Strong Coordination Over Multi-Hop Line Networks Using Channel Resolvability Codebooks

Badri N. Vellambi, *Senior Member, IEEE*, Joerg Kliewer, *Senior Member, IEEE*, and Matthieu R. Bloch, *Senior Member, IEEE*

***Abstract*— We analyze the problem of strong coordination over a multi-hop line network in which the node initiating the coordination is a terminal network node. We assume that each node has access to a certain amount of randomness that is local to the node, and that the nodes also have shared common randomness, which are used together with explicit hop-by-hop communication to achieve information-theoretic strong coordination. We derive the trade-offs among the required rates of communication on the network links, the rates of local randomness available at network nodes, and the rate of common randomness to realize strong coordination. We present an achievable coding scheme built using multiple layers of channel resolvability codes, and establish several settings in which this scheme offers the best possible trade-offs among network resources.**

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

I. Introduction

DECENTRALIZED control is an essential feature in almost all large-scale networks such as the Internet, surveillance systems, sensor networks, traffic and power grid networks. Control in such networks is achieved in a distributed fashion by coordinating various actions and response signals of interest. Communication between various parts of the network serves as an effective means to establish coordination. Coordination can generally be enabled via the following two modes of communication.

- Coordination of a system through *explicit* communication where communication signals extrinsic to the control and coordination of the system are sent from one part of system to another to specifically coordinate/control the system [3]. In this case, the signals for achieving coordination are conveyed additional to the signals used for communication.
- Coordination of a system through *implicit* communication where the signals inherently sent from one part of the system to another during its natural operation are also used to coordinate/control the system [4]–[6]. For example, consider robot soccer, where coordination with other robots as well as the playing of the game is managed simultaneously by the actions/moves each robot undertakes [7]. In this case, coordination is achieved implicitly by pre-sharing a strategy/codebook; each robot can use this pre-shared information in real time to guage what future actions are going to be taken by other robots and consequently align its own actions accordingly. Note that no (explicit) communication signal other than the actions are employed here.

The problem of coordination through (either modes of) communication is very closely tied to a slew of information-theoretic problems, including intrinsic randomness, channel resolvability, and random number generation [8]–[10], channel simulation and synthesis [10]–[14], and distributed random variable generation [15], [16]. Consequently, many ideas for the design of codes for these problems heavily feature in the design of coordination codes. Two notions of coordination have been studied in the literature:

- *empirical* coordination, where the aim is to closely match the empirical distribution of the actions/signals at network nodes with a prescribed target histogram/probability mass function; and
- *strong* coordination, where the aim is the generation of actions at various network nodes that are collectively required to resemble the output of a jointly correlated source. In this setting, by observing the actions of the network nodes, a statistician cannot determine (with significant confidence) as to whether the actions were generated by a jointly correlated source, or from a (strong) coordination scheme.

A compendious introduction to the fundamental limits and optimal coding strategies for empirical and strong coordination in many canonical networks (e.g., one-hop, broadcast, relay networks) can be found in [3]. However, the majority of the networks considered therein comprised of two or three terminals. The limits and means of the empirical coordination of a discrete memoryless source with a receiver connected by a point-to-point noisy channel was explored in [17]–[19].

Manuscript received April 8, 2016; revised May 31, 2017; accepted September 13, 2017. Date of publication November 1, 2017; date of current version January 18, 2018. This work was supported by NSF under Grant CCF-1440014, Grant CCF-1439465, and Grant CCF-1320304. This paper was published in part at the 2015 IEEE Information Theory Workshop [1] and the 50th Annual Conference on Information Sciences and Systems [2].

B. N. Vellambi is with the Research School of Computer Science, Australian National University, Canberra, ACT 2601, Australia (e-mail: badri.n.vellambi@ieee.org).

J. Kliewer is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: jkliewer@njit.edu).

M. R. Bloch is with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: matthieu.bloch@ece.gatech.edu).

Communicated by H. H. Permuter, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2017.2768529

The effects of causality of encoding and channel feedback were investigated in [19], and the benefits of channel state information available acausally at the encoder was explored in [17], [18].

Coordination over the more general three-terminal setting in the presence of a relay was considered in [20]–[22]. Inner and outer bounds of the required rates of communication for coordination were derived in [20] and [21]. It must be noted that [20] focuses on strong coordination and only one-way communication, whereas [21] focuses on strong coordination and two-way communication with actions required only at the end terminals (and not at the relay). Inner and outer bounds for the required rates of communication for coordination over a noiseless triangular network with relay was studied in [22]. The fundamental limits and optimal schemes for empirical coordination with implicit communication over multiple-access channels with state were explored in [23] and [24].

In this work, we quantify the network resources required for achieving *strong* correlation in multi-hop line networks. By network resources, we mean three quantities required for establishing strong coordination: (a) the rates of hop-by-hop communication between network nodes; (b) the rate of randomness locally available at each node; and (c) the required rate of common randomness shared by all network nodes. The problem considered here is closely related to strong coordination problems investigated in [20], [25], and [26]. As will be clear in due course, a solution to strong coordination code is devised by first solving an allied problem of distributed generation of actions (correlated random variables), and then appropriately inverting the operation at one of the nodes. Hence, the strong coordination problem and the allied action-generation problem are closely related to channel synthesis and the problem of generating correlated random variables considered in [13]–[15], and our ideas and results parallel those in these works.

In [26], the strong coordination rate region for two- and multi-hop line networks is characterized under the secrecy constraint that an eavesdropper does not additionally learn anything about the joint statistics of the actions even when they observe the communication on the network links. This work does not consider this additional secrecy requirement. It presents a general achievability scheme that is optimal (i.e., one can, in theory, derive codes using our general achievability scheme for any point in the underlying capacity region) in the following cases:

- when there is sufficient common randomness shared by all the nodes in the network;
- when the intermediate nodes operate in a *functional* regime in which intermediate-node processing is a deterministic function of the incoming messages and the common randomness alone; and
- when common randomness is absent, and the actions form a Markov chain that is aligned with the network topology.

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

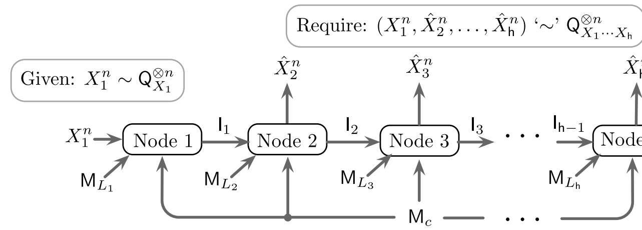

Fig. 1. The strong coordination problem setup.

Finally, this work is concluded in Section VI followed by appendices containing ancillary results and the proofs of relevant claims made in Section IV.

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 refer to node indices. In case of double subscripts, the first indicates the node index, and the next indicates the component (i.e., time) index. Given a finite set S, $\mathsf{unif}(S)$ denotes the uniform probability mass function (pmf) on the set S. Given a pmf p_X, $\mathsf{supp}(p_X)$ indicates the support of p_X, and $T_\varepsilon^n[p_X]$ denotes the set of all ε-letter typical sequences of length n [27]. Given two pmfs p and q on the same alphabet $\mathcal{X}$, with $\mathsf{supp}(p) \subseteq \mathsf{supp}(q)$, $\mathsf{D}_{\mathsf{KL}}(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$. Given an event E, $\mathbb{P}(E)$ denotes the probability of occurrence of the event E. The expectation operator is denoted by $\mathbb{E}[\cdot]$. Lastly, $p_{X_1\cdots X_k}^{\otimes n}$ denotes the pmf of n i.i.d. random k-tuples, with each k-tuple correlated according to pmf $p_{X_1\cdots X_k}$.

III. Problem Definition

Consider the problem of the control of traffic signaling systems over a network of busy roads. The network of signaling systems have to route vehicular traffic smoothly, efficiently, and nearly instantanously. To do so, the signaling systems must be coordinated with one another taking into consideration the periodic trends in traffic patterns. The signaling systems have to agree on a pre-shared strategy to effect coordination, which can potentially be randomized to incorporate random deviations in the underlying traffic patterns. The line coordination problem studied in this work is loosely based on such an application.

As a starting point, this work focuses on a multi-hop line network consisting of h nodes (Nodes $1, \ldots, \mathsf{h}$) as in Fig. 1. The communication between signaling systems is modeled as noiseless bit pipes that connect Node i with Node $i+1$, $1 \le i < \mathsf{h}$. The overall aim is to enable strong coordination of the signaling pattern of the h nodes according to a design joint pmf $\mathsf{Q}_{X_1\cdots X_{\mathsf{h}}}$ that is assumed to be given. It is assumed that Node 1 is specified an action sequence $\{X_{1,i}\}_{i\in\mathbb{N}}$ modeled by an i.i.d process Q_{X_1} distributed over a finite set $\mathcal{X}_1$. Nodes are assumed to possess local randomness as well as common randomness shared by all h nodes for use in establishing strong coordination. The amount of randomness required serves as a measure of how much complexity is required to realize strong coordination according to the specified pattern 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 $\hat{X}_i^n \in \mathcal{X}_i^n$ at Nodes i, $i > 2$ satisfying the following condition: the joint pmf of actions $(X_1^n, \hat{X}_2^n, \ldots, \hat{X}_{\mathsf{h}}^n)$ and $\mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}^{\otimes n}$ (i.e., the joint pmf of n symbols output by a given discrete memoryless source corresponding to the design joint pmf $\mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}$) are nearly indistinguishable under the variational distance metric. The overall aim is to characterize the required rates of communication messages, and local and common randomness to achieve such strong coordination. The following definitions are now in order.

Definition 1: Given joint pmf $\mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}$ and $\varepsilon > 0$, a strong coordination ε-code of length n at rate tuple $\mathbf{R} \triangleq (\mathsf{R}_c, \mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1}, \rho_1, \ldots, \rho_{\mathsf{h}}) \in \mathbb{R}^{+2\mathsf{h}}$ 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 $\psi_1, \ldots, \psi_{\mathsf{h}-1}$, and $\mathsf{h}-1$ action-generating functions $\phi_2, \ldots, \phi_{\mathsf{h}}$ such that:

- Randomness constraints:

$$\text{[Common]}\quad \mathsf{M}_c \sim \mathsf{unif}([\![1, 2^{n\mathsf{R}_c}]\!]),$$
$$\text{[Local]}\quad \mathsf{M}_{L_i} \sim \mathsf{unif}([\![1, 2^{n\rho_i}]\!]), \quad 1 \le i \le \mathsf{h}.$$

- Message-generation and action-generation constraints:

$$\mathsf{I}_1 \triangleq \psi_1(\mathsf{M}_{L_1}, X_1^n, \mathsf{M}_c) \in [\![1, 2^{n\mathsf{R}_1}]\!],$$
$$\mathsf{I}_j \triangleq \psi_j(\mathsf{M}_{L_j}, \mathsf{I}_{j-1}, \mathsf{M}_c) \in [\![1, 2^{n\mathsf{R}_j}]\!], \quad 2 \le j < \mathsf{h},$$
$$\hat{X}_j^n \triangleq \phi_j(\mathsf{M}_{L_j}, \mathsf{I}_{j-1}, \mathsf{M}_c), \quad 2 \le j \le \mathsf{h}.$$

- Strong coordination constraint:

$$\left\| \mathsf{Q}_{X_1}^{\otimes n} Q_{\hat{X}_2^n \cdots \hat{X}_{\mathsf{h}}^n | X_1^n} - \mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}^{\otimes n} \right\| \le \varepsilon, \tag{1}$$

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

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

Let us suppose that the design joint pmf $\mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}$ is such that $H(X_2, \ldots, X_{\mathsf{h}} | X_1) = 0$, i.e., $X_2, \ldots, X_{\mathsf{h}}$ are (deterministic) functions of X_1. In this simplistic case, there is no need for local or common randomness, and the strong coordination problem becomes purely a communication problem with the following rate region.

Remark 1: If $H(X_2, \ldots, X_{\mathsf{h}} | X_1) = 0$, then strong coordination is achievable at $(\mathsf{R}_c, \mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1}, \rho_1, \ldots, \rho_h) \in \mathbb{R}^{+2\mathsf{h}}$ iff for each $\ell = 1, \ldots, \mathsf{h}$, $\mathsf{R}_\ell \ge H(X_{\ell+1}, \ldots, X_{\mathsf{h}})$. ■

So without loss of generality, we may assume for the rest of this work that the above remark does not apply to the given pmf $\mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}$. Before we proceed to the results, we introduce three possible modes of operation for intermediate nodes. These three modes, highlighted in Fig. 2, differ on how an intermediate node generates a message for its next-hop node, and are as follows.

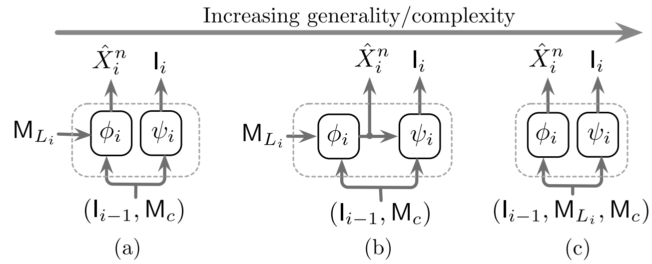

Fig. 2. Three possible encoder structures. (a) Functional. (b) Action-dependent. (c) Unrestricted.

- In the *functional* mode given in Fig. 2(a), the outgoing message at each intermediate node is generated from the incoming message and common randomness, i.e., the local randomness at an intermediate node is used only to generate the action corresponding to the node.
- In the *action-dependent* mode given in Fig. 2(b), the intermediate node uses the incoming message, and local and common randomness to generate the node's action. The outgoing message is then generated using the incoming message, common randomness, and the generated action. In this mode, local randomness at a node can affect the next-hop message only through the generated action; and finally,
- In the *unrestricted* mode given in Fig. 2(c), both the action and the next-hop message generated at an intermediate node depend on the incoming message, and local and common randomness.

In theory, the set of rate vectors achievable using the unrestricted mode is a superset of those achievable using the action-dependent mode, which is in turn a superset of those achievable by the functional mode. Further, these inclusions are, in general, strict (see the discussion at the end of Section IV). Before we present an achievable coding scheme for strong coordination, we present the following lemmas, which characterize the rate-transfer arguments allowable in the strong coordination problem at hand. These lemmas formalize the intuitive ideas described in their proofs.

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

A. For any $1 \le \ell \le \mathsf{h}$ and $0 \le \delta \le \rho_\ell$, strong coordination is also achievable under the unrestricted mode of operation using a common randomness rate $\mathsf{R}_c + \delta$, local randomness rates $(\rho_1, \ldots, \rho_{\ell-1}, \rho_\ell - \delta, \rho_{\ell+1}, \ldots, \rho_{\mathsf{h}})$ and communication rates $(\mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1})$; and
B. For any $1 < \ell \le \mathsf{h}$ and $0 \le \delta \le \rho_\ell$, strong coordination is also achievable under the unrestricted mode of operation using a common randomness rate R_c, local randomness rates $(\rho_1, \ldots, \rho_{\ell-1} + \delta, \rho_\ell - \delta, \rho_{\ell+1}, \ldots, \rho_{\mathsf{h}})$ and

communication rates $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{R}_3, \ldots, \mathsf{R}_{\ell-2}, \mathsf{R}_{\ell-1} + \delta, \mathsf{R}_\ell, \ldots, \mathsf{R}_{\mathsf{h}-1})$.

Proof: The first rate-transfer argument follows from the fact that a part of common randomness can always be used by precisely one node in the network to boost its local randomness. The second rate transfer follows from that fact that unused/excess local randomness at a node can be transmitted to the next-hop node to boost its local randomness in the unrestricted mode of operation. ■

Lemma 2: If strong coordination is achievable in the action-dependent (or functional) mode of operation using a common randomness rate R_c, local randomness rates $(\rho_1, \ldots, \rho_\mathsf{h})$ and communication rates $(\mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1})$, then:

A. For any $1 \leq \ell \leq \mathsf{h}$ and $0 \leq \delta \leq \rho_\ell$, strong coordination is also achievable under the action-dependent (or functional) mode of operation using a common randomness rate $\mathsf{R}_c + \delta$, local randomness rates $(\rho_1, \ldots, \rho_{\ell-1}, \rho_\ell - \delta, \rho_{\ell+1}, \ldots, \rho_\mathsf{h})$ and communication rates $(\mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1})$; and

B. For any $1 < \ell \leq \mathsf{h}$ and $0 \leq \delta \leq \rho_\ell$, strong coordination is also achievable under the action-dependent (or functional) mode of operation using a common randomness rate R_c, local randomness rates $(\rho_1 + \delta, \ldots, \rho_{\ell-1}, \rho_\ell - \delta, \rho_{\ell+1}, \ldots, \rho_\mathsf{h})$ and communication rates $(\mathsf{R}_1 + \delta, \mathsf{R}_2 + \delta, \ldots, \mathsf{R}_{\ell-1} + \delta, \mathsf{R}_\ell, \ldots, \mathsf{R}_{\mathsf{h}-1})$.

Proof: The first rate-transfer argument follows from the same argument as that for the first rate-transfer argument in Lemma 1. The second rate transfer follows from that fact that unused/excess local randomness at only the first node can be transmitted to any other node to boost its local randomness. Note that intermediate nodes cannot forward unused local randomness to downstream nodes when operating in the functional or action-dependent modes. ■

IV. Inner Bound: Achievability

In this first section of our results, we present an inner bound (achievability result) for the strong coordination rate region in two stages. First, we present the achievability for a two-hop network, and then present the achievability for the general multi-hop network. The derived inner bound holds for any finite-alphabet auxiliary RVs as long as they meet a certain decomposition, and a bound on their cardinalities is not generally known/derived. However, for the settings in Sec. V where inner and outer bounds match, the cardinalities of the auxiliary RVs can bounded using Carathéodory's theorem [28].

A. Inner Bound: An Achievable Scheme for $\mathsf{h} = 3$

The approach for the design of strong coordination codes combines ideas from channel resolvability codes [3], [10], [16] and channel synthesis [13]. In order to design a strong coordination code, we look at an allied problem of generating h actions $\hat{X}_1^n, \ldots, \hat{X}_\mathsf{h}^n$ from uniform and independent random variables (a.k.a. indices) such that the joint pmf of the generated actions $Q_{\hat{X}_1^n \cdots \hat{X}_\mathsf{h}^n}$ satisfies:

$$\left\| Q_{\hat{X}_1^n \cdots \hat{X}_\mathsf{h}^n} - \mathsf{Q}^{\otimes n}_{X_1 \cdots X_\mathsf{h}} \right\|_1 \leq \varepsilon. \tag{2}$$

The intuition behind the conversion of the strong coordination problem to the allied source generation problem can be traced to the following remark derived from a result of Han and Verdú [11]:

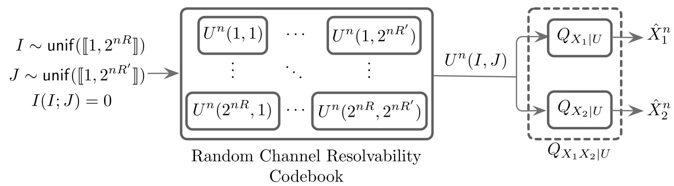

Fig. 3. The approach for the allied action-generation problem for $\mathsf{h} = 2$.

Remark 2: Given a joint pmf Q_{UY}, let $\mathcal{C} = \{C^n(1), \ldots, C^n(2^{nR})\}$ be a random channel codebook of 2^{nR} codewords with each codeword selected i.i.d. using Q_U, i.e., $C^n(i) \sim Q_U^{\otimes n}$, $i = 1, \ldots, 2^{nR}$. Let $P_{Y^n}(\cdot) = \sum_{i=1}^n \frac{1}{2^{nR}} Q_{Y|U}^{\otimes n}(\cdot|C^n(i))$ denote the pmf of the channel output when a codeword from this codebook is selected uniformly at random and transmitted over the DMC $Q_{Y|U}$. If $R > I(U; Y)$, then

$$\lim_{n\to\infty} \mathbb{E} \left\| P_{Y^n} - Q_Y^{\otimes n} \right\|_1 = 0, \tag{3}$$

where the expectation is over all codebook realizations. ■

Owing to the similarity of (2) and (3), the latter can indeed be viewed as the generation of actions of a single node via a codebook.

In [3], Cuff et al. used the above result to devise a two-node strong coordination scheme by analyzing the setup described in Fig. 3 that generates a pair of jointly correlated actions. In this approach,

- a channel resolvability code is constructed using an auxiliary random variable U and a pair of parallel channels $Q_{X_1X_2|U} = Q_{X_1|U} Q_{X_2|U}$ to incorporate the fact that the actions are generated at two different nodes;
- the channel resolvability codebook is arranged in a table and the codewords are selected by two independent uniformly distributed indices, one of which represents the message communicated, and the other corresponds to the shared common randomness;
- the code rates are chosen so that: (a) the joint pmf of the actions induced by the codebook can be made close to the i.i.d. distribution in the sense of (2), and (b) the index I and X_1^n are nearly independent, i.e., the mutual information between I and the generated action X_1^n can be made as close to zero as required; and lastly,
- once a realization of the channel resolvability codebook that meets the above requirements is fixed, a strong coordination scheme is generated by inverting the operation at Node 1 (i.e., instead of generating the action given I and J, Node 1 generates I given common randomness J and the action X_1^n specified by nature using the conditional pmf $P_{I|J,X_1^n}$ derived from the chosen codebook).

By analyzing the above two requirements, it was shown in [3] that a strong coordination scheme of communication rate R and the common randomness rate R_c exists iff there exists an auxiliary RV U such that $X_1 \leftrightarrow U \leftrightarrow X_2$, and

$$\mathsf{R} > I(U; X_1), \tag{4a}$$

$$\mathsf{R} + \mathsf{R}_c > I(U; X_1, X_2). \tag{4b}$$

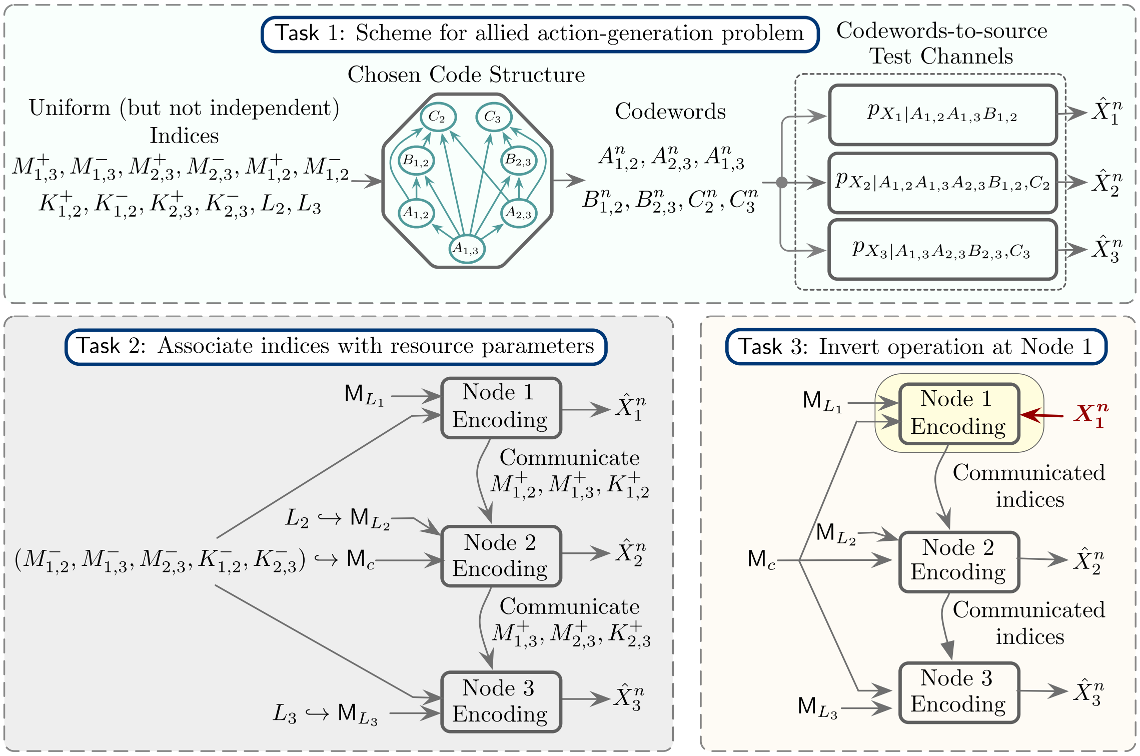

Fig. 4. Transforming a solution to the problem of generating h sources to a solution for the strong coordination problem.

We now proceed to present an achievable scheme for the $\mathsf{h} = 3$ case, and will serve as an illustrative example for the general multi-hop setting, especially since this case contains all the intricacies and difficulties of the general multi-hop setting. As in the $\mathsf{h} = 2$ case, the approach for $\mathsf{h} \geq 3$ devises a solution for the allied action-generation problem and then inverts the operation at Node 1. This approach has the following three distinct parts.

- Task 1: The first task is to devise a scheme to generate the $\mathsf{h} = 3$ actions, which is termed as the *allied action-generation problem*. To do so, first, a suitable structure of auxiliary RVs is chosen, and a codebook structure based on the chosen auxiliary RV structure is constructed. Appropriately distributed indices are used to select the codewords from the codebook, and suitable *test channels* are used to generate the h actions satisfying the above strong coordination requirement. Note that the auxiliary RV and codebook structure, and the corresponding test channels must be such that the actions are generated in a distributed fashion, incorporating the fact that, in an application, these actions may in fact be decided/undertaken by separate entities at different locations.
- Task 2: The next task is to assign roles to indices by partitioning the set of indices as common randomness, local randomness at each node, and messages to be communicated between nodes.
- Task 3: The last task is to then invert the operation at Node 1, which transforms the operation of generating the action at Node 1 to generating the messages intended for communication from the specified action and the shared randomness.

An illustration of the three steps for the three-node setting is given in Fig. 4. Note that much of the detail presented therein such as the exact structure and form of the auxiliary RVs, test channels and the assignments to the network resources (communication, local randomness and common randomness rates) will be elaborated in due course. In the figure, we notate $X \hookrightarrow Y$ to indicate that the RV X is associated as a part of the RV Y, and hence a part of Y is used to realize X. As of now, the figure is only intended to indicate the overall procedure. However, we will repeatedly refer back to this figure (and the tasks) as we develop various technical aspects of the strong coordination scheme.

It would be natural to extend the two-node approach of Fig. 3 to three nodes by the introduction of two auxiliary RVs U, V such that

$$X_1 \leftrightarrow U \leftrightarrow (X_2, V)$$
$$(X_1, U, X_2) \leftrightarrow V \leftrightarrow X_3,$$

where U plays the role of coordinating X_1 and X_2, and V is '*generated*' at Node 2 for coordinating with the actions of Node 1 and Node 2 with that of Node 3. The aims is to devise a scheme where only Node 2 knows and uses both U and V codewords. Simple at it may be, we do not have a way to realize such a scheme. The complication arises because we require all the indices used to select the codewords to be jointly uniform to exploit the channel resolvability result (Remark 2), which is possible if the random codebook for either U or V is constructed conditionally based on that of the other. Consequently, for a node to be able to identify the codeword of the variable with the conditional codebook, it must also know the codeword of the variable on which the conditional codebook is constructed, which then means that either Node 1 or Node 3 knows both (U and V) codewords, which contradicts our goal of devising a scheme where only Node 2 knows both U and V codewords.

One possible way to circumvent this issue is to introduce an auxiliary RV for each pair of nodes. For $\mathsf{h} = 3$ nodes, we introduce 3 auxiliary RVs $A_{1,3}, A_{1,2}, A_{2,3}$, where each $A_{i,j}$ is envisaged as the auxiliary RV whose codeword index (in part or in full) is conveyed from Node i to Node j, and hence can be used by Nodes $i, i+1, \dots, j$ to generate their actions. Since $A_{1,2}$ is known only to Nodes 1 and 2, $A_{2,3}$ is known only to Nodes 2 and 3, and $A_{1,3}$ is known to all nodes, the codebooks for $A_{1,2}$ and $A_{2,3}$ can be constructed conditionally on that of $A_{1,3}$; however, to ensure that the codewords for $A_{1,3}, A_{1,2}, A_{2,3}$ via jointly uniform random indices, we additionally impose the following chain.

$$A_{1,2} \leftrightarrow A_{1,3} \leftrightarrow A_{2,3}. \quad (5)$$

Note that this chain is imposed so as to exploit a suitable three-source extension of the channel resolvability result (Remark 2), where we use uniform random indices to select the three codewords, and then use the codewords and appropriate test channels to generate the three actions. If we take this approach, we can generate the three actions as the marginals of a joint pmf that takes the following form:

$$Q_{A_{1,3}} Q_{A_{1,2}|A_{1,3}} Q_{A_{2,3}|A_{1,3}} Q_{X_1|A_{1,2}A_{1,3}} \times Q_{X_2|A_{1,2}A_{1,3}A_{2,3}} Q_{X_3|A_{1,3}A_{2,3}}, \quad (6)$$

where the three parallel test channels that are used to generate the actions from the three auxiliary RVs are given by $Q_{X_1|A_{1,2}A_{1,3}}$, $Q_{X_2|A_{1,2}A_{1,3}A_{2,3}}$, and $Q_{X_3|A_{1,3}A_{2,3}}$.

However, we can devise an improved achievability scheme if for each $i = 1, 2$, we introduce an additional auxiliary RV $B_{i,i+1}$ that is generated by Node i and is intended for Node $i+1$. Let the joint pmf between the 5 auxiliary RVs and the 3 actions decompose as follows.

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

Note that for $i = 1, 2$, RVs $A_{i,i+1}$ and $B_{i,i+1}$ are both generated by Node i and intended for Node $i+1$; despite this similarity, the roles played by these auxiliary RVs are quite different. The following discussion highlights the difference between these auxiliary RVs.

Discussion 1: Suppose that we build a scheme with auxiliary RVs $A_{1,2}, A_{1,3}, A_{2,3}$, i.e, we set $B_{1,2}$ and $B_{2,3}$ as constant RVs. Then the joint pmf that we can emulate is given by (6). Since we have $A_{1,2} \leftrightarrow A_{1,3} \leftrightarrow A_{2,3}$ we see that the joint pmf can be rearranged as

$$Q_{A_{1,3}} Q_{X_1,A_{1,2}|A_{1,3}} Q_{X_2|A_{1,2}A_{1,3},A_{2,3}} Q_{A_{2,3},X_3|A_{1,3}}. \quad (7)$$

Therefore, when $B_{1,2}$ and $B_{2,3}$ are set as constant RVs, no matter what the choices for $A_{1,2}$ and $A_{2,3}$ are, X_1 and X_3 must be conditionally independent given $A_{1,3}$, which effectively is a restriction on the choice of $A_{1,3}$. Now, suppose that we build a scheme with auxiliary RVs $A_{1,3}, B_{1,2}, B_{2,3}$, i.e, we set $A_{1,2}$ and $A_{2,3}$ as constant RVs. Then the joint pmf that we can emulate is given by

$$Q_{A_{1,3},X_1,B_{1,2}} Q_{X_2|B_{1,2},A_{1,3}} Q_{B_{2,3}|X_2,A_{1,3}} Q_{X_3|B_{2,3},A_{1,3}}, \quad (8)$$

which does not necessarily imply the conditional independence of X_1 and X_3 given $A_{1,3}$. Hence, employing non-trivial $(B_{1,2}, B_{2,3})$ or $(A_{1,2}, A_{2,3})$ allows for different choices for $A_{1,3}$, which in turn could potentially translate into different resource requirements. ■

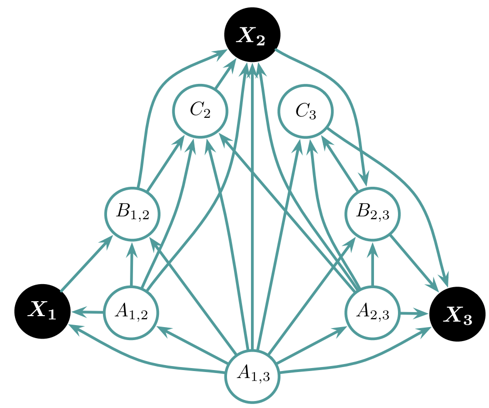

Fig. 5. An illustration of the structure of auxiliary random variables and actions when $\mathsf{h} = 3$.

Now, we finally supplement the suite of 5 auxiliary RVs with two more C_2 and C_3 which, unlike the A and the B RVs, are not communicated between nodes, but are local to Nodes 2 and 3, respectively. Their role is solely to enable us to quantify the amount of local randomness required at Nodes 2 and 3. The reader might find it helpful to know that our eventual choice will be $C_i = X_i$ for $i > 1$. This also explains the obvious absence of an auxiliary RV named C_1 since Node 1's action is specified by nature, and only actions at Nodes 2 and 3 are generated by the scheme. Finally, the joint pmf of the 7 auxiliary RVs and the three actions that we would like to emulate is given as follows:

$$Q_{A_{1,2}A_{1,3}X_1} Q_{A_{2,3}|A_{1,3}} Q_{B_{1,2}|X_1A_{1,2}A_{1,3}} Q_{C_2X_2|A_{1,2}A_{1,3}A_{2,3}B_{1,2}} \times Q_{B_{2,3}|A_{2,3}A_{1,3}X_2} Q_{C_3,X_3|A_{2,3}A_{1,3}B_{2,3}} \quad (9)$$

A graphical illustration of the dependencies among the ten RVs for $\mathsf{h} = 3$ is given in Fig. 5. In this illustration, each RV is conditionally independent of the rest given its neighbors (in the undirected sense). It will be precisely this dependency between the auxiliary RVs that will be used to build the codebooks.

To build the codebooks, we first fix a pmf that decomposes as in (9). We then select 12 rates:

- column rates $(\mu^+_{1,3}, \mu^+_{1,2}, \mu^+_{2,3})$ and row rates $(\mu^-_{1,3}, \mu^-_{1,2}, \mu^-_{2,3})$ for $A_{1,3}$, $A_{1,2}$, and $A_{2,3}$ codebooks, respectively;
- column rates (κ^+_1, κ^+_2) and row rates (κ^-_1, κ^-_2) for $B_{1,2}$, and $B_{2,3}$ codebooks, respectively; and
- rates λ_2, and λ_3 for C_2, and C_3 codebooks, respectively;

Using these rates, we construct the codebooks in the following order:

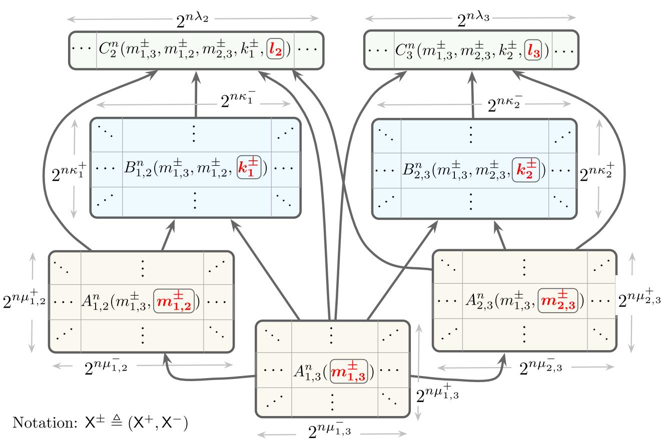

Fig. 6. An illustration of the structure of codebooks for the $\mathsf{h} = 3$ setting.

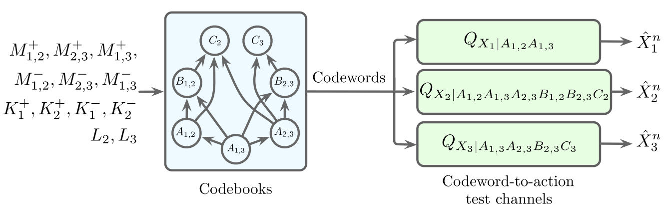

Fig. 7. A *fallacious* attempt at generating the three actions.

- Construct a random codebook of $2^{n(\mu_{1,3}^+ + \mu_{1,3}^-)}$ codewords each generated using $Q_{A_{1,3}}$ and arranged in a table of $2^{n\mu_{1,3}^+}$ rows and $2^{n\mu_{1,3}^-}$ columns.
- For each $A_{1,3}$ codeword, construct a random (conditional) codebook for $A_{1,2}$ with $2^{n(\mu_{1,2}^+ + \mu_{1,2}^-)}$ codewords each generated using $Q_{A_{1,2}|A_{1,3}}$, and arranged in a table of $2^{n\mu_{1,2}^+}$ rows and $2^{n\mu_{1,2}^-}$ columns.
- Similarly, construct random codebooks for $A_{2,3}$ using $Q_{A_{2,3}|A_{1,3}}$.
- For each pair of $A_{1,3}$ and $A_{1,2}$ codewords, construct a random (conditional) codebook for $B_{1,2}$ with $2^{n(\kappa_1^+ + \kappa_1^-)}$ codewords each generated using $Q_{B_{1,2}|A_{1,2}A_{1,3}}$, and arranged in a table of $2^{n\mu_{1,2}^+}$ rows and $2^{n\mu_{1,2}^-}$ columns.
- Similarly, construct random codebooks for $B_{2,3}$ using $Q_{B_{2,3}|A_{2,3}A_{1,3}}$.
- For each tuple of $A_{1,3}$, $A_{1,2}$, $A_{2,3}$, and $B_{1,2}$ codewords, generate a random codebook of $2^{n\lambda_2}$ codewords using $Q_{C_2|A_{1,2}A_{2,3}A_{1,3}B_{1,2}}$.
- Similarly, construct random codebooks for C_3 using $Q_{C_3|A_{2,3}A_{1,3}B_{2,3}}$

Note that the codebooks for $A_{1,3}$, $A_{1,2}$, $A_{2,3}$, $B_{1,2}$, and $B_{2,3}$ are tabular as opposed to being rectangular. Now, for Task 1, one can *naïvely* attempt to use twelve independent and uniformly distributed random indices to select the 7 auxiliary RV codewords and generate the three actions using appropriate *test* channels as illustrated in Fig. 7. Specifically, independent and uniformly distributed indices $\boldsymbol{M}^{\pm} \triangleq (M_{1,2}^+, M_{1,2}^-, M_{1,3}^+, M_{1,3}^-, M_{2,3}^+, M_{2,3}^-)$, $K_1^+, K_1^-, K_2^+, K_2^-, L_2$, and L_3 can be used to select codewords from $A_{1,2}$, $A_{1,3}$, $A_{2,3}$, $B_{1,2}$, $B_{2,3}$, C_2 and C_3 codebooks and three test channels can be used to generate the nodes' actions. One can hope that the three actions derived thus have the right joint statistics.

However, this approach *does not* yield the right joint pmf of actions due to the manner in which codebooks are constructed. To see this, note that not all joint pmfs that decompose as in (8) satisfy $I(B_{1,2}; B_{2,3}|A_{1,2}A_{1,3}A_{2,3}) = 0$. However, since the codebooks for $B_{1,2}$ and $B_{2,3}$ are constructed conditionally on those of $(A_{1,2}, A_{1,3})$ and $(A_{2,3}, A_{1,3})$, respectively, selecting codewords uniformly from the five codebooks will not ensure that the empirical distribution of $(A_{1,2}, A_{1,3}, A_{2,3}, B_{1,2}, B_{2,3})$ derived from the randomly selected codewords matches the marginal derived from the joint pmf in (8). Hence, an alternate and more intricate means to generate actions from the codebooks has to be devised.

Our approach to generating the three actions involves breaking the action-generation problem to the following three generalized channel resolvability problems that are illustrated in Fig. 8.

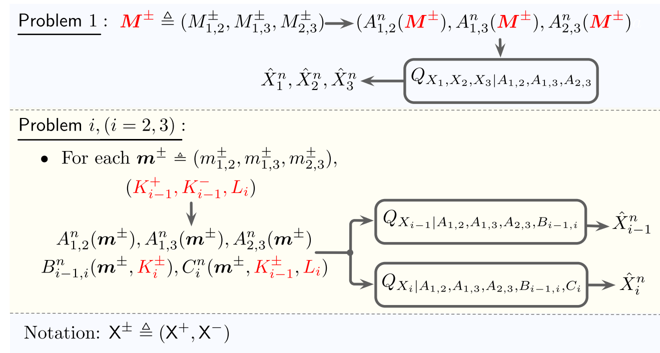

Fig. 8. The 3 subproblems for the $\mathsf{h} = 3$ setting.

- In the first problem, we let random vector of indices $\boldsymbol{M}^{\pm} \triangleq (M_{1,2}^+, M_{1,2}^-, M_{1,3}^+, M_{1,3}^-, M_{2,3}^+, M_{2,3}^-)$ to be uniform over its alphabet. Let this random vector be used to select the corresponding codeword triple $A_{1,2}^n(\boldsymbol{M}^{\pm}), A_{1,3}^n(\boldsymbol{M}^{\pm}), A_{2,3}^n(\boldsymbol{M}^{\pm})$ and let the actions be generated using the test channel $Q_{X_1X_2X_3|A_{1,2}A_{1,3}A_{2,3}}$. The resultant joint pmf of the generated actions is then given by

$$\widehat{Q}^{(1)}_{\hat{X}_1^n\hat{X}_2^n\hat{X}_3^n} \triangleq \frac{\sum_{\boldsymbol{m}^{\pm}} Q^{\otimes n}_{X_1X_2X_3|A_{1,2}A_{1,3}A_{2,3}}\left(\cdot \,\middle|\, \begin{matrix} A_{1,2}^n(\boldsymbol{m}^{\pm}) \\ A_{1,3}^n(\boldsymbol{m}^{\pm}) \\ A_{2,3}^n(\boldsymbol{m}^{\pm}) \end{matrix}\right)}{2^{n(\mu_{1,2}^+ + \mu_{1,2}^- + \mu_{1,3}^+ + \mu_{1,3}^- + \mu_{2,3}^+ + \mu_{2,3}^-)}}.$$

The aim in this problem is to find constraints on the six rates $\mu_{1,2}^+$, $\mu_{1,2}^-$, $\mu_{1,3}^+$, $\mu_{1,3}^-$, $\mu_{2,3}^+$, and $\mu_{2,3}^-$ such that:

1) the joint pmf $\widehat{Q}^{(1)}_{\hat{X}_1^n\hat{X}_2^n\hat{X}_3^n}$ is *close* to the design pmf $\mathsf{Q}^{\otimes n}_{X_1X_2X_3}$, i.e.,

$$\lim_{n\to\infty} \mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}(\widehat{Q}^{(1)}_{\hat{X}_1^n\hat{X}_2^n\hat{X}_3^n} \,\|\, \mathsf{Q}^{\otimes n}_{X_1X_2X_3})\right] = 0. \quad (10)$$

2) $\hat{X}_1^n$ is nearly independent of random indices $\boldsymbol{M}^- \triangleq (M_{1,2}^-, M_{1,3}^-, M_{2,3}^-)$, i.e.,

$$\lim_{n\to\infty} \frac{\sum_{\boldsymbol{m}^-} \mathbb{E}\left[\mathsf{D}_{\text{KL}}(\widehat{Q}^{(1)}_{\hat{X}_1^n|\boldsymbol{M}^-}(\cdot|\boldsymbol{m}^-) \parallel \widehat{Q}^{(1)}_{\hat{X}_1^n})\right]}{2^{n(\mu_{1,2}^- + \mu_{1,3}^- + \mu_{2,3}^-)}} = 0, \tag{11}$$

where $\widehat{Q}^{(1)}_{\hat{X}_1^n|\boldsymbol{M}^-}(\cdot|\tilde{\boldsymbol{m}}^-)$ is the conditional pmf of $\hat{X}_1^n$ given $\boldsymbol{M}^- = \tilde{\boldsymbol{m}}^-$ given by (12) at the top of the next page.

Note that we have expectations in both constraints due to the random construction of the codebooks. Since our overall goal is to generate the actions whose joint statistics nearly match that of an i.i.d. source $\mathsf{Q}_{X_1X_2X_3}$, the first constraint seems natural. The second constraint, however, appears rather unnecessary at first glance. It is an artifact of the proposed scheme and not a requirement of the problem per se. It is imposed so as to be able to eventually interpret the random indices corresponding to row selection (i.e., the $(\cdot)^-$ indices) as originating from the common randomness M_c, which by the setup, is independent of the action specified at Node 1.

- The second problem is a generalized channel resolvability problem in which we focus on generating a pair of actions $(\hat{X}_1^n, \hat{X}_2^n)$. First suppose that we fix a vector of indices $\boldsymbol{m}^\pm$ that selects the three A codewords. Now, let (K_1^+, K_1^-, L_2) be jointly uniform indices that together with $\boldsymbol{m}^\pm$ identify unique codewords for $A_{1,2}$, $A_{1,3}$, $A_{2,3}$, $B_{1,2}$, and C_2 from the respective codebooks. Let these random codewords be used to generate the pair of actions $(\hat{X}_1^n, \hat{X}_2^n)$ using the test channel $Q_{X_1X_2|A_{1,2}A_{1,3}A_{2,3}B_{1,2}C_2}$, which decomposes into two parallel channels $Q_{X_1|A_{1,2}A_{1,3}A_{2,3}B_{1,2}}$ and $Q_{X_2|A_{1,2}A_{1,3}A_{2,3}B_{1,2}C_2}$. The resultant joint pmf of the generated actions $\widehat{Q}^{(2,\boldsymbol{m}^\pm)}_{\hat{X}_1^n\hat{X}_2^n}$ is given by (13) at the top of the next page.

In this problem, we aim to find constraints on codebook rates κ_1^+, κ_1^- and λ_2 such that:

  - for almost all $\boldsymbol{m}^\pm$, pmf $\widehat{Q}^{(2,\boldsymbol{m}^\pm)}_{\hat{X}_1^n\hat{X}_2^n}(\cdot,\cdot)$ is *close* to the pmf $Q^{\otimes n}_{X_1X_2X_3|A_{1,2}A_{1,3}A_{2,3}}\left(\cdot \middle| \begin{matrix} A_{1,2}^n(\boldsymbol{m}^\pm) \\ A_{1,3}^n(\boldsymbol{m}^\pm) \\ A_{2,3}^n(\boldsymbol{m}^\pm) \end{matrix}\right)$ in the sense of (15) at the top of the next page.
  - for almost all $\boldsymbol{m}^\pm$, the action $\hat{X}_1^n$ and K_1^- are nearly independent, which is enforced by satisfying (16). Note that the conditional pmf $\widehat{Q}^{(2,\boldsymbol{m}^\pm)}_{\hat{X}_1^n|K_1^-}$ in (16) is defined in (14), as shown at the top of the next page.

As in the previous problem, the second constraint is imposed with the goal of eventually being able to interpret K_1^- as a part of the common randomness M_c. The first constraint ensures that the codebooks for $B_{1,2}$ and C_2 are sufficiently large to ensure that the pmf pair of action sequences appear independent (in time) conditioned on the A-codewords. The need for this constraint will become clear shortly when we describe our solution to the allied action-generation problem.

- Lastly, as is evident from Fig. 8, the third problem is identical to the second problem except that each subscript index is incremented by 1. Analogous to the previous problem, here we focus on generating action sequence pair $(\hat{X}_2^n, \hat{X}_3^n)$ using the $B_{2,3}$ and C_3 codebooks requiring two constraints similar to (15) and (16). The joint pmf of the pair of action sequences for each realization of $\boldsymbol{m}^\pm$ in this case is denoted by $\widehat{Q}^{(3,\boldsymbol{m}^\pm)}_{\hat{X}_2^n\hat{X}_3^n}(\cdot,\cdot)$.

Now, let us suppose that we can find codebook rates such that the above six constraints (two for each problem) are met. Let us define a scheme for generating the three sequence of action as follows:

- Let $\boldsymbol{M}^\pm = (M_{1,2}^+, M_{1,2}^-, M_{1,3}^+, M_{1,3}^-, M_{2,3}^+, M_{2,3}^-)$ be a random vector uniform over its alphabet, i.e., any subset of components constitute of uniform and independent random variables.
- Let $\hat{X}_1^n$ be the action sequence generated by transmitting $\boldsymbol{A}^n(\boldsymbol{M}^\pm) \triangleq (A_{1,2}^n(\boldsymbol{M}^\pm), A_{1,3}^n(\boldsymbol{M}^\pm), A_{2,3}^n(\boldsymbol{M}^\pm))$ over the channel $Q_{X_1|A_{1,2}A_{1,3}A_{2,3}} = Q_{X_1|A_{1,2}A_{1,3}}$. The resulting pmf of $\hat{X}_1^n$ is given by Problem 1 to be $\widehat{Q}^{(1)}_{\hat{X}_1^n}$.
- Using the realizations of $\boldsymbol{M}^\pm$ and $\hat{X}_1^n$, generate $(K_1^\pm, L_2)$ distributed according to $Q_{K_2^\pm,L_2|\hat{X}_1^n,\boldsymbol{M}^\pm}$ induced by Problem 2, and transmit codewords $\boldsymbol{A}^n(\boldsymbol{M}^\pm)$, $B_{1,2}^n(\boldsymbol{M}^\pm, K_1^\pm)$ and $C_2(\boldsymbol{M}^\pm, K_1^\pm, L_2)$ over the channel $Q_{X_2|\boldsymbol{A},B_{1,2},C_2}$ to generate $\hat{X}_2^n$.
- Similar to the previous step, using the realizations of $\boldsymbol{M}^\pm$ and $\hat{X}_2^n$, generate $(K_2^\pm, L_3)$ distributed according to $Q_{K_2^\pm,L_3|X_2^n,\boldsymbol{M}^\pm}$ induced by Problem 3, and transmit codewords $\boldsymbol{A}^n(\boldsymbol{M}^\pm)$, $B_{2,3}^n(\boldsymbol{M}^\pm, K_2^\pm)$ and $C_2^n(\boldsymbol{M}^\pm, K_2^\pm, L_3)$ over the channel $Q_{X_3|\boldsymbol{A},B_{2,3},C_3}$ to generate $\hat{X}_3^n$.

At this point, the following fact must be highlighted. First, by dividing the problem into three, we have ensured that even though the codebooks for $B_{1,2}$ and $B_{1,3}$ are constructed conditionally independent given $\boldsymbol{A} = (A_{1,2}, A_{1,3}, A_{2,3})$, the codeword selection is such that the conditional empirical distribution of $(B_{1,2}, B_{1,3})$ given $\boldsymbol{A}$ matches correlation $Q_{B_{1,2},B_{1,3}|\boldsymbol{A}}$ induced by (8), thereby avoiding the issue in the fallacious attempt given in Fig. 7. In other words, in the above approach, $(\boldsymbol{M}^\pm, K_1^\pm, L_2)$ is a collection of independent and uniform random indices, $(\boldsymbol{M}^\pm, K_2^\pm, L_3)$ is a collection of independent and uniform random indices, but not $(\boldsymbol{M}^\pm, K_1^\pm, K_2^\pm)$. It is not difficult to see that the above scheme yields actions jointly correlated according to pmf $\check{Q}_{\hat{X}_1^n\hat{X}_2^n\hat{X}_3^n}$ given by (17) at the top of the next page.

Now, if we can identify the codebook rate conditions for the three problems, and prove that $\check{Q}_{\hat{X}_1^n\hat{X}_2^n\hat{X}_3^n}$ can be made arbitrarily close to $\mathsf{Q}^{\otimes n}_{X_1X_2X_3}$ under the variational distance metric as n is allowed to grow unboundedly, then we have a solution for the allied action-generation problem. Theorems 1 and 2, and Lemma 3, detailed in Section IV-B, establish exactly that.

In specific, Theorem 1 of Sec. IV-B3 guarantees that (10) and (11) of Problem 1 hold provided the following eight rate conditions are met, of which the first four are required to meet (10), and the last four, to meet (11).

$$\mu_{1,3}^+ + \mu_{1,3}^- > I(X_1X_2X_3; A_{1,3}) \tag{18a}$$

$$\widehat{Q}^{(1)}_{\hat{X}_1^n|\boldsymbol{M}^-}(\cdot|\tilde{\boldsymbol{m}}^-) \triangleq \sum_{\boldsymbol{m}^\pm:\boldsymbol{m}^-=\tilde{\boldsymbol{m}}^-} \frac{Q^{\otimes n}_{X_1|A_{1,2}A_{1,3}A_{2,3}}(\cdot|A^n_{1,2}(\boldsymbol{m}^\pm), A^n_{1,3}(\boldsymbol{m}^\pm), A^n_{2,3}(\boldsymbol{m}^\pm))}{2^{n(\mu^+_{1,2}+\mu^+_{1,3}+\mu^+_{2,3})}}. \quad (12)$$

$$\widehat{Q}^{(2,\boldsymbol{m}^\pm)}_{\hat{X}_1^n\hat{X}_2^n}(\cdot,\cdot) = \frac{\sum_{k_1^+,k_1^-,l_2} Q^{\otimes n}_{X_1X_2|A_{1,2}A_{1,3}A_{2,3}B_{1,2}C_2}(\cdot,\cdot|A^n_{1,2}(\boldsymbol{m}^\pm)A^n_{1,3}(\boldsymbol{m}^\pm)A^n_{2,3}(\boldsymbol{m}^\pm)B^n_{1,2}(\boldsymbol{m}^\pm,k_1^\pm)C^n_2(\boldsymbol{m}^\pm,k_1^\pm,l_2))}{2^{n(\kappa_1^++\kappa_1^-+\lambda_2)}}, \quad (13)$$

$$\widehat{Q}^{(2,\boldsymbol{m}^\pm)}_{\hat{X}_1^n|K_1^-}(\cdot|k_1^-) = \sum_{k_1^+} \frac{Q^{\otimes n}_{X_1|A_{1,2}A_{1,3}A_{2,3}B_{1,2}}(\cdot|A^n_{1,2}(\boldsymbol{m}^\pm), A^n_{1,3}(\boldsymbol{m}^\pm), A^n_{2,3}(\boldsymbol{m}^\pm), B^n_{1,2}(\boldsymbol{m}^\pm,k_1^+,k_1^-))}{2^{n\kappa_1^+}}. \quad (14)$$

$$\lim_{n\to\infty}\sum_{\boldsymbol{m}^\pm} \frac{\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\widehat{Q}^{(2,\boldsymbol{m}^\pm)}_{\hat{X}_1^n\hat{X}_2^n}\middle\| Q^{\otimes n}_{X_1X_2|A_{1,2}A_{1,3}A_{2,3}}\left(\cdot\,|A^n_{1,2}(\boldsymbol{m}^\pm), A^n_{1,3}(\boldsymbol{m}^\pm), A^n_{2,3}(\boldsymbol{m}^\pm)\right)\right)\right]}{2^{n(\mu^+_{1,2}+\mu^-_{1,2}+\mu^+_{1,3}+\mu^-_{1,3}+\mu^+_{2,3}+\mu^-_{2,3})}} = 0. \quad (15)$$

$$\lim_{n\to\infty}\sum_{\boldsymbol{m}^\pm,k_1^-} \frac{\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\widehat{Q}^{(2,\boldsymbol{m}^\pm)}_{\hat{X}_1^n|K_1^-}(\cdot|k_1^-)\middle\| Q^{\otimes n}_{X_1|A_{1,2}A_{1,3}A_{2,3}}\left(\cdot\,|A^n_{1,2}(\boldsymbol{m}^\pm), A^n_{1,3}(\boldsymbol{m}^\pm), A^n_{2,3}(\boldsymbol{m}^\pm)\right)\right)\right]}{2^{n(\mu^+_{1,2}+\mu^-_{1,2}+\mu^+_{1,3}+\mu^-_{1,3}+\mu^+_{2,3}+\mu^-_{2,3}+\kappa_1^-)}} = 0. \quad (16)$$

$$\breve{Q}_{\hat{X}_1^n\hat{X}_2^n\hat{X}_3^n}(x_1^n,x_2^n,x_3^n) \triangleq \sum_{\boldsymbol{m}^\pm} \frac{Q^{\otimes n}_{X_1|A_{1,2}A_{1,3}A_{2,3}}(x_1^n|\boldsymbol{A}^n(\boldsymbol{m}^\pm)) \cdot \frac{\widehat{Q}^{(2,\boldsymbol{m}^\pm)}_{\hat{X}_1^n\hat{X}_2^n}(x_1^n,x_2^n)}{\widehat{Q}^{(2,\boldsymbol{m}^\pm)}_{\hat{X}_1^n}(x_1^n)} \cdot \frac{\widehat{Q}^{(3,\boldsymbol{m}^\pm)}_{\hat{X}_2^n\hat{X}_3^n}(x_2^n,x_3^n)}{\widehat{Q}^{(3,\boldsymbol{m}^\pm)}_{\hat{X}_2^n}(x_2^n)}}{2^{n(\mu^+_{1,2}+\mu^-_{1,2}+\mu^+_{1,3}+\mu^-_{1,3}+\mu^+_{2,3}+\mu^-_{2,3})}}. \quad (17)$$

$$\begin{bmatrix}\mu^+_{1,2}+\mu^-_{1,2}\\ +\mu^+_{1,3}+\mu^-_{1,3}\end{bmatrix} > I(X_1X_2X_3;A_{1,2}A_{1,3}) \quad (18b)$$
$$\begin{bmatrix}\mu^+_{1,3}+\mu^-_{1,3}\\ +\mu^+_{2,3}+\mu^-_{2,3}\end{bmatrix} > I(X_1X_2X_3;A_{1,3}A_{2,3}) \quad (18c)$$
$$\begin{bmatrix}\mu^+_{1,2}+\mu^-_{1,2}+\mu^+_{1,3}\\ +\mu^-_{1,3}+\mu^+_{2,3}+\mu^-_{2,3}\end{bmatrix} > I(X_1X_2X_3;A_{1,2}A_{1,3}A_{2,3}) \quad (18d)$$
$$\mu^+_{1,3} > I(X_1;A_{1,3}) \quad (18e)$$
$$\mu^+_{1,2}+\mu^+_{1,3} > I(X_1;A_{1,2}A_{1,3}) \quad (18f)$$
$$\mu^+_{1,3}+\mu^+_{2,3} > I(X_1;A_{1,3}A_{2,3}) \quad (18g)$$
$$\mu^+_{1,2}+\mu^+_{1,3}+\mu^+_{2,3} > I(X_1;A_{1,2}A_{1,3}A_{2,3}) \quad (18h)$$

Next, Theorem 2 of Sec. IV-B3 guarantees that (15) and (16) of Problem 2, and the respective conditions of Problem 3 hold provided the following six rate conditions are met.

$$\kappa_1^++\kappa_1^-+\lambda_2 > I(X_1X_2X_3;B_{1,2}C_2|A_{1,2}A_{1,3}A_{2,3}) \quad (19a)$$
$$\kappa_1^++\kappa_1^- > I(X_1X_2X_3;B_{1,2}|A_{1,2}A_{1,3}A_{2,3}) \quad (19b)$$
$$\kappa_1^+ > I(X_1;B_{1,2}|A_{1,2}A_{1,3}A_{2,3}) \quad (19c)$$
$$\kappa_2^++\kappa_2^-+\lambda_3 > I(X_1X_2X_3;B_{2,3}C_3|A_{1,2}A_{1,3}A_{2,3}) \quad (19d)$$
$$\kappa_2^++\kappa_2^- > I(X_1X_2X_3;B_{2,3}|A_{1,2}A_{1,3}A_{2,3}) \quad (19e)$$
$$\kappa_2^+ > I(X_1;B_{2,3}|A_{1,2}A_{1,3}A_{2,3}) \quad (19f)$$

Thus, from the above, we see that if the codebook rates satisfy the fourteen constraints in (18) and (19), then as a consequence of Lemma 3 of Sec. IV-B4, the above approach is a solution to the allied action-generation problem. This solution can be translated to a solution to our strong coordination problem by inverting the operation at Node 1 (i.e., by selecting corresponding indices from the specified action, instead of generating the action for Node 1 from random indices). The resulting solution to the strong coordination problem is as follows.

- **A1** The nodes agree on a joint pmf that satisfies (9) with additional constraints $C_2 = X_2$ and $C_3 = X_3$.
- **A2** The nodes collectively decide on the twelve codebook rates meeting the fourteen constraints in (18) and (19), and agree on a realization of the codebooks.
- **A3** The nodes use disjoint parts of the shared common randomness to identify random indices specifying the columns for A- and B-codewords, i.e., all nodes extract $(M^-_{1,2}, M^-_{1,3}, M^-_{2,3}, K^-_1, K^-_2)$ from M_c.
- **A4** Given X_1^n and the selected $(M^-_{1,2}, M^-_{1,3}, M^-_{1,2})$, Node 1 generates random indices $(M^+_{1,2}, M^+_{1,3}, M^+_{1,2})$ using the pmf $Q_{M^+_{1,2}M^+_{1,3}M^+_{1,2}|X_1^nM^-_{1,2}M^-_{1,3}M^-_{1,2}}$ induced by Problem 1. The randomness required to generate this random indices is obtained by using the local randomness available at Node 1. Since $(M^-_{1,2}, M^-_{1,3}, M^-_{1,2})$ is extracted from M_c, $(M^-_{1,2}, M^-_{1,3}, M^-_{1,2})$ and X_1^n are necessarily independent. We must therefore explicitly enforce the independence of $(M^-_{1,2}, M^-_{1,3}, M^-_{1,2})$ and X_1^n in our design, and hence the requirement of (11) in Problem 1.
- **A5** Node 1 then uses $(M^\pm_{1,2}, M^\pm_{1,2}, K_1^+)$ and actions X_1^n to generate an index K_1^+ distributed according to $Q_{K_1^+|X_1^nM^\pm_{1,2},M^\pm_{1,3},K_1^-}$ induced by Problem 2. Again, the randomness required to generate this index is derived from the local randomness available at Node 1.
- **A6** Node 1 then conveys $(M^+_{1,2}, M^+_{1,3}, K_1^+)$ to Node 2.
- **A7** Node 2 uses the received indices $(M^+_{1,2}, M^+_{1,3}, K_1^+)$ and the column indices identified from the common randomness to identify the codewords for $A_{1,2}$, $A_{1,3}$, and $B_{1,2}$. Additionally, it uses its local randomness to

select a uniform random index L_2 independent of all other indices.

A8 Node 2 then uses the received/selected indices to identify codewords for $A_{1,2}$, $A_{1,3}$, $B_{1,2}$, and C_2, and declares its sequence of actions X_2^n to be the same as $C_2^n(M_{1,2}^{\pm}, M_{1,3}^{\pm}, M_{2,3}^{\pm}, K_1^{\pm}, L_2)$.

A9 Node 2 then uses the generated action X_2^n and the indices $(M_{1,2}^{\pm}, M_{1,3}^{\pm}, M_{2,3}^{\pm}, K_2^{-})$ and its local randomness to generate index K_2^+ distributed according to pmf $Q_{K_2^+|X_2^n M_{1,2}^{\pm}, M_{1,3}^{\pm}, M_{2,3}^{\pm}, K_1^-}$ induced by Problem 2. Just as in A2, it is for this step that we require (16) to hold.

A10 Node 2 then conveys $(M_{1,3}^+, M_{2,3}^+, K_2^+)$ to Node 3.

A11 Node 3 generates a uniform random index L_3 independent of all other indices using its local randomness.

A12 Node 3 uses the received messages and the column indices extracted from local and common randomness to identify $A_{1,3}$, $A_{2,3}$, $B_{2,3}$ and C_3 codewords. As in the case of Node 2, Node 3 declares its actions to be the selected codeword $C_3^n(M_{1,3}^{\pm}, M_{2,3}^{\pm}, K_2^{\pm}, L_3)$.

At this juncture, the reader will be able to corroborate the details of the above scheme with those presented in Fig. 4. A straightforward computation then reveals that the joint pmf of the actions generated by the above scheme can be made vanishingly close to $\breve{Q}_{\hat{X}_1^n \hat{X}_2^n \hat{X}_3^n}$ as n grows, which in turn can be made arbitrarily close to $\mathsf{Q}_{X_1,X_2,X_3}^{\otimes n}$ if (a) the codebooks satisfy (18) and (19), and (b) the local and common randomness rates are sufficiently large to ensure that the above steps can be realized. We therefore require the following:

1) The common randomness rate should be sufficiently large so as to perform Step **A3**, which requires
$$\mathsf{R}_c \ge \mu_{1,2}^- + \mu_{1,3}^- + \mu_{2,3}^- + \kappa_1^- + \kappa_2^-. \tag{20}$$

2) The local randomness at Node 1 is required to select the indices $(M_{1,2}^+, M_{1,3}^+)$ at Step **A3** and K_1^+ at Step **A4**. The rates of local randomness required to achieve these random index selection is quantified by Theorems 6 and 7 of Appendices D and E, respectively, to be:
$$\rho_1 \ge \Big(\mu_{1,2}^+ + \mu_{1,3}^+ - I(X_1; A_{1,2}, A_{1,3})\Big) + \big(\kappa_1^+ - I(X_1; B_{1,2}|A_{1,2}, A_{1,3})\big). \tag{21}$$

3) The local randomness at Node 2 is required to select L_2 at Step **A7** and K_2^+ at Step **A9**. Again, by Theorem 7 of Appendix E, we see that the required rate of local randomness is
$$\rho_2 \ge \lambda_2 + \kappa_2^+ - I(X_1; B_{2,3}|A_{1,2}, A_{1,3}, A_{2,3}). \tag{22}$$

4) The local randomness at Node 2 is only required to select L_3, which necessitates
$$\rho_3 \ge \lambda_3. \tag{23}$$

5) Lastly, the communication rates on the two links must be sufficient for the conveyance of the corresponding indices in Steps **A6** and **A10**, for which we require
$$\mathsf{R}_1 \ge \mu_{1,2}^+ + \mu_{1,3}^+ + \kappa_1^+, \tag{24}$$
$$\mathsf{R}_2 \ge \mu_{1,3}^+ + \mu_{2,3}^+ + \kappa_2^+. \tag{25}$$

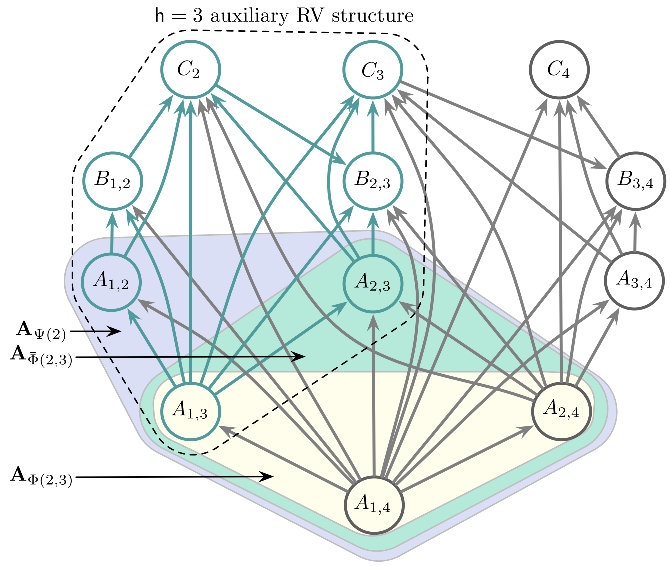

Fig. 9. An illustration of the auxiliary RV structure for $\mathsf{h} = 4$.

Piecing all the results together, we obtain the implicit achievability result for $\mathsf{h} = 3$ setting that a rate tuple $(\mathsf{R}_c, \mathsf{R}_1, \mathsf{R}_2, \rho_1, \rho_2, \rho_3)$ is achievable if we can find twelve codebook rates satisfying (18) and (19) such that (20)-(25) also hold. Lastly, we can possibly enlarge the space of achievable tuples by incorporating the rate-transfer arguments of Lemma 1 thereby yielding all rate tuples achievable using the above strong coordination scheme. Note that since we have not imposed any constraints on the operation of intermediate node, the constraints correspond to the unrestricted mode. We can derive similar implicit achievability results for schemes based on functional and action-dependent modes. Detailed treatment for these two modes are only provided in the general multi-hop setting in the following section.

B. Inner Bound: An Achievable Scheme for General Multi-Hop Line Networks

The approach for strong coordination in general multi-hop networks is a natural extension of the two-hop scheme presented in Section IV-A. As in the two-hop setting, we divide the problem into three tasks, the first of which considers the appropriate allied action-generation problem. This problem is solved by first choosing an appropriate structure for auxiliary RVs, suitably constructing codebooks that reflect the dependencies in the chosen RV structure, and subdividing the problem appropriately to solve each of them separately and then by piecing them together to find the general solution to the action-generation problem. Next, we assign each index corresponding to an auxiliary RV codeword to either common or local randomness at a particular node, and lastly, we invert the operation at Node 1, and identify the required rates of communication, and local and common randomness. The reader is redirected to Fig. 4 for an illustration of this approach.

1) Choice of Auxiliary Random Variables (Task 1): We use $\binom{\mathsf{h}}{2} + 2\mathsf{h} - 2$ auxiliary RVs in a specific way to generate the h actions in the allied problem. For an illustration of the auxiliary RV structure for the allied action-generation problem when $\mathsf{h} = 3$ and $\mathsf{h} = 4$, the reader is directed to Figs. 5 and 9, respectively. The details of the general auxiliary RV structure are as follows:

- There are three different groups of Auxiliary RVs collectively indicated by the letters A, B, and C. There are $\binom{\mathsf{h}}{2}$ random variables $\{A_{i,j} : 1 \le i < j \le \mathsf{h}\}$, $\mathsf{h} - 1$ random variables $\{B_{i,i+1} : 1 \le i < \mathsf{h}\}$, and $\mathsf{h}-1$ random variables $\{C_i : 2 \le i \le \mathsf{h}\}$.
- For $1 \le i < j \le \mathsf{h}$, auxiliary RV $A_{i,j}$ represents message generated at Node i intended for Node j, and hence can be used by Nodes $i, i+1, \ldots, j$.
- As a generalization of (5), for each $1 \le i < j \le \mathsf{h}$, we impose the following Markov chain.[1]

$$A_{i,j} \leftrightarrow \boldsymbol{A}_{\Phi(i,j)} \leftrightarrow \boldsymbol{A} \setminus \{A_{i,j}\}, \tag{26}$$

where we use the following notation.

$$\begin{aligned}
\mathcal{F} &\triangleq \{(i,j) : 1 \le i < j \le \mathsf{h}\},\\
\Phi(i,j) &\triangleq \left\{(i',j') : i' \le i < j \le j'\right\} \setminus \{(i,j)\},\\
\boldsymbol{A}_S &\triangleq \{A_s\}_{s \in S}, \quad S \subseteq \mathcal{F},\\
\boldsymbol{A} &\triangleq \boldsymbol{A}_{\mathcal{F}}.
\end{aligned}$$

Note that $\boldsymbol{A}_{\Phi(i,j)}$ represents RVs generated by nodes prior to and including Node i and are intended for Node j or nodes situated after Node j *except* RV $A_{i.j}$. Fig. 9 presents an illustration of $\boldsymbol{A}_{\Phi(2,3)}$ for $\mathsf{h} = 4$. In the figure, the subset of auxiliary RVs that are connected by directed edges with heads at RV $A_{i,j}$ form $\boldsymbol{A}_{\Phi(i,j)}$. Since Nodes $i, \ldots, j$ have access to the codewords corresponding to $\boldsymbol{A}_{\Phi(i,j)}$, we can allow arbitrary correlation between $A_{i,j}$ and $\boldsymbol{A}_{\Phi(i,j)}$. Further, for each $(i',j') \notin \Phi(i,j) \cup \{(i,j)\}$, there is at least one k' such that $i \le k' \le j$ and Node k' does not have access to $A_{i',j'}$. Hence, $\boldsymbol{A}_{\Phi(i,j)\cup\{(i,j)\}}$ is the unique maximal subset of auxiliary RVs that Nodes $i, \ldots, j$ together share. The Markov conditions of (26) generalize (5) in the two-hop setting, and are required to ensure that jointly uniform random indices can be used to select auxiliary RV codewords (from a suitable codebook setup) whose joint empirical pmf matches that of the design pmf.
- For each $i = 1, \ldots, \mathsf{h} - 1$, auxiliary RV $B_{i,i+1}$ is generated by Node i and is intended for Node $i+1$ one hop away. As in the two-hop setting, the auxiliary RVs $A_{i,i+1}$ and $B_{i,i+1}$ play different roles (see Discussion 1 of Section IV-A); auxiliary random variables $\{A_{i,j} : 1 \le i < j \le \mathsf{h}\}$ are selected prior to the generation of any action sequence, i.e., $\hat{X}_2^n, \ldots, \hat{X}_{\mathsf{h}}^n$. However, for each $1 \le i < \mathsf{h}$, $B_{i,i+1}$ is selected after the action $\hat{X}_i^n$ is generated. The curious reader might wonder if the fact that auxiliary RVs $A_{i,i+1}$ and $B_{i,i+1}$ play different roles is reason enough to introduce the B RVs. The fact that B auxiliary RVs are indeed essential and that they strictly improve the achievability scheme is established at the end of this work in Section V-C.
- Lastly, the $\mathsf{h} - 1$ auxiliary RVs $\{C_2, \ldots, C_{\mathsf{h}}\}$ are introduced solely to quantify the use of local randomness required at each node in the network. The indices and the codewords corresponding to these auxiliary RVs are not communicated between nodes.

[1]The authors thank the anonymous reviewer who suggested approaching the problem via a combination of Slepian-Wolf coding/random binning exploiting of [15, Lemma 1] (also of [30, Th. 1]) that does not require the Markov assumptions in (26). The random binning approach however, in theory, requires at least h^3 binning operations/rates. In this work, we have restricted our focus on an achievable scheme using channel resolvability codebooks.

The joint pmf of the actions and auxiliary RVs $Q_{\boldsymbol{A},B_{1,2},\ldots,B_{\mathsf{h}-1,\mathsf{h}},C_2,\ldots,C_{\mathsf{h}},X_1,\ldots X_{\mathsf{h}}}$ that we aim to emulate must factorize as

$$Q_{A_{1,2}\cdots A_{\mathsf{h}-1,\mathsf{h}}} Q_{X_1|\boldsymbol{A}_{\Psi(1)}} \times \prod_{j=1}^{\mathsf{h}-1} \left(\begin{array}{r} Q_{B_{j,j+1}|X_j \boldsymbol{A}_{\overline{\Phi}(j,j+1)}} Q_{C_{j+1}|\boldsymbol{A}_{\Psi(j+1)} B_{j,j+1}} \\ \times Q_{X_{j+1}|\boldsymbol{A}_{\Psi(j+1)} B_{j,j+1} C_{j+1}} \end{array} \right), \tag{28}$$

where $Q_{A_{1,2},\ldots,A_{\mathsf{h}-1,\mathsf{h}}}$ satisfies the conditions described by (26) and the marginal $Q_{X_1\cdots X_{\mathsf{h}}} = \mathsf{Q}_{X_1\cdots X_{\mathsf{h}}}$. In (28), we let

$$\overline{\Phi}(i,j) \triangleq \Phi(i,j) \cup \{(i,j)\}, \tag{29a}$$
$$\Psi(i) \triangleq \{(i',j') \in \mathcal{F} : i' \le i \le j'\}. \tag{29b}$$

Note that $\boldsymbol{A}_{\overline{\Phi}(i,j)} = \boldsymbol{A}_{\Phi(i,j)\cup\{(i,j)\}}$ is exactly the set of all A auxiliary RVs that Nodes $i, \ldots, j$ have access to, and $\boldsymbol{A}_{\Psi(i)}$ represents all auxiliary RVs generated by Nodes $1, \ldots, i$ intended for Nodes $i, \ldots, \mathsf{h}$. An illustration of $\overline{\Phi}(2,3)$, $\Psi(2)$ for $h = 4$ can be found in Fig. 9.

As in the two-hop case, while it is preferable that there be only one RV per hop that encapsulates completely the role of the message conveyed on a hop, we do not have the tools to devise an achievable scheme with such a property. The joint pmf in (28) is the most general structure of RVs for which we are able to devise an achievable scheme using the channel resolvability codebook approach. We now present the precise codebook structure and construction that emulates (28).

2) Codebook Construction (Task 1): The construction of codebooks that incorporate the specific structure of auxiliary RVs is accomplished using the following (total) ordering of index pairs. Codebooks for A auxiliary RVs are constructed starting from the leftmost index pair.

$$\begin{aligned}(1,\mathsf{h}) &\succ (1,\mathsf{h}-1) \succ \cdots \succ (1,2) \succ (2,\mathsf{h}) \succ \cdots \succ (2,3)\\ &\succ (3,\mathsf{h}) \succ \cdots \succ (3,4) \succ \cdots \succ (\mathsf{h}-1,\mathsf{h}).\end{aligned}$$

To define the codebooks, we define the rates for each codebooks as in Table I. As in the two-hop case, we assign two rates for the codebooks for each A and B auxiliary RV and one for each C codebook. For A or B auxiliary RV codebooks, the rates with superscript $+$ will correspond to messages communicated over edges, and the rates with the superscript $-$ will correspond to indices extracted from common randomness shared by all nodes.

We use the following notation.

$$\begin{aligned}
\mathcal{M}_S^{\pm} &\triangleq \bigtimes_{(i',j')\in S} \left(\mathcal{M}_{i',j'}^{+} \times \mathcal{M}_{i',j'}^{-}\right), \quad S \subseteq \mathcal{F},\\
m_{i,j}^{\pm} &\triangleq (m_{i,j}^{+}, m_{i,j}^{-}), \quad (i,j) \in \mathcal{F},\\
\boldsymbol{m}_S^{\mathsf{x}} &\triangleq \{m_s^{\mathsf{x}}\}_{s\in S}, \quad S \subseteq \mathcal{F} \text{ and } \mathsf{x} \in \{+,-,\pm\},\\
\boldsymbol{m}^{\mathsf{x}} &\triangleq \boldsymbol{m}_{\mathcal{F}}^{\mathsf{x}}, \quad \mathsf{x} \in \{+,-,\pm\}.
\end{aligned}$$

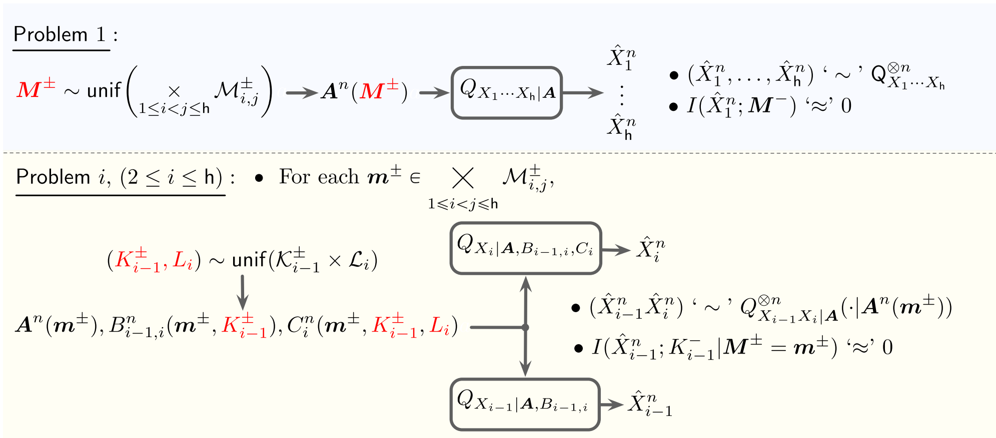

Fig. 10. The h subproblems.

TABLE I
CODEBOOK PARAMETERS AND NOTATION

Auxiliary RV	Codebook indices, their rates and alphabets	
$A_{i,j}$ $((i,j)\in\mathcal{F})$	$(\mu^+_{i,j},\mu^-_{i,j})\in[0,\infty)\times[0,\infty)$	
	$\mathcal{M}^+_{i,j} \triangleq [\![1,2^{n\mu^+_{i,j}}]\!]$; $\mathcal{M}^-_{i,j} \triangleq [\![1,2^{n\mu^-_{i,j}}]\!]$; $\mathcal{M}^{\pm}_{i,j} \triangleq \mathcal{M}^+_{i,j}\times\mathcal{M}^-_{i,j}$; $m^{\pm}_{i,j} \triangleq (m^+_{i,j},m^-_{i,j})\in\mathcal{M}^{\pm}_{i,j}$	
$B_{i,i+1}$ $(1\le i<\mathsf{h})$	$(\kappa^+_i,\kappa^-_i)\in[0,\infty)\times[0,\infty)$	
	$\mathcal{K}^+_i \triangleq [\![1,2^{n\kappa^+_i}]\!]$; $\mathcal{K}^-_i \triangleq [\![1,2^{n\kappa^-_i}]\!]$; $\mathcal{K}^{\pm}_i \triangleq \mathcal{K}^+_i\times\mathcal{K}^-_i$; $k^{\pm}_i \triangleq (k^+_i,k^-_i)\in\mathcal{K}^{\pm}_i$	
C_i $(1<i\le\mathsf{h})$	$\lambda_i\in[0,\infty)$	
	$\mathcal{L}_i \triangleq [\![1,2^{n\lambda_i}]\!]$	$l_i\in\mathcal{L}_i$

The codebooks for the auxiliary RVs are constructed as follows.

B1 For each $m^{\pm}_{1,\mathsf{h}} \in \mathcal{M}^+_{1,\mathsf{h}}\times\mathcal{M}^-_{1,\mathsf{h}}$, generate codeword $A^n_{1,\mathsf{h}}(m^{\pm}_{1,\mathsf{h}}) \sim Q^{\otimes n}_{A_{1,\mathsf{h}}}$ independently.

B2 For $(i,j),(i',j')\in\mathcal{F}$ such that $(i,j)\succ(i',j')$, the codebook for $A_{i,j}$ is constructed before that for $A_{i',j'}$. By design, the codebook for $A_{i',j'}$ is constructed after those for $A_{i'',j''}$, $(i'',j'')\in\Phi(i',j')$.

B3 For each $1\le i<j\le\mathsf{h}$, $\boldsymbol{m}^{\pm}_{\Phi(i,j)}\in\mathcal{M}^{\pm}_{\Phi(i,j)}$, generate $A^n_{i,j}(\boldsymbol{m}^{\pm}_{\Phi(i,j)}) \sim Q^{\otimes n}_{A_{i,j}|\boldsymbol{A}_{\Phi(i,j)}}(\cdot|\boldsymbol{A}^n_{\Phi(i,j)}(\boldsymbol{m}^{\pm}_{\Phi(i,j)}))$ independently using previously selected $\boldsymbol{A}^n_{\Phi(i,j)}(\boldsymbol{m}^{\pm}_{\Phi(i,j)})$.

B4 For each $i=1,2,\dots,\mathsf{h}-1$, and index tuple $(\boldsymbol{m}^{\pm}_{\Phi(i,i+1)},k^{\pm}_i)\in\mathcal{M}^{\pm}_{\Phi(i,i+1)}\times\mathcal{K}^{\pm}_i$, independently generate random codeword $B^n_{i,i+1}(\boldsymbol{m}^{\pm}_{\Phi(i,i+1)},k^{\pm}_i)$ distributed according to $Q^{\otimes n}_{B_{i,i+1}|\boldsymbol{A}_{\Phi(i,i+1)}}(\cdot|\boldsymbol{A}^n_{\Phi(i,i+1)}(\boldsymbol{m}^{\pm}_{\Phi(i,i+1)}))$.

B5 For $1<i\le\mathsf{h}$, and index tuple $(\boldsymbol{m}^{\pm}_{\Psi(i)},\boldsymbol{k}^{\pm}_{i-1},l_i)\in\mathcal{M}^{\pm}_{\Psi(i)}\times\mathcal{K}^{\pm}_{i-1}\times\mathcal{L}_i$, independently generate random codeword $C^n_i(\boldsymbol{m}^{\pm}_{\Psi(i)},k^{\pm}_{i-1},l_i)$ distributed according to $Q^{\otimes n}_{C_i|\boldsymbol{A}_{\Psi(i)}B_{i-1,i}}(\cdot|\boldsymbol{A}^n_{\Psi(i)}(\boldsymbol{m}^{\pm}_{\Psi(i)}),B^n_{i-1,i}(\boldsymbol{m}^{\pm}_{\Psi(i)},k^{\pm}_{i-1}))$.

While the above description is notation-heavy, it is, in reality, a multi-hop generalization of the description of the codebooks for $\mathsf{h}=3$ given in Section IV-A and Fig. 6. For the sake of succinctness, we introduce the following notation:

$$\boldsymbol{A}^n(\boldsymbol{m}^{\pm}) \triangleq \left\{A^n_{i,j}(\boldsymbol{m}^{\pm}_{\Phi(i,j)})\right\}_{(i,j)\in\mathcal{F}}$$

and for $1<i\le h$,

$$B^n_{i-1,i}(\boldsymbol{m}^{\pm},k^{\pm}_{i-1}) \triangleq B^n_{i-1,i}(\boldsymbol{m}^{\pm}_{\Phi(i-1,i)},k^{\pm}_{i-1}),$$
$$C^n_i(\boldsymbol{m}^{\pm},k^{\pm}_{i-1},l_i) \triangleq C^n_i(\boldsymbol{m}^{\pm}_{\Psi(i)},k^{\pm}_{i-1},l_i).$$

Note that we have so far neither specified the rates $\mu^+_{i,j}$, $\mu^-_{i,j}$, κ^+_i, κ^-_i, and λ_i in the above description, nor have we described how the codewords are going to be selected for generating the actions. In the following, we will identify the required rates so that appropriate channel resolvability code design techniques can be employed to generate the actions.

3) Identifying Codebook Rates (Task 1): Similar to the two-hop setting, the allied action-generation problem is decomposed into h problems whose solutions will be pieced together to form a solution for the allied action-generation problem. Consider h problems illustrated in Fig. 10, and formally defined below.

Definition 3: Problem 1 pertains to characterization of the rates required for codebooks corresponding to auxiliary RVs $\{A_{i,j}:(i,j)\in\mathcal{F}\}$. Let a realization of the A codebooks be fixed. Let $\boldsymbol{M}^{\pm} \sim \mathsf{unif}\left(\bigtimes_{(i,j)\in\mathcal{F}}\mathcal{M}^{\pm}_{i,j}\right)$ be used to select the A-codewords, which are then transmitted through the DMC $Q_{X_1\cdots X_\mathsf{h}|\boldsymbol{A}}$ to obtain $\hat{X}^n_1,\dots,\hat{X}^n_\mathsf{h}$. Let $\widehat{Q}^{(1)}_{\hat{X}^n_1\cdots\hat{X}^n_\mathsf{h}\boldsymbol{M}^{\pm}}$ denote the induced joint pmf. Then,

$$\widehat{Q}^{(1)}_{\hat{X}^n_1\cdots\hat{X}^n_\mathsf{h}} \triangleq \sum_{\boldsymbol{m}^{\pm}} \frac{Q^{\otimes n}_{X_1\cdots X_\mathsf{h}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))}{2^{n\sum_{(i,j)\in\mathcal{F}}(\mu^+_{i,j}+\mu^-_{i,j})}} = \sum_{\boldsymbol{m}^{\pm}} \frac{Q^{\otimes n}_{X_1|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\prod_{j=2}^{\mathsf{h}}\frac{Q^{\otimes n}_{X_jX_{j-1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))}{Q^{\otimes n}_{X_{j-1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))}}{2^{n\sum_{(i,j)\in\mathcal{F}}(\mu^+_{i,j}+\mu^-_{i,j})}} \quad (33)$$

denote the pmf of the output $(\hat{X}^n_1,\dots,\hat{X}^n_\mathsf{h})$ from the channel $Q_{X_1\cdots X_\mathsf{h}|\boldsymbol{A}}$ when the codewords are selected from the codebooks uniformly at random. Problem 1 aims to derive conditions on codebook rates $\{(\mu^+_{i,j}, \mu^-_{i,j})\}_{(i,j)\in\mathcal{F}}$ such that:

$$\lim_{n\to\infty}\left(\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}(\widehat{Q}^{(1)}_{\hat{X}^n_1\cdots\hat{X}^n_{\mathsf{h}}} \parallel \mathsf{Q}^{\otimes n}_{X_1\cdots X_{\mathsf{h}}})\right]\right) = 0, \quad (34)$$

$$\lim_{n\to\infty}\left(\sum_{\boldsymbol{m}^-}\frac{\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}(\widehat{Q}^{(1)}_{\hat{X}^n_1|\boldsymbol{M}^-}(\cdot|\boldsymbol{m}^-) \parallel \widehat{Q}^{(1)}_{\hat{X}^n_1})\right]}{2^{n\sum_{(i,j)\in\mathcal{F}}\mu^-_{i,j}}}\right) = 0, \quad (35)$$

where $\widehat{Q}^{(1)}_{\hat{X}^n_1|\boldsymbol{M}^-}(\cdot|\boldsymbol{m}^-)$ is the conditional pmf of $\hat{X}^n_1$ given $\boldsymbol{M}^- = \boldsymbol{m}^-$, and similar to (12), is given by

$$\widehat{Q}^{(1)}_{\hat{X}^n_1|\boldsymbol{M}^-}(\cdot|\tilde{\boldsymbol{m}}^-) \triangleq \sum_{\boldsymbol{m}^\pm:\boldsymbol{m}^-=\tilde{\boldsymbol{m}}^-}\frac{Q^{\otimes n}_{X_1|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^\pm))}{2^{n\sum_{(i,j)\in\mathcal{F}}\mu^+_{i,j}}}.$$

■

As in the two-hop setting the first constraint concerns our goal of generating actions that are statistically indistinguishable from those of a DMS $\mathsf{Q}_{X_1\cdots X_{\mathsf{h}}}$, whereas the second is imposed with the eventual goal of identifying $\boldsymbol{M}^- \triangleq \{M^-_{i,j}\}_{(i,j)\in\mathcal{F}}$ as random indices extracted from the randomness common to all nodes, which by the nature of the setup is independent of the specified action. The following result provides conditions on the A codebook rates that ensure that the two constraints of Problem 1 are satisfied.

Theorem 1: Suppose that the finite-alphabet auxiliary RV[2] codebooks are constructed using Steps **B1-B5**. For $S \subsetneq \mathcal{F}$, let $\mathcal{J}_S \triangleq \{(i,j) : \overline{\Phi}(i,j)\cap S \neq \emptyset\}$. Then,

- (34) is met if for each $S \subseteq \mathcal{F}\setminus\{(1,\mathsf{h})\}$,

$$\sum_{s\in\mathcal{F}\setminus S}(\mu^+_s + \mu^-_s) > I\big(X_1,\ldots,X_{\mathsf{h}}; A_{\mathcal{F}\setminus\mathcal{J}_S}\big). \quad (36)$$

- (35) is met if for each $S \subseteq \mathcal{F}\setminus\{(1,\mathsf{h})\}$,

$$\sum_{s\in\mathcal{F}\setminus S}\mu^+_s > I\big(X_1; A_{\mathcal{F}\setminus\mathcal{J}_S}\big). \quad (37)$$

Proof: See Appendix A. ■

The two conditions of the above theorem are the *natural* multi-variable generalizations of the two-node coordination capacity [3] (also see (4)). The first constraint prescribes a minimum size for each of the codebooks, which appears as a sum of row- and colum-rates for codebooks, whereas the second condition prescribes the minimum row-rate for the codebooks, or equivalently, the maximum binning possible for each of the codebooks, since the column rates will be eventually interpreted as indices extracted from common randomness as in the earlier two-hop setting. The theorem prescribes $2^{\binom{\mathsf{h}}{2}}$ rate conditions that need to be met for Problem 1, however, as in the example below, many equations can be redundant. In fact, the only non-trivial constraints correspond to those sets S for which $S = \mathcal{J}_S$. When imagined pictorially using Fig. 9, the non-trivial constraints correspond to sets S for which there is no edge from an element of S^c to an element of S.

Example 1: Consider the rate constraints imposed by (36) for $S = \{(1,3)\}$, $\{(1,2),(1,3)\}$, $\{(1,3),(2,3)\}$ and $\{(1,2),(1,3),(2,3)\}$ corresponding to $\mathsf{h} = 4$. For each of these choices for S, $\mathcal{F}\setminus\mathcal{J}_S = \{(1,4),(2,4),(3,4)\}$, and therefore the RHS of the constraints imposed by (36) are identical. However, the LHS terms differ, and of these 4 constraints, only the one corresponding to $S = \{(1,2),(1,3),(2,3)\}$ is non-trivial (i.e., the constraints corresponding to the remaining three choices for S are redundant), and is given by

$$\begin{bmatrix}\mu^+_{1,4} + \mu^-_{1,4} + \mu^+_{2,4}\\ +\mu^-_{2,4} + \mu^+_{3,4} + \mu^-_{3,4}\end{bmatrix} > I(\{X_i\}^4_{i=1}; \{A_{i,4}\}^3_{i=1}).$$

■

Now that we have identified rate constraints that meet (34), by invoking Pinsker's and Jensen inequalities [28], [31], the following result can be established.

Remark 3: When the rate constraints given by (36) are met,

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}&\left\|\widehat{Q}^{(1)}_{\hat{X}^n_1\cdots\hat{X}^n_{\mathsf{h}}} - \mathsf{Q}^{\otimes n}_{X_1\cdots X_{\mathsf{h}}}\right\|_1\\ &= \lim_{n\to\infty}\mathbb{E}\left\|\sum_{\boldsymbol{m}^\pm}\frac{Q^{\otimes n}_{X_1\cdots X_{\mathsf{h}}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^\pm))}{2^{\sum_{(i,j)\in\mathcal{F}}n(\mu^+_{i,j}+\mu^-_{i,j})}} - \mathsf{Q}^{\otimes n}_{X_1\cdots X_{\mathsf{h}}}\right\|_1 \quad (38)\\ &= \lim_{n\to\infty}\mathbb{E}\left\|\frac{\sum_{\boldsymbol{m}^\pm}Q^{\otimes n}_{X_1|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^\pm))\prod_{j=2}^{\mathsf{h}}\frac{Q^{\otimes n}_{X_jX_{j-1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^\pm))}{Q^{\otimes n}_{X_{j-1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^\pm))}}{2^{n\sum_{(i,j)\in\mathcal{F}}(\mu^+_{i,j}+\mu^-_{i,j})}}\right.\\ &\qquad\left. - \mathsf{Q}^{\otimes n}_{X_1\cdots X_{\mathsf{h}}}\right\|_1 = 0, \quad (39)\end{aligned}$$

where the expectation is over only the A-codebooks. ■

Thus, by choosing the A-codebook rates satisfying the constraints of Theorem 1, we are guaranteed the existence of codebooks using which we can generate the h actions. At this point, it might seem that the generation of the h actions is complete, and can proceed with identifying the indices used for codeword selection with the appropriate resources, and inverting the operation at Node 1. However, this is not the case since our aim is to devise a scheme that generates the actions in a *distributed* fashion (cf. Task 1 of Section IV, and Fig. 4). We do not yet have distributed generation of actions, since $Q_{X_1,\ldots,X_{\mathsf{h}}|\boldsymbol{A}}$ does not decompose into h parallel channels, i.e.,

$$Q_{X_1,\ldots,X_{\mathsf{h}}|\boldsymbol{A}} = Q_{X_1|\boldsymbol{A}}\prod_{i=2}^{n}Q_{X_i|\boldsymbol{A},X_{i-1}} \neq Q_{X_1|\boldsymbol{A}}\prod_{i=2}^{n}Q_{X_i|\boldsymbol{A}}.$$

In order to decompose the channel $Q_{X_1,\ldots,X_{\mathsf{h}}|\boldsymbol{A}}$ into h parallel channels, we need to use B- and C-codebooks, and we do that via the following $\mathsf{h}-1$ subproblems.

Definition 4: For $i = 2,\ldots,\mathsf{h}$, Problem i quantifies the codebook rates correspon for auxiliary RVs $B_{1,2},\ldots,B_{\mathsf{h}-1,\mathsf{h}}$ and $C_2,\ldots,C_{\mathsf{h}}$. Fix a realization of all codebooks, $i \in \{2,\ldots,\mathsf{h}\}$, and $\boldsymbol{m}^\pm \in \mathcal{M}^\pm_{\mathcal{F}}$. Let $(K^\pm_{i-1}, L_i) \sim \mathsf{unif}\left(\mathcal{K}^+_{i-1}\times\mathcal{K}^-_{i-1}\times\mathcal{L}_i\right)$ and the given $\boldsymbol{m}^\pm$ be used to select codewords $\boldsymbol{A}^n(\boldsymbol{m}^\pm)$, $B^n_{i-1,i}(\boldsymbol{m}^\pm, K^\pm_{i-1})$ and $C^n_i(\boldsymbol{m}^\pm, K^\pm_{i-1}, L_i)$, which are then transmitted over $Q_{X_{i-1}X_i|\boldsymbol{A}B_{i-1,i}C_i}$ to obtain

[2]If instead of a KL divergence requirement, our requirements were formulated using variational distance between the corresponding distributions, we could have used the techniques developed in [13] and [30] to establish the result for general (i.e., not necessarily finite) auxiliary RV alphabets. However, in this work, we restrict ourselves to RVs with finite alphabets.

$$\widehat{Q}^{(i,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{i-1}\hat{X}^n_i}(\cdot,\cdot) = \sum_{\boldsymbol{k}^{\pm}_{i-1},l_i} \frac{Q^{\otimes n}_{X_{i-1}X_i|\boldsymbol{A}B_{i-1,i}C_i}(\cdot,\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}), B^n_{i-1,i}(\boldsymbol{m}^{\pm},k^{\pm}_{i-1}), C^n_i(\boldsymbol{m}^{\pm},k^{\pm}_{i-1},l_i))}{2^{n(\kappa^+_{i-1}+\kappa^-_{i-1}+\lambda_i)}}. \tag{40}$$

$$\widehat{Q}^{(i,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{i-1}|K^-_{i-1}}(\cdot|k^-_{i-1}) = \sum_{\tilde{k}^{\pm}_{i-1}:\tilde{k}^-_{i-1}=k^-_{i-1}} \frac{Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}B_{i-1,i}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}), B^n_{i-1,i}(\boldsymbol{m}^{\pm},\tilde{k}^{\pm}_{i-1}))}{2^{n\kappa^+_{i-1}}}. \tag{41}$$

$$\lim_{n\to\infty}\sum_{\boldsymbol{m}^{\pm}} \frac{\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\widehat{Q}^{(i,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{i-1}\hat{X}^n_i}\,\middle\|\, Q^{\otimes n}_{X_{i-1}X_i|\boldsymbol{A}}\left(\,\cdot\,|\boldsymbol{A}^n(\boldsymbol{m}^{\pm})\right)\right)\right]}{2^{n\sum_{(i,j)\in\mathcal{F}}(\mu^+_{i,j}+\mu^-_{i,j})}} = 0. \tag{42}$$

$$\lim_{n\to\infty}\sum_{\boldsymbol{m}^{\pm},k^-_{i-1}} \frac{\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\widehat{Q}^{(i,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{i-1}|K^-_{i-1}}(\cdot|k^-_{i-1})\,\middle\|\, Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}}\left(\,\cdot\,|\boldsymbol{A}^n(\boldsymbol{m}^{\pm})\right)\right)\right]}{2^{n\left(\kappa^-_{i-1}+\sum_{(i,j)\in\mathcal{F}}(\mu^+_{i,j}+\mu^-_{i,j})\right)}} = 0. \tag{43}$$

$\hat{X}^n_{i-1}$ and $\hat{X}^n_i$. Let $\widehat{Q}^{(i,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{i-1}\hat{X}^n_i K^{\pm}_{i-1}L_i}$ denote the induced joint pmf of the indices used for codeword selection and the pair of channel outputs, whose marginal corresponding of the pair of channel outputs is given in (40) at the top of this page. Since $X_{i-1} \leftrightarrow (\boldsymbol{A}, B_{i-1,i}) \leftrightarrow C_i$, for $i \in \{2,\ldots,\mathsf{h}\}$, $\boldsymbol{m}^{\pm} \in \mathcal{M}^{\pm}_{\mathcal{F}}$ and $k^-_{i-1} \in \mathcal{K}^-_{i-1}$, we can define a conditional pmf $\widehat{Q}^{(i,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{i-1}|K^-_{i-1}}(\cdot|k^-_{i-1})$ as in (41) at the top of this page. The aim of $\mathsf{Problem}\ i$, $i = 2,\ldots,\mathsf{h}$, is to derive conditions on $\kappa^+_{i-1}, \kappa^-_{i-1}, \lambda_i$ such that (42) and (43), are shown at the top of this page, are met. ■

$\mathsf{Problem}\ i$, $2 \le i \le \mathsf{h}$, also poses two constraints, each similar to the corresponding ones in $\mathsf{Problem\ 1}$.

- The first constraint enables us to approximate the conditional pmfs that appear within the product in (33) by conditional pmfs in (40) derived from outputs of suitable channel resolvability codes.
- The second constraint ensures for each $\boldsymbol{m}^{\pm}$ the action $\hat{X}^n_{i-1}$ generated by averaging over the $B_{i-1,i}$ and C_i codebooks is nearly independent of K^-_{i-1}, which will be needed when we transform the scheme for the allied action-generation problem into a scheme for strong coordination, at which time, K^-_{i-1} will be viewed as a part of common randomness.

The following result characterizes sufficient conditions on the codebook rates for (42) and (43) to hold.

Theorem 2: Fix $i \in \{2,\ldots,\mathsf{h}\}$. Let the auxiliary RV codebooks be constructed using **B1**-**B5**. Then, (42) holds if the rates are chosen such that

$$\kappa^+_{i-1} + \kappa^-_{i-1} + \lambda_i > I(X_{i-1}, X_i; B_{i-1,i}, C_i|\boldsymbol{A}), \tag{44a}$$

$$\kappa^+_{i-1} + \kappa^-_{i-1} > I(X_{i-1}, X_i; B_{i-1,i}|\boldsymbol{A}). \tag{44b}$$

Further, (43) is met provided

$$\kappa^+_{i-1} > I(X_{i-1}; B_{i-1,i}|\boldsymbol{A}). \tag{45}$$

Proof: See Appendix B. ■

Analogous to Remark 3, an application of Pinsker's and Jensen inequalities to (42) yields the following.

Remark 4: When the rate constraints (44a) and (44b) are met for each $i = 2,\ldots,\mathsf{h}$,

$$\lim_{n\to\infty} \frac{\sum_{\boldsymbol{m}^{\pm}}\mathbb{E}\left\|\widehat{Q}^{(i,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{i-1}\hat{X}^n_i} - Q^{\otimes n}_{X_{i-1}X_i|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\right\|_1}{2^{n\sum_{(i,j)\in\mathcal{F}}(\mu^+_{i,j}+\mu^-_{i,j})}} = 0, \tag{46}$$

where $\widehat{Q}^{(i,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{i-1}\hat{X}^n_i}$ is defined in (40). ■

We are now ready to combine together the solutions for the h problems to obtain a code for generating the h sources from the auxiliary RVs via their codebooks.

4) A Solution for the Allied Action-generation Problem ($\mathsf{Task\ 1}$): We start with the following result that enables to replace ideal conditional pmfs in the product term of (39) with those obtained from channel resolvability codes using the solutions to $\mathsf{Problem}\ i$, $i = 2,\ldots,\mathsf{h}$.

Lemma 3: For the random codebook construction given in **B1**-**B5**, if the rate constraints given in Theorems 1 and 2 are met, then, in addition to (39) and (46), the following also holds.

$$\sum_{x^n_1,\ldots,x^n_{\mathsf{h}}} \mathbb{E}\left|\sum_{\boldsymbol{m}^{\pm}} \frac{Q^{\otimes n}_{X_1|\boldsymbol{A}}(x^n_1|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))}{2^{n\sum_{(i,j)\in\mathcal{F}}(\mu^+_{i,j}+\mu^-_{i,j})}} \prod_{j=2}^{\mathsf{h}} \frac{\widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{j-1}\hat{X}^n_j}(x^n_{j-1},x^n_j)}{\widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{j-1}}(x^n_{j-1})} - \mathsf{Q}^{\otimes n}_{X_1\cdots X_{\mathsf{h}}}(x^n_1,\ldots,x^n_{\mathsf{h}})\right| \overset{n\to\infty}{\longrightarrow} 0. \tag{47}$$

Proof: See Appendix C. ■

The key to devising a solution to the allied action-generation problem lies in the summation term in (47). Let us take a closer look at this term, since it actually yields a distributed solution to the action-generation problem as detailed in Fig. 11. Suppose that the codebooks for the auxiliary RVs are constructed according to **B1**-**B5**. Let $\boldsymbol{M}^{\pm} \sim \mathsf{unif}\left(\mathcal{M}^{\pm}_{\mathcal{F}}\right)$ be used to select codewords for $A_{1,2},\ldots,A_{\mathsf{h}-1,\mathsf{h}}$ from their codebooks. Then,

- The first term $Q^{\otimes n}_{X_1|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{M}^{\pm}))$ is the pmf of the output when $\boldsymbol{A}^n(\boldsymbol{M}^{\pm})$ is fed into DMC $Q_{X_1|\boldsymbol{A}}$. By (28), $X_1 \leftrightarrow (A_{1,2},\ldots,A_{1,\mathsf{h}}) \leftrightarrow \boldsymbol{A}$. So, even though we use the channel $Q_{X_1|\boldsymbol{A}}$, it is effectively $Q_{X_1|A_{1,2},A_{1,3}\ldots,A_{1,\mathsf{h}}}$, and

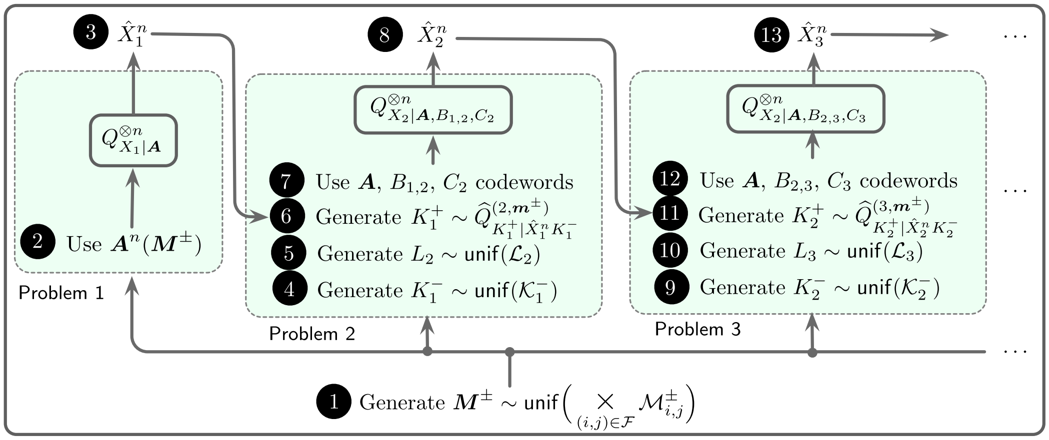

Fig. 11. A solution to the allied action-generation problem in the multi-hop setting.

only codewords $\{A^n_{1,j}(\boldsymbol{M}^{\pm}) : j > 1\}$ affect the output and its distribution.

- Consider the product term in (47). From Fig. 10, we see that the conditional pmf $\frac{\widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{j-1}\hat{X}^n_j}(x^n_{j-1},x^n_j)}{\widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{j-1}}(x^n_{j-1})}$ is achieved by: (1) generating $(K^{\pm}_{j-1}, L_j) \sim \widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{K^{\pm}_{j-1}L_j|\hat{X}^n_{j-1}}(\cdot,\cdot\,|\,x^n_{j-1})$, (2) selecting the $\boldsymbol{A}$, $B_{j-1,j}$, and C_j codewords corresponding to $\boldsymbol{m}^{\pm}$ and the realizations of $(K^{\pm}_{j-1}, L_j)$, and (3) transmitting the selected codewords through $Q_{X_j|\boldsymbol{A},B_{j-1,j},C_j}$ to obtain $\hat{X}^n_j$. Notice from Fig. 10 that L_j is used only to determine the codeword for C_j, and is independent of $K^{\pm}_{j-1}$ and $\hat{X}^n_{j-1}$. Thus, generating $(K^{\pm}_{j-1}, L_j) \sim \widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{K^{\pm}_{j-1}L_j|\hat{X}^n_{j-1}}(\cdot,\cdot\,|\,x^n_{j-1})$ is the same as generating $K^{\pm}_{j-1} \sim \widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{K^{\pm}_{j-1}|\hat{X}^n_{j-1}}(\cdot\,|\,x^n_{j-1})$ and independently generating L_j according to a uniform distribution. Further, from the second constraint (43) enforced in Problem j, we also ensure that K^-_{j-1} is nearly independent of $\hat{X}^n_{j-1}$. Thus, generating $K^{\pm}_{j-1} \sim \widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{K^{\pm}_{j-1}|\hat{X}^n_{j-1}}(\cdot\,|\,x^n_{j-1})$ is nearly the same as generating K^-_{j-1} independently according to a uniform distribution, and then generating $K^+_{j-1} \sim \widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{K^+_{j-1}|\hat{X}^n_{j-1},K^-_{j-1}}$.

Combining the above facts, we see that

$$\sum_{\boldsymbol{m}^{\pm}} \frac{Q^{\otimes n}_{X_1|\boldsymbol{A}}(x^n_1|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))}{2^{n\sum_{(i,j)\in\mathcal{F}}(\mu^+_{i,j}+\mu^-_{i,j})}} \prod_{j=2}^{\mathsf{h}} \frac{\widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{j-1}\hat{X}^n_j}(x^n_{j-1},x^n_j)}{\widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{j-1}}(x^n_{j-1})}$$

is the joint pmf of the h actions generated as in Fig. 11. In light of Lemma 3 above, we then infer that the illustration in Fig. 11 *is indeed* a solution to the allied action-generation problem. We are finally ready to state the achievable scheme for strong coordination using that of the allied action-generation problem.

5) A Scheme for Strong Coordination and the Resources Required (Task 2 and Task 3): Now that we have essentially completed the design of a scheme that generates the h actions, we are done with Task 1 described in Fig. 4. Before we present the details of Task 2 of Fig. 4 that relates to identifying the resources used for corresponding codebook rates, we detail Task 3 that relates to inverting the operation at Node 1 to generate the messages from the specified action. The strong coordination scheme derived from the above action-generation scheme is as follows.

C1 Pick a pmf meeting (28) with the additional constraint that $C_i = X_i$ for $i > 1$. Let $\{(\mu^+_{i,j}, \mu^-_{i,j}) : (i,j) \in \mathcal{F}\}$, $\{(\kappa^+_i, \kappa^-_i) : 1 \le i < \mathsf{h}\}$, and $\{\lambda_i : 1 < i \le \mathsf{h}\}$ satisfy (36), (37), (44), and (45).

C2 Generate codebooks for $\{A_{i,j}\}_{(i,j)\in\mathcal{F}}$, $\{B_{i,i+1}\}_{i=1}^{\mathsf{h}-1}$, and $\{C_i\}_{i=2}^{\mathsf{h}}$ via Steps **B1**-**B5**.

C3 Generate an instance $(\{m^-_{i,j}\}_{(i,j)\in\mathcal{F}}, \{k^-_i\}_{1\le i<\mathsf{h}})$ of RVs $(\{M^-_{i,j}\}_{(i,j)\in\mathcal{F}}, \{K^-_i\}_{1\le i<\mathsf{h}})$ distributed according to

$$\mathsf{unif}\Bigg(\bigtimes_{(i,j)\in\mathcal{F}} [\![1, 2^{n\mu^-_{i,j}}]\!] \times \bigtimes_{1\le\ell<\mathsf{h}} [\![1, 2^{n\kappa^-_\ell}]\!]\Bigg).$$

These RVs are assumed to be extracted from the common randomness available to all nodes.

C4 Generate an instance $\{m^+_{i,j} : 2 \le i < j \le \mathsf{h}\}$ of $\{M^+_{i,j}\}_{2\le i<j\le\mathsf{h}} \sim \mathsf{unif}\Bigg(\bigtimes_{2\le i<j\le\mathsf{h}} [\![1, 2^{n\mu^+_{i,j}}]\!]\Bigg)$. Random index $M^+_{i,j}$ is assumed to be generated by Node i.

C5 Given $X^n_1 = x^n_1$, Node 1 uses the joint pmf of $(\boldsymbol{M}^{\pm}, \hat{X}^n_1)$ obtained from Problem 1 to generate instances $(m^+_{1,2}, m^+_{1,3}, \dots, m^+_{1,\mathsf{h}})$ of $(M^+_{1,2}, M^+_{1,3}, \dots, M^+_{1,\mathsf{h}})$ distributed according to

$$\widehat{Q}^{(1)}_{M^+_{1,2}M^+_{1,3}\cdots M^+_{1,\mathsf{h}}|\hat{X}^n_1 M^-_{1,2}M^-_{1,3}\cdots M^-_{1,\mathsf{h}}}(\cdot|x^n_1, m^-_{1,2}, \dots, m^-_{1,\mathsf{h}}).$$

Let $\boldsymbol{m}^{\pm}$ be the instance of the realized and generated indices in Steps **C3**-**C5**.

C6 For $i = 2, \dots, \mathsf{h}-1$, repeat the following four steps in the given order, and for increasing i.

C7 Using the joint pmf obtained from Problem $i+1$, Node i generates a realization k^+_i of $K^+_i \in [\![1, 2^{n\kappa^+_i}]\!]$ such that $K^+_i \sim \widehat{Q}^{(i+1,\boldsymbol{m}^{\pm})}_{K^+_i|\hat{X}^n_i,\boldsymbol{M}^{\pm},K^-_i}(\cdot|\hat{x}^n_i, k^-_i)$.

C8 Node i forwards $\{m^+_{i',j'} : i' \le i < j'\}$, and k^+_i to Node $i+1$.

C9 Node $i+1$ generates realization l_{i+1} of $L_{i+1} \sim \mathsf{unif}([\![1, 2^{n\lambda_{i+1}}]\!])$ independent of all other RVs.

C10 Declare $\hat{X}^n_{i+1} \triangleq C^n_{i+1}(\boldsymbol{m}^{\pm}, k^+_i, k^-_i, l_{i+1})$.

What remains now is the task of associating the rates of the various indices to that of the resources (i.e., communication rates, and (local and common) randomness rates), which is Task 2 as described in Fig. 4. Note that this association requires the identification of the resource required to realize Steps **C3**, **C4**, **C5**, **C7**, **C8**, and **C9**, and depends on the mode of intermediate node operation illustrated in Fig. 2. This association for each mode is as follows.

a) Unrestricted mode: In this mode, the following associations are allowable.

- Communication Rate: Step **C8** quantifies the transmission by all nodes. For $1 \le i < \mathsf{h}$, Node i communicates $\{M^+_{i',j'} : i' \le i < j'\}$, and K^+_i to Node $i+1$. Thus,

$$\mathsf{R}_i \ge \kappa^+_i + \sum_{(i',j')\in\Psi(i)} \mu^+_{i',j'}, \quad 1 \le i < \mathsf{h}. \tag{48}$$

- Common randomness: Common randomness is needed only to identify $\{M^-_{i,j} : (i,j) \in \mathcal{F}\}$ and $\{K^-_i : 1 \le i < \mathsf{h}\}$ in Step **C3**, and hence, we require

$$\mathsf{R}_c \ge (\kappa^-_1 + \cdots + \kappa^-_{\mathsf{h}-1}) + \sum_{(i,j)\in\mathcal{F}} \mu^-_{i,j}. \tag{49}$$

- Local randomness at Node 1: At Node 1, we need to use local randomness at Steps **C5** and **C7** to generate indices $M^+_{1,2}, \ldots, M^+_{1,\mathsf{h}}$ and K^+_1 so as to select $A_{1,2}, \ldots, A_{1,\mathsf{h}}$ and $B_{1,2}$ codewords using the given realization of X^n_1, and available common randomness. Using Theorem 6 of Appendix D, we can identify the amount of local randomness needed to generate these indices to be

$$\rho_1 \ge \kappa^+_1 - I(X_1; A_{1,2}, \ldots, A_{1,\mathsf{h}}, B_{1,2}) + \sum_{j=2}^{\mathsf{h}} \mu^+_{1,j}. \tag{50}$$

- Local randomness at intermediate Nodes: For $2 \le i < \mathsf{h}$, the local randomness at Node i serves three purposes: (1) to select messages $\{M^+_{i,j} : j > i\}$ at Step **C4**, which requires a rate of $\mu^+_{i,j}$; (2) to generate K^+_i at Step **C7**, which by Theorem 7 of Appendix E, requires a rate of $\kappa^+_i - I(X_i; B_{i,i+1}|\boldsymbol{A})$; and lastly, (3) to generate L_i at Step **C9** for use in generating the Node i's action, which requires a rate λ_i. Thus, for $1 < \ell < \mathsf{h}$,

$$\rho_\ell \ge \left(\kappa^+_\ell - I(X_\ell; B_{\ell,\ell+1}|\boldsymbol{A})\right) + \lambda_\ell + \sum_{j=\ell+1}^{\mathsf{h}} \mu^+_{\ell,j}. \tag{51}$$

- Local randomness intermediate Node h: Local randomness ρ_h is only needed to generate L_h at Step **C9** to output action $\hat{X}^n_\mathsf{h}$ at Step **C10**. Thus,

$$\rho_\mathsf{h} \ge \lambda_\mathsf{h}. \tag{52}$$

b) Functional mode: In this mode, local randomness at a node other than Node 1 can only be used to generate the node's actions. This imposes two constraints specific to this mode.

1. From Step **C8** we see that for $1 \le i < \mathsf{h}$, Node i transmits $\{M^+_{i',j'} : i' \le i < j'\}$, and K^+_i to Node $i+1$. For $i > 1$, K^+_i is selected based on the action for Node i (see (28) and Step **C7**), which can depend on the local randomness of Node i. Hence, for $i > 1$, K^+_i can depend on the local randomness of Node i. Since in this mode, communicated messages cannot depend on local randomness of intermediate nodes, we must require
$$H(B_{i,i+1}) = \kappa^+_i = \kappa^-_i = 0, \quad i > 1. \tag{53}$$
Further, $B_{1,2}$ can be absorbed into $A_{1,2}$ since both are generated using Node 1's actions. Thus, without loss of generality, we can assume that (53) holds for $i = 1$ as well.
2. Random indices $\{M^+_{i,j} : (i,j) \in \mathcal{F}\}$ in Step **C4** must be selected using either the incoming message or the common randomness. Hence, for any $i > 1$, it must be true that Node $i-1$ is also aware of $M^+_{i,j}$. Thus, Node $i-1$ is also aware of the exact the chosen $A_{i,j}$ codeword. Proceeding inductively, we can argue that Node 1 must be aware of the $A_{i,j}$ codeword. Thus, for $i > 1$ and for each $j = 3, \ldots, \mathsf{h}$, we can embed auxiliary RVs $A_{2,j}, \ldots, A_{j-1,j}$ into auxiliary RV $A_{1,j}$ without affecting the communication, local randomness or common randomness requirements. Thus,
$$H(A_{i,j}) = \mu^+_{i,j} = \mu^-_{i,j} = 0, \quad 2 < i < j. \tag{54}$$

Incorporating these two conditions, we obtain the following resource requirements for this mode.

$$\mathsf{R}_i \ge \mu^+_{1,i+1} + \cdots + \mu^+_{1,\mathsf{h}}, \quad 1 \le i < \mathsf{h}, \tag{55a}$$
$$\mathsf{R}_c \ge \mu^-_{1,i+1} + \cdots + \mu^-_{1,\mathsf{h}}, \tag{55b}$$
$$\rho_\ell \ge \begin{cases} \sum_{j=2}^{\mathsf{h}} \mu^+_{1,j} - I(X_1; A_{1,2}, \ldots, A_{1,\mathsf{h}}) & \ell = 1 \\ \lambda_\ell & \ell > 1. \end{cases} \tag{55c}$$

c) Action-dependent mode: As in the case of the functional mode, two conditions specific to the action-dependent mode arise due to the restrictions intermediate node operation.

- Just as in the functional mode, here too it can be shown that auxiliary RVs $A_{2,3}, \ldots, A_{\mathsf{h}-1,\mathsf{h}}$ can be absorbed into auxiliary RVs $A_{1,2}, \ldots, A_{1,\mathsf{h}}$. Thus, without loss of generality, we may assume that
$$H(A_{i,j}) = \mu^+_{i,j} = \mu^-_{i,j} = 0, \quad 2 < i < j. \tag{56}$$
- Since an intermediate node's local randomness cannot be used to generate the next-hop messages (see Fig. 2), the randomness required to implement Step **C7** must be extracted from either the incoming message and the common randomness. Proceeding inductively, we can infer that the randomness usable from the incoming message must indeed originate from Node 1 and must be communicated in a hop-by-hop fashion. Thus, for

$i = 1, \ldots, \mathsf{h}-1$, the link between Node i and Node $i+1$ must be used to communicate $\{M^{+}_{1,\ell} : \ell > i\}$, K^{+}_{i}, and the randomness required for selecting $\{K^{+}_{\ell} : \ell > i\}$ at Step **C7** for all downstream nodes.

Piecing together these requirements, we obtain the following resource requirements. For $1 \le i < \mathsf{h}$,

$$\mathsf{R}_c \ge \sum_{k=2}^{\mathsf{h}} \mu^{-}_{1,k} + \sum_{\ell=1}^{\mathsf{h}-1} \kappa^{-}_{\ell}, \tag{57a}$$

$$\mathsf{R}_i \ge \kappa^{+}_{i} + \sum_{k=i+1}^{\mathsf{h}} \mu^{+}_{1,k} + \sum_{\ell=i+1}^{\mathsf{h}-1} (\kappa^{+}_{\ell} - I(X_\ell; B_{\ell,\ell+1}|\boldsymbol{A})), \tag{57b}$$

$$\rho_1 \ge \left(\begin{array}{c} \sum_{j=2}^{\mathsf{h}} \mu^{+}_{1,j} - I(X_1; A_{1,2}, \ldots, A_{1,\mathsf{h}}) \\ + \sum_{\ell=1}^{\mathsf{h}-1} (\kappa^{+}_{\ell} - I(X_\ell; B_{\ell,\ell+1}|\boldsymbol{A})) \end{array} \right), \tag{57c}$$

$$\rho_\ell \ge \lambda_\ell, \quad \ell > 1. \tag{57d}$$

6) Inner Bound: Now that we have completed the three steps, the general inner bound to the capacity region achievable only by schemes in the (a) functional mode is given by the conditions in (55) along with the rate-transfer arguments allowed by Lemma 2; (b) action-dependent mode is given by the conditions in (57) along with the rate-transfer arguments allowed by Lemma 2; and (c) in the unrestricted mode is given by the conditions in (48)-(52) along with the rate-transfer arguments allowed by Lemma 1. In each mode, the codebook parameters must also satisfy (36), (37), (44), and (45).

V. Capacity Regions

In this section, we derive outer bounds for specific settings that establish the optimality of previously derived achievable scheme i.e., the trade-offs achieved by proposed scheme cannot be surpassed/improved by any other scheme.

A. Functional-Mode Capacity Region

We begin this section with the capacity result characterizing the trade-offs among network resources to establish strong coordination using exclusively the functional mode of intermediate node operation.

Theorem 3: A rate point $(\mathsf{R}_c, \mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1}, \rho_1, \ldots, \rho_{\mathsf{h}})$ is achievable with the functional mode *if and only if* there exist auxiliary RVs $Z_2, \ldots, Z_h$ jointly correlated with the actions according to pmf $Q_{X_1,\ldots,X_{\mathsf{h}},Z_2,Z_3,\ldots,Z_h}$ such that:

1. $Q_{X_1,\ldots,X_{\mathsf{h}}} = \mathsf{Q}_{X_1,\ldots,X_{\mathsf{h}}}$;
2. $Q_{X_1,\ldots,X_{\mathsf{h}},Z_2,Z_3,\ldots,Z_{\mathsf{h}}} = Q_{X_1,Z_2,Z_3,\ldots,Z_{\mathsf{h}}} \prod_{\ell=2}^{\mathsf{h}} Q_{X_\ell|Z_\ell,\ldots,Z_{\mathsf{h}}}$; and
3. for $1 \le i < \mathsf{h}$, $S \subseteq \{i+1, \ldots, \mathsf{h}\}$, and $T \subseteq \{2, \ldots, \mathsf{h}\}$,

$$\mathsf{R}_i \ge I(X_1; Z_{i+1}, \ldots, Z_{\mathsf{h}}),$$

$$\mathsf{R}_c + \mathsf{R}_i + \sum_{s\in S} \rho_s \ge I(X_1, \ldots, X_{\mathsf{h}}; X_S, Z_{i+1}, \ldots, Z_{\mathsf{h}}),$$

$$\mathsf{R}_c + \sum_{s\in T\cup\{1\}} \rho_s \ge I(X_2, \ldots, X_{\mathsf{h}}; Z_2, \ldots, Z_{\mathsf{h}}, X_T|X_1),$$

Proof: We begin with the achievability part. Given $Q_{X_1\cdots X_{\mathsf{h}}Z_2Z_3\cdots Z_{\mathsf{h}}}$, consider the achievable scheme of Section IV-B with the following choices:

$$A_{i,j} \triangleq \text{constant}, \quad 1 < i < j \le \mathsf{h}, \tag{58a}$$

$$B_{i,i+1} \triangleq \text{constant}, \quad 1 \le i < \mathsf{h}, \tag{58b}$$

$$A_{1,\ell} \triangleq Z_\ell, \quad 1 < \ell < \mathsf{h}, \tag{58c}$$

$$C_\ell \triangleq X_\ell, \quad 1 < \ell < \mathsf{h}, \tag{58d}$$

$$\lambda_\ell \triangleq H(X_\ell|Z_2, \ldots, Z_{\mathsf{h}}), \quad 1 < \ell < \mathsf{h}. \tag{58e}$$

For the above choice of auxiliary variables, note that the decomposition in condition 2 of Theorem 3 is in accordance with the joint pmf structure in (28). Using the above assignments and constraints in (55) for the functional setting, we see that the following rate region is achievable.

$$\mathsf{R}_c \ge \mu^{-}_{1,2} + \ldots + \mu^{-}_{1,\mathsf{h}},$$
$$\mathsf{R}_i \ge \mu^{+}_{1,i+1} + \ldots + \mu^{+}_{1,\mathsf{h}}, \quad 1 \le i < \mathsf{h},$$
$$\rho_i \ge \begin{cases} \mu^{+}_{1,2} + \ldots + \mu^{+}_{1,\mathsf{h}} - I(X_1; Z_2, \ldots, Z_{\mathsf{h}}) & i = 1 \\ H(X_i|Z_i, \ldots, Z_{\mathsf{h}}) & i > 1, \end{cases}$$

where the code parameters $\mu^{+}_{1,2}, \ldots, \mu^{+}_{1,\mathsf{h}}, \mu^{-}_{1,2}, \ldots, \mu^{-}_{1,\mathsf{h}}$ can take non-negative values satisfying the following conditions derived in Theorems 1 and 2:

$$\sum_{k=i}^{\mathsf{h}} (\mu^{+}_{1,i} + \mu^{-}_{1,i}) \ge I(X_1, \ldots, X_{\mathsf{h}}; Z_i, \ldots, Z_{\mathsf{h}}), \quad 1 < i \le \mathsf{h},$$

$$\sum_{k=i}^{\mathsf{h}} \mu^{+}_{1,i} \ge I(X_1; Z_i, \ldots, Z_{\mathsf{h}}), \quad 1 < i \le \mathsf{h}.$$

Now, applying the rate-transfer arguments in Lemma 2 to the above region, we see that the achievable region includes the following region.

$$\mathsf{R}_c \ge \mu^{-}_{1,2} + \ldots + \mu^{-}_{1,\mathsf{h}} + \delta_1 + \ldots + \delta_h,$$

$$\mathsf{R}_i \ge \sum_{k=i+1}^{\mathsf{h}} (\mu^{+}_{1,k} + \varepsilon_k), \quad 1 \le i < \mathsf{h},$$

$$\rho_1 \ge \sum_{k=2}^{\mathsf{h}} (\mu^{+}_{1,k} + \varepsilon_k) - I(X_1; Z_2, \ldots, Z_{\mathsf{h}}) - \delta_1,$$

$$\rho_i \ge H(X_i|Z_i, \ldots, Z_{\mathsf{h}}) - \delta_i - \varepsilon_i, \quad 1 < i \le \mathsf{h},$$

$$\sum_{k=i}^{\mathsf{h}} (\mu^{+}_{1,i} + \mu^{-}_{1,i}) \ge I(X_1, \ldots, X_{\mathsf{h}}; Z_i, \ldots, Z_{\mathsf{h}}), \quad 1 < i \le \mathsf{h},$$

$$\sum_{k=i}^{\mathsf{h}} \mu^{+}_{1,i} \ge I(X_1; Z_i, \ldots, Z_{\mathsf{h}}), \quad 1 < i \le \mathsf{h},$$

where in addition to the non-negativity constraints of the code parameters, we also impose

$$\delta_j \ge 0, \quad 1 \le j \le \mathsf{h},$$
$$\varepsilon_j \ge 0, \quad 1 < j \le \mathsf{h}.$$

In the above, δ_i denotes the portion of common randomness that is used only by Node i as its local randomness (see the first rate-transfer condition of Lemma 2), and ε_i denotes the portion of local randomness of Node 1 that is communicated to Node i to be used as its local randomness (see the second rate-transfer condition of Lemma 2). Finally, an application of Fourier-Motzkin elimination [28, Appendix D] to dispose of the code and rate-transfer parameters yields the required result.

Now, for the converse part, suppose that rate tuple $\mathbf{R} \triangleq (\mathsf{R}_c, \mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1}, \rho_1, \rho_2, \ldots, \rho_{\mathsf{h}})$ is achievable. Fix $\varepsilon > 0$ and an ε-code of length n operating at $\mathbf{R}$ that outputs $\hat{X}_i^n$ at Node i, $i > 1$. Then,

$$\left\|\mathsf{Q}_{X_1}^{\otimes n} Q_{\hat{X}_2^n \cdots \hat{X}_{\mathsf{h}}^n | X_1^n} - \mathsf{Q}_{X_1 \cdots X_{\mathsf{h}}}^{\otimes n}\right\|_1 \leq \varepsilon. \tag{61}$$

For notational ease, denote $\hat{X}_1^n \triangleq X_1^n$. Since (61) holds, we infer from [16, Sec. V.A] that for any $S \subseteq \{1, \ldots, \mathsf{h}\}$ and $i \in \{1, \ldots, \mathsf{h}\}$,

$$H(\{\hat{X}_j^n\}_{j \in S}) \geq \sum_{k=1}^{n} H(\{\hat{X}_{j,k}\}_{j \in S}) - n\delta'_{n,\varepsilon}, \tag{62}$$

$$H(\{\hat{X}_j^n\}_{j \in S} | \hat{X}_i^n) \geq \sum_{k=1}^{n} H(\{\hat{X}_{j,k}\}_{j \in S} | \hat{X}_{i,k}) - n\delta'_{n,\varepsilon}, \tag{63}$$

for some $\delta'_{n,\varepsilon} \to 0$ as $\varepsilon \to 0$. Then, for any $i \in [\![1, \mathsf{h}-1]\!]$,

$$\begin{aligned}
n\mathsf{R}_i &\geq H(\mathsf{I}_i) \geq H(\mathsf{I}_i | \mathsf{M}_c) \\
&\geq I(\hat{X}_1^n; \mathsf{I}_i | \mathsf{M}_c) \\
&\overset{\text{(a)}}{=} I(\hat{X}_1^n; \mathsf{I}_i, \ldots, \mathsf{I}_{\mathsf{h}-1} | \mathsf{M}_c) \\
&\overset{\text{(b)}}{=} I(\hat{X}_1^n; \mathsf{M}_c, \mathsf{I}_i, \ldots, \mathsf{I}_{\mathsf{h}-1}) \\
&\overset{\text{(c)}}{=} \sum_{k=1}^{n} I(\hat{X}_{1,k}; \mathsf{M}_c, \{\mathsf{I}_j\}_{j=i}^{\mathsf{h}-1}, \hat{X}_1^{k-1}) \\
&\overset{\text{(d)}}{=} \sum_{k=1}^{n} I(\hat{X}_{1,k}; \{Z_{j,k}\}_{j=i+1}^{\mathsf{h}}) \\
&\overset{\text{(e)}}{=} nI\big(\hat{X}_{1,U}; \{Z_{j,U}\}_{j=i+1}^{\mathsf{h}} \mid U\big) \\
&\overset{\text{(f)}}{=} nI\big(\hat{X}_{1,U}; \{Z_j^*\}_{j=i+1}^{\mathsf{h}}\big),
\end{aligned}$$

where

- (a) follows from the fact that in functional mode, local randomness at intermediate nodes is not used for generating next-hop messages. Hence, I_{k+1} is a function of only M_c and I_k for any $k > 1$.
- (b) because $\hat{X}_1^n = X_1^n$, and M_c are independent;
- (c) because $\hat{X}_1^n = X_1^n$ is i.i.d.;
- (d) by defining $Z_{j,k} \triangleq (\mathsf{M}_c, \mathsf{I}_{j-1}, \hat{X}_1^{k-1})$ for $2 \leq j \leq \mathsf{h}$ and $1 \leq k \leq n$;
- (e) by defining time-sharing RV $U \sim \mathsf{unif}([\![1, n]\!])$; and
- (f) by setting $Z_j^* \triangleq (U, Z_{j,U})$ for $2 \leq j \leq \mathsf{h}$, and since $\hat{X}_{1,U}$ and U are independent.

Lastly, we need to ensure that

$$\hat{X}_{1,U} \leftrightarrow \{Z_k^*\}_{k=2}^{\mathsf{h}} \leftrightarrow (\hat{X}_{2,U}, \ldots, \hat{X}_{\mathsf{h},U}), \tag{64}$$

and for $1 < i \leq \mathsf{h}$,

$$\hat{X}_{i,U} \leftrightarrow \{Z_k^*\}_{k=i}^{\mathsf{h}} \leftrightarrow (\{Z_j^*\}_{j=2}^{i-1}, \{\hat{X}_{j,U} : j \neq i\}). \tag{65}$$

To do that, note that

$$\begin{aligned}
0 &\overset{(a)}{=} I(\hat{X}_1^n, \mathsf{M}_c, \mathsf{M}_{L_1}; \mathsf{M}_{L_2}, \ldots, \mathsf{M}_{L_{\mathsf{h}}}) \\
&\overset{(b)}{=} I(\hat{X}_1^n, \mathsf{M}_c, \mathsf{M}_{L_1}, \mathsf{I}_1, \ldots, \mathsf{I}_{\mathsf{h}-1}; \mathsf{M}_{L_2}, \ldots, \mathsf{M}_{L_{\mathsf{h}}}) \\
&\geq I(\hat{X}_1^n; \{\mathsf{M}_{L_\ell}\}_{\ell=2}^{\mathsf{h}} | \mathsf{M}_c, \mathsf{I}_1, \ldots, \mathsf{I}_{\mathsf{h}-1}) \\
&= \sum_{k=1}^{n} I(X_{1,k}; \{\mathsf{M}_{L_\ell}\}_{\ell=2}^{\mathsf{h}} | \mathsf{M}_c, \mathsf{I}_1, \ldots, \mathsf{I}_{\mathsf{h}-1}, \hat{X}_1^{k-1}) \\
&\overset{(c)}{=} \sum_{k=1}^{n} I(\hat{X}_{1,k}; \{\mathsf{M}_{L_\ell}, \hat{X}_\ell^n\}_{\ell=2}^{\mathsf{h}} | \mathsf{M}_c, \{\mathsf{I}_\ell\}_{\ell=1}^{\mathsf{h}-1}, \hat{X}_1^{k-1}) \\
&\geq \sum_{k=1}^{n} I(\hat{X}_{1,k}; \hat{X}_{2,k}, \ldots, \hat{X}_{\mathsf{h},k} | \mathsf{M}_c, \{\mathsf{I}_\ell\}_{\ell=1}^{\mathsf{h}-1}, \hat{X}_1^{k-1}) \\
&= nI(\hat{X}_{1,U}; \hat{X}_{2,U}, \ldots, \hat{X}_{\mathsf{h},U} | \{Z_k^*\}_{k=2}^{\mathsf{h}}),
\end{aligned}$$

where in (a) we use the independence of the local randomness of Nodes $1, \ldots \mathsf{h}$, common randomness shared by all nodes, and the action specified at Node 1; in (b) we use the fact that I_1 is a function of $(\mathsf{M}_c, \mathsf{M}_{L_1}, \hat{X}_1^n)$, and for any $k = 1, \ldots, \mathsf{h}-2$, I_{k+1} is a function of $(\mathsf{M}_c, \mathsf{I}_k)$ (and by induction, a function of M_c, M_{L_1}, and $\hat{X}_1^n$) in the functional mode; and in (c) we use the fact that $\hat{X}_j^n$ is a function of $(\mathsf{M}_c, \mathsf{M}_{L_j}, \mathsf{I}_{j-1})$ for $j > 1$. Similarly, for $i > 1$ and $k \in \{1, \ldots, n\}$,

$$\begin{aligned}
0 &= I(\mathsf{M}_{L_i}; \mathsf{M}_c, \{\mathsf{M}_{L_\ell}\}_{\ell=1}^{i-1}, \{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^{\mathsf{h}}, \hat{X}_1^n) \\
&\overset{(a)}{\geq} I(\mathsf{M}_{L_i}; \mathsf{M}_c, \{\mathsf{I}_\ell\}_{\ell=1}^{\mathsf{h}-1}, \mathsf{M}_{L_1}, \hat{X}_1^{k-1}, \{\hat{X}_{j,k} : j \neq i\}) \\
&\geq I(\mathsf{M}_{L_i}; \{\mathsf{I}_\ell\}_{\ell=1}^{i-2}, \{\hat{X}_{j,k} : j \neq i\} \,|\, \mathsf{M}_c, \{\mathsf{I}_\ell\}_{\ell=i-1}^{\mathsf{h}-1}, \hat{X}_1^{k-1}) \\
&\overset{(b)}{\geq} I(\hat{X}_{i,k}; \{\mathsf{I}_\ell\}_{\ell=1}^{i-2}, \{\hat{X}_{j,k} : j \neq i\} \,|\, \mathsf{M}_c, \{\mathsf{I}_\ell\}_{\ell=i-1}^{\mathsf{h}-1}, \hat{X}_1^{k-1}),
\end{aligned}$$

where in (a) we first use: (i) for $j \geq 1$, I_j is a function of $(\mathsf{M}_c, \mathsf{M}_{L_1}, \hat{X}_1^n)$ in the functional mode, and (ii) $\hat{X}_j^n$ is a function of $(\mathsf{M}_c, \mathsf{I}_{j-1}, \mathsf{M}_{L_j})$ for $j > 1$, and then drop appropriate terms from the mutual information term. Similarly, in (b), we introduce $\hat{X}_{i,k}$ since it is a function of $(\mathsf{M}_c, \mathsf{I}_{i-1}, \mathsf{M}_{L_i})$, and then drop M_{L_i}. Finally, summing over all k yields $I(\hat{X}_{i,U}; \{Z_j^*\}_{j=2}^{i-1}, \{\hat{X}_{j,U} : j \neq i\} \mid \{Z_k^*\}_{k=i}^{\mathsf{h}}) = 0$ for $i > 1$, and consequently, the chain in (65).

Next, pick $i \in \{1, \ldots, \mathsf{h}-1\}$ and subset $S \subseteq [\![i+1, \mathsf{h}]\!]$. Now, consider (66)-(67) shown at the top of the next page, where

- (a) follows due to functional mode of operation, and because $H(\hat{X}_j^n | \mathsf{M}_c, \mathsf{I}_{j-1}, \mathsf{M}_{L_j}) = 0$ for $j > 1$;
- (b) follows by dropping $\{\mathsf{M}_{L_s}\}_{s \in S}$ from the mutual information term;
- (c) uses the chain rule, and then (62) to get rid of the conditioning;
- (d) follows by introducing the uniform time-sharing RV U; and lastly,
- (e) follows by defining $\delta''_{n,\varepsilon} \triangleq \delta'_{n,\varepsilon} + I(\{\hat{X}_{\ell,U}\}_{\ell=1}^{\mathsf{h}}; U)$, and by noting that (61) ensures that $(\hat{X}_1^n, \ldots, \hat{X}_{\mathsf{h}}^n)$ are nearly i.i.d., we can invoke [16, eq. (5)] to infer

$$I(\hat{X}_{1,U}, \ldots, \hat{X}_{1,U}; U) \overset{\varepsilon \to 0}{\longrightarrow} 0. \tag{70}$$

$$
\begin{aligned}
n\big(\mathsf{R}_c + \mathsf{R}_i + \sum_{s\in S} \rho_s\big) &\geq H(\mathsf{M}_c, \mathsf{I}_i, \{\mathsf{M}_{L_s}\}_{s\in S}) && (66)\\
&\overset{(a)}{=} H(\mathsf{M}_c, \mathsf{I}_i, \dots, \mathsf{I}_{\mathsf{h}-1}, \{\mathsf{M}_{L_s}\}_{s\in S}\{\hat{X}_s^n\}_{s\in S})\\
&\geq I(\hat{X}_1^n, \dots, \hat{X}_{\mathsf{h}}^n; \mathsf{M}_c, \mathsf{I}_i, \dots, \mathsf{I}_{\mathsf{h}-1}, \{\mathsf{M}_{L_s}\}_{s\in S}, \{\hat{X}_s^n\}_{s\in S})\\
&\overset{(b)}{\geq} I(\hat{X}_1^n, \dots, \hat{X}_{\mathsf{h}}^n; \mathsf{M}_c, \mathsf{I}_i, \dots, \mathsf{I}_{\mathsf{h}-1}, \{\hat{X}_s^n\}_{s\in S})\\
&\overset{(c)}{\geq} \sum_{k=1}^{n} I\big(\{\hat{X}_{\ell,k}\}_{\ell=1}^{\mathsf{h}}; \mathsf{M}_c, \mathsf{I}_i, \dots, \mathsf{I}_{\mathsf{h}-1}, \{\hat{X}_\ell^{k-1}\}_{\ell=1}^{\mathsf{h}}, \{\hat{X}_s^n\}_{s\in S}\big) - n\delta'_{n,\varepsilon}\\
&\geq \sum_{k=1}^{n} I\big(\{\hat{X}_{\ell,k}\}_{\ell=1}^{\mathsf{h}}; \{Z_{j,k}\}_{j=i+1}^{\mathsf{h}}, \{\hat{X}_{s,i}\}_{s\in S}\big) - n\delta'_{n,\varepsilon}\\
&\overset{(d)}{=} nI\big(\{\hat{X}_{\ell,U}\}_{\ell=1}^{\mathsf{h}}; \{Z_{j,U}\}_{j=i+1}^{\mathsf{h}}, \{\hat{X}_{s,U}\}_{s\in S} \mid U\big) - n\delta'_{n,\varepsilon}\\
&\overset{(e)}{\geq} nI\big(\{\hat{X}_{\ell,U}\}_{\ell=1}^{\mathsf{h}}; \{Z_j^*\}_{j=i+1}^{\mathsf{h}}, \{\hat{X}_{s,U}\}_{s\in S}\big) - n\delta''_{n,\varepsilon}. && (67)
\end{aligned}
$$

$$
\begin{aligned}
n\big(\mathsf{R}_c + \rho_1 + \sum_{s\in S} \rho_s\big) &\geq H(\mathsf{M}_c, \mathsf{M}_{L_1}\{\mathsf{M}_{L_s}\}_{s\in S}) && (68)\\
&\geq I(\hat{X}_1^n, \dots, \hat{X}_{\mathsf{h}}^n; \mathsf{M}_c, \mathsf{M}_{L_1}\{\mathsf{M}_{L_s}\}_{s\in S})\\
&\overset{(a)}{=} I(\{\hat{X}_\ell^n\}_{\ell=2}^{\mathsf{h}}; \mathsf{M}_c, \mathsf{M}_{L_1}, \{\mathsf{M}_{L_s}\}_{s\in S}|\hat{X}_1^n)\\
&\overset{(b)}{=} I(\{\hat{X}_\ell^n\}_{\ell=2}^{\mathsf{h}}; \mathsf{M}_c, \mathsf{M}_{L_1}, \{\mathsf{M}_{L_s}\}_{s\in S}, \{\mathsf{I}_\ell\}_{\ell=1}^{\mathsf{h}-1}|\hat{X}_1^n)\\
&\overset{(c)}{=} I(\{\hat{X}_\ell^n\}_{\ell=2}^{\mathsf{h}}; \mathsf{M}_c, \mathsf{M}_{L_1}, \{\mathsf{M}_{L_s}\}_{s\in S}, \{\mathsf{I}_\ell\}_{\ell=1}^{\mathsf{h}-1}, \{\hat{X}_s^n\}_{s\in S}|\hat{X}_1^n)\\
&\overset{(d)}{\geq} \sum_{k=1}^{n} I(\{\hat{X}_{\ell,k}\}_{\ell=2}^{\mathsf{h}}; \mathsf{M}_c, \mathsf{M}_{L_1}, \{\mathsf{I}_\ell\}_{\ell=1}^{\mathsf{h}-1}, \{\hat{X}_s^n\}_{s\in S}, \{\hat{X}_\ell^{k-1}\}_{\ell=1}^{\mathsf{h}}, \hat{X}_1^n|\hat{X}_{1,k}) - n\delta'_{n,\varepsilon}\\
&\geq \sum_{k=1}^{n} I(\{\hat{X}_{\ell,k}\}_{\ell=1}^{\mathsf{h}}; \mathsf{M}_c, \{\mathsf{I}_\ell\}_{\ell=1}^{\mathsf{h}-1}, \{\hat{X}_{s,k}\}_{s\in S}, \hat{X}_1^{k-1}|\hat{X}_{1,k}) - n\delta'_{n,\varepsilon}\\
&\overset{(e)}{\geq} \sum_{k=1}^{n} I(\{\hat{X}_{\ell,k}\}_{\ell=1}^{\mathsf{h}}; \{Z_{j,k}\}_{j=2}^{\mathsf{h}}, \{\hat{X}_{s,k}^n\}_{s\in S}|\hat{X}_{1,k}) - n\delta'_{n,\varepsilon}\\
&\overset{(f)}{\geq} nI(\{\hat{X}_{\ell,U}\}_{\ell=1}^{\mathsf{h}}; \{Z_{j,U}\}_{j=2}^{\mathsf{h}}, \{\hat{X}_{s,U}^n\}_{s\in S}|\hat{X}_{1,U}, U) - n\delta'_{n,\varepsilon}\\
&\overset{(g)}{\geq} nI(\{\hat{X}_{\ell,U}\}_{\ell=1}^{\mathsf{h}}; \{Z_j^*\}_{j=2}^{\mathsf{h}}, \{\hat{X}_{s,U}^n\}_{s\in S}|\hat{X}_{1,U}) - n\tilde{\delta}_{n,\varepsilon}, && (69)
\end{aligned}
$$

Next, for $S \subseteq \{2, \dots, \mathsf{h}\}$, consider (68)-(69) shown at the top of this page, where

(a) follows because common and local randomness are independent of the action specified at Node 1;

(b) holds since $H(\mathsf{I}_1|\,\mathsf{M}_c, \mathsf{M}_{L_1}, \hat{X}_1^n) = 0$, and due to the functional mode of operation at other nodes;

(c) follows because $H(\hat{X}_j^n|\mathsf{M}_c, \mathsf{M}_{L_j}, \mathsf{I}_{j-1}) = 0$ for $j > 1$;

(d) uses the chain rule, then drops $\{\mathsf{M}_{L_s}\}_{s\in S}$, and lastly employs (63) to get rid of the conditioning;

(e) follows by introducing the auxiliary RVs defined earlier;

(f) follows by introducing the uniform time-sharing RV U; and lastly,

(g) follows by defining $\tilde{\delta}_{n,\varepsilon} \triangleq \delta'_{n,\varepsilon} + I(\{\hat{X}_{\ell,U}\}_{\ell=1}^{\mathsf{h}}; U|\hat{X}_{1,U})$, and from (70), we see that

$$
I(\{\hat{X}_{\ell,U}\}_{\ell=1}^{\mathsf{h}}; U|\hat{X}_{1,U}) \leq I(\{\hat{X}_{\ell,U}\}_{\ell=1}^{\mathsf{h}}; U) \overset{\varepsilon\to 0}{\longrightarrow} 0. \quad (71)
$$

Lastly, since we have the correct structure for the auxiliary RVs, we can just as in [3], restrict their cardinalities using Carathéodory's theorem. The proof is then complete by limiting $\varepsilon \to 0$, by invoking the continuity of the information functional, the compactness of the space of joint pmfs of the actions and the auxiliary RVs, and from the facts that $\delta''_{n,\varepsilon} \to 0$ and $\tilde{\delta}_{n,\varepsilon} \to 0$. ■

Remark 5: Unlike usual source-coding outer-bound proofs, the joint pmf of $(\hat{X}_{1,U}, \dots, \hat{X}_{\mathsf{h},U})$ *depends* on ε. However, $\big\|Q_{\hat{X}_{1,U},\dots,\hat{X}_{\mathsf{h},U}} - \mathsf{Q}_{X_1\cdots X_{\mathsf{h}}}\big\|_1 \to 0$ as $\varepsilon \to 0$. ■

The above result automatically yields the trade-offs between common randomness rate and communication rates when local randomness is absent at Nodes $2, \dots, \mathsf{h}$, since in this case, intermediate nodes can only operate in is the functional mode. In this case, Theorem 3 yields the following.

Remark 6: When $\rho_2 = \cdots = \rho_{\mathsf{h}} = 0$, the trade-offs between common randomness rate and the rates of communication are given by:

$$
\begin{aligned}
\mathsf{R}_\ell &\geq I(X_1; X_{\ell+1}, \dots, X_{\mathsf{h}}), \quad 1 \leq \ell < \mathsf{h}, && \text{(72a)}\\
\mathsf{R}_c + \mathsf{R}_\ell &\geq H(X_{\ell+1}, \dots, X_{\mathsf{h}}), \quad 1 \leq \ell < \mathsf{h}, && \text{(72b)}\\
\mathsf{R}_c + \rho_1 &\geq H(X_2, \dots, X_{\mathsf{h}}|X_1). && \text{(72c)}
\end{aligned}
$$

■

We conclude this section with the following result that argues that when the common randomness shared by all the nodes is sufficiently large, it is sufficient to focus on strong coordination schemes where intermediate nodes operate in the functional mode. Hence, the trade-offs defined by Theorem 3 approximate the strong coordination capacity region at large rates of common randomness.

Theorem 4: Suppose that the rate of common randomness available to all nodes is large, i.e., $\mathsf{R}_c > H(X_2, \ldots, X_{\mathsf{h}}|X_1)$. Then, the requirements for local randomness rates and communication rates are decoupled, and are given by:

$$\rho_i \geq 0, \quad i = 1, \ldots, \mathsf{h}, \tag{73a}$$
$$\mathsf{R}_i \geq I(X_1; X_{i+1}, \ldots, X_{\mathsf{h}}), \quad i = 1, \ldots, \mathsf{h}-1. \tag{73b}$$

Further, in this regime, it suffices to focus on functional schemes alone.

Proof: For the achievability, consider the functional scheme of Theorem 3 with the following choices:

$$A_{1,i} = X_i, \quad i = 2, \ldots, \mathsf{h}, \tag{74a}$$
$$\mu^+_{1,i} = \begin{cases} I(X_1; X_i|X_{i+1}, \ldots, X_{\mathsf{h}}), & 1 < i < \mathsf{h} \\ I(X_1; X_{\mathsf{h}}), & i = \mathsf{h}, \end{cases} \tag{74b}$$
$$\mu^-_{1,i} = H(X_i|X_{i+1}, \ldots, X_{\mathsf{h}}, X_1), \quad i = 2, \ldots, \mathsf{h}. \tag{74c}$$

From (55), we see that the least rate of common randomness required for this achievable scheme is $\sum_{i=2}^{\mathsf{h}} \mu^-_{1,i} = H(X_2, \ldots, X_{\mathsf{h}}|X_1)$, and the communication rate between Node i and Node $i+1$ is $I(X_1; X_{i+1}, \ldots, X_{\mathsf{h}})$. Further, for this choice, we do not require Nodes $1, \ldots, \mathsf{h}$ to possess any local randomness, thus establishing the achievability of the region claimed.

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 $(X_{i+1}^n, \ldots, X_{\mathsf{h}}^n)$. For the one-hop network, the smallest rate of communication is $I(X_1; X_{i+1}, \ldots, X_{\mathsf{h}})$ [3, Th. 10]. ■

B. Markov Actions

In this part, we investigate the case where the h actions form a Markov chain that is aligned with the network topology, i.e.,

$$X_1 \leftrightarrow X_2 \leftrightarrow \cdots \leftrightarrow X_{\mathsf{h}-1} \leftrightarrow X_{\mathsf{h}}. \tag{75}$$

For this specific class of actions, while we do not have a complete characterization of the optimal trade-offs among the common and local randomness at various nodes and the communication rates on the links, we derive two partial characterizations that correspond to the two extreme cases of common randomness rates. More specifically, we quantify the required rates of local randomness and the communication rates on each link of the network when:

(a) common randomness is *sufficiently large*, i.e., $\mathsf{R}_c > H(X_2, \ldots, X_{\mathsf{h}}|X_1)$; and
(b) common randomness is absent;

The first setting when the common randomness is sufficiently large is a direct consequence of Theorem 4, and the corresponding result for Markov actions is as follows.

Remark 7: The requirements on local randomness rates and communication rates when the rate of common randomness $\mathsf{R}_c > H(X_2, \ldots, X_{\mathsf{h}}|X_1)$ are given by:

$$\rho_i \geq 0, \quad i = 1, \ldots, \mathsf{h}, \tag{76a}$$
$$\mathsf{R}_i \geq I(X_1; X_{i+1}), \quad i = 1, \ldots, \mathsf{h}-1. \tag{76b}$$

Further, it suffices to use functional schemes to achieve the above lower bounds. ■

The following is the main result of this section that quantifies the trade-offs in the absence of common randomness.

Theorem 5: Strong coordination is achievable at local randomness rates $(\rho_1, \ldots, \rho_{\mathsf{h}})$ and communication rates $(\mathsf{R}_1, \ldots, \mathsf{R}_{\mathsf{h}-1})$ and zero common randomness rate (i.e., in the absence of common randomness) provided there exist auxiliary RVs $Z_1, \ldots, Z_{\mathsf{h}-1}$ such that

$$X_1 \leftrightarrow Z_1 \leftrightarrow X_2 \leftrightarrow Z_2 \leftrightarrow \cdots \leftrightarrow X_{\mathsf{h}-1} \leftrightarrow Z_{\mathsf{h}-1} \leftrightarrow X_{\mathsf{h}},$$

and for each $1 \leq i \leq j \leq \mathsf{h}$,

$$\mathsf{R}_i + \sum_{k=i+1}^{j} \rho_k \geq \left[\begin{array}{r} H(\{X_\ell\}_{\ell=i+1}^{j}|X_i) + I(X_i; Z_i) \\ + I(X_{j+1}; Z_j|X_j) \end{array} \right], \tag{77a}$$
$$\sum_{k=1}^{j} \rho_k \geq H(\{X_\ell\}_{\ell=2}^{j}|X_1) + I(X_{j+1}; Z_j|X_j). \tag{77b}$$

Note that in (77), $X_{\mathsf{h}+1}$ and Z_{h} are viewed as constant RVs.

Proof: For the achievable part of the proof, pick $Q_{X_1,\ldots,X_{\mathsf{h}},Z_1,\ldots,Z_{\mathsf{h}-1}}$ such that the above Markov chain holds, and $Q_{X_1,\ldots,X_{\mathsf{h}}} = \mathsf{Q}_{X_1,\ldots,X_{\mathsf{h}}}$. To build a code using this joint pmf, we adapt the code design of Section IV-B with the following assignments:

$$A_{i,j} \triangleq \text{constant}, \quad 1 \leq i < j \leq \mathsf{h}, \tag{78a}$$
$$B_{i,i+1} \triangleq Z_i, \quad 1 \leq i < \mathsf{h}, \tag{78b}$$
$$C_i \triangleq X_i, \quad 1 < i \leq \mathsf{h}. \tag{78c}$$

such that the joint pmf of actions and the auxiliary RVs are:

$$Q_{X_1} Q_{B_{1,2}|X_1} \prod_{j=2}^{\mathsf{h}} \left(Q_{C_j|B_{j-1,j}} Q_{X_j|B_{j-1,j}C_j} Q_{B_{j,j+1}|X_j} \right),$$

with $Q_{X_1,\ldots,X_{\mathsf{h}}} = \mathsf{Q}_{X_1,\ldots,X_{\mathsf{h}}}$. Note that this assignment meets the decomposition specified in (28). Now, from the analysis in Section IV-B, and specifically from Theorem 2 we see that we can build a strong coordination code with the following codebook parameters.

$$(\mu^+_{i,j}, \mu^-_{i,j}) \triangleq (0, 0), \quad 1 \leq i < j \leq \mathsf{h},$$
$$(\kappa^+_i, \kappa^-_i) \triangleq \left(I(X_i X_{i+1}; B_{i,i+1}), 0 \right), \quad 1 \leq i < \mathsf{h},$$
$$\lambda_i \triangleq I(X_{i-1} X_i; C_i|B_{i-1,i}), \quad 1 < i \leq \mathsf{h}.$$

Now, using the assignments (for the unrestricted mode of intermediate-node operation) in Section IV-B5, we infer that a code can be built with the following common, local and communication rates.

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

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.

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

where the rate-transfer variables δ_i's are subject to non-negativity constraints. An application of Fourier-Motzkin elimination to dispose of the rate-transfer variables yields the requisite rate region.

Now, to prove the converse, suppose that there exists a scheme requiring a local randomness rate of ρ_i at Node i and a communication rate of R_i from Node i to Node $i+1$ such that the joint pmf of the actions satisfies

$$\left\| Q_{\hat{X}_1^n \cdots \hat{X}_\mathsf{h}^n} - \mathsf{Q}^{\otimes n}_{X_1 \cdots X_\mathsf{h}} \right\|_1 \le \varepsilon, \tag{84}$$

where $\hat{X}_1^n = X_1^n$ is the action specified at Node 1. Then, for any $1 \le j < \mathsf{h}$, the argument in (81)-(82) shown at the top of the next page holds, where

(a) follows because I_1 is a function of $(\hat{X}_1^n, \mathsf{M}_{L_1})$ in the absence of common randomness, 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 (63), since the actions are nearly i.i.d.;

(d) by defining $Y_j \triangleq \mathsf{I}_j$;

(e) by introducing a time-sharing variable U that is uniform over $\{1, \ldots, n\}$;

(f) from by setting $\bar{Y}_j \triangleq (Y_j, U)$ and defining

$$\tilde{\delta}_{n,\varepsilon} \triangleq \delta'_{n,\varepsilon} + I(\hat{X}_{2,U}, \ldots, \hat{X}_{\mathsf{h},U}; U|X_{1,U}), \tag{85}$$

which due to (71), is guaranteed to vanish as we let $\varepsilon \to 0$; and finally,

(g) follows by defining for $j = 1, \ldots, \mathsf{h}-1$,

$$\bar{\delta}_{j,n,\varepsilon} \triangleq \tilde{\delta}_{n,\varepsilon} + I(\hat{X}_{j+1,U}; \{\hat{X}_{\ell,U}\}_{\ell=1}^{j-1}|\hat{X}_{j,U}), \tag{86}$$

which, due to the Markovity of the actions (75) and Remark 5, is also guaranteed to vanish as $\varepsilon \to 0$.

While this establishes (77b) for $j < \mathsf{h}$, the argument for $j = \mathsf{h}$ follows from the same argument by interpreting I_h, $X_{\mathsf{h}+1}$, and Y_h to be constant RVs.

Now, to prove (77a), we proceed as follows. Let $1 \le i \le j < \mathsf{h}$. Then, the argument (87)-(88) shown at the top of the page after next holds, where

(a) holds because I_{k+1} is a function of $(\mathsf{I}_k, \mathsf{M}_{L_{k+1}})$ in the absence of common randomness for any $k \ge 1$.

(b) follows since $\hat{X}_\ell^n$ is a function of $(\mathsf{I}_{\ell-1}, \mathsf{M}_{L_\ell})$ for $\ell = i+1, \ldots, j$;

(c) follows from a time-sharing argument with auxiliary RVs $\{\bar{Y}_j : 1 \le j < \mathsf{h}\}$ defined previously; and

(d) follows by the use of $\bar{\delta}_{j,n,\varepsilon}$ defined in (86).

Also, as before, the proof of (77a) for $j = \mathsf{h}$ follows similarly by by setting I_h, $X_{\mathsf{h}+1}$, and Y_h as constants.

We are left to establish the Markov condition to be met by the actions and the auxiliary RVs. Note that the auxiliary RVs as they are defined do not satisfy $X_1 \leftrightarrow \bar{Y}_1 \leftrightarrow X_2 \leftrightarrow \cdots \leftrightarrow \bar{Y}_{\mathsf{h}-1} \leftrightarrow X_\mathsf{h}$. This is because our choice of the auxiliary RV $\bar{Y}_j = \mathsf{I}_j$ ensures that we have conditional independence of actions at adjacent nodes given the message communicated over the hop connecting the two nodes (i.e., $X_{j,U} \leftrightarrow \bar{Y}_j \leftrightarrow X_{j+1,U}$ for all $j = 1, \ldots, \mathsf{h}-1$); however, we are not guaranteed the conditional independence of messages conveyed on adjacent hops conditioned on the action of the node in-between (i.e., we do not have $\bar{Y}_j \leftrightarrow X_{j+1,U} \leftrightarrow \bar{Y}_{j+1}$). Note however that the information functionals in (77a) and (77b) only contain one auxiliary RV. Hence, it is possible to define a new set of auxiliary RVs that would both satisfy the long chain in the claim and preserve the information functionals. To do so, define RVs $\tilde{X}_k$, $k = 1, \ldots, \mathsf{h}$, and Z_j, $j = 1, \ldots, \mathsf{h}-1$, such that their joint pmf is given by

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

Note that we have $\tilde{X}_1 \leftrightarrow Z_1 \leftrightarrow \tilde{X}_2 \leftrightarrow \cdots \leftrightarrow Z_{\mathsf{h}-1} \leftrightarrow \tilde{X}_\mathsf{h}$, and further for $1 \le i < j \le \mathsf{h}$,

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

The proof is completed by bounding the sizes of the auxiliary RVs $\{Z_j\}_{j=1}^{\mathsf{h}-1}$, and then by letting $\varepsilon \to 0$, which ensures that $Q_{\tilde{X}_1,\ldots,\tilde{X}_\mathsf{h}} \to \mathsf{Q}_{X_1,\ldots,X_\mathsf{h}}$, and that each infinitesimal $\bar{\delta}_{j,n,\varepsilon}$, $1 \le j < \mathsf{h}$ vanishes. ■

We conclude this section on outer bounds with a short discussion on the indispensability of auxiliary RVs $\{A_{i,j} : (i,j) \in \mathcal{F}\}$ and $\{B_{i,i+1} : i = 1, \ldots \mathsf{h}-1\}$.

C. The Need for $\{A_{i,j} : (i,j) \in \mathcal{F}\}$ and $\{B_{i,i+1} : 1 \le i < \mathsf{h}\}$

So far in the capacity results derived above, we have always set $\{A_{i,j} : (i,j) \in \mathcal{F}\}$ to be constant RVs. A natural question then is: *are these RVs are even useful?* The following example should establish the need for non-trivial choices for these RVs. Suppose that $U \sim \mathsf{Bern}(0.5)$ and $V = U \oplus Z$, where $Z \sim \mathsf{Bern}(p)$ for some $0 < p < 0.5$ independent of U.

$$
\begin{aligned}
n\sum_{k=1}^{j}\rho_k &\geq H(\mathsf{M}_{L_1},\dots,\mathsf{M}_{L_j}) \qquad (81)\\
&\geq H(\mathsf{M}_{L_1},\dots,\mathsf{M}_{L_j}|\hat{X}_1^n)\\
&\geq I(\{\hat{X}_\ell^n\}_{\ell=2}^{j+1};\{\mathsf{M}_{L\ell}\}_{\ell=1}^{j}|\hat{X}_1^n)\\
&\overset{(a)}{\geq} I(\{\hat{X}_\ell^n\}_{\ell=2}^{j+1};\{\mathsf{I}_\ell\}_{\ell=1}^{j},\{\mathsf{M}_{L\ell}\}_{\ell=1}^{j}|\hat{X}_1^n)\\
&= \sum_{k=1}^{n} I(\{\hat{X}_{\ell,k}^n\}_{\ell=2}^{j+1};\{\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+1})\\
&\overset{(b)}{\geq} \sum_{k=1}^{n} I(\{\hat{X}_{\ell,k}^n\}_{\ell=2}^{j+1};\{\mathsf{I}_\ell\}_{\ell=1}^{j},\{\hat{X}_{\ell,k}\}_{\ell=2}^{j}\mid \hat{X}_1^n,\{\hat{X}_\ell^{k-1}\}_{\ell=2}^{j+1})\\
&\overset{(c)}{\geq} \sum_{k=1}^{n} I(\{\hat{X}_{\ell,k}^n\}_{\ell=2}^{j+1};\{\mathsf{I}_\ell\}_{\ell=1}^{j},\{\hat{X}_{\ell,k}\}_{\ell=2}^{j},\hat{X}_1^n,\{\hat{X}_\ell^{k-1}\}_{\ell=2}^{j+1}|\hat{X}_{1,k})-n\delta'_{n,\varepsilon}\\
&\geq \sum_{k=1}^{n} I(\{\hat{X}_{\ell,k}^n\}_{\ell=2}^{j+1};\mathsf{I}_j,\{\hat{X}_{\ell,k}\}_{\ell=2}^{j}|\hat{X}_{1,k})-n\delta'_{n,\varepsilon}\\
&\overset{(d)}{=} \sum_{k=1}^{n} I(\{\hat{X}_{\ell,k}^n\}_{\ell=2}^{j+1};Y_j,\{\hat{X}_{\ell,k}\}_{\ell=2}^{j}|\hat{X}_{1,k})-n\delta'_{n,\varepsilon}\\
&\overset{(e)}{=} nI(\{\hat{X}_{\ell,U}^n\}_{\ell=2}^{j+1};Y_j,\{\hat{X}_{\ell,U}\}_{\ell=2}^{j}|\hat{X}_{1,U},U)-n\delta'_{n,\varepsilon}\\
&\overset{(f)}{\geq} nI(\{\hat{X}_{\ell,U}^n\}_{\ell=2}^{j+1};\bar{Y}_j,\{\hat{X}_{\ell,U}\}_{\ell=2}^{j}|\hat{X}_{1,U})-n\tilde{\delta}_{n,\varepsilon}\\
&= nH(\{\hat{X}_{\ell,U}^n\}_{\ell=2}^{j}|\hat{X}_{1,U})+nI(\hat{X}_{j+1,U};\bar{Y}_j|\{\hat{X}_{\ell,U}^n\}_{\ell=1}^{j})-n\tilde{\delta}_{n,\varepsilon}\\
&\overset{(g)}{=} nH(\{\hat{X}_{\ell,U}^n\}_{\ell=2}^{j}|\hat{X}_{1,U})+nI(\hat{X}_{j+1,U};\bar{Y}_j,\{\hat{X}_{\ell,U}^n\}_{\ell=1}^{j-1}|\hat{X}_{j,U})-n\bar{\delta}_{j,n,\varepsilon}\\
&\geq nH(\{\hat{X}_{\ell,U}^n\}_{\ell=2}^{j}|\hat{X}_{1,U})+nI(\hat{X}_{j+1,U};\bar{Y}_j|\hat{X}_{j,U})-n\bar{\delta}_{j,n,\varepsilon}, \qquad (82)
\end{aligned}
$$

Consider a $\mathsf{h}=5$ setting where $X_1=X_3=X_5=U$ and $X_2=X_4=V$. Suppose that strong coordination is to be achieved with no shared randomness, and no local randomness at Node 1. From the general structure of the joint pmf of the actions and the auxiliary RVs given in (28), we infer that the following chains *must* hold.

$$X_1 \leftrightarrow (A_{1,2},A_{1,3},A_{1,4},A_{1,5},B_{1,2}) \leftrightarrow X_5$$
$$X_2 \leftrightarrow (A_{1,4},A_{1,5},A_{2,3},A_{2,4},A_{2,5},A_{3,4},A_{3,5},X_3) \leftrightarrow X_4$$

In the absence of any randomness at Node 1, the only non-trivial choice for $B_{1,2}$ and $A_{1,j}$, $j>1$, is U. Since $X_1=X_5=U$, at least one of $A_{1,2}$, $A_{1,3}$, $A_{1,4}$, $A_{1,5}$, $B_{1,2}$ must be U. Further, since $X_2=X_4=V$,

$$
\begin{aligned}
0&=H(V|A_{1,4},A_{1,5},A_{2,3},A_{2,4},A_{2,5},A_{3,4},A_{3,5},X_3)\\
&=H(U\oplus Z\mid U,A_{2,3},A_{2,4},A_{2,5},A_{3,4},A_{3,5})\\
&=H(Z\mid U,A_{2,3},A_{2,4},A_{2,5},A_{3,4},A_{3,5}).
\end{aligned}
$$

Since Z is independent of U, a non-trivial choice for $(A_{2,3},A_{2,4},A_{2,5},A_{3,4},A_{3,5})$ is a must. Though contrived, this example illustrates that in some settings allowed by the problem formulation, we do require auxiliary RVs to be *generated* at intermediate nodes to be non-trivial. It is straightforward to extend this to arbitrary hop lengths and auxiliary RVs.

The auxiliary RVs $\{A_{i,j}:(i,j)\in\mathcal{F}\}$ have a seemingly natural purpose: Node i uses $A_{i,j}$ to coordinate its actions with that of Node j. However, the need for $\{B_{i,i+1}:1\leq i<\mathsf{h}\}$ is technical, and arises from the fact that not all joint pmfs for $\{A_{i,j}:(i,j)\in\mathcal{F}\}$ can be realized via a scheme based on channel resolvability codebooks. For example, their joint pmf must satisfy the chains in (26). For simpilicity, let us focus on the following setting. Let random variables V_1 and V_2 be jointly correlated according to some Q_{V_1,V_2} that has full support and $I(V_1;V_2)>0$. Let us focus on the $\mathsf{h}=3$ setting, where: (a) actions $X_1=V_1$, $X_2=(V_1,V_2)$, and $X_3=V_2$; (b) common randomness is absent; and (c) the local randomness at each node is large, say $\rho_i>H(V_1,V_2)$, $i=1,2,3$. Since, $X_1\leftrightarrow X_2\leftrightarrow X_3$, the rates for communication required for strong coordination as specified by Theorem 5 are

$$\mathsf{R}_i\geq H(V_i),\quad i\in\{1,2\}. \qquad (90)$$

An achievable code for the corner point of the above region can be constructed by setting $A_{1,2}=A_{1,3}=A_{2,3}=$ constant, and by choosing $B_{1,2}=X_1=V_1$ and $B_{2,3}=V_2=X_3$. We will now show that it is impossible to attain the corner point $(\mathsf{R}_1,\mathsf{R}_2)=(H(V_1),H(V_2))$ by use of only $A_{1,2}$, $A_{1,3}$, $A_{2,3}$.

Using the rate expressions for the unrestricted mode of operation at Node 2 given in Section IV-B5, we see that we can build a code with common randomness rate $\mathsf{R}_c=0$, local randomness rates $\rho_1=\rho_2=\rho_3=H(V_1,V_2)$, and communication rates $\mathsf{R}_1=H(V_1)$ and $\mathsf{R}_2=H(V_2)$ with only auxiliary RVs $A_{1,2}$, $A_{1,3}$, $A_{2,3}$ if there exists a joint pmf

$$
\begin{aligned}
n\mathsf{R}_i + n\sum_{k=i+1}^{j}\rho_k &\geq H(\mathsf{I}_i, \{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^{j}) \qquad (87)\\
&\overset{(a)}{\geq} H(\{\mathsf{I}_\ell\}_{\ell=i}^{j}, \{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^{j})\\
&\geq I(\{\hat{X}_\ell^n\}_{\ell=i}^{j+1}; \{\mathsf{I}_\ell\}_{\ell=i}^{j}, \{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^{j})\\
&\geq I(\hat{X}_i^n; \{\mathsf{I}_\ell\}_{\ell=i}^{j}, \{\mathsf{M}_{L_\ell}\}_{\ell=i+1}^{j}) + 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)\\
&\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)\\
&\overset{(b)}{=} 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)\\
&\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)\\
&= \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{(62),(63)}{\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) - 2n\delta'_{n,\varepsilon}\\
&\overset{(c)}{\geq} nI(\hat{X}_{i,U}; \bar{Y}_i) + nI(\{\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}) - 2n\tilde{\delta}_{n,\varepsilon}\\
&= nI(\hat{X}_{i,U}; \bar{Y}_i) + nH(\{\hat{X}_{\ell,U}\}_{\ell=i+1}^{j}|\hat{X}_{i,U}) + nI(\hat{X}_{j+1,U}; \bar{Y}_j|\{\hat{X}_{\ell,U}\}_{\ell=i}^{j}) - 2n\tilde{\delta}_{n,\varepsilon},\\
&\overset{(d)}{\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, \{\hat{X}_{\ell,U}\}_{\ell=i}^{j-1}|\hat{X}_{j,U}) - 2\bar{\delta}_{j,n,\varepsilon}\big)\\
&\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|\hat{X}_{j,U}) - 2\bar{\delta}_{j,n,\varepsilon}\big), \qquad (88)
\end{aligned}
$$

$Q_{X_1X_2X_3A_{1,2},A_{1,3},A_{2,3}}$ such that

$$A_{1,2} \leftrightarrow A_{1,3} \leftrightarrow A_{2,3}, \qquad (91)$$

$$V_1 = X_1 \leftrightarrow (A_{1,2}, A_{1,3}) \leftrightarrow (X_2, X_3) = (V_1, V_2), \qquad (92)$$

$$V_2 = X_3 \leftrightarrow (A_{1,3}, A_{2,3}) \leftrightarrow (X_1, X_2) = (V_1, V_2). \qquad (93)$$

provided there exist codebook rates that satisfy the following conditions imposed by the unrestricted mode of operation and by Theorems 1 and 2, i.e,

$$
\begin{aligned}
\mathsf{R}_c &\overset{(49)}{=} \mu^-_{1,2} + \mu^-_{2,3} + \mu^-_{1,3} = 0, &(94a)\\
\mathsf{R}_1 &\overset{(48)}{=} \mu^+_{1,2} + \mu^+_{1,3} = H(V_1), &(94b)\\
\mathsf{R}_2 &\overset{(48)}{=} \mu^+_{2,3} + \mu^+_{1,3} = H(V_2), &(94c)\\
\mu^+_{1,3} &\geq I(V_1, V_2; A_{1,3}), &(94d)\\
\mu^+_{1,3} + \mu^+_{1,2} &\geq I(V_1, V_2; A_{1,2}, A_{1,3}), &(94e)\\
\mu^+_{1,3} + \mu^+_{2,3} &\geq I(V_1, V_2; A_{1,3}, A_{2,3}), &(94f)\\
\mu^+_{1,3} + \mu^+_{1,2} + \mu^+_{2,3} &\geq I(V_1, V_2; A_{1,2}, A_{1,3}, A_{2,3}). &(94g)
\end{aligned}
$$

Then, it must be true that

$$I(V_2; A_{1,3}|V_1) = 0, \qquad (95)$$

since

$$
\begin{aligned}
I(V_2; A_{1,2}A_{1,3}|V_1) &= I(V_1, V_2; A_{1,2}A_{1,3}V_1) - H(V_1)\\
&\overset{(92)}{=} I(V_1, V_2; A_{1,2}A_{1,3}) - H(V_1)\\
&\overset{(94b),(94e)}{\leq} \mathsf{R}_1 - H(V_1) = 0.
\end{aligned}
$$

Similarly, $I(V_1; A_{1,3}|V_2) = 0$, since

$$
\begin{aligned}
I(V_1; A_{1,3}A_{2,3}|V_2) &= I(V_1, V_2; A_{1,3}A_{2,3}V_2) - H(V_2)\\
&\overset{(93)}{=} I(V_1, V_2; A_{1,3}A_{2,3}) - H(V_2)\\
&\overset{(94c),(94f)}{\leq} \mathsf{R}_2 - H(V_2) = 0.
\end{aligned}
$$

Thus, we have $A_{1,3} \leftrightarrow V_1 \leftrightarrow V_2$ and $V_1 \leftrightarrow V_2 \leftrightarrow A_{1,3}$. Since Q_{V_1,V_2} has full support, it follows that for any $(a_{1,3}, v_1, v_2) \in \mathcal{A}_{1,3} \times \mathcal{V}_1 \times \mathcal{V}_2$, we have

$$
\begin{aligned}
Q_{A_{1,3}|V_1}(a_{1,3}|v_1) &= \frac{Q_{A_{1,3}V_1V_2}(a_{1,3}, v_1, v_2)}{\mathsf{Q}_{V_1V_2}(v_1, v_2)}\\
&= Q_{A_{1,3}|V_2}(a_{1,3}|v_2). \qquad (96)
\end{aligned}
$$

Hence, for any $(a_{1,3}, v_1) \in \mathcal{A}_{1,3} \times \mathcal{V}_1$,

$$
\begin{aligned}
Q_{A_{1,3}|V_1}(a_{1,3}|v_1) &= \sum_{v_2} Q_{A_{1,3}|V_1}(a_{1,3}|v_1)\mathsf{Q}_{V_2}(v_2)\\
&\overset{(96)}{=} \sum_{v_2} Q_{A_{1,3}|V_2}(a_{1,3}|v_2)\mathsf{Q}_{V_2}(v_2)\\
&= Q_{A_{1,3}}(a_{1,3}).
\end{aligned}
$$

Hence, $A_{1,3}$ is independent of V_1. Combining this fact with (95), we see that

$$I(V_1, V_2; A_{1,3}) = 0. \qquad (97)$$

Since (92) and (93) imply that V_1 is a function of $(A_{1,2}, A_{1,3})$, and V_2 is a function of $(A_{1,3}, A_{2,3})$, it then follows that

$$
\begin{aligned}
0 &< I(V_1; V_2) \overset{(97)}{=} I(V_1; V_2, A_{1,3}) \overset{(97)}{=} I(V_1; V_2|A_{1,3})\\
&\leq I(A_{1,2}; A_{2,3}|A_{1,3}), \qquad (98)
\end{aligned}
$$

which is a contradiction since (98) violates (91). Hence, we cannot achieve this corner point with the use of $A_{1,2}, A_{1,3}, A_{2,3}$ alone. This above argument also establishes that the corner point is not achievable using the functional mode of operation at intermediate nodes, thereby establishing the following fact.

Remark 8: The portion of the strong coordination capacity region achievable by schemes in the functional mode of operation is, in general, a *strict* subset of the strong coordination capacity region.

VI. Conclusion

In this work, we have analyzed the communication and randomness resources required to establish strong coordination over a multi-hop line network. To derive an achievable scheme, we first build an intricate multi-layer structure of channel resolvability codes to generate the actions at all the nodes, which is then appropriately inverted to obtain a strong coordination code. The resultant strong coordination code is not generally optimal (i.e., it is not known to achieve the optimal trade-offs among common randomness rate, local randomness rates, and hop-by-hop communication rates); however, it is shown to achieve the best trade-offs in several settings, including when all intermediate nodes operate under a functional regime, and when common randomness is plentiful.

The need for an intricate multi-layer scheme stems from a basic limitation in our understanding of the design of strong coordination codes for general multi-terminal problems: unlike in typicality-based schemes for a multi-user settings, where we can use joint typicality as the criterion to use a received message (at some intermediate network node) to select a codeword for transmission, we do not have a similar criterion here to translate messages from one hop to the next. In other words, we do not have a covering-type result to translate codewords of one channel resolvability codebook to another directly. However, by the operation of inverting actions to select messages, we have been able to translate between messages selecting codewords from auxiliary RV codebooks for specific patterns of correlation among the RVs. However, a general approach allowing arbitrary correlation patterns using channel resolvability codebooks is still missing.

Appendix

A. Proof of Theorem 1

Before we proceed, we first use the following notation to simplify the analysis.

$$\boldsymbol{Y} \triangleq (X_1, \ldots, X_{\mathsf{h}}),$$
$$\hat{\boldsymbol{Y}} \triangleq (\hat{X}_1, \ldots, \hat{X}_{\mathsf{h}}),$$
$$\mathsf{N} \triangleq 2^{n \sum_{(i,j)\in\mathcal{F}} (\mu^+_{i,j}+\mu^-_{i,j})}.$$

Let for any $S \subseteq \mathcal{F}$, $\mathcal{J}_S \triangleq \left\{(i,j) : \overline{\Phi}(i,j) \cap S \neq \emptyset\right\}$. Now, to find the conditions on the rates, we proceed in a fashion similar to [20] and [25]. The notation and arguments for the manipulations in (100)-(104) shown at the top of the next page are as follows:

- (101) follows by the use of the law of iterated expectations, where the inner conditional expectation denotes the expectation over all random codeword constructions $\{\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm) : \tilde{\boldsymbol{m}}^\pm \neq \boldsymbol{m}^\pm\}$ given codeword $\boldsymbol{A}^n(\boldsymbol{m}^\pm)$;
- (102) follows from Jensen's inequality; and
- (103) follows by splitting the inner sum in (102) according to the indices where $\boldsymbol{m}^\pm$ and $\tilde{\boldsymbol{m}}^\pm$ differ. Let $\Gamma(\boldsymbol{m}^\pm, \tilde{\boldsymbol{m}}^\pm) \triangleq \{s \in \mathcal{F} : m^\pm_s \neq \tilde{m}^\pm_s\}$. For any pair of indices $(\boldsymbol{m}^\pm, \tilde{\boldsymbol{m}}^\pm)$, the following hold:
  - if $(i,j) \notin \mathcal{J}_{\Gamma(\boldsymbol{m}^\pm,\tilde{\boldsymbol{m}}^\pm)}$, then $A^n_{i,j}(\boldsymbol{m}^\pm) = A^n_{i,j}(\tilde{\boldsymbol{m}}^\pm)$. This is because if $(i,j) \notin \mathcal{J}_{\Gamma(\boldsymbol{m}^\pm,\tilde{\boldsymbol{m}}^\pm)}$, then by definition, $\boldsymbol{m}^\pm_{\Phi(i,j)} = \tilde{\boldsymbol{m}}^\pm_{\Phi(i,j)}$, and both $A^n_{i,j}(\boldsymbol{m}^\pm)$ and $A^n_{i,j}(\tilde{\boldsymbol{m}}^\pm)$ denote the $A_{i,j}$ codeword corresponding to $\boldsymbol{m}^\pm_{\Phi(i,j)} = \tilde{\boldsymbol{m}}^\pm_{\Phi(i,j)}$.
  - if $(i,j) \in \mathcal{J}_{\Gamma(\boldsymbol{m}^\pm,\tilde{\boldsymbol{m}}^\pm)}$, then the random variables $A^n_{i,j}(\boldsymbol{m}^\pm)$, $A^n_{i,j}(\tilde{\boldsymbol{m}}^\pm)$ are conditionally independent given $\{A^n_{i,j}(\boldsymbol{m}^\pm) : (i,j) \notin \mathcal{J}_{\Gamma(\boldsymbol{m}^\pm,\tilde{\boldsymbol{m}}^\pm)}\}$, which by the earlier remark, is exactly the same as $\{A^n_{i,j}(\tilde{\boldsymbol{m}}^\pm) : (i,j) \notin \mathcal{J}_{\Gamma(\boldsymbol{m}^\pm,\tilde{\boldsymbol{m}}^\pm)}\}$.

Combining both, we see that

$$\boldsymbol{A}^n(\boldsymbol{m}^\pm) \leftrightarrow A^n_{\mathcal{J}^c_{\Gamma(\boldsymbol{m}^\pm,\tilde{\boldsymbol{m}}^\pm)}}(\tilde{\boldsymbol{m}}^\pm) \leftrightarrow \boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm). \tag{105}$$

Given $\boldsymbol{m}^\pm$, let $\mathcal{H}_{\boldsymbol{m}^\pm,S} \triangleq \left\{\tilde{\boldsymbol{m}}^\pm : \Gamma(\tilde{\boldsymbol{m}}^\pm, \boldsymbol{m}^\pm) = S\right\}$. Then, we see that

$$\mathbb{E}\left[\sum_{\tilde{\boldsymbol{m}}^\pm} \frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm))}{\mathsf{N}Q^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}\middle|\boldsymbol{A}^n(\boldsymbol{m}^\pm)\right] \tag{106}$$

$$= \sum_S \sum_{\tilde{\boldsymbol{m}}^\pm \in \mathcal{H}_{\boldsymbol{m}^\pm,S}} \frac{\mathbb{E}\left[Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm))\middle|\boldsymbol{A}^n(\boldsymbol{m}^\pm)\right]}{\mathsf{N}Q^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}. \tag{107}$$

Note that if $(1,\mathsf{h}) \in S$, then $\mathcal{J}_S = \mathcal{F}$ and hence $\boldsymbol{A}^n(\boldsymbol{m}^\pm)$ and $\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm)$ are independent, and hence

$$\sum_{S:(1,\mathsf{h})\in S} \sum_{\tilde{\boldsymbol{m}}^\pm \in \mathcal{H}_{\boldsymbol{m}^\pm,S}} \frac{\mathbb{E}\left[Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm))\middle|\boldsymbol{A}^n(\boldsymbol{m}^\pm)\right]}{\mathsf{N}Q^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}$$
$$= \sum_{S:(1,\mathsf{h})\in S} \sum_{\tilde{\boldsymbol{m}}^\pm \in \mathcal{H}_{\boldsymbol{m}^\pm,S}} \frac{1}{\mathsf{N}} \leq 1. \tag{108}$$

Further, when $(1,\mathsf{h}) \notin S$, then $\mathcal{J}_S \subsetneq \mathcal{F}$. Using the chain in (105), we see that when $(1,\mathsf{h}) \notin S$,

$$\sum_{\tilde{\boldsymbol{m}}^\pm \in \mathcal{H}_{\boldsymbol{m}^\pm,S}} \frac{\mathbb{E}\left[Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm))\middle|\boldsymbol{A}^n(\boldsymbol{m}^\pm)\right]}{\mathsf{N}Q^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}$$
$$= \sum_{\tilde{\boldsymbol{m}}^\pm \in \mathcal{H}_{\boldsymbol{m}^\pm,S}} \frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}_{\mathcal{J}^c_S}}(\boldsymbol{y}^n|\boldsymbol{A}^n_{\mathcal{J}^c_S}(\boldsymbol{m}^\pm))}{\mathsf{N}Q^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}. \tag{109}$$

Combining the above arguments, we see that

$$\mathbb{E}\left[\sum_{\tilde{\boldsymbol{m}}^\pm} \frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm))}{\mathsf{N}Q^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}\middle|\boldsymbol{A}^n(\boldsymbol{m}^\pm)\right] \tag{110}$$

$$\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}(Q^{(1)}_{\hat{\boldsymbol{Y}}^n} \parallel \mathsf{Q}^{\otimes n}_{\boldsymbol{Y}})\right] = \mathbb{E}\left[\sum_{\boldsymbol{y}^n}\left(\frac{\sum_{\boldsymbol{m}^\pm} Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\boldsymbol{m}^\pm))}{\mathsf{N}}\right)\log\left(\frac{\sum_{\tilde{\boldsymbol{m}}^\pm} Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm))}{\mathsf{N}\mathsf{Q}^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}\right)\right] \tag{100}$$

$$= \sum_{\boldsymbol{y}^n,\boldsymbol{m}^\pm}\mathbb{E}\left[\frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\boldsymbol{m}^\pm))}{\mathsf{N}}\,\mathbb{E}\left[\log\left(\sum_{\tilde{\boldsymbol{m}}^\pm}\frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm))}{\mathsf{N}\mathsf{Q}^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}\right)\middle|\boldsymbol{A}^n(\boldsymbol{m}^\pm)\right]\right] \tag{101}$$

$$\le \sum_{\boldsymbol{y}^n,\boldsymbol{m}^\pm}\mathbb{E}\left[\frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\boldsymbol{m}^\pm))}{\mathsf{N}}\log\left(\mathbb{E}\left[\sum_{\tilde{\boldsymbol{m}}^\pm}\frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm))}{\mathsf{N}\mathsf{Q}^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}\middle|\boldsymbol{A}^n(\boldsymbol{m}^\pm)\right]\right)\right] \tag{102}$$

$$\le \sum_{\boldsymbol{y}^n,\boldsymbol{m}^\pm}\mathbb{E}\left[\frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\boldsymbol{m}^\pm))}{\mathsf{N}}\log\left(1+\sum_{S:(1,\mathsf{h})\notin S}\frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}_{\mathcal{J}^{\mathsf{c}}_S}}(\boldsymbol{y}^n|\boldsymbol{A}^n_{\mathcal{J}^{\mathsf{c}}_S}(\boldsymbol{m}^\pm))}{\left(2^{n\sum_{s\notin S}(\mu^+_s+\mu^-_s)}\right)\mathsf{Q}^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}\right)\right] \tag{103}$$

$$\le \sum_{\boldsymbol{y}^n,\boldsymbol{a}^n} Q^{\otimes n}_{\boldsymbol{Y}\boldsymbol{A}}(\boldsymbol{y}^n,\boldsymbol{a}^n)\log\left[1+\sum_{S:(1,\mathsf{h})\notin S}\frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}_{\mathcal{J}^{\mathsf{c}}_S}}(\boldsymbol{y}^n|\boldsymbol{a}^n_{\mathcal{J}^{\mathsf{c}}_S})}{2^{n\left(\sum_{s\notin S}(\mu^+_s+\mu^-_s)\right)}\mathsf{Q}^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}\right]. \tag{104}$$

$$\le 1+\sum_{S:(1,\mathsf{h})\notin S}\sum_{\tilde{\boldsymbol{m}}^\pm\in\mathcal{H}_{\boldsymbol{m}^\pm,S}}\frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}_{\mathcal{J}^{\mathsf{c}}_S}}(\boldsymbol{y}^n|\boldsymbol{A}^n_{\mathcal{J}^{\mathsf{c}}_S}(\boldsymbol{m}^\pm))}{\mathsf{N}\mathsf{Q}^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}$$

$$= 1+\sum_{S:(1,\mathsf{h})\notin S}\frac{|\mathcal{H}_{\boldsymbol{m}^\pm,S}|\,Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}_{\mathcal{J}^c_S}}(\boldsymbol{y}^n|\boldsymbol{A}^n_{\mathcal{J}^c_S}(\boldsymbol{m}^\pm))}{\mathsf{N}\mathsf{Q}^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}$$

$$\stackrel{(a)}{\le} 1+\sum_{S:(1,\mathsf{h})\notin S}\frac{Q^{\otimes n}_{\boldsymbol{Y}|\boldsymbol{A}_{\mathcal{J}^c_S}}(\boldsymbol{y}^n|\boldsymbol{A}^n_{\mathcal{J}^{\mathsf{c}}_S}(\boldsymbol{m}^\pm))}{2^{n\sum_{s\notin S}(\mu^+_s+\mu^-_s)}\,\mathsf{Q}^{\otimes n}_{\boldsymbol{Y}}(\boldsymbol{y}^n)}, \tag{111}$$

where in (a) we use a counting argument that yields

$$|\mathcal{H}_{\boldsymbol{m}^\pm,S}| \le 2^{n\left(\sum_{s\in S}(\mu^+_s+\mu^-_s)\right)}. \tag{112}$$

Finally, the required rate conditions can be derived from (104) by splitting the outer sum depending on whether $(\boldsymbol{y}^n,\boldsymbol{a}^n)\in T^n_\varepsilon[Q_{\boldsymbol{Y}\boldsymbol{A}}]$ or not. The sum for atypical realizations in (104) is no more than

$$\mathbb{P}\left[(\boldsymbol{Y}^n,\boldsymbol{A}^n)\notin T^n_\varepsilon[Q_{\boldsymbol{Y}\boldsymbol{A}}]\right]\cdot\log\left(1+2^{\mathsf{h}^2}\eta_{\boldsymbol{Y}}^{-n}\right), \tag{113}$$

where $\eta_{\boldsymbol{Y}} = \min_{\boldsymbol{y}\in\mathsf{supp}(X_1,\ldots,X_\mathsf{h})}\mathsf{Q}_{\boldsymbol{Y}}(\boldsymbol{y})$. This term goes to zero as $n\to\infty$. The contribution from typical realizations can be made to vanish asymptotically, if for each $S\subseteq\mathcal{F}$, $\{(\mu^+_{i,j},\mu^-_{i,j}):(i,j)\in\mathcal{F}\}$ satisfy:

$$\sum_{s\notin S}(\mu^+_s+\mu^-_s) > I\left(\boldsymbol{Y};A_{\mathcal{J}^{\mathsf{c}}_S}\right) = I\left(X_1,\ldots,X_\mathsf{h};A_{\mathcal{J}^{\mathsf{c}}_S}\right).$$

That completes the proof of sufficient conditions for meeting (34). Now, to ensure that (35) is met, we note that by the random construction of the codebooks,

$$\sum_{\boldsymbol{m}^-}\frac{\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}(\widehat{Q}^{(1)}_{\hat{X}^n_1|\boldsymbol{M}^-}(\cdot|\boldsymbol{m}^-)\parallel\widehat{Q}^{(1)}_{\hat{X}^n_1})\right]}{2^{n\left(\sum_{(i,j)\in\mathcal{F}}\mu^-_{i,j}\right)}}$$

$$= \mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}(\widehat{Q}^{(1)}_{\hat{X}^n_1|\boldsymbol{M}^-}(\cdot|\underline{\mathbf{1}})\parallel\widehat{Q}^{(1)}_{\hat{X}^n_1})\right]$$

$$= \mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}(\widehat{Q}^{(1)}_{\hat{X}^n_1|M_\mathsf{h}}(\cdot|\underline{\mathbf{1}})\parallel Q^{\otimes n}_{\hat{X}_1}) - \mathsf{D}_{\mathsf{KL}}(\widehat{Q}^{(1)}_{\hat{X}^n_1}\parallel Q^{\otimes n}_{\hat{X}_1})\right],$$

where $\underline{\mathbf{1}}$ denotes the all-one vector of length $|\mathcal{F}| = \binom{\mathsf{h}}{2}$. Note that the analysis in (100)-(104) yields conditions when the the second term in the above equation vanishes as we let n diverge. So, we only need to focus on the first term. To do so, we proceed just as in the first part of this proof.

$$\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}(\widehat{Q}^{(1)}_{\hat{X}^n_1|\boldsymbol{M}^-}(\cdot|\underline{\mathbf{1}})\parallel\mathsf{Q}^{\otimes n}_{\boldsymbol{X}_1})\right]$$

$$\stackrel{(a)}{=}\mathbb{E}\left[\sum_{x^n_1}\left(\left(\frac{\sum_{\boldsymbol{m}^\pm:\boldsymbol{m}^-=\underline{\mathbf{1}}} Q^{\otimes n}_{X_1|\boldsymbol{A}}(x^n_1|\boldsymbol{A}^n(\boldsymbol{m}^\pm))}{\mathsf{N}'}\right)\times\log\left(\frac{\sum_{\tilde{\boldsymbol{m}}^\pm:\tilde{\boldsymbol{m}}^-=\underline{\mathbf{1}}} Q^{\otimes n}_{X_1|\boldsymbol{A}}(x^n_1|\boldsymbol{A}^n(\tilde{\boldsymbol{m}}^\pm))}{\mathsf{N}'\mathsf{Q}^{\otimes n}_{X_1}(x^n_1)}\right)\right)\right]$$

$$\stackrel{(b)}{\le}\sum_{x_1{}^n,\boldsymbol{a}^n} Q^{\otimes n}_{\boldsymbol{X_1}\boldsymbol{A}}(x^n_1,\boldsymbol{a}^n)\log\left[\sum_{S\subseteq\mathcal{F}}\frac{Q^{\otimes n}_{X_1}|\boldsymbol{A}_{\mathcal{J}^{\mathsf{c}}_S}(x^n_1|\boldsymbol{a}^n_{\mathcal{J}^{\mathsf{c}}_S})}{2^{n\left(\sum_{s\notin S}\mu^+_s\right)}\mathsf{Q}^{\otimes n}_{X_1}(x^n_1)}\right],$$

where

- (a) follows by notating $\mathsf{N}' = 2^{n\sum_{(i,j)\in\mathcal{F}}\mu^+_{i,j}}$; and
- (b) follows from steps identical to those between (100) and (104). The sole difference is that $(X^n_1,\ldots,X^n_\mathsf{h})$ is replaced by X^n_1, and the sums correspond to only all possible values taken by $\boldsymbol{M}^+$, since $\boldsymbol{M}^-=\underline{\mathbf{1}}$.

As before, the sum of terms in the equation above corresponding to atypical sequences yields a quantity no more than

$$\mathbb{P}\left[(X^n_1,\boldsymbol{A}^n)\notin T^n_\varepsilon[Q_{X_1\boldsymbol{A}}]\right]\cdot\log\left(1+2^{\mathsf{h}^2}\eta^{-n}_{X_1}\right),$$

where $\eta_{X_1} = \min_{\boldsymbol{y}\in\mathsf{supp}(X_1)}\mathsf{Q}_{X_1}(x_1)$; this quantity vanishes as $n\to\infty$. On the other hand, the contribution from typical sequences can be made arbitrarily small if for any $S\subseteq\mathcal{F}$,

$$\sum_{s\notin S}\mu^+_s > I\left(X_1;A_{\mathcal{J}^{\mathsf{c}}_S}\right).$$

B. Proof of Theorem 2

We proceed in a way similar to the proof of Theorem 1. We use the following notation in this proof.

$$\begin{aligned} \boldsymbol{D}_i^n(\boldsymbol{m}^{\pm}, k_{i-1}^{\pm}, l_i) &\triangleq \left(B_{i-1,i}^n(\boldsymbol{m}^{\pm}, k_{i-1}^{\pm}), C_i^n(\boldsymbol{m}^{\pm}, k_{i-1}^{\pm}, l_i)\right), \\ \boldsymbol{Y}_i &\triangleq (X_{i-1}, X_i), \\ \ell_i &\triangleq (k_{i-1}^{\pm}, l_i), \\ \mathsf{N}_i &\triangleq 2^{n(\kappa_{i-1}^{+}+\kappa_{i-1}^{-}+\lambda_i)}. \end{aligned}$$

Now, to find the conditions on the rates, we proceed in a fashion similar to [20] and [25]. The notation and arguments for the manipulations in (115)-(116) shown at the bottom of the next page are as follows:

- (a) follows since the codebooks for $\boldsymbol{A}$, $B_{i-1,i}$, and C_i are generated in an i.i.d. fashion.
- in (b), the expectation is over the codebooks for $B_{i-1,i}$, and C_i and the realization of $\boldsymbol{A}^n(\underline{\mathbf{1}})$.
- (c) follows by the use of the law of iterated expectations, where the inner conditional expectation is over all random codeword constructions $\{\boldsymbol{D}_i^n(\underline{\mathbf{1}}, \ell_i'') : \ell_i'' \neq \ell_i'\}$ conditioned on codeword $\boldsymbol{D}_i^n(\underline{\mathbf{1}}, \ell_i')$;
- (d) follows from Jensen's inequality; and
- similar to (103), (e) follows by splitting the inner summation according to the components where $\ell_i' \triangleq ({k_{i-1}^{\pm}}', l_i')$ and $\ell_i'' \triangleq ({k_{i-1}^{\pm}}'', l_i'')$ differ. Unity is an upper bound when the expectation is evaluated for terms corresponding to ${k_{i-1}^{\pm}}' \neq {k_{i-1}^{\pm}}''$, the second term is the result when the expectation is evaluated for $\ell_i' = \ell_i''$, and lastly, the third is the result from terms for which ${k_{i-1}^{\pm}}' = {k_{i-1}^{\pm}}''$ and $l_i' \neq l_i''$.

Finally, the rate conditions can be extracted from (116) by splitting the outer sum depending on whether $(\boldsymbol{y}^n, \boldsymbol{a}^n, b^n, c^n) \in T_\varepsilon^n[Q_{\boldsymbol{Y}_i \boldsymbol{A} \boldsymbol{D}_i}]$ or not. The sum for non-typical realizations in (116) is no more than

$$\mathbb{P}\left[(\boldsymbol{Y}^n, \boldsymbol{A}^n, \boldsymbol{D}_i^n) \notin T_\varepsilon^n[Q_{\boldsymbol{Y}_i \boldsymbol{A} \boldsymbol{D}_i}]\right] \cdot \log\left(1 + 2\eta_{\boldsymbol{Y}_i}^{-n}\right),$$

where $\eta_{\boldsymbol{Y}_i} = \min_{\boldsymbol{y} \in \mathsf{supp}(X_{i-1}, X_i)} \mathsf{Q}_{\boldsymbol{Y}_i}(\boldsymbol{y})$; this term goes to zero as $n \to \infty$. The contribution from typical realizations can be observed to vanish asymptotically, provided

$$\begin{aligned} \kappa_{i-1}^{+} + \kappa_{i-1}^{-} + \lambda_i &> I(X_{i-1}, X_i; B_{i-1,i}, C_i | \boldsymbol{A}), \\ \kappa_{i-1}^{+} + \kappa_{i-1}^{-} &> I(X_{i-1}, X_i; B_{i-1,i} | \boldsymbol{A}). \end{aligned}$$

Now, to ensure (43), let $\mathsf{N}_i' \triangleq 2^{n(\kappa_{i-1}^{+}+\lambda_i)}$. Consider then the argument from (117)-(118) shown at the bottom of the next page, where

- (a) follows since the codebooks for $\boldsymbol{A}$, $B_{i-1,i}$, and C_i are generated in an i.i.d. fashion.
- in (b), the expectation is over the portion of $B_{i-1,i}$-, and C_i-codebooks corresponding to $k_{i-1}^{-} = 1$ and the codeword $\boldsymbol{A}^n(\underline{\mathbf{1}})$.
- (c) follows from arguments similar to (115)-(116).

Lastly, by splitting the contributions of typical and non-typical sequences, the terms in (118) can be shown to vanish if:

$$\begin{aligned} \kappa_{i-1}^{+} + \lambda_i &> I(X_{i-1}; B_{i-1,i}, C_i | \boldsymbol{A}) \overset{(28)}{=} I(X_{i-1}; B_{i-1,i} | \boldsymbol{A}), \\ \kappa_{i-1}^{+} &> I(X_{i-1}; B_{i-1,i} | \boldsymbol{A}). \end{aligned}$$

The proof is complete by noting that the former constraint is redundant.

C. Proof of Lemma 3

Fix $\varepsilon > 0$, and let $\mathsf{N} \triangleq 2^{n \sum_{(i,j) \in \mathcal{F}} (\mu_{i,j}^{+} + \mu_{i,j}^{-})}$. Recall from Remarks 3 and 4 that

$$\lim_{n\to\infty} \mathbb{E}\left[\| \widehat{Q}^{(1)}_{\hat{X}_1^n \cdots \hat{X}_{\mathsf{h}}^n} - \mathsf{Q}^{\otimes n}_{X_1 \cdots X_{\mathsf{h}}} \|_1 + \sum_{i=2}^{\mathsf{h}} \sum_{\boldsymbol{m}^{\pm}} \frac{\| \widehat{Q}^{(i,\boldsymbol{m}^{\pm})}_{\hat{X}_{i-1}^n \hat{X}_i^n} - Q^{\otimes n}_{X_{i-1} X_i | \boldsymbol{A}}(\cdot | \boldsymbol{A}^n(\boldsymbol{m}^{\pm})) \|_1}{\mathsf{N}} \right] = 0. \quad (126)$$

Let for $j = 1, \ldots, \mathsf{h} - 1$,

$$\delta_i \triangleq \lim_{n\to\infty} \left(\frac{1}{\mathsf{N}} \sum_{\boldsymbol{m}^{\pm}} \mathbb{E} \left\| Q^{\otimes n}_{X_i | \boldsymbol{A}}(\cdot | \boldsymbol{A}^n(\boldsymbol{m}^{\pm})) \prod_{j=i+1}^{\mathsf{h}} \widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}_j^n | \hat{X}_{j-1}^n} - Q^{\otimes n}_{X_i \cdots X_{\mathsf{h}} | \boldsymbol{A}}(\cdot | \boldsymbol{A}^n(\boldsymbol{m}^{\pm})) \right\|_1 \right),$$

where

$$\hat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}_j^n | \hat{X}_{j-1}^n} \triangleq \frac{\hat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}_{j-1}^n \hat{X}_j^n}}{\hat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}_{j-1}^n}}.$$

First, consider $\delta_{\mathsf{h}-1}$, which is proven to be zero due to the argument in (121)-(122). Then, one can argue that for $j > 1$, $\delta_{\mathsf{h}-j} \leq \delta_{\mathsf{h}-j+1}$ using the argument in (123)-(125). Note that in (124), we have used the fact that (28) implies $X_i \leftrightarrow (X_{i+1}, \boldsymbol{A}) \leftrightarrow (X_{i+2}, \ldots, X_{\mathsf{h}})$, which then allows us to eliminate X_{h-j} within the third variational distance term. By induction, we have $\delta_1 = 0$. Finally, by transferring the summation inside the norm, the following can be shown.

$$\lim_{n\to\infty} \frac{\left[\mathbb{E} \left\| \sum_{\boldsymbol{m}^{\pm}} \left(Q^{\otimes n}_{X_i | \boldsymbol{A}}(\cdot | \boldsymbol{A}^n(\boldsymbol{m}^{\pm})) \prod_{j=i+1}^{\mathsf{h}} \widehat{Q}^{(j,\boldsymbol{m}^{\pm})}_{\hat{X}_j^n | \hat{X}_{j-1}^n} \right) - \sum_{\boldsymbol{m}^{\pm}} Q^{\otimes n}_{X_i \cdots X_{\mathsf{h}} | \boldsymbol{A}}(\cdot | \boldsymbol{A}^n(\boldsymbol{m}^{\pm})) \right\|_1 \right]}{\mathsf{N}} = 0.$$

The proof is then complete by invoking the triangle inequality to combine the above result with (38).

D. Message Selection at Node 1

Let pmf $Q_{D_1, D_2 Y}$ be given. Consider the a nested channel resolvability code for generating $Y \sim Q_Y$ via the channel $Q_{Y|D_1,\ldots D_2}$. Let the codebook structure be as given in Fig. 12, where the codebook for D_i is constructed randomly using $Q_{D_i | D_{i-1} \cdots D_1}$. Suppose that the rates of the codebooks satisfy

$$\nu_1 + \cdots + \nu_i > I(Y; D_1, \ldots, D_i), \quad i = 1, \ldots, k. \quad (127)$$

Using the approaches in [1] or [15], it can be shown that

$$\lim_{n\to\infty} \mathbb{E}\left[\| Q_{\hat{Y}^n} - Q_Y^{\otimes n} \|_1\right] = 0,$$

where

$$Q_{\hat{Y}^n} \triangleq \sum_{l_1,\ldots,l_k} \frac{Q^{\otimes n}_{Y|D_1 \cdots D_k}(\cdot | D_1^n(l_1), \ldots, D_k^n(l_1, \ldots, l_k))}{2^{n(\nu_1 + \cdots + \nu_k)}}.$$

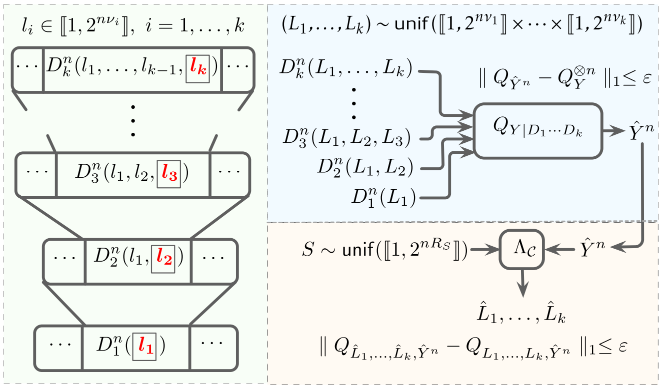

Fig. 12. A nested codebook structure for channel resolvability and the codeword selection problem.

Now suppose that we aim to realize a random selection of indices $(\hat{L}_1, \ldots, \hat{L}_k)$ using a function $\Lambda_{\mathcal{C}}$ that depends on the codebooks $\mathcal{C}$ and takes as inputs, an independent and uniform random seed S and the output of the channel $\hat{Y}^n$. Suppose that we require the conditional pmf of the selected indices given the channel output to match the *a posteriori* probability of the indices $(L_1, \ldots, L_k)$ given $\hat{Y}^n$, i.e.,

$$\mathbb{E}\left[\| Q_{\hat{L}_1,\ldots,\hat{L}_k,\hat{Y}^n} - Q_{L_1,\ldots,L_k,\hat{Y}^n} \|_1\right] \leq \varepsilon.$$

The following result characterizes the rate of random seed required to realize such a random selection.

Theorem 6: Fix $n \in \mathbb{N}$. Consider the random codebook structure given in Fig. 12 with rates $\nu_1, \ldots, \nu_k$ that satisfy (127). Let $\hat{Y}^n$ denote the output of the channel when the input is a codeword tuple that is uniformly selected from channel resolvability codebook. Let $S \sim \mathsf{unif}([\![1, 2^{nR_S}]\!])$, where

$$R_S > \nu_1 + \cdots + \nu_k - I(Y; D_1, \ldots, D_k).$$

Then, there exists a function $\Lambda_{\mathcal{C}} : \mathcal{Y}^n \times [\![1, 2^{nR_S}]\!] \to [\![1, 2^{n\nu_1}]\!] \times \cdots \times [\![1, 2^{n\nu_k}]\!]$ (that depends on the instance of the realized codebooks) such that $(\hat{L}_1, \ldots, \hat{L}_k) \triangleq \Lambda_{\mathcal{C}}(\hat{Y}^n, S)$ satisfies:

$$\lim_{n\to\infty} \mathbb{E}\left[\| Q_{\hat{L}_1,\ldots,\hat{L}_k,\hat{Y}^n} - Q_{L_1,\ldots,L_k,\hat{Y}^n} \|_1\right] = 0, \quad (128)$$

where the expectation is over all random codebooks.

$$\begin{aligned}
&\sum_{\boldsymbol{m}^\pm} \frac{\mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\widehat{Q}^{(i,\boldsymbol{m}^\pm)}_{\hat{\boldsymbol{Y}}_i^n} \,\|\, Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^\pm))\right)\right]}{2^{n(\mu^+_{1,2}+\mu^-_{1,2}+\cdots+\mu^+_{\mathsf{h}-1,\mathsf{h}}+\mu^-_{\mathsf{h}-1,\mathsf{h}})}} \\
&\stackrel{(a)}{=} \mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\widehat{Q}^{(i,\underline{\mathbf{1}})}_{\hat{\boldsymbol{Y}}_i^n} \,\|\, Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\underline{\mathbf{1}}))\right)\right] \qquad (115)\\
&\stackrel{(b)}{=} \mathbb{E}\left[\sum_{\boldsymbol{y}^n}\left(\frac{\sum_{\ell'_i} Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}\boldsymbol{D}_i}(\boldsymbol{y}^n|\boldsymbol{A}^n(\underline{\mathbf{1}})\boldsymbol{D}_i^n(\underline{\mathbf{1}},\ell'_i))}{\mathsf{N}_i}\right)\log\left(\frac{\sum_{\ell''_i} Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}\boldsymbol{D}_i}(\boldsymbol{y}^n|\boldsymbol{A}^n(\underline{\mathbf{1}})\boldsymbol{D}_i^n(\underline{\mathbf{1}},\ell''_i))}{\mathsf{N}_i \mathsf{Q}^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\underline{\mathbf{1}}))}\right)\right] \\
&\stackrel{(c)}{=} \sum_{\boldsymbol{y}^n,\ell'_i} \mathbb{E}\left[\frac{Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}\boldsymbol{D}_i}(\boldsymbol{y}^n|\boldsymbol{A}^n(\underline{\mathbf{1}})\boldsymbol{D}_i^n(\underline{\mathbf{1}},\ell'_i))}{\mathsf{N}_i}\,\mathbb{E}\left[\log\frac{\sum_{\ell''_i} Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}\boldsymbol{D}_i}(\boldsymbol{y}^n|\boldsymbol{A}^n(\underline{\mathbf{1}})\boldsymbol{D}_i^n(\underline{\mathbf{1}},\ell''_i))}{\mathsf{N}_i \mathsf{Q}^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\underline{\mathbf{1}}))}\,\middle|\,\begin{array}{c}\boldsymbol{A}^n(\underline{\mathbf{1}})\\ \boldsymbol{D}_i^n(\underline{\mathbf{1}},\ell'_i)\end{array}\right]\right] \\
&\stackrel{(d)}{\leq} \sum_{\boldsymbol{y}^n,\ell'_i} \mathbb{E}\left[\frac{Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}\boldsymbol{D}_i}(\boldsymbol{y}^n|\boldsymbol{A}^n(\underline{\mathbf{1}})\boldsymbol{D}_i^n(\underline{\mathbf{1}},\ell'_i))}{\mathsf{N}_i}\log\mathbb{E}\left[\frac{\sum_{\ell''_i} Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}\boldsymbol{D}_i}(\boldsymbol{y}^n|\boldsymbol{A}^n(\underline{\mathbf{1}})\boldsymbol{D}_i^n(\underline{\mathbf{1}},\ell''_i))}{\mathsf{N}_i \mathsf{Q}^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{A}^n(\underline{\mathbf{1}}))}\,\middle|\,\begin{array}{c}\boldsymbol{A}^n(\underline{\mathbf{1}})\\ \boldsymbol{D}_i^n(\underline{\mathbf{1}},\ell'_i)\end{array}\right]\right] \\
&\stackrel{(e)}{\leq} \sum_{\substack{\boldsymbol{y}^n,\boldsymbol{a}^n\\ b^n,c^n}} Q^{\otimes n}_{\boldsymbol{Y}_i\boldsymbol{A}\boldsymbol{D}_i}(\boldsymbol{y}^n,\boldsymbol{a}^n,b^n,c^n)\log\left[1+\frac{Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}\boldsymbol{D}_i}(\boldsymbol{y}^n|\boldsymbol{a}^n,b^n,c^n)}{\mathsf{N}_i\, Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{a}^n)}+\frac{Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}B_{i-1,i}}(\boldsymbol{y}^n|\boldsymbol{a}^n,b^n)}{\frac{\mathsf{N}_i}{2^{n\lambda_i}}\, Q^{\otimes n}_{\boldsymbol{Y}_i|\boldsymbol{A}}(\boldsymbol{y}^n|\boldsymbol{a}^n)}\right]. \qquad (116)
\end{aligned}$$

$$\begin{aligned}
&\frac{\sum_{\boldsymbol{m}^\pm,k^-_{i-1}} \mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\widehat{Q}^{(i,\boldsymbol{m}^\pm)}_{\hat{X}^n_{i-1}|K^-_{i-1}}(\cdot|k^-_{i-1}) \,\|\, Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}}\left(\cdot\,|\boldsymbol{A}^n(\boldsymbol{m}^\pm)\right)\right)\right]}{2^{n((\mu^+_{1,2}+\mu^-_{1,2})+\cdots+(\mu^+_{\mathsf{h}-1,\mathsf{h}}+\mu^-_{\mathsf{h}-1,\mathsf{h}})+\kappa^-_{i-1})}} \qquad (117)\\
&\stackrel{(a)}{=} \mathbb{E}\left[\mathsf{D}_{\mathsf{KL}}\left(\widehat{Q}^{(i,\underline{\mathbf{1}})}_{\hat{X}^n_{i-1}|K^-_{i-1}}(\cdot|1)\middle\| Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}}\left(\cdot\,|\boldsymbol{A}^n(\underline{\mathbf{1}})\right)\right)\right] \\
&\stackrel{(b)}{=} \mathbb{E}\left[\sum_{\boldsymbol{y}^n}\frac{\sum_{\ell'_i:\tilde{k}^-_{i-1}=1} Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}\boldsymbol{D}_i}(x^n_{i-1}|\boldsymbol{A}^n(\underline{\mathbf{1}})\boldsymbol{D}_i^n(\underline{\mathbf{1}},\hat{\ell}_i))}{\mathsf{N}'_i}\log\frac{\sum_{\tilde{\ell}_i:\tilde{k}^-_{i-1}=1} Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}\boldsymbol{D}_i}(x^n_{i-1}|\boldsymbol{A}^n(\underline{\mathbf{1}})\boldsymbol{D}_i^n(\underline{\mathbf{1}},\tilde{\ell}_i))}{\mathsf{N}'_i\mathsf{Q}^{\otimes n}_{X_{i-1}|\boldsymbol{A}}(x^n_{i-1}|\boldsymbol{A}^n(\underline{\mathbf{1}}))}\right] \\
&\stackrel{(c)}{\leq} \sum_{\substack{\boldsymbol{x}^n_{i-1},\boldsymbol{a}^n\\ b^n,c^n}} Q^{\otimes n}_{X_{i-1}\boldsymbol{A}\boldsymbol{D}_i}(\boldsymbol{x}^n_{i-1},\boldsymbol{a}^n,b^n,c^n)\log\left[1+\frac{Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}\boldsymbol{D}_i}(x^n_{i-1}|\boldsymbol{a}^n,b^n,c^n)}{\mathsf{N}'_i\, Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}}(x^n_{i-1}|\boldsymbol{a}^n)}+\frac{Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}B_{i-1,i}}(x^n_{i-1}|\boldsymbol{a}^n,b^n)}{\frac{\mathsf{N}'_i}{2^{n\lambda_i}}\, Q^{\otimes n}_{X_{i-1}|\boldsymbol{A}}(x^n_{i-1}|\boldsymbol{a}^n)}\right]. \qquad (118)
\end{aligned}$$

$$\delta_{\mathsf{h}-1} \triangleq \lim_{n\to\infty}\left(\frac{1}{\mathsf{N}}\sum_{\boldsymbol{m}^{\pm}}\mathbb{E}\left\|Q^{\otimes n}_{X_{\mathsf{h}-1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\widehat{Q}^{(\mathsf{h},\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}}|\hat{X}^n_{\mathsf{h}-1}} - Q^{\otimes n}_{X_{\mathsf{h}-1},X_{\mathsf{h}}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\right\|_1\right) \tag{121}$$

$$\le \lim_{n\to\infty}\left[\begin{array}{l}\frac{1}{\mathsf{N}}\sum_{\boldsymbol{m}^{\pm}}\mathbb{E}\left\|Q^{\otimes n}_{X_{\mathsf{h}-1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\widehat{Q}^{(\mathsf{h},\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}}|\hat{X}^n_{\mathsf{h}-1}} - \widehat{Q}^{(\mathsf{h},\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}-1}\hat{X}^n_{\mathsf{h}}}\right\|_1 \\ +\ \frac{1}{\mathsf{N}}\sum_{\boldsymbol{m}^{\pm}}\mathbb{E}\left\|\widehat{Q}^{(\mathsf{h},\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}-1}\hat{X}^n_{\mathsf{h}}} - Q^{\otimes n}_{X_{\mathsf{h}-1},X_{\mathsf{h}}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\right\|_1\end{array}\right]$$

$$= \lim_{n\to\infty}\left[\begin{array}{l}\frac{1}{\mathsf{N}}\sum_{\boldsymbol{m}^{\pm}}\mathbb{E}\left\|Q^{\otimes n}_{X_{\mathsf{h}-1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm})) - \widehat{Q}^{(\mathsf{h},\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}-1}}\right\|_1 \\ +\ \frac{1}{\mathsf{N}}\sum_{\boldsymbol{m}^{\pm}}\mathbb{E}\left\|\widehat{Q}^{(\mathsf{h},\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}-1}\hat{X}^n_{\mathsf{h}}} - Q^{\otimes n}_{X_{\mathsf{h}-1},X_{\mathsf{h}}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\right\|_1\end{array}\right] \overset{(126)}{=} 0. \tag{122}$$

$$\delta_{h-j} \triangleq \lim_{n\to\infty}\frac{\left(\sum_{\boldsymbol{m}^{\pm}}\left\|Q^{\otimes n}_{X_{\mathsf{h}-j}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\prod_{s=\mathsf{h}-j+1}^{\mathsf{h}}\widehat{Q}^{(s,\boldsymbol{m}^{\pm})}_{\hat{X}^n_s|\hat{X}^n_{s-1}} - Q^{\otimes n}_{X_{\mathsf{h}-j}\cdots X_{\mathsf{h}}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\right\|_1\right)}{\mathsf{N}} \tag{123}$$

$$\le \lim_{n\to\infty}\frac{\left[\begin{array}{l}\sum_{\boldsymbol{m}^{\pm}}\left\|\left[Q^{\otimes n}_{X_{\mathsf{h}-j}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm})) - \widehat{Q}^{(\mathsf{h}-j+1,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}-j}}\right]\prod_{s=\mathsf{h}-j+1}^{\mathsf{h}}\widehat{Q}^{(s,\boldsymbol{m}^{\pm})}_{\hat{X}^n_s|\hat{X}^n_{s-1}}\right\|_1 \\ +\sum_{\boldsymbol{m}^{\pm}}\left\|\left[\widehat{Q}^{(\mathsf{h}-j+1,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}-j}\hat{X}^n_{\mathsf{h}-j+1}} - Q^{\otimes n}_{X_{\mathsf{h}-j}X_{\mathsf{h}-j+1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\right]\prod_{s=\mathsf{h}-j+2}^{\mathsf{h}}\widehat{Q}^{(s,\boldsymbol{m}^{\pm})}_{\hat{X}^n_s|\hat{X}^n_{s-1}}\right\|_1 \\ +\sum_{\boldsymbol{m}^{\pm}}\left\|Q^{\otimes n}_{X_{\mathsf{h}-j}X_{\mathsf{h}-j+1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\prod_{s=\mathsf{h}-j+2}^{\mathsf{h}}\widehat{Q}^{(s,\boldsymbol{m}^{\pm})}_{\hat{X}^n_s|\hat{X}^n_{s-1}} - Q^{\otimes n}_{X_{\mathsf{h}-j}\cdots X_{\mathsf{h}}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\right\|_1\end{array}\right]}{\mathsf{N}}$$

$$\le \lim_{n\to\infty}\frac{\left[\begin{array}{l}\sum_{\boldsymbol{m}^{\pm}}\left\|Q^{\otimes n}_{X_{\mathsf{h}-j}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm})) - \widehat{Q}^{(\mathsf{h}-j+1,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}-j}}\right\|_1 \\ +\sum_{\boldsymbol{m}^{\pm}}\left\|\widehat{Q}^{(\mathsf{h}-j+1,\boldsymbol{m}^{\pm})}_{\hat{X}^n_{\mathsf{h}-j}\hat{X}^n_{\mathsf{h}-j+1}} - Q^{\otimes n}_{X_{\mathsf{h}-j}X_{\mathsf{h}-j+1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\right\|_1 \\ +\sum_{\boldsymbol{m}^{\pm}}\left\|Q^{\otimes n}_{X_{\mathsf{h}-j+1}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\prod_{s=\mathsf{h}-j+2}^{\mathsf{h}}\widehat{Q}^{(s,\boldsymbol{m}^{\pm})}_{\hat{X}^n_s|\hat{X}^n_{s-1}} - Q^{\otimes n}_{X_{\mathsf{h}-j+1}\cdots X_{\mathsf{h}}|\boldsymbol{A}}(\cdot|\boldsymbol{A}^n(\boldsymbol{m}^{\pm}))\right\|_1\end{array}\right]}{\mathsf{N}} \tag{124}$$

$$\overset{(126)}{=} \delta_{h-j+1}. \tag{125}$$

Proof: Let $K \triangleq |\mathcal{D}_1||\mathcal{D}_2|\cdots|\mathcal{D}_k||\mathcal{Y}|$. Choose $\delta, \varepsilon > 0$ such that

$$R_S - \sum_{\ell=1}^{k}\nu_\ell + I(Y; D_1, \dots, D_k) - 4\delta\log_2 K > \varepsilon. \tag{129}$$

Let $(L_1, \dots, L_k) \sim \mathsf{unif}([\![1, 2^{n\nu_1}]\!] \times \cdots \times [\![1, 2^{n\nu_k}]\!])$, $L^i \triangleq (L_1, \dots, L_i)$ and $\hat{L}^i \triangleq (\hat{L}_1, \dots, \hat{L}_i)$ for $1 \le i \le k$. By the random codebook construction, it follows that $(D_1^n(L_1), D_2^n(L^2), \dots, D_k^n(L^k), \hat{Y}^n)$ is indistinguishable from the output of a DMS $Q_{D_1\cdots D_kY}$. Hence, by [27, Th. 1.1], it follows that

$$\mathbb{P}\left[(D_1^n(L_1), D_2^n(L^2), \dots, D_k^n(L^k), \hat{Y}^n) \notin T_\delta^n[Q_{D_1\cdots D_kY}]\right] \le 2Ke^{-n\delta^2\eta}, \tag{130}$$

where $\eta \triangleq \min_{Q_{D_1\cdots D_kY}\in\mathsf{supp}(Q_{D_1\cdots D_kY})} Q_{D_1\cdots D_kY}(d_1, \dots, d_k, y)$.

Now, given any particular realization of codebooks $\mathcal{C}_\circ \triangleq \{(d_1^n(l_1), \dots, d_l^n(l^k))\}_{l^k\in[\![1,2^{n\nu_1}]\!]\times\cdots\times[\![1,2^{n\nu_k}]\!]}$, and channel output $y^n \in \mathcal{Y}^n$, let

$$\mathcal{N}_{\mathcal{C}_\circ}(y^n) \triangleq \left\{l^k : (d_1^n(l_1), \dots, d_k^n(l^k), y^n) \in T_\delta^n[Q_{D_1\cdots D_kY}]\right\}. \tag{131}$$

Then, by Lemma 4 below, for sufficiently large n.

$$\mathbb{E}\left[|\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)|\right] \le (k+1)2^{n\left(\sum_{\ell=1}^{k}\nu_\ell - I(Y;D_1\dots,D_k) + 2\delta\log_2 K\right)}.$$

Define $\mathcal{G}$ to be the set of all codebook realizations $\mathcal{C}_\circ \triangleq \{(d_1^n(l_1), \dots, d_l^n(l^k)) : l^k \in [\![1, 2^{n\nu_1}]\!] \times \cdots \times [\![1, 2^{n\nu_k}]\!]\}$ such that the following two conditions are met:

$$\mathbb{P}\left[(d_1^n(L_1), \dots, d_k^n(L^k), \hat{Y}^n) \notin T_\delta^n[Q_{D_1\cdots D_kY}]\,\big|\,\mathcal{C} = \mathcal{C}_\circ\right] \le \sqrt{2Ke^{-n\delta^2\eta}},$$

$$\frac{\mathbb{E}\left[|\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)|\,\big|\,\mathcal{C} = \mathcal{C}_\circ\right]}{k+1} \le 2^{n\left(\sum_{\ell=1}^{k}\nu_\ell - I(Y;D_1\cdots D_k) + 3\delta\log_2 K\right)}.$$

By Markov's inequality, we then have

$$\mathbb{P}[\mathcal{C} \notin \mathcal{G}] \le \sqrt{2Ke^{-n\delta^2\eta}} + 2^{-n\delta\log_2 K}. \tag{132}$$

Pick $\mathcal{C}^* \triangleq \left\{(d_1^{*n}(l_1), \dots, d_l^{*n}(l^k))\right\}_{l^k\in[\![1,2^{n\nu_1}]\!]\times\cdots\times[\![1,2^{n\nu_k}]\!]} \in \mathcal{G}$ and let $\mathcal{G}_{\mathcal{C}^*}$ to be the set of channel outputs y^n such that the following two conditions are met.

$$\mathbb{P}\left[(d_1^{*n}(L_1),\ldots,d_k^{*n}(L^k),\hat{Y}^n)\notin T_\delta^n[Q_{D_1\cdots D_kY}]\Big|\,\substack{\hat{Y}^n=y^n\\ \mathcal{C}=\mathcal{C}^*}\right] \leq \sqrt[4]{2Ke^{-n\delta^2\eta}}, \tag{133}$$
$$|\mathcal{N}_{\mathcal{C}^*}(y^n)| \leq (k+1)2^{n\left(\nu_1+\cdots+\nu_k-I(Y;D_1\ldots,D_k)+4\delta\log_2 K\right)}.$$

Then, by Markov's inequality, it follows that

$$\mathbb{P}[\hat{Y}^n\notin\mathcal{G}_{\mathcal{C}^*}|\mathcal{C}=\mathcal{C}^*]\leq\sqrt[4]{2Ke^{-n\delta^2\eta}}+2^{-n\delta\log_2 K}. \tag{134}$$

Further, it also follows that for each $y^n\in\mathcal{G}_{\mathcal{C}^*}$,

$$\sum_{l^k\notin\mathcal{N}_{\mathcal{C}^*}(y^n)} Q_{L_1\cdots L_k\hat{Y}^n}(l^k,y^n)\overset{(133)}{\leq}\sqrt[4]{2Ke^{-n\delta^2\eta}}, \tag{135}$$

where $Q_{L_1\cdots L_k\hat{Y}^n}$ is the joint pmf between indices and the output induced by $\mathcal{C}^*$. Then, from Lemma 5 below, for each $y^n\in\mathcal{G}_{\mathcal{C}^*}$, we can find $f_{y^n}:[\![1,2^{nR_S}]\!]\to[\![1,2^{n\nu_1}]\!]\times\cdots\times[\![1,2^{n\nu_k}]\!]$ such that

$$\begin{aligned}\| Q_{f_{y^n}(S)} &- Q_{L_1\cdots L_k|\hat{Y}^n=y^n}\|_1\\ &\leq \frac{|\mathcal{N}_{\mathcal{C}^*}(y^n)|}{2^{nR_S}}+\sqrt[4]{2Ke^{-n\delta^2\eta}}\\ &\leq (k+1)\frac{2^{n\left(\sum_{\ell=1}^k\nu_\ell-I(D_1\ldots,D_k;Y)+4\delta\log_2 S\right)}}{2^{nR_s}}+\sqrt[4]{2Ke^{-n\delta^2\eta}}\\ &\overset{(129)}{\leq}(k+1)2^{-n\varepsilon}+\sqrt[4]{2Ke^{-n\delta^2\eta}},\end{aligned}$$

where $S\sim\mathsf{unif}([\![1,2^{nR_S}]\!])$. Pick $l^{*k}\in[\![1,2^{n\nu_1}]\!]\times\cdots\times[\![1,2^{n\nu_k}]\!]$. Now, we can piece together these functions to define

$$\Lambda_{\mathcal{C}^*}(y^n,S)\triangleq\begin{cases}f_{y^n}(S), & y^n\in\mathcal{G}_{\mathcal{C}^*}\\ l^{*k}, & y^n\notin\mathcal{G}_{\mathcal{C}^*}.\end{cases} \tag{136}$$

By construction, we now have

$$\sum_{y^n\in\mathcal{G}_{\mathcal{C}^*}} Q_{\hat{Y}^n|\mathcal{C}=\mathcal{C}^*}(y^n)\,\|\,Q_{\Lambda_{\mathcal{C}^*}(y^n,S)}-Q_{L^k|\hat{Y}^n=y^n}\|_1 \leq (k+1)2^{-n\varepsilon}+\sqrt[4]{2Ke^{-n\delta^2\eta}}. \tag{137}$$

Combining the above bound with (134) and the fact that the variation distance between two pmfs is bounded above by 2, we obtain

$$\sum_{y^n} Q_{\hat{Y}^n|\mathcal{C}=\mathcal{C}^*}(y^n)\,\|\,Q_{\Lambda_{\mathcal{C}^*}(y^n,S)}-Q_{L^k|\hat{Y}^n=y^n}\|_1 \leq (k+1)2^{-n\varepsilon}+3\sqrt[4]{2Ke^{-n\delta^2\eta}}+2\cdot 2^{-n\delta\log_2 K}.$$

Since the RHS does not depend on the choice of $\mathcal{C}^*$ in $\mathcal{G}$, it follows that

$$\mathbb{E}\left[\|\,Q_{\Lambda_{\mathcal{C}}(y^n,S)}-Q_{L^k|\hat{Y}^n=y^n}\|_1\,\Big|\,\mathcal{C}\in\mathcal{G}\right] \leq (k+1)2^{-n\varepsilon}+3\sqrt[4]{2Ke^{-n\delta^2\eta}}+2\cdot 2^{n\delta\log_2 K}.$$

Next, using the fact that the variational distance between two pmfs is no more than 2, we also have

$$\mathbb{E}\left[\|\,Q_{\Lambda_{\mathcal{C}}(y^n,S)}-Q_{L^k|\hat{Y}^n=y^n}\|_1\,\Big|\,\mathcal{C}\notin\mathcal{G}\right]\leq 2. \tag{138}$$

Finally, combining the above two equations with (132) completes the claim. ■

Lemma 4: Consider the codebook construction of Fig. 12 with codebook sizes satisfying (127). Let $\mathcal{N}_{\mathcal{C}}(\cdot)$ and K be as in the proof of Theorem 6. Then, for n sufficiently large,

$$\mathbb{E}\left[|\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)|\right]\leq(k+1)2^{n\left(\sum_{\ell=1}^k\nu_\ell-I(Y;D_1\ldots,D_k)-2\delta\log_2 K\right)}.$$

Proof: Due to the random construction of the codebooks,

$$\mathbb{E}\left[|\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)|\right]=\sum_{\mathring{l}^k}\mathbb{E}\left[\mathbb{1}\{\mathring{l}^k\in\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)\}\big|L^k=(1,\ldots,1)\right].$$

To evaluate the conditional expectation, we partition the space $[\![1,2^{n\nu_1}]\!]\times\cdots\times[\![1,2^{n\nu_k}]\!]$ as follows.

$$\mathcal{S}_i=\left\{l^k:\begin{array}{ll}l_j=1, & j<i\\ l_{j+1}\neq 1, & j=i\end{array}\right\},\quad i=0,\ldots,k. \tag{139}$$

Note that $\bigcup_{j=1}^k\mathcal{S}_i=[\![1,2^{n\nu_1}]\!]\times\cdots\times[\![1,2^{n\nu_k}]\!]$. By the random nature of codebook construction, we have

$$\begin{aligned}\mathbb{P}&\left[(l_1,\ldots,l_k)\in\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)\,\big|L^k=(1,\ldots,1)\right]\\ &=\mathbb{P}\left[(l_1',\ldots,l_k')\in\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)\,\big|L^k=(1,\ldots,1)\right].\end{aligned} \tag{140}$$

for any pair of tuples $(l_1,\ldots,l_k),(l_1',\ldots,l_k')\in\mathcal{S}_j$, $j=0,\ldots,k$. Let for $j=0,1,\ldots,k$, $\ell^k(j)$ be chosen such that $\ell^k(j)\in\mathcal{S}_j$, and thus, due to (140), we have

$$\begin{aligned}\mathbb{E}[|\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)|]&=\sum_{j=0}^k|\mathcal{S}_j|\,\mathbb{P}\Big[\ell^k(j)\in\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)\,\big|L^k=(1,\ldots,1)\Big]\\ &\leq\sum_{j=0}^k\left[\prod_{\imath>j}2^{n\nu_\imath}\right]\eta_j,\end{aligned} \tag{141}$$

where we let $\eta_j\triangleq\mathbb{P}\left[\ell^k(j)\in\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)\big|L^k=(1,\ldots,1)\right]$, $j=0,\ldots,k$. Clearly, $\eta_k\leq 1$, and η_0 is exactly the probability that realizations $(D_1^n,D_2^n,\ldots,D_k^n)\sim Q_{D_1\ldots D_k}^{\otimes n}$ and $Y^n\sim Q_Y^{\otimes n}$ selected independent of one another are jointly δ-letter typical. Thus, by [27, Th. 1.1], it follows that

$$\begin{aligned}\eta_0&=\sum_{(d_1^n,\ldots,d_k^n,y^n)\in T_\delta^n[Q_{D_1\cdots D_kY}]}Q_{D_1\cdots D_k}(d_1^n,\ldots,d_k^n)Q_Y(y^n)\\ &\leq 2^{-nI(Y;D_1,\ldots,D_k)+n\delta(H(D_1,\ldots,D_k,Y)+H(D_1,\ldots,D_k)+H(Y))}\\ &\leq 2^{-n(I(Y;D_1,\ldots,D_k)-2\delta\log_2 K)}.\end{aligned} \tag{142}$$

Now, when $0<j<k$, we observe that η_j is the probability that $(D_1^n,D_2^n,\ldots,D_k^n)\sim Q_{D_1\ldots D_k}^{\otimes n}$ and Y^n (i.e., the output when $(D_1^n,\ldots,D_j^n)$ is sent through the channel $Q_{Y|D_1,\ldots,D_j}$ are jointly δ-letter typical. Therefore, by use of [27, Ths. 1.1 and 1.2], we see that

$$\eta_j\leq 2^{-n(I(D_1,\ldots,D_k;Y|D_1,\ldots,D_j)-2\delta\log_2 K)}. \tag{143}$$

Finally, combining (141), (142), and (143), we obtain

$$\mathbb{E}\left[|\mathcal{N}_{\mathcal{C}}(\hat{Y}^n)|\right]\leq 1+\sum_{j=0}^{k-1}\frac{2^{n\left(\sum_{\ell=j+1}^k\nu_\ell\right)}}{2^{n\left(I(Y;D_1\ldots,D_k|D_1\ldots,D_j)-2\delta\log_2 K\right)}}.$$

The claim then follows since (127) implies that

$$\max_{0\le j<k}\left[\sum_{\ell=j+1}^{k}\nu_\ell - I(Y;D_1\ldots,D_k|D_1\ldots,D_j)\right] = \sum_{\ell=1}^{k}\nu_\ell - I(Y;D_1,\ldots,D_k).$$

■

Lemma 5: Let Q be a pmf on a finite set $\mathcal{A}$ such that there exists $\mathcal{B}\subseteq\mathcal{A}$ with $|\mathcal{B}|=M$ and $\sum_{b\in\mathcal{B}}Q(b)\ge 1-\varepsilon$ for $0<\varepsilon<1$. Now, suppose that $L\sim\mathsf{unif}([\![1,\ell]\!])$. Then, there exists $f:[\![1,\ell]\!]\to\mathcal{A}$ such that $Q_{f(L)}$, the pmf of $f(L)$, satisfies $\| Q_{f(L)}-Q\|_1\le\varepsilon+\frac{M}{\ell}$.

Proof: Let $b_1\preceq b_2\preceq\cdots\preceq b_M$ be an ordering of $\mathcal{B}$. Let $p_0=0$, and for $1\le i\le M$, let $p_i\triangleq\sum_{j=1}^{i}Q(b_j)$ denote the cumulative mass function. Now, let $N_i\triangleq\lfloor p_i\ell\rfloor$, $i=0,\ldots,M$, and let $f:[\![1,N_M]\!]\to\mathcal{B}$ be defined by the pre-images via $f^{-1}(b_i)=\{N_{i-1}+1,\ldots,N_i\}$, $i=1,\ldots,M$. Now, by construction, we have

$$0\le p_i-\mathbb{P}[f(L)\in\{b_1,\ldots,b_i\}]\le\ell^{-1},\quad i=1,\ldots,M.$$

Consequently, we also have for any $i=1,\ldots,M$,

$$-\ell^{-1}\le p_i-p_{i-1}-Q_{f(L)}(b_i) \tag{144}$$
$$= Q(b_i)-Q_{f(L)}(b_i)\le\ell^{-1}. \tag{145}$$

Hence, we see that

$$\sum_{a\in\mathcal{A}}|Q(a)-Q_{f(L)}(a)| = \left(\sum_{i=1}^{M}|Q(b_i)-Q_{f(L)}(b_i)| + \mathbb{P}[A\notin\mathcal{B}]\right) \overset{(145)}{\le} \frac{M}{\ell}+\varepsilon. \tag{146}$$

■

E. Message Selection at Nodes $2,\ldots,\mathsf{h}-1$

Let pmf $p_{D_1,D_2,Y}$ and $n\in\mathbb{N}$ be given. Suppose that $D_1^n\sim Q_{D_1}^{\otimes n}$. Now, generate a codebook for D_2 such that codewords $\{D_2^n(i_2): i_2\in[\![1,2^{n\nu_2}]\!]\}$ are selected independently with each codeword selected using $\prod_{k=1}^{n}Q_{D_2|D_{1,k}}$. Suppose that $\nu_2>I(Y;D_2|D_1)$. Then, it can be shown that

$$\lim_{n\to\infty}\mathbb{E}\left[\|\hat{Q}_{\hat{Y}^n|D_1^n}-Q_{Y|D_1}^{\otimes n}(\cdot\mid D_1^n)\|_1\right]=0, \tag{147}$$

where

$$\hat{Q}_{\hat{Y}^n|D_1^n}\triangleq\sum_{l_2=1}^{2^{n\nu_2}}\frac{Q_{Y|D_1,D_2}^{\otimes n}(\cdot|D_1^n,D_2^n(l_2))}{2^{n\nu_2}}. \tag{148}$$

Now suppose that as in Appendix D, one would like to characterize the amount of randomness required to generate randomly $\hat{L}_2$ using a function $\Lambda_{\mathcal{C}}$ (that depends on the codebooks $\mathcal{C}$) that takes as inputs a uniform random seed S, the output of the channel $\hat{Y}^n$, and the actual L_2' that was used to generate the channel output. We want $\hat{L}_2$ to mimic L_2 and that the joint correlation of the RV $\hat{L}_2$ and the realized $\hat{Y}^n$ is arbitrarily close to $Q_{L_2,\hat{Y}^n}$, i.e.,

$$\lim_{n\to\infty}\mathbb{E}\left[\|Q_{\hat{L}_2,\hat{Y}^n}-Q_{L_2,\hat{Y}^n}\|_1\right]=0. \tag{149}$$

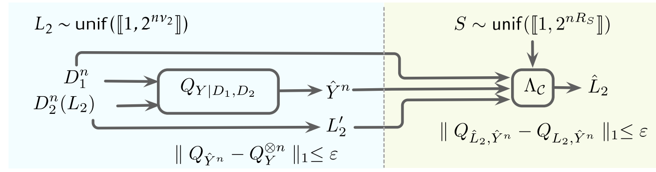

Fig. 13. A nested codebook structure for channel resolvability and the codeword selection problem.

The following result characterizes the rate of randomness required to realize this random index selection.

Theorem 7: Consider the above codebook setup with $\nu_2>I(Y;D_2|D_1)$. Let $\hat{Y}^n$ denote the channel output in the setup of Fig. 13. Let

$$R_S>\nu_2-I(Y;D_2|D_1).$$

Then, there exists a function $\Lambda_{\mathcal{C}}:\mathcal{Y}^n\times[\![1,2^{nR_S}]\!]\times\mathcal{D}_1^n\to[\![1,2^{n\nu_2}]\!]$ (that depends on the instance of the realized codebooks) such that $\hat{L}_2\triangleq\Lambda_{\mathcal{C}}(\hat{Y}^n,S,D_1^n)$ satisfies (149).

Proof: The proof mirrors exactly those of Theorem 6 and associated lemmas, and is omitted. ■

References

[1] B. N. Vellambi, J. Kliewer, and M. Bloch, "Strong coordination over multi-hop line networks," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2015, pp. 192–196.

[2] B. N. Vellambi, J. Kliewer, and M. Bloch, "Strong coordination over a line when actions are Markovian," in *Proc. 50th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2015, pp. 412–417.

[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] P. Grover and A. Sahai, "Implicit and explicit communication in decentralized control," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2010, pp. 278–285.

[5] P. Cuff and L. Zhao, "Coordination using implicit communication," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2011, pp. 467–471.

[6] G. Ranade and A. Sahai, "Implicit communication in multiple-access settings," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2011, pp. 998–1002.

[7] D. Vail and M. Veloso, "Dynamic multi-robot coordination," in *Multi-Robot Systems: From Swarms to Intelligent Automata*, vol. 2. Dordrecht, The Netherlands: Kluwer, 2003, pp. 87–100.

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

[9] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distortion theory," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 63–86, Jan. 1996.

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

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

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

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

[14] M. H. Yassaee, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 3053–3057.

[15] A. Gohari and V. Anantharam, "Generating dependent random variables over networks," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2011, pp. 698–702.
[16] P. Cuff, "Communication requirements for generating correlated random variables," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2008, pp. 1393–1397.
[17] M. L. Treust, "Correlation between channel state and information source with empirical coordination constraint," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2014, pp. 272–276.
[18] M. L. Treust, "Empirical coordination with two-sided state information and correlated source and state," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 466–470.
[19] M. L. Treust, "Empirical coordination with channel feedback and strictly causal or causal encoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 471–475.
[20] M. R. Bloch and J. Kliewer, "Strong coordination over a three-terminal relay network," in *Proc. IEEE Inf. Theory Workshop*, Nov. 2014, pp. 646–650.
[21] F. Haddadpour, M. H. Yassaee, A. Gohari, and M. R. Aref, "Coordination via a relay," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 3048–3052.
[22] A. Bereyhi, M. Bahrami, M. Mirmohseni, and M. R. Aref, "Empirical coordination in a triangular multiterminal network," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2149–2153.
[23] B. Larrousse and S. Lasaulce, "Coded power control: Performance analysis," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 3040–3044, doi: 10.1109/ISIT.2013.6620784.
[24] B. Larrousse, S. Lasaulce, and M. Wigger, "Coordination in state-dependent distributed networks: The two-agent case," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 979–983.
[25] M. Bloch and J. Kliewer, "Strong coordination over a line network," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2319–2323.
[26] S. Satpathy and P. Cuff, "Secure cascade channel synthesis," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2955–2959. [Online]. Available: http://arxiv.org/abs/1305.4314v1
[27] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, nos. 4–5, pp. 265–444, 2007.
[28] A. E. Gamal and Y.-H. Kim, *Network Information Theory*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[29] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6760–6786, Nov. 2014.
[30] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, Apr. 2006.
[31] I. Cisiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless System*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

Badri N. Vellambi (M'06–SM'16) 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 was a post-doctoral research fellow with the Institute for Telecommunications Research, University of South Australia, between 2008 and 2015. During 2015-2016, he was a post-doctoral research associate with the New Jersey Institute of Technology, Newark, NJ, USA. Since 2017, he has been a post-doctoral research fellow at the Research School of Computer Science, Australian National University, Canberra, Australia. His current research interests include information theory, data compression, channel coding, statistical learning, reinforcement learning and artificial intelligence.

Joerg Kliewer (S'97–M'99–SM'04) received the Dipl.-Ing. (M.Sc.) degree in electrical engineering from Hamburg University of Technology, Hamburg, Germany, in 1993 and the Dr.-Ing. degree (Ph.D.) in electrical engineering from the University of Kiel, Germany, in 1999, respectively.

From 1993 to 1998, he was a research assistant at the University of Kiel, and from 1999 to 2004, he was a senior researcher and lecturer with the same institution. In 2004, he visited the University of Southampton, U.K., for one year, and from 2005 until 2007, he was with the University of Notre Dame, IN, as a visiting assistant professor. From 2007 until 2013 he was with New Mexico State University, Las Cruces, NM, most recently as an associate professor. He is now with the New Jersey Institute of Technology, Newark, NJ, as an associate professor. His research interests span information and coding theory, graphical models, and statistical algorithms, which includes applications to networked communication and security, data storage, and biology.

Dr. Kliewer was the recipient of a Leverhulme Trust Award and a German Research Foundation Fellowship Award in 2003 and 2004, respectively. He was an Associate Editor of the IEEE TRANSACTIONS on COMMUNICATIONS from 2008 until 2014, and since 2015 serves as an Area Editor for the same journal. He is also an Associate Editor of the IEEE TRANSACTIONS on INFORMATION THEORY since 2017 and a member of the editorial board of the IEEE INFORMATION THEORY SOCIETY NEWSLETTER since 2012.

Matthieu R. Bloch (M'08–SM'16) is an Associate Professor in the School of Electrical and Computer Engineering. He received the engineering degree from Supélec, Gif-sur-Yvette, France, the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 2003, the Ph.D. degree in engineering Science from the Université de Franche-Comté, Besançon, France, in 2006, and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology in 2008. In 2008-2009, he was a postdoctoral research associate at the University of Notre Dame, South Bend, IN. Since July 2009, Dr. Bloch has been on the faculty of the School of Electrical and Computer Engineering, and from 2009 to 2013 Dr. Bloch was based at Georgia Tech Lorraine. His research interests are in the areas of information theory, error-control coding, wireless communications, and cryptography. Dr. Bloch is a senior member of the IEEE and has served on the organizing committee of several international conferences; he was also the chair of the Online Committee of the IEEE Information Theory Society from 2011 to 2014 and he has been on the Board of Governors of the IEEE Information Theory Society since January 2016. He is presently an Associate Editor for the IEEE TRANSACTIONS on INFORMATION THEORY. He is the co-recipient of the IEEE Communications Society and IEEE Information Theory Society 2011 Joint Paper Award and the co-author of the textbook Physical-Layer Security: From Information Theory to Security Engineering published by Cambridge University Press.