^N X^N) \right)$$

$$+ \sum_{j \in \mathcal{V}_{A|C}^c \backslash \mathcal{V}_A^c} \left(H(U_1^j|U_1^{1:j-1}C^N) - H(U_1^j|U_1^{1:j-1}C^N X^N) \right)$$

$$\stackrel{(i)}{\leq} (|\mathcal{V}_C^c| + |\mathcal{V}_{C|X}| + |\mathcal{V}_{A|XC}| + |\mathcal{V}_{A|C}^c|)\delta_N \leq 2N\delta_N$$

where

(a) holds by invertibility of G_n;

(b) - (d) follows from the chain rule of the KL divergence [28];

(e) results from the definitions of the conditional distributions in (8), and (9);

(f) follows from the definitions of the index sets as shown in Figures 4 and 5;

(g) results from the encoding of $\widetilde{U}_{1_i}^N$ and $\widetilde{U}_{2_i}^N$ bit-by-bit at $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively, with uniformly distributed randomness bits and message bits. These bits are generated by applying successive cancellation encoding using previous bits and side information with conditional distributions defined in (8) and (9);

(h) holds by the one-to-one relation between U_2^N and C^N;

(i) follows from the sets defined in (5) and (6).

■

Lemma 2. *For block $i \in [\![1,k]\!]$, we have*

$$\mathbb{D}(\tilde{P}_{X_i^N Y_i^N} || Q_{X^N Y^N})$$

$$\leq \mathbb{D}(\tilde{P}_{X_i^N A_i^N C_i^N B_i^N \widehat{C}_i^N Y_i^N} || Q_{X^N A^N C^N B^N \widehat{C}^N Y^N}) \leq \delta_N^{(2)}$$

where $\delta_N^{(2)} \triangleq \mathcal{O}(\sqrt{N^3 \delta_N})$.

Proof. Consider the argument shown at the top of the following page. In this argument:

(a) - (b) results from the Markov chain $X^N - A^N C^N - B^N \widehat{C}^N - Y^N$;

(c) follows from [17, Lemma 16] where

$$\hat{\delta}_N^{(2)} \triangleq -N \log(\mu_{XACB\widehat{C}Y}) \sqrt{2 \ln 2} \sqrt{2N \delta_N},$$

$$\mu_{XACB\widehat{C}Y} \triangleq \min^*_{x,y,a,c,b,\hat{c}} (Q_{XACB\widehat{C}Y});$$

(d) follows from the chain rule of KL divergence [28];

(e) holds by Lemma 1 and [17, Lemma 14] where

$$\delta_N^{(1)} \triangleq -N \log(\mu_{XAC}) \sqrt{2 \ln 2} \sqrt{2N \delta_N},$$

$$\mu_{XAC} \triangleq \min^*_{x,a,c}(Q_{XAC});$$

(f) follows from the chain rule of KL divergence [28];

(g) holds by [17, Lemma 14], where

$$\mu_{ACB\widehat{C}} \triangleq \min^*_{a,c,b,\hat{c}}(Q_{ACB\widehat{C}}),$$

$$\mu_{YB\widehat{C}} \triangleq \min^*_{y,b,\hat{c}}(Q_{YB\widehat{C}});$$

(h) holds by bounding the terms
$\mathbb{D}(Q_{B^N \widehat{C}^N|A^N C^N} || \tilde{P}_{B_i^N \widehat{C}_i^N|A_i^N C_i^N})$, and
$\mathbb{D}(Q_{Y^N|B^N \widehat{C}^N} || \tilde{P}_{Y_i^N|B_i^N \widehat{C}_i^N})$, as follows:

$$\mathbb{D}(Q_{B^N \widehat{C}^N|A^N C^N} || \tilde{P}_{B_i^N \widehat{C}_i^N|A_i^N C_i^N})$$

$$\stackrel{(a)}{=} \mathbb{D}(Q_{B^N|A^N} Q_{\widehat{C}^N|B^N} || Q_{B^N|A^N} \tilde{P}_{\widehat{C}_i^N|B_i^N})$$

$$= \mathbb{D}(Q_{\widehat{C}^N|B^N} || \tilde{P}_{\widehat{C}_i^N|B_i^N})$$

$$\stackrel{(b)}{=} \mathbb{D}(Q_{\widehat{U}^N|B^N} || \tilde{P}_{\widehat{U}_i^N|B_i^N})$$

$$\stackrel{(c)}{=} \sum_{j=1}^N \mathbb{E}_{Q_{U_2^{1:j-1}B^N}} \left[\mathbb{D}(Q_{U_2^j|U_2^{1:j-1}B^N} || \tilde{P}_{U_{2_i}^j|U_{2_i}^{1:j-1}B_i^N}) \right]$$

$$\stackrel{(d)}{=} \sum_{j \in \mathcal{V}_C^c} \mathbb{E}_{Q_{U_2^{1:j-1}B^N}} \left[\mathbb{D}(Q_{U_2^j|U_2^{1:j-1}B^N} || \tilde{P}_{U_{2_i}^j|U_{2_i}^{1:j-1}B_i^N}) \right]$$

$$+ \sum_{j \in \mathcal{H}_{C|B} \cup \mathcal{V}_{C|X}} \mathbb{E}_{Q_{U_2^{1:j-1}B^N}} \left[\mathbb{D}(Q_{U_2^j|U_2^{1:j-1}B^N} || \tilde{P}_{U_{2_i}^j|U_{2_i}^{1:j-1}B_i^N}) \right]$$

$$\stackrel{(e)}{=} \sum_{j \in \mathcal{V}_C^c} \left(H(U_2^j|U_2^{1:j-1}) - H(U_2^j|U_2^{1:j-1}B^N) \right)$$

$$+ \sum_{j \in \mathcal{H}_{C|B} \cup \mathcal{V}_{C|X}} \left(1 - H(U_2^j|U_2^{1:j-1}B^N) \right)$$

$$\stackrel{(f)}{\leq} |\mathcal{V}_C^c|\delta_N + |\mathcal{H}_{C|B} \cup \mathcal{V}_{C|X}|\delta_N \leq N\delta_N,$$

where

(a) results from the Markov chain $C - A - B - \widehat{C}$ and the fact that $\tilde{P}_{B_i^N|A_i^N} = Q_{B^N|A^N}$;

(b) holds by the one-to-one relation between U_2^N and C^N;

(c) follows from the chain rule of KL divergence [28];

$$\mathbb{D}(\tilde{P}_{X_i^N A_i^N C_i^N B_i^N \widehat{C}_i^N Y_i^N} || Q_{X^N A^N C^N B^N \widehat{C}^N Y^N})$$

$$= \mathbb{D}(\tilde{P}_{Y_i^N | X_i^N A_i^N C_i^N B_i^N \widehat{C}_i^N} \tilde{P}_{X_i^N A_i^N C_i^N B_i^N \widehat{C}_i^N} || Q_{Y^N | X^N A^N C^N B^N \widehat{C}^N} Q_{X^N A^N C^N B^N \widehat{C}^N})$$

$$\overset{(a)}{=} \mathbb{D}(\tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N} \tilde{P}_{X_i^N A_i^N C_i^N B_i^N \widehat{C}_i^N} || Q_{Y^N | B^N \widehat{C}^N} Q_{X^N A^N C^N B^N \widehat{C}^N})$$

$$= \mathbb{D}(\tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N} \tilde{P}_{B_i^N \widehat{C}_i^N | X_i^N A_i^N C_i^N} \tilde{P}_{X_i^N A_i^N C_i^N} || Q_{Y^N | B^N \widehat{C}^N} Q_{B^N \widehat{C}^N | X^N A^N C^N} Q_{X^N A^N C^N})$$

$$\overset{(b)}{=} \mathbb{D}(\tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N} \tilde{P}_{B_i^N \widehat{C}_i^N | A_i^N C_i^N} \tilde{P}_{X_i^N A_i^N C_i^N} || Q_{Y^N | B^N \widehat{C}^N} Q_{B^N \widehat{C}^N | A^N C^N} Q_{X^N A^N C^N})$$

$$\overset{(c)}{\leq} \hat{\delta}_N^{(2)} + \mathbb{D}(\tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N} \tilde{P}_{B_i^N \widehat{C}_i^N | A_i^N C_i^N} \tilde{P}_{X_i^N A_i^N C_i^N} || \tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N} \tilde{P}_{B_i^N \widehat{C}_i^N | A_i^N C_i^N} Q_{X^N A^N C^N})$$

$$+ \mathbb{D}(\tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N} \tilde{P}_{B_i^N \widehat{C}_i^N | A_i^N C_i^N} Q_{X^N A^N C^N} || Q_{Y^N | B^N \widehat{C}^N} Q_{B^N \widehat{C}^N | A^N C^N} Q_{X^N A^N C^N})$$

$$\overset{(d)}{=} \hat{\delta}_N^{(2)} + \mathbb{D}(\tilde{P}_{X_i^N A_i^N C_i^N} || Q_{X^N A^N C^N}) + \mathbb{D}(\tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N} \tilde{P}_{B_i^N \widehat{C}_i^N | A_i^N C_i^N} || Q_{Y^N | B^N \widehat{C}^N} Q_{B^N \widehat{C}^N | A^N C^N})$$

$$\overset{(e)}{\leq} \hat{\delta}_N^{(2)} + \delta_N^{(1)} + \mathbb{D}(\tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N} \tilde{P}_{B_i^N \widehat{C}_i^N | A_i^N C_i^N} || Q_{Y^N | B^N \widehat{C}^N} Q_{B^N \widehat{C}^N | A^N C^N})$$

$$\overset{(f)}{=} \hat{\delta}_N^{(2)} + \delta_N^{(1)} + \mathbb{D}(\tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N} || Q_{Y^N | B^N \widehat{C}^N}) + \mathbb{D}(\tilde{P}_{B_i^N \widehat{C}_i^N | A_i^N C_i^N} || Q_{B^N \widehat{C}^N | A^N C^N})$$

$$\overset{(g)}{\leq} \hat{\delta}_N^{(2)} + \delta_N^{(1)} - N \log(\mu_{YB\widehat{C}}) \sqrt{2 \ln 2} \sqrt{\mathbb{D}(Q_{Y^N | B^N \widehat{C}^N} || \tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N})}$$

$$- N \log(\mu_{ACB\widehat{C}}) \sqrt{2 \ln 2} \sqrt{\mathbb{D}(Q_{B^N \widehat{C}^N | A^N C^N} || \tilde{P}_{B_i^N \widehat{C}_i^N | A_i^N C_i^N})}$$

$$\overset{(h)}{\leq} \hat{\delta}_N^{(2)} + \delta_N^{(1)} - N \log(\mu_{YB\widehat{C}}) \sqrt{2 \ln 2} \sqrt{N \delta_N} - N \log(\mu_{ACB\widehat{C}}) \sqrt{2 \ln 2} \sqrt{N \delta_N}$$

(d) - (e) results from the definitions of the conditional distributions in (10);

(f) follows from the sets defined in (5).

$$\mathbb{D}(Q_{Y^N | B^N \widehat{C}^N} || \tilde{P}_{Y_i^N | B_i^N \widehat{C}_i^N})$$

$$\overset{(a)}{=} \sum_{j=1}^{N} \mathbb{E}_{Q_{T^{1:j-1} B_i^N \widehat{C}_i^N}} \left[\mathbb{D}(Q_{T^j | T^{1:j-1} B^N \widehat{C}^N} || \tilde{P}_{T^j | T^{1:j-1} B_i^N \widehat{C}_i^N}) \right]$$

$$\overset{(b)}{=} \sum_{j \in \mathcal{V}_{Y|BC}} \mathbb{E}_{Q_{T^{1:j-1} B_i^N \widehat{C}_i^N}} \left[\mathbb{D}(Q_{T^j | T^{1:j-1} B^N \widehat{C}^N} || \tilde{P}_{T^j | T^{1:j-1} B_i^N \widehat{C}_i^N}) \right]$$

$$\overset{(c)}{=} \sum_{j \in \mathcal{V}_{Y|BC}} \left(\log |\mathcal{Y}| - H(T^j | T^{1:j-1} B^N C^N) \right)$$

$$\overset{(d)}{\leq} |\mathcal{V}_{Y|BC}| \delta_N \leq N \delta_N,$$

where

(a) follows from the chain rule of KL divergence [28];

(b) - (c) results from the definitions of the conditional distribution in (11);

(d) follows from the set defined in (7).

∎

Now, Lemmas 3 and 4 provide the independence between two consecutive blocks and the independence between all blocks based on the results of Lemma 2.

Lemma 3. *For block $i \in [\![2, k]\!]$, we have*

$$\mathbb{D}(\tilde{P}_{X_{i-1:i}^N Y_{i-1:i}^N \bar{J}_1} || \tilde{P}_{X_{i-1}^N Y_{i-1}^N \bar{J}_1} \tilde{P}_{X_i^N Y_i^N}) \leq \delta_N^{(3)}$$

where $\delta_N^{(3)} \triangleq \mathcal{O}(\sqrt[4]{N^{15} \delta_N})$.

Proof. We reuse the proof of [17, Lemma 3] with substitutions $q_{U^{1:N}} \leftarrow Q_{C^N}$, $q_{Y^{1:N}} \leftarrow Q_{X^N Y^N}$, $\tilde{p}_{U_i^{1:N}} \leftarrow \tilde{P}_{C_i^N}$,

$\tilde{p}_{Y^{1:N}} \leftarrow \tilde{P}_{Y_i^N X_i^N}$, and $\bar{R}_1 \leftarrow \bar{J}_1$. This will result in the Markov chain $X_{i-1}^N \widetilde{Y}_{i-1}^N - \bar{J}_1 - X_i^N \widetilde{Y}_i^N$ replacing the chain in [17, Lemma 3]. ∎

Lemma 4. *We have*

$$\mathbb{D}\left(\tilde{P}_{X_{1:k}^N Y_{1:k}^N} || \prod_{i=1}^{k} \tilde{P}_{X_i^N Y_i^N} \right) \leq (k-1) \delta_N^{(3)}$$

where $\delta_N^{(3)}$ is defined in Lemma 3.

Proof. We reuse the proof of [17, Lemma 4] with substitutions $\tilde{p}_{Y_i^{1:N}} \leftarrow \tilde{P}_{X_i^N Y_i^N}$, and $\bar{R}_1 \leftarrow \bar{J}_1$. This will result in the Markov chain $X_{1:i-2}^N \widetilde{Y}_{1:i-2}^N - \bar{J}_1 X_{i-1}^N \widetilde{Y}_{i-1}^N - X_i^N \widetilde{Y}_i^N$ replacing the chain in [17, Lemma 4]. ∎

Finally, by the results of Lemma 4 we can show in Lemma 5 that the target distribution $Q_{X^N Y^N}$ is approximated asymptotically over all blocks jointly.

Lemma 5. *We have*

$$\mathbb{D}\left(\tilde{P}_{X_{1:k}^N Y_{1:k}^N} || Q_{X^{1:kN} Y^{1:kN}} \right) \leq \delta_N^{(4)}.$$

where $\delta_N^{(4)} \triangleq \mathcal{O}(k^{3/2} N^{23/8} \delta_N^{1/8})$

Proof. We reuse the proof of [17, Lemma 5] with substitutions $q_{Y^{1:N}} \leftarrow Q_{X^N Y^N}$, $\tilde{p}_{Y_i^{1:N}} \leftarrow \tilde{P}_{Y_i^N X_i^N}$. ∎

Theorem 2. *The polar coding scheme described in Algorithms 1, 2 achieves the region stated in Theorem 1. It satisfies (1) for a binary input DMC channel and a target distribution q_{XY} defined over $\mathcal{X} \times \mathcal{Y}$, with an axillary random variable C defined over the binary alphabet.*

Proof. The common randomness rate R_o is given as

$$\frac{|\bar{J}_1| + |J_{1:k}|}{kN} = \frac{|\mathcal{V}_{C|XY}| + k|\mathcal{V}_{C|X} \setminus \mathcal{V}_{C|XY}|}{Nk}$$

$$= \frac{|\mathcal{V}_{C|XY}|}{kN} + \frac{|\mathcal{V}_{C|X} \setminus \mathcal{V}_{C|XY}|}{N}$$

$$\xrightarrow{N\to\infty} \frac{H(C|XY)}{k} + I(Y;C|X)$$

$$\xrightarrow{k\to\infty} I(Y;C|X). \tag{12}$$

The communication rate R_c is given as

$$\frac{k|\mathcal{F}_5 \cup \mathcal{F}_3|}{kN} = \frac{k|\mathcal{V}_C \setminus \mathcal{V}_{C|X}|}{Nk} = \frac{|\mathcal{V}_C \setminus \mathcal{V}_{C|X}|}{N}$$

$$\xrightarrow{N\to\infty} I(X;C), \tag{13}$$

whereas R_a can be written as

$$\frac{|\mathcal{V}_{A|CXY}| + k|\mathcal{F}_8|}{kN} = \frac{|\mathcal{V}_{A|CXY}| + k|\mathcal{V}_{A|C} \setminus \mathcal{V}_{A|CX}|}{kN}$$

$$= \frac{|\mathcal{V}_{A|CXY}|}{kN} + \frac{|\mathcal{V}_{A|C} \setminus \mathcal{V}_{A|CX}|}{N}$$

$$\xrightarrow{N\to\infty} I(A;X|C) + \frac{H(A|CXY)}{k}$$

$$\xrightarrow{k\to\infty} I(A;X|C). \tag{14}$$

The rates of local randomness ρ_1 and ρ_2, respectively, are given as

$$\rho_1 = \frac{k|\mathcal{F}_6|}{kN} = \frac{k|\mathcal{V}_{A|CX} \setminus \mathcal{V}_{A|CXY}|}{Nk} = \frac{|\mathcal{V}_{A|CX} \setminus \mathcal{V}_{A|CXY}|}{N}$$

$$\xrightarrow{N\to\infty} I(A;Y|CX) \quad \text{and} \tag{15}$$

$$\rho_2 = \frac{k|V_{Y|BC}|}{kN} \xrightarrow{N\to\infty} H(Y|BC). \tag{16}$$

Finally we see that conditions (2a)-(2g) are satisfied by (12)-(16). Hence, given R_a, R_o, R_c satisfying Theorem 1, based on Lemma 5 and Pinsker's inequality we have

$$\mathbb{E}\big[||\tilde{P}_{X_{1:k}^N Y_{1:k}^N} - Q_{X^{1:kN}Y^{1:kN}}||_{TV}\big]$$

$$\leq \mathbb{E}\Big[\sqrt{2\mathbb{D}(\tilde{P}_{X_{1:k}^N Y_{1:k}^N}||Q_{X^{1:kN}Y^{1:kN}})}\Big]$$

$$\leq \sqrt{2\mathbb{E}\big[\mathbb{D}(\tilde{P}_{X_{1:k}^N Y_{1:k}^N}||Q_{X^{1:kN}Y^{1:kN}})\big]} \xrightarrow{N\to\infty} 0. \tag{17}$$

As a result, from (17) there exists an $N \in \mathbb{N}$ for which the polar code-induced pmf between the pair of actions satisfies the strong coordination condition is given by (1). ∎

REFERENCES

[1] E. Soljanin, "Compressing quantum mixed-state sources by sending classical information," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2263–2275, Aug. 2002.

[2] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.

[3] A. A. Gohari and V. Anantharam, "Generating dependent random variables over networks," in *Proc. IEEE Inf. Theory Workshop*, Paraty, Brazil, Oct. 2011, pp. 698–702.

[4] F. Haddadpour, M. H. Yassaee, A. Gohari, and M. R. Aref, "Coordination via a relay," in *Proc. IEEE Int. Sympos. on Inform. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 3048–3052.

[5] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[6] A. Bereyhi, M. Bahrami, M. Mirmohseni, and M. R. Aref, "Empirical coordination in a triangular multi-terminal network," in *Proc. IEEE Int. Sympos. on Inform. Theory*, Istanbul, Turkey, 2013, pp. 2149–2153.

[7] M. R. Bloch and J. Kliewer, "Strong coordination over a line network," in *Proc. IEEE Int. Sympos. on Inform. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2319–2323.

[8] ——, "Strong coordination over a three-terminal relay network," in *Proc. IEEE Inf. Theory Workshop*, Hobart, Australia, Nov. 2014, pp. 646–650.

[9] M. H. Yassaee, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 2964–2982, 2015.

[10] B. N. Vellambi, J. Kliewer, and M. R. Bloch, "Strong coordination over multi-hop line networks," *arXiv preprint*, 2016. [Online]. Available: http://arxiv.org/abs/1602.09001

[11] P. Cuff and C. Schieler, "Hybrid codes needed for coordination over the point-to-point channel," in *Proc. 49th Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, Sep. 2011, pp. 235–239.

[12] F. Haddadpour, M. H. Yassaee, S. Beigi, A. Gohari, and M. R. Aref, "Simulation of a channel with another channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2659–2677, May 2017.

[13] G. Cervia, L. Luzzi, M. L. Treust, and M. Bloch, "Strong coordination of signals and actions over noisy channels," in *Proc. IEEE Int. Sympos. on Inform. Theory*, Aachen, Germany, June 2017, pp. 2835–2839.

[14] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[15] ——, "Source polarization," in *Proc. IEEE Int. Sympos. on Inform. Theory*, June 2010, pp. 899–903.

[16] M. R. Bloch, L. Luzzi, and J. Kliewer, "Strong coordination with polar codes," in *50th Annual Allerton Conference on Commununication, Control, and Computing*, Monticello, IL, USA, Oct. 2012, pp. 565–571.

[17] R. A. Chou, M. Bloch, and J. Kliewer, "Empirical and strong coordination via soft covering with polar codes," *arXiv preprint*, 2016. [Online]. Available: http://arxiv.org/abs/1608.08474

[18] R. Blasco-Serrano, R. Thobaben, and M. Skoglund, "Polar codes for coordination in cascade networks," in *International Zurich Seminar on Communications (IZS)*. Eidgenössische Technische Hochschule Zürich, 2012, pp. 55–58.

[19] G. Cervia, L. Luzzi, M. R. Bloch, and M. L. Treust, "Polar coding for empirical coordination of signals and actions over noisy channels," in *Proc. IEEE Inf. Theory Workshop*, Sept 2016, pp. 81–85.

[20] S. A. Obead, B. N. Vellambi, and J. Kliewer, "Strong coordination over noisy channels: Is separation sufficient?" in *Proc. IEEE Int. Sympos. on Inform. Theory*, Aachen, Germany, June 2017, pp. 2840–2844.

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

[22] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, 1993.

[23] E. Şaşoğlu, E. Telatar, and E. Arikan, "Polarization for arbitrary discrete memoryless channels," in *Proc. IEEE Inf. Theory Workshop*, Oct 2009, pp. 144–148.

[24] R. A. Chou and M. R. Bloch, "Polar coding for the broadcast channel with confidential messages: A random binning analogy," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2410–2429, May 2016.

[25] M. Mondelli, S. H. Hassani, I. Sason, and R. L. Urbanke, "Achieving Marton's region for broadcast channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 783–800, Feb 2015.

[26] E. E. Gad, Y. Li, J. Kliewer, M. Langberg, A. A. Jiang, and J. Bruck, "Asymmetric error correction and flash-memory rewriting using polar codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4024–4038, July 2016.

[27] G. D. Forney Jr., "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proc. 41st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct 2003, pp. 430–439.

[28] T. M. Cover and J. Thomas, *Elements of Information Theorey*, 2nd ed., ser. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, 2006.