Strong Coordination Over Noisy Channels

Sarah A. Obead, *Graduate Student Member, IEEE*, Badri N. Vellambi, *Senior Member, IEEE*, and Jörg Kliewer, *Senior Member, IEEE*

***Abstract*— We study the problem of strong coordination of the actions of two nodes X and Y that communicate over a discrete memoryless channel (DMC) such that the actions follow a prescribed joint probability distribution. We propose two novel random coding schemes and a polar coding scheme for this noisy strong coordination problem, and derive inner and outer bounds for the respective strong coordination capacity region. The first scheme is a joint coordination-channel encoding scheme that utilizes the randomness provided by the communication channel to reduce the amount of local randomness required to generate the sequence of actions at Node Y. Based on this random coding scheme, we provide a characterization of the capacity region for a special case of the noisy strong coordination setup, namely, when the DMC is a deterministic channel. The second scheme exploits separate coordination and channel encoding where local randomness is extracted from the channel after decoding. Moreover, by leveraging the random coding results for this problem, we present an example in which the proposed joint encoding scheme is able to strictly outperform the separate encoding scheme in terms of achievable communication rate for the same amount of injected randomness into both systems. Thus, we establish the sub-optimality of the separation of strong coordination and channel encoding with respect to the communication rate over the DMC in this problem. Finally, the third scheme is a joint coordination-channel polar coding scheme for strong coordination. We show that polar codes are able to achieve the 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 employed together to simulate another DMC.**

***Index Terms*— Strong coordination, joint source-channel coding, channel resolvability, superposition coding, polar codes.**

I. Introduction

A FUNDAMENTAL problem in decentralized networks is to coordinate activities of different nodes with the goal of reaching a state of agreement. The problem of communication-based coordination of multi-node systems arises in numerous applications including autonomous robots, smart traffic control, and distributed computing such as distributed games and grid computing [3]. Coordination is understood to be the ability to arrive at a prescribed joint distribution of actions at all nodes in the network. Several theoretical and applied studies on multi-node coordination have targeted questions on how nodes exchange information and how their actions can be correlated to achieve a desired overall behavior. Two types of coordination have been addressed in the literature – *empirical* coordination where the normalized histogram of induced joint actions is required to be close to a prescribed target distribution, and *strong* coordination, where the induced sequence of joint actions of all the nodes is required to be statistically close (i.e., nearly indistinguishable) from a chosen target probability mass function (pmf).

Recently, a significant amount of work has been devoted to finding the capacity regions of various coordination network problems based on both empirical and strong coordination [3]–[8]. Bounds on the capacity region for the point-to-point case were obtained in [9] under the assumption that the nodes communicate in a bidirectional fashion in order to achieve coordination. A similar framework was adopted and improved in [10]. In [6], [8], [11], the authors addressed inner and outer bounds for the capacity region of a three-terminal network in the presence of a relay. The work of [6] was later extended in [7], [12] to derive a precise characterization of the strong coordination region for multi-hop networks.

While the majority of recent works on coordination have considered noise-free communication channels, coordination over noisy channels has received only little attention in the literature so far. However, notable exceptions are [13]–[15]. In [13], joint empirical coordination of the channel inputs/outputs of a noisy communication channel with source and reproduction sequences is considered, and in [14], the notion of strong coordination is used to simulate a discrete memoryless channel (DMC) via another channel. Recently, the authors of [15] explored the strong coordination variant of the problem investigated in [13] when two-sided channel state information is present and side information is available at the decoder.

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 [16], [17]. For example, polar coding for strong point-to-point coordination is addressed in [18], [19], and for empirical coordination in cascade networks in [20], respectively. The only existing design of polar codes for the

Manuscript received August 14, 2018; revised October 7, 2020; accepted February 12, 2021. Date of publication March 2, 2021; date of current version April 21, 2021. This work was supported by NSF under Grant CCF-1440014, Grant CCF-1439465, and Grant CNS-1526547. This article was presented in part at the 2017 IEEE International Symposium on Information Theory (ISIT) [1] and in part at the 55th Annual Allerton Conference [2]. *(Corresponding author: Sarah A. Obead.)*

Sarah A. Obead and Jörg Kliewer are with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: sao23@njit.edu; jkliewer@njit.edu).

Badri N. Vellambi was with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA. He is now with the Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH 45221 USA (e-mail: badri.vellambi@uc.edu).

Communicated by A. Gohari, Associate Editor At Large.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2021.3063140.

Digital Object Identifier 10.1109/TIT.2021.3063140

noisy empirical coordination case [21] is based on the joint source-channel coordination approach in [13]. A construction based on polar codes for the noisy *strong* coordination problem has been first presented in our previous conference work [2], which is part of this paper.

In this work, we consider the point-to-point coordination setup illustrated in Fig. 1, where in contrast to [13] and [15] only source and reproduction sequences at two different nodes (X and Y) are coordinated by means of a suitable communication scheme over a DMC. Specifically, we propose two novel achievable coding schemes for this noisy coordination scenario, derive inner and outer bounds to the underlying strong coordination capacity region, and provide the capacity region for a special case of the noisy strong coordination setup. In particular, we characterize the capacity region for the case when the DMC is deterministic. Finally, we design an explicit low-complexity nested polar coding scheme that achieves the inner bound of the point-to-point noisy coordination capacity region.

The first scheme is a joint coordination channel encoding scheme that utilizes randomness provided by the DMC to reduce the local randomness required in generating the action sequence at Node Y (see Fig. 1). It is worth mentioning that the noisy point-to-point coordination setup is in fact equivalent to the average case model of asymptotically simulating a DMC using another DMC of [14]. Due to this equivalency, the proposed joint scheme is related to the scheme in [14]; however, the presented scheme exhibits a significantly different codebook construction adapted to our coordination framework. Our scheme requires to explicitly quantify of the amount of common randomness shared by the two nodes as well as the local randomness at each of the two nodes. In contrast to [14], where local randomness is assumed to be infinite and inner and outer bounds for the achievable rates are presented accordingly, we consider common and local randomness to be limited resources similar to [5], [22]–[24]. Consequently, we establish a trade-off between the use of communication, common and local randomness rates. To this end, we propose a solution that achieves strong coordination over noisy channels via the soft covering principle [5]. Unlike to solutions inspired by random-binning[1] , a soft covering based solution is able to quantify the local randomness required at the encoder and decoder to generate correlated action sequences. Note that quantifying the amount of local randomness is also absent from the analysis in [15].

Our second achievable scheme exploits separate coordination and channel encoding where local randomness is extracted from the channel after decoding. In a separation-based encoding scheme, we consider a two-stage method for solving the noisy strong coordination problem. Specifically, an outer strong coordination code for noise-free links is conveyed over the DMC through an inner capacity achieving channel code. Finally, to enhance the performance of this benchmark separation-based scheme, a channel randomness extraction stage is added at the decoder to supplement the local randomness required in generating the action sequence at Node Y. Moreover, when the noisy channel and the correlation between X to Y are both given by binary symmetric channels (BSCs), we study the effect of the capacity of the noisy channel on the sum rate of common and local randomness. We conclude this section by showing that a joint coordination-channel encoding scheme is able to strictly outperform a separation-based scheme[2] in terms of achievable communication rate if the same amount of randomness is injected into the system in the high-capacity regime for the BSC, i.e., $C \to 1$. This example reveals that separate coordination and channel encoding is indeed sub-optimal in the context of strong coordination under the additional constraint of minimizing the communication rate.

Lastly, the third scheme is a joint coordination-channel polar coding scheme that employs nested codebooks similar to polar codes for the broadcast channel [26]. We show that our proposed construction provides an equivalent constructive alternative for strong coordination over noisy channels. Here, by equivalent we mean that for every rate point for which one can devise a random joint coordination-channel code, one can also devise a polar coding scheme with significantly lower encoding and decoding complexity. Also, our proposed polar coding scheme employs the soft covering principle [19] and offers a constructive solution to a channel simulation problem, where a DMC is employed to simulate another DMC in the presence of shared randomness [14].

The remainder of the paper is organized as follows: Section II outlines the notation. The problem of strong coordination over a noisy communication link is presented in Section III. We then derive achievability results for the noisy point-to-point coordination in Section IV for the joint random-coding scheme, discuss the characterization of the capacity region for a special case of the noisy strong coordination setup, and derive a general outer bound to the capacity region. Section V presents the separate encoding scheme with randomness extraction. In Section VI, we present numerical results for the proposed joint and separate coordination and channel encoding schemes, establishing the sub-optimality of the separate encoding scheme when the target joint distribution is described by a doubly binary symmetric source and the noisy channel by a BSC, respectively. In Section VII we propose a joint coordination-channel polar code construction and a proof that this construction achieves the random coding inner bound. Finally, some conclusions are drawn in Section VIII.

II. Notation

Throughout the paper, we denote a discrete random variable with upper-case letters (e.g., X) and its realization with lower case letters (e.g., x). The alphabet size of the random variable X is denoted as $|\mathcal{X}|$. We use $[\![n]\!]$ to denote the set $\{1,\dots,n\}$ for $n \in \mathbb{N}$. Similarly, we use X_k^n to denote the finite sequence $\{X_{k,1}, X_{k,2},\dots, X_{k,n}\}$ and $X_k^{i:j}$ to denote

[1]To the best of our knowledge, random binning based tools have not been fully developed yet to quantify the local randomness required at the encoder and decoder. However, initial steps along these lines has been taken in [25, Theorem 4].

[2]Note that when defining separation we also consider the number of channel uses, i.e., the communication rate, as a quantity of interest besides the communication reliability, i.e., the probability of decoding error.

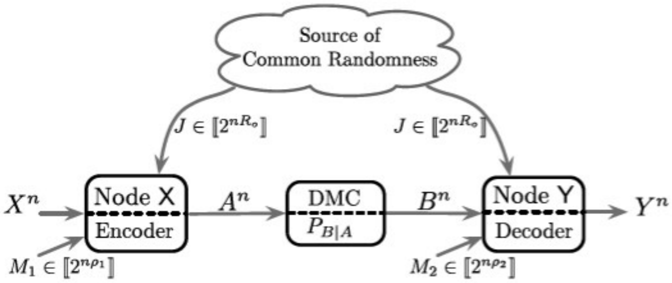

Fig. 1. Point-to-point strong coordination over a DMC.

$\{X_{k,i}, X_{k,i+1}, \dots, X_{k,j}\}$ such that $1 \leq i \leq j \leq n$. Given $\mathcal{A} \subset [\![n]\!]$, we let $X^n[\mathcal{A}]$ denote the components X_i such that $i \in \mathcal{A}$. We use boldface upper-case letters (e.g., $\mathbf{X}$) to denote matrices. We denote the source polarization transform as $\mathbf{G}_n = R\mathbf{F}^{\otimes n}$, where R is the bit-reversal mapping defined in [16], $\mathbf{F} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, and $\mathbf{F}^{\otimes n}$ denotes the n-th Kronecker product of $\mathbf{F}$. The binary entropy function is denoted as $h_2(\cdot)$, and the indicator function by $\mathbb{1}(\cdot)$. $\mathbb{P}[A]$ is the probability that the event A occurs. The pmf of the discrete random variable X is denoted as $P_X(x)$. However, we sometime use the lower case notation (e.g., $p_X(x)$) to distinguish target pmfs or alternative definitions. 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, between two distributions $P_X(x)$ and $Q_X(x)$ defined over an alphabet $\mathcal{X}$. Given a pmf $P_X(x)$ we let $\min_x^*(P_X) = \min_{x \in \mathcal{X}} \{P_X(x) : P_X(x) > 0\}$. $\mathcal{T}_\epsilon^n(P_X)$ denotes the set of ϵ-strongly letter-typical sequences of length n. We let $P_{X_1 X_2 \dots X_k}^n$ denotes the pmf of n i.i.d. random variables $X_1, X_2, \dots, X_k$, associated with the pmf $P_{X_1 X_2 \dots X_k}$. Finally, Markov chains, satisfying $P_{XYZ} = P_{XY} P_{Z|Y}$, are denoted by $X - Y - Z$.

III. Problem Definition

The point-to-point coordination setup we consider 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 ρ_ℓ, $\ell = 1, 2$. Thus, in designing a *block* scheme to coordinate n actions of the nodes, we assume $J \in [\![2^{nR_o}]\!]$, and $M_\ell \in [\![2^{n\rho_\ell}]\!]$, $k = 1, 2$, and we wish to communicate a codeword $A^n(I)$ over the DMC $P_{B|A}$ to Node Y, where I denotes the (appropriately selected) coordination message. The *codeword* $A^n(I)$ 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 channel output B^n, common randomness J, and local randomness M_2. We assume that the common randomness is independent of the action specified at Node X. A tuple (R_o, ρ_1, ρ_2) is deemed *achievable* if for each $\epsilon > 0$, there exists $n \in \mathbb{N}$ and a (strong coordination) coding scheme such that the joint pmf of actions $\hat{P}_{X^n Y^n}$ induced by this scheme and the n-fold product[3] of the desired joint pmf Q_{XY}^n are ϵ-*close* in total variation, i.e.,

$$\|\hat{P}_{X^n Y^n} - Q_{XY}^n\|_{TV} < \epsilon. \tag{1}$$

[3]This is the joint pmf of n i.i.d. copies of $(X,Y) \sim Q_{XY}$.

Remark 1: One straightforward observation is when the action at Node Y is a (deterministic) function of the action supplied to Node X by nature, i.e., $Y = f(X)$ and $H(Y|X) = 0$. For this case, the setup of strong coordination depicted in Fig. 1 is equivalent to the problem of lossless compression and communication of the source Y^n over a noisy communication channel. In fact, this particular case corresponds to the problem of simulating a noiseless link from a DMC where the common randomness is known to be useless (see [14], [27]). This equivalence can also be seen in terms the total variation distance between the code-induced joint pmf and the desired joint pmf and the probability of decoding error as follows

$$\begin{aligned}\|\hat{P}_{X^n Y^n} - Q_{XY}^n\|_{TV} &= \|Q_X^n \hat{P}_{Y^n|X^n} - Q_{XY}^n\|_{TV} \\ &= \sum_{x^n, y^n \neq f(x^n)} Q_X^n(x^n) \hat{P}_{Y^n|X^n}(y^n|x^n) \\ &\qquad - Q_{XY}^n(x^n, y^n) \\ &= \mathbb{P}[Y^n \neq f(X^n)],\end{aligned}$$

where it is well known that the probability of decoding error $\mathbb{P}[Y^n \neq f(X^n)] \to 0$, thus the strong coordination condition in (1) is satisfied, if $H(Y) \leq \mathbb{C}_{P_{B|A}}$. Here, $\mathbb{C}_{P_{B|A}}$ is the channel capacity for the channel $P_{B|A}$ defined as $\mathbb{C}_{P_{B|A}} \triangleq \max_{P_A} I(A;B)$.

We now present the achievable strong coordination schemes.

IV. Joint Coordination Channel Encoding

The first contribution of this work is the following characterization of the inner and outer bounds to the strong coordination capacity region.

Theorem 1 (Inner Bound): A tuple (R_o, ρ_1, ρ_2) is achievable for the noisy strong coordination setup in Fig. 1 if there exist auxiliary random variables (C, A) jointly correlated with the actions (X, Y) according to $P_{XYABC} = P_{AC} P_{X|AC} P_{B|A} P_{Y|BC}$, such that the marginal distribution $P_{XY} = Q_{XY}$, and

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

In the following, we characterize an outer bound to the strong coordination capacity region.

Theorem 2 (Outer Bound): If a tuple (R_o, ρ_1, ρ_2) is achievable for the noisy strong coordination setup in Fig. 1, then there must exist auxiliary random variables (C, A, D) jointly correlated with the actions (X, Y) according to $P_{ABCDXY} = P_{AX} P_{B|A} P_{CDY|ABX}$, such that the marginal distribution $P_{XY} = Q_{XY}$,

$$R_o + \rho_1 > I(Y; ACD|X) - H(B|A), \tag{3a}$$
$$R_o > I(XY; CD) - I(B; CD), \tag{3b}$$
$$I(X; C) < I(B; C), \tag{3c}$$

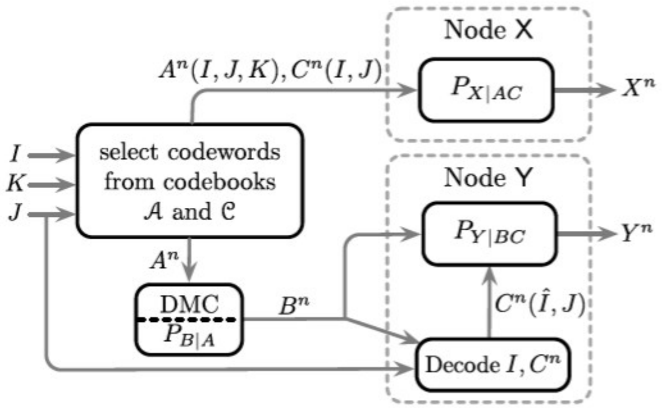

Fig. 2. A joint scheme for the allied problem.

where $|\mathcal{C}| \leq |\mathcal{A}||\mathcal{B}||\mathcal{X}||\mathcal{Y}|+5$, and $|\mathcal{D}| \leq (|\mathcal{A}||\mathcal{B}||\mathcal{X}||\mathcal{Y}|+5)|\mathcal{A}||\mathcal{B}||\mathcal{X}||\mathcal{Y}|$.

In following subsections, we first construct a joint coordination channel encoding scheme that achieve the rates of Theorem 1. Then, we show that for a special case the constructed scheme is optimal followed by the derivation of the outer bound of Theorem 2.

A. Inner Bound: Achievability

This scheme follows an approach similar to those in [3], [6], [7], [12] where coordination codes are designed based on allied channel resolvability problems [28]. The structure of the allied problem pertinent to the coordination problem at hand is given in Fig. 2. The aim of the allied problem is to generate[4] n symbols for two correlated sources X^n and Y^n whose joint statistics is close to Q^n_{XY} as defined by (1). To do so, we employ three independent and uniformly distributed messages I, K, and J and two codebooks $\mathcal{A}$ and $\mathcal{C}$ as shown in Fig. 2. To define the two codebooks, consider auxiliary random variables $A \in \mathcal{A}$ and $C \in \mathcal{C}$ jointly correlated with (X,Y) as $P_{XYABC} = P_{AC}P_{X|AC}P_{B|A}P_{Y|BC}$ and with marginal distribution $P_{XY} = Q_{XY}$.

From this factorization it can be seen that the scheme consists of two *reverse* test channels $P_{X|AC}$ and $P_{Y|AC}$ used to generate the sources from the codebooks. In particular, $P_{Y|AC} = \sum_b P_{B=b|A}P_{Y|B=b,C}$, i.e., the randomness of the DMC contributes to the randomized generation of Y^n.

Generating X^n and Y^n from I, K, J represents a complex channel resolvability problem with the following ingredients:

i) Nested codebooks: Codebook $\mathcal{C}$ of size $2^{n(R_o+R_c)}$ is generated i.i.d. according to pmf P_C, i.e., $C^n_{ij} \sim \prod_{l=1}^n P_C(\cdot)$ for all $(i,j) \in \mathcal{I}\times\mathcal{J}$. Codebook $\mathcal{A}$ is generated by randomly selecting $A^n_{ijk} \sim \prod_{l=1}^n P_{A|C}(\cdot|C^n_{ij})$ for all $(i,j,k) \in \mathcal{I}\times\mathcal{J}\times\mathcal{K}$, where $\mathcal{I} \triangleq [\![2^{nR_c}]\!]$, $\mathcal{J} \triangleq [\![2^{nR_o}]\!]$, and $\mathcal{K} \triangleq [\![2^{nR_a}]\!]$.

ii) Encoding functions:
$C^n : [\![2^{nR_c}]\!] \times [\![2^{nR_o}]\!] \to \mathcal{C}^n$,
$A^n : [\![2^{nR_c}]\!] \times [\![2^{nR_o}]\!] \times [\![2^{nR_a}]\!] \to \mathcal{A}^n$.

iii) 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^n_{IJ} and A^n_{IJK} from codebooks $\mathcal{C}$ and $\mathcal{A}$.

iv) The selected codewords C^n_{IJ} and A^n_{IJK} are then passed through DMC $P_{X|AC}$ at Node X, while at Node Y, codeword A^n_{IJK} is sent through DMC $P_{B|A}$ whose output B^n is used to decode codeword $C^n_{\hat{I}J}$ and both are then passed through DMC $P_{Y|BC}$ to obtain Y^n.

Since the codewords are randomly chosen, the induced joint pmf of the generated actions and codeword indices in the allied problem is itself a random variable and depends on the random codebook. Given a realization of the codebooks

$$\mathsf{C} \triangleq (\mathcal{A},\mathcal{C}) = \left\{ a^n_{ijk}, c^n_{ij} : \substack{i\in[\![2^{nR_c}]\!] \\ j\in[\![2^{nR_o}]\!] \\ k\in[\![2^{nR_a}]\!]} \right\}, \tag{4}$$

the code-induced joint pmf of the actions and codeword indices in the allied problem is given by

$$\mathring{P}_{X^nY^nIJK}(x^n,y^n,i,j,k) \triangleq \frac{P^n_{X|AC}(x^n|a^n_{ijk}c^n_{ij})}{2^{n(R_c+R_o+R_a)}} \times \Big(\sum_{b^n,\hat{i}} P^n_{B|A}(b^n|a^n_{ijk})\mathsf{P}_{\hat{I}|B^nJ}(\hat{i}|b^n,j)P^n_{Y|BC}(y^n|b^n c^n_{\hat{i}j})\Big), \tag{5}$$

where $\mathsf{P}_{\hat{I}|B^nJ}$ denotes the operation of decoding the index I using the common randomness and the channel output at Node Y. Note that the indices for the C-codeword that generate X and Y sequences in (5) can be different since the decoding of the index I at Node Y may fail. We are done if we accomplish the following tasks:

(1) identify conditions on R_o, R_c, R_a under which the code-induced pmf $\mathring{P}_{X^nY^n}$ is *close* to the design pmf Q^n_{XY} with respect to total variation, i.e.,

$$\lim_{n\to\infty} \mathbb{E}_{\mathsf{C}}\left[||\mathring{P}_{X^nY^n} - Q^n_{XY}\|_{TV}\right] = 0$$

(2) devise a strong coordination scheme by inverting the operation at Node X in Fig. 2 by enforcing independence between the action sequence X^n and the common randomness J. This translates to identifying the conditions on R_c, R_a under which the code-induced joint distribution of X^n and J, $\check{P}_{X^nJ}$ is *close* to the product of the marginal distributions, i.e.,

$$\lim_{n\to\infty} \mathbb{E}_{\mathsf{C}}\left[||\check{P}_{X^nJ} - Q^n_X P_J||_{TV}\right] = 0,$$

and the mutual information between the action sequence X^n and the common randomness J is asymptotically equal to zero, i.e., $\lim_{n\to\infty} I(X^n;J) = 0$.

Note that in the above, we let $\mathbb{E}_{\mathsf{C}}$ to denote the expectation over the random realization of the codebooks. Moreover, during our analysis, we bound the total variation distance closeness via bounding the Kullback-Leibler (KL) divergence of the relative distributions then utilizing Pinsker's inequality [29] to relate the two measurements. This techniques, first used

[4]Note that, at this point, we assume that the action sequence X^n is not available as an input to Node X but is generated as an output. We will show later on in this subsection that this operation can be reversed to map back to the original problem setup.

in [30], is commonly used in analyzing strong coordination problems (e.g., [6], [7] and [15]), since the required analysis can be carried out more easily than the total variation analysis. This will be done in following sections by subdividing the analysis of the allied problem.

1) Resolvability Constraints: Assuming that the decoding of I and the codeword C^n_{IJ} occurs perfectly at Node Y, we see that the code-induced joint pmf induced by the allied scheme for the realization of the codebook C in (4) is

$$\check{P}_{X^nY^nIJK}(x^n, y^n, i, j, k) = \frac{P^n_{X|AC}(x^n|a^n_{ijk}c^n_{ij})}{2^{n(R_c+R_o+R_a)}} \times \Big(\sum_{b^n} P^n_{B|A}(b^n|a^n_{ijk}) P^n_{Y|BC}(y^n|b^n c^n_{ij}) \Big). \quad (6)$$

The following result quantifies when the above induced distribution is close to the n-fold product of the design pmf Q_{XY}.

Lemma 1 (Resolvability Constraints): The total variation between the code-induced pmf $\check{P}_{X^nY^n}$ in (6) and the desired pmf Q^n_{XY} asymptotically vanishes, i.e., $\mathbb{E}_{\mathsf{C}}\big[\|\check{P}_{X^nY^n} - Q^n_{XY}\|_{TV}\big] \to 0$ as $n \to \infty$, if

$$R_a + R_o + R_c > I(XY; AC), \quad (7)$$
$$R_o + R_c > I(XY; C). \quad (8)$$

Note that in the above, we let $\mathbb{E}_{\mathsf{C}}$ to denote the expectation over the random realization of the codebooks.

Proof: In the following, we drop the subscripts from the pmfs for simplicity, e.g., $P^n_{X|AC}(x^n|A^n_{ijk}, C^n_{ij})$ will be denoted by $P(x^n|A^n_{ijk}, C^n_{ij})$, and $Q^n_{XY}(x^n, y^n)$ will be denoted by $Q(x^n, y^n)$, respectively. Let $R \triangleq R_a + R_c + R_o$, and choose $\epsilon > 0$. Consider the derivation for $\mathbb{E}_{\mathsf{C}}\big[\mathbb{D}(\check{P}_{X^nY^n}||Q^n_{XY})\big]$ shown at the bottom of the next page.

In this argument:

- (a) follows from the law of iterated expectations, where the inner conditional expectation denotes the expectation over all random codewords $\{A^n_{i'j'k'}, C^n_{i',j'} : i' \neq i, j' \neq j, k' \neq k\}$ given the codewords $A^n_{i,j,k}, C^n_{i,j}$. Note that we have used (a^n_{ijk}, c^n_{ij}) to denote the codewords corresponding to the indices (i, j, k), and $(a^n_{i'j'k'}, c^n_{i'j'})$ to denote the codewords corresponding to the indices (i', j', k').
- (b) follows from Jensen's inequality [31].
- (c) follows from dividing the inner summation over the indices (i', j', k') into three subsets based on the indices (i, j, k) from the outer summation.
- (d) follows from taking the expectation within the subsets in (c) such that when
  - $(i', j') = (i, j), (k' \neq k)$: $a^n_{i'j'k'}$ is conditionally independent of a^n_{ijk} following the nature of the codebook construction (i.e., i.i.d. at random);
  - $(i', j') \neq (i, j)$: both codewords (a^n_{ijk}, c^n_{ij}) are independent of $(a^n_{i'j'k'}, c^n_{i'j'})$ regardless of the value of k. As a result, the expected value of the induced distribution with respect to the input codebooks is the desired distribution Q^n_{XY} [3].
- (e) follows from
  - $(i', j', k') = (i, j, k)$: there is only one pair of codewords (a^n_{ijk}, c^n_{ij});
  - when $(k' \neq k)$ while $(i', j') = (i, j)$ there are $(2^{nR_a} - 1)$ indices in the sum;
  - $(i', j') \neq (i, j)$: the number of the indices is at most 2^{nR}.
- (f) results from splitting the outer summation: The first summation contains typical sequences and is bounded by using the probabilities of the typical set. The second summation contains the tuple of sequences when the pair of actions sequences x^n, y^n and codewords c^n, a^n are not ϵ-jointly typical (i.e., $(x^n, y^n, a^n, c^n) \notin \mathcal{T}^n_\epsilon(P_{XYAC})$). This sum is upper bounded following [6] with $\mu_{XY} = \min^*_{x,y}\big(P_{XY}(x, y)\big)$.
- (g) follows from Chernoff bound on the probability that a sequence is not strongly typical [32] where $\mu_{XYAC} = \min^*_{x,y,a,c}(P_{XYAC}(x, y, a, c))$ and $\delta(\epsilon)$ denotes a positive function of ϵ that vanishes as n goes to infinity, i.e., $\delta(\epsilon) \to 0$ as $n \to \infty$.
- (h) Consequently, the contribution of typical sequences can be made asymptotically smaller than some $\epsilon' > 0$ if

  $$R_a + R_o + R_c > I(XY; AC), \quad R_o + R_c > I(XY; C),$$

  while the second term converges to zero exponentially fast with n, i.e.,

  $$\big(2|\mathcal{X}||\mathcal{Y}||\mathcal{A}||\mathcal{C}|e^{-n\epsilon^2\mu_{XYAC}}\big)\log(2\mu^{-n}_{XY} + 1) \xrightarrow{n\to\infty} 0$$

  and $\epsilon' \to 0$ as $n \to \infty$.

Finally, if (7) and (8) are satisfied, by applying Pinsker's inequality [29] we have

$$\mathbb{E}_{\mathsf{C}}\big[||\check{P}_{X^nY^n} - Q^n_{XY}||_{TV}\big] \leq \mathbb{E}_{\mathsf{C}}\Big[\sqrt{2\mathbb{D}(\check{P}_{X^nY^n}||Q^n_{XY})}\Big] \leq \sqrt{2\mathbb{E}_{\mathsf{C}}\big[\mathbb{D}(\check{P}_{X^nY^n}||Q^n_{XY})\big]} \xrightarrow{n\to\infty} 0. \quad (9)$$

■

Remark 2: Given $\epsilon > 0$, R_a, R_o, R_c satisfying (7) and (8), it follows from (9) that there exist an $n \in \mathbb{N}$ and a random codebook realization for which the code-induced pmf between the indices and the pair of actions satisfies

$$||\check{P}_{X^nY^n} - Q^n_{XY}||_{TV} < \epsilon. \quad (10)$$

2) Decodability Constraint: Since the operation at Node Y in Fig. 2 involves the decoding of I and thereby the codeword $C^n(I, J)$ using B^n and J, the induced distribution of the scheme for the allied problem that is given in (5) will not match that of (6) unless and until we ensure that the decoding succeeds with high probability as $n \to \infty$. The following lemma quantifies the necessary rate for this decoding to succeed asymptotically almost always.

Lemma 2 (Decodability Constraint): Let $\hat{I}, C^n_{\hat{I}J}$ be the output of a typicality-based decoder that uses common randomness J to decode the index I and the sequence C^n_{IJ} from B^n. Let $\mathbb{P}[\hat{I} \neq I]$ be the probability that the decoding fails for a realization of the random codebook. If the rate for the index I satisfies $R_c < I(B; C)$ then,

i) $\mathbb{E}_{\mathsf{C}}\big[\mathbb{P}[\hat{I} \neq I]\big] \to 0$ as $n \to \infty$, and

ii) $\lim_{n\to\infty} \mathbb{E}_{\mathsf{C}}\big[\|\check{P}_{X^nY^nIJK} - \mathring{P}_{X^nY^nIJK}\|_{TV}\big] = 0.$

Proof: We start the proof of *i)* by calculating the average probability of error, averaged over all codewords in the codebook and averaged over all random codebook realizations as follows:

$$\begin{aligned}\mathbb{E}_{\mathsf{C}}\left[\mathbb{P}[\hat{I}\neq I]\right] &= \sum_{\mathsf{C}} P_{\mathsf{C}}(\mathsf{c})\mathbb{P}[\hat{I}\neq I]\\ &= \sum_{\mathsf{C}} P_{\mathsf{C}}(\mathsf{c})\sum_{i,j,k}\frac{1}{2^{nR}}\mathbb{P}\left[\hat{I}\neq I\middle|\substack{I=i\\J=j\\K=k}\right]\\ &= \sum_{i,j,k}\frac{1}{2^{nR}}\sum_{\mathsf{C}} P_{\mathsf{C}}(\mathsf{c})\mathbb{P}\left[\hat{I}\neq I\middle|\substack{I=i\\J=j\\K=k}\right]\\ &\overset{(a)}{=} \mathbb{P}\left[\hat{I}\neq I\middle|\substack{I=1\\J=1\\K=1}\right],\end{aligned} \tag{11}$$

where in (a) we have used the fact that the conditional probability of error is independent of the triple of indices due to the i.i.d nature of the codebook construction. Also, due to the random construction and the properties of jointly typical

$$\begin{aligned}
&\mathbb{E}_{\mathsf{C}}\left[\mathbb{D}(\check{P}_{X^nY^n}\|Q^n_{XY})\right]\\
&= \mathbb{E}_{\mathsf{C}}\left[\sum_{x^n,y^n}\left(\sum_{i,j,k}\frac{P(x^n|A^n_{ijk},C^n_{ij})P(y^n|A^n_{ijk},C^n_{ij})}{2^{nR}}\right)\log\left(\sum_{i',j',k'}\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}Q(x^n,y^n)}\right)\right]\\
&= \mathbb{E}_{\mathsf{C}}\left[\sum_{x^n,y^n}\sum_{i,j,k}\left(\frac{P(x^n|A^n_{ijk},C^n_{ij})P(y^n|A^n_{ijk},C^n_{ij})}{2^{nR}}\log\left(\sum_{i',j',k'}\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}Q(x^n,y^n)}\right)\right)\right]\\
&\overset{(a)}{=}\sum_{x^n,y^n}\sum_{i,j,k}\mathbb{E}_{A^n_{ijk}C^n_{ij}}\left[\frac{P(x^n|A^n_{ijk},C^n_{ij})P(y^n|A^n_{ijk},C^n_{ij})}{2^{nR}}\mathbb{E}_{\text{rest}}\left[\log\left(\sum_{i',j',k'}\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}Q(x^n,y^n)}\right)\middle|A^n_{ijk}C^n_{ij}\right]\right]\\
&\overset{(b)}{\leq}\sum_{x^n,y^n}\sum_{i,j,k}\mathbb{E}_{A^n_{ijk}C^n_{ij}}\left[\frac{P(x^n|A^n_{ijk},C^n_{ij})P(y^n|A^n_{ijk},C^n_{ij})}{2^{nR}}\log\left(\mathbb{E}_{A^n_{i'j'k'}C^n_{i'j'}}\left[\sum_{i',j',k'}\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}Q(x^n,y^n)}\middle|A^n_{ijk}C^n_{ij}\right]\right)\right]\\
&\overset{(c)}{=}\sum_{x^n,y^n}\sum_{i,j,k}\sum_{a^n_{ijk},c^n_{ij}}\frac{P(x^n,y^n,a^n_{ijk},c^n_{ij})}{2^{nR}}\log\left(\mathbb{E}_{A^n_{i'j'k'}C^n_{i'j'}}\left[\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}Q(x^n,y^n)}\middle|A^n_{ijk}C^n_{ij}\right]\right.\\
&\qquad+\sum_{\substack{i',j',k':\\(i',j')=(i,j),(k'\neq k)}}\mathbb{E}_{A^n_{i'j'k'}C^n_{i'j'}}\left[\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}Q(x^n,y^n)}\middle|A^n_{ijk}C^n_{ij}\right]\\
&\qquad\left.+\sum_{\substack{i',j',k':\\(i',j')\neq(i,j)}}\mathbb{E}_{A^n_{i'j'k'}C^n_{i'j'}}\left[\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}Q(x^n,y^n)}\middle|A^n_{ijk}C^n_{ij}\right]\right)\\
&\overset{(d)}{=}\sum_{x^n,y^n}\sum_{i,j,k}\sum_{a^n_{ijk},c^n_{ij}}\frac{P(x^n,y^n,a^n_{ijk},c^n_{ij})}{2^{nR}}\log\left(\frac{P(x^n,y^n|a^n_{ijk},c^n_{ij})}{2^{nR}Q(x^n,y^n)}+\sum_{\substack{i',j',k':\\(i',j')=(i,j),(k'\neq k)}}\frac{P(x^n,y^n|c^n_{ij})}{2^{nR}Q(x^n,y^n)}\right.\\
&\qquad\left.+\sum_{\substack{i',j',k':\\(i',j')\neq(i,j)}}\frac{Q(x^n,y^n)}{2^{nR}Q(x^n,y^n)}\right)\\
&\overset{(e)}{\leq}\sum_{x^n,y^n,a^n_{ijk},c^n_{ij}}P(x^n,y^n,a^n_{ijk},c^n_{ij})\log\left(\frac{P(x^n,y^n|a^n_{ijk},c^n_{ij})}{2^{nR}Q(x^n,y^n)}+(2^{nR_a})\frac{P(x^n,y^n|c^n_{ij})}{2^{nR}Q(x^n,y^n)}+1\right)\\
&\overset{(f)}{\leq}\left[\sum_{(x^n,y^n,a^n,c^n)\in\mathcal{T}^n_\epsilon(P_{XYAC})}P(x^n,y^n,a^n,c^n)\log\left(\frac{2^{-nH(XY|AC)(1-\epsilon)}}{2^{nR}2^{-nH(XY)(1+\epsilon)}}+\frac{2^{-nH(XY|C)(1-\epsilon)}}{2^{n(R_o+R_c)}2^{-nH(XY)(1+\epsilon)}}+1\right)\right]\\
&\qquad+\mathbb{P}\big((x^n,y^n,a^n,c^n)\notin\mathcal{T}^n_\epsilon(P_{XYAC})\big)\log(2\mu^{-n}_{XY}+1)\\
&\overset{(g)}{\leq}\left[\sum_{(x^n,y^n,a^n,c^n)\in\mathcal{T}^n_\epsilon(P_{XYAC})}P(x^n,y^n,a^n,c^n)\log\left(\frac{2^{n(I(XY;AC)+\delta(\epsilon))}}{2^{nR}}+\frac{2^{n(I(XY;C)+\delta(\epsilon))}}{2^{n(R_o+R_c)}}+1\right)\right]\\
&\qquad+\left(2|\mathcal{X}||\mathcal{Y}||\mathcal{A}||\mathcal{C}|e^{-n\epsilon^2\mu_{XYAC}}\right)\log(2\mu^{-n}_{XY}+1)\\
&\overset{(h)}{\leq}\epsilon'.
\end{aligned}$$

set, we have

$$\mathbb{P}\big((A^n_{111}, B^n, C^n_{11}) \in \mathcal{T}^n_\epsilon(P_{ABC})\big) \xrightarrow{n\to\infty} 1.$$

We now continue the proof by constructing the sets for each j and $b^n \in \mathcal{B}^n$ that Node Y will construct to identify the transmitted index as

$$\hat{S}_{j,b^n,\mathsf{C}} \triangleq \{i : (b^n, c^n_{ij}) \in \mathcal{T}^n_\epsilon(P_{BC})\}.$$

The set $\hat{S}_{j,b^n,\mathsf{C}}$ consists of indices $i \in I$ such that for a given common randomness index $J = j$ and channel realization $B^n = b^n$, the sequences (b^n, c^n_{ij}) are jointly-typical. Assuming $(i,j,k) = (1,1,1)$ was realized, and if $\hat{S}_{1,b^n,\mathsf{C}} = \{1\}$, then the decoding will be successful. The probability of this event occurring is divided into two steps as follows.

- First, assuming $(i,j,k) = (1,1,1)$ was realized, for successful decoding, 1 must be an element of $\hat{S}_{J,B^n,\mathsf{C}}$. The probability of this event can be bounded as follows:

$$\begin{aligned}
&\mathbb{E}_{\mathsf{C}}\Big[\mathbb{P}\Big[I \in \hat{S}_{J,B^n,\mathsf{C}}\Big|\substack{I=1\\J=1\\K=1}\Big]\Big]\\
&= \sum_{a^n,b^n,c^n}\Big(P^n_C(c^n)P^n_{A|C}(a^n|c^n)P^n_{B|A}(b^n|a^n)\\
&\qquad\qquad \times \mathbb{1}\big((c^n,b^n) \in \mathcal{T}^n_\epsilon(P_{BC})\big)\Big)\\
&= \sum_{b^n,c^n} P^n_{BC}(b^n,c^n)\mathbb{1}\big((b^n,c^n) \in \mathcal{T}^n_\epsilon(P_{BC})\big)\\
&\overset{(a)}{\geq} 1-\delta(\epsilon) \xrightarrow{n\to\infty} 1,
\end{aligned}$$

where (a) follows from the properties of jointly typical sets and $\delta(\epsilon) \to 0$ as $n \to \infty$.
- Next, assuming again that $(i,j,k) = (1,1,1)$ was realized, for successful decoding, no index greater than or equal to 2 must be an element of $\hat{S}_{J,B^n,\mathsf{C}}$. The probability of this event can be bounded as follows:

$$\begin{aligned}
&\mathbb{E}_{\mathsf{C}}\mathbb{P}\Big[\hat{S}_{J,B^n,\mathsf{C}} \cap \{2,\ldots,2^{nR_c}\} = \emptyset\Big|\substack{I=1\\J=1\\K=1}\Big]\\
&= 1 - \sum_{i'\neq 1}\mathbb{E}_{\mathsf{C}}\mathbb{P}\Big[i' \in \hat{S}_{J,B^n,\mathsf{C}}\Big|\substack{I=1\\J=1\\K=1}\Big]\\
&= 1 - \sum_{i'\neq 1}\mathbb{P}[(C^n_{i'1}, B^n) \in \mathcal{T}^n_\epsilon(P_{BC})]\\
&\overset{(a)}{\geq} 1 - \sum_{i'\neq 1} 2^{-n(I(B;C)-\delta(\epsilon))}\\
&= 1 - (2^{nR_c}-1)2^{-n(I(B;C)-\delta(\epsilon))}\\
&= 1 - 2^{-n(I(B;C)-R_c-\delta(\epsilon))} + 2^{-nI(B;C)}\\
&\overset{(b)}{\geq} 1-\delta(\epsilon) \xrightarrow{n\to\infty} 1,
\end{aligned}$$

where (a) follows from the packing lemma [33], and (b) results if $R_c < I(B;C) - \delta(\epsilon)$ and sufficiently large n yield $\delta(\epsilon) \to 0$. Then from (11), the claim in *i)* follows as given by

$$\begin{aligned}
\mathbb{E}_{\mathsf{C}}\big[\mathbb{P}[\hat{I} \neq I]\big] &= \mathbb{E}_{\mathsf{C}}\mathbb{P}\Big[\hat{I} \neq I\Big|\substack{I=1\\J=1\\K=1}\Big]\\
&\leq \Big(\mathbb{E}_{\mathsf{C}}\mathbb{P}\Big[I \notin \hat{S}_{J,B^n,\mathsf{C}}\Big|\substack{I=1\\J=1\\K=1}\Big]\\
&\quad + \mathbb{E}_{\mathsf{C}}\mathbb{P}\Big[\hat{S}_{J,B^n,\mathsf{C}} \cap \{2,\ldots,2^{nR_c}\} \neq \emptyset\Big|\substack{I=1\\J=1\\K=1}\Big]\Big)\\
&\xrightarrow{n\to\infty} 0.
\end{aligned}$$

Finally, the proof of *ii)* follows in a straightforward manner. If the previous two conditions are met, then $\mathbb{E}_{\mathsf{C}}[\mathbb{P}[\hat{I} \neq I]] \to 0$ and $\mathbb{E}_{\mathsf{C}}[P_{\hat{I}|B^nJ}(\hat{i}|b^n,j)] \to \delta_{I\hat{I}}$, where $\delta_{I\hat{I}}$ denotes the Kronecker delta. Consequently, the claim then follows by simple algebraic manipulation of (5) and (6) as

$$\lim_{n\to\infty}\mathbb{E}_{\mathsf{C}}\big[\|\check{P}_{X^nY^nIJK} - \mathring{P}_{X^nY^nIJK}\|_{TV}\big] = 0. \tag{12}$$

■

3) Independence Constraint: We complete modifying the allied structure in Fig. 2 to mimic to the original problem with a final step. By assumption, we have a natural independence between the action sequence X^n and the common randomness J. As a result, the joint distribution over X^n and J in the original problem is a product of the marginal distributions Q^n_X and P_J. To mimic this behavior in the scheme for the allied problem, we artificially enforce independence by ensuring that the mutual information between X^n and J vanishes. This process is outlined in Lemma 3.

Lemma 3 (Independence constraint): Consider the scheme for the allied problem given in Fig. 2. Both $I(J;X^n) \to 0$ and $\mathbb{E}_{\mathsf{C}}\big[||\check{P}_{X^nJ} - Q^n_X P_J||_{TV}\big] \to 0$ as $n \to \infty$ if the code rates satisfy

$$R_a + R_c > I(X;AC), \tag{13}$$
$$R_c > I(X;C). \tag{14}$$

The proof of Lemma 3, shown in Appendix A, builds on the results of Section IV-A2 and the proof of Lemma 1 in Section IV-A1, resulting in

$$\begin{aligned}
\mathbb{E}_{\mathsf{C}}\big[||\check{P}_{X^nJ} - Q^n_X P_J||_{TV}\big] &\leq \mathbb{E}_{\mathsf{C}}\Big[\sqrt{2\mathbb{D}(\check{P}_{X^nJ}||Q^n_X P_J)}\,\Big]\\
&\leq \sqrt{2\mathbb{E}_{\mathsf{C}}\big[\mathbb{D}(\check{P}_{X^nJ}||Q^n_X P_J)\big]} \xrightarrow{n\to\infty} 0.
\end{aligned} \tag{15}$$

Remark 3: Given $\epsilon > 0$, R_a, R_c meeting (13) and (14), it follows from (15) that there exists an $n \in \mathbb{N}$ and a random codebook realization for which the code-induced pmf between the common randomness J and the actions of Node X satisfies

$$||\check{P}_{X^nJ} - Q^n_X P_J||_{TV} < \epsilon. \tag{16}$$

In the original problem of Fig. 1, the input action sequence X^n and the index J from the common randomness source are available and the A- and C-codewords are to be selected. Now, to devise a scheme for the strong coordination problem, we proceed as follows. We let Node X choose indices I and K (and, consequently, the A- and C-codewords) from the realized X^n and J using the conditional distribution $\mathring{P}_{I,K|X^n,J}$. The joint pmf of the actions and the indices is then given by

$$\hat{P}_{X^nY^nIJK} \triangleq Q^n_X P_J \mathring{P}_{I,K|X^n,J}\mathring{P}_{Y^n|I,J,K}. \tag{17}$$

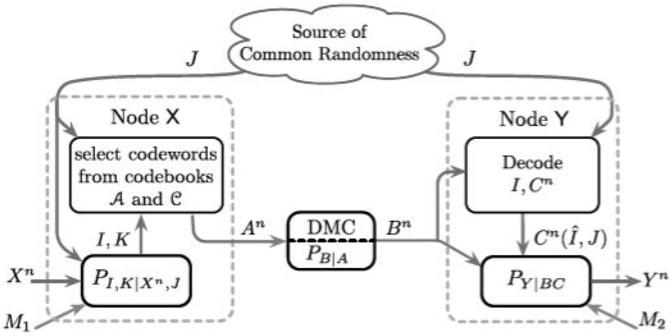

Fig. 3. The joint strong coordination encoding scheme.

As a result, from the allied scheme of Fig. 2 we obtain the joint scheme illustrated in Fig. 3.
Finally, we can argue that

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{C}}[\|\hat{P}_{X^n,Y^n} - Q^n_{XY}\|_{TV}] = 0, \tag{18}$$

since the total variation between the marginal pmf $\hat{P}_{X^n,Y^n}$ and the design pmf Q^n_{XY} can be bounded as

$$\begin{aligned}
&\|\hat{P}_{X^nY^n} - Q^n_{XY}\|_{TV}\\
&\quad\overset{(a)}{\leq} \|\hat{P}_{X^nY^n} - \check{P}_{X^nY^n}\|_{TV} + \|\check{P}_{X^nY^n} - Q^n_{XY}\|_{TV}\\
&\quad\overset{(b)}{\leq} \|\hat{P}_{X^nY^nIJK} - \check{P}_{X^nJ}\mathring{P}_{IKY^n|X^n,J}\|_{TV}\\
&\qquad + \|\check{P}_{X^nY^n} - Q^n_{XY}\|_{TV}\\
&\quad\overset{(c)}{=} \|Q^n_X P_J - \check{P}_{X^nJ}\|_{TV} + \|\check{P}_{X^nY^n} - Q^n_{XY}\|_{TV}
\end{aligned}$$

where
(a) follows from the triangle inequality;
(b) follows from (17) and [5, Lemma V.1];
(c) follows from [5, Lemma V.2].
Note that the terms on the RHS of the above equation can be made vanishingly small provided the resolvability, decodability, and independence conditions are met. Thus, by satisfying the conditions stated in Lemmas 1-3, the coordination scheme defined by (17) achieves strong coordination asymptotically between Nodes X and Y by communicating over the DMC $P_{B|A}$. Note that since the operation at Nodes X and Y amount to index selection according to $\mathring{P}_{I,K|X^n,J}$, and generation of Y^n using the DMC $P_{Y|BC}$, both operations are randomized. The last step consists in viewing the local randomness as the source of randomness in the operations at Nodes X and Y. This is detailed in the following paragraph.

4) Local Randomness Rates: As seen from Fig. 3, at Node X, local randomness M_1 is employed to randomize the selection of indices (I, K) by synthesizing the channel $\mathring{P}_{IK|X^nJ}$ whereas Node Y utilizes its local randomness M_2 to generate the action sequence Y^n by simulating the channel $P_{Y|BC}$. Using the list decoding and likelihood arguments of [7, Section IV.B], [5, Section III.E], we can argue that for any given realizations of J, the minimum rate of local randomness required for the probabilistic selection of indices I, K can be derived by quantifying the number of A and C codewords (equally identifying a list of index tuples (I, K)) jointly typical with $X^n = x^n$. Quantifying the list size as in [7] yields $\rho_1 \geq R_a + R_c - I(X; AC)$. At Node Y, the necessary local randomness for the generation of the action sequence is bounded by the channel simulation rate of DMC $P_{Y|BC}$ [23]. Thus, $\rho_2 \geq H(Y|BC)$. Combining the local randomness rates constraints with the constraints in Lemmas 1-3, we obtain the inner bound to the strong coordination region

$$R_a + R_o + R_c > I(XY; AC), \tag{19a}$$
$$R_o + R_c > I(XY; C), \tag{19b}$$
$$R_a + R_c > I(X; AC), \tag{19c}$$
$$R_c > I(X; C), \tag{19d}$$
$$R_c < I(B; C), \tag{19e}$$
$$\rho_1 > R_a + R_c - I(X; AC), \tag{19f}$$
$$\rho_2 > H(Y|BC). \tag{19g}$$

In this work, we are in particular interested in both local and common randomness rates. Therefore, we deploy Fourier-Motzkin elimination [34] on the rate constraints of (19) to obtain the rates constraints of Theorem 1.

Remark 4: Due to the equivalence between the noisy point-to-point strong coordination problem and the average DMC simulation problem, the achievable strong coordination rates of Theorem 1 are equivalent to the channel simulation rates derived in [14, Theorem 4] when we assume infinite local randomness. That is, when ρ_1 and ρ_2 are infinite, the rate constraints of Theorem 1 reduce to

$$R_o + I(B; C) > I(XY; C),$$
$$I(X; C) < I(B; C).$$

B. Capacity Region for a Special Case

Here we characterize a special case of the proposed joint coordination-channel encoding scheme in which the scheme is optimal.

1) Deterministic Channel: This is the case when the channel output B is a deterministic function of the channel input A i.e., $H(B|A) = 0$. Although this special case is discussed in the context of simulating a DMC channel over a deterministic channel [14], we present this case in the context of our achievable construction with rates as stated in Theorem 1.

In this case, we select the auxiliary random variables A and C as follows. Let $C = (U, A)$ and select A independent of X, Y, and U with P_A to be the capacity achieving input distribution of the channel $P_{B|A}$, i.e., $\mathbb{C}_{P_{B|A}} = \max_{P_A} H(B)$. Let U be an auxiliary random variable related to X and Y via the Markov chain $X - U - Y$. As a result, the joint distribution of Theorem 1 takes the form $P_U P_A P_{B|A} P_{X|U} P_{Y|U}$ and the problem reduces to a two-terminal strong coordination over a noiseless channel and a separate channel coding problem. Accordingly, from Theorem 1, the following rates are achievable.

$$R_o + \rho_1 \overset{(a)}{\geq} I(Y; U|X) \tag{20a}$$

$$\begin{aligned} R_o &\overset{(b)}{\geq} I(XY; U) - I(A; B)\\ &= I(XY; U) - \mathbb{C}_{P_{B|A}} \end{aligned} \tag{20b}$$

$$I(X;U) \overset{(c)}{\leq} \mathbb{C}_{P_{B|A}} \tag{20c}$$

$$\rho_2 \overset{(d)}{\geq} H(Y|U) \tag{20d}$$

where (a)-(d) follows from the choice of A and $C = (U, A)$; (b)-(c) follows from the fact that the channel $P_{B|A}$ is deterministic and from the selection of P_A to be the capacity achieving input distribution.

Now, the optimality of (20a)-(20d) follows in a straightforward way from channel coding for and strong coordination over noise-free channels [7, Theorem 3] for the special case of a single hop.

C. Outer Bound: Converse

Let $\epsilon > 0$ and the tuple (R_o, ρ_1, ρ_2) is achievable by the length-n code that induces a distribution $\hat{P}_{X^n,Y^n}$ that satisfies

$$\|\hat{P}_{X^n,Y^n} - P_{XY}^{\otimes n}\|_{TV} \leq \epsilon, \tag{21}$$

Let T be is a time-sharing random variable uniformly distributed on the set $\{1,\ldots,n\}$ and independent of the induced joint distribution. We infer from [35, Sec.V.A] the following useful inequalities,

$$|H(X^n,Y^n) - \sum_{t=1}^{n} H(X_t,Y_t)| \leq n\delta_\epsilon,$$
$$|H(Y^n|X^n) - \sum_{t=1}^{n} H(Y_t|X_t)| \leq n\delta_\epsilon,$$
$$I(X_T,Y_T;T) \leq n\delta\epsilon, \tag{22}$$

for some $\delta_\epsilon \to 0$ as $\epsilon \to 0$.

Step 1, consider the constraint of (3a) for $R_o + \rho_1$.

$$\begin{aligned}
n(R_o+\rho_1) &= H(JM_1) \\
&\geq I(X^nY^n;JM_1) \\
&\overset{(a)}{=} I(Y^n;JM_1|X^n) \\
&\overset{(b)}{=} I(Y^n;A^nJM_1|X^n) \\
&\geq I(Y^n;A^nJ|X^n) \\
&\overset{(c)}{=} \sum_{t=1}^{n} \left[I(Y^t;B_{t+1}^nA^nJ|X^n) - I(Y^{t-1};B_t^nA^nJ|X^n)\right] \\
&\overset{(d)}{=} \sum_{t=1}^{n} \left[I(Y_t;B_{t+1}^nA^nJ|X^nY^{t-1}) \right.\\
&\qquad \left. -I(Y^{t-1};B_t|X^nB_{t+1}^nA^nJ)\right] \\
&\overset{(e)}{\geq} \sum_{t=1}^{n} \left[I(Y_t;B_{t+1}^nA^nJX^{t-1}Y^{t-1}|X_t) \right.\\
&\qquad \left. -I(Y^{t-1};B_t|X^nB_{t+1}^nA^nJ)\right] - n\delta_\epsilon \\
&\overset{(f)}{\geq} \sum_{t=1}^{n} \left[I(Y_t;A_t(JB_{t+1}^nX^{t-1})Y^{t-1}|X_t) \right.\\
&\qquad \left. -I(Y^n;B_t|X^nB_{t+1}^nA^nJ)\right] - n\delta_\epsilon \\
&\overset{(g)}{=} \sum_{t=1}^{n} I(Y_t;A_t(JB_{t+1}^nX^{t-1})Y^{t-1}|X_t) \\
&\qquad - I(Y^n;B^n|X^nA^nJ) - n\delta_\epsilon \\
&\overset{(h)}{\geq} \sum_{t=1}^{n} I(Y_t;A_t(JB_{t+1}^nX^{t-1})Y^{t-1}|X_t) \\
&\qquad - H(B^n|A^n) - n\delta_\epsilon \\
&\overset{(i)}{=} \sum_{t=1}^{n} \left[I(Y_t;A_t(JB_{t+1}^nX^{t-1})Y^{t-1}|X_t) - H(B_t|A_t)\right] - n\delta_\epsilon \\
&\overset{(k)}{=} \sum_{t=1}^{n} \left[I(Y_t;A_tC_tD_t|X_t) - H(B_t|A_t)\right] - n\delta_\epsilon \\
&= nI(Y_T;A_TC_TD_T|X_TT) - nH(B_T|A_TT) - n\delta_\epsilon \\
&\overset{(l)}{=} nI(Y_T;A_TC_TD_TT|X_T) - nI(Y_T;T|X_T) - n\delta_\epsilon \\
&\qquad - nH(B_T|A_TT) \\
&\overset{(m)}{\geq} nI(Y;ACD|X) - nH(B|A) - n\delta_\epsilon',
\end{aligned}$$

where

- (a) follows from that fact that X^n and (J, M_1) are independent of each other;
- (b) follows since A^n is a function of (X^n, J, M_1);
- (c) follows by introducing a telescoping sum of which only one term from the previous equation remains uncanceled;
- (d) follows by canceling the common term $I(Y^{t-1}; B_{t+1}^n A^n J|X^n)$ from both terms in the previous equation;
- (e) follows since (X^n, Y^n) are nearly i.i.d., and hence (22) applies;
- (f) follows by dropping (A^{t-1}, A_{t+1}^n) from the first term, and replacing Y^n in place of Y^{t-1} in the second;
- (g) follows by combining all the negative terms;
- (h) follows by upper bounding the second term by a conditional entropy term;
- (i) follows from the fact that B_t is conditionally independent of every other B and A variables given A_t;
- (k) follows from the definition of the auxiliary random variables $C_t \triangleq (X^{t-1}B_{t+1}^nJ)$ and $D_t \triangleq Y^{t-1}$;
- (l) follows from the definition of the auxiliary random variable $C \triangleq (C_T, T)$, and the fact that T is independent of B_T;
- (m) follows by defining $\delta_\epsilon' \triangleq I(Y_T;T|X_T) + \delta_\epsilon$ where by [35, Sec.V.A] $\delta_\epsilon' \to 0$ with $\epsilon \to 0$.

Step 2, we now move to the constraint of (3b) for R_o.

$$\begin{aligned}
nR_o &= H(J) \geq I(X^nY^n;J) \\
&\overset{(a)}{=} \sum_{t=1}^{n} \left[I(X^tY^t;B_{t+1}^nJ) - I(X^{t-1}Y^{t-1};B_t^nJ)\right] \\
&\overset{(b)}{=} \sum_{t=1}^{n} \left[I(X_tY_t;B_{t+1}^nJ|X^{t-1}Y^{t-1}) \right.\\
&\qquad \left. -I(X^{t-1}Y^{t-1};B_t|B_{t+1}^nJ)\right] \\
&\overset{(c)}{\geq} \sum_{t=1}^{n} \left[I(X_tY_t;B_{t+1}^nJX^{t-1}Y^{t-1}) \right.\\
&\qquad \left. -I(X^{t-1}Y^{t-1};B_t|B_{t+1}^nJ)\right] - n\delta_\epsilon \\
&\overset{(d)}{\geq} \sum_{t=1}^{n} \left[I(X_tY_t;(JB_{t+1}^nX^{t-1})Y^{t-1}) \right.\\
&\qquad \left. -I(B_t;X^{t-1}Y^{t-1}B_{t+1}^nJ)\right] - n\delta_\epsilon
\end{aligned}$$

$$\overset{(e)}{=} \sum_{t=1}^{n} \left[I(X_tY_t; (JB_{t+1}^{n}X^{t-1})Y^{t-1}) - I(B_t; (JB_{t+1}^{n}X^{t-1})Y^{t-1}) \right] - n\delta_\epsilon$$
$$\overset{(f)}{=} \sum_{t=1}^{n} [I(X_tY_t; C_tD_t) - I(B_t; C_tD_t)] - n\delta_\epsilon$$
$$= nI(X_TY_T; C_TD_T|T) - nI(B_T; C_TD_T|T) - n\delta_\epsilon$$
$$\geq nI(X_TY_T; C_TD_TT) - I(B_T; C_TD_TT) - n\delta_\epsilon^{'}$$
$$= nI(XY; CD) - nI(B; CD) - n\delta_\epsilon^{'},$$

where

(a) follows by introducing a telescoping sum of which only one term from the previous equation remains uncanceled;
(b) follows by canceling the common term $I(X^{t-1}Y^{t-1}; B_{t+1}^{n}J)$ from both terms in the previous equation;
(c) follows since (X^n, Y^n) are nearly i.i.d., and hence (22) applies;
(d) follows by magnifying the second term suitably;
(e) follows by combining appropriate terms;
(f) follows from the definition of the auxiliary random variables C_T and D_T.

Step 3, finally we consider the constraint of (3c)

$$I(C; X) \overset{(a)}{=} I(C_TT; X_T)$$
$$\overset{(b)}{=} I(C_T; X_T|T)$$
$$= \frac{1}{n}\sum_{t=1}^{n} I(C_t; X_t)$$
$$\overset{(c)}{=} \frac{1}{n}\sum_{t=1}^{n} I(X^{t-1}B_{t+1}^{n}J; X_t)$$
$$\overset{(d)}{=} \frac{1}{n}\sum_{t=1}^{n} I(B_{t+1}^{n}; X_t|X^{t-1}J)$$
$$\overset{(e)}{=} \frac{1}{n}\sum_{t=1}^{n} I(X^{t-1}; B_t|B_{t+1}^{n}J)$$
$$\overset{(f)}{\leq} \frac{1}{n}\sum_{t=1}^{n} I(X^{t-1}B_{t+1}^{n}J; B_t)$$
$$= \frac{1}{n}\sum_{t=1}^{n} I(C_t; B_t)$$
$$= I(C_T; B_T|T)$$
$$\leq I(C_TT; B_T)$$
$$= I(C; B),$$

where

(a) follows from the definition of the auxiliary random variable C,
(b) holds due to the fact that X_T is independent of T;
(c) follows from the definition of the auxiliary random variable C_T;
(d) follows from the fact that X^n is i.i.d and independent of the common randomness J;
(e) follows from Csiszár sum identity [33, Sec 2.3];
(f) follows from chain rule of mutual information.

Finally, we prove the cardinality bound for the outer bound of Theorem 2. To bound the cardinality of the auxiliary random variables, we only need to worry about the cardinalities C and D, since those of A and B are specified by the channel.

Cardinality bounds: To bound the cardinality of C, note that the RHS of the equations of the outer bound in (3) are preserved if we preserve:

i) the distribution $P_{XYAB}(x, y, a, b)$ for each tuple (x, y, a, b);
ii) the information functionals $H(YD|ACX)$ and $H(D|ACYX)$ (together, they preserve $H(Y|ACDX)$),
iii) the information functionals $H(XYD|C)$ and $H(D|C)$ (together, they preserve $H(XY|CD)$),
iv) and the information functionals $H(BD|C)$ and $H(D|C)$ (together, they preserve $H(B|CD)$).

By viewing each of these quantities as a function of P_C and $P_{XYABD|C}$, and by invoking the Support Lemma [33], we can see that the size of alphabet of C can be restricted to $|\mathcal{X}||\mathcal{Y}||\mathcal{A}||\mathcal{B}| - 1 + 6 = |\mathcal{X}||\mathcal{Y}||\mathcal{A}||\mathcal{B}| + 5$. Once this is done, we can preserve the distribution of $P_{XYABC}(x, y, a, b, c)$ for each tuple (x, y, a, b, c) by another application of the Support Lemma [33], which will allow us to restrict the size of D to at most $|\mathcal{X}||\mathcal{Y}||\mathcal{A}||\mathcal{B}||\mathcal{C}| = (|\mathcal{X}||\mathcal{Y}||\mathcal{A}||\mathcal{B}| + 5)|\mathcal{X}||\mathcal{Y}||\mathcal{A}||\mathcal{B}|$. ■

V. Separate Coordination-Channel Encoding Scheme With Randomness Extraction

In the previous section, we have presented a joint coordination-channel encoding scheme that utilizes the randomness provided by the DMC. This randomness is required to reduce the amount of local randomness needed to generate the sequence of actions at Node Y. As a basis for comparison, we will now introduce a separate encoding scheme that involves randomness extraction from the channel to supplement the local randomness required at Node Y.

In this scheme we consider a two-stage method for achieving strong coordination over noisy channels. As depicted in Fig. 4, an outer code represented by a strong coordination code is designed to coordinate the sequence of actions between the two nodes, i.e., achieve strong coordination as defined in Section III. The coordination encoder generates, based on the common randomness index J, a coordination message I at rate R_c to encode the action sequence X^n for Node Y. This message is then reliably communicated over the DMC at rate R_a using an inner capacity achieving channel code.

The following theorem describes an inner bound to the strong coordination region using the separate encoding scheme with randomness extraction.

Theorem 3: There exists an achievable separate coordination-channel encoding scheme for the noisy strong coordination setup in Fig 4 such that (1) is satisfied if

$$R_o + \rho_1 \geq I(Y; U|X), \tag{23a}$$
$$R_o \geq I(XY; U) - I(A; B), \tag{23b}$$
$$I(A; B) \geq I(X; U), \tag{23c}$$
$$\rho_2 \geq \max\big(0, H(Y|U) - H(B|A)\big). \tag{23d}$$

where U is an auxiliary random variable jointly correlated with the actions (X, Y), i.e., $X - U - Y$.

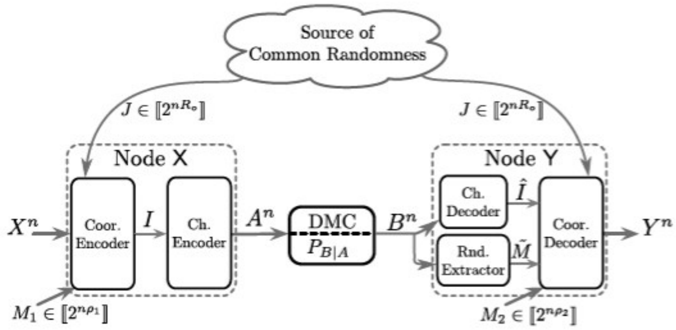

Fig. 4. Separate encoding for point-to-point strong coordination over a DMC.

A. Inner Bound: Achievability

We first use a $(2^{nR_c}, 2^{nR_o}, n)$ strong coordination code for noise-free channels [3] to a generate coordination message I of rate R_c. A noiseless strong coordination code consists from a codebook $\mathcal{U}$ of size $2^{n(R_o+R_c)}$ generated i.i.d. according to pmf P_U, i.e., $U_{ij}^n \sim \prod_{l=1}^n P_U(\cdot)$ for all $(i,j) \in \mathcal{I} \times \mathcal{J}$ where $\mathcal{I} \triangleq [\![2^{nR_c}]\!]$, $\mathcal{J} \triangleq [\![2^{nR_o}]\!]$, an encoding function $I : \mathcal{X}^n \times [\![2^{nR_o}]\!] \to [\![2^{nR_c}]\!]$ and a decoding function $Y^n : [\![2^{nR_c}]\!] \times [\![2^{nR_o}]\!] \to \mathcal{Y}^n$. Such a code exists and satisfies (1) (see [3], [5], [35], [36]) if and only if the rates R_o, R_c satisfy

$$R_c + R_o \geq I(XY;U), \tag{24a}$$
$$R_c \geq I(X;U), \tag{24b}$$

and U is selected such that $X - U - Y$ forms a Markov chain. This coordination message I is then communicated over the noisy channel using a rate-R_a channel code over m channel uses with codebook $\mathcal{A}$. Hence, $R_c = \lambda R_a$, where $\lambda = m/n$. The probability of decoding error, i.e, $\mathbb{P}_{e_I}^{(m)} \triangleq \mathbb{P}[I \neq \hat{I}]$, can be made vanishingly small if $R_a < I(A;B)$. Then, from the channel decoder output $\hat{I}$ and the common randomness message J we reconstruct the coordination sequence U^n and pass it though a test channel $P(Y|U)$ to generate the action sequence at Node Y.

At this point, we have obtained the necessary conditions on the coordination rate R_c, the common randomness rate R_o and the communication rate R_a to achieve strong coordination over the noisy DMC. Now, similar to the joint scheme, we can quantify the local randomness at both nodes [5], [7]. At node X, this yields $\rho_1 \geq R_c - I(X;U)$. At Node Y, the necessary local randomness for the generation of the action sequence is bounded by the channel simulation rate of the DMC $P_{Y|U}$ [23], i.e., $\rho_2 \geq H(Y|U)$.

Remark 5: Note that this separate encoding scheme without the randomness extraction phase can be constructed as a special case of the joint coordination-channel scheme in Fig. 3 by simply choosing $C = U$, $P_{AC} = P_A P_U$, and A being independent of X, Y, and C.

In the following we: (a) consider the randomness extraction phase at the decoder; (b) derive the necessary condition on the communication rate R_a; and (c) quantify the rate of the extracted randomness. The derived condition on the communication rate R_a guarantees that the randomness extracted from the channel output provides (nearly) uniformly distributed bits that are independent of the message communicated over the channel.

1) Intrinsic Randomness Extraction: In the following we consider channel transmission with a super-block consisting of κ blocks of length m, where $\kappa \in \mathbb{N}$ is a large but fixed number. At the transmitter, the coordination super-message I^κ is split as $I^\kappa = (I_1, \ldots, I_\kappa)$ where each I_i has mR_a bits. Then in each sub-block $i = 1, \ldots, \kappa$, the transmitter sends the coordination message I_i using an optimal channel code of block-length m. From the channel coding theorem we obtain that $\mathbb{P}_{e_I}^{(m)} \leq 2^{-m\epsilon'}$ for some $\epsilon' > 0$ if $R_a < I(A;B)$. Consequently, it follows that, $\mathbb{P}[I^\kappa \neq \hat{I}^\kappa] \leq \kappa 2^{-m\epsilon'}$ for fixed κ. We can then utilize a randomness extractor on the channel super-block output $B^{\kappa m}$ to supplement the local randomness available at Node Y. The following lemma establishes a much needed fact about the randomness extractions stage that the channel output can be used to not only decode, but also to extract randomness (at a particular rate) that is guaranteed to be nearly independent of the message conveyed over the channel.

Lemma 4: Consider the separation based scheme over a DMC $P_{B|A}$ where an optimal channel code of length m symbols is used κ times. If $R_a < I(A;B)$, then we can extract a nearly uniformly distributed randomness $\tilde{M}$ with alphabet $[\![2^{\kappa m\tilde{R}}]\!]$ of rate $\tilde{R} \leq H(B|A)$ use such that:

i) $I(I, \hat{I}; \tilde{M}) \to 0$
ii) $\|\frac{1}{|\tilde{\mathcal{M}}|} - P_{\tilde{M}}\|_{TV} < 2^{-\kappa\beta}$ for some $\beta > 0$.

Proof: Observe that the messages and the sequence of channel outputs $B^{m\kappa} \in \mathcal{B}^{m\kappa}$

$$(I^\kappa, B^{m\kappa}) = \Big\{(I_j, B_{(j-1)m+1:jm})\Big\}_{j=1}^{\kappa}$$

are κ i.i.d. copies according to $P_{I_1 B^m}$. We consider $(I^\kappa, B^{m\kappa})$ as a discrete memoryless multiple source (DMMS). We can then utilize the intrinsic randomness extraction results of [37] that guarantees the following.

For any $\delta > 0$, there exist $\alpha > 0$ and $\beta > 0$ such that for sufficiently large κ any $\mathcal{S} \subset \mathcal{B}^{m\kappa}$ with $P(\mathcal{S}) > 2^{-\kappa\alpha}$ has a bin (color) mapping $\phi_\kappa : \mathcal{S} \to \tilde{\mathcal{M}}$, $\tilde{\mathcal{M}} \triangleq [\![2^{\kappa m\tilde{R}}]\!]$ such that for $\phi_\kappa(B^{m\kappa}) = \tilde{M}$:

$$\left\|P_{\tilde{M}} - \frac{1}{|\tilde{\mathcal{M}}|}\right\|_{TV} \overset{(a)}{<} 2^{-\kappa\beta},$$
$$\|P_{I^\kappa \tilde{M}} - P_{I^\kappa} P_{\tilde{M}}\|_{TV} \overset{(b)}{<} 2^{-\kappa\beta},$$

provided:

$$\begin{aligned} m\tilde{R} &\overset{(c)}{\leq} H(B^m|I_1) - \delta \\ &= H(B^m, I_1) - H(I_1) - \delta \\ &= mH(B) - mR_a + H(I_1|B^m) - \delta \\ &\overset{(d)}{\leq} mH(B) - mR_a + H(\mathbb{P}_{e_I}^{(m)}) + \mathbb{P}_{e_I}^{(m)} \log|\mathcal{I}| - \delta, \end{aligned}$$

where (a)-(c) follow directly from [37, Proposition 1] and (d) follows from Fano's inequality. Consequently,

$$\begin{aligned} \tilde{R} &\leq H(B) - R_a + \frac{H(\mathbb{P}_{e_I}^{(m)}) + \mathbb{P}_{e_I}^{(m)} \log|\mathcal{I}| - \delta}{m} \\ &\overset{m\to\infty}{\longrightarrow} H(B) - R_a, \end{aligned}$$

Therefore, when R_a is close to $I(A;B)$, one can extract randomness of rate $\tilde{R}$ close to $H(B|A)$ bits/channel use. Moreover, if $|\tilde{\mathcal{M}}| \geq 4$ [37, Lemma 1] then

$$I(I^\kappa;\tilde{M}) \leq \|P_{I^\kappa \tilde{M}} - P_{I^\kappa} P_{\tilde{M}}\|_{TV} \log \frac{|\tilde{\mathcal{M}}|}{\|P_{I^\kappa \tilde{M}} - P_{I^\kappa} P_{\tilde{M}}\|_{TV}}$$

$$I(I^\kappa;\tilde{M}) \leq 2^{-\kappa\beta} \left(\log|\tilde{\mathcal{M}}| + o(1)\right).$$

Finally, we have

$$\begin{aligned} I(I,\hat{I};\tilde{M}) &= I(I;\tilde{M}) + I(\hat{I};\tilde{M}|I) \\ &= \tfrac{1}{\kappa} I(I^\kappa;\tilde{M}) + H(\hat{I}|I) - H(\hat{I}|I,\tilde{M}) \\ &\leq \tfrac{1}{\kappa} 2^{-\kappa\beta}\left(\log|\tilde{\mathcal{M}}| + o(1)\right) + H(\hat{I}|I) \\ &\leq \tfrac{1}{\kappa} 2^{-\kappa\beta}\left(\log|\tilde{\mathcal{M}}| + o(1)\right) + H(\mathbb{P}_{e_I}^{(m)}) + \mathbb{P}_{e_I}^{(m)} \log|\mathcal{I}| \\ &\overset{\kappa,m\to\infty}{\longrightarrow} 0. \end{aligned}$$

■

Remark 6: Randomness extraction from an arbitrary source was first studied in [38], and then in [37] and [39]. Note that the results of channel randomness extraction with asymptotic independence in [39] can also be used to prove Lemma 4.

Now, we set $\lambda = 1$ to facilitate a comparison with the joint scheme from Section IV and supplement the local randomness at Node Y with the randomness extracted from the channel noise. Thus, $\rho_2 \geq \max\left(0, H(Y|U) - H(B|A)\right)$. Combining the obtained local randomness rate constraints with the constraints of (24) for the outer strong coordination scheme and the inner channel code, we obtain the inner bound to the strong coordination region

$$R_o + R_c \geq I(XY;U), \tag{25a}$$
$$R_c \geq I(X;U), \tag{25b}$$
$$R_c \leq I(A;B), \tag{25c}$$
$$\rho_1 \geq R_c - I(X;U), \tag{25d}$$
$$\rho_2 \geq \max\left(0, H(Y|U) - H(B|A)\right). \tag{25e}$$

Followed by performing Fourier-Motzkin elimination [34], we obtain the inner bound to the strong coordination region described in Theorem 3. The proof follows in a straightforward way from the proof of Theorem 1 with the selection of auxiliary random variables C, A and the decomposition of P_{AC} as stated in Remark 5 combined with Lemma 4, and thus is omitted.

VI. An Example

In the following we compare the performance of the joint scheme in Section IV and the separation-based scheme in Section V using a simple example. Specifically, we let X to be a Bernoulli-$\frac{1}{2}$ source, the communication channel $P_{B|A}$ to be a BSC with crossover probability p_o (BSC(p_o)), and the conditional distribution $P_{Y|X}$ to be a BSC(p).

A. Basic Separation Scheme With Randomness Extraction

To derive the rate constraints for the basic separation scheme, we consider[5] $X - U - Y$ with $U \sim$Bernoulli-$\frac{1}{2}$, $P_{U|X}$ =BSC(p_1), and $P_{Y|U}$ =BSC(p_2), $p_2 \in [0,p]$, $p_1 = \dfrac{p-p_2}{1-2p_2}$. Using this to obtain the mutual information terms required for Theorem 3, we get

$$I(X;U) = 1 - h_2(p_1),\ I(A;B) = 1 - h_2(p_o), \tag{26a}$$
$$I(XY;U) = 1 + h_2(p) - h_2(p_1) - h_2(p_2), \tag{26b}$$
$$\text{and } H(Y|U) = h_2(p_2). \tag{26c}$$

After a round of Fourier-Motzkin elimination by using (26a)-(26c) in Theorem 3, we obtain the following constraints for the achievable region using the separation-based scheme with randomness extraction:

$$R_o + \rho_1 + \rho_2 \geq h_2(p) - \min\left(h_2(p_2), h_2(p_o)\right), \tag{27a}$$
$$h_2(p_1) \geq h_2(p_o). \tag{27b}$$

Note that (27a) presents the achievable sum rate constraint for the required randomness in the system.

B. Joint Scheme

The rate constraints for the joint scheme are constructed in two stages. First, we derive the scheme for the codebook cardinalities $|\mathcal{A}| = 2$ and $|\mathcal{C}| = 2$, an extension to larger $|\mathcal{C}|$ is straightforward but more tedious (see Figs. 5 and 6).[6] The joint scheme correlates the codebooks while ensuring that the decodability constraint (19e) is satisfied. To find the best tradeoff between these two features, we find the joint distribution P_{AC} that maximizes $I(B;C)$. For $|\mathcal{C}| = 2$ this is simply given by $P_{A|C} = \delta_{ac}$, where δ_{ac} denotes the Kronecker delta. Then, the distribution $P_X P_{CA|X} P_{B|A} P_{Y|BC}$ that produces the boundary of the strong coordination region for the joint scheme is formed by cascading two BSCs and another symmetric channel, yielding the Markov chain $X - (C,A) - (C,B) - Y$, with the channel transition matrices

$$P_{CA|X} = \begin{bmatrix} 1-p_1 & 0 & 0 & p_1 \\ p_1 & 0 & 0 & 1-p_1 \end{bmatrix}, \tag{28}$$

$$P_{CB|CA} = \begin{bmatrix} 1-p_o & p_o & 0 & 0 \\ 0 & 0 & p_o & 1-p_o \end{bmatrix}, \tag{29}$$

$$P_{Y|CB} = \begin{bmatrix} 1-\alpha & 1-\beta & \beta & \alpha \\ \alpha & \beta & 1-\beta & 1-\alpha \end{bmatrix}^T \tag{30}$$

for some $\alpha, \beta \in [0,1]$.

[5]For a doubly symmetric binary source (DSBS) given by $P_{XY} = P_X P_{Y|X}$ where $P_{Y|X}$ is a BSC(p), the BSC(p) is realized through a cascade of two BSCs with an intermediate random variable U. The resulting joint distribution of (X,Y,U) produce the boundary of the rate region, satisfies the Markov chain $X - U - Y$, and the marginal distribution for (X,Y) is P_{XY} [40]. Finally, since the BSC input X is given by $X \sim$Bernoulli-$\frac{1}{2}$ it follows that $U \sim$Bernoulli-$\frac{1}{2}$.

[6]Note that these cardinalities are not necessarily optimal. They are, however, analytically feasible, provide a good intuition about the performance of the scheme, and satisfy the bound $|\mathcal{C}| \leq |\mathcal{A}||\mathcal{B}||\mathcal{X}||\mathcal{Y}| + 5 = 21$ as given in Theorem 2.

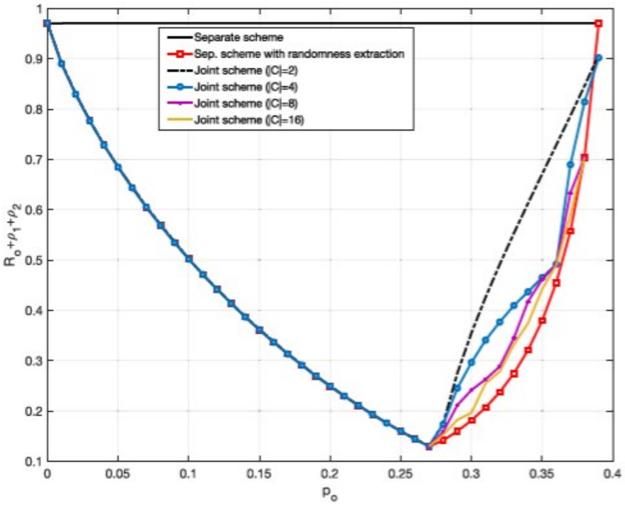

Fig. 5. Randomness sum rate vs. BSC crossover probability p_0 for target distribution $Q_{Y|X} = \text{BSC}(0.4)$.

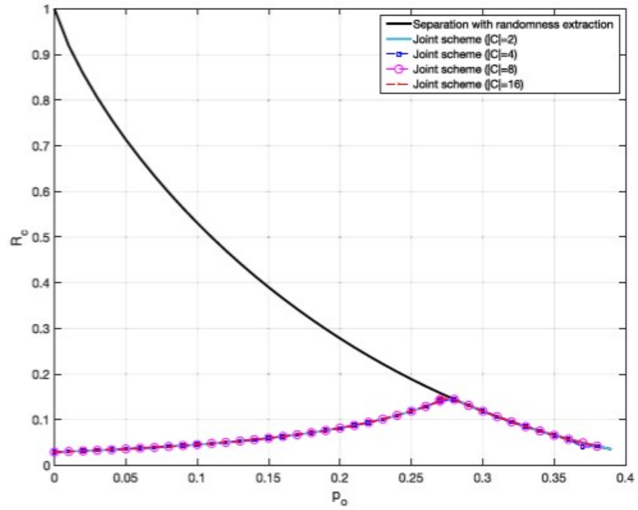

Fig. 6. Communication rate vs. BSC crossover probability p_0 for target distribution $Q_{Y|X} = \text{BSC}(0.4)$.

Then, the mutual information terms required for Theorem 1 can be expressed with $p_2 \triangleq (1-p_o)\alpha + p_o\beta$ as

$$\begin{aligned} I(X;AC) &= I(X;C) = 1 - h_2(p_1), \\ I(XY;AC) &= I(XY;C) = 1 + h_2(p) - h_2(p_1) - h_2(p_2), \\ I(B;C) &= 1 - h_2(p_o), \text{ and} \\ H(Y|BC) &= p_o h_2(\beta) + (1-p_o) h_2(\alpha). \end{aligned}$$

To find the minimum achievable sum rate, we minimize the rate constraints in Theorem 1 with respect to the parameters p_2, α, and β as follows:

$$R_o+\rho_1+\rho_2 = \min_{p_2,\alpha,\beta}\big(h_2(p) - h_2(p_2) + (1-p_o)h_2(\alpha) + p_o h_2(\beta)\big)$$
$$\text{subject to} \quad \begin{array}{rcl} h_2(p_1) &>& h_2(p_o), \\ p &=& p_1 - 2p_1p_2 + p_2. \end{array} \tag{31}$$

C. Numerical Results

Fig. 5 presents a comparison between the minimum randomness sum rate $R_o + \rho_1 + \rho_2$ required to achieve coordination using both the joint and the separate scheme with randomness extraction. The communication channel is given by BSC(p_o), and the target distribution is set as $Q_{Y|X} = \text{BSC}(0.4)$. The rates for the joint scheme are obtained by solving the optimization problem in (31). Similar results are obtained for the joint scheme with $|\mathcal{C}| > 2$. For the separate encoding scheme we choose p_2 such that $h_2(p_1) = h_2(p_0)$ to maximize the amount of extracted channel randomness. We also include the performance of the separate coding scheme without randomness extraction.

As can be seen from Fig. 5, both the joint scheme and the separate scheme with randomness extraction provide the same sum rate $R_o + \rho_1 + \rho_2$ for $p_o \leq p'_o$ where $p'_o \triangleq \frac{1-\sqrt{1-2p}}{2}$. We also observe that for noisier channels with $p_o > p'_o$ the joint scheme approaches the performance of the separate encoding scheme when the cardinality of C is increased. The increase of $R_o + \rho_1 + \rho_2$ for $p_o > p'_o$ is due to the fact that in this regime the channel provides more than sufficient randomness for simulating the action sequence Y^n via the test channel $P_{Y|BC}$ (see Fig. 3). As a result, the parameters α and β associated with $P_{Y|BC}$ must be adjusted to ensure that (1) is still satisfied. As p_o increases further, the required total randomness of the joint scheme approaches the one for the basic separate scheme again.

Note that when $p_o = p'_o$, the distributions P_{XYU} and P_{XYAC} of the separate encoding and joint coding scheme, respectively, become the distributions which achieve Wyner's common information [40]. Here, Wyner's common information is defined as $C(X;Y) \triangleq \min_{U:X-U-Y} I(XY;U)$, with $U = (A, C)$ for the case of the joint scheme. In other words, the two BSCs $P_{U|X}$ =BSC(p_1) and $P_{Y|U}$ =BSC(p_2) in Section VI. A are both identical to the communication channel $P_{B|A}$ and their cascade is equal to the target $Q_{Y|X}$. Let us first consider the separate encoding scheme. Here, we fix $p_1 = p_o$ to maximize the randomness extracted from the channel. Therefore, when $p_1 = p'_o$, i.e., the DMC cross-over probability is equal to $p_o = p'_o$, by solving $p_1 = \frac{p-p_2}{1-2p_2}$ for p_2 results in $p_2 = \frac{1-\sqrt{1-2p}}{2} = p'_o$. At this point, Wyner's common information is achieved, i.e., $I(XY;U)$ is minimized. As a result, we obtain the minimum coordination rate when no common randomness is available [3], [5], i.e., $R_o = 0$. Combining that with the fact that we extract $H(Y|U)$ from the channel, which leads to a local randomness rate of $\rho_2 = H(Y|U) - H(B|A) = 0$, results in the minimum randomness sum rate as observed in Fig. 5. On the other hand, let us consider the joint scheme for the special case of $|\mathcal{C}| = 2$. In this case, in order to exploit the channel randomness we fix $p_2 = p_o$ to maximize the amount of implicitly extracted randomness. This is done by selecting $\alpha = 0$ and $\beta = 1$ in (30) up until $p_o \leq p'_o$. As a result, the test channel $P_{Y|BC}$ becomes a deterministic channel, i.e., $\rho_2 = 0$. At $p_o = p'_o$ the joint distribution P_{XYAC} becomes the distribution which achieves Wyner's common information with cascade BSCs: $P_{CA|X}$ as a BSC(p_1) with $p_1 = p'_o$, $P_{Y|CA}$ as a BSC(p'_o) and we obtain again the minimum randomness sum rate.

Fig. 6 provides a comparison of the communication rate i.e., the number of channel uses per n time slots required to

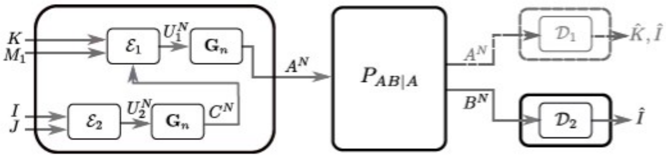

Fig. 7. Block diagram of the superposition polar code.

achieve strong coordination, for both schemes. Note that the joint scheme provides significantly smaller rates than the separation scheme with randomness extraction for $p_o \leq p_o'$, independently of the cardinality of $|\mathcal{C}|$. Thus, in this regime joint coordination-channel coding provides an advantage in terms of communication cost and outperforms a separation-based scheme for the same amount of randomness injected into the system.

VII. Nested Polar Code for Strong Coordination Over Noisy Channels

Since the proposed joint coordination-channel coding scheme, displayed in Fig. 3, is based on a channel resolvability framework, we adopt a channel resolvability-based polar construction for noise-free strong coordination [19] in combination with polar coding for the degraded broadcast channel [26]. In this section, we propose a strong coordination scheme based on polar coding that achieves the inner bound stated in Theorem 1. For some $N \triangleq 2^n, n \in \mathbb{N}$, let the joint pmf of actions induced by the polar coding scheme be $\tilde{P}_{X^N Y^N}$. For strong coordination coding scheme, $\tilde{P}_{X^N Y^N}$ must be *close* in total variation to the N i.i.d. copies of desired joint pmf $(X,Y) \sim Q_{XY}$, Q_{XY}^N, i.e.,

$$\|\tilde{P}_{X^N Y^N} - Q_{XY}^N\|_{TV} < \epsilon. \tag{32}$$

Theorem 4: For a binary input DMC and a target distribution Q_{XY} defined over $\mathcal{X} \times \mathcal{Y}$, with an auxiliary random variable C defined over the binary alphabet, there exist a polar coding scheme that satisfies (32) and achieves the region stated in Theorem 1.

We now construct the polar coding scheme of Theorem 4.

A. Coding Scheme

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

Consider a 2-user *physically* degraded discrete memoryless broadcast channel (DM-BC) $P_{AB|A}$ in Fig. 7 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 $P_{A|A}$ (we denote this as $P_{A|A} \succ P_{B|A}$). We construct the nested polar coding scheme in a similar fashion as in [26] as this mimics the nesting of the codebooks $\mathcal{C}$ and $\mathcal{A}$ in Step i) of the random coding construction in Section IV-A. Here, the second (weaker) user is able to recover an estimate $\hat{I}$ for its intended message I, while the first (stronger) user is able to recover estimates $\hat{K}, \hat{I}$ for both messages K and I, respectively. 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., $P_{A|C} \succ P_{B|C}$) [26, Lemma 3]. Note that we let $\widehat{C}$ be the random variable resulting from recovering C^N at Node Y from $\hat{I}$ and the shared randomness message J. Let $\mathbf{V}$ be a matrix of the selected codewords A^N and C^N as

$$\mathbf{V} \triangleq \begin{bmatrix} A^N \\ C^N \end{bmatrix}. \tag{33}$$

Now, apply the polar linear transformation $\mathbf{G}_n$, where $\mathbf{G}_n$ is defined in Section II, as

$$\mathbf{U} \triangleq \begin{bmatrix} U_1^N \\ U_2^N \end{bmatrix} = \mathbf{V}\mathbf{G}_n, \tag{34}$$

where the joint distribution of the random variables in $\mathbf{U}$ is given by $Q_{U_1\ U_2}^N(u_1^N, u_2^N) = Q_{AC}^N(u_1^N \mathbf{G}_n, u_2^N \mathbf{G}_n)$. First, consider $C^N \triangleq U_2^N \mathbf{G}_n$ from (33) and (34) where U_2^N is generated by the second encoder $\mathcal{E}_2$ in Fig. 7. For $\beta < \frac{1}{2}$ and $\delta_N \triangleq 2^{-N^\beta}$ we define the very high and high entropy sets

$$\mathcal{V}_C \triangleq \{i \in [\![N]\!] : H(U_{2,i}|U_2^{i-1}) > 1 - \delta_N\}, \tag{35a}$$

$$\mathcal{V}_{C|X} \triangleq \{i \in [\![N]\!] : H(U_{2,i}|U_2^{i-1} X^N) > 1 - \delta_N\} \subseteq \mathcal{V}_C, \tag{35b}$$

$$\mathcal{V}_{C|XY} \triangleq \{i \in [\![N]\!] : H(U_{2,i}|U_2^{i-1} X^N Y^N) > 1 - \delta_N\} \subseteq \mathcal{V}_{C|X}, \tag{35c}$$

$$\mathcal{H}_{C|B} \triangleq \{i \in [\![N]\!] : H(U_{2,i}|U_2^{i-1} B^N) > \delta_N\}, \tag{35d}$$

$$\mathcal{H}_{C|A} \triangleq \{i \in [\![N]\!] : H(U_{2,i}|U_2^{i-1} A^N) > \delta_N\}, \tag{35e}$$

which by [42, 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. 8. Note that the set $\mathcal{H}_{C|B}$ (exemplary denoted in red in Fig. 8) 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 $\mathcal{H}_{C|A} \subseteq \mathcal{H}_{C|B}$ and $\mathcal{H}_{C|B}^c \subseteq \mathcal{H}_{C|A}^c$, respectively. This ensures that the polarization indices are guaranteed to be aligned (i.e., $\mathcal{L}_2 \subseteq \mathcal{L}_1$) [43], [26, Lemma 4]. As a consequence,

[†]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 [41].

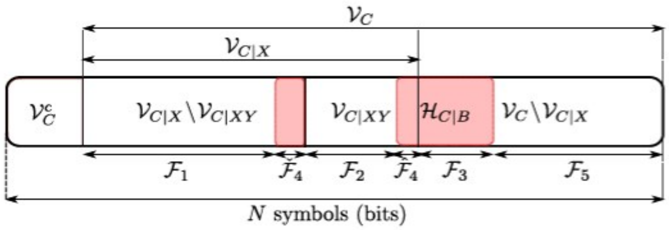

Fig. 8. Index sets for codeword C.

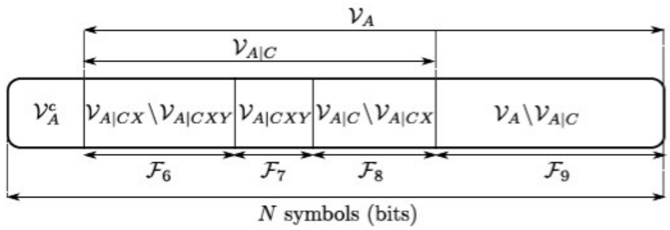

Fig. 9. Index sets for codeword A.

the bits decodable by the weaker user are also decodable by the stronger user.

Accordingly, in terms of the polarization sets in (35a)-(35d) we define the sets combining channel resolvability for strong coordination and broadcast channel construction as

$$\begin{aligned}
\mathcal{F}_1 &\triangleq (\mathcal{V}_{C|X} \setminus \mathcal{V}_{C|XY}) \cap \mathcal{H}_{C|B}^c,\\
\mathcal{F}_2 &\triangleq \mathcal{V}_{C|XY} \cap \mathcal{H}_{C|B}^c,\\
\mathcal{F}_3 &\triangleq \mathcal{V}_{C|X}^c \cap \mathcal{H}_{C|B} = \mathcal{H}_{C|B} \setminus \mathcal{H}_{C|BX},\\
\mathcal{F}_4 &\triangleq \mathcal{V}_{C|X} \cap \mathcal{H}_{C|B} = \mathcal{H}_{C|BX},\\
\hat{\mathcal{F}}_4 &\triangleq \mathcal{H}_{C|BXY},\\
\check{\mathcal{F}}_4 &\triangleq \mathcal{H}_{C|BX} \backslash \mathcal{H}_{C|BXY},\\
\mathcal{F}_5 &\triangleq (\mathcal{V}_C \setminus \mathcal{V}_{C|X}) \cap \mathcal{H}_{C|B}^c.
\end{aligned}$$

Now, consider $A^N \triangleq U_1^N \mathbf{G}_n$ (see (33) and (34)), where U_1^N is generated by the first encoder $\mathcal{E}_1$ with C^N as a side information as seen in Fig. 7. We define the very high entropy sets illustrated in Fig. 9 as

$$\mathcal{V}_A \triangleq \{i \in [\![N]\!] : H(U_{1,i}|U_1^{i-1}) > 1-\delta_N\}, \tag{36a}$$

$$\mathcal{V}_{A|C} \triangleq \{i \in [\![N]\!] : H(U_{1,i}|U_1^{i-1}C^N) > 1-\delta_N\} \subseteq \mathcal{V}_A, \tag{36b}$$

$$\mathcal{V}_{A|CX} \triangleq \{i \in [\![N]\!] : H(U_{1,i}|U_1^{i-1}C^N X^N) > 1-\delta_N\} \subseteq \mathcal{V}_{A|C}, \tag{36c}$$

$$\mathcal{V}_{A|CXY} \triangleq \{i \in [\![N]\!] : H(U_{1,i}|U_1^{i-1}C^N X^N Y^N) > 1-\delta_N\} \subseteq \mathcal{V}_{A|CX}, \tag{36d}$$

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. 8, 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 $\mathcal{H}_{A|A} = \emptyset$.

Similarly, in terms of the polarization sets in (36a)-(36d) we define the sets combining channel resolvability for strong coordination and broadcast channel construction as shown in Fig. 9

$$\begin{aligned}
\mathcal{F}_6 &\triangleq \mathcal{V}_{A|CX} \setminus \mathcal{V}_{A|CXY}, & \mathcal{F}_8 &\triangleq \mathcal{V}_{A|C} \setminus \mathcal{V}_{A|CX},\\
\mathcal{F}_7 &\triangleq \mathcal{V}_{A|CXY}, & \mathcal{F}_9 &\triangleq \mathcal{V}_A \setminus \mathcal{V}_{A|C}.
\end{aligned}$$

Finally, we define the sequence T^N as the polar linear transformation of Y^N i.e., $T^N \triangleq Y^N \mathbf{G}_n$. Now consider $Y^N = T^N \mathbf{G}_n$. By invertibility of $\mathbf{G}_n$ we define the very high entropy set:

$$\mathcal{V}_{Y|BC} \triangleq \{i \in [\![N]\!] : H(T_i|T^{i-1}B^N C^N) > \log|\mathcal{Y}| - \delta_N\}, \tag{37}$$

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. 3 as a source polarization operation [17], [19]. Note that here, we let $\mathbf{G}_n$ be a polar code generator matrix defined appropriately based on the alphabet of Y, e.g., if $|\mathcal{Y}|$ is the prime number $q \geq 2$, $\mathbf{G}_n$ is as defined in Section II. However, the matrix operation is now carried out in the Galois field $GF(q)$ and the entropy terms of the polarization sets are calculated with respect to base-q logarithms [17, Theorem 4]. 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 resulting in a storage complexity of $\mathcal{O}(kN)$ and a time complexity of $\mathcal{O}(kN \log N)$. In Algorithm 1 we use the *tilde* notation (i.e., $\widetilde{U}_1^N, \widetilde{U}_2^N$, $\widetilde{A}^N$, and $\widetilde{C}^N$) to denote the change in the statistics of the length-N random variables (i.e., U_1^N, U_2^N, A^N, and C^N) as a result of inserting uniformly distributed message and randomness bits at specific indices during encoding. Since for strong coordination the goal is to approximate a target joint distribution with the 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 [19]. The encoding scheme also leverages a block chaining construction [21], [42]–[44] to achieve the rates stated in Theorem 1.

More precisely, as demonstrated in Fig. 3, we are interested in successfully recovering the message I that is intended for the channel of the weak user $P_{B|A}$ in Fig. 7. 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. 10). At encoder $\mathcal{E}_2$, the set $\mathcal{F}_3$ of block $i \in [\![k]\!]$ is embedded in the reliably decodable bits of $\mathcal{F}_1 \cup \mathcal{F}_2$ of the following block $i+1$. This is possible by following the decodability constraint (see (19d), (19e) 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 $\mathcal{F}_2$ [21]. However, since these sets originally contain uniformly distributed common randomness J [19], the bits of $\mathcal{F}_3$ can be embedded while maintaining

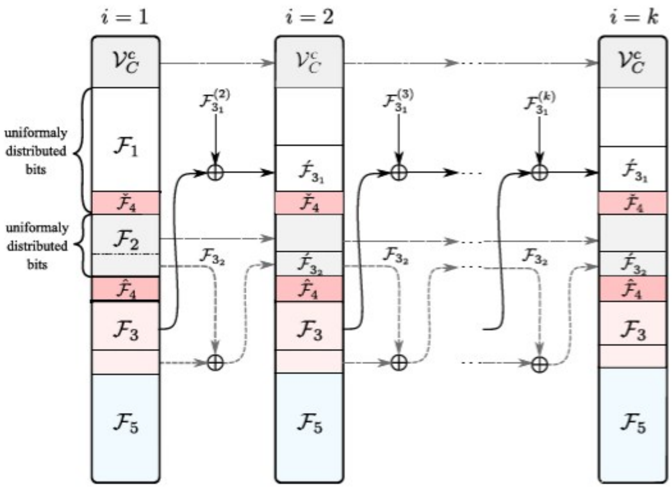

Fig. 10. Chaining construction for block encoding.

the uniformity of the randomness by taking advantage of the Crypto Lemma [45, Lemma 2], [46, Lemma 3.1]. 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. Note that since uniformly random bits are reused to convey information bits, chaining can be seen as a derandomization strategy.

2) Decoding: The decoder is described in Algorithm 2. In Algorithm 2, we use the *hat* notation, i.e., $\widehat{U}_2^N$ and $\widehat{C}^N$, to distinguish the reconstruction of the N-length random variables, i.e., U_2^N and consequently C_2^N, from the corresponding quantities at the encoder. Recall that we are only interested in the message $\hat{I}$ intended for the weak user channel given by $P_{B|A}$ in Fig. 7. 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 in Section VII-A. The analysis is based on the KL divergence which upper bounds the total variation in (32) by Pinsker's inequality. We start the analysis with a set of sequential lemmas. In particular, Lemma 5 is useful to show

Algorithm 1 Encoding Algorithm at Node X for Strong Coordination

Input: $X_{1:k}^N$, uniformly distributed local randomness bits $M_{1:k}$ of size $k|\mathcal{F}_6|$, common randomness bits reused over k blocks $\bar{J} = (\bar{J}_1, \bar{J}_2)$ of sizes $|\mathcal{F}_2 \cup \hat{\mathcal{F}}_4|$, and $|\mathcal{F}_7|$, respectively, and uniformly distributed common randomness bits for each block $J_{1:k}$, each 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. 7 constructs $\widetilde{U}_{2_i}^N$ bit-by-bit as follows:
   if $i = 1$ **then**
   - $\widetilde{U}_{2_i}^N[\mathcal{F}_1 \cup \check{\mathcal{F}}_4] \leftarrow J_i$
   - $\widetilde{U}_{2_i}^N[\mathcal{F}_2 \cup \hat{\mathcal{F}}_4] \leftarrow \bar{J}_1$

   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 $\tilde{P}_{U_{2_i,j}|U_{2_i}^{j-1}X_i^N}$ defined by

$$\tilde{P}_{U_{2_i,j}|U_{2_i}^{j-1}X_i^N} \triangleq \begin{cases} Q_{U_{2,j}|U_2^{j-1}} & j \in \mathcal{V}_C^c, \\ Q_{U_{2,j}|U_2^{j-1}X^N} & j \in \mathcal{F}_3 \cup \mathcal{F}_5. \end{cases} \tag{38}$$

3. $\widetilde{C}_i^N \leftarrow \widetilde{U}_{2_i}^N \mathbf{G}_n$
4. $\mathcal{E}_1$ in Fig. 7 constructs $\widetilde{U}_{1_i}^N$ bit-by-bit as follows:
   - $\widetilde{U}_{1_i}^N[\mathcal{F}_6] \leftarrow M_{1_i}$
   - $\widetilde{U}_{1_i}^N[\mathcal{F}_7] \leftarrow \bar{J}_2$
   - Given X_i^N and $\widetilde{C}_i^N$, successively draw the remaining components of $\widetilde{U}_{1_i}^N$ according to $\tilde{P}_{U_{1_i,j}|U_{1_i}^{j-1}C_i^N X_i^N}$ defined by

$$\tilde{P}_{U_{1_i,j}|U_{1_i}^{j-1}C_i^N X_i^N} \triangleq \begin{cases} Q_{U_{1,j}|U_1^{j-1}} & j \in \mathcal{V}_A^c, \\ Q_{U_{1,j}|U_1^{j-1}C^N} & j \in \mathcal{F}_9, \\ Q_{U_{1,j}|U_1^{j-1}C^N X^N} & j \in \mathcal{F}_8. \end{cases} \tag{39}$$

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

in Lemma 6 that the strong coordination scheme based on channel resolvability holds for each block individually regardless of the randomness recycling. Note that, in the current Section VII and the associated Appendices B and C, we refrain from using the N-fold product notation of joint, and conditional distribution, e.g., respectively, $Q_{X^N Y^N} = Q_{XY}^N$ and $Q_{U_2^N|X^N} = Q_{U_2|X}^N$ to unify the notation across conditional distributions.

Lemma 5: For block $i \in [\![k]\!]$, 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 $\bar{J}_1$ of sizes $|\mathcal{F}_2 \cup \hat{\mathcal{F}}_4|$ reused over k blocks, "fresh" uniformly distributed common randomness $J_{1:k}$ each of size $k|\check{\mathcal{F}}_4 \cup \mathcal{F}_1|$ for all k blocks and shared with Node X.

Output: $\widetilde{Y}_{1:k}^N$

1. **For** block $i = k, \ldots, 1$ **do**
2. $\mathcal{D}_2$ in Fig. 7 constructs $\widehat{U}_{2_i}^N$ bit-by-bit as follows:
   - $(\widehat{U}_{2_i}^N[(\mathcal{F}_1 \setminus \acute{\mathcal{F}}_{3_1}) \cup \check{\mathcal{F}}_4], \mathcal{F}_{3_1}^{(i)}) \leftarrow J_i$
   - $(\widehat{U}_{2_i}^N[(\mathcal{F}_2 \setminus \acute{\mathcal{F}}_{3_2}) \cup \hat{\mathcal{F}}_4], \mathcal{F}_{3_2}) \leftarrow \bar{J}_1$
   - 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

$$\tilde{P}_{U_{2_i,j}|U_{2_i}^{j-1}B_i^N} \triangleq \begin{cases} Q_{U_{2,j}|U_2^{j-1}} & j \in \mathcal{V}_C^c, \\ Q_{U_{2,j}|U_2^{j-1}B_i^N} & j \in \acute{\mathcal{F}}_{3_2} \cup \acute{\mathcal{F}}_{3_1} \cup \mathcal{F}_5. \end{cases} \tag{40}$$

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. $\widehat{C}_i^N \leftarrow \widehat{U}_{2_i}^N \mathbf{G}_n$
6. Channel simulation: given $\widehat{C}_i^N$ and $B_i{}^N$, successively draw the components of $\widetilde{T}_i^N$ according to

$$\tilde{P}_{T_{i,j}|T_i^{j-1}B_i^N C_i^N} \triangleq \begin{cases} 1/|\mathcal{Y}| & j \in \mathcal{V}_{Y|BC}, \\ Q_{T_j|T^{j-1}B^N C^N} & j \in \mathcal{V}_{Y|BC}^c. \end{cases} \tag{41}$$

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

Proof: We have

$$\begin{aligned}
&\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}) \\
&= \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{(b)}{=} \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{(c)}{=} \sum_{j=1}^N \mathbb{E}_{Q_{U_2^{j-1} X^N}}\left[\mathbb{D}(Q_{U_{2,j}|U_2^{j-1} X^N} || \tilde{P}_{U_{2_i,j}|U_{2_i}^{j-1} X_i^N})\right] \\
&+ \sum_{j=1}^N \mathbb{E}_{Q_{U_1^{j-1} U_2^N X^N}}\left[\mathbb{D}(Q_{U_{1,j}|U_1^{j-1} U_2^N X^N} || \tilde{P}_{U_{1_i,j}|U_{1_i}^{j-1} U_{2_i}^N X_i^N})\right] \\
&\overset{(d)}{=} \sum_{j \notin \mathcal{F}_3 \cup \mathcal{F}_5} \mathbb{E}_{Q_{U_2^{j-1} X^N}}\left[\mathbb{D}(Q_{U_{2,j}|U_2^{j-1} X^N} || \tilde{P}_{U_{2_i,j}|U_{2_i}^{j-1} X_i^N})\right] \\
&+ \sum_{j \notin \mathcal{F}_8} \mathbb{E}_{Q_{U_1^{j-1} U_2^N X^N}}\left[\mathbb{D}(Q_{U_{1,j}|U_1^{j-1} U_2^N X^N} || \tilde{P}_{U_{1_i,j}|U_{1_i}^{j-1} U_{2_i}^N X_i^N})\right] \\
&\overset{(e)}{=} \sum_{j \in \mathcal{V}_C^c \cup \mathcal{V}_{C|X}} \mathbb{E}_{Q_{U_2^{j-1} X^N}}\left[\mathbb{D}(Q_{U_{2,j}|U_2^{j-1} X^N} || \tilde{P}_{U_{2_i,j}|U_{2_i}^{j-1} X_i^N})\right] \\
&+ \sum_{j \in \mathcal{V}_A^c \cup \mathcal{V}_{A|CX} \cup \mathcal{V}_A \setminus \mathcal{V}_{A|C}} \mathbb{E}_{Q_{U_1^{j-1} U_2^N X^N}}\Big[\\
&\qquad \mathbb{D}(Q_{U_{1,j}|U_1^{j-1} U_2^N X^N} || \tilde{P}_{U_{1_i,j}|U_{1_i}^{j-1} U_{2_i}^N X_i^N})\Big] \\
&\overset{(f)}{=} \sum_{j \in \mathcal{V}_C^c} \Big(H(U_{2,j}|U_2^{j-1}) - H(U_{2,j}|U_2^{j-1} X^N)\Big) \\
&\quad + \sum_{j \in \mathcal{V}_{C|X}} \Big(1 - H(U_{2,j}|U_2^{j-1} X^N)\Big) \\
&\quad + \sum_{j \in \mathcal{V}_A^c} \Big(H(U_{1,j}|U_1^{j-1}) - H(U_{1,j}|U_1^{j-1} U_2^N X^N)\Big) \\
&\quad + \sum_{j \in \mathcal{V}_{A|CX}} \Big(1 - H(U_{1,j}|U_1^{j-1} U_2^N X^N)\Big) \\
&\quad + \sum_{j \in \mathcal{V}_{A|C}^c \setminus \mathcal{V}_A^c} \Big(H(U_{1,j}|U_1^{j-1} U_2^N) - H(U_{1,j}|U_1^{j-1} U_2^N X^N)\Big) \\
&\overset{(g)}{=} \sum_{j \in \mathcal{V}_C^c} \Big(H(U_{2,j}|U_2^{j-1}) - H(U_{2,j}|U_2^{j-1} X^N)\Big) \\
&\quad + \sum_{j \in \mathcal{V}_{C|X}} \Big(1 - H(U_{2,j}|U_2^{j-1} X^N)\Big) \\
&\quad + \sum_{j \in \mathcal{V}_A^c} \Big(H(U_{1,j}|U_1^{j-1}) - H(U_{1,j}|U_1^{j-1} C^N X^N)\Big) \\
&\quad + \sum_{j \in \mathcal{V}_{A|CX}} \Big(1 - H(U_{1,j}|U_1^{j-1} C^N X^N)\Big) \\
&\quad + \sum_{j \in \mathcal{V}_{A|C}^c \setminus \mathcal{V}_A^c} \Big(H(U_{1,j}|U_1^{j-1} C^N) - H(U_{1,j}|U_1^{j-1} C^N X^N)\Big) \\
&\overset{(h)}{\leq} (|\mathcal{V}_C^c| + |\mathcal{V}_{C|X}| + |\mathcal{V}_{A|XC}| + |\mathcal{V}_{A|C}^c|)\delta_N \leq 2N\delta_N,
\end{aligned}$$

where

- (a) holds by invertibility of $\mathbf{G}_n$;
- (b) - (c) follows from the chain rule of the KL divergence [31];
- (d) results from the definitions of the conditional distributions in (38), and (39);
- (e) follows from the definitions of the index sets as shown in Figs. 8 and 9;
- (f) 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 (38) and (39);
- (g) holds by the one-to-one relation between U_2^N and C^N;
- (h) follows from the sets defined in (35) and (36), for $\delta_N \triangleq 2^{-N^\beta}$ and $\beta < \frac{1}{2}$.

■

Lemma 6: For block $i \in [\![k]\!]$, we have

$$\begin{aligned}
&\mathbb{D}(\tilde{P}_{X_i^N Y_i^N} || Q_{X^N Y^N}) \\
&\quad \leq \mathbb{D}(\tilde{P}_{X_i^N A_i^N C_i^N B_i^N \widehat{C}_i^N Y_i^N} || Q_{X^N A^N C^N B^N \widehat{C}^N Y^N}) \leq \delta_N^{(1)}
\end{aligned}$$

where $\delta_N^{(1)} \triangleq \mathcal{O}(\sqrt{N^3 \delta_N})$, $\delta_N \triangleq 2^{-N^\beta}$ and $\beta < \frac{1}{2}$.

Proof: Consider the argument shown at the bottom of the 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 [19, Lemma 16] where

$$\delta_N^{(2)} \triangleq -N \log(\mu_{XACB\widehat{C}Y})\sqrt{2\ln 2}\sqrt{2N\delta_N},$$
$$\mu_{XACB\widehat{C}Y} \triangleq \min^*_{x,y,a,c,b,\hat{c}}(Q_{XACB\widehat{C}Y});$$

(d) follows from the chain rule of KL divergence [31];

(e) holds by Lemma 5 and [19, Lemma 14] where

$$\hat{\delta}_N^{(2)} \triangleq -N \log(\mu_{XAC})\sqrt{2\ln 2}\sqrt{2N\delta_N},$$
$$\mu_{XAC} \triangleq \min^*_{x,a,c}(Q_{XAC});$$

(f) follows from the chain rule of KL divergence [31];

(g) holds by [19, 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 $\mathbb{D}(Q_{B^N\widehat{C}^N|A^NC^N}||\tilde{P}_{B_i^N\widehat{C}_i^N|A_i^NC_i^N})$, 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:

- First, we show that $\mathbb{D}(Q_{B^N\widehat{C}^N|A^NC^N}||\tilde{P}_{B_i^N\widehat{C}_i^N|A_i^NC_i^N}) \leq N\delta_N$ by the following argument:

$$\begin{aligned}
&\mathbb{D}(Q_{B^N\widehat{C}^N|A^NC^N}||\tilde{P}_{B_i^N\widehat{C}_i^N|A_i^NC_i^N})\\
&\overset{(a)}{=} \mathbb{D}(Q_{B^N|A^N}Q_{\widehat{C}^N|B^N}||Q_{B^N|A^N}\tilde{P}_{\widehat{C}_i^N|B_i^N})\\
&= \mathbb{D}(Q_{\widehat{C}^N|B^N}||\tilde{P}_{\widehat{C}_i^N|B_i^N})\\
&\overset{(b)}{=} \mathbb{D}(Q_{\widehat{U}^N|B^N}||\tilde{P}_{\widehat{U}_i^N|B_i^N})\\
&\overset{(c)}{=} \sum_{j=1}^N \mathbb{E}_{Q_{U_2^{j-1}B^N}}\left[\mathbb{D}(Q_{U_{2,j}|U_2^{j-1}B^N}||\tilde{P}_{U_{2_i,j}|U_{2_i}^{j-1}B_i^N})\right]\\
&\overset{(d)}{=} \sum_{j\in\mathcal{V}_C^c} \mathbb{E}_{Q_{U_2^{j-1}B^N}}\left[\mathbb{D}(Q_{U_{2,j}|U_2^{j-1}B^N}||\tilde{P}_{U_{2_i,j}|U_{2_i}^{j-1}B_i^N})\right]\\
&\quad + \sum_{j\in\mathcal{H}_{C|B}\cup\mathcal{V}_{C|X}} \mathbb{E}_{Q_{U_2^{j-1}B^N}}\left[\mathbb{D}(Q_{U_{2,j}|U_2^{j-1}B^N}||\tilde{P}_{U_{2_i,j}|U_{2_i}^{j-1}B_i^N})\right]\\
&\overset{(e)}{=} \sum_{j\in\mathcal{V}_C^c}\Big(H(U_{2,j}|U_2^{j-1}) - H(U_{2,j}|U_2^{j-1}B^N)\Big)\\
&\quad + \sum_{j\in\mathcal{H}_{C|B}\cup\mathcal{V}_{C|X}}\Big(1 - H(U_{2,j}|U_2^{j-1}B^N)\Big)\\
&\overset{(f)}{\leq} |\mathcal{V}_C^c|\delta_N + |\mathcal{H}_{C|B}\cup\mathcal{V}_{C|X}|\delta_N \leq N\delta_N,
\end{aligned}$$

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 [31];

(d) - (e) results from the definitions of the conditional distributions in (40);

(f) follows from the sets defined in (35).

- Next, we show that $\mathbb{D}(Q_{Y^N|B^N\widehat{C}^N}||\tilde{P}_{Y_i^N|B_i^N\widehat{C}_i^N}) \leq N\delta_N$ with the following derivation:

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

$$\begin{aligned}
&\mathbb{D}(\tilde{P}_{X_i^NA_i^NC_i^NB_i^N\widehat{C}_i^NY_i^N}||Q_{X^NA^NC^NB^N\widehat{C}^NY^N})\\
&= \mathbb{D}(\tilde{P}_{Y_i^N|X_i^NA_i^NC_i^NB_i^N\widehat{C}_i^N}\tilde{P}_{X_i^NA_i^NC_i^NB_i^N\widehat{C}_i^N}||Q_{Y^N|X^NA^NC^NB^N\widehat{C}^N}Q_{X^NA^NC^NB^N\widehat{C}^N})\\
&\overset{(a)}{=} \mathbb{D}(\tilde{P}_{Y_i^N|B_i^N\widehat{C}_i^N}\tilde{P}_{X_i^NA_i^NC_i^NB_i^N\widehat{C}_i^N}||Q_{Y^N|B^N\widehat{C}^N}Q_{X^NA^NC^NB^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^NA_i^NC_i^N}\tilde{P}_{X_i^NA_i^NC_i^N}||Q_{Y^N|B^N\widehat{C}^N}Q_{B^N\widehat{C}^N|X^NA^NC^N}Q_{X^NA^NC^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^NC_i^N}\tilde{P}_{X_i^NA_i^NC_i^N}||Q_{Y^N|B^N\widehat{C}^N}Q_{B^N\widehat{C}^N|A^NC^N}Q_{X^NA^NC^N})\\
&\overset{(c)}{\leq} \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^NC_i^N}\tilde{P}_{X_i^NA_i^NC_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^NC_i^N}Q_{X^NA^NC^N})\\
&\qquad + \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^NC_i^N}Q_{X^NA^NC^N}||Q_{Y^N|B^N\widehat{C}^N}Q_{B^N\widehat{C}^N|A^NC^N}Q_{X^NA^NC^N})\\
&\overset{(d)}{=} \delta_N^{(2)} + \mathbb{D}(\tilde{P}_{X_i^NA_i^NC_i^N}||Q_{X^NA^NC^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^NC_i^N}||Q_{Y^N|B^N\widehat{C}^N}Q_{B^N\widehat{C}^N|A^NC^N})\\
&\overset{(e)}{\leq} \delta_N^{(2)} + \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^NC_i^N}||Q_{Y^N|B^N\widehat{C}^N}Q_{B^N\widehat{C}^N|A^NC^N})\\
&\overset{(f)}{=} \delta_N^{(2)} + \hat{\delta}_N^{(2)} + \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^NC_i^N}||Q_{B^N\widehat{C}^N|A^NC^N})\\
&\overset{(g)}{\leq} \delta_N^{(2)} + \hat{\delta}_N^{(2)} - 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})}\\
&\qquad - N\log(\mu_{ACB\widehat{C}})\sqrt{2\ln 2}\sqrt{\mathbb{D}(Q_{B^N\widehat{C}^N|A^NC^N}||\tilde{P}_{B_i^N\widehat{C}_i^N|A_i^NC_i^N})}\\
&\overset{(h)}{\leq} \delta_N^{(2)} + \hat{\delta}_N^{(2)} - 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}
\end{aligned}$$

$$\overset{(b)}{=}\sum_{j\in\mathcal{V}_{Y|BC}}\mathbb{E}_{Q_{T^{j-1}B_i^N C_i^N}}\Big[\mathbb{D}(Q_{T_j|T^{j-1}B^N\widehat{C}^N}||\tilde{P}_{T_j|T^{j-1}B_i^N\widehat{C}_i^N})\Big]$$
$$\overset{(c)}{=}\sum_{j\in\mathcal{V}_{Y|BC}}\Big(\log|\mathcal{Y}|-H(T_j|T^{j-1}B^NC^N)\Big)$$
$$\overset{(d)}{\le}|\mathcal{V}_{Y|BC}|\delta_N$$
$$\le N\delta_N,$$

where

(a) follows from the chain rule of KL divergence [31];
(b) - (c) results from the definitions of the conditional distribution in (41);
(d) follows from the set defined in (37). ■

Now, Lemmas 7 and 8 provide the independence between two consecutive blocks and the independence between all blocks, respectively, based on the results of Lemma 6.

Lemma 7: For block $i\in[\![2,k]\!]$, we have

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

where $\delta_N^{(3)}\triangleq\mathcal{O}(\sqrt[4]{N^{15}\delta_N})$, $\delta_N\triangleq 2^{-N^\beta}$ and $\beta<\frac{1}{2}$.

The proof of Lemma 7 can be found in Appendix B.

Lemma 8: We have

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

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

The proof of Lemma 8 can be found in Appendix C.

Finally, by the results of Lemma 8 we can show in Lemma 9 that the target distribution $Q_{X^NY^N}$ is approximated asymptotically over all blocks jointly.

Lemma 9: We have

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

where $\delta_N^{(4)}\triangleq\mathcal{O}(k^{3/2}N^{23/8}\delta_N^{1/8})$, $\delta_N\triangleq 2^{-N^\beta}$ and $\beta<\frac{1}{2}$.

Proof: We reuse the proof of [19, Lemma 5] with substitutions $q_{Y^{1:N}}\leftarrow Q_{X^NY^N}$, $\tilde{p}_{Y_i^{1:N}}\leftarrow\tilde{P}_{Y_i^N X_i^N}$. ■

C. Proof of Theorem 4

We can now proceed to show that the polar coding scheme described in Algorithms 1, 2 achieves the region stated in Theorem 1 and satisfies (32).

Proof: The common randomness rate R_o is given as

$$\begin{aligned}\frac{|\bar{J}_1|+|J_{1:k}|}{kN}&=\frac{|\mathcal{V}_{C|XY}|+k|\mathcal{V}_{C|X}\setminus\mathcal{V}_{C|XY}|}{Nk}\\&=\frac{|\mathcal{V}_{C|XY}|}{kN}+\frac{|\mathcal{V}_{C|X}\setminus\mathcal{V}_{C|XY}|}{N}\\&\xrightarrow{N\to\infty}\frac{H(C|XY)}{k}+I(Y;C|X)\\&\xrightarrow{k\to\infty}I(Y;C|X).\end{aligned}\tag{42}$$

The communication rate R_c is given as

$$\begin{aligned}\frac{k|\mathcal{F}_5\cup\mathcal{F}_3|}{kN}&=\frac{k|\mathcal{V}_C\setminus\mathcal{V}_{C|X}|}{Nk}=\frac{|\mathcal{V}_C\setminus\mathcal{V}_{C|X}|}{N}\\&\xrightarrow{N\to\infty}I(X;C),\end{aligned}\tag{43}$$

whereas R_a can be written as

$$\begin{aligned}\frac{|\mathcal{V}_{A|CXY}|+k|\mathcal{F}_8|}{kN}&=\frac{|\mathcal{V}_{A|CXY}|+k|\mathcal{V}_{A|C}\setminus\mathcal{V}_{A|CX}|}{kN}\\&=\frac{|\mathcal{V}_{A|CXY}|}{kN}+\frac{|\mathcal{V}_{A|C}\setminus\mathcal{V}_{A|CX}|}{N}\\&\xrightarrow{N\to\infty}I(A;X|C)+\frac{H(A|CXY)}{k}\\&\xrightarrow{k\to\infty}I(A;X|C).\end{aligned}\tag{44}$$

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

$$\begin{aligned}\rho_1&=\frac{k|\mathcal{F}_6|}{kN}=\frac{|\mathcal{V}_{A|CX}\setminus\mathcal{V}_{A|CXY}|}{N}\\&\xrightarrow{N\to\infty}I(A;Y|CX),\end{aligned}\tag{45}$$

$$\rho_2=\frac{k|V_{Y|BC}|}{kN}\xrightarrow{N\to\infty}H(Y|BC).\tag{46}$$

Finally we see that conditions (19a)-(19g) are satisfied by (42)-(46). Hence, given R_a, R_o, R_c satisfying Theorem 1, based on Lemma 9 and Pinsker's inequality [29] we have

$$\begin{aligned}&\mathbb{E}\big[||\tilde{P}_{X_{1:k}^N Y_{1:k}^N}-Q_{X^{1:kN}Y^{1:kN}}||_{TV}\big]\\&\le\mathbb{E}\Big[\sqrt{2\mathbb{D}(\tilde{P}_{X_{1:k}^N Y_{1:k}^N}||Q_{X^{1:kN}Y^{1:kN}})}\Big]\\&\le\sqrt{2\mathbb{E}\big[\mathbb{D}(\tilde{P}_{X_{1:k}^N Y_{1:k}^N}||Q_{X^{1:kN}Y^{1:kN}})\big]}\xrightarrow{N\to\infty}0.\end{aligned}\tag{47}$$

As a result, from (47) there exists an $N\in\mathbb{N}$ for which the polar code-induced pmf between the pair of actions satisfies the strong coordination condition in (32). ■

VIII. Summary and Concluding Remarks

In this paper, we have investigated a fundamental question regarding communication-based coordination: Is separate coordination and channel coding optimal in the context of point-to-point strong coordination? In particular, we considered a two-node strong coordination setup with a DMC as the communication link. To that extent, we presented achievability and infeasibility results for this setting and constructed a general joint coordination-channel encoding scheme based on random codes. We also provided a capacity result for a special case of the noisy strong coordination setup where the discrete memoryless communication channel is a deterministic channel. The proof technique underlying our joint coding scheme is based on channel resolvability, a technique which is widely used in analyzing strong coordination problems. In addition, we presented a benchmark scheme based on separate coordination and channel encoding and utilized randomness extraction to improve its performance. In this scheme, randomness is extracted from the channel at the decoder. In addition, by leveraging our random coding results, we presented an example for coordinating a doubly binary symmetric source over a binary symmetric communication channel in which the proposed joint scheme outperforms a separation-based scheme in terms of achievable communication rate. As a result, we conclude that a separation-based scheme, even if it exploits randomness extraction from the communication channel, is sub-optimal for this problem. Finally, we have also proposed a constructive coding scheme based on polar codes for the noisy two-node network that can achieve all the rates that the joint scheme can, where achievability is guaranteed asymptotically. Although this work yields some insight in coordination over noisy communication links, a tighter converse bound to establish the optimality of the presented joint encoding scheme is still open.

Appendix A
Proof of Lemma 3

The proof of Lemma 3 leverages the results from Section IV-A2. The bound on $||\check{P}_{X^n J} - Q_X^n P_J||_{TV}$ is obtained in a similar manner as in the proof of Lemma 1. Note that here we also drop the subscripts from the pmfs for simplicity, e.g., $P^n_{X|AC}(x^n|A^n_{ijk}, C^n_{ij})$ will be denoted by $P(x^n|A^n_{ijk}, C^n_{ij})$, and $Q_X^n(x^n)$ will be denoted by $Q(x^n)$ in the following.

Proof of Lemma 3: Consider the argument shown at the bottom of the page. In this argument: $\epsilon' > 0$, $\epsilon' \to 0$ as $n \to \infty$.

(a) follows from the law of iterated expectations. Note that we have used (a^n_{ijk}, c^n_{ij}) to denote the codewords corresponding to the indices (i,j,k) and $(a^n_{i'j'k'}, c^n_{i'j'})$ to denote the codewords corresponding to the indices (i',j',k'), respectively;

(b) follows from Jensen's inequality [31];

(c) follows from dividing the inner summation over the indices (i',k') into three subsets based on the indices (i,j,k) from the outer summation;

(d) results from taking the conditional expectation within the subsets in (c);

(e) follows from
- $(i',j',k') = (i,j,k)$: there is only one pair of codewords represented by the indices A^n_{ijk}, C^n_{ij} corresponding to x^n;
- when $(k' \neq k)$ and $(i',j') = (i,j)$ there are $(2^{nR_a} - 1)$ indices in the sum;
- $(i',j',k') \neq (i,j,k)$: the number of the indices is at most $2^{n(R_a+R_c)}$. Moreover, $P(x^n)$ is less than ϵ

$$
\begin{aligned}
\mathbb{E}_{\mathsf{C}}[\mathbb{D}(\check{P}_{X^n J}||Q_X^n P_J)] &= \mathbb{E}_{\mathsf{C}}\left[\sum_{x^n,j}\Big(\sum_{i,k}\frac{P(x^n|A^n_{ijk},C^n_{ij})}{2^{nR}}\Big)\log\Big(\sum_{i',k'}\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}Q(x^n)P(j)}\Big)\right]\\
&= \sum_{x^n}\sum_{i,j,k}\mathbb{E}_{\mathsf{C}}\left[\Big(\frac{P(x^n|A^n_{ijk},C^n_{ij})}{2^{nR}}\Big)\log\Big(\sum_{i',k'}\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{n(R_a+R_c)}Q(x^n)}\Big)\right]\\
&\overset{(a)}{=} \sum_{x^n}\sum_{i,j,k}\mathbb{E}_{A^n_{ijk},C^n_{ij}}\left[\Big(\frac{P(x^n|A^n_{ijk},C^n_{ij})}{2^{nR}}\Big)\mathbb{E}_{A^n_{i'j'k'}C^n_{i'j'}}\left[\log\Big(\sum_{i',k'}\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{n(R_a+R_c)}Q(x^n)}\Big)\Big|A^n_{ijk}C^n_{ij}\right]\right]\\
&\overset{(b)}{\leq} \sum_{x^n}\sum_{i,j,k}\mathbb{E}_{A^n_{ijk}C^n_{ij}}\left[\Big(\frac{P(x^n|A^n_{ijk},C^n_{ij})}{2^{nR}}\Big)\log\Big(\mathbb{E}_{A^n_{i'j'k'}C^n_{i'j'}}\Big[\sum_{i',k'}\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{n(R_a+R_c)}Q(x^n)}\Big|A^n_{ijk}C^n_{ij}\Big]\Big)\right]\\
&\overset{(c)}{=} \sum_{x^n}\sum_{a^n_{ijk},c^n_{ij}}\sum_{i,j,k}\frac{P(x^n,a^n_{ijk},c^n_{ij})}{2^{nR}}\log\Bigg(\sum_{\substack{i',k':\\(i',j',k')=(i,j,k)}}\mathbb{E}_{A^n_{i'j'k'}C^n_{i'j'}}\Big[\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{n(R_a+R_c)}Q(x^n)}\Big|A^n_{ijk}C^n_{ij}\Big]\\
&\quad+\sum_{\substack{i',k':\\(i',j')=(i,j),(k'\neq k)}}\mathbb{E}_{A^n_{i'j'k'}C^n_{i'j'}}\Big[\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{n(R_a+R_c)}Q(x^n)}\Big|A^n_{ijk}C^n_{ij}\Big]+\sum_{\substack{i',j',k':\\(i',j',k')\neq(i,j,k)}}\mathbb{E}_{A^n_{i'j'k'}C^n_{i'j'}}\Big[\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{n(R_a+R_c)}Q(x^n)}\Big|A^n_{ijk}C^n_{ij}\Big]\Bigg)\\
&\overset{(d)}{=} \sum_{x^n}\sum_{a^n_{ijk},c^n_{ij}}\sum_{i,j,k}\frac{P(x^n,a^n_{ijk},c^n_{ij})}{2^{nR}}\log\Bigg(\frac{P(x^n|a^n_{ijk},c^n_{ij})}{2^{n(R_a+R_c)}Q(x^n)}+\sum_{\substack{i',k':\\(i',j')=(i,j),(k'\neq k)}}\frac{P(x^n|c^n_{ij})}{2^{n(R_a+R_c)}Q(x^n)}+\sum_{\substack{i',k':\\(i',j',k')\neq(i,j,k)}}\frac{P(x^n)}{2^{n(R_a+R_c)}Q(x^n)}\Bigg)\\
&\overset{(e)}{\leq} \sum_{x^n}\sum_{a^n_{ijk},c^n_{ij}}P(x^n a^n_{ijk},c^n_{ij})\log\left(\frac{P(x^n|a^n_{ijk},c^n_{ij})}{2^{n(R_a+R_c)}Q(x^n)}+(2^{nR_a})\frac{P(x^n|c^n_{ij})}{2^{n(R_a+R_c)}Q(x^n)}+1\right)\\
&\overset{(f)}{\leq} \left[\sum_{(x^n,a^n,c^n)\in\mathcal{T}^n_\epsilon(P_{XAC})}P(x^n,a^n,c^n)\log\left(\frac{2^{-nH(X|AC)(1-\epsilon)}}{2^{n(R_a+R_o)}2^{-nH(X)(1+\epsilon)}}+\frac{2^{-nH(X|C)(1-\epsilon)}}{2^{nR_c}2^{-nH(X)(1+\epsilon)}}+1\right)\right]\\
&\qquad+\mathbb{P}\big((x^n,a^n,c^n)\notin\mathcal{T}^n_\epsilon(P_{XAC})\big)\log(2\mu_X^{-n}+1)\\
&\overset{(g)}{\leq} \left[\sum_{(x^n,a^n,c^n)\in\mathcal{T}^n_\epsilon(P_{XAC})}P(x^n,a^n,c^n)\log\left(\frac{2^{n(I(X;AC)+\delta(\epsilon))}}{2^{n(R_c+R_a)}}+\frac{2^{n(I(X;C)+\delta(\epsilon))}}{2^{n(R_c)}}+1\right)\right]\\
&\qquad+\big(2|\mathcal{X}||\mathcal{A}||\mathcal{C}|e^{-n\epsilon^2\mu_{XAC}}\big)\log(2\mu_X^{-n}+1)\\
&\overset{(h)}{\leq} \epsilon',
\end{aligned}
$$

close to $Q(x^n)$ as a consequence of Lemma 1 and [47, Lemma 16].

(f) results from splitting the outer summation: The first summation contains typical sequences and is bounded by using the probabilities of the typical set. The second summation contains the tuple of sequences when the action sequence x^n and codewords c^n, a^n, represented here by the indices (i, j, k), are not ϵ-jointly typical (i.e., $(x^n, a^n, c^n) \notin \mathcal{T}_\epsilon^n(P_{XAC})$). This sum is upper bounded following [6] with $\mu_X = \min_x^*(P_X)$.

(g) follows from the Chernoff bound on the probability that a sequence is not strongly typical [32].

(h) consequently, the contribution of typical sequences can be asymptotically made small if

$$R_a + R_c \geq I(X;AC), \quad R_c \geq I(X;C).$$

The second term converges to zero exponentially fast with n [32], and following Pinsker's inequality [29] we have

$$\begin{aligned}\mathbb{E}_{\mathsf{C}}\left[||\check{P}_{X^nJ} - Q_X^n P_J||_{TV}\right] &\leq \mathbb{E}_{\mathsf{C}}\left[\sqrt{2\mathbb{D}(\check{P}_{X^nJ}||Q_X^n P_J)}\right] \\ &\leq \sqrt{2\mathbb{E}_{\mathsf{C}}\left[\mathbb{D}(\check{P}_{X^nJ}||Q_X^n P_J)\right]} \leq \sqrt{2\epsilon'}.\end{aligned}$$

■

Appendix B
Proof of Lemma 7

We reuse the proof of [19, Lemma 3] with substitutions $q_{U^{1:N}} \leftarrow Q_{C^N}$, $q_{Y^{1:N}} \leftarrow Q_{X^NY^N}$, $\tilde{p}_{U_i^{1:N}} \leftarrow \tilde{P}_{C_i^N}$, $\tilde{p}_{Y_i^{1:N}} \leftarrow \tilde{P}_{Y_i^N X_i^N}$, and $\bar{R}_1 \leftarrow \bar{J}_1$. This results 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 [19, Lemma 3].

Proof of Lemma 7:

$$\begin{aligned}
&H(U_2^N[\mathcal{V}_{C|XY}]|X^NY^N) - H(\widetilde{U}_{2_i}^N[\mathcal{V}_{C|XY}]|X_i^N\widetilde{Y}_i^N) \\
&= H(U_2^N[\mathcal{V}_{C|XY}]X^NY^N) - H(\widetilde{U}_{2_i}^N[\mathcal{V}_{C|XY}]X_i^N\widetilde{Y}_i^N) \\
&\quad - H(X^NY^N) + H(X_i^N\widetilde{Y}_i^N) \\
&\overset{(a)}{\leq} \mathbb{D}(\tilde{P}_{U_{2_i}^N[\mathcal{V}_{C|XY}]X_i^NY_i^N}||Q_{U_2^N[\mathcal{V}_{C|XY}]X^NY^N}) \\
&\quad + \Big(N^3\log(|\mathcal{X}||\mathcal{Y}||\mathcal{C}|)\sqrt{2\ln 2} \\
&\qquad \times \sqrt{\mathbb{D}(\tilde{P}_{U_{2_i}^N[\mathcal{V}_{C|XY}]X_i^NY_i^N}||Q_{U_2^N[\mathcal{V}_{C|XY}]X^NY^N})}\Big) \\
&\quad + \mathbb{D}(Q_{X^NY^N}||\tilde{P}_{X_i^NY_i^N}) \\
&\quad + N^2\log(|\mathcal{X}||\mathcal{Y}|)\sqrt{2\ln 2}\sqrt{\mathbb{D}(\tilde{P}_{X_i^NY_i^N})||Q_{X^NY^N}} \\
&\overset{(b)}{\leq} \mathbb{D}(\tilde{P}_{U_{2_i}^N X_i^NY_i^N}||Q_{U_2^N X^NY^N}) \\
&\quad + N^3\log(|\mathcal{X}||\mathcal{Y}||\mathcal{C}|)\sqrt{2\ln 2}\sqrt{\mathbb{D}(\tilde{P}_{U_{2_i}^N X_i^NY_i^N}||Q_{U_2^N X^NY^N})} \\
&\quad + \mathbb{D}(Q_{X^NY^N}||\tilde{P}_{X_i^NY_i^N}) \\
&\quad + N^2\log(|\mathcal{X}||\mathcal{Y}|)\sqrt{2\ln 2}\sqrt{\mathbb{D}(\tilde{P}_{X_i^NY_i^N})||Q_{X^NY^N}} \\
&\overset{(c)}{\leq} \delta_N^{(1)} + N^3\log(|\mathcal{X}||\mathcal{Y}||\mathcal{C}|)\sqrt{2\ln 2}\sqrt{\delta_N^{(1)}} \\
&\quad - N\log(\mu_{XY})\sqrt{2\ln 2}\sqrt{\mathbb{D}(\tilde{P}_{X_i^NY_i^N}||Q_{X^NY^N})} \\
&\quad + N^2\log(|\mathcal{X}||\mathcal{Y}|)\sqrt{2\ln 2}\sqrt{\mathbb{D}(\tilde{P}_{X_i^NY_i^N}||Q_{X^NY^N})} \\
&\leq \delta_N^{(1)} + N^3\log(|\mathcal{X}||\mathcal{Y}||\mathcal{C}|)\sqrt{2\ln 2}\sqrt{\delta_N^{(1)}} \\
&\quad - N\log(\mu_{XY})\sqrt{2\ln 2}\sqrt{\delta_N^{(1)}} + N^2\log(|\mathcal{X}||\mathcal{Y}|)\sqrt{2\ln 2}\sqrt{\delta_N^{(1)}} \\
&\leq \hat{\delta}_N^{(3)},
\end{aligned}$$

where

(a) follows from [19, Lemma 17];

(b) follows from the chain rule of KL divergence [31];

(c) follows from Lemma 6 and [19, Lemma 14].

Hence, for block $i \in [\![2, k]\!]$, we have

$$\begin{aligned}
&\mathbb{D}(\tilde{P}_{X_{i-1:i}^N Y_{i-1:i}^N \bar{J}_1}||\tilde{P}_{X_{i-1}^N Y_{i-1}^N \bar{J}_1}\tilde{P}_{X_i^NY_i^N}) \\
&= I(X_{i-1}^N\widetilde{Y}_{i-1}^N\bar{J}_1; X_i^N\widetilde{Y}_i^N) \\
&= I(X_i^N\widetilde{Y}_i^N; \bar{J}_1) + I(X_{i-1}^N\widetilde{Y}_{i-1}^N; X_i^N\widetilde{Y}_i^N|\bar{J}_1) \\
&= I(X_i^N\widetilde{Y}_i^N; \bar{J}_1) \\
&= I(X_i^N\widetilde{Y}_i^N; U_{2_i}^N[\mathcal{V}_{C|XY}]) \\
&= H(U_{2_i}^N[\mathcal{V}_{C|XY}]) - H(\widetilde{U}_{2_i}^N[\mathcal{V}_{C|XY}]|X_i^N\widetilde{Y}_i^N) \\
&\overset{(a)}{\leq} |\mathcal{V}_{C|XY}|\log(|\mathcal{C}|) - H(U_2^N[\mathcal{V}_{C|XY}]|X^NY^N) + \hat{\delta}_N^{(3)} \\
&\overset{(b)}{\leq} |\mathcal{V}_{C|XY}| - \sum_{j\in\mathcal{V}_{C|XY}} H(U_{2,j}|X^NY^NU^{j-1}) + \hat{\delta}_N^{(3)} \\
&\leq |\mathcal{V}_{C|XY}| - |\mathcal{V}_{C|XY}|(1-\delta_N) + \hat{\delta}_N^{(3)} \\
&= |\mathcal{V}_{C|XY}|\delta_N + \hat{\delta}_N^{(3)} \\
&\leq N\delta_N + \hat{\delta}_N^{(3)} \leq \delta_N^{(3)},
\end{aligned}$$

where (a) follows from (48), and (b) follows from the definition of the high entropy sets (35). ■

Appendix C
Proof of Lemma 8

We reuse the proof of [19, Lemma 4] with substitutions $\tilde{p}_{Y^{1:N}} \leftarrow \tilde{P}_{X_i^NY_i^N}$, and $\bar{R}_1 \leftarrow \bar{J}_1$. This will result in the Markov chain $X_{1:i-2}^N\widetilde{Y}_{1:i-2}^N - \bar{J}_1X_{i-1}^N\widetilde{Y}_{i-1}^N - X_i^N\widetilde{Y}_i^N$, replacing the chain in [19, Lemma 4].

Proof of Lemma 8:

$$\begin{aligned}
&\mathbb{D}\Big(\tilde{P}_{X_{1:k}^NY_{1:k}^N}||\prod_{i=1}^k \tilde{P}_{X_i^NY_i^N}\Big) \\
&\overset{(a)}{=} \sum_{i=2}^k I(X_i^N\widetilde{Y}_i^N; X_{1:i-1}^N\widetilde{Y}_{1:i-1}^N) \\
&\leq \sum_{i=2}^k I(X_i^N\widetilde{Y}_i^N; X_{1:i-1}^N\widetilde{Y}_{1:i-1}^N\bar{J}_1) \\
&= \sum_{i=2}^k I(X_i^N\widetilde{Y}_i^N; X_{i-1}^N\widetilde{Y}_{i-1}^N\bar{J}_1) \\
&\qquad + I(X_i^N\widetilde{Y}_i^N; X_{2:i-1}^N\widetilde{Y}_{2:i-1}^N|X_{i-1}^N\widetilde{Y}_{i-1}^N\bar{J}_1) \\
&\overset{(b)}{=} \sum_{i=2}^k \mathbb{D}(\tilde{P}_{X_{i-1:i}^NY_{i-1:i}^N\bar{J}_1}||\tilde{P}_{X_{i-1}^NY_{i-1}^N\bar{J}_1}\tilde{P}_{X_i^NY_i^N}) \\
&\overset{(c)}{\leq} \sum_{i=2}^k \delta_N^{(3)} = (k-1)\delta_N^{(3)},
\end{aligned}$$

where
(a) follows from [19, Lemma 15].
(b) holds by the Markov chain $X_{1:i-2}^N \widetilde{Y}_{1:i-2}^N - \bar{J}_1 X_{i-1}^N \widetilde{Y}_{i-1}^N - X_i^N \widetilde{Y}_i^N$.
(c) follows from Lemma 7.

■

REFERENCES

[1] S. A. Obead, B. N. Vellambi, and J. Kliewer, "Strong coordination over noisy channels: Is separation sufficient?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2840–2844.

[2] S. A. Obead, J. Kliewer, and B. N. Vellambi, "Joint coordination-channel coding for strong coordination over noisy channels based on polar codes," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Oct. 2017, pp. 580–587.

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

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

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

[6] M. R. Bloch and J. Kliewer, "Strong coordination over a three-terminal relay network," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Hobart, TAS, Australia, Nov. 2014, pp. 646–650.

[7] B. N. Vellambi, J. Kliewer, and M. R. Bloch, "Strong coordination over multi-hop line networks using channel resolvability codebooks," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1132–1162, Feb. 2018.

[8] A. Bereyhi, M. Bahrami, M. Mirmohseni, and M. R. Aref, "Empirical coordination in a triangular multiterminal network," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2149–2153.

[9] 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.

[10] 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, Jun. 2015.

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

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

[13] P. Cuff and C. Schieler, "Hybrid codes needed for coordination over the point-to-point channel," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Sep. 2011, pp. 235–239.

[14] 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.

[15] G. Cervia, L. Luzzi, M. L. Treust, and M. R. Bloch, "Strong coordination of signals and actions over noisy channels with two-sided state information," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4681–4708, Aug. 2020.

[16] 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, Jul. 2009.

[17] E. Arikan, "Source polarization," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 899–903.

[18] M. R. Bloch, L. Luzzi, and J. Kliewer, "Strong coordination with polar codes," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Oct. 2012, pp. 565–571.

[19] R. A. Chou, M. R. Bloch, and J. Kliewer, "Empirical and strong coordination via soft covering with polar codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5087–5100, Jul. 2018.

[20] R. Blasco-Serrano, R. Thobaben, and M. Skoglund, "Polar codes for coordination in cascade networks," in *Int. Zurich Seminar Commun. (IZS)*. Zürich, Switzerland: ETH Zürich, 2012, pp. 55–58.

[21] 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 (ITW)*, Sep. 2016, pp. 81–85.

[22] B. N. Vellambi, J. Kliewer, and M. R. Bloch, "Strong coordination over multi-hop line networks," in *Proc. IEEE Inf. Theory Workshop Fall (ITW)*, Jeju Island, South Korea, Oct. 2015, pp. 192–196.

[23] Y. Steinberg and S. Verdu, "Channel simulation and coding with side information," *IEEE Trans. Inf. Theory*, vol. 40, no. 3, pp. 634–646, May 1994.

[24] M. R. Bloch and J. Kliewer, "On secure communication with constrained randomization," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 1172–1176.

[25] M. Mojahedian, M. Aref, and A. Gohari, "On the equivalency of reliability and security metrics for wireline networks," *Scientia Iranica, Trans. D, Comput. Sci. Eng. Elect. Eng.*, vol. 26, no. 3, pp. 1749–1761, 2019.

[26] 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.

[27] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, "Zero-error channel capacity and simulation assisted by non-local correlations," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5509–5523, Aug. 2011.

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

[29] M. Pinsker, "Information and information stability of random variables and processes," in *London-Amsterdam XII*, vol. 243, A. Feinstein, Ed. San Francisco, CA, USA: Holden-Day, 1964.

[30] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3989–4001, Jun. 2011.

[31] T. M. Cover and J. Thomas, *Elements of Information Theorey* (Wiley Series in Telecommunications and Signal Processing), 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[32] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, nos. 4–5, pp. 265–444, 2008.

[33] A. El Gamal and Y. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[34] G. B. Dantzig and B. C. Eaves, "Fourier-Motzkin elimination and its dual," *J. Combinat. Theory A*, vol. 14, no. 3, pp. 288–297, May 1973.

[35] P. Cuff, "Communication requirements for generating correlated random variables," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 1393–1397.

[36] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2926–2959, May 2014.

[37] I. Csiszár, "Almost independence and secrecy capacity," *Problems Inf. Transmiss.*, vol. 32, no. 1, pp. 40–47, 1996.

[38] S. Vembu and S. Verdu, "Generating random bits from an arbitrary source: Fundamental limits," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1322–1332, Sep. 1995.

[39] M. Bloch, "Channel intrinsic randomness," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 2607–2611.

[40] A. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 163–179, Mar. 1975.

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

[42] 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.

[43] 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.

[44] E. En 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, Jul. 2016.

[45] 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. Allerton Conf.*, Monticello, IL, USA, Oct. 2003, pp. 430–439.

[46] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*, 1st ed. New York, NY, USA: Cambridge Univ. Press, 2011.

[47] P. Cuff, "Communication in networks for coordinating behavior," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 2009.

Sarah A. Obead (Graduate Student Member, IEEE) received the B.S. degree in information technology from the University of Benghazi (Garyounis University), Benghazi, Libya, in 2010, and the M.Sc. degree in telecommunication from the New Jersey Institute of Technology, Newark, NJ, USA, in 2016. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology. Her current research interests lie in the areas of information theory and coding theory with applications to private information retrieval and private computation.

Badri N. Vellambi (Senior Member, IEEE) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology-Madras, Chennai, India, in 2002, and the M.S. degree in electrical engineering, the M.S. degree in mathematics, and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2005, 2008, and 2008, respectively. He held Post-Doctoral Research Fellowship positions with the Institute for Telecommunications Research, University of South Australia, the New Jersey Institute of Technology, Newark, NJ, USA, and the Research School of Computer Science, Australian National University, Canberra, Australia, from August 2008 to August 2018. He is currently an Assistant Professor with the Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH, USA. His current research interests include information theory, data compression, channel coding, statistical learning, reinforcement learning, and artificial intelligence.

Jörg Kliewer (Senior Member, IEEE) received the Dr. Ing. degree (Ph.D.) in electrical engineering from the University of Kiel, Germany, in 1999. From 1993 to 1998, he was a Research Assistant at the University of Kiel, where he was a Senior Researcher and Lecturer from 1999 to 2004. In 2004, he visited the University of Southampton, U.K., for one year. From 2005 to 2007, he was with the University of Notre Dame, Notre Dame, IN, USA, as a Visiting Assistant Professor. From 2007 to 2013, he was with New Mexico State University, Las Cruces, NM, USA, most recently as an Associate Professor. He is currently with the New Jersey Institute of Technology, Newark, NJ, USA, as a Professor. His research interests span information theory, machine learning, graphical models, and statistical algorithms with applications to secure and private communication and data storage. He has been a member of the Editorial Board of the IEEE Information Theory Society Newsletter since 2012. He is a recipient of a Leverhulme Trust Award, a German Research Foundation Fellowship Award, and a Fulbright Scholarship. He was an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2008 to 2014. Since 2015, he has been serving as an Area Editor for the same journal. He has also been an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2017 to 2020.