Strong Coordination over a Line when Actions are Markovian

Badri N. Vellambi, Jörg Kliewer
New Jersey Institute of Technology, Newark, NJ 07102
Email: {badri.n.vellambi, jkliewer}@njit.edu

Matthieu R. Bloch
Georgia Institute of Technology, Atlanta, GA 30332
Email: matthieu.bloch@ece.gatech.edu

***Abstract*—We analyze the problem of strong coordination over a multi-hop line network when the actions to be generated by the nodes satisfy a Markov chain that is matched to the network topology. We devise and prove the optimality of two schemes that cover the portions of the capacity region corresponding to unlimited or no common randomness shared by all nodes.**

***Index Terms*—Strong coordination, channel resolvability, channel synthesis, line network.**

I. Introduction

Recently, the problem of coordinating actions of autonomous agents in a distributed fashion has become an important aspect in networked communication. The goal here is to minimize the required communication overhead to facilitate coordination according to a prescribed target distribution [1]. Two models of coordination have been studied in the literature: (a) *empirical* coordination, where the histogram of induced actions at the agents is required to follow the target distribution, and (b) *strong* coordination, where the sequence of joint actions is required to be statistically indistinguishable from that generated by a source corresponding to the target distribution. As both strong and empirical coordination require lower communication rates when compared to a straightforward transmission of the action sequences, these schemes can be useful in applications such as distributed control, multi-agent based exploration and surveillance, and distributed random variable generation [2].

The fundamental limits for both coordination models are known for the point-to-point case [1], [3], [4]. Recent works have focused on the three-terminal relay channel setting. In particular, inner and outer bounds on the required rate of communication have been derived for one- and two-way communication in [5] and [6], respectively. Recently, the problem of strong coordination over a multi-hop line network was studied under two assumptions: (a) each node possesses local randomness that only it has access to; and (b) common randomness is available to all the nodes [7]. The trade-offs among the rates of communication, and of local and common randomness have been quantified in the setting where the node initiating the coordination possesses unlimited local randomness, and next-hop message generation at any intermediate node does not use its local randomness.

In this work, we focus on the setting in [7] with the additional requirement that the actions of the network nodes form a Markov chain that is matched to the network topology. For this special setting, we study the two extreme cases of unlimited or no shared randomness common to all nodes. The fundamental tradeoffs between the required local randomness at each node, and the rates of communication required between nodes are characterized in both extreme cases. We note that the optimal schemes in both extreme cases are different from one another, and conclude by introducing a joint scheme that combines the features of both optimal schemes.

The remainder of this work is organized as follows. Section II presents the notation used, Section III presents the formal definition of the strong coordination problem studied, and Section IV presents the main results of this work.

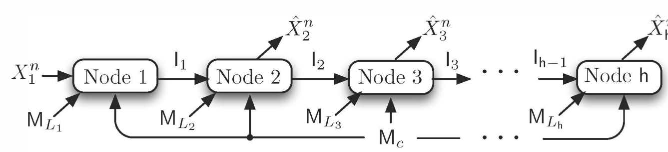
This work is supported by NSF grants CCF-1440014, CCF-1439465, and CCF-1320304.

II. Notation

For $m, n \in \mathbb{N}$ with $m < n$, $[\![m, n]\!] \triangleq \{m, m+1, \ldots, n\}$. Uppercase letters (e.g., X, Y) denote random variables (RVs), and the respective script versions (e.g., $\mathcal{X}$, $\mathcal{Y}$) denote their alphabets. In this work, all alphabets are assumed to be finite. Lowercase letters denote the realizations of random variables (e.g., x, y). Superscripts indicate the length of vectors. Single subscripts always indicate the node indices. In case of double subscripts, the first indicates the node index, and the next, the component index. The uniform p.m.f on a finite set S is given by $\mathsf{unif}(S)$. The variational distance between two p.m.f.s p and q over the same set is given by $\mathbb{V}(p, q)$. Lastly, $p^{\otimes n}_{X_1 X_2 \cdots X_k}$ denotes the p.m.f. of n i.i.d random vectors, with each random vector distributed according to p.m.f. $p_{X_1 X_2 \cdots X_k}$.

III. Problem Definition and Preliminary Results

The line coordination problem is a multi-hop extension of the one studied in [9], and is depicted in Fig. 1. For the sake of completeness, the problem is formally defined here.



Given: $X_1^n \sim \mathsf{Q}_{X_1}^{\otimes n}$ Require: $(X_1^n, \hat{X}_2^n, \ldots, \hat{X}_{\mathsf{h}}^n)$ '$\sim$' $\mathsf{Q}^{\otimes n}_{X_1 \cdots X_{\mathsf{h}}}$

Fig. 1. Problem Setup.

A line network with h nodes (Nodes $1, \ldots, \mathsf{h}$) and $\mathsf{h} - 1$ links (modeled as noiseless bit pipes) that connect Node i with Node $i + 1$, $1 \le i < \mathsf{h}$, is given. Node 1 is specified an action sequence $\{X_{1,i}\}_{i \in \mathbb{N}}$, an i.i.d process with each component distributed according to p.m.f. Q_{X_1} over a finite set $\mathcal{X}_1$. Nodes are assumed to possess local randomness, as well

978-1-4673-9457-4/16/$31.00 ©2016 IEEE

as common randomness shared by all h nodes to enable *strong coordination* using block codes. A block code (of length n) uses n symbols of the specified action (i.e., X_1^n), and common and local randomness to generate actions $\hat{X}_i^n$ at Nodes i, $i > 2$ satisfying the following condition: the joint statistics of actions $(X_1^n, \hat{X}_2^n, \ldots, \hat{X}_{\mathsf{h}}^n)$ and those of n symbols output by discrete memoryless source $\mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}^{\otimes n}$ are nearly indistinguishable under the variational distance metric. The overall aim is to quantify the required rates of communication, and of common and local randomness to achieve strong coordination. In this work, we focus solely on the case where the joint p.m.f. of the actions satisfies the following Markov chain.

$$X_1 — X_2 — X_3 — \cdots — X_{\mathsf{h}}. \tag{1}$$

The following definitions are now in order.

Definition 1: Given a p.m.f. $\mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}$ and $\varepsilon > 0$, and rate tuple $\mathbf{R} = (\mathsf{R}_c, \mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1}, \rho_1, \ldots, \rho_{\mathsf{h}}) \subseteq {\mathbb{R}^+}^{2\mathsf{h}}$, a strong coordination ε-code of length n at $\mathbf{R}$ is a collection of $\mathsf{h}+1$ independent and uniform random variables $(\mathsf{M}_c, \mathsf{M}_{L_1}, \ldots, \mathsf{M}_{L_{\mathsf{h}}})$, $\mathsf{h}-1$ message-generating functions $\{\psi_j\}_{j=1}^{\mathsf{h}-1}$, and $\mathsf{h}-1$ action-generating functions $\{\phi_j\}_{j=2}^{\mathsf{h}}$ subject to:

- Common and local randomness constraints:

$$\mathsf{M}_c \sim \mathsf{unif}(\llbracket 1, 2^{n\mathsf{R}_c} \rrbracket), \tag{2}$$

$$\mathsf{M}_{L_i} \sim \mathsf{unif}(\llbracket 1, 2^{n\rho_i} \rrbracket), \quad i = 1, \ldots, \mathsf{h}. \tag{3}$$

- Message-generation and action-generation constraints:

$$\mathsf{I}_j \triangleq \begin{cases} \psi_1(\mathsf{M}_{L_1}, X_1^n, \mathsf{M}_c) \in \llbracket 1, 2^{n\mathsf{R}_1} \rrbracket, & i = 2 \\ \psi_j(\mathsf{M}_{L_j}, \mathsf{I}_{j-1}, \mathsf{M}_c) \in \llbracket 1, 2^{n\mathsf{R}_j} \rrbracket, & 2 \le j < \mathsf{h} \end{cases}, \tag{4}$$

$$\hat{X}_j^n \triangleq \phi_j(\mathsf{M}_{L_j}, \mathsf{I}_{j-1}, \mathsf{M}_c), \quad 2 \le j \le \mathsf{h}. \tag{5}$$

- Strong coordination constraint:

$$\mathbb{V}\big(\mathsf{Q}_{X_1}^{\otimes n} \hat{Q}_{\hat{X}_2^n \cdots \hat{X}_{\mathsf{h}}^n | X_1^n}, \mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}^{\otimes n}\big) \le \varepsilon, \tag{6}$$

where $\hat{Q}_{\hat{X}_2^n \cdots \hat{X}_{\mathsf{h}}^n | X_1^n}$ is the conditional p.m.f. of the actions at Nodes $2, \ldots, \mathsf{h}$ induced by the code. ■

Definition 2: Strong coordination of actions according to a joint p.m.f. $\mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}$ is said to be *achievable* at a rate tuple $\mathbf{R} \triangleq (\mathsf{R}_c, \mathsf{R}_1, ..., \mathsf{R}_{\mathsf{h}-1}, \rho_1, ..., \rho_{\mathsf{h}}) \subseteq {\mathbb{R}^+}^{2\mathsf{h}}$ if for any $\varepsilon > 0$, there exists a strong coordination ε-code of some length $n \in \mathbb{N}$ at $\mathbf{R}$. The strong coordination capacity region is defined to be the closure of the set of all achievable rate vectors. ■

Before we proceed to the main results, we present a preliminary result that summarizes the *rate-transfer* arguments natural to the problem formulation, followed by a brief introduction of the one-hop strong coordination scheme, which we will use repeatedly to build a code for the multi-hop setting at hand.

Lemma 1: If strong coordination is achievable using communication rates $(\mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1})$, common randomness rate R_c, and local randomness rates $(\rho_1, \ldots, \rho_{\mathsf{h}})$, then:

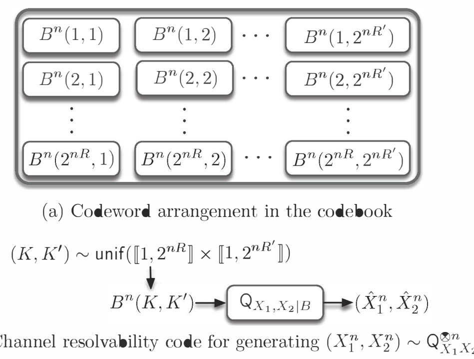
A. For any $1 \le \ell \le \mathsf{h}$ and $\varepsilon \le \rho_\ell$, strong coordination is also achievable using communication rates $(\mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1})$, common randomness rate $\mathsf{R}_c + \varepsilon$, and local randomness rates $(\rho_1, \ldots, \rho_{\ell-1}, \rho_\ell - \varepsilon, \rho_{\ell+1}, \ldots, \rho_{\mathsf{h}})$; and

B. For any $1 < \ell \le \mathsf{h}$ and $\delta \le \rho_\ell$, strong coordination is also achievable using communication rates $(\mathsf{R}_1, \ldots, \mathsf{R}_{\ell-2}, \mathsf{R}_{\ell-1} + \delta, \mathsf{R}_\ell, \ldots, \mathsf{R}_{\mathsf{h}-1})$, common randomness rate R_c, and local randomness rates $(\rho_1, \ldots, \rho_{\ell-1} + \delta, \rho_\ell - \delta, \rho_{\ell+1}, \ldots, \rho_{\mathsf{h}})$.

Proof: The proofs are straightforward and are omitted. ■

Point-to-point (one-hop) Strong Coordination

Since the coding strategies for the multi-hop line network will be based on the repeated use of the one-hop (point-to-point) strong coordination code, it will be beneficial to briefly describe the optimal strategy for the one-hop setting. To build a strong coordination code for a p.m.f. Q_{X_1,X_2}, we first choose an auxiliary RV B such that the joint p.m.f $\mathsf{Q}_{X_1X_2B}$ decomposes as $\mathsf{Q}_B\mathsf{Q}_{X_1|B}\mathsf{Q}_{X_2|B}$. Then, for a sufficiently large n, we generate $2^{n(R+R')}$ B-codewords in an i.i.d. fashion using the marginal Q_B, and arrange it in a matrix of 2^{nR} rows and $2^{nR'}$ columns as depicted in Fig. 2. Suppose now that we



(a) Codeword arrangement in the codebook

(b) Channel resolvability code for generating $(X_1^n, X_2^n) \sim \mathsf{Q}_{X_1X_2}^{\otimes n}$

Fig. 2. A channel resolvability code for generating actions.

select a codeword uniformly at random (using uniform RVs K and K' to select the row and column, respectively), and define the actions of the two nodes to be the output of the channel $\mathsf{Q}_{X_1,X_2|B}$ when the input is the chosen codeword. Then, from the well known results on channel resolvability [10] and coordination capacity [1], we are guaranteed that

$$\mathbb{V}(Q_{\hat{X}_1^n,\hat{X}_2^n}, \mathsf{Q}_{X_1,X_2}^{\otimes n}) \stackrel{n\to\infty}{\longrightarrow} 0, \tag{7}$$

$$\mathbb{V}(Q_{\hat{X}_1^n,K'}, \mathsf{Q}_{\hat{X}_1^n} \cdot \mathsf{unif}\llbracket 1, 2^{nR'} \rrbracket)) \stackrel{n\to\infty}{\longrightarrow} 0, \tag{8}$$

provided $R + R' > I(X_1, X_2; B)$ and $R > I(X_1; B)$.

Note that the channel resolvability code induces a joint p.m.f. $Q_{K,K',\hat{X}_1^n,X_2^n}$. To design a (randomized) strong coordination code, we view K' and R' as parts of common randomness and its rate, and proceed as follows. Given instances of the specified action at Node 1, and common randomness, say x_1^n and k', respectively, we generate an instance of $K \sim Q_{K|\hat{X}_1^n = x_1^n, K' = k'}$, and transmit it to Node 2. From Lemma 2 of the Appendix, we see that generating this instance needs a local randomness rate of $\rho_1 > R - I(X_1; B)$.

Now, upon receiving the instance of K, Node 2 identifies the codeword chosen by Node 1, since it is privy to common randomness $K' = k'$. Node 2 then uses the channel $Q_{X_2|B}$ to generate its output, which by the result on channel synthesis [11], can be achieved if $\rho_2 > H(X_2|B)$.

Upon using the allowable rate transfers of Lemma 1, and assigning the coding parameters to the system resource parameters (of Definition 1), we see that $(\mathsf{R}_c, \mathsf{R}_1, \rho_1, \rho_2)$ is achievable if:

$$\begin{aligned} \mathsf{R}_c &> R' + \delta_1 + \delta_2 \\ \mathsf{R}_1 &> R + \varepsilon_2 \\ \rho_1 &> R - I(X_1; B) + \varepsilon_2 - \delta_1 \\ \rho_2 &> H(X_2|B) - \varepsilon_2 - \delta_2 \\ R + R' &> I(X_1, X_2; B) \\ R &> I(X_1; B), \end{aligned} \tag{9}$$

where δ_i, $i = 1, 2$, is the variable enabling rate transfer from common randomness to the local randomness at Node i, and ε_2 is the variable enabling rate transfer from the local randomness at Node 1 to the local randomness at Node 2. Lastly, applying Fourier-Motzkin elimination to get rid of $R, R', \delta_1, \delta_2, \varepsilon_2$ yields the achievable region to be:

$$\mathsf{R}_1 > I(X_1; B) \tag{10}$$

$$\mathsf{R}_c + \rho_1 > I(X_2; B|X_1) \tag{11}$$

$$\mathsf{R}_c + \rho_1 + \rho_2 > H(X_2|X_1) \tag{12}$$

$$\mathsf{R}_c + \mathsf{R}_1 > I(X_1, X_2; B) \tag{13}$$

$$\mathsf{R}_c + \mathsf{R}_1 + \rho_2 > I(X_1, X_2; B, X_2), \tag{14}$$

whose optimality can be shown by arguments similar to [7].

IV. New Results on Fundamental Tradeoffs

We partition this section into three parts; the first is devoted to the case where the common randomness shared by nodes in the multi-hop network is sufficiently large, and the second to the case where there is no shared randomness. In the last, we shed light as to why the optimal scheme in the intermediary setting where common randomness is limited is non-trivial.

A. *Sufficiently Large Common Randomness*

Let us quantify the portion of the capacity region corresponding to *sufficiently large* common randomness, by which we mean $\mathsf{R}_c > H(X_2, \ldots, X_h|X_1)$. Since the total entropy embedded in the actions of Nodes $2, \ldots, \mathsf{h}$ equals $H(X_2, \ldots, X_h|X_1)$, it is, *a priori*, believable that strong coordination can be achieved without requiring local randomness.

To do so, let us build a nested set of channel resolvability codes with $\mathsf{h} - 1$ stages akin to our approach in [7]. Fix $\varepsilon > 0$. For the first stage, let us built a point-to-point strong coordination code between Node 1 and Node h using a one-hop strong coordination code with $B = X_\mathsf{h}$. For this choice, Node 1 decides what action Node h must take. From (10)-(14), we see that coordination between Node 1 and Node h is possible by: (1) expending a portion of common randomness of rate $H(X_\mathsf{h}|X_1) + \frac{\varepsilon}{2}$ for selecting the second of the indices corresponding to the X_h-codeword; (2) communicating a message of rate $I(X_1; X_\mathsf{h}) + \frac{\varepsilon}{2}$ from Node 1 to Node h; and (3) by expending a chunk of common randomness of size ε as the local randomness at Node 1 to enable the selection of the communicated message according to the correct conditional distribution. Note that we do not expend any local randomness at Node j, $j \geq 1$.

Now, for the second layer of coding, we build a point-to-point strong coordination code between Node 1 and Node $\mathsf{h} - 1$ using, again, a one-hop strong coordination code with $B = X_{\mathsf{h}-1}$. Since the first stage enables Nodes 2,..., $\mathsf{h} - 1$ to also be aware of the chosen action for Node h, we can build the codebook described in Fig. 2 by using $\mathsf{Q}_{X_{\mathsf{h}-1}|X_\mathsf{h}} = \mathsf{Q}_{X_{\mathsf{h}-1}|X_\mathsf{h}}$ instead of $\mathsf{Q}_{X_{\mathsf{h}-1}}$. Note that, in fact, we are building a codebook for *every* codeword in the first-stage codebook. Again, for this choice, we can show that we achieve coordination, which is now among Nodes $1, \mathsf{h} - 1$ and h by: (1) expending a portion of common randomness of rate $H(X_{\mathsf{h}-1}|X_\mathsf{h}, X_1) + \frac{\varepsilon}{2}$ for selecting the second of the indices corresponding to the $X_{\mathsf{h}-1}$-codeword; (2) communicating a message of rate $I(X_1; X_{\mathsf{h}-1}|X_\mathsf{h}) + \frac{\varepsilon}{2}$ from Node 1 to Node $\mathsf{h} - 1$; and (3) by expending a chunk of common randomness of size ε as the local randomness at Node 1 to enable the selection of the communicated message according to the correct conditional distribution. Again, note that we do not expend any local randomness at Node j, $j \geq 1$.

Until we complete the $\mathsf{h} - 1^{\text{th}}$ stage for generating the action of Node 2, we proceed inductively by constructing at each stage, exponentially many codebooks, one for each sequence of codewords for all prior stages. So for $1 < s < \mathsf{h}$, the s^{th} codebook is constructed by using $\mathsf{Q}_{X_{\mathsf{h}-s+1}|X_{\mathsf{h}-s+2},\ldots,X_\mathsf{h}}$; this codebook is used to generate the action at Node $\mathsf{h} - s + 1$, and uses: (1) a portion of common randomness of rate $H(X_{\mathsf{h}-s+1}|X_{\mathsf{h}-s+2}, \ldots, X_\mathsf{h}, X_1) + \frac{\varepsilon}{2}$ for selecting the second of the indices corresponding to the $X_{\mathsf{h}-s+1}$-codeword; (2) communicating a message of rate $I(X_1; X_{\mathsf{h}-s+1} \mid X_{\mathsf{h}-s+2}, \ldots, X_\mathsf{h})$ from Node 1 to Node $\mathsf{h} - s + 1$ and (3) a chunk of common randomness of size ε as the local randomness at Node 1 to enable the selection of the communicated message according to the correct conditional distribution. Again, note that we do not expend any local randomness at Node j, $j \geq 1$.

Note that ε can be chosen vanishingly small. So, the total resources expended are:

$$\mathsf{R}_c = H(X_2, \ldots, X_\mathsf{h} \mid X_1), \tag{15}$$

$$\mathsf{R}_i = I(X_1; X_{i+1}, \ldots, X_\mathsf{h}), \quad 1 \leq i < \mathsf{h}, \tag{16}$$

$$\rho_i = 0, \qquad 1 \leq i \leq \mathsf{h}, \tag{17}$$

which corroborates our intuition that no local randomness should be needed when common randomness is sufficiently large. Lastly, the optimality of this scheme is evident from the cut-set argument that for $1 \leq i < \mathsf{h}$, the rate between Node i and Node $i + 1$ can be no smaller than the smallest rate in a one-hop network where the first node is specified the action X_1^n and the second node requires the action $(X_{i+1}^n, \ldots, X_\mathsf{h}^n)$. This smallest rate obtained from (10) equals:

$$\min_{X_1 - B - (X_{i+1}, \ldots, X_\mathsf{h})} I(X_1; B) \triangleq I(X_1; X_{i+1}, \ldots, X_\mathsf{h}).$$

Remark 1: The set of achievable communication and local randomness rates given by (15) and (16), and its optimality for $\mathsf{R}_c > H(X_2, \ldots, X_\mathsf{h}|X_1)$ holds for any p.m.f. $\mathsf{Q}_{X_1,\ldots,X_\mathsf{h}}$. However, when the p.m.f. satisfies the Markov chain in (1), the rate conditions simplify to $\mathsf{R}_i > I(X_1; X_{i+1})$, $1 \le i < \mathsf{h}$.

B. No Common Randomness

For strong coordination in the absence of common randomness, we again make repeated use of the one-hop scheme with one major difference. Instead of a nested codebook setup, the multi-hop strong coordination code in this case is the juxtaposition of $\mathsf{h}-1$ one-hop strong coordination codes, i.e., we use X_1^n and a one-hop code to generate X_2^n, and use the generated X_2^n and another one-hop code to generate X_3^n, and so on. To, achieve this, we pick $Q_{X_1,\ldots,X_\mathsf{h},Z_1,\ldots,Z_{\mathsf{h}-1}}$ such that:

$$Q_{X_1,\ldots,X_\mathsf{h}} = \mathsf{Q}_{X_1,\ldots,X_\mathsf{h}} \tag{18}$$

$$X_1 — Z_1 — X_2 — Z_2 — \cdots — Z_{\mathsf{h}-1} — Z_\mathsf{h} \tag{19}$$

By generating and using the codebook for Z_i to coordinate X_i with X_{i+1} independent of all other codebooks' generations and usages, we can ensure that the tuple of h actions indeed meets the variational distance criterion imposed by (6). Since we are working in the no common randomness regime, we see from (10)-(14), that implementation of the $\mathsf{h}-1$ individual one-hop coordination schemes will require the following resources.

- for $1 \le j < \mathsf{h}$, a communication rate of $I(X_j, X_{j+1}; Z_j)$ between Nodes j and $j+1$;
- a local randomness rate of $I(X_2; Z_1|X_1)$ at Node 1 for implementing its message selection;
- for $1 < j < \mathsf{h}$, a local randomness rate of $H(X_j|Z_{j-1})$ and $I(X_{j+1}; Z_j|X_j)$ for generating the action at Node j, and for implementing the selection of message intended for Node $j+1$, respectively; and
- a local randomness rate of $H(X_\mathsf{h}|Z_{\mathsf{h}-1})$ for generating the action at Node h.

Thus, it is possible to build a strong coordination code with:

$$\mathsf{R}_c \triangleq 0,$$
$$\rho_j \triangleq \begin{cases} I(X_2; Z_1|X_1), & j=1 \\ I(X_{j+1}; Z_j|X_j) + H(X_j|Z_{j-1}), & 1<j<\mathsf{h}, \\ H(X_\mathsf{h}|Z_{\mathsf{h}-1}), & j=\mathsf{h} \end{cases}$$
$$\mathsf{R}_j \triangleq I(X_j, X_{j+1}; Z_j), \qquad 1 \le j < \mathsf{h}.$$

Since the rate-transfer from ρ_j to ρ_{j-1} is allowable by communicating local randomness from Node $j-1$ to Node j (Lemma 1), we see that the following randomness and communication rates also suffice to achieve strong coordination.

$$\mathsf{R}_c \triangleq 0,$$
$$\rho_j \triangleq \begin{cases} I(X_2; Z_1|X_1) + \delta_2, & j=1 \\ I(X_{j+1}; Z_j|X_j) + H(X_j|Z_{j-1}) - \delta_j + \delta_{j+1}, & 1<j<\mathsf{h} \\ H(X_\mathsf{h}|Z_{\mathsf{h}-1}) - \delta_j, & j=\mathsf{h} \end{cases}$$
$$\mathsf{R}_j \triangleq I(X_j, X_{j+1}; Z_j) + \delta_j, \qquad 1 \le j < \mathsf{h},$$

where the rate-transfer variables δ_i's are non-negative. As before, an application of Fourier-Motzkin elimination to dispose of the rate-transfer variables yields the following rate region. For each $1 \le i < j \le \mathsf{h}$,

$$\mathsf{R}_i + \sum_{k=i+1}^{j} \rho_k \ge \begin{pmatrix} I(X_i; Z_i) + I(X_{j+1}; Z_j|X_j) \\ + H(\{X_k\}_{k=i+1}^{j}|X_i) \end{pmatrix}, \tag{20}$$

$$\sum_{k=1}^{j} \rho_k \ge I(\{X_k\}_{k=2}^{j+1}; Z_j, X_1, \ldots, X_j|X_1), \tag{21}$$

where we define $X_{\mathsf{h}+1}$ and Z_h to be constant random variables. In this case of no common randomness, we can show that the above region and the coding scheme are indeed optimal.

Theorem 1: Strong coordination is achievable in the absence of common randomness if and only if there exist auxiliary random variables $Z_1, \ldots, Z_{\mathsf{h}-1}$ satisfying (18), (19) such that the communication and local randomness rates meet (20) and (21) for $1 \le i < j \le \mathsf{h}$.

Proof: Suppose that there exists a scheme requiring a local randomness rate of $\rho_i + \varepsilon$ at Node i and a communication rate of $\mathsf{R}_i + \varepsilon$ from Node i to Node $i+1$ such that the joint p.m.f. of the actions satisfies:

$$\mathbb{V}(Q_{\hat{X}_1^n \cdots \hat{X}_\mathsf{h}^n}, \mathsf{Q}_{X_1 \cdots X_\mathsf{h}}^{\otimes n}) \le \varepsilon, \tag{22}$$

where for notational ease, we let $\hat{X}_1^n = X_1^n$ to be the action specified at Node 1. Since (22) holds, $(\hat{X}_1^n, \hat{X}_2^n, \ldots, \hat{X}_\mathsf{h}^n)$ are nearly i.i.d, and from [2, Sec. V. A], we infer that for any $S \subseteq \{1, \ldots, \mathsf{h}\}$ and $i \in \{1, \ldots, \mathsf{h}\}$,

$$H(\{\hat{X}_j^n\}_{j\in S}) \ge \sum_{k=1}^{n} H(\{\hat{X}_{j,k}\}_{j\in S}) - n\delta'_{n,\varepsilon} \tag{23}$$

$$H(\{\hat{X}_j^n\}_{j\in S}|\hat{X}_i^n) \ge \sum_{k=1}^{n} H(\{\hat{X}_{j,k}\}_{j\in S}|\hat{X}_{i,k}) - n\delta'_{n,\varepsilon} \tag{24}$$

for some $\delta'_{n,\varepsilon} \to 0$ as $\varepsilon \to 0$. Then, for any $1 \le j < \mathsf{h}$, the arguments in (25)-(34) hold, where:

- (a) follows because I_1 is a function of X_1^n and M_{L_1} and for $i = 2, \ldots, j$, I_j is a function of I_{j-1} and M_{L_j};
- (b) follows from two steps: 1) introducing action variables $\{\hat{X}_{\ell,k}\}_{\ell=2}^{j}$, since they are functions of $\{\mathsf{I}_\ell\}_{l=1}^{j-1}$ and $\{\mathsf{M}_{L_\ell}\}_{\ell=2}^{j}$, and then 2) by dropping $\{\mathsf{M}_{L_\ell}\}_{\ell=2}^{j}$;
- (c) follows from (24), since the actions are nearly i.i.d.;
- (d) by defining $Y_j \triangleq I_j$;
- (e) by introducing a time-sharing variable U that is uniform over $\{1, \ldots, n\}$; and
- (f) from by setting $\bar{Y}_j \triangleq (Y_j, U)$ and

$$\tilde{\delta}_{n,\varepsilon} \triangleq \delta'_{n,\varepsilon} + I(\hat{X}_{2,U}, \ldots, \hat{X}_{\mathsf{h},U}; U|X_{1,U}), \tag{35}$$

which vanishes as $\varepsilon \to 0$, since (22) ensures that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} Q_{\hat{X}_{1,i},\ldots,\hat{X}_{\mathsf{h},i}} - \mathsf{Q}_{X_1,\ldots,X_\mathsf{h}} \right\|_1 \le \varepsilon, \tag{36}$$

which then implies that $I(\hat{X}_{1,U}, \ldots, \hat{X}_{\mathsf{h},U}; U) \stackrel{\varepsilon\to 0}{\longrightarrow} 0$.

This completes the argument for (21) when $j < \mathsf{h}$. The argument for $j = \mathsf{h}$ follows from the above by setting I_h, Y_h as constant RVs, and by setting $\bar{Y}_\mathsf{h} \triangleq U$.

To show (20), we proceed as follows. Let $1 \le i < j < \mathsf{h}$. Then, (37)-(44) also hold, where (g) follows from a time-

$$
\begin{aligned}
n(\rho_1+\cdots+\rho_j)+jn\varepsilon &\geq H(\mathsf{M}_{L_1},\ldots,\mathsf{M}_{L_j}) \geq H(\mathsf{M}_{L_1},\ldots,\mathsf{M}_{L_j}|\hat{X}_1^n) && (25)\\
&\geq I(\hat{X}_2^n,\ldots,\hat{X}_{j+1}^n;\mathsf{M}_{L_1},\ldots,\mathsf{M}_{L_j}|\hat{X}_1^n) && (26)\\
&\overset{(a)}{\geq} I(\hat{X}_2^n,\ldots,\hat{X}_{j+1}^n;\mathsf{I}_1,\ldots,\mathsf{I}_j,\mathsf{M}_{L_1},\ldots,\mathsf{M}_{L_j}|\hat{X}_1^n) && (27)\\
&= \sum_{k=1}^n I(\hat{X}_{2,k},\ldots,\hat{X}_{j+1,k};\{\mathsf{I}_\ell\}_{\ell=1}^j,\{\mathsf{M}_{L_\ell}\}_{\ell=1}^j \mid \hat{X}_1^n,\{\hat{X}_\ell^{k-1}\}_{\ell=2}^j) && (28)\\
&\overset{(b)}{\geq} \sum_{k=1}^n I(\hat{X}_{2,k},\ldots,\hat{X}_{j+1,k};\{\mathsf{I}_\ell\}_{\ell=1}^j,\{\hat{X}_{\ell,k}\}_{\ell=1}^j \mid \hat{X}_1^n,\{\hat{X}_\ell^{k-1}\}_{\ell=2}^j) && (29)\\
&\overset{(c)}{\geq} \sum_{k=1}^n I(\hat{X}_{2,k},\ldots,\hat{X}_{j+1,k};\{\mathsf{I}_\ell\}_{\ell=1}^j,\{\hat{X}_{\ell,k}\}_{\ell=1}^j,\hat{X}_1^n,\{\hat{X}_\ell^{k-1}\}_{\ell=2}^{j+1}|\hat{X}_{1,k}) - n\delta'_{n,\varepsilon} && (30)\\
&\geq \sum_{k=1}^n I(\hat{X}_{2,k},\ldots,\hat{X}_{j+1,k};\mathsf{I}_j,\{\hat{X}_{\ell,k}\}_{\ell=1}^j|\hat{X}_{1,k}) - n\delta'_{n,\varepsilon} && (31)\\
&\overset{(d)}{=} \sum_{k=1}^n I(\hat{X}_{2,k},\ldots,\hat{X}_{j+1,k};Y_j,\{\hat{X}_{\ell,k}\}_{\ell=1}^j|\hat{X}_{1,k}) - n\delta'_{n,\varepsilon} && (32)\\
&\overset{(e)}{=} nI(\hat{X}_{2,U},\ldots,\hat{X}_{j+1,U};Y_j,\{\hat{X}_{\ell,U}\}_{\ell=1}^j|\hat{X}_{1,U},U) - n\delta'_{n,\varepsilon} && (33)\\
&\overset{(f)}{\geq} nI(\hat{X}_{2,U},\ldots,\hat{X}_{j+1,U};\bar{Y}_j,\{\hat{X}_{\ell,U}\}_{\ell=1}^j|\hat{X}_{1,U}) - n\tilde{\delta}_{n,\varepsilon}, && (34)
\end{aligned}
$$

$$
\begin{aligned}
n\mathsf{R}_i+n(\rho_{i+1}+\cdots+\rho_j)+n(j-i+1) &\geq H(\mathsf{I}_i,\{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^j) = H(\{\mathsf{I}_\ell\}_{\ell=i}^j,\{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^j) && (37)\\
&\geq I(\{\hat{X}_\ell^n\}_{\ell=i}^{j+1};\{\mathsf{I}_\ell\}_{\ell=i}^j,\{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^j) && (38)\\
&\geq I(\hat{X}_i^n;\mathsf{I}_i) + I(\{\hat{X}_\ell^n\}_{\ell=i+1}^{j+1};\{\mathsf{I}_\ell\}_{\ell=i}^j,\{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^j \mid \hat{X}_i^n) && (39)\\
&= I(\hat{X}_i^n;\mathsf{I}_i) + I(\{\hat{X}_\ell^n\}_{\ell=i+1}^{j+1};\{\mathsf{I}_\ell\}_{\ell=i}^j,\{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^j,\{\hat{X}_\ell^n\}_{\ell=i+1}^j \mid \hat{X}_i^n) && (40)\\
&\geq I(\hat{X}_i^n;\mathsf{I}_i) + I(\{\hat{X}_\ell^n\}_{\ell=i+1}^{j+1};\mathsf{I}_j,\{\hat{X}_\ell^n\}_{\ell=i+1}^j \mid \hat{X}_i^n) && (41)\\
&= \sum_{k=1}^n \Big(I(\hat{X}_{i,k};\mathsf{I}_i|\hat{X}_i^{k-1}) + I(\{\hat{X}_{\ell,k}\}_{\ell=i+1}^{j+1};\mathsf{I}_j,\{\hat{X}_\ell^n\}_{\ell=i+1}^j \mid \hat{X}_i^n,\{\hat{X}_\ell^{k-1}\}_{\ell=i+1}^{j+1})\Big)\\
&\overset{(23),(24)}{\geq} \sum_{k=1}^n \Big(I(\hat{X}_{i,k};\mathsf{I}_i) + I(\{\hat{X}_{\ell,k}\}_{\ell=i+1}^{j+1};\mathsf{I}_j,\{\hat{X}_{\ell,k}\}_{\ell=i+1}^j \mid \hat{X}_{i,k})\Big) - 2\delta'_{n,\varepsilon} && (42)\\
&\overset{(g)}{\geq} n\big(I(\hat{X}_{i,U};\bar{Y}_i) + I(\{\hat{X}_{\ell,U}\}_{\ell=i+1}^{j+1};\bar{Y}_j,\{\hat{X}_{\ell,U}\}_{\ell=i+1}^j \mid \hat{X}_{i,U})\big) - 2\tilde{\delta}_{n,\varepsilon}, && (43)\\
&\overset{(h)}{\geq} n\big(I(\hat{X}_{i,U};\bar{Y}_i) + H(\{\hat{X}_{\ell,U}\}_{\ell=i+1}^j|\hat{X}_{i,U}) + I(\hat{X}_{j+1,U};\bar{Y}_j \mid \hat{X}_{j,U})\big) - \bar{\delta}_{n,\varepsilon}, && (44)
\end{aligned}
$$

---

sharing argument and the auxiliary RV assignment in step (f) previously, and (h) holds by setting

$$\bar{\delta}_{n,\varepsilon} = 2\tilde{\delta}_{n,\varepsilon} + I(\hat{X}_{j+1,U};\{\hat{X}_{k,U}\}_{k=i}^{j-1}|\hat{X}_{j,U}), \quad (45)$$

which vanishes as $\varepsilon \to 0$. Further, as before, the proof for $j=\mathsf{h}$ follows by setting I_h, Y_h as constants and by setting $\bar{Y}_\mathsf{h} \triangleq U$. We are left to prove that the Markov condition in (19) is met by the actions and the auxiliary RVs. Note that as per the definitions of the auxiliary RVs, we do not satisfy the long chain in (19). The joint p.m.f. decomposes as follows.

$$Q_{\bar{Y}_1} Q_{X_{1,U}|\bar{Y}_1} \prod_{j=1}^{\mathsf{h}-1} Q_{\hat{X}_{j+1,U},\bar{Y}_{j+1}|\bar{Y}_j}. \quad (46)$$

In other words, we have $X_{j,U} — \bar{Y}_j — X_{j+1,U}$ for all $j = 1,\ldots,\mathsf{h}-1$, but not $\bar{Y}_j — X_{j+1,U} — \bar{Y}_{j+1}$. However, the information functional terms in (20) and (21) contain only one auxiliary RV each. Hence, it is possible to define a new set of auxiliary RVs that satisfies the long chain in the claim, and preserves the information functionals. To do so, define RVs $\{\tilde{X}_k\}_{k=1}^{\mathsf{h}}$ and $\{Z_j\}_{j=1}^{\mathsf{h}-1}$, such that their joint p.m.f. is given by

$$
\begin{aligned}
&Q_{\tilde{X}_1,\ldots,\tilde{X}_\mathsf{h}} \triangleq Q_{\hat{X}_{1,U},\ldots,\hat{X}_{\mathsf{h},U}} && (47)\\
&Q_{Z_1,\ldots,Z_{\mathsf{h}-1}|\tilde{X}_1,\ldots,\tilde{X}_\mathsf{h}}(z_1,\ldots,z_{\mathsf{h}-1}|x_1,\ldots,x_h)\\
&\qquad \triangleq \prod_{j=1}^{\mathsf{h}-1} Q_{\bar{Y}_j|\hat{X}_{j,U}\hat{X}_{j+1,U}}(z_j|x_j,x_{j+1}). && (48)
\end{aligned}
$$

Then, we have $\tilde{X}_1 — Z_1 — \tilde{X}_2 — \cdots — Z_{\mathsf{h}-1} — \tilde{X}_\mathsf{h}$. Further, for $1 \leq i < j \leq \mathsf{h}$,

$$
\begin{aligned}
H(\tilde{X}_{i+1},\ldots,\tilde{X}_j|\tilde{X}_i) &= H(\hat{X}_{i+1,U},\ldots,\hat{X}_{j,U}|\hat{X}_{i,U}), && (49)\\
I(\tilde{X}_{j+1};Z_j|\tilde{X}_j) &= I(\hat{X}_{j+1,U};\bar{Y}_j|\hat{X}_{j,U}), && (50)\\
I(\tilde{X}_j;Z_j) &= I(\hat{X}_{j,U};\bar{Y}_j). && (51)
\end{aligned}
$$

The proof is then complete by first limiting the size of the auxiliary RVs $\{Z_j\}_{j=1}^{\mathsf{h}-1}$, and then limiting $\varepsilon \to 0$, which ensures that the infinitesimals in (34) and (44) vanish, and $Q_{\tilde{X}_1,\ldots,\tilde{X}_\mathsf{h}} \to \mathsf{Q}_{X_1,\ldots,X_\mathsf{h}}$. ■

C. Neither Sufficiently Large Nor Zero Common Randomness

The optimal coding strategies for the two extremal cases of common randomness are different from one another. In addition, each strategy is sub-optimal for the other extreme case of common randomness. This can be seen as follows.

- Suppose that we use the juxtaposition of $\mathsf{h}-1$ one-hop schemes when $\mathsf{R}_c > H(X_2, \ldots, X_\mathsf{h}|X_1)$. Since the trade-offs for the one-hop scheme are governed by (10)-(14), the rates on hops connecting Node i and Node $i+1$ in this scheme can be no smaller than $I(X_i; X_{i+1})$, which can be strictly greater than $I(X_1; X_{i+1})$. Thus, enforcing the conditional independence of next-hop and previous-hop messages at an intermediate node given the node's action *is* sub-optimal, even when the joint p.m.f. of the actions satisfies the Markov chain in (1).
- Consider a two-hop network where $X_2 = (V_1, V_2)$, $X_1 = V_1$, $X_3 = V_2$, and $H(V_2) < H(V_1)$. Now, suppose that we use a code with the nested structure when common randomness is absent. Since the nested structure requires Node 1 to initiate the selection of all codewords (corresponding to the auxiliary random variables), no matter what the choices for auxiliary random variables are, and no matter how much local randomness each node possesses, the communication rates resulting from the nested scheme *must* satisfy $\mathsf{R}_1 \geq \mathsf{R}_2 \geq \cdots \geq \mathsf{R}_{\mathsf{h}-1}$. However, the juxtaposition scheme with $Z_1 = V_1$ and $Z_2 = V_2$ would ensure that $\mathsf{R}_1 < \mathsf{R}_2$ is achievable when sufficient local randomness is available at each node. Hence, the nested structure is sub-optimal when $\mathsf{R}_c = 0$.

To ensure optimality at both extremes of common randomness and good performance in between, we need to intertwine the two schemes so that the joint scheme can be optimal at both extremes of common randomness. To see how this could be possible, allow us to illustrate with a two-hop network. Since we need both Node 1 and Node 2 to have the ability to generate and select messages, we could consider a three message/auxiliary random setup, where message $M_{i,j}$, $1 \leq i < j \leq 3$ originates at Node i and forwarded until it reaches at Node j, and is used to select to a codeword from a codebook corresponding to an auxiliary random variable $A_{i,j}$. Clearly, by the way the messages are set up, the codebook for $A_{1,2}$ can be constructed conditional on that of $A_{1,3}$. Similarly, the codebook for $A_{2,3}$ can be constructed conditional on that of $A_{1,3}$. So, in effect, it would be possible to construct a codebook structure with the following joint p.m.f. decomposition.

$$Q_{A_{1,3}} Q_{A_{1,2}|A_{1,3}} Q_{A_{2,3}|A_{1,3}} Q_{X_1|A_{1,2},A_{1,3}} \times Q_{X_2|A_{1,2},A_{2,3}A_{1,3}} Q_{X_3|A_{2,3}A_{1,3}}$$

Using this setup with only auxiliary RVs $A_{1,2}$ and $A_{1,3}$ yields the nested codebook structure, whereas using only $A_{1,2}$ and $A_{2,3}$ yields the juxtaposition codebook structure. However, if $A_{1,3}$ is a constant, RVs $A_{1,2}$ and $A_{2,3}$ are independent. Hence, only those juxtaposition codebook structures corresponding to independent auxiliary RVs can be derived from this setup. The incorporation of juxtaposition codebooks designed using dependent auxiliary RV choices, and the characterization of the achievable trade-offs among communication rates, and local and common randomness rates using this joint setup are ongoing research tasks.

Appendix A
A Technical Result

Lemma 2: Consider the a channel resolvability code for generating $Y \sim Q_Y$ via the channel $Q_{Y|B}$. Let $R > I(B;Y)$ and $K \sim \mathsf{unif}([\![1, 2^{nR}]\!])$. Suppose that the codebook contains 2^{nR} B-codewords constructed randomly using Q_B. Let $\hat{Y}^n$ denote the output when $B^n(K)$ is fed into the channel $Q_{Y|B}$. Let $Q_{K,\hat{Y}^n}$ denote the p.m.f. induced by the code. Let $S \sim \mathsf{unif}([\![1, 2^{n\varrho}]\!])$ for $\varrho > R - I(B;Y)$. Then, there exists a function $\Lambda_\mathcal{C} : \mathcal{Y}^n \times [\![1, 2^{n\varrho}]\!] \to [\![1, 2^{nR}]\!]$ (that depends on the instance of the realized codebook) such that:

$$\lim_{n\to\infty} \mathbb{E}_\mathcal{C}\, \mathbb{V}(Q_{\Lambda_\mathcal{C}(\hat{Y}^n,S),\hat{Y}^n}, Q_{K,\hat{Y}^n}) = 0, \tag{52}$$

where $\mathbb{E}_\mathcal{C}$ denotes the expectation over all random codebooks.

Proof: An outline of the proof is as follows. Since the relationship between the selected codeword $B^n(K)$ and $\hat{Y}^n$ are governed by the channel law, with high probability, they will be jointly typical. By enlisting all codewords jointly typical with a realized $\hat{Y}^n$, it can be shown that the list size grows exponentially in n, with the exponent equalling $R - I(B;Y)$. We can then argue that S can be used to select one codeword from this list depending on their *a posteriori* probabilities. Since the first-order terms in the exponents of these *a posteriori* probabilities are identical, we can argue the rate of S required to enable the selection coincides with the exponent of the list size. ■

References

[1] P. Cuff, H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.

[2] P. Cuff, "Communication requirements for generating correlated random variables," in *2008 IEEE International Symposium on Information Theory*, July 2008, pp. 1393–1397.

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

[4] M. H. Yassaee, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," in *2012 IEEE International Symposium on Information Theory*, July 2012, pp. 3053–3057.

[5] M. Bloch and J. Kliewer, "Strong coordination over a three-terminal relay network," in *2014 IEEE Information Theory Workshop*, Nov. 2014, pp. 646–650.

[6] F. Haddadpour, M. H. Yassaee, A. Gohari, and M. R. Aref, "Coordination via a relay," in *2012 IEEE International Symposium on Information Theory*, July 2012, pp. 3048–3052.

[7] B. N. Vellambi, J. Kliewer, and M. Bloch, "Strong coordination over multihop line networks," *2015 IEEE Information Theory Workshop*, pp. 192–196, 2015.

[8] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.

[9] M. Bloch and J. Kliewer, "Strong coordination over a line network," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 2319–2323.

[10] T. S. Han, *Information-Spectrum Methods in Information Theory*, 1st ed. Springer, 2003.

[11] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.