Exploiting the N-to-1 Mapping in Compress-and-Forward Relaying

Kevin Luo, *Member, IEEE*, Ramy H. Gohary, *Senior Member, IEEE*, and Halim Yanikomeroglu, *Senior Member, IEEE*

***Abstract*—In this paper, a forward decoding procedure is developed for the compress-and-forward (CF) relaying scheme. This procedure uses a layered framework and is based on exploiting a feature of the N-to-1 mapping inherent in the underlying Wyner–Ziv binning. It is shown that exploiting this mapping enables the relaxation of the constraint on the rate of the relay codewords representing the bin indices. For the cooperative multimessage network, the proposed procedure achieves the same rate region as the short-message noisy network coding (SNNC) scheme. However, this procedure is more advantageous for other networks including the two networks presented herein. The first network is a relay chain one with two destinations, whereas the second network is a partially cooperative multimessage one with three destinations. In both networks, side information is available to a subset of the decoding nodes, but not to the rest of the nodes, and in both cases, the network benefits from the relaxation of the rate of the CF bin indices. This relaxation results in rate regions larger than those achieved by the conventional CF and SNNC.**

***Index Terms*—Relay channel, compress-and-forward, short-message noisy network coding, Wyner-Ziv binning, layered framework, side information, relay chain, partially cooperative multimessage network.**

I. Introduction

COMPRESS-AND-FORWARD (CF) [1] is a classical relaying scheme for communicating over relay channels. In conventional CF, the source transmits a new codeword in each time block. The relay uses a pre-designed codebook to generate descriptions of its received signal. Using Wyner-Ziv binning, the codewords in the codebook of this description are randomly partitioned into non-overlapping bins, which results in an N-to-1 mapping from the description codewords to the bin indices. In each block, the CF relay provides a description of its received signal and sends its bin index in the next block to facilitate decoding at the receiver. Decoding at the CF receiver comprises three steps [1]: 1) decoding the bin index from the relay; 2) using the received signal as side information to decode the relay description codeword in the bin; and 3) recovering the transmitted codeword from the source with the facilitation of the relay description. The last two steps can also be performed jointly [2].

Manuscript received July 12, 2014; revised June 11, 2015; accepted September 30, 2015. Date of publication October 16, 2015; date of current version December 18, 2015. This work was supported in part by BlackBerry Ltd., Canada, and in part by the Collaborative Research and Development (CRD) Grant and the Discovery Grant (DG) programs of the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors are with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: wenluo@sce.carleton.ca; gohary@sce.carleton.ca; halim@sce.carleton.ca).

Communicated by Y. Liang, 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.2015.2491941

CF relaying was originally proposed for the three-node relay channel, but was later extended to channels with multiple relays [3]–[5]. CF is also known to be capacity achieving for various classes of relay channels [6]–[10].

Akin to CF is the noisy network coding (NNC) scheme provided in [11]. The philosophy that underlies this scheme resembles, to some extent, that of CF. However, there are three differences between the CF scheme and the NNC one. First, in contrast with random binning used in CF, in NNC the relay transmits a codeword that bears a 1-to-1 correspondence with the description codeword. Second, in NNC the source uses repetitive transmission, wherein one long message is encoded over a large number of blocks. This is in contrast with conventional CF, wherein a new short message is transmitted by the source in each time block. Finally, in NNC, the received signals in all time blocks are concatenated and decoded jointly; in CF the decoding is performed on a forward block-by-block basis.

Repetitive encoding of long messages over a large number of blocks incurs significant delay, which renders short messaging more desirable. Variations of the original NNC that use short message encoding (SNNC) were proposed in [12]–[15]. Despite their differences, both NNC and SNNC decoding use the inherent 1-to-1 mapping between the description codewords at the relay and its transmitted codewords. This 1-to-1 mapping can be seen as a special case of the general Wyner-Ziv binning with equal rate of the Wyner-Ziv codes and the description codewords at the relay.

In [12], SNNC was studied for the standard three-node relay channel when forward and backward decoding are used. In [13] and [14], it was shown that SNNC yields the same rate region as NNC when either backward decoding or joint decoding with concatenated blocks is used in the multimessage network considered therein.

Forward decoding that uses the SNNC codebook structure was investigated for the multimessage network in [15]. This decoding is based on ordered partitions of the nodes, thereby resulting in a set of constraints on the achievable rate. Using a geometric approach, it was shown that, in the multimessage network considered in [15], there exist ordered partitions that yield a rate region that coincides with the one achieved by NNC.

In the standard three-node relay channel, conventional CF, NNC and SNNC achieve the same rate. However, in more general multimessage networks, NNC and SNNC achieve rate regions larger than that achieved by conventional CF.

In this paper, we consider a multimessage network similar to the one considered in [11], [13], and [15]. For this network, we develop a decoding procedure based on the conventional CF codebook structure. The proposed procedure uses multiple layers of the joint typicality sets for decoding. At each layer, a subset of the transmitted codewords and the codewords of the relay descriptions are successfully decoded. The codebooks that contains these successfully decoded codewords are used to construct the joint typicality set at the next decoding layer. Such sequential construction of the joint typicality set at each layer provides a hierarchical structure whereby the joint typicality set at one layer is a proper subset of that at the layer below it. This construction provides in effect a systematic method for obtaining the ordered partitions conceived in [15]. In addition to the layered framework, the proposed decoding procedure exploits the N-to-1 mapping that underlies the Wyner-Ziv binning. We show that this procedure is able to achieve the same rate region as that achieved by SNNC in the multimessage network. However, because of the N-to-1 mapping, this procedure enables the relaxation of the constraint on the rate of the bin indices, which will be shown to be beneficial in certain cases. In particular, we will provide hereinafter instances in which the N-to-1 mapping yields rate advantages. A key feature of those instances is that side information from the source is available only to subset of the nodes. This can be caused, for instance, when the link between the source and the receiving node is broken. The lack of side information may induce rate loss when the relay description is used to recover the CF bin indices. In such situations it is more beneficial for the receiving node to recover the CF bin indices directly without side information, which enables the node to take advantage of the relaxed constraint on the rate of the CF bin indices. To investigate the gain of the relaxation of the rate constraint on the CF bin indices, we consider two networks. In the first network, a source broadcasts a common message to two destinations, which are assisted by a chain of two cascaded relays, cf. [16]. In this chain, the first relay receives signals from the source, uses the CF strategy and forwards the bin indices of its description codewords to the second relay and the two destinations. The second relay only receives signals from the first relay and uses the decode-and-forward (DF) strategy to assist the destinations in recovering the bin indices transmitted by the first relay. In the second network, a source S sends common messages to two receivers D_1 and D_2 with the assistance of a relay R. The relay R cooperates with S in the transmission to D_1 and D_2, and also sends independent messages to its own destination D_3. Unlike D_1 and D_2, destination D_3 does not have direct link from S and hence the link between R and D_3 is a standard point-to-point one. Without the signal from S, destination D_3 can only recover the message from R without the cooperation from S. Therefore, the network is only partially cooperative. It will be shown that for the considered networks, the relaxation of the rate constraint provided by the N-to-1 mapping enables larger rate regions to be achieved.

Notation: Regular face upper and lower case letters will refer to random variables, and their corresponding realizations, respectively. Boldface letters will refer to length-n sequences, and the calligraphic font will be used to refer to sets of nodes. A sequence $\mathbf{x}$ of an index s transmitted or selected by a node d_k in block b is denoted by $\mathbf{x}_k(s_{k,b})$. A sequence of random variables of node index d_k is denoted by $\mathbf{X}_k$. A tuple of random variables is denoted by $X_{\mathcal{A}} \triangleq (X_k : d_k \in \mathcal{A})$. The sum rate of the codebooks of a set of nodes $\mathcal{A}$ is denoted by $R_{\mathcal{A}} \triangleq \sum_{d_k \in \mathcal{A}} R_{d_k}$.

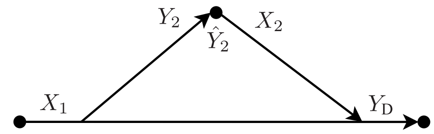

Fig. 1. Standard three-node relay channel (SNNC or CF).

II. Preliminaries

We begin by reviewing the achievable rate of SNNC [13] (with 1-to-1 mapping) and conventional CF (with N-to-1 mapping) [2] in the standard three-node relay channel [1] shown in Fig. 1. Our goal in this section is to gain insight into the relay transmission rate and its impact to the achievable rate.

For both SNNC and conventional CF, the source sends $\mathbf{X}_1$ in each block and the relay and the receiver receive $\mathbf{Y}_2$ and $\mathbf{Y}_{\mathrm{D}}$, respectively. Upon receiving $\mathbf{Y}_2$, the relay obtains a description, $\hat{\mathbf{Y}}_2$, of its received signal, which is mapped to $\mathbf{X}_2$. The relay then sends $\mathbf{X}_2$ in the next block to the receiver to facilitate decoding. The difference between SNNC and conventional CF in the codebook structure is the relationship between $\hat{\mathbf{Y}}_2$ and $\mathbf{X}_2$. In SNNC, each $\hat{\mathbf{y}}_2$ is mapped to a distinct $\mathbf{x}_2$, whereas in conventional CF, potentially multiple $\hat{\mathbf{y}}_2$ are assigned to one bin index which is mapped to an $\mathbf{x}_2$.

Let $R_1, \check{R}_2$ and $\hat{R}_2$ be the rates of the codewords that represent the source message, the relay bin index and the relay description of its received signal, respectively. Consider the probability mass functions (pmfs) of the form

$$p(x_1, x_2, y_{\mathrm{D}}, y_2, \hat{y}_2) = p(x_1)p(x_2)p(y_{\mathrm{D}}, y_2|x_1, x_2)p(\hat{y}_2|x_2, y_2).$$

In SNNC [13], [15], $\check{R}_2 = \hat{R}_2$, and the following rate can be achieved:

$$R_1 \le I(X_1; \hat{Y}_2, Y_{\mathrm{D}}|X_2), \tag{1a}$$
$$R_1 \le I(X_1, X_2; Y_{\mathrm{D}}) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_{\mathrm{D}}). \tag{1b}$$

In conventional CF, $\check{R}_2 \le \hat{R}_2$, and the rate satisfying (1a) and the following constraints is achievable, cf. [2, Sec. 16.7]:

$$R_1 \le I(X_1; Y_{\mathrm{D}}|X_2) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_{\mathrm{D}}) + \check{R}_2, \tag{2a}$$
$$\check{R}_2 \le I(X_2; Y_{\mathrm{D}}). \tag{2b}$$

Choosing $\check{R}_2 = I(X_2; Y_{\mathrm{D}})$ in (2b) maximizes R_1 and yields the same rate expression as SNNC.

For the three-node relay channel, it was shown in [2, Remark 16.3] that the achievable rate is maximized when the two constraints in (1) are equal, which yields

$$I(X_2; Y_{\mathrm{D}}) = I(\hat{Y}_2; Y_2|X_2, Y_{\mathrm{D}}). \tag{3}$$

It can be seen that when (3) is satisfied, CF is able to achieve the same rate as SNNC and yields $\check{R}_2 = I(X_2; Y_{\mathrm{D}}) =$

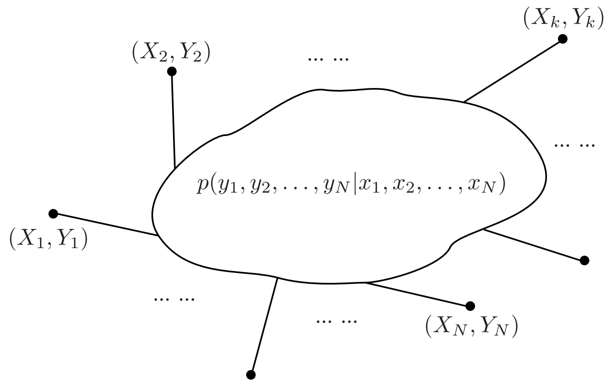

Fig. 2. An N-node multimessage network.

$I(\hat{Y}_2; Y_2|X_2, Y_{\mathrm{D}}) \leq I(\hat{Y}_2; Y_2|X_2) \leq \hat{R}_2$, which contrasts the condition in SNNC, wherein $\check{R}_2 = \hat{R}_2$.

In the next section, it will be shown that the result that CF achieves the same rate as SNNC but yields a lower rate on the bin indices can be extended to more general multimessage networks.

III. A Layered Forward Decoding Procedure for Multimessage Networks

In a relay network with multiple receivers, when SNNC (with 1-to-1 mapping) is used, the rate of the codewords representing the bin indices at the relay is an intermediate parameter and can be eliminated from the expressions of the achievable rate region. In contrast, when conventional CF (with N-to-1 mapping) is used in the presence of multiple receivers, the rate of the codewords representing the bin indices at the relay cannot be readily eliminated. The decoding of the relay transmitted codewords at each receiver imposes a constraint on the rate of the bin indices. This constraint appears in the achievable rate expressions and induces a rate loss in comparison with SNNC.

In this section, we analyze a CF forward decoding procedure in which two component strategies are combined: the N-to-1 mapping characteristic of conventional CF [1] and the sliding-window decoding of the relay messages characteristic of SNNC, see e.g., [15]. Combining these components results in a new decoding procedure that subsumes conventional CF and SNNC, and therefore enables potentially higher rates to be achieved. Indeed, subsequent developments in Theorems 1 and 2 show that combining these components yields the same rate expressions as NNC and SNNC but with a more relaxed constraint on rate of the bin indices at the relays. The advantage of the relaxed constraint on the rate of the bin indices will be illustrated in detail in the next section.

Consider the multimessage network shown in Fig. 2. The network contains a set of nodes $\mathcal{N} = \{1, \ldots, N\}$, each of which acts as a source, a receiver and a relay. As a source, node $d_k \in \mathcal{N}$ sends an independent common message through the transmission of $\mathbf{X}_k$ to the set of its destinations $\mathcal{D}_{d_k} \subseteq \mathcal{N}$. The set of nodes that wish to send messages to d_k is denoted by $\mathcal{S}_{d_k} \subseteq \mathcal{N}$. As a relay to assist the transmission of other nodes, d_k provides a description, $\hat{\mathbf{Y}}_k$, of its received signal $\mathbf{Y}_{d_k}$, and facilitates the decoding at other nodes through the transmission of $\mathbf{X}_k$. Cooperation between nodes in this network is based on the facilitation provided by the encoding procedure at each node and the use of the received signal as side information in the decoding procedure at each receiver.

In recovering the messages from the nodes in $\mathcal{S}_{d_k}$, the receiver at d_k can treat the information of the messages from the nodes in $\mathcal{N} \setminus \mathcal{S}_{d_k}$ as interference. Using this approach and SNNC (with 1-to-1 mapping), expressions for the achievable regions are provided in [15]. In Theorem 1 herein, this result is extended to the codebook structure that bears the general N-to-1 mapping. In contrast, in Theorem 2, an achievable rate region is provided for the case when the receiver at d_k treats the interference from the nodes in $\mathcal{N} \setminus \mathcal{S}_{d_k}$ as noise instead of attempting to decode it.

Theorem 1: Let $(\times_{k=1}^{N} \mathcal{X}_k, p(y^N|x^N), \times_{k=1}^{N} \mathcal{Y}_{d_k})$ be the general discrete memoryless multimessage network. A rate tuple $(R_1, \ldots, R_N)$ is achievable if:

$$\begin{aligned} R_{\mathcal{S}} \leq \min_{d_k \in \mathcal{S}^c \cap \mathcal{D}_{\mathcal{S}}} & I(X_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{S}^c}) \\ & - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}}, \end{aligned} \quad (4)$$

for all subsets $\mathcal{S} \subset \mathcal{N}$ and $\check{\mathcal{S}} \subseteq \mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$, such that $\mathcal{S}^c \cap \mathcal{D}_{\mathcal{S}} \neq \emptyset$; the set $\mathcal{D}_{\mathcal{S}} \triangleq \cup_{d_l \in \mathcal{S}} \mathcal{D}_{d_l}$, $\check{\mathcal{S}}^c \triangleq \mathcal{S}^c \setminus \check{\mathcal{S}}$. □

The implications of the inequalities in (4) can be inferred by considering a three node network. In such a network, there are two possibilities for $\mathcal{S}$: $\mathcal{S} = \{1\}$ and $\mathcal{S} = \{1, 2\}$. For the first possibility, there are two cases for $\check{\mathcal{S}}$: $\check{\mathcal{S}} = \emptyset$ and $\check{\mathcal{S}} = \{2\}$. Using these possibilities in (4) yields the following bounds on the source rate, R_1:

for $\mathcal{S} = \{1\}, \quad \mathcal{S}^c = \{2\}, \quad \check{\mathcal{S}} = \emptyset, \quad \check{\mathcal{S}}^c = \{2\}$,

$$R_1 \leq I(X_1; \hat{Y}_2, Y_{\mathrm{D}} | X_2); \quad (5a)$$

for $\mathcal{S} = \{1, 2\}, \quad \mathcal{S}^c = \emptyset, \quad \check{\mathcal{S}} = \check{\mathcal{S}}^c = \emptyset$,

$$R_1 \leq I(X_1, X_2; Y_{\mathrm{D}}) - I(\hat{Y}_2; Y_2 | X_1, X_2, Y_{\mathrm{D}}); \quad (5b)$$

for $\mathcal{S} = \{1\}, \quad \mathcal{S}^c = \{2\}, \quad \check{\mathcal{S}} = \{2\}, \quad \check{\mathcal{S}}^c = \emptyset$,

$$R_1 \leq I(X_1; Y_{\mathrm{D}} | X_2) - I(\hat{Y}_2; Y_2 | X_1, X_2, Y_{\mathrm{D}}) + \check{R}_2. \quad (5c)$$

The first two bounds are identical to the standard CF bounds, cf. [2, Sec. 16.7] and (1a) and (1b). For the third bound we note that choosing $\check{R}_2 \geq I(\hat{Y}_2; Y_2 | X_2, Y_{\mathrm{D}})$ reduces (5c) to (5a), and choosing $\check{R}_2 \geq I(X_2; Y_{\mathrm{D}})$ reduces (5c) to (5b). This implies that for this scenario, (5c) is redundant and subsequently, the proposed scheme does not yield an advantage beyond conventional CF in the three node network.

The expression on the right hand side of (4) can be regarded as a generalization of the rate achieved by NNC and SNNC. In particular, as shown in Remark 2, when $R_{\check{\mathcal{S}}}$ satisfies the conditions in (12), the rate expression in Theorem 1 reduces to the one achieved by NNC and SNNC. As such, using the proposed signalling strategy in particular networks can in general yield a rate region that includes the rate region that can be achieved by the NNC and SNNC schemes; the additional advantage of the proposed scheme follows from exploiting the N-to-1 mapping as elucidated in detail in the proof of Theorem 1 and the examples in Sect. IV-A and IV-B.

Proof: The decoding procedure uses strong joint typicality [17] and features a layered framework. For layer 1 in this framework, the receiver considers the codebooks of all the nodes in the network and constructs the set containing the codewords that are jointly typical with the signal received in a particular block. For layer 2, the receiver considers only those codebooks that correspond to exactly one codeword in the joint typicality set in layer 1, and subsequently constructs the set containing the codewords that are jointly typical with the signal received in the following block. Hence, the codebooks considered at layer 2 is only a subset of those at layer 1. Subsequent layers are constructed in a similar way until all the relay description codewords are successfully decoded. (Further discussions on the number of decoding layers will be provided a the end of Sect. III.) Using the joint typicality sets at all the layers jointly, the receiver recovers the source messages. The details are as follows.

Codebook Generation: For node d_k, generate $2^{n(R_k+\check{R}_k)}$ independent identically distributed (i.i.d.) $\mathbf{x}_k(m_k, s_k)$, each according to the distribution $p(\mathbf{x}_k) = \prod_{i=1}^{n} p(x_{ki})$, $m_k \in [1 : 2^{nR_k}]$, $s_k \in [1 : 2^{n\check{R}_k}]$; for each $\mathbf{x}_k(m_k, s_k)$, generate $2^{n\hat{R}_k}$ i.i.d. $\hat{\mathbf{y}}_k(z_k|m_k, s_k)$, each according to the distribution $p(\hat{\mathbf{y}}_k|\mathbf{x}_k) = \prod_{i=1}^{n} p(\hat{y}_{ki}|x_{ki})$, $z_k \in [1 : 2^{n\hat{R}_k}]$.

Random Binning: For node d_k, randomly partition the set $\{1, \cdots, 2^{n\hat{R}_k}\}$ into $2^{n\check{R}_k}$ bins. Let $s_k = \mathcal{B}_k(z_k)$ denote the N-to-1 mapping corresponding to random binning at d_k.

Encoding: Let b be the current block. At the end of block b, node d_k

- finds z_k such that $(\hat{\mathbf{y}}_k(z_k|m_{k,b}, s_{k,b}), \mathbf{x}_k(m_{k,b}, s_{k,b}), \mathbf{y}_{d_k}(b))$ are jointly ϵ-typical. By the covering lemma in [2], such a z_k exists as $n \to \infty$ if

$$\hat{R}_k \geq I(\hat{Y}_k; Y_{d_k}|X_k). \tag{6}$$

  If more than one such z_k exist, choose the smallest z_k and let $z_{k,b} = z_k$;
- determines $s_k = \mathcal{B}_k(z_{k_b})$ and lets $s_{k,b+1} = s_k$.

Codewords $\mathbf{x}_k(m_{k,b}, s_{k,b})$ are sent in block b from all $d_k \in \mathcal{N}$.

Decoding Procedure: We provide a layered forward decoding procedure for d_k.

Let i, j denote the decoding layer and the block number of the received signal used at layer i, respectively. Let ℓ_{d_k} be the layer at which the decoding at d_k ends. We have

$$j = b - \ell_{d_k} + i. \tag{7}$$

Using this relationship, we drop the block number j from the expressions in the analysis of the decoding procedure when it is clear. Furthermore, in analyzing the decoding at d_k, node identity d_k is omitted from the subscript of ℓ_{d_k} ($\ell \triangleq \ell_{d_k}$) and various sets of nodes when it is clear.

Let $\mathcal{L}$ be the maximum number of the decoding layers in the network, i.e., $\mathcal{L} \triangleq \max_{d_k \in \mathcal{N}} \ell_{d_k}$. At any $d_k \in \mathcal{N}$,

$$1 \leq i \leq \ell \leq \mathcal{L} \leq N - 1. \tag{8}$$

Consider the receiver at d_k. Assuming that in block $b \geq \ell$, $s_{\mathcal{N},b-\ell}$ has been successfully recovered, the receiver at d_k

- constructs the following jointly ϵ-typical set:

$$(\mathbf{x}_{\mathcal{A}_{m,i}}, \hat{\mathbf{y}}_{\mathcal{A}_{z(J),i}}, \mathbf{y}_{d_k}(b - \ell + i)). \tag{9}$$

  (The sets of nodes $\mathcal{A}_{m,i}$ and $\mathcal{A}_{z(J),i}$ for $i \leq \ell$ will be made clear below.)
- forms the set $\mathcal{A}_{m,i+1}$: for each node d_l in this set, there exists a unique $\mathbf{x}_{d_l}(\hat{m}_{d_l})$ in the jointly ϵ-typical set in (9), and $d_l \in \mathcal{A}_{m,i}$.
- forms the set $\mathcal{A}_{z(J^c),i+1}$: which contains the node d_l, $d_l \in \mathcal{A}_{m,i+1} \cap \mathcal{A}^c_{z(J),j}$, for all $j < i$; (The set $\mathcal{A}^c_{z(J),j}$ is defined below.)
- forms the set $\mathcal{A}^c_{z(J),i+1}$: for each node d_l in this set, there exist multiple $\hat{\mathbf{y}}_{d_l}(\hat{z}_{d_l}|\hat{m}_{d_l})$ in the jointly ϵ-typical set in (9), and $d_l \in \mathcal{A}_{m,i+1} \setminus \mathcal{A}_{z(J^c),i+1}$;
- forms the set $\mathcal{A}_{z(J),i+1}$: for each node d_l in this set, there exists a unique $\hat{\mathbf{y}}_{d_l}(\hat{z}_{d_l}|\hat{m}_{d_l})$ in the jointly ϵ-typical set in (9), and $d_l \in \mathcal{A}_{m,i+1} \setminus \mathcal{A}_{z(J^c),i+1}$;
- proceeds to layer $i + 1$ if $\mathcal{A}^c_{z(J),i+1} \neq \emptyset$;
- ends at layer i when $\mathcal{A}^c_{z(J),i+1} = \emptyset$.

  Using the jointly ϵ-typical sets in ℓ layers jointly, the receiver at d_k declares that $m_{\mathcal{S}_{d_k}} = \hat{m}_{\mathcal{S}_{d_k}}$ was sent in block $b-\ell$; and that $s_{\mathcal{N}} = \hat{s}_{\mathcal{N}}$ was sent in block $b-\ell+1$.

Let $\mathcal{A}_{\ell_m} \triangleq \mathcal{A}_{m,\ell+1} = \mathcal{A}_{z(J),\ell+1} \cup \mathcal{A}_{z(J^c),\ell+1}$ and $\mathcal{A}_{\ell_z} \triangleq \mathcal{A}_{z(J),\ell+1}$. Furthermore, let $\mathcal{A}_{m,1} \triangleq \mathcal{N}$, $\mathcal{A}_{z(J),1} \triangleq \mathcal{N}$, $\mathcal{A}^c_{z(J),1} \triangleq \emptyset$ and $\mathcal{A}_{z(J^c),1} \triangleq \emptyset$. In Lemmas 1 and 2, we provide useful properties of the above sets.

Lemma 1: By definition, the sets formed by d_k in the decoding procedure have the following properties, for $i \leq \ell$,

1. $\mathcal{A}^c_{m,i+1} = \mathcal{A}_{m,i} \setminus \mathcal{A}_{m,i+1}$;
2. $\mathcal{A}_{z(J),i}$, $\mathcal{A}^c_{z(J),i}$ and $\mathcal{A}_{z(J^c),i}$ are disjoint, $\mathcal{A}_{z(J),i} \cup \mathcal{A}^c_{z(J),i} \cup \mathcal{A}_{z(J^c),i} = \mathcal{A}_{m,i}$;
3. $\mathcal{A}_{\ell_m} \subseteq \mathcal{A}_{m,i} \subseteq \mathcal{A}_{m,j}$, for $i > j$;
4. $(\mathcal{A}_{z(J),i} \cup \mathcal{A}^c_{z(J),i}) = \mathcal{A}_{z(J),i-1} \subseteq \mathcal{A}_{z(J),j}$, for $i > j$;
5. $\mathcal{A}_{z(J),i} \cap \mathcal{A}_{z(J^c),i+1} = \emptyset$;
6. $\mathcal{A}_{\ell_z} \subseteq \mathcal{A}_{z(J),i+1} \subseteq \mathcal{A}_{z(J),i}$;
7. For the receiver at d_k, $d_k \in \mathcal{D}_{\mathcal{S}_{d_k}}$ and $d_k \in \mathcal{A}_{m,i}$, $d_k \in \mathcal{A}_{z(J),i}$. Hence, $d_k \in \mathcal{A}_{\ell_m}$, $d_k \in \mathcal{A}_{\ell_z}$.
8. $\mathcal{A}^c_{z(J),\ell+1} = \emptyset$.

Lemma 2: For the sets formed by the receiver, define $\mathcal{A}_{z(J),i\setminus i+1} \triangleq ((\mathcal{A}^c_{m,i+1} \setminus \mathcal{A}^c_{z(J),i}) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}^c_{z(J),i+1}$. The following equality holds:

$$\mathcal{A}_{z(J),i\setminus i+1} = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}. \tag{10}$$

Proof: See details in Appendix A. ■

Analysis of the Probability of Error: See detailed analysis in Appendix B. ■

Now, we make three remarks.

Remark 1: When $\mathcal{A}_{\ell_m} = \mathcal{A}_{\ell_z}$, we have $\check{\mathcal{S}} = \mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z} = \emptyset$ and $\check{\mathcal{S}}^c = \mathcal{S}^c$. The result in Theorem 1 reduces to the following simplified form:

$$R(\mathcal{S}) \leq I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k}|X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}}|X_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}), \tag{11}$$

for $\mathcal{S} \cap \mathcal{S}_{d_k} \neq \emptyset$.

The simplified form of the achievable rate region coincides with that of NNC and SNNC.

Remark 2: The rate region described in Theorem 1 reduces to the simplified form in (11) when $\check{R}_{\check{\mathcal{S}}}$ satisfies the following condition:

$$\check{R}_{\check{\mathcal{S}}} \geq \min\{I(X_{\mathcal{S}^c\setminus\check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\check{\mathcal{S}}^c}) - R_{\mathcal{S}^c\setminus\check{\mathcal{S}}^c}, I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}}|X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k})\}. \quad (12)$$

Proof: See detailed proof in Appendix C. ■

Remark 3: Theorem 1 shows that the proposed scheme achieves the same rate region as NNC and SNNC but with reduced rates of codewords representing the bin indices. This reduction will be shown to result in a rate advantage in Sect. IV.

Proof: To expose the reduction in the rate of the bin indices, it suffices to show that the right hand side of (12) is upper bounded by $\hat{R}_{\check{\mathcal{S}}}$. We have

$$\begin{aligned} I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}}|X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) &= \sum_{i\in\check{\mathcal{S}}} I(\hat{Y}_i; Y_{\check{\mathcal{S}}}|X_{\mathcal{S}^c}, \hat{Y}_{\mathcal{K}_i}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &= \sum_{i\in\check{\mathcal{S}}} I(\hat{Y}_i; Y_i|X_{\mathcal{S}^c}, \hat{Y}_{\mathcal{K}_i}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &\leq \sum_{i\in\check{\mathcal{S}}} I(\hat{Y}_i; Y_i|X_i) \\ &\leq \hat{R}_{\check{\mathcal{S}}}, \end{aligned}$$

where we have used $\check{\mathcal{S}} \subseteq \mathcal{S}^c$. Clearly, the lower bound on $\check{R}_{\check{\mathcal{S}}}$ is lower than $\hat{R}_{\check{\mathcal{S}}}$. ■

We now consider the case in which the transmitted codewords from the nodes in $\mathcal{N}\setminus\mathcal{S}_{d_k}$ can only provide information about the bin indices to facilitate decoding at d_k, and the receiver at d_k treats the information representing the message indices in $\mathbf{X}_{\mathcal{N}\setminus\mathcal{S}_{d_k}}$ as noise. Using this approach, we provide the following theorem.

Theorem 2: Let $(\times_{k=1}^N \mathcal{X}_k, p(y^N|x^N), \times_{k=1}^N \mathcal{Y}_{d_k})$ be the general discrete memoryless multimessage network. A rate tuple $(R_1, \ldots, R_N)$ is achievable if

$$R_{\mathcal{T}} \leq I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{N}\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}}, \quad (13)$$

for all subsets $\mathcal{S} \subset \mathcal{N}$, $\check{\mathcal{S}} \subseteq \mathcal{S}^c = \mathcal{N}\setminus\mathcal{S}$ and $\mathcal{T} \subseteq \mathcal{S}_{d_k}$, where $\mathcal{T} \neq \emptyset$; the subset $\check{\mathcal{S}}^c = \mathcal{S}^c\setminus\check{\mathcal{S}}$ and $\mathcal{T}^c = \mathcal{S}_{d_k}\setminus\mathcal{T}$. □

Similar to the case of Theorem 1, we note that the expression on the right hand side of (13) can be regarded as a generalization of the rate achieved by NNC. In particular, when $R_{\check{\mathcal{S}}}$ satisfies the criteria in (18), the right hand side of (13) reduces to (11), cf. Remark 5 and [11, Th. 3]. Hence, in general, the signalling strategy that underlies Theorem 2 offers the potential of achieving rate regions that include those achieved by the corresponding NNC scheme; the advantage of that signalling strategy follows from exploiting the N-to-1 mapping as elucidated in detail in the proof of Theorem 2 and the example in Sect. IV-B.

Proof: Using the same philosophy as that used in the proof of Theorem 1, the decoding procedure herein also features a layered framework. Details are provided below.

Codebook Generation: For node d_k, generate $2^{n\check{R}_k}$ i.i.d. $\mathbf{u}_k(s_k)$, each according to the distribution $p(\mathbf{u}_k) = \prod_{i=1}^n p(u_{ki})$, $s_k \in [1 : 2^{n\check{R}_k}]$; for each $\mathbf{u}_k(s_k)$, generate 2^{nR_k} i.i.d. $\mathbf{x}_k(m_k|s_k)$, each according to the distribution $p(\mathbf{x}_k|\mathbf{u}_{ki}) = \prod_{i=1}^n p(x_{ki}|s_{ki})$, $m_k = [1 : 2^{nR_k}]$; for each $\mathbf{u}_k(s_k)$, generate $2^{n\hat{R}_k}$ i.i.d. $\hat{\mathbf{y}}_k(z_k|s_k)$, each according to the distribution $p(\hat{\mathbf{y}}_k|\mathbf{u}_k) = \prod_{i=1}^n p(\hat{y}_{ki}|s_{ki})$, $z_k \in [1 : 2^{n\hat{R}_k}]$.

Random Binning: For node d_k, randomly partition the set $\{1, \cdots, 2^{n\hat{R}_k}\}$ into $2^{n\check{R}_k}$ bins. Let $s_k = \mathcal{B}_k(z_k)$ denote the N-to-1 mapping at d_k as the result of binning.

Encoding: Let b be the current block. At the end of block b, node d_k

- finds an index z_k such that $(\hat{\mathbf{y}}_k(z_k|s_{k,b}), \mathbf{u}_k(s_{k,b}), \mathbf{y}_{d_k}(b))$ are jointly ϵ-typical. By the covering lemma in [2], such a z_k exists as $n \to \infty$ if
$$\hat{R}_k \geq I(\hat{Y}_k; Y_{d_k}|U_k). \quad (14)$$
If more than one such z_k exist, choose the smallest z_k and let $z_{k,b} = z_k$;
- determines $s_k = \mathcal{B}_k(z_{k_b})$ and lets $s_{k,b+1} = s_k$.

Codewords $\mathbf{x}_k(m_{k,b}|s_{k,b})$ are sent in block b from all $d_k \in \mathcal{N}$.

Decoding Procedure: Similar to the procedure provided in the proof of Theorem 1, the procedure herein employs the layered forward decoding strategy, and the relationships in (7) and (8) hold. The main difference between the two procedures lies in the sets formed at the receiver at each decoding layer.

Consider the receiver at d_k. Assuming that in block b, $s_{\mathcal{N},b-\ell}$ has been successfully recovered, where $\ell \triangleq \ell_{d_k}$, the receiver at d_k

- constructs the following jointly ϵ-typical set:
$$(\mathbf{x}_{\mathcal{A}_{m,i}}, \mathbf{u}_{\mathcal{A}_{s,i}}, \hat{\mathbf{y}}_{\mathcal{A}_{z(J),i}}, \mathbf{y}_{d_k}(b-\ell+i)). \quad (15)$$
(The definition of the sets of nodes $\mathcal{A}_{m,i}$, $\mathcal{A}_{s,i}$ and $\mathcal{A}_{z(J),i}$ for $i \leq \ell$ is given below.)
- forms the set $\mathcal{A}_{s,i+1}$: for each node d_l in this set, there exists a unique $\mathbf{u}_{d_l}(\hat{s}_{d_l})$ in the jointly ϵ-typical set in (15), and $d_l \in \mathcal{A}_{s,i}$;
- forms the set $\mathcal{A}_{m,i+1}$: for each node d_l in this set, there exists a unique $\mathbf{x}_{d_l}(\hat{m}_{d_l}|\hat{s}_{d_l})$ in the jointly ϵ-typical set in (15), and $d_l \in \mathcal{A}_{m,i} \cap \mathcal{A}_{s,i+1}$;
- forms set $\mathcal{A}_{z(J^c),i+1}$, which contains the node $d_l \in \mathcal{A}_{s,i+1} \cap \mathcal{A}^c_{z(J),j}$, for all $j < i$; (The set $\mathcal{A}^c_{z(J),j}$ is defined below.)
- forms the set $\mathcal{A}^c_{z(J),i+1}$: for each node d_l in this set, there exist multiple $\hat{\mathbf{y}}_{d_l}(\hat{z}_{d_l}|\hat{s}_{d_l})$ in the jointly ϵ-typical set in (15), and $d_l \in \mathcal{A}_{s,i+1}\setminus\mathcal{A}_{z(J^c),i+1}$;
- forms the set $\mathcal{A}_{z(J),i+1}$: for each node d_l in this set, there exists a unique $\hat{\mathbf{y}}_{d_l}(\hat{z}_{d_l}|\hat{s}_{d_l})$ in the jointly ϵ-typical set in (15), and $d_l \in \mathcal{A}_{s,i+1}\setminus\mathcal{A}_{z(J^c),i+1}$;
- proceeds to layer $i+1$ if $\mathcal{A}^c_{z(J),i+1} \neq \emptyset$.
- ends at layer i when $\mathcal{A}_{z(J),i+1} = \emptyset$.
Using the jointly ϵ-typical sets in ℓ layers jointly, the receiver at d_k declares that $m_{\mathcal{S}_{d_k}} = \hat{m}_{\mathcal{S}_{d_k}}$ were sent in block $b-\ell$; and that $s_{\mathcal{N}} = \hat{s}_{\mathcal{N}}$ were sent in block $b-\ell+1$.

Let $\mathcal{A}_{\ell_m} \triangleq \mathcal{A}_{m,\ell+1}$, $\mathcal{A}_{\ell_s} \triangleq \mathcal{A}_{s,\ell+1} = \mathcal{A}_{z(J),\ell+1} \cup \mathcal{A}_{z(J^c),\ell+1}$ and $\mathcal{A}_{\ell_z} \triangleq \mathcal{A}_{z(J),\ell+1}$. Furthermore, let $\mathcal{A}_{m,1} \triangleq \mathcal{S}_{d_k}$, $\mathcal{A}_{s,1} \triangleq \mathcal{N}$, $\mathcal{A}_{z(J),1} \triangleq \mathcal{N}$, $\mathcal{A}^c_{z(J),1} \triangleq \emptyset$ and $\mathcal{A}_{z(J^c),1} \triangleq \emptyset$. In Lemmas 3 and 4 we provide the counterparts of Lemmas 1 and 2 for the decoding procedure of Theorem 2.

Lemma 3: The sets formed by the receiver at $d_k \in \mathcal{N}$ in the decoding procedure have the following properties, for $i \leq \ell$,

1. $\mathcal{A}^c_{m,i+1} = \mathcal{A}_{m,i} \setminus \mathcal{A}_{m,i+1}$, $\mathcal{A}^c_{s,i+1} = \mathcal{A}_{s,i} \setminus \mathcal{A}_{s,i+1}$;
2. $\mathcal{A}_{z(J),i}$, $\mathcal{A}^c_{z(J),i}$ and $\mathcal{A}_{z(J^c),i}$ are disjoint, $\mathcal{A}_{z(J),i} \cup \mathcal{A}^c_{z(J),i} \cup \mathcal{A}_{z(J^c),i} = \mathcal{A}_{s,i}$;
3. $\mathcal{A}_{\ell_m} \subseteq \mathcal{A}_{m,i} \subseteq \mathcal{A}_{m,j}$, $\mathcal{A}_{\ell_s} \subseteq \mathcal{A}_{s,i} \subseteq \mathcal{A}_{s,j}$, for $i > j$;
4. $(\mathcal{A}_{z(J),i} \cup \mathcal{A}^c_{z(J),i}) = \mathcal{A}_{z(J),i-1} \subseteq \mathcal{A}_{z(J),j}$, for $i > j$;
5. $\mathcal{A}_{z(J),i} \cap \mathcal{A}_{z(J^c),i+1} = \emptyset$;
6. $\mathcal{A}_{\ell_z} \subseteq \mathcal{A}_{z(J),i+1} \subseteq \mathcal{A}_{z(J),i}$;
7. For the receiver at $\forall d_k$, $d_k \in \mathcal{D}_{\mathcal{S}_{d_k}}$ and $d_k \in \mathcal{A}_{m,i}$, $d_k \in \mathcal{A}_{z(J),i}$. Hence, $d_k \in \mathcal{A}_{\ell_m}$, $d_k \in \mathcal{A}_{\ell_z}$.
8. $\mathcal{A}^c_{z(J),\ell+1} = \emptyset$;
9. $\mathcal{A}_{z(J),i} \subseteq \mathcal{A}_{s,i}$, $\mathcal{A}_{m,i} \subseteq \mathcal{A}_{s,i}$.

Lemma 4: For the sets formed by the receiver, define $\mathcal{A}_{z(J),i\setminus i+1} \triangleq ((\mathcal{A}^c_{s,i+1} \setminus \mathcal{A}^c_{z(J),i}) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}^c_{z(J),i+1}$. The following equality holds:

$$\mathcal{A}_{z(J),i\setminus i+1} = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}. \tag{16}$$

Proof: Replacing $\mathcal{A}^c_{m,i+1}$ and $\mathcal{A}_{m,i}$ by $\mathcal{A}^c_{s,i+1}$ and $\mathcal{A}_{s,i}$, respectively, the lemma can be proved using a technique similar to the one in the proof of Lemma 2. Details are omitted. ■

Analysis of the Probability of Error: See details in Appendix D. ■

Next, we make three remarks.

Remark 4: When $\mathcal{A}_{\ell,s} = \mathcal{A}_{\ell,z}$, we have $\check{\mathcal{S}} = \mathcal{A}_{\ell_s} \setminus \mathcal{A}_{\ell,z} = \emptyset$ and $\check{\mathcal{S}}^c = \mathcal{S}^c$. The result in Theorem 2 reduces to the following simplified form:

$$\begin{aligned} R_{\mathcal{T}} \leq\ & I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) \\ & - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}). \end{aligned} \tag{17}$$

for $\mathcal{T} \neq \emptyset$. □

The simplified form of the achievable rate region coincides with [11, Th. 3].

Remark 5: The rate region described in Theorem 2 reduces to the simplified form in (17) when $\check{R}_{\check{\mathcal{S}}}$ satisfies the following condition:

$$\begin{aligned} \check{R}_{\check{\mathcal{S}}} \geq \min\{ & I(U_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\check{\mathcal{T}}^c}, U_{\check{\mathcal{S}}^c}), \\ & I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k})\}. \end{aligned} \tag{18}$$

Proof: Detailed proof is provided in Appendix E. ■

Remark 6: Theorem 2 shows that the proposed scheme achieves the same rate region as NNC but with reduced bin indices rates. This reduction will be shown to result in a rate advantage in Sect. IV-B.

Proof: To expose the reduction in the rate of the bin indices, it suffices to show that the right hand side of (18) is upper bounded by $\hat{R}_{\check{\mathcal{S}}}$. Towards that end, we write

$$\begin{aligned} & I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ & = \sum_{i \in \check{\mathcal{S}}} I(\hat{Y}_i; Y_{\check{\mathcal{S}}} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\mathcal{K}_i}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ & = \sum_{i \in \check{\mathcal{S}}} I(\hat{Y}_i; Y_i | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\mathcal{K}_i}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ & \leq \sum_{i \in \check{\mathcal{S}}} I(\hat{Y}_i; Y_i | U_i) \\ & \leq \hat{R}_{\check{\mathcal{S}}}, \end{aligned}$$

where we have used $\check{\mathcal{S}} \subseteq \mathcal{S}^c$. Clearly, the lower bound on $\check{R}_{\check{\mathcal{S}}}$ is lower than $\hat{R}_{\check{\mathcal{S}}}$. ■

Since the 1-to-1 mapping satisfies the condition in (18), the achievable rate region of the simplified form in (17) can also be obtained by using short message encoding with 1-to-1 mapping. This observation extends the results provided in [13] and [15] for SNNC.

To obtain an upper bound on the maximum number of decoding layers, we recall that the decoding procedure described in the proofs of Theorems 1 and 2 progresses from one layer to the next depending on the outcome of a joint-typicality test. In particular, if at a given layer i, multiple description codewords of a particular node are found to be in the joint-typicality set, the decoding procedure excludes the codebooks corresponding to the multiple description codewords and progresses to the next layer, i.e., layer $i + 1$. The decoding procedure stops once it reaches a layer with no multiple description codewords of any node in the joint-typicality set. From this procedure, it can be seen that the number of candidate description codebooks examined in each layer decreases strictly from one layer to the next, which implies that the maximum number of layers, $\mathcal{L}$, is upper bounded by $N - 1$. Hence, the decoding delay in the network can be upper bounded by $N - 1$ blocks. In block 1, each node uses a message index and a known bin index in encoding. From block 2, each node performs the encoding as described in the procedure provided in the proofs of the theorems. To end the transmission, each node continues to encode and transmit for $N - 2$ blocks using a known message index and the bin index of the description of its received signal in the previous block.

Finally, we note that for the cooperative multimessage network shown in Fig. 2, the results of using N-to-1 mapping in Theorem 1 and 2 do not yield a rate gain in comparison with their simplified forms which can also be achieved by SNNC (with 1-to-1 mapping). However, our goal is not to show the rate advantage of the N-to-1 mapping in this network, but rather to show its advantage in the network instances considered in the next section.

IV. Exploiting the Gain of N-to-1 Mapping

In this section, we consider two network instances in which side information is only available to a subset, but not to the rest of the receiving nodes. We will show that for these instances the relaxation of the bin indices rate constraints resulting from the N-to-1 mapping renders the decoding procedures

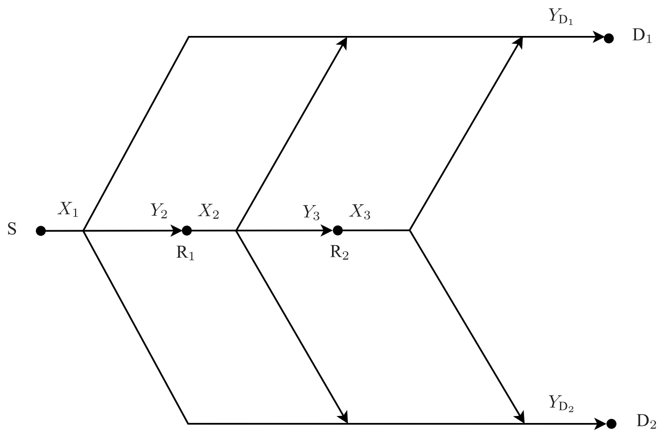

Fig. 3. A two-destination broadcast relay chain network.

in Theorems 1 and 2 more advantageous than their SNNC counterparts.

The first network is a relay chain one in which a source broadcasts a common message to two destinations, which are assisted by a chain of two cascaded relays. In this chain, the first relay receives signals from the source, uses the CF strategy and forwards the bin indices of its description codewords to the second relay and the two destinations. The second relay only receives signals from the first relay and uses the DF strategy to assist the destinations in recovering the bin indices transmitted by the first relay. In this case, the relaxation of the rate of the bin indices of the CF description codewords provided by Theorem 1 benefits the decoding at the second relay and is able to provide a rate gain over SNNC and conventional CF.

The second network is a multimessage one in which, in addition to the source messages, a CF relay has its own independent messages to send to a separate receiver that has no access to side information. The transmission of the bin indices by the CF relay causes interference to this receiver. We show that the N-to-1 mapping that underlies Theorems 1 and 2 results in a reduction in the rate of the bin indices, thereby providing a rate gain over other schemes that use 1-to-1 mapping.

We now analyze the achievable rates of these two networks.

A. Achievable Rate of a Broadcast Relay Chain Network

We now describe the two-destination relay network shown in Fig. 3. In this network, a source S wishes to send a common message to two destinations D_1 and D_2 through the transmission of $\mathbf{X}_1$ with the assistance of a CF relay R_1 and a DF relay R_2 in a chain. The received signals at relay R_1 and R_2 are denoted by $\mathbf{Y}_i, i = 2, 3$, respectively, and the received signal $\mathbf{Y}_3$ is independent of $\mathbf{X}_1$. The received signals at D_1 and D_2 are denoted by $\mathbf{Y}_{D_i}, i = 1$ and 2, respectively.

Without R_2, the network reduces to a broadcast relay channel, cf. [18]. In that case, using N-to-1 mapping yields the same achievable rate R_1 as 1-to-1 mapping. However, since $\mathbf{Y}_3$ is independent of $\mathbf{X}_1$, correct decoding at R_2 imposes a constraint on the rate of the CF bin indices.

We evaluate three relaying strategies in the analysis of the achievable rate of this network. In all the considered strategies, node S uses the standard CF codebook structure. The relay R_2 uses the standard DF codebook structure and procedure.[1] The difference between these three strategies lies in the way that the relay R_1 operates. In particular, in

- Strategy 1, the decoding procedure combines DF decoding and the decoding procedure of Theorem 1;
- Strategy 2, the decoding procedure combines DF decoding and the decoding procedure of SNNC; and in
- Strategy 3, the decoding procedure combines DF decoding and the decoding procedure of conventional CF.

The detailed procedures are provided in Appendix F.

Next, we provide the achievable rate expressions for these strategies for the discrete memoryless case and the Gaussian case.

1) The Discrete Memoryless Case: The achievable rate corresponding to the above relaying strategies for the discrete memoryless case are provided in the following corollary.

Corollary 1: For the discrete memoryless network in Fig. 3, $(\mathcal{X}_1, p(y_2|x_1)p(y_3|x_2)p(y_{D_1}, y_{D_2}|x_1, x_2, x_3), \mathcal{Y}_2 \times \mathcal{Y}_3 \times \mathcal{Y}_{D_1} \times \mathcal{Y}_{D_2})$, the rate R_1 is achievable, where

- for Strategy 1,

$$\begin{aligned} R_1 \le \sup \min_{i=1,2} \min\{&I(X_1; \hat{Y}_2, Y_{D_i}|X_2, X_3), \\ &I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) \\ &+ \min\{I(X_2; Y_3|X_3), I(X_2, X_3; Y_{D_i})\}\}; \end{aligned} \tag{19}$$

- for Strategy 2,

$$\begin{aligned} R_1 \le \sup \min_{i=1,2} \min\{&I(X_1; \hat{Y}_2, Y_{D_i}|X_2, X_3), \\ &I(X_1, X_2, X_3; Y_{D_i}) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}), \end{aligned} \tag{20}$$

subject to

$$I(\hat{Y}_2; Y_2|X_2, X_3) \le I(X_2; Y_3|X_3);$$

- for Strategy 3,

$$\begin{aligned} R_1 \le \sup \min_{i=1,2} \{\min\{&I(X_1; \hat{Y}_2, Y_{D_i}|X_2, X_3), \\ &I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) \\ &+ \min\{I(X_2; Y_3|X_3), \min_{i=1,2} I(X_2, X_3; Y_{D_i})\}\}\}, \end{aligned} \tag{21}$$

where, for all strategies, the supremum is taken over the pmfs of the form

$$\begin{aligned} &p(x_1, x_2, x_3, y_2, y_3, y_{\mathcal{D}_S}, \hat{y}_2) = p(x_1)p(x_2|x_3)p(x_3) \\ &p(y_2|x_1)p(y_3|x_2)p(y_{D_1}, y_{D_2}|x_1, x_2, x_3)p(\hat{y}_2|x_2, x_3, y_2). \end{aligned}$$

□

Proof: See details in Appendix. F ■

We note that Strategies 2 and 3 can be regarded as special cases of Strategy 1. Hence, Strategy 1 offers the potential

[1] Herein we provide the achievable rate expressions when R_2 operates in the DF mode. The counterparts of the expressions when R_2 operates in the CF mode can be readily obtained using the procedure in Theorem 1, SNNC and conventional CF. However, to maintain focus, the expressions pertaining to this case are not presented.

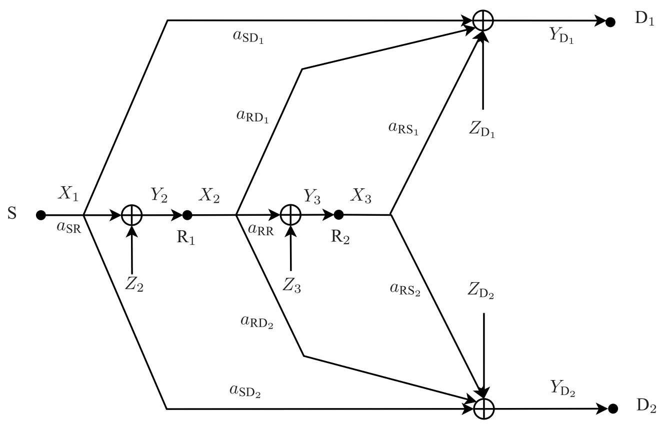

Fig. 4. A two-destination Gaussian broadcast relay network with a relay chain, where Z' is an independent additive noise and Z_2, Z_3, Z_{D_1} and $Z_{D_2} \sim \mathcal{N}(0, 1)$.

of yielding a higher achievable rate than Strategies 2 and 3. We will next show that this is actually the case when the network is Gaussian.

2) The Gaussian Case: Now, we compare the considered strategies for the Gaussian network depicted in Fig. 4. The network shown in this figure is composed of scalar channel coefficients, independent additive Gaussian noises and Gaussian codebooks with average power constraints. As shown in Fig. 4, the transmitted signals from nodes S, R_1 and R_2 are denoted by $X_i \sim \mathcal{N}(0, P_i)$, where P_i is the average transmit power, $i = 1, 2, 3$, respectively. The gain of the S-to-R_1 and S-to-D_i links are denoted by a_{SR} and $a_{SD_i}, i = 1, 2$. The gain of the R_1-to-R_2 and R_1-to-D_i links are denoted by a_{RR} and $a_{RD_i}, i = 1, 2$, respectively. The gain of the R_2-to-D_i link is denoted by $a_{RS_i}, i = 1, 2$. The independent additive noises on the S-to-R_1, and R_1-to-R_2 links are denoted by Z_2 and Z_3, respectively, and that at the receiver D_i is denoted by $Z_{D_i}, i = 1, 2$. All noises are Gaussian distributed with zero mean and unit variance. Using this notation, the received signals at R_1, R_2 and $D_i, i = 1, 2$, can be expressed as

$$\begin{aligned}
Y_2 &= a_{SR}X_1 + Z_2,\\
Y_3 &= a_{RR}X_2 + Z_3,\\
Y_{D_1} &= a_{SD_1}X_1 + a_{RD_1}X_2 + a_{RS_1}X_3 + Z_{D_1},\\
Y_{D_2} &= a_{SD_2}X_1 + a_{RD_2}X_2 + a_{RS_2}X_3 + Z_{D_2}. \qquad (22)
\end{aligned}$$

Denoting the description of the received signal at R_1 by $\hat{Y}_2 = Y_2 + N'$, where $N' \sim \mathcal{N}(0, N')$ [19], we now define following signal-to-noise ratios (SNRs):

$$\begin{aligned}
\gamma_{SR} &= a_{SR}^2 P_1, \quad \gamma_{RR} = a_{RR}^2 P_2,\\
\gamma_{SD_i} &= a_{SD_i}^2 P_1, \quad \gamma_{RD_i} = a_{RD_i}^2 P_2\\
\gamma_{RS_i} &= a_{RS_i}^2 P_3, \quad \gamma' = N', \quad i = 1, 2.
\end{aligned}$$

Let $\mathcal{C}(x) = \frac{1}{2}\log_2(1 + x)$ and let ρ be the correlation coefficient between X_2 and X_3, i.e., $\rho = \frac{E(X_2X_3)}{\sqrt{P_2P_3}}$. Using the technique in [20], we have

$$\begin{aligned}
I(X_1; \hat{Y}_2, Y_{D_i}|X_2, X_3) &= \mathcal{C}\Big(\frac{\gamma_{SR}}{1+\gamma'} + \gamma_{SD_i}\Big),\\
I(X_1; Y_{D_i}|X_2, X_3) &= \mathcal{C}(\gamma_{SD_i}),\\
I(X_2, X_3; Y_{D_i}) &= \mathcal{C}\Big(\frac{\gamma_{RD_i} + \gamma_{RS_i} + 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}}{1+\gamma_{SD_i}}\Big),\\
I(X_1, X_2, X_3; Y_{D_i}) &= \mathcal{C}(\gamma_{SD_i} + \gamma_{RD_i} + \gamma_{RS_i}\\
&\quad + 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}),\\
I(X_2; Y_3|X_3) &= \mathcal{C}((1-\rho^2)\gamma_{RR}),\\
I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) &= \mathcal{C}(1/\gamma'),\\
I(\hat{Y}_2; Y_2|X_2, X_3) &= \mathcal{C}\Big(\frac{1+\gamma_{SR}}{\gamma'}\Big).
\end{aligned}$$

Using these results in Corollary 1, we have the following proposition.

Proposition 1: For the network shown in Fig. 4, the rate R_1 is achievable, where

- for Strategy 1:

$$\begin{aligned}
R_1 \le \max_{\rho,\gamma'} \min_{i=1,2} \min\Big\{&\mathcal{C}\Big(\frac{\gamma_{SR}}{1+\gamma'} + \gamma_{SD_i}\Big), \mathcal{C}(\gamma_{SD_i}) - \mathcal{C}\Big(\frac{1}{\gamma'}\Big)\\
&+ \min\Big\{\mathcal{C}\Big(\frac{\gamma_{RD_i} + \gamma_{RS_i} + 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}}{1+\gamma_{SD_i}}\Big),\\
&\mathcal{C}((1-\rho^2)\gamma_{RR})\Big\}\Big\};
\end{aligned}$$

- for Strategy 2:

$$\begin{aligned}
R_1 \le \max_{\rho,\gamma'} \min_{i=1,2} \min\Big\{&\mathcal{C}\Big(\frac{\gamma_{SR}}{1+\gamma'} + \gamma_{SD_i}\Big),\\
&\mathcal{C}(\gamma_{SD_i} + \gamma_{RD_i} + \gamma_{RS_i}\\
&+ 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}) - \mathcal{C}(1/\gamma')\Big\},
\end{aligned}$$

subject to

$$\mathcal{C}\Big(\frac{1+\gamma_{SR}}{\gamma'}\Big) \le \mathcal{C}((1-\rho^2)\gamma_{RR});$$

- for Strategy 3:

$$\begin{aligned}
R_1 \le \max_{\rho,\gamma'} \min_{i=1,2} \Big\{&\min\Big\{\mathcal{C}\Big(\frac{\gamma_{SR}}{1+\gamma'} + \gamma_{SD_i}\Big),\\
&\mathcal{C}(\gamma_{SD_i}) - \mathcal{C}(1/\gamma')\\
&+ \min\Big\{\mathcal{C}((1-\rho^2)\gamma_{RR}),\\
&\min_{i=1,2}\Big(\frac{\gamma_{RD_i} + \gamma_{RS_i} + 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}}{1+\gamma_{SD_i}}\Big)\Big\}\Big\}\Big\},
\end{aligned}$$

where, for all strategies, $\rho \in [-1, 1], \gamma' \ge 0$.

To illustrate the advantage of Strategy 1, in Fig. 5 we plot the achievable rates provided in Proposition 1 for the three strategies when $\gamma_{SR} = 2$, $\gamma_{SD_1} = 1$, $\gamma_{SD_2} = 2$, $\gamma_{RD_1} = 2$, $\gamma_{RD_2} = 1$, $\gamma_{RS_1} = 2$, $\gamma_{RS_2} = 1$ and $0 \le \gamma_{RR} \le 4$.

From Fig. 5 it can be seen that when $\gamma_{RR} \le 0.7$, Strategies 1 and 3 achieve the same rate, which is higher than that achieved by Strategy 2. For $\gamma_{RR} > 0.7$, Strategy 1 achieves a higher rate than both Strategies 2 and 3. Fig. 5 also shows that for $\gamma_{RR} \le 1.6$, Strategy 3 achieves a higher rate than Strategy 2, and for $\gamma_{RR} > 1.6$, Strategy 2 achieves a higher rate than Strategy 3. Note that if γ_{RR} is sufficiently large, the constraint on γ' in the achievable rate of Strategy 2 becomes inactive. In that case, Strategies 1 and 2 yield the same rate.

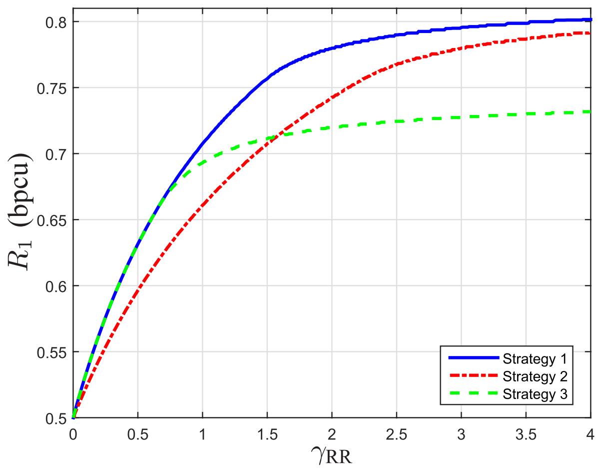

Fig. 5. Achievable rates of a two-destination Gaussian broadcast relay chain network in Fig. 4, $\gamma_{\mathrm{SR}} = 2, \gamma_{\mathrm{SD}_1} = 1, \gamma_{\mathrm{SD}_2} = 2, \gamma_{\mathrm{RD}_1} = 2, \gamma_{\mathrm{RD}_2} = 1, \gamma_{\mathrm{RS}_1} = 2, \gamma_{\mathrm{RS}_2} = 1$.

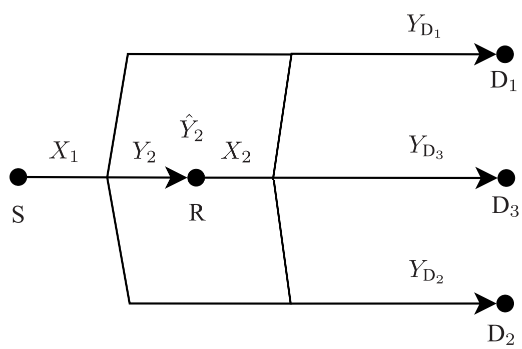

Fig. 6. A two-message three-receiver partially cooperative network.

B. Achievable Rate Region of a Partially Cooperative Multimessage Network

In this section, we consider an example that belongs to a class of multimessage networks. In this class, each source node wishes to send an independent message to its destinations with the assistance of relay nodes. Each relay node has its own independent message and wishes to send it to its own destinations through the direct link without the assistance of other nodes. The destinations of the relay nodes do not have direct links from other nodes. The set of destinations of the source nodes and the set of destinations of the relay nodes are disjoint. Each destination recovers its intended messages without collaboration.

Fig. 6 shows an example of such a network. In this example, source S wishes to send a common message to two receivers D_1 and D_2, with the assistance of node R. In addition, node R has its own independent message and wishes to send it to a third destination, D_3. Since the only link to D_3 is the R-to-D_3 one, node R sends its message to D_3 without being assisted by S. Hence, the network is only a partially cooperative one.

For the network in Fig. 6, we consider a relaying scheme in which node R facilitates decoding at D_1 and D_2 by transmitting the bin index of a description of its received signal. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be the codewords sent by S and R, respectively. Let $\mathbf{Y}_2$ and $\mathbf{Y}_{\mathrm{D}_i}$ be the received signal at R and $\mathrm{D}_i, i = 1, 2, 3$, respectively, and let $\hat{\mathbf{Y}}_2$ be the codewords corresponding to the description of R of its received signal. We use R_1, R_2, $\check{R}_2$ and $\hat{R}_2$ to denote the rate of $\mathcal{X}_1$, the rate of the independent message sent from R to D_3, the rate of the relay bin indices and the rate of the description codebook at the relay, respectively.

Without D_3, the network reduces to a broadcast relay channel, cf. [18]. In that case, Theorems 1 and 2 imply that using either the N-to-1 or the 1-to-1 mapping yields the same achievable rate, R_1. However, the presence of D_3 and the fact that its received signal, $\mathbf{Y}_{\mathrm{D}_3}$, does not contain information about $\mathbf{X}_1$ implies that these rates are not necessarily identical. To explore this possibility, we consider the following decoding strategies:

1. Use SNNC codebook structure at S and R. Use SNNC decoding procedure (Theorem 1 with $\check{R}_2 = \hat{R}_2$) at D_i, $i = 1, 2, 3$. The corresponding achievable rate region is denoted by $\mathcal{R}_1$.
2. Use SNNC codebook structure at S and R. Only decode desired codewords at $\mathrm{D}_i, i = 1, 2, 3$, and treat the undesired signal as noise (Theorem 2 with $\check{R}_2 = \hat{R}_2$). The corresponding achievable rate region is denoted by $\mathcal{R}_2$.
3. Use CF codebook structure at S and R. Use the decoding procedure in Theorem 1 at D_1 and D_2, and directly recover the intended message from R at D_3. The corresponding achievable rate region is denoted by $\mathcal{R}_3$.
4. Use CF codebook structure at S and R. Use the decoding procedure in Theorem 2 at D_1 and D_2, and directly recover the intended message from R at D_3. The corresponding achievable rate region is denoted by $\mathcal{R}_4$.

Next, we provide the achievable rate expressions for these strategies for the discrete memoryless case and the Gaussian case.

1) The Discrete Memoryless Case: In the following corollary we provide expressions for the rate regions that can be achieved by each of the above strategies in the discrete memoryless case:

Corollary 2: For the discrete memoryless network in Fig. 6 $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_{\mathrm{D}_1}, y_{\mathrm{D}_2}|x_1, x_2)p(y_{\mathrm{D}_3}|x_2), \mathcal{Y}_2 \times \mathcal{Y}_{\mathrm{D}_1} \times \mathcal{Y}_{\mathrm{D}_2} \times \mathcal{Y}_{\mathrm{D}_3})$, consider fixed pmf of the form:

- for Strategies 1 and 3,

$$\begin{aligned} p(x_1, x_2, \hat{y}_2, y_2, y_{\mathcal{D}_\mathrm{S}}, y_{\mathrm{D}_3}) &= p(x_1)p(x_2) \\ &\quad \times p(\hat{y}_2|x_2, y_2)p(y_2, y_{\mathrm{D}_1}, y_{\mathrm{D}_2}|x_1, x_2)p(y_{\mathrm{D}_3}|x_2). \end{aligned}$$

- for Strategies 2 and 4,

$$\begin{aligned} p(x_1, x_2, u, \hat{y}_2, y_2, y_{\mathcal{D}_\mathrm{S}}, y_{\mathrm{D}_3}) &= p(x_1)p(x_2|u)p(u) \\ &\quad \times p(\hat{y}_2|u, y_2)p(y_2, y_{\mathrm{D}_1}, y_{\mathrm{D}_2}|x_1, x_2)p(y_{\mathrm{D}_3}|x_2). \end{aligned}$$

Using Strategy 1, the rate pair (R_1, R_2) is achievable, where

$$R_1 + R_2 \leq \min_{i=1,2}\{I(X_1, X_2; Y_{\mathrm{D}_i}) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_{\mathrm{D}_i})\}, \tag{26a}$$

$$R_1 \leq \min_{i=1,2} I(X_1; \hat{Y}_2, Y_{\mathrm{D}_i}|X_2), \tag{26b}$$

$$R_1 + R_2 \leq I(X_2; Y_{\mathrm{D}_3}) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_{\mathrm{D}_3}). \tag{26c}$$

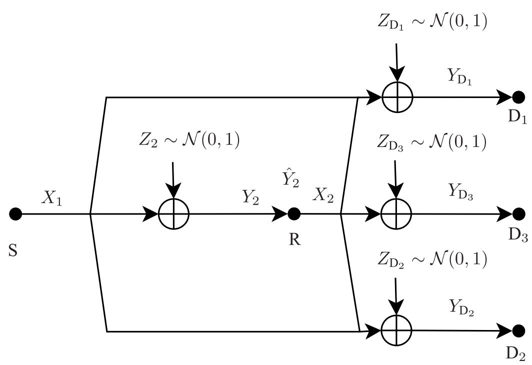

Fig. 7. Gaussian particularization of the network in Fig. 6.

Using Strategy 2, the rate pair (R_1, R_2) is achievable, where

$$R_1 \le \min_{i=1,2}\{I(X_1, U_2; Y_{\mathrm{D}_i}) - I(\hat{Y}_2; Y_2|X_1, U_2, Y_{\mathrm{D}_i})\}, \quad (27a)$$

$$R_1 \le \min_{i=1,2} I(X_1; \hat{Y}_2, Y_{\mathrm{D}_i}|U_2), \quad (27b)$$

$$R_2 \le \min\{I(X_2; Y_{\mathrm{D}_3}|U), I(X_2; Y_{\mathrm{D}_3}) - I(\hat{Y}_2; Y_2|U_2, Y_{\mathrm{D}_3})\}. \quad (27c)$$

Using Strategy 3, the rate pair (R_1, R_2) is achievable, if the sum rate and R_1 satisfy (26a) and (26b), respectively, and the following constraints on R_1 and R_2 are satisfied:

$$R_1 \le I(X_1; Y_{\mathrm{D}_i}|X_2) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_{\mathrm{D}_i}) + \check{R}_2, \quad (28a)$$

$$R_2 \le I(X_2; Y_{\mathrm{D}_3}) - \check{R}_2. \quad (28b)$$

Using Strategy 4, the rate pair (R_1, R_2) is achievable, if R_1 satisfies (27a) and (27b), and the following constraints on R_1 and R_2 are satisfied:

$$R_1 \le I(X_1; Y_{\mathrm{D}_i}|U_2) - I(\hat{Y}_2; Y_2|X_1, U_2, Y_{\mathrm{D}_i}) + \check{R}_2, \quad (29a)$$

$$R_2 \le \min\{I(X_2; Y_{\mathrm{D}_3}|U_2), I(X_2; Y_{\mathrm{D}_3}) - \check{R}_2\}. \quad (29b)$$

□

Next, we will particularize the network in Fig. 6 to the Gaussian case.

2) The Gaussian Case: Now we consider the case that each link in Fig. 6 is an additive white Gaussian channel with i.i.d. zero mean unit variance Gaussian noises Z_2 at R, and Z_{D_i} and at D_i, $i = 1, 2, 3$, respectively. Nodes S and R are assumed to use Gaussian codebooks with average transmit power constraints. This case is shown in Fig. 7. For constructing the codebook of node R in Strategies 2 and 4, we use $\alpha_0 \in [0, 1]$ to represent the fraction of power that R allocates to transmit the bin index and $\alpha_1 = 1 - \alpha_0$ to represent the fraction of power that R allocates to transmit its own message index. The SNRs of the S-to-R, S-to-D_i and R-to-D_i links are denoted by $\gamma_{\mathrm{SR}}, \gamma_{\mathrm{SD}_i}$ and γ_{RD_i}, $i = 1, 2, 3$, respectively. The variance of the additional noise in the relay description of its received signal is denoted by γ' [19]. Using these notations and a technique similar to the one in [20], Corollary 2 can be readily used to obtain expressions for achievable rates on this network. These expressions are recorded in the following proposition.

Proposition 2: For the Gaussian network shown in Fig. 7, the rate pair (R_1, R_2) is achievable, where for

- Strategy 1,

$$R_1 + R_2 \le \min_{i=1,2} \mathcal{C}(\gamma_{\mathrm{SD}_i} + \gamma_{\mathrm{RD}_i}) - \mathcal{C}(1/\gamma'), \quad (30a)$$

$$R_1 \le \min_{i=1,2} \mathcal{C}\left(\frac{\gamma_{\mathrm{SR}}}{1+\gamma'} + \gamma_{\mathrm{SD}_i}\right), \quad (30b)$$

$$R_1 + R_2 \le \mathcal{C}(\gamma_{\mathrm{RD}_3}) - \mathcal{C}(1/\gamma'); \quad (30c)$$

- Strategy 2,

$$R_1 \le \min_{i=1,2} \mathcal{C}\Big(\frac{\gamma_{\mathrm{SD}_i} + \alpha_0\gamma_{\mathrm{RD}_i}}{1 + (1-\alpha_0)\gamma_{\mathrm{RD}_i}}\Big) - \mathcal{C}(1/\gamma'), \quad (31a)$$

$$R_1 \le \mathcal{C}\Big(\frac{\gamma_{\mathrm{SR}}}{1+\gamma'} + \frac{\gamma_{\mathrm{SD}_i}}{1 + (1-\alpha_0)\gamma_{\mathrm{RD}_i}}\Big), \quad (31b)$$

$$R_2 \le \min\Big\{\mathcal{C}\big((1-\alpha_0)\gamma_{\mathrm{RD}_3}\big), \mathcal{C}(\gamma_{\mathrm{RD}_3}) - \mathcal{C}\Big(\frac{1+\gamma_{\mathrm{SR}}}{\gamma'}\Big)\Big\}; \quad (31c)$$

- Strategy 3,

$$R_1,\ R_2 \text{ satisfy (30a) and (30b)}, \quad (32a)$$

$$R_1 \le \min_{i=1,2} \mathcal{C}(\gamma_{\mathrm{SD}_i}) - \mathcal{C}(1/\gamma') + \check{R}_2, \quad (32b)$$

$$R_2 \le \mathcal{C}(\gamma_{\mathrm{RD}_3}) - \check{R}_2, \quad (32c)$$

$$\check{R}_2 \le \mathcal{C}\Big(\frac{1+\gamma_{\mathrm{SR}}}{\gamma'}\Big); \quad (32d)$$

- Strategy 4,

$$R_1,\ R_2 \text{ satisfy (31a), (31b) and (32d)}, \quad (33a)$$

$$R_1 \le \min_{i=1,2} \mathcal{C}\Big(\frac{\gamma_{\mathrm{SD}_i}}{1 + (1-\alpha_0)\gamma_{\mathrm{RD}_i}}\Big) - \mathcal{C}(1/\gamma') + \check{R}_2, \quad (33b)$$

$$R_2 \le \min\big\{\mathcal{C}\big((1-\alpha_0)\gamma_{\mathrm{RD}_3}\big), \mathcal{C}\big(\gamma_{\mathrm{RD}_3}\big) - \check{R}_2\big\}, \quad (33c)$$

where, for all strategies, $\alpha_0 \in [0, 1]$.

We now compare the rate expressions of Strategies 1 and 3. From Remark 2, constraint (32b) on R_1 in Strategy 3 coincides with (30b) in Strategy 1 when $\check{R}_2 = \max_{i=1,2} \mathcal{C}\big(\frac{1}{\gamma'} + \frac{\gamma_{\mathrm{SR}}}{\gamma'(1+\gamma_{\mathrm{SD}_i})}\big)$. Without loss of generality, assume $\gamma_{\mathrm{SD}_1} \le \gamma_{\mathrm{SD}_2}$. Then let $\check{R}_2 = \mathcal{C}\left(\frac{1}{\gamma'} + \frac{\gamma_{\mathrm{SR}}}{\gamma'(1+\gamma_{\mathrm{SD}_1})}\right)$. From the decoding at D_1 and D_2, the constraint on R_1 from the decoding at D_1 is tighter than that at D_2 in both strategies according to (30b). It can be seen that the sum rate constraint from (32b) and (32c) in Strategy 3, $\mathcal{C}(\gamma_{\mathrm{SD}_1}) + \mathcal{C}(\gamma_{\mathrm{RD}_3}) - \mathcal{C}(1/\gamma')$, is more relaxed than that in (30c) in Strategy 1. Note that (30a) is a common sum rate constraint in both strategies. Hence the sum rate constraint in Strategy 3 is more relaxed than that in Strategy 1 in general. This implies that in general, the constraint on R_2 in Strategy 3 is more relaxed than that in Strategy 1 for the same constraint on R_1. Therefore, we obtain $\mathcal{R}_1 \subseteq \mathcal{R}_3$.

To compare the rate expressions of Strategies 2 and 4, it can be shown that from Remark 5, constraint (33b) reduces to (31a) when $\check{R}_2 = \max_{i=1,2} \mathcal{C}\left(\frac{1}{\gamma'} + \frac{\gamma_{\mathrm{SR}}(1+(1-\alpha_0)\gamma_{\mathrm{RD}_i})}{\gamma'(1+(1-\alpha_0)\gamma_{\mathrm{RD}_i}+\gamma_{\mathrm{SD}_i})}\right) \le \mathcal{C}\left(\frac{1}{\gamma'} + \frac{\gamma_{\mathrm{SR}}}{\gamma'}\right)$. Using this in (33c) implies that Strategy 4 yields a more relaxed constraint on R_2 in comparison with the constraint in (31c) for Strategy 2. Hence, we have $\mathcal{R}_2 \subseteq \mathcal{R}_4$.

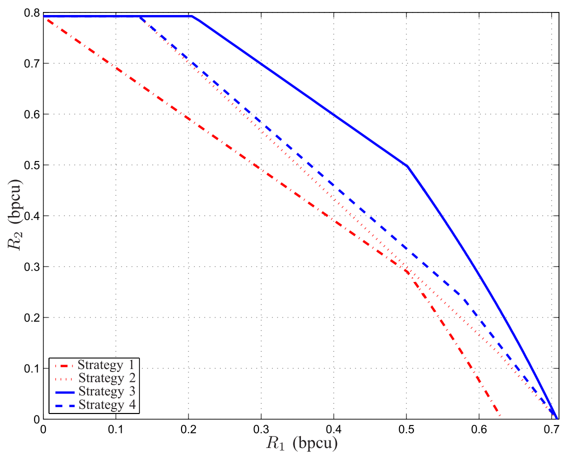

Fig. 8. Achievable rate region for the network in Figure 7. ($\gamma_{\text{SR}} = 2$, $\gamma_{\text{SD}_1} = 1$, $\gamma_{\text{RD}_1} = 4$, $\gamma_{\text{SD}_2} = 2$, $\gamma_{\text{RD}_2} = 1$, $\gamma_{\text{RD}_3} = 1$.)

Fig. 8 provides the achievable rate regions of each strategy for an SNR instance in which $\gamma_{\text{SR}} = 2$, $\gamma_{\text{SD}_1} = 1$, $\gamma_{\text{RD}_1} = 4$, $\gamma_{\text{SD}_2} = 2$, $\gamma_{\text{RD}_2} = 1$, $\gamma_{\text{RD}_3} = 1$. For this instance, it can be seen from Fig. 8 that $\mathcal{R}_1 \subseteq \mathcal{R}_2 \subseteq \mathcal{R}_4 \subseteq \mathcal{R}_3$. Strategy 1 in which D_1 and D_2 treat the undesired signal from R as interference, in comparison with Strategy 2 in which D_1 and D_2 treat the undesired signal from R as noise, yields smaller achievable rate region. Interestingly, under this SNR condition, treating the undesired signal as interference at D_1 and D_2 in Strategy 3 does not induce additional rate loss in comparison with treating it as noise in Strategy 4. Hence it is beneficial to use Strategy 3 which yields a larger achievable rate region.

V. Conclusion

In this paper, we provided a layered forward decoding procedure that enables exploiting the N-to-1 mapping that underlies CF relaying. This procedure relaxes the rate constraint on the bin indices, and is subsequently able to yield a rate advantage over CF-based schemes that use the 1-to-1 mapping. To illustrate the advantage of this procedure, we considered two networks, a two-destination broadcast relay chain network and a partially cooperative multimessage network. In both networks, side information is only available to a subset of the receiving nodes, but not to the rest of the receiving nodes. Our findings are confirmed by numerical evaluation of Gaussian instances of these networks.

Appendix A
Proof of Lemma 2

To prove the lemma, consider

$$\begin{aligned}
&\mathcal{A}_{z(J),i\setminus i+1}\\
&\quad= (\mathcal{A}^c_{m,i+1} \setminus \mathcal{A}^c_{z(J),i} \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}^c_{z(J),i+1}\\
&\quad\overset{\text{(a)}}{=} (\mathcal{A}^c_{m,i+1} \setminus (\mathcal{A}^c_{z(J),i} \cup \mathcal{A}_{z(J^c),i})) \cup \mathcal{A}^c_{z(J),i+1}\\
&\quad= (\mathcal{A}^c_{m,i+1} \setminus (\mathcal{A}_{m,i} \setminus \mathcal{A}_{z(J),i})) \cup \mathcal{A}^c_{z(J),i+1}\\
&\quad\overset{\text{(b)}}{=} (\mathcal{A}^c_{m,i+1} \cap \mathcal{A}_{z(J),i}) \cup (\mathcal{A}^c_{m,i+1} \setminus \mathcal{A}_{m,i}) \cup \mathcal{A}^c_{z(J),i+1}\\
&\quad= (\mathcal{A}^c_{m,i+1} \cap \mathcal{A}_{z(J),i}) \cup \mathcal{A}^c_{z(J),i+1}\\
&\quad\overset{\text{(c)}}{=} (\mathcal{A}^c_{m,i+1} \cup \mathcal{A}^c_{z(J),i+1}) \cap (\mathcal{A}_{z(J),i} \cup \mathcal{A}^c_{z(J),i+1})\\
&\quad\overset{\text{(d)}}{=} (\mathcal{A}^c_{m,i+1} \cup \mathcal{A}^c_{z(J),i+1}) \cap \mathcal{A}_{z(J),i}\\
&\quad\overset{\text{(e)}}{=} ((\mathcal{A}_{m,i} \setminus \mathcal{A}_{m,i+1}) \cup \mathcal{A}^c_{z(J),i+1}) \cap \mathcal{A}_{z(J),i}\\
&\quad= (\mathcal{A}_{m,i} \setminus (\mathcal{A}_{z(J^c),i+1} \cup \mathcal{A}_{z(J),i+1})) \cap \mathcal{A}_{z(J),i}\\
&\quad\overset{\text{(f)}}{=} (\mathcal{A}_{z(J),i} \cap \mathcal{A}_{m,i}) \setminus (\mathcal{A}_{z(J^c),i+1} \cup \mathcal{A}_{z(J),i+1})\\
&\quad= \mathcal{A}_{z(J),i} \setminus (\mathcal{A}_{z(J^c),i+1} \cup \mathcal{A}_{z(J),i+1})\\
&\quad= \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J^c),i+1} \setminus \mathcal{A}_{z(J),i+1}\\
&\quad\overset{\text{(g)}}{=} \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}.
\end{aligned}$$

The proof uses the properties in Lemma 1. In particular, (a) follows from Property 2; (b) follows Property 1 such that $\mathcal{A}^c_{m,i+1} \setminus \mathcal{A}_{m,i} = \emptyset$; (c) follows from Property 4; (d) follows from Property 1; (e) follows from Property 2 and 3; (f) follows from Property 2; (g) follows from Property 5.

Appendix B
Analysis of Probability of Error for Theorem 1

Without loss of generality, assume that for any node $d_k \in \mathcal{N}$, $m_l = 1$, $s_l = 1$ were transmitted and $z_l = 1$ was selected in block $b - \ell, b - \ell + 1, \ldots, b$.

We begin the analysis of the probability of error by providing the following lemma, which applies to any decoding layer i at any given decoding node d.

Lemma 5: Let $\mathcal{U}$ and $\mathcal{V}$ be all the nodes whose codewords $\mathbf{X}_{\mathcal{U}}$ and $\hat{\mathbf{Y}}_{\mathcal{V}}$ are considered at a given layer of the decoding procedure. Let the codewords of $\mathcal{M} \subseteq \mathcal{U}$ and $\mathcal{Z} \subseteq \mathcal{V}$ be $\mathbf{X}_{\mathcal{M}} \subseteq \mathbf{X}_{\mathcal{U}}$ and $\hat{\mathbf{Y}}_{\mathcal{Z}} \subseteq \hat{\mathbf{Y}}_{\mathcal{V}}$, respectively, where for each node in $\mathcal{M}$ and each node in $\mathcal{Z}$ there are multiple codewords that lie in the joint typicality set at the considered layer. For any sets $\mathcal{G} \subseteq \mathcal{M}$ and $\mathcal{F} \subseteq \mathcal{Z}$,

$$\begin{aligned}
&P((\mathbf{X}_{\mathcal{G}}((m,s) \neq (1,1)), \hat{\mathbf{Y}}_{\mathcal{F}}(z \neq 1), \mathbf{X}_{\mathcal{M}\setminus\mathcal{G}}, \hat{\mathbf{Y}}_{\mathcal{Z}\setminus\mathcal{F}},\\
&\qquad \mathbf{x}_{\mathcal{U}\setminus\mathcal{M}}, \hat{\mathbf{y}}_{\mathcal{V}\setminus\mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)}) \le 2^{n(R_{\mathcal{M}}+\check{R}_{\mathcal{M}}+\hat{R}_{\mathcal{Z}}-\mathcal{I}_0)}; \quad (34)
\end{aligned}$$

and

$$\begin{aligned}
&P((\mathbf{x}_{\mathcal{G}}(1,1), \hat{\mathbf{y}}_{\mathcal{F}}(1), \mathbf{X}_{\mathcal{M}\setminus\mathcal{G}}, \hat{\mathbf{Y}}_{\mathcal{Z}\setminus\mathcal{F}}, \mathbf{x}_{\mathcal{U}\setminus\mathcal{M}}, \hat{\mathbf{y}}_{\mathcal{V}\setminus\mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)})\\
&\le 2^{n(R_{\mathcal{M}}+\check{R}_{\mathcal{M}}+\hat{R}_{\mathcal{Z}}-\mathcal{I}_0)}, \quad (35)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{I}_0 &= I(X_{\mathcal{M}}; X_{\mathcal{U}\setminus\mathcal{M}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d)\\
&\quad + \sum_{i\in\mathcal{Z}} I(\hat{Y}_i; X_{\mathcal{U}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i), \quad (36)
\end{aligned}$$

where $\hat{Y}_{\mathcal{K}_i} \triangleq \hat{Y}_{\{d_{i'} \in \mathcal{Z}: i' < i,\}}$, for $i \in \mathcal{Z}$.

Proof: Using joint typicality lemma in [2, Sect.2.5.1], the first statement in the lemma can be readily obtained.

To prove the second the statement in the lemma, consider that the probability $P(\mathbf{X}_{\mathcal{G}}((m,s) \neq (1,1)), \hat{\mathbf{Y}}_{\mathcal{F}}(z \neq 1), \mathbf{x}_{\mathcal{U}\setminus\mathcal{M}}, \hat{\mathbf{y}}_{\mathcal{V}\setminus\mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)})$ can be upper bounded by

$$2^{n(R_{\mathcal{G}}+\check{R}_{\mathcal{G}}+\hat{R}_{\mathcal{F}}-\mathcal{I}')}, \text{ where} \quad (37a)$$

$$\begin{aligned}
\mathcal{I}' &= I(X_{\mathcal{G}}; X_{\mathcal{U}\setminus\mathcal{M}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d)\\
&\quad + \sum_{i\in\mathcal{F}} I(\hat{Y}_i; X_{\mathcal{U}\setminus\mathcal{M}\cup\mathcal{G}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i). \quad (37b)
\end{aligned}$$

Note that if $R_{\mathcal{G}} + \check{R}_{\mathcal{G}} + \hat{R}_{\mathcal{F}} \le \mathcal{I}'$, the probability $P(\mathbf{X}_{\mathcal{G}}((m,s) \neq (1,1)), \hat{\mathbf{Y}}_{\mathcal{F}}(z \neq 1), \mathbf{x}_{\mathcal{U}\setminus\mathcal{M}}, \hat{\mathbf{y}}_{\mathcal{V}\setminus\mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)})$ vanishes as $n \to 0$, which contradicts the definition of $\mathcal{G}$ and $\mathcal{F}$. Hence (37a) and (37b) provide

$$R_{\mathcal{G}} + \check{R}_{\mathcal{G}} + \hat{R}_{\mathcal{F}} > \mathcal{I}'. \quad (38)$$

Now, the probability $P((\mathbf{x}_{\mathcal{G}}((1,1)), \hat{\mathbf{y}}_{\mathcal{F}}(1), \mathbf{X}_{\mathcal{M}\setminus\mathcal{G}}, \hat{\mathbf{Y}}_{\mathcal{Z}\setminus\mathcal{F}}, \mathbf{x}_{\mathcal{U}\setminus\mathcal{M}}, \hat{\mathbf{y}}_{\mathcal{V}\setminus\mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)})$ can be upper bounded by

$$\begin{aligned} &P((\mathbf{x}_{\mathcal{G}}((1,1)),\hat{\mathbf{y}}_{\mathcal{F}}(1),\mathbf{X}_{\mathcal{M}\setminus\mathcal{G}},\hat{\mathbf{Y}}_{\mathcal{Z}\setminus\mathcal{F}},\mathbf{x}_{\mathcal{U}\setminus\mathcal{M}},\hat{\mathbf{y}}_{\mathcal{V}\setminus\mathcal{Z}},\mathbf{Y}_d)\in\mathcal{A}_\epsilon^{(n)}) \\ &\le 2^{n(R_{\mathcal{M}\setminus\mathcal{G}}+\check{R}_{\mathcal{M}\setminus\mathcal{G}}+\hat{R}_{\mathcal{Z}\setminus\mathcal{F}}-\mathcal{I}_1)}, \end{aligned} \tag{39}$$

where

$$\begin{aligned} \mathcal{I}_1 &= I(X_{\mathcal{M}\setminus\mathcal{G}}; X_{\mathcal{U}\setminus\mathcal{M}\cup\mathcal{G}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}\cup\mathcal{F}}, Y_d) \\ &\quad + \sum_{i\in\mathcal{Z}\setminus\mathcal{F}} I(\hat{Y}_i; X_{\mathcal{U}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}\cup\mathcal{F}}, \hat{Y}_{\mathcal{K}_i}, Y_d|X_i) \\ &= I(X_{\mathcal{M}}; X_{\mathcal{U}\setminus\mathcal{M}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d) - I(X_{\mathcal{G}}; \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d|X_{\mathcal{U}\setminus\mathcal{M}}) \\ &\quad + I(X_{\mathcal{M}\setminus\mathcal{G}}; \hat{Y}_{\mathcal{F}}|X_{\mathcal{U}\setminus\mathcal{M}\cup\mathcal{G}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d) \\ &\quad + \sum_{i\in\mathcal{Z}} I(\hat{Y}_i; X_{\mathcal{U}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d|X_i) \\ &\quad - \sum_{i\in\mathcal{F}} I(\hat{Y}_i; X_{\mathcal{U}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d|X_i) \\ &= I(X_{\mathcal{M}}; X_{\mathcal{U}\setminus\mathcal{M}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d) \\ &\quad + \sum_{i\in\mathcal{Z}} I(\hat{Y}_i; X_{\mathcal{U}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d|X_i) \\ &\quad - I(X_{\mathcal{G}}; \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d|X_{\mathcal{U}\setminus\mathcal{M}}) \\ &\quad + I(X_{\mathcal{M}\setminus\mathcal{G}}; \hat{Y}_{\mathcal{F}}|X_{\mathcal{U}\setminus\mathcal{M}\cup\mathcal{G}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d) \\ &\quad - \sum_{i\in\mathcal{F}} I(\hat{Y}_i; X_{\mathcal{U}\setminus\mathcal{M}\cup\mathcal{G}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d|X_i) \\ &\quad - \sum_{i\in\mathcal{F}} I(\hat{Y}_i; X_{\mathcal{M}\setminus\mathcal{G}}|X_{\mathcal{U}\setminus\mathcal{M}\cup\mathcal{G}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d) \\ &= I(X_{\mathcal{M}}; X_{\mathcal{U}\setminus\mathcal{M}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d) \\ &\quad + \sum_{i\in\mathcal{Z}} I(\hat{Y}_i; X_{\mathcal{U}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d|X_i) \\ &\quad - I(X_{\mathcal{G}}; \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, Y_d|X_{\mathcal{U}\setminus\mathcal{M}}) \\ &\quad - \sum_{i\in\mathcal{F}} I(\hat{Y}_i; X_{\mathcal{U}\setminus\mathcal{M}\cup\mathcal{G}}, \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d|X_i) \\ &= I_0 - I' \qquad (40) \\ &\ge I_0 - R_{\mathcal{G}} - \check{R}_{\mathcal{G}} - R_{\mathcal{F}}, \qquad (41) \end{aligned}$$

where (40) follows from (36) and (37b), and (41) follows from (38). Substituting (41) in (39) yields the result of the second statement in the lemma, which completes the proof of the lemma. ■

We note that by definition, the following relationship holds between the sets in Lemma 5 and the sets defined in the decoding procedure:

$$\mathcal{U} = \mathcal{A}_{m,i}, \quad \mathcal{V} = \mathcal{A}_{m,i} \setminus \mathcal{A}_{z(J^c),i+1}, \tag{42a}$$
$$\mathcal{M} = \mathcal{A}_{m,i+1}^c, \quad \mathcal{Z} = \mathcal{A}_{z(J),i+1}^c. \tag{42b}$$

Using Properties 4 and 5 in Lemma 1, set $\mathcal{Z}$ can also be written as

$$\mathcal{Z} = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{z(J^c),i+1}. \tag{43}$$

Furthermore, using Property 2 in Lemma 1, we have

$$\begin{aligned} \mathcal{V} \setminus \mathcal{Z} &= (\mathcal{A}_{m,i} \setminus \mathcal{A}_{z(J^c),i+1}) \setminus \mathcal{A}_{z(J),i+1}^c \\ &= \mathcal{A}_{z(J),i+1}. \end{aligned} \tag{44}$$

Now we analyze the probability of error. Define the following events for layer i in the decoding procedure:

$$\begin{aligned} \mathcal{E}_{i,1} &= \{(\mathbf{x}_{\mathcal{A}_{m,i}}(1,1), \hat{\mathbf{y}}_{z(J),i}(1|1,1), \mathbf{Y}_{d_k}(b-\ell+i)) \notin \mathcal{A}_\epsilon^{(n)}\}; \\ \mathcal{E}_{i,2} &= \{(\mathbf{X}_{\mathcal{A}_{m,i+1}^c}(\hat{m},\hat{s}), \mathbf{x}_{\mathcal{A}_{m,i+1}}(1,1), \hat{\mathbf{Y}}_{\mathcal{A}_{z(J)}^c,i+1}(\hat{z}|1,1), \\ &\qquad \hat{\mathbf{Y}}_{(\mathcal{A}_{m,i+1}^c\setminus\mathcal{A}_{z(J),i}^c)\setminus\mathcal{A}_{z(J^c),i}}(\hat{z}|\hat{m},\hat{s}), \hat{\mathbf{y}}_{\mathcal{A}_{z(J),i+1}}(1|1,1), \\ &\qquad \mathbf{Y}_{d_k}(b-\ell+i)) \in \mathcal{A}_\epsilon^{(n)}, \text{for some } \hat{m},\hat{z},\hat{s} \in \mathcal{B}(\hat{z})\}. \end{aligned}$$

The receiver at d_k makes an error if any event in $\mathcal{E} \triangleq (\cup_i \mathcal{E}_{i,1}) \bigcup (\cap_i \mathcal{E}_{i,2})$ occurs for some $\hat{m} \neq 1, \hat{s} \neq 1$. Using the union bound, the probability of error is given by $P(\mathcal{E}) = P((\cup_i \mathcal{E}_{i,1}) \bigcup (\cap_i \mathcal{E}_{i,2})) = P(\cup_i \mathcal{E}_{i,1}) + P(\cap_i \mathcal{E}_{i,2}) \le \sum_i P(\mathcal{E}_{i,1}) + P(\cap_i \mathcal{E}_{i,2})$. By the conditional typicality lemma in [2], $P(\mathcal{E}_{i,1}) \to 0$ as $n \to \infty$. Now we upper bound $P(\cap_i \mathcal{E}_{i,2})$. Consider the probability $P(\mathcal{E}_{i,2})$ for the case that $(\hat{m},\hat{s}) \neq (1,1)$ for $\mathbf{x}_l(\hat{m},\hat{s}), d_l \in \mathcal{A}_{m,i+1}^c$, which can be bounded by

$$P(\mathcal{E}_{i,2}) \le \sum_{m_{\mathcal{A}_{m,i}}, s_{\mathcal{A}_{m,i}}, z_{\mathcal{A}_{z,i}}} 2^{-n\beta_i}, \text{ where} \tag{45a}$$

$$\beta_i = \sum_{j=1}^{3} \beta_{i,j}, \text{ and} \tag{45b}$$

$$\begin{aligned} \beta_{i,1} &= I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}|X_{\mathcal{A}_{m,i+1}}), \\ \beta_{i,2} &= \sum_{i\in(\mathcal{A}_{m,i+1}^c\setminus\mathcal{A}_{z(J),i}^c)\setminus\mathcal{A}_{z(J^c),i}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}|X_i), \\ \beta_{i,3} &= \sum_{i\in\mathcal{A}_{z(J),i+1}^c} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{((\mathcal{A}_{m,i+1}^c\setminus\mathcal{A}_{z(J),i}^c)\setminus\mathcal{A}_{z(J^c),i})\cup \mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}|X_i). \end{aligned}$$

Note that using chain rule and Lemma 2, we can rewrite β_i in (45b) as

$$\begin{aligned} \beta_i &= I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}|X_{\mathcal{A}_{m,i+1}}) \\ &\quad + \sum_{i\in\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}|X_i), \end{aligned}$$

which can be shown, by substituting the sets using (42), (43) and (44), to have same the form as (36) in Lemma 5.

For node $d_l \in \mathcal{G} \subseteq \mathcal{M} \triangleq \mathcal{A}_{m,i+1}^c$ and $d_k \in \mathcal{F} \subseteq \mathcal{Z} \triangleq \mathcal{A}_{z(J),i+1}^c$, multiple $\mathbf{x}_l(\hat{m},\hat{s})$ and $\hat{\mathbf{y}}_k(\hat{z}|\hat{m},\hat{s})$ are found in the joint typicality set, respectively, at layer i. For the case $(\hat{m},\hat{s}) \neq (1,1)$ and $\hat{z} \neq 1$, the probability of the event is given by $P(\mathcal{E}_{i,2})$, for which an upper bound is provided in (45). For the case $(\hat{m},\hat{s}) = (1,1)$ and the case $\hat{z} = 1$, substituting sets in Lemma 5 using (42) provides that the probability of these cases can also be upper bounded by (45).

Hence, we can upper bound the probability $P(\cap_i \mathcal{E}_{i,2})$ as

$$\begin{aligned}
&P(\bigcap_{i=1}^{\ell} \mathcal{E}_{i,2}) \\
&\leq \prod_{i=1}^{\ell} \sum_{m_{\mathcal{A}_{m,i}}, s_{\mathcal{A}_{m,i}}, z_{\mathcal{A}_{z,i}}} P(\mathcal{E}_{i,2}) \\
&\leq \prod_{i=1}^{\ell} \sum_{m_{\mathcal{A}^c_{m,i+1}}, s_{\mathcal{A}^c_{m,i+1}}, z_{\mathcal{A}_{z(J),i\setminus i+1}}} 2^{-n\beta_i} \\
&\leq \prod_{i=1}^{\ell-1} \prod_{l \in m_{\mathcal{A}^c_{m,i+1}}} 2^{nR_l} \prod_{l \in \mathcal{A}_{z(J),i\setminus i+1}} 2^{n\hat{R}_l} \cdot 2^{-n\sum_{j=1}^{3}\beta_{i,j}} \\
&\quad \prod_{l \in m_{\mathcal{A}^c_{\ell_m}}} 2^{nR_l} \prod_{l \in \mathcal{A}_{\ell_m}\setminus\mathcal{A}_{\ell_z}} 2^{-n\check{R}_l} \prod_{l \in \mathcal{A}_{z(J),\ell-1\setminus\ell}} 2^{n\hat{R}_l} \cdot 2^{-n\sum_{j=1}^{3}\beta_{i,j}} \\
&= 2^{nR_{\mathcal{A}_{m,1}\setminus\mathcal{A}_{\ell_m}}} \cdot 2^{-n\check{R}_{\mathcal{A}_{\ell_m}\setminus\mathcal{A}_{\ell_z}}} \\
&\quad \cdot 2^{n\sum_{i=1}^{\ell}(\hat{R}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} - \sum_{j=1}^{3}\beta_{i,j})}. \qquad (46)
\end{aligned}$$

Using chain rule and the definition of $\mathcal{A}_{\ell_z}$, we can expand $\beta_{i,1}$ as

$$\begin{aligned}
\beta_{i,1} &= I(X_{\mathcal{A}^c_{m,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),\ell+1}}, Y_{d_k}|X_{\mathcal{A}_{m,i+1}}) \\
&\quad + I(X_{\mathcal{A}^c_{m,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),i+1}\setminus\mathcal{A}_{z(J),\ell+1}}|X_{\mathcal{A}_{m,i+1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell+1}}) \\
&= I(X_{\mathcal{A}^c_{m,i+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{m,i+1}}) \\
&\quad + I(X_{\mathcal{A}^c_{m,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),i+1}\setminus\mathcal{A}_{\ell_z}}|X_{\mathcal{A}_{m,i+1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}). \qquad (47)
\end{aligned}$$

We define

$$\begin{aligned}
\beta_{i,1,1} &= I(X_{\mathcal{A}^c_{m,i+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{m,i+1}}), \\
\beta_{i,1,2} &= I(X_{\mathcal{A}^c_{m,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),i+1}\setminus\mathcal{A}_{\ell_z}}|X_{\mathcal{A}_{m,i+1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}).
\end{aligned}$$

Now, consider the summation of $\beta_{i,1,1}$ from two consecutive decoding layers ℓ and $\ell-1$:

$$\begin{aligned}
&(X_{\mathcal{A}_{m,\ell}\setminus\mathcal{A}_{m,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{m,\ell+1}}) \\
&\quad + I(X_{\mathcal{A}^c_{m,\ell}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{m,\ell}\setminus\mathcal{A}_{m,\ell+1}}, X_{\mathcal{A}_{m,\ell+1}}) \\
&= I(X_{\mathcal{A}_{m,\ell}\setminus\mathcal{A}_{m,\ell+1}}, X_{\mathcal{A}^c_{m,\ell}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{m,\ell+1}}) \\
&= I(X_{\mathcal{A}_{m,\ell-1}\setminus\mathcal{A}_{m,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{m,\ell+1}})
\end{aligned}$$

where the first equality follows the chain rule and in the last equality we have used

$$(\mathcal{A}_{m,\ell}\setminus\mathcal{A}_{m,\ell+1}) \cup \mathcal{A}^c_{m,\ell} = \mathcal{A}_{m,\ell-1}\setminus\mathcal{A}_{m,\ell+1}$$

by Property 1 and 3 in Lemma 1.

Using the same technique iteratively, it can be shown that $\sum_{i=\tilde{i}}^{\ell}\beta_{i,1,1} = I(X_{\mathcal{A}_{m,\tilde{i}}\setminus\mathcal{A}_{m,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{m,\ell+1}})$. Hence, $\sum_{i=1}^{\ell}\beta_{i,1,1}$ yields

$$\begin{aligned}
\sum_{i=1}^{\ell}\beta_{i,1,1} &= I(X_{\mathcal{A}_{m,1}\setminus\mathcal{A}_{m,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{m,\ell+1}}) \\
&= I(X_{\mathcal{A}_{m,1}\setminus\mathcal{A}_{\ell_m}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{\ell_m}}). \qquad (48)
\end{aligned}$$

Next, consider the summation of $\beta_{i,2}$ and $\beta_{i,3}$ in (46) for all possible $\hat{z}$. We define $\beta_{i,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}}$, where

$$\begin{aligned}
&\beta_{i,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} \\
&= \sum_{i\in(\mathcal{A}^c_{m,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z,i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}|X_i) - \hat{R}_i \\
&\quad + \sum_{i\in\mathcal{A}^c_{z,i+1}} (-\hat{R}_i + I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}, \\
&\qquad \hat{Y}_{((\mathcal{A}^c_{m,i+1}\setminus\mathcal{A}^c_{z,i})\setminus\mathcal{A}_{z(J^c),i})\cup\mathcal{A}_{z(J),i+1}}|X_i)) \\
&\leq \sum_{i\in(\mathcal{A}^c_{m,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}} -I(\hat{Y}_i; Y_i|X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}) \\
&\quad - \sum_{i\in\mathcal{A}^c_{z,i+1}} I(\hat{Y}_i; Y_i|X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}, \\
&\qquad \hat{Y}_{((\mathcal{A}^c_{m,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i})\cup\mathcal{A}_{z(J),i+1}}) \\
&= -I(\hat{Y}_{(\mathcal{A}^c_{m,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}}; Y_{(\mathcal{A}^c_{m,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}} \\
&\qquad |X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}) \\
&\quad - I(\hat{Y}_{\mathcal{A}^c_{z(J),i+1}}; Y_{\mathcal{A}^c_{z(J),i+1}} \\
&\qquad |X_{\mathcal{A}_{m,i}}, \hat{Y}_{((\mathcal{A}^c_{m,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i})\cup\mathcal{A}_{z(J),i+1}}, Y_{d_k}) \\
&= -I(\hat{Y}_{(\mathcal{A}^c_{m,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}}, \hat{Y}_{\mathcal{A}^c_{z(J),i+1}}; \\
&\qquad Y_{(\mathcal{A}^c_{m,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}}, Y_{\mathcal{A}^c_{z(J),i+1}} \\
&\qquad |X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}) \\
&= -I(\hat{Y}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}}; Y_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} \\
&\qquad |X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}),
\end{aligned}$$

where in the first inequality we have used (6), the second and third equalities follow the chain rule, and in the last equality we have used (10) from Lemma 2.

By definition, $\mathcal{A}_{z(J),\ell+1} = \mathcal{A}_{\ell_z}$, hence $\mathcal{A}_{z(J),\ell+1}\setminus\mathcal{A}_{\ell_z} = \emptyset$ and $\beta_{\ell,1,2} = 0$. Now, for layer ℓ and $\ell-1$, we calculate

$$\begin{aligned}
&\beta_{\ell,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{z(J),\ell+1}} + \beta_{\ell-1,1,2} \\
&\quad + \beta_{\ell-1,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}} \\
&\leq -I(\hat{Y}_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}}|X_{\mathcal{A}_{m,\ell}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
&\quad + I(X_{\mathcal{A}^c_{m,\ell}}; \hat{Y}_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}}|X_{\mathcal{A}_{m,\ell}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
&\quad - I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}}; Y_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}} \\
&\qquad X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell}}, Y_{d_k}) \\
&= -I(\hat{Y}_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}}|X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
&\quad - I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}}; Y_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}} \\
&\qquad |X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell}}, Y_{d_k}) \\
&= -I(\hat{Y}_{(\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell})\cup(\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z})}; \\
&\qquad Y_{(\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell})\cup(\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z})} \\
&\qquad |X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
&= -I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{\ell_z}}|X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}), \qquad (49)
\end{aligned}$$

where the first and second equalities follow the chain rule, and in the last equality we have used

$$(\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}) \cup (\mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{\ell_z}) = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{\ell_z} \quad (50)$$

for $\forall i \leq \ell$, which is a direct consequence from Property 6 in Lemma 1.

Using this technique iteratively, we obtain $\sum_{i=\tilde{i}}^{\ell} \beta_{i,1,2} + \beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} \leq -I(\hat{Y}_{\mathcal{A}_{z(J),\tilde{i}} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\tilde{i}} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\tilde{i}}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k})$. Consider the summation over all ℓ layers, we have

$$\sum_{i=1}^{\ell} \beta_{i,1,2} + \beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} \leq -I(\hat{Y}_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}). \quad (51)$$

Using (48) and (51), the probability in (46) can be upper bounded by

$$\begin{aligned} &P(\cap_i \mathcal{E}_{i,2}) \\ &\leq 2^{n(R_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}} - \check{R}_{\mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}})} \cdot 2^{-n(\sum_{i=1}^{\ell} \beta_{i,1,1} + \sum_{i=1}^{\ell} (\beta_{i,1,2} + \beta_{i,2U3}))} \\ &= 2^{n(R_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}} - \mathcal{I}(\mathcal{A}))}, \end{aligned}$$

where $\mathcal{I}(\mathcal{A})$ is given by

$$\begin{aligned} &I(X_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}) \\ &\quad - I(\hat{Y}_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) + \check{R}_{\mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}}. \end{aligned}$$

Using $\mathcal{A}_{m,1} = \mathcal{A}_{z(J),1}$ by definition, $\mathcal{I}(\mathcal{A})$ can be written as

$$\begin{aligned} &I(X_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}) \\ &\quad - I(\hat{Y}_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) + \check{R}_{\mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}}. \end{aligned}$$

Let $\mathcal{S} \triangleq \mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}, \mathcal{S}^c \triangleq \mathcal{A}_{\ell_m}$, hence $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$. Let $\check{\mathcal{S}} \triangleq \mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}, \check{\mathcal{S}}^c \triangleq \mathcal{A}_{\ell_z}$, hence $\check{\mathcal{S}}^c = \mathcal{A}_{\ell_m} \setminus \check{\mathcal{S}}$. The above result yields that $P(\mathcal{E}) \to 0$ as $n \to \infty$, if

$$\begin{aligned} R_{\mathcal{S}} \leq\ &I(X_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{S}^c}) \\ &- I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}}. \end{aligned}$$

Using the technique similar to the one in [13], it can be shown that for $\mathcal{S} \cap \mathcal{S}_{d_k} = \emptyset$, the constraints in the above inequality can be dropped and $d_k \in \mathcal{D}_{\mathcal{S}}$. By Property 7 in Lemma 1, $d_k \in \mathcal{A}_{\ell_m} = \mathcal{S}^c$. Hence $d_k \in \mathcal{S}^c \cap \mathcal{D}_{\mathcal{S}}$. This completes the proof of the theorem.

Appendix C
Proof of Remark 2

To prove the lemma, first we note that when

$$\check{R}_{\check{\mathcal{S}}} \geq I(X_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\check{\mathcal{S}}^c}) - R_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}, \quad (52)$$

the probability of error tends to 0 as $n \to \infty$ if the rate expression in (4) satisfies

$$\begin{aligned} R_{\mathcal{N} \setminus \check{\mathcal{S}}^c} &= R_{\mathcal{S} \cup (\mathcal{S}^c \setminus \check{\mathcal{S}}^c)} \\ &\leq I(X_{\mathcal{S} \cup (\mathcal{S}^c \setminus \check{\mathcal{S}}^c)}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\check{\mathcal{S}}^c}) \\ &\quad - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &= I(X_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\check{\mathcal{S}}^c}) \\ &\quad - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \end{aligned}$$

where we have used

$$\mathcal{S} \cup (\mathcal{S}^c \setminus \check{\mathcal{S}}^c) = \mathcal{N} \setminus \check{\mathcal{S}}^c$$

since $\check{\mathcal{S}}^c \subseteq \mathcal{S}^c$ and $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$. Redefine $\mathcal{S} \triangleq \mathcal{N} \setminus \check{\mathcal{S}}^c$ and $\mathcal{S}^c \triangleq \check{\mathcal{S}}^c$, we have

$$R_{\mathcal{S}} \leq I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}),$$

which is the simplified form of the rate expression (11).

Next, in general the rate expression in (4) can be modified to the following form:

$$\begin{aligned} R_{\mathcal{S}} \leq\ &I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) \\ &- I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &- I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ =\ &I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) \\ &+ I(Y_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &- I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ &- I(\hat{Y}_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ =\ &I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) - (\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}) \\ &+ \check{R}_{\check{\mathcal{S}}} - I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}} | X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}). \end{aligned} \quad (53)$$

where the first two steps follow the chain rule; in the last equality we have also used chain rule and $\mathcal{S}^c \setminus \check{\mathcal{S}}^c = \check{\mathcal{S}}$ and $(\mathcal{N} \setminus \check{\mathcal{S}}^c) \setminus (\mathcal{S}^c \setminus \check{\mathcal{S}}^c) = \mathcal{S}$, all by definition.

It can be seen that if

$$\check{R}_{\check{\mathcal{S}}} \geq I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}} | X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \quad (54)$$

the rate expression in (53) reduces to its simplified form (11).

Appendix D
Analysis of Probability of Error for Theorem 2

Without loss of generality, assume that for any node $d_k \in \mathcal{N}$, $m_l = 1$, $s_l = 1$ were transmitted and $z_l = 1$ was selected in block $b - \ell, b - \ell + 1, \ldots, b$.

We begin the analysis of the probability of error by providing the following lemma, which is a counterpart of Lemma 5 and applies to any decoding layer at any decoding node d.

Lemma 6: Let $\mathcal{U}, \mathcal{W}$ and $\mathcal{V}$ be all the nodes whose codewords $\mathbf{X}_{\mathcal{U}}, \mathbf{U}_{\mathcal{W}}$ and $\hat{\mathbf{Y}}_{\mathcal{V}}$ are considered at a given layer of the decoding procedure. Let the codewords of $\mathcal{M} \subseteq \mathcal{U}, \mathcal{J} \subseteq \mathcal{W}$ and $\mathcal{Z} \subseteq \mathcal{V}$ be $\mathbf{X}_{\mathcal{M}} \subseteq \mathbf{X}_{\mathcal{U}}, \mathbf{U}_{\mathcal{J}} \subseteq \mathbf{U}_{\mathcal{W}}$ and $\hat{\mathbf{Y}}_{\mathcal{Z}} \subseteq \hat{\mathbf{Y}}_{\mathcal{V}}$, respectively, where for each node in $\mathcal{M}$, each node in $\mathcal{W}$ and each node in $\mathcal{Z}$ there are multiple codewords that lie in the joint typicality set at the considered layer. For any sets $\mathcal{G} \subseteq \mathcal{M}, \mathcal{H} \subseteq \mathcal{J}$ and $\mathcal{F} \subseteq \mathcal{Z}$,

$$\begin{aligned} &P((\mathbf{X}_{\mathcal{G}}(m \neq 1), \mathbf{U}_{\mathcal{H}}(s \neq 1), \hat{\mathbf{Y}}_{\mathcal{F}}(z \neq 1), \mathbf{X}_{\mathcal{M} \setminus \mathcal{G}}, \mathbf{U}_{\mathcal{J} \setminus \mathcal{H}}, \hat{\mathbf{Y}}_{\mathcal{Z} \setminus \mathcal{F}}, \\ &\quad \mathbf{x}_{\mathcal{U} \setminus \mathcal{M}}, \mathbf{u}_{\mathcal{W} \setminus \mathcal{J}}, \hat{\mathbf{y}}_{\mathcal{V} \setminus \mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)}) \leq 2^{n(R_{\mathcal{M}} + \check{R}_{\mathcal{J}} + \hat{R}_{\mathcal{Z}} - \mathcal{I}_0)}; \end{aligned} \quad (55)$$

and

$$\begin{aligned} &P((\mathbf{x}_{\mathcal{G}}(1), \mathbf{u}_{\mathcal{H}}(1), \hat{\mathbf{y}}_{\mathcal{F}}(1), \mathbf{X}_{\mathcal{M} \setminus \mathcal{G}}, \mathbf{U}_{\mathcal{J} \setminus \mathcal{H}}, \hat{\mathbf{Y}}_{\mathcal{Z} \setminus \mathcal{F}}, \\ &\quad \mathbf{x}_{\mathcal{U} \setminus \mathcal{M}}, \mathbf{u}_{\mathcal{W} \setminus \mathcal{J}}, \hat{\mathbf{y}}_{\mathcal{V} \setminus \mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)}) \leq 2^{n(R_{\mathcal{M}} + \check{R}_{\mathcal{J}} + \hat{R}_{\mathcal{Z}} - \mathcal{I}_0)}, \end{aligned} \quad (56)$$

where

$$\mathcal{I}_0 = I(X_{\mathcal{M}}, U_{\mathcal{J}}; \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \quad Y_d | X_{\mathcal{U}\setminus\mathcal{M}}, U_{\mathcal{W}\setminus\mathcal{J}}) + \sum_{i\in\mathcal{Z}} I(\hat{Y}_i; X_{\mathcal{U}}, U_{\mathcal{W}}, \quad \hat{Y}_{\mathcal{V}\setminus\mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | U_i), \quad (57)$$

Proof: Using the following substitution in Lemma 5,

$$(X_{\mathcal{M}}, U_{\mathcal{J}}) \triangleq X_{\mathcal{M}} \text{ and } (X_{\mathcal{U}}, U_{\mathcal{W}}) \triangleq X_{\mathcal{U}},$$

the result in Lemma 6 can be readily obtained. ∎

We note that by definition, the following relationship holds between the sets in Lemma 6 and the sets defined in the decoding procedure for Theorem 2:

$$\mathcal{U} = \mathcal{A}_{m,i}, \quad \mathcal{W} = \mathcal{A}_{s,i}, \quad \mathcal{V} = \mathcal{A}_{s,i} \setminus \mathcal{A}_{z(J^c),i+1}, \quad (58a)$$

$$\mathcal{M} = \mathcal{A}^c_{m,i+1}, \quad \mathcal{J} = \mathcal{A}^c_{s,i+1}, \quad \mathcal{Z} = \mathcal{A}^c_{z(J),i+1}. \quad (58b)$$

Using Properties 4 and 5 in Lemma 1, set $\mathcal{Z}$ can also be written as

$$\mathcal{Z} = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{z(J^c),i+1}. \quad (59)$$

Furthermore, using Property 2 in Lemma 1, we have

$$\mathcal{V} \setminus \mathcal{Z} = (\mathcal{A}_{s,i} \setminus \mathcal{A}_{z(J^c),i+1}) \setminus \mathcal{A}^c_{z(J),i+1} = \mathcal{A}_{z(J),i+1}. \quad (60)$$

Now we analyze the probability of error. Define the following events for layer i in the decoding procedure:

$$\mathcal{E}_{i,1} = \{(\mathbf{x}_{\mathcal{A}_{m,i}}(1|1), \mathbf{u}_{\mathcal{A}_{s,i}}(1), \hat{\mathbf{y}}_{z(J),i}(1|1), \mathbf{Y}_{d_k}(b-\ell+i)) \notin \mathcal{A}^{(n)}_\epsilon\};$$

$$\mathcal{E}_{i,2} = \{(\mathbf{U}_{\mathcal{A}^c_{s,i+1}}(\hat{s}), \mathbf{u}_{\mathcal{A}_{s,i+1}}(1), \mathbf{X}_{\mathcal{A}^c_{m,i+1}}(\hat{m}|\hat{s}), \mathbf{X}_{\mathcal{A}^c_{m,i+1}}(\hat{m}|1), \mathbf{x}_{\mathcal{A}_{m,i+1}}(1|1), \hat{\mathbf{Y}}_{\mathcal{A}^c_{z(J)},i+1}(\hat{z}|1), \hat{\mathbf{Y}}_{(\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}}(\hat{z}|\hat{s}), \hat{\mathbf{y}}_{\mathcal{A}_{z(J),i+1}}(1|1), \mathbf{Y}_{d_k}(b-\ell+i)) \in \mathcal{A}^{(n)}_\epsilon, \text{ for some } \hat{m}, \hat{z} \text{ and } \hat{s} \in \mathcal{B}(\hat{z})\}.$$

The receiver at d_k makes an error if any event in $\mathcal{E} \triangleq (\cup_i \mathcal{E}_{i,1}) \bigcup (\cap_i \mathcal{E}_{i,2})$ for some $\hat{m} \neq 1, \hat{s} \neq 1$ occurs. Using the union bound, the probability of error is given by $P(\mathcal{E}) = P((\cup_i \mathcal{E}_{i,1}) \bigcup (\cap_i \mathcal{E}_{i,2})) = P(\cup_i \mathcal{E}_{i,1}) + P(\cap_i \mathcal{E}_{i,2}) \leq \sum_i P(\mathcal{E}_{i,1}) + P(\cap_i \mathcal{E}_{i,2})$. By the conditional typicality lemma in [2], $P(\mathcal{E}_{i,1}) \to 0$ as $n \to \infty$. Now we upper bound $P(\cap_i \mathcal{E}_{i,2})$. Consider the probability $P(\mathcal{E}_{i,2})$ for the case that $\hat{m} \neq 1$ for $\mathbf{x}_l(\hat{m}|\hat{s}), d_l \in \mathcal{A}^c_{m,i+1}$ and $\hat{s} \neq 1$ for $\mathbf{u}_l(\hat{s}), d_l \in \mathcal{A}^c_{s,i+1}$, which can be bounded by

$$P(\mathcal{E}_{i,2}) \leq \sum_{m_{\mathcal{A}_{m,i}}, s_{\mathcal{A}_{s,i}}, z_{\mathcal{A}_{z,i}}} 2^{-n\beta_i}, \text{ where} \quad (61a)$$

$$\beta_i = \sum_{j=1}^{3} \beta_{i,j}, \text{ and} \quad (61b)$$

$$\beta_{i,1} = I(X_{\mathcal{A}^c_{m,i+1}}, U_{\mathcal{A}^c_{s,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}),$$

$$\beta_{i,2} = \sum_{i\in(\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | U_i),$$

$$\beta_{i,3} = \sum_{i\in\mathcal{A}^c_{z(J),i+1}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{((\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i})\cup\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | U_i).$$

Note that using chain rule and Lemma 2, we can rewrite β_i in (61b) as

$$\beta_i = I(X_{\mathcal{A}^c_{m,i+1}}, U_{\mathcal{A}^c_{s,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}) + \sum_{i\in\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | X_i),$$

which can be shown, by substituting the sets using (58), (59) and (60), to have same the form as (57) in Lemma 6.

For node $d_l \in \mathcal{G} \subseteq \mathcal{M} \triangleq \mathcal{A}^c_{m,i+1}, d_h \in \mathcal{H} \subseteq \mathcal{J} \triangleq \mathcal{A}^c_{s,i+1}$ and $d_k \in \mathcal{F} \subseteq \mathcal{Z} \triangleq \mathcal{A}^c_{z(J),i+1}$, multiple $\mathbf{x}_l(\hat{m}|\hat{s}), \mathbf{u}(\hat{s})$ and $\hat{\mathbf{y}}_k(\hat{z}|\hat{s})$ are found in the joint typicality set, respectively, at layer i. For the case $\hat{m} \neq 1, \hat{s} \neq 1$ and $\hat{z} \neq 1$, the probability of the event is given by $P(\mathcal{E}_{i,2})$, for which an upper bound is provided in (61). For the case $\hat{m} = 1$, the case $\hat{s} = 1$ and the case $\hat{z} = 1$, substituting sets in Lemma 6 using (58) provides that the probability of these cases can also be upper bounded by (45).

Now consider the probability of event $\mathcal{E}_{i,2}$ over the ℓ layers:

$$\begin{aligned} P(\bigcap_{i=1}^{\ell} \mathcal{E}_{i,2}) &\leq \prod_{i=1}^{\ell} \sum_{m_{\mathcal{A}_{m,i}}, s_{\mathcal{A}_{s,i}}, z_{\mathcal{A}_{z,i}}} P(\mathcal{E}_{i,2}) \\ &\leq \prod_{i=1}^{\ell} \sum_{m_{\mathcal{A}^c_{m,i+1}}, s_{\mathcal{A}^c_{s,i+1}}, z_{\mathcal{A}_{z(J),i\setminus i+1}}} 2^{-n\beta_i} \\ &\leq \prod_{i=1}^{\ell-1} \prod_{l\in m_{\mathcal{A}^c_{m,i+1}}} 2^{nR_l} \prod_{l\in\mathcal{A}_{z(J),i\setminus i+1}} 2^{n\hat{R}_l} \cdot 2^{-n\sum_{j=1}^{3}\beta_{i,j}} \\ &\quad \prod_{l\in m_{\mathcal{A}^c_{\ell_m}}} 2^{nR_l} \prod_{l\in\mathcal{A}_{\ell_s}\setminus\mathcal{A}_{\ell_z}} 2^{-n\check{R}_l} \prod_{l\in\mathcal{A}_{z(J),\ell-1\setminus\ell}} 2^{n\hat{R}_l} \cdot 2^{-n\sum_{j=1}^{3}\beta_{i,j}} \\ &= 2^{nR_{\mathcal{A}_{m,1}\setminus\mathcal{A}_{\ell_m}}} \cdot 2^{-n\check{R}_{\mathcal{A}_{\ell_s}\setminus\mathcal{A}_{\ell_z}}} \cdot 2^{n\sum_{i=1}^{\ell}(\hat{R}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} - \sum_{j=1}^{3}\beta_{i,j})}. \end{aligned} \quad (62)$$

Using chain rule and the definition of $\mathcal{A}_{\ell_z}$, we can expand $\beta_{i,1}$ as

$$\begin{aligned} \beta_{i,1} &= I(X_{\mathcal{A}^c_{m,i+1}}, U_{\mathcal{A}^c_{s,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),\ell+1}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}) \\ &\quad + I(X_{\mathcal{A}^c_{m,i+1}}, U_{\mathcal{A}^c_{s,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),i+1}\setminus\mathcal{A}_{z(J),\ell+1}} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell+1}}) \\ &= I(X_{\mathcal{A}^c_{m,i+1}}, U_{\mathcal{A}^c_{s,i+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}) \\ &\quad + I(X_{\mathcal{A}^c_{m,i+1}}, U_{\mathcal{A}^c_{s,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),i+1}\setminus\mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}). \end{aligned} \quad (63)$$

We define

$$\beta_{i,1,1} = I(X_{\mathcal{A}^c_{m,i+1}}, U_{\mathcal{A}^c_{s,i+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}})$$

$$\beta_{i,1,2} = I(X_{\mathcal{A}^c_{m,i+1}}, U_{\mathcal{A}^c_{s,i+1}}; \hat{Y}_{\mathcal{A}_{z(J),i+1}\setminus\mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}).$$

Now, consider the summation of $\beta_{i,1,1}$ from two consecutive decoding layers ℓ and $\ell-1$:

$$\begin{aligned}
&(X_{\mathcal{A}_{m,\ell}\setminus\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell}\setminus\mathcal{A}_{s,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell+1}}) \\
&\quad + I(X_{\mathcal{A}^c_{m,\ell}}, U_{\mathcal{A}^c_{s,\ell}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell}\setminus\mathcal{A}_{m,\ell+1}}, X_{\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell}\setminus\mathcal{A}_{s,\ell+1}}, U_{\mathcal{A}_{s,\ell+1}}) \\
&= I(X_{\mathcal{A}_{m,\ell}\setminus\mathcal{A}_{m,\ell+1}}, X_{\mathcal{A}^c_{m,\ell}}, U_{\mathcal{A}_{s,\ell}\setminus\mathcal{A}_{s,\ell+1}}, U_{\mathcal{A}^c_{s,\ell}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell+1}}) \\
&= I(X_{\mathcal{A}_{m,\ell-1}\setminus\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell-1}\setminus\mathcal{A}_{s,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell+1}})
\end{aligned}$$

where the first equality follows the chain rule and in the last equality we have used

$$(\mathcal{A}_{m,\ell}\setminus\mathcal{A}_{m,\ell+1}) \cup \mathcal{A}^c_{m,\ell} = \mathcal{A}_{m,\ell-1}\setminus\mathcal{A}_{m,\ell+1},$$
$$(\mathcal{A}_{s,\ell}\setminus\mathcal{A}_{s,\ell+1}) \cup \mathcal{A}^c_{s,\ell} = \mathcal{A}_{s,\ell-1}\setminus\mathcal{A}_{s,\ell+1}$$

by Property 1 and 3 in Lemma 3.

Using this technique iteratively, we obtain $\sum_{i=\tilde{i}}^{\ell} \beta_{i,1,1} = I(X_{\mathcal{A}_{m,\tilde{i}}\setminus\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{s,\tilde{i}}\setminus\mathcal{A}_{\ell_s}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{\ell_s}})$. Consider the summation over all ℓ layers, we have

$$\sum_{i=1}^{\ell} \beta_{i,1,1} = I(X_{\mathcal{A}_{m,1}\setminus\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{s,1}\setminus\mathcal{A}_{\ell_s}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{\ell_s}}). \tag{64}$$

Next, consider the summation of $\beta_{i,2}$ and $\beta_{i,3}$ in (62) for all possible $\hat{z}$. We define $\beta_{i,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}}$, where

$$\begin{aligned}
&\beta_{i,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} \\
&= \sum_{i\in(\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z,i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | U_i) - \hat{R}_i \\
&\quad + \sum_{i\in\mathcal{A}^c_{z,i+1}} -\hat{R}_i + I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}, \hat{Y}_{((\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z,i})\setminus\mathcal{A}_{z(J^c),i})\cup\mathcal{A}_{z(J),i+1}} | U_i) \\
&\le \sum_{i\in(\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}} -I(\hat{Y}_i; Y_i | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}) \\
&\quad - \sum_{i\in\mathcal{A}^c_{z,i+1}} I(\hat{Y}_i; Y_i | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}, \hat{Y}_{((\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i})\cup\mathcal{A}_{z(J),i+1}}) \\
&= -I(\hat{Y}_{(\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}}; Y_{(\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}} | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}) \\
&\quad - I(\hat{Y}_{\mathcal{A}^c_{z(J),i+1}}; Y_{\mathcal{A}^c_{z(J),i+1}} | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{((\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i})\cup\mathcal{A}_{z(J),i+1}}, Y_{d_k}) \\
&= -I(\hat{Y}_{(\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}}, \hat{Y}_{\mathcal{A}^c_{z(J),i+1}}; Y_{(\mathcal{A}^c_{s,i+1}\setminus\mathcal{A}^c_{z(J),i})\setminus\mathcal{A}_{z(J^c),i}}, Y_{\mathcal{A}^c_{z(J),i+1}} | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}) \\
&= -I(\hat{Y}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}}; Y_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}),
\end{aligned}$$

where in the first inequality we have used (14), the second and third equalities follow the chain rule, and in the last equality we have used (16) from Lemma 4.

By definition, $\mathcal{A}_{z(J),\ell+1} = \mathcal{A}_{\ell_z}$, hence $\mathcal{A}_{z(J),\ell+1}\setminus\mathcal{A}_{\ell_z} = \emptyset$ and $\beta_{\ell,1,2} = 0$. Now, for layer ℓ and $\ell-1$, we calculate

$$\begin{aligned}
&\beta_{\ell,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{z(J),\ell+1}} + \beta_{\ell-1,1,2} + \beta_{\ell-1,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}} \\
&\le -I(\hat{Y}_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell}}, U_{\mathcal{A}_{s,\ell}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
&\quad + I(X_{\mathcal{A}^c_{m,\ell}}, U_{\mathcal{A}^c_{s,\ell}}; \hat{Y}_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell}}, U_{\mathcal{A}_{s,\ell}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
&\quad - I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}}; Y_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell}}, Y_{d_k}) \\
&= -I(\hat{Y}_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
&\quad - I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}}; Y_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell}} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell}}, Y_{d_k}) \\
&= -I(\hat{Y}_{(\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell})\cup(\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z})}; Y_{(\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{z(J),\ell})\cup(\mathcal{A}_{z(J),\ell}\setminus\mathcal{A}_{\ell_z})} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
&= -I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell-1}\setminus\mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}),
\end{aligned} \tag{65}$$

where the first and second equalities follow chain rule, and in the last equality we have used a property similar to (50), which is a direct consequence of Property 6 of Lemma 3.

Using this technique iteratively, we obtain $\sum_{i=\tilde{i}}^{\ell} \beta_{i,1,2} + \beta_{i,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} \le -I(\hat{Y}_{\mathcal{A}_{z(J),\tilde{i}}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\tilde{i}}\setminus\mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\tilde{i}}}, U_{\mathcal{A}_{s,\tilde{i}}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k})$.

Consider the summation over all ℓ layers, we have

$$\sum_{i=1}^{\ell} \beta_{i,1,2} + \beta_{i,2\cup3} - \hat{R}_{\mathcal{A}_{z(J),i}\setminus\mathcal{A}_{z(J),i+1}} = -I(\hat{Y}_{\mathcal{A}_{z(J),1}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),1}\setminus\mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, U_{\mathcal{A}_{s,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}). \tag{66}$$

Using (64) and (66), the probability in (62) can be upper bounded by

$$\begin{aligned}
&P(\cap_i \mathcal{E}_{i,2}) \\
&\le 2^{n(R_{\mathcal{A}_{m,1}\setminus\mathcal{A}_{\ell_m}} - \check{R}_{\mathcal{A}_{\ell_s}\setminus\mathcal{A}_{\ell,z}})} \cdot 2^{-n(\sum_{i=1}^{\ell}\beta_{i,1,1} + \sum_{i=1}^{\ell}(\beta_{i,1,2}+\beta_{i,2U3}))} \\
&= 2^{n(R_{\mathcal{A}_{m,1}\setminus\mathcal{A}_{\ell_m}} - \mathcal{I}(\mathcal{A}))},
\end{aligned}$$

where $\mathcal{I}(\mathcal{A})$ is given by

$$\begin{aligned}
&I(X_{\mathcal{A}_{m,1}\setminus\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{s,1}\setminus\mathcal{A}_{\ell_s}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{\ell_s}}) \\
&\quad - I(\hat{Y}_{\mathcal{A}_{z(J),1}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),1}\setminus\mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, U_{\mathcal{A}_{s,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
&\quad + \check{R}_{\mathcal{A}_{\ell_s}\setminus\mathcal{A}_{\ell,z}}.
\end{aligned}$$

Using $\mathcal{A}_{s,1} = \mathcal{A}_{z(J),1}$ by definition, $\mathcal{I}(\mathcal{A})$ can be written as

$$I(X_{\mathcal{A}_{m,1}\setminus\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{s,1}\setminus\mathcal{A}_{\ell_s}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}|X_{\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{\ell_s}}) \\ -I(\hat{Y}_{\mathcal{A}_{s,1}\setminus\mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{s,1}\setminus\mathcal{A}_{\ell_z}}|X_{\mathcal{A}_{m,1}}, U_{\mathcal{A}_{s,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k})+\check{R}_{\mathcal{A}_{\ell_s}\setminus\mathcal{A}_{\ell,z}}.$$

Let $\mathcal{T} \triangleq \mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}, \mathcal{T}^c \triangleq \mathcal{A}_{\ell_m}$. Let $\mathcal{S} \triangleq \mathcal{A}_{s,1} \setminus \mathcal{A}_{\ell_s}$, $\mathcal{S}^c \triangleq \mathcal{A}_{\ell_s}$, hence $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$. Let $\check{\mathcal{S}} \triangleq \mathcal{A}_{\ell_s} \setminus \mathcal{A}_{\ell_z}, \check{\mathcal{S}}^c \triangleq \mathcal{A}_{\ell_z}$, hence $\check{\mathcal{S}}^c = \mathcal{A}_{\ell_s} \setminus \check{\mathcal{S}}$. The above result yields that $P(\mathcal{E}) \to 0$ as $n \to \infty$, if

$$R_{\mathcal{T}} \le I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) \\ -I(\hat{Y}_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{N}\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}}.$$

Assume that the constraint is violated for $\mathcal{T} = \emptyset$, we have

$$I(\hat{Y}_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{N}\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ > I(U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{S}^c}) + \check{R}_{\check{\mathcal{S}}}. \quad (67)$$

In this case, the constraint on $R_{\mathcal{T}}$ can be bounded by:

$$\begin{aligned} R_{\mathcal{T}} &\le I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) \\ &\quad - I(\hat{Y}_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{N}\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ &= I(X_{\mathcal{T}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) \\ &\quad + I(U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{S}^c}) \\ &\quad - I(\hat{Y}_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{N}\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ &< I(X_{\mathcal{T}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}), \end{aligned}$$

where the last inequality follows from (67). This implies that the receiver can treat the signal from the nodes in $\mathcal{S} \setminus \mathcal{T}$ as noise. Hence, the constraint for the case $\mathcal{T} = \emptyset$ can be dropped. This completes the proof of the theorem.

Appendix E
Proof of Remark 5

To prove the lemma, first we note that when

$$\check{R}_{\check{\mathcal{S}}} \ge I(U_{\mathcal{S}^c\setminus\check{\mathcal{S}}^c}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k}|X_{\check{\mathcal{T}}^c}, U_{\check{\mathcal{S}}^c}), \quad (68)$$

the rate expression in (13) yields

$$\begin{aligned} R_{\mathcal{T}} &\le I(X_{\mathcal{T}}, U_{\mathcal{S}\cup\mathcal{S}^c\setminus\check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\check{\mathcal{S}}^c}) \\ &\quad -I(\hat{Y}_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{N}\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &= I(X_{\mathcal{T}}, U_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\check{\mathcal{S}}^c}) \\ &\quad -I(\hat{Y}_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{N}\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \end{aligned}$$

where we have used $\check{\mathcal{S}}^c \subseteq \mathcal{S}^c$ and $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$. Redefine $\mathcal{S} \triangleq \mathcal{N} \setminus \check{\mathcal{S}}^c$ and $\mathcal{S}^c \triangleq \check{\mathcal{S}}^c$ yields the simplified form of the rate expression (17).

Next, in general, the rate expression in (13) can be modified to the following form:

$$\begin{aligned} R_{\mathcal{T}} &\le I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) \\ &\quad - I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c\setminus\check{\mathcal{S}}^c}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &\quad - I(\hat{Y}_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{N}\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ &= I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) \\ &\quad + I(Y_{\mathcal{S}^c\setminus\check{\mathcal{S}}^c}; \hat{Y}_{\mathcal{S}^c\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &\quad - I(\hat{Y}_{\mathcal{N}\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{N}\setminus\check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ &\quad - I(\hat{Y}_{\mathcal{S}^c\setminus\check{\mathcal{S}}^c}; Y_{\mathcal{S}^c\setminus\check{\mathcal{S}}^c}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &= I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k}|X_{\mathcal{S}^c}) \\ &\quad - (\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}) \\ &\quad + \check{R}_{\check{\mathcal{S}}} - I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \end{aligned}$$

where in the last equality we have used $\check{\mathcal{S}}^c = \mathcal{S}^c \setminus \check{\mathcal{S}}$ and $(\mathcal{N}\setminus\check{\mathcal{S}}^c)\setminus(\mathcal{S}^c\setminus\check{\mathcal{S}}^c) = \mathcal{S}$ due to $\check{\mathcal{S}}^c \subseteq \mathcal{S}^c \subseteq \mathcal{N}$ and $\mathcal{S}^c = \mathcal{N}\setminus\mathcal{S}$. It can be seen that if

$$\check{R}_{\check{\mathcal{S}}} \ge I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \quad (69)$$

the rate expression reduces to its simplified form (11).

Conditions (68) and (69) together provide the desired result.

Appendix F
Proof of Corollary 1

Codebook generation and the encoding procedure incorporates those in Theorem 1 and the standard DF [1]. The detailed procedures are provided herein for completeness.

a) Codebook generation:

- Generate 2^{nR_1} i.i.d $\mathbf{x}_1(m_1)$, each according to distribution $p(\mathbf{x}_1) = \prod_{i=1}^n p(x_{1i}), m_1 \in [1 : 2^{nR_1}]$.
- Generate $2^{n\check{R}_3}$ i.i.d $\mathbf{x}_3(s_3)$, each according to distribution $p(\mathbf{x}_3) = \prod_{i=1}^n p(x_{3i}), s_3 \in [1 : 2^{n\check{R}_3}]$.
- For each $\mathbf{x}_3(s_3)$, generate $2^{n\check{R}_2}$ i.i.d $\mathbf{x}_2(s_2)$, each according to distribution $p(\mathbf{x}_2|\mathbf{x}_3) = \prod_{i=1}^n p(x_{2i}|x_{3i}), s_2 \in [1 : 2^{n\check{R}_2}]$.
- For each $(\mathbf{x}_2(s_2), \mathbf{x}_3(s_3))$ pair, generate $2^{n\hat{R}_2}$ i.i.d $\hat{\mathbf{y}}_2(z_2)$, each according to distribution $p(\hat{\mathbf{y}}_2|\mathbf{x}_2, \mathbf{x}_3) = \prod_{i=1}^n p(\hat{y}_{2i}|x_{2i}, x_{3i}), z_2 \in [1 : 2^{n\hat{R}_2}]$.

b) Random Binning:

- Randomly partition the set $\{1, 2, \cdots, 2^{n\hat{R}_2}\}$ into $2^{n\check{R}_2}$ bins. Let $s_2 = \mathcal{B}_2(z_2)$ denote the N-to-1 mapping as the results of binning.
- Randomly partition the set $\{1, 2, \cdots, 2^{n\check{R}_2}\}$ into $2^{n\check{R}_3}$ bins. Let $s_3 = \mathcal{B}_2(s_2)$ denote the N-to-1 mapping as the results of binning.

c) Encoding: In block j,

- source S encodes $\mathbf{x}_1(m_{1,j+1})$;
- relay R_1 finds an index z such that $(\mathbf{x}_2(s_j), \mathbf{x}_3(s_j), \hat{\mathbf{y}}_2(z|s_{2,j}, s_{3,j}), \mathbf{y}_2(j))$ are jointly ϵ-typical. Such a z exists as $n \to \infty$ if

$$\hat{R}_2 \ge I(\hat{Y}_2; Y_2|X_2, X_3). \quad (70)$$

If there exist more than one such z, choose the smallest one and let $z_{2,j} = z$;

- relay R_1 finds the bin index $s_{2,j+1} = \mathcal{B}_2(z_{2,j})$ and $s_{3,j+2} = \mathcal{B}_2(s_{2,j+1})$.
- relay R_2 finds a unique index s_2 such that $(\mathbf{x}_2(s_2), \mathbf{x}_3(s_{3,j}), \mathbf{y}_3(j))$ are jointly ϵ-typical. The probability of error tends to 0 as $n \to \infty$ if

$$\check{R}_2 \le I(X_2; Y_3|X_3), \quad (71)$$

and obtains $s_{3,j+1} = \mathcal{B}_2(s_2)$;
- codewords $\mathbf{x}_1(m_{1,j}), \mathbf{x}_2(s_{2,j})$ and $\mathbf{x}_3(s_{3,j})$ are sent into the channel.

d) Decoding and probability of error: Now, we upper bound the probability of erroneous decoding at the destinations and analyze the constraints on the rate of bin indices $\check{R}_2$ and $\check{R}_3$.

- Using Strategy 1, the decoding procedure at the destinations $D_i, i = 1, 2$, partially follows the one in Theorem 1. In addition, at each decoding layer, the codeword transmitted by the relay R_2 that represents the bin index of s_2, i.e., $\mathbf{x}_3(s_3)$ where $s_3 = \mathcal{B}_2(s_2)$ for each s_2, must be jointly typical with the received signal at the next layer. Hence, using Theorem 1 and (70), when the following constraints are satisfied, the probability of error tends to 0 as $n \to \infty$.

$$R_1 \le I(X_1; \hat{Y}_2, Y_{D_i}|X_2, X_3), \quad (72a)$$

$$R_1 \le I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) + \check{R}_2, \quad (72b)$$

$$R_1 \le I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) + I(X_2; Y_{D_i}|X_3) + \check{R}_3, \quad (72c)$$

$$R_1 \le I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) + I(X_2, X_3; Y_{D_i}). \quad (72d)$$

Now we simplify these constraints. First, it can be seen that when

$$\check{R}_3 \ge \check{R}_2 - \min_{i=1,2} I(X_2; Y_{D_i}|X_3), \quad (73)$$

the constraint (72c) is more relaxed than (72b). We note that random binning imposes $\check{R}_3 \le \check{R}_2$, which is satisfied under condition (73). Hence (72c) can be dropped without inducing additional constraint.

Next, we consider the constraint (72b) in two cases.

- case 1: $I(\hat{Y}_2; Y_2|X_2, X_3, Y_{D_i}) \le I(X_2; Y_3|X_3)$. In this case, choosing $I(\hat{Y}_2; Y_2|X_2, X_3, Y_{D_i}) \le \check{R}_2 \le I(X_2; Y_3|X_3)$ renders the constraint (72b) to be more relaxed than (72a).
- case 2: $I(\hat{Y}_2; Y_2|X_2, X_3, Y_{D_i}) > I(X_2; Y_3|X_3)$. In this case, choosing $\check{R}_2 \le I(X_2; Y_3|X_3) < I(\hat{Y}_2; Y_2|X_2, X_3, Y_{D_i})$ renders the constraint (72b) to be more relaxed than

$$R_1 \le I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) + I(X_2; Y_3|X_3). \quad (74)$$

Constraint (72d) and (74) provide

$$R_1 \le I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) + \min\{I(X_2; Y_3|X_3), I(X_2, X_3; Y_{D_i})\}. \quad (75)$$

In both cases, constraint (71) and the condition $\check{R}_2 \le \hat{R}_2$ which is imposed by random binning are satisfied. Therefore, constraint (72b) can be dropped, and (72a) and (75) yield (19).

- Using Strategy 2, the 1-to-1 mapping implies $\hat{R}_2 = \check{R}_2$, and hence (70) and (71) impose the following constraint:

$$I(\hat{Y}_2; Y_2|X_2, X_3) \le \hat{R}_2 = \check{R}_2 \le I(X_2; Y_3|X_3). \quad (76)$$

However, for the case $\check{R}_2 = \hat{R}_2$, constraint (72b) is more relaxed than (72a). Furthermore, constraint (72c) can be dropped when selecting

$$\check{R}_3 \ge \max_{i=1,2} \min\{I(X_3; Y_{D_i}), \check{R}_2 - I(\check{Y}_2; Y_{D_i}|X_2, X_3, Y_{D_i}) - I(X_2; Y_{D_i}|X_3)\}. \quad (77)$$

Constraints (72a), (72d) and (76) yield (20) in the corollary.

- Using Strategy 3, the recovery of s_2 at the destination D_i implies that

$$\check{R}_2 \le \min_{i=1,2}\{I(X_2; Y_{D_i}|X_3) + \min\{\check{R}_3, I(X_3; Y_{D_i})\}\}, \quad (78)$$

in which the rate $\check{R}_3$ can be chosen as

$$\check{R}_3 \ge I(X_3; Y_{D_i}), \quad (79)$$

such that the right hand side of (78) is maximized, and we obtain

$$\check{R}_2 \le \min_{i=1,2} I(X_2, X_3; Y_{D_i}). \quad (80)$$

If (71) is not binding, it can be seen from (79) and (80) that $\check{R}_3 \le \check{R}_2$ can be satisfied. On the other hand, if (71) is binding, we can choose $\check{R}_3 \le \check{R}_2 = I(X_2; Y_3|X_3)$. In both cases, the constraint $\check{R}_3 \le \check{R}_2$ imposed by random binning can be satisfied and we have

$$\check{R}_2 \le \min\{I(X_2; Y_3|X_3), \min_{i=1,2}\{I(X_2, X_3; Y_{D_i})\}\}. \quad (81)$$

Next, the destinations use s_2 to recover m_1. The probability of erroneous decoding tends to 0 as $n \to \infty$ if

$$R_1 \le I(X_1; \hat{Y}_2, Y_{D_i}|X_2, X_3) \quad (82)$$

$$R_1 \le I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) + \check{R}_2. \quad (83)$$

Hence, when choosing

$$\check{R}_2 = \min\{I((\hat{Y}_2; Y_2|X_2, X_3, Y_{\mathrm{D}_i}), \min\{I(X_2; Y_3|X_3), \min_{i=1,2}\{I(X_2, X_3; Y_{\mathrm{D}_i})\}\}\},$$

which satisfies the constraint $\check{R}_2 \le \hat{R}_2$ imposed by random binning. The constraints (82) and (83) yield (21).

This completes the proof of the corollary.

References

[1] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[2] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[3] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[4] M. H. Yassaee and M. R. Aref, "Generalized compress-and-forward strategy for relay networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 2683–2687.

[5] X. Wu and L.-L. Xie, "On the optimal compressions in the compress-and-forward relay schemes," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2613–2628, May 2013.

[6] T. M. Cover and Y.-H. Kim, "Capacity of a class of deterministic relay channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 591–595.

[7] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. 45th Allerton Conf. Commun, Control, Comput.*, Monticello, IL, USA, Sep. 2007, pp. 129–135.

[8] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.

[9] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 921–930, Mar. 2009.

[10] H.-F. Chong and M. Motani, "The capacity of several new classes of semi-deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6397–6404, Oct. 2011.

[11] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[12] G. Kramer and J. Hou, "Short-message quantize-forward network coding," in *Proc. 8th Int. Workshop Multi-Carrier Syst. Solutions (MC-SS)*, Munich, Germany, May 2011, pp. 1–3.

[13] J. Hou and G. Kramer, "Short message noisy network coding for multiple sources," in *Proc. IEEE Int. Symp. Inf. Theory*, Boston, MA, USA, Jul. 2012, pp. 1677–1681.

[14] J. Hou and G. Kramer. (Aug. 2013). "Short message noisy network coding with a decode-forward option." [Online]. Available: http://arxiv-web3.library.cornell.edu/abs/1304.1692v2

[15] M. H. Yassaee and M. R. Aref, "Slepian–Wolf coding over cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3462–3482, Jun. 2011.

[16] P. Razaghi and W. Yu, "Parity forwarding for multiple-relay networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 158–173, Jan. 2009.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[18] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3517–3535, Oct. 2007.

[19] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

[20] K. Luo, R. H. Gohary, and H. Yanikomeroglu, "Analysis of the generalized DF-CF for Gaussian relay channels: Decode or compress?" *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1810–1821, May 2013.

Kevin Luo (S'02–M'06) received the B.Eng. degree in the Department of Information Engineering from Xidian University, Shaanxi, China in 1991, the M.Sc. degree in the Department of Electrical and Computer Engineering from University of Calgary, Alberta, Canada in 2004, and is currently pursuing his Ph.D. degree in the Department of Systems and Computer Engineering from Carleton University, Ontario, Canada.

Mr. Luo was a member of wireless systems group in Nortel Networks from 2004 to 2009, where he conducted research and development on the fourth-generation wireless communication systems.

His main research interests include information theory, signal processing and cross-layer design with their applications to cooperative communications in wireless networks.

Ramy H. Gohary (S'02–M'06–SM'13) is an adjunct research professor and a senior research associate with the Systems and Computer Engineering Department at Carleton University. He received the B.Sc. (Hons.) degree from Assiut University, Egypt in 1996, the M.Sc. degree from Cairo University, Egypt, in 2000, and the Ph.D. degree from McMaster University, Ontario, Canada in 2006, all in electronics and communications engineering.

Dr. Gohary is the co-inventor of four US patents, and the author of more than thirty well-cited IEEE journal papers. He is also the referee for more than ten scientific IEEE journals, and a member of the technical program committees of seven international IEEE conferences.

Dr. Gohary is a registered Limited Engineering Licensee (LEL) in the province of Ontario.

Dr. Gohary's research interests include analysis and design of MIMO and cooperative wireless communication systems, applications of optimization and geometry in signal processing and communications, information theoretic aspects of multiuser communication systems, and applications of iterative detection and decoding techniques in multiple antenna and multiuser systems.

Halim Yanikomeroglu (S'96–M'98–SM'12) was born in Giresun, Turkey, in 1968. He received the B.Sc. degree in electrical and electronics engineering from the Middle East Technical University, Ankara, Turkey, in 1990, and the M.A.Sc. degree in electrical engineering (now ECE) and the Ph.D. degree in electrical and computer engineering from the University of Toronto, Canada, in 1992 and 1998, respectively.

During 1993-1994, he was with the R&D Group of Marconi Kominikasyon A.S., Ankara, Turkey. Since 1998 he has been with the Department of Systems and Computer Engineering at Carleton University, Ottawa, Canada, where he is now a Full Professor. His research interests cover many aspects of wireless technologies with a special emphasis on cellular networks. He has given a high number of tutorials and invited talks on wireless technologies in the leading international conferences. In recent years, his research has been funded by Huawei, Telus, Blackberry, Samsung, Communications Research Centre of Canada (CRC), and Nortel. This collaborative research resulted in over 20 patents (granted and applied). Dr. Yanikomeroglu has been involved in the organization of the IEEE Wireless Communications and Networking Conference (WCNC) from its inception, including serving as Steering Committee Member as well as the Technical Program Chair or Co-Chair of WCNC 2004 (Atlanta), WCNC 2008 (Las Vegas), and WCNC 2014 (Istanbul). He was the General Co-Chair of the IEEE Vehicular Technology Conference Fall 2010 held in Ottawa. He has served in the editorial boards of the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He was the Chair of the IEEE's Technical Committee on Personal Communications (now called Wireless Technical Committee). He is a Distinguished Lecturer for the IEEE Communications Society (2015-2016) and the IEEE Vehicular Technology Society (2011-2015).

Dr. Yanikomeroglu is a recipient of the IEEE Ottawa Section Outstanding Educator Award in 2014, Carleton University Faculty Graduate Mentoring Award in 2010, the Carleton University Graduate Students Association Excellence Award in Graduate Teaching in 2010, and the Carleton University Research Achievement Award in 2009. Dr. Yanikomeroglu spent the 2011-2012 academic year at TOBB University of Economics and Technology, Ankara, Turkey, as a Visiting Professor. He is a registered Professional Engineer in the province of Ontario, Canada.