On the Capacity of Wireless 1-Hop Intersession Network Coding — A Broadcast Packet Erasure Channel Approach

Chih-Chun Wang, *Member, IEEE*

***Abstract*—Motivated by practical wireless network protocols, this paper focuses on wireless intersession network coding (INC) over a 1-hop neighborhood, of which the exact capacity region remains an open problem. Towards better understanding of the capacity, this work first models the wireless overhearing events by *broadcast packet erasure channels* that are memoryless and stationary. Since most INC gain is resulted from destinations overhearing packets transmitted by other sources, this work then focuses exclusively on the *2-staged INC schemes*, which fully capture the throughput benefits of overheard message side information (MSI) through the use of one-time feedback but refrain from exploiting the broadcast spatial diversity gain of channel output feedback. Under this setting, a capacity outer bound is provided for any number of M coexisting unicast sessions. For the special cases of $M \leq 3$, it is shown that the outer bound can be achieved and is indeed the capacity. To quantify the tightness of the outer bound for $M \geq 4$, a capacity inner bound for general M is provided. Both the outer and inner bounds can be evaluated by any linear programming solver. Numeric results show that for $4 \leq M \leq 5$ with randomly chosen channel parameters, the difference between the outer and inner bounds is within 1% for 96.7% of the times. Focusing exclusively on the benefits of MSI, the results in this paper can also be viewed as the generalization of index-coding capacity from wireline broadcast with binary alphabets to wireless broadcast with high-order alphabets.**

***Index Terms*—Broadcast channel capacity, message side information, network code alignment, packet erasure channels, wireless 1-hop network coding.**

I. Introduction

The throughput benefits of network coding (NC) have been demonstrated and elegantly characterized in [28] for the setting of coding over a single multicast session, termed intrasession NC. However, since most traffic is unicast for which intrasession NC provides no capacity gains (cf. [21], [28] for wireline networks and [11], [19], [36] for wireless erasure networks with acknowledgment), the throughput benefits of NC can be fully realized only when considering both coding within a single multicast session [28] and coding over multiple unicast sessions, the so-called *intersession network coding* (INC) [41]. Determining the capacity of INC is a notoriously challenging problem even for wireline networks [14]–[16]. For example, when considering the integral setting with an unbounded number of coexisting sessions, the complexity of determining the INC capacity is NP-complete [27], [41]. To mitigate its complexity, recent practical wireless INC schemes [20] focus on INC over a 1-hop neighborhood. Despite its simplified setting and attractive complexity savings, the exact capacity region of 1-hop INC remains an open problem and several attempts have since been made to quantify some suboptimal achievable rate regions [6], [10], [26], [33], [35], [39], [47]. Motivated by practical wireless protocols, this paper provides a new model and the corresponding analysis that make progress towards answering the following open questions: Exactly (or at most) how much throughput improvement one can expect from INC in a wireless 1-hop neighborhood over non-coding solutions; and how to achieve (or approach) the corresponding INC capacity.

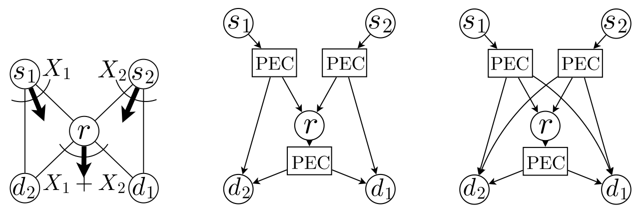
Most 1-hop INC schemes [20], [44] take advantage of the broadcast nature of wireless media. Take Fig. 1(a) for example. Two unicast sessions (s_1, d_1) and (s_2, d_2) would like to send packets X_1 and X_2, respectively, through a common relay node r. Assuming half-duplex and full interference, each node can either transmit or receive, and no two nodes can transmit simultaneously. Also assuming that there is no noise in the wireless links and each transmitted packet always arrives at the target receiver successfully, without network coding four time slots are necessary to complete the two sessions: $s_1 \stackrel{[X_1]}{\longrightarrow} r$, $s_2 \stackrel{[X_2]}{\longrightarrow} r$, $r \stackrel{[X_1]}{\longrightarrow} d_1$, and $r \stackrel{[X_2]}{\longrightarrow} d_2$. [20] observes that if we further assume that each transmitted packet can always be overheard by more than one node, then one can complete the transmission in three time slots by sending packets: $s_1 \stackrel{[X_1]}{\longrightarrow} \{r, d_2\}$, $s_2 \stackrel{[X_2]}{\longrightarrow} \{r, d_1\}$, and $r \stackrel{[X_1+X_2]}{\longrightarrow} \{d_1, d_2\}$. See Fig. 1(a) for illustration. Destinations d_1 and d_2 can then decode the desired packets by subtracting the overheard *side information* X_2 and X_1, respectively, from the coded message $[X_1 + X_2]$. This simple illustrative example quickly becomes a challenging problem when the wireless channels are not perfect. What if d_i cannot overhear all the transmissions from s_j for distinct $i, j \in \{1, 2\}$? What if a coded packet broadcast by r cannot reach both d_1 and d_2 simultaneously? How much throughput improvement can we still expect from INC? The setting becomes even more complicated when considering the random overhearing events[1]

C.-C. Wang is with the Center for Wireless Systems and Applications (CWSA), School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47906, USA. E-mail: chihw@purdue.edu.

This work was supported in part by the NSF Awards CCF-0845968 and CNS-0905331.

Part of the results was published in ISIT 2010. Manuscript received ?? ??, 2010; revised ?? ??, 2010.

[1]The number of types of different overhearing events is generally proportional to $M2^M$ as will be clear in Section II.



(a) (b) (c)

Fig. 1. (a) The optimal 1-hop INC with no channel randomness. (b) The PEC-based canonical wireless 1-hop relay network with no direct communication between s_i and d_i. (c) A similar setting with direct (s_i, d_i) communication among M coexisting sessions (s_i, d_i), $i \in [M] \triangleq \{1, \cdots, M\}$ for general $M > 2$ values.

This capacity problem is related to the Gaussian broadcast channel (GBC) problem with message side information (MSI) [22], [25], [32], [37], [42], [46]. Namely, wireless 1-hop INC can be modeled as the following two-staged scheme: In the first stage, all sources s_j, $j \in [M]$, transmit. The destinations d_i can store the overheard transmission of Stage-1 and use it as MSI in the subsequent stage. In the end of Stage 1, a one-time feedback is sent from all d_i back to r so that the relay r knows which MSI is available at which destination. In the second stage, the relay node r is facing a *broadcast channel problem* such that r would like to broadcast M data streams (one for each session) to M destinations, respectively, while taking advantage of the MSI available at d_i, $i \in [M]$. Although the GBC capacity without MSI is well known for general M values, the corresponding capacity with general MSI is currently known only for the simplest case of $M \leq 2$ [42] and remains an open problem for $M \geq 3$.

Another related setting that is developed independently in the computer science community is the *index coding* problem [3], [7]. Index coding considers a similar broadcast with MSI problem (i.e. the Stage-2 transmission from r to $\{d_i\}$) but focuses on noise-free binary channels and each session having 1 bit of information to transmit, which is different from the GBC settings of additive Gaussian noise with each session transmitting at a real-valued rate R_i. Analytical comparison between the 1-bit setting of index coding and the corresponding multi-bit counterpart can be found in [1]. Other related results that explore the benefits of MSI in unit-edge-capacity networks can be found in [2], [9].

Similar to the existing works, this paper models the wireless 1-hop INC problem by the same 2-stage setting as in GBCs with MSI and in index coding. On the other hand, to capture the packet-by-packet behavior in wireless networks, we consider broadcast packet erasure channels (PECs) [11], [12], [17], [24], [34], [40], [45] instead of the GBC or the noise-free channel in index coding. Each transmission symbol (a packet) is chosen from a finite field $\mathsf{GF}(q)$ for some sufficiently large q and our objective is to maximize the achievable real-valued rate R_i, which is different from the symbol power constraints in GBCs and from the binary-alphabet, integral setting of index coding. Unlike the existing GBC and index coding results that focus mainly on the broadcast problem, we further take into account how s_i should transmit in the first stage in order to optimize the overheard MSI that best benefits the subsequent broadcast stage.[2]

Focusing on the 2-staged setting, a capacity outer bound is provided for any number M of coexisting unicast sessions and any overhearing events modeled by PECs that are memoryless and stationary. For the cases of $M \leq 3$, it is shown that the outer bound meets the capacity, which is a step forward from the $M = 2$ capacity characterization in the GBC with MSI. To quantify the tightness of the outer bound for $M \geq 4$, a capacity inner bound for general M values is provided, which is based on the concept of *code alignment* in the finite field (also see [13], [38], [40], [43]) that is in parallel with the interference alignment technique [4] in the Euclidean space. Both bounds can be computed by linear programming (LP) solvers for any M value. Numeric results show that for $4 \leq M \leq 5$ with randomly chosen PEC parameters, the gap between the two bounds is within 1% for 96.7% of the times, which demonstrates the effectiveness of our capacity bounds from a practical perspective.

The rest of this paper is organized as follows. Section II provides the formal problem formulation. Section III states the main results, which include the capacity region of 2-staged wireless 1-hop INC for $M \leq 3$ sessions and a pair of outer and inner bounds for general M sessions. Section IV presents the equivalent LP-based form of the outer bound. Section V proves the capacity characterization for $M \leq 3$, while Section VI proves the achievability results for general M values. The results of numerical experiments and some observation remarks are included in Section VII. Section VIII concludes the paper.

II. Problem Formulation

A. Packet Erasure Channels (PECs)

For any positive integer K, we use $[K] \triangleq \{1, 2, \cdots, K\}$ to denote the set of integers from 1 to K, and use $2^{[K]}$ to denote the collection of all subsets of $[K]$.

A 1-to-K PEC takes an input $W \in \mathsf{GF}(q)$ and outputs a K-dimensional vector in $\mathbf{Z} \triangleq (Z_1, \cdots, Z_K) \in (\{W\} \cup \{*\})^K$, where the k-th coordinate Z_k being "$*$" denotes that the transmitted symbol W does not reach the k-th receiver (thus being erased). We also assume that there is no other type of noise, i.e., the individual output Z_k is either equal to the input W or an erasure "$*$." The *success probabilities* of a 1-to-K PEC from a single source s to K destinations are described by 2^K non-negative parameters: $p_{s \to T\overline{[K]\backslash T}}$ for all $T \in 2^{[K]}$ such that $\sum_{T \in 2^{[K]}} p_{s \to T\overline{[K]\backslash T}} = 1$ and for all $w \in \mathsf{GF}(q)$, and

$$\mathsf{Prob}\left(\{k \in [K] : Z_k = w\} = T \,|\, W = w\right) = p_{s \to T\overline{[K]\backslash T}}.$$

That is, $p_{s \to T\overline{[K]\backslash T}}$ denotes the probability that the transmitted symbol W is received *by and only by* the receivers in T. We

[2]It is worth mentioning that the 2-staged setting in this work does incur some suboptimality as we refrain the relay r from exploiting the spatial diversity gain of broadcast channels with feedback [31]. On the other hand, the 2-staged setting fully captures the benefits that destinations could overhear packets transmitted from other sources, which is a major contribution of the capacity gain of wireless INC. See Section II-C for further discussion.

can further generalize this notation by defining

$$p_{s\to T_1\overline{T_2}} = \sum_{\forall T\in 2^{[K]}:T_1\subseteq T, T_2\subseteq([K]\backslash T)} p_{s\to T\overline{[K]\backslash T}}.$$

for any pair of disjoint sets $T_1, T_2 \in 2^{[K]}$. Namely, $p_{s\to T_1\overline{T_2}}$ is the probability that the transmitted symbol W is received *by* all the receivers in T_1 and *not by* any receivers in T_2. This notation $p_{s\to T_1\overline{T_2}}$ can be viewed as a marginal success probability that does not care whether the receivers outside $(T_1 \cup T_2)$ receive the transmitted symbol or not.

The following notation will also be used frequently in this work. For all $T \in 2^{[K]}$, we define

$$p_{s;\cup T} = \sum_{\forall T'\in 2^{[K]}:T'\cap T\neq\emptyset} p_{s\to T'\overline{[K]\backslash T'}}.$$

That is, $p_{s;\cup T}$ is the probability that *at least one of the receivers in* T successfully receives the transmitted symbol W. By convention, we define $p_{s;\cup\emptyset} = 0$. For example, when $K = 2$,

$$p_{s;\cup\{1,2\}} = p_{s\to\{1\}\overline{\{2\}}} + p_{s\to\{2\}\overline{\{1\}}} + p_{s\to\{1,2\}\overline{\emptyset}}$$

is the probability that at least one of the two receivers receives the transmitted symbol W. It can be verified by definition that both the following two notations

$$\forall k \in [K],\ p_{s\to\{k\}\overline{\emptyset}} = p_{s;\cup\{k\}} \tag{1}$$

describe the marginal probability that the k-th receiver receives W successfully. We sometimes use $p_{s;k}$ as shorthand for (1).

We can repeatedly use the channel for n time slots and let $W(t)$ and $\mathbf{Z}(t)$ denote the input and output for the t-th time slot. We assume that the 1-to-K broadcast PEC is memoryless and time-invariant, i.e., for any given functions $w(\cdot) : [n] \mapsto \mathsf{GF}(q)$, and $T(\cdot) : [n] \mapsto 2^{[K]}$,

$$\begin{aligned}\mathsf{Prob}\,(\forall t\in[n], \{k : Z_k(t) = w(t)\} = T(t)\\ |\forall t\in[n], W(t) = w(t)) = \prod_{t=1}^{n} p_{s\to T(t)\overline{[K]\backslash T(t)}}.\end{aligned}$$

Note that the above setting allows the success events among different receivers to be dependent, also defined as *spatial dependence*. For example, when two logical receivers are situated in the same physical node, we simply set the $p_{s\to T\overline{[K]\backslash T}}$ parameters to allow perfect correlation between the success events of the two receivers.

B. The Canonical 1-Hop Relay Network With No Direct Path

We consider the following canonical 1-hop relay network. Such a network consists of M source/destination pairs (s_i, d_i), $\forall i \in [M]$ and a common relay node r. Source s_i would like to send nR_i information packets $X_{i,1}$ to X_{i,nR_i} to destination d_i through the common relay r for some positive integer n and fractional rate R_i. For notational simplicity, we define the information vector $\mathbf{X}_i \triangleq (X_{i,1}, \cdots, X_{i,nR_i})$.

The sources, destinations, and the relay are connected by broadcast PECs. More specifically, each s_i is associated with a 1-to-M PEC that sends a symbol $W_i \in \mathsf{GF}(q)$ from s_i to $\{d_j : \forall j \in [M]\backslash i\} \cup \{r\}$. For each s_i, the outputs of its 1-to-M PEC is denoted by $Z_{i\to j}$, $j \in [M]\backslash i$ and by $Z_{i\to r}$, which refer to the outputs arriving at destination d_j and the relay r, respectively.[3] The common relay r is also associated with a 1-to-M PEC that sends a symbol $W_r \in \mathsf{GF}(q)$ from r to $\{d_j : \forall j \in [M]\}$. The corresponding output symbol that arrives at destination d_j is denoted by $Z_{r\to j}$, for all $j \in [M]$.

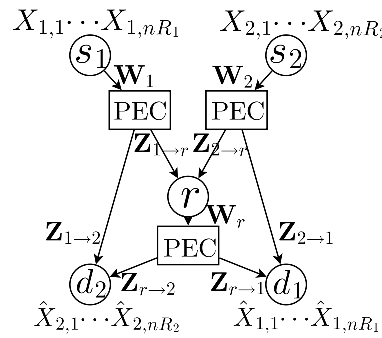


Fig. 2. Illustration of the canonical 1-hop relay network with no direct communication between the source and destination (s_i, d_i) for the case of $M = 2$.

We consider a block network code setting. Namely, each of the sources s_1 to s_M and the relay r can use their associated PECs for exactly n times. We use the following vector notation to denote the $(M+1)n$ transmitted symbols and their corresponding outputs. For all $i \in [M]$, we define

$$\begin{aligned}\mathbf{W}_i &\triangleq (W_i(1), W_i(2), \cdots, W_i(n)).\\ \mathbf{Z}_{i\to r} &\triangleq (Z_{i\to r}(1), Z_{i\to r}(2), \cdots, Z_{i\to r}(n))\\ \mathbf{W}_r &\triangleq (W_r(1), W_r(2), \cdots, W_r(n)).\\ \mathbf{Z}_{r\to i} &\triangleq (Z_{r\to i}(1), Z_{r\to i}(2), \cdots, Z_{r\to i}(n)),\end{aligned}$$

and for all $i, j \in [M]$ and $i \neq j$, we define

$$\mathbf{Z}_{i\to j} \triangleq (Z_{i\to j}(1), Z_{i\to j}(2), \cdots, Z_{i\to j}(n)),$$

for which we use the input argument "(t)", $t = 1, \cdots, n$, to distinguish the n channel usages. An illustration of a canonical 1-hop relay network with $M = 2$ is provided in Fig. 2. In this work, we also assume that all the PECs are memoryless and stationary. The $(M+1)$ PEC channels can thus be described by the joint success probabilities $p_{r\to\cdot}$ and $p_{s_i\to\cdot}$ for all $i \in [M]$ as described in Section II-A. For example, if $M = 4$, $p_{r\to\{1,4\}\overline{\{2,3\}}}$ denotes the probability that a packet sent by r is received by d_1 and d_4 but not by d_2 and d_3. Similarly, $p_{s_2\to\{3,4\}\overline{\{1,r\}}}$ denotes the probability that a packet sent by s_2 is received by d_3 and d_4 but not by d_1 and r. The channel statistics $p_{r\to\cdot}$ and $p_{s_i\to\cdot}$ are assumed to be known to all network nodes, which could be achieved through a training period.

[3]Most existing wireless protocols, including the practical INC protocol in [20], assume a single route and focus on transmit all packets to its immediate downstream neighbor. Motivated by this observation, we preclude any 2-hop transmission directly from s_i to d_i in the canonical setting. The recent concept of opportunistic routing, which allows 2-hop overhearing directly, will be discussed in Section II-D.

Before defining the 2-staged network coding capacity, we define the *channel status vector* of the PEC outputs $\mathbf{Z}_{i\to j}$ by

$$\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}} \triangleq (1_{\{Z_{i\to j}(1)\neq *\}}, \cdots, 1_{\{Z_{i\to j}(n)\neq *\}}),$$

where $1_{\{\cdot\}}$ is the indicator function. Namely, $\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}}$ is an n-dimensional binary vector such that its t-th coordinate is zero if and only if in the t-th channel usage of source s_i, destination d_j receives an erasure.

A 2-staged 1-hop intersession network code is then defined by $(M+1)$ encoding functions $g_{s_i}(\cdot)$ and $g_r(\cdot)$:

$$\forall i \in [M],\ \mathbf{W}_i = g_{s_i}(\mathbf{X}_i) \tag{2}$$

$$\mathbf{W}_r = g_r(\{\mathbf{Z}_{i\to r}, \mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}} : \forall i,j \in [M], i\neq j\}), \tag{3}$$

and M decoding functions $g_{d_j}(\cdot)$:

$$\forall j \in [M],\ \hat{\mathbf{X}}_j = g_{d_j}(\{\mathbf{Z}_{i\to j}, \mathbf{Z}_{r\to j} : \forall i \in [M]\backslash j\}). \tag{4}$$

Namely, each source s_i generates the coded packet $\mathbf{W}_i$ based on its information packets $\mathbf{X}_i$, see (2). Then each destination d_j is allowed to feedback the channel status vectors $\{\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}} : \forall i \in [M]\backslash j\}$ back to the relay r. The relay r will make its coding decision based on what it receives from the sources s_i and on the overhearing patterns reported by the destinations d_j through the channel status vectors, see (3). In the end, each destination d_j decodes the original packets $\hat{\mathbf{X}}_j$ based on the packets received from the sources $\{s_i : \forall i \in [M]\}$ and the relay r, see (4). The encoding and decoding functions can be linear or nonlinear.

Assume that the information symbols $\mathbf{X}_i = (X_{i,1}, \cdots, X_{i,nR_i})$ are independently and uniformly distributed over $\mathsf{GF}(q)$ for all $i \in [M]$. We can then define the achievability of any 2-staged 1-hop intersession network code as follows.

Definition 1: A rate vector $(R_1, \cdots, R_M)$ is achievable if for any $\epsilon > 0$, there exists a 2-staged 1-hop intersession network code satisfying

$$\forall i \in [M],\ \mathsf{P}\left(\hat{\mathbf{X}}_i \neq \mathbf{X}_i\right) < \epsilon,$$

for sufficiently large n and sufficiently large underlying finite field $\mathsf{GF}(q)$.

Definition 2: The 2-staged INC capacity region of a canonical 1-hop relay network of M sessions is defined as the closure of all achievable rate vectors $(R_1, \cdots, R_M)$.

C. Comparison to The Practical COPE Protocol

The definition of a 1-hop intersession network code in (2) to (4) models closely the following 2-stage wireless 1-hop network coding protocol. That is, all sources s_i and the relay r have full knowledge of the PEC parameters before transmission. Each (s_i, d_i) session has nR_i information packets $X_{i,1}, \cdots, X_{i,nR_i}$ to transmit. In the first stage, the sources s_i, $i \in [M]$, take turns and each s_i uses the 1-to-M PEC for n times and sends out n coded symbols generated from the information symbols[4] $X_{i,1}, \cdots, X_{i,nR_i}$, possibly in a non-linear fashion. Due to the randomness of wireless channels, some packets will arrive at r and/or be overheard by d_j and some will not. In the beginning of the second stage, all destinations $\{d_i : i \in [M]\}$ report to the relay r the reception status of the packets sent in the first stage so that r is fully aware of the MSI available at individual d_i. The relay r then uses its 1-to-M PEC for n times and sends out coded packets generated from the packets received by r during the first stage, possibly in a non-linear fashion. No feedback is allowed in the middle of the second stage. In practice, feedback is costly and the "reception reports" may only be sent infrequently [20]. Our setting thus also models the extreme case that a one-time exchange of reception/channel status bit maps is allowed only in the end of Stage 1. After the second stage, each d_i computes the *decoded symbols* $\hat{\mathbf{X}}_i \triangleq (\hat{X}_{i,1}, \cdots, \hat{X}_{i,nR_i})$ based on all the symbols received by d_i.

This 2-stage scheme can be viewed as a fully coordinated, delay elastic approximation of the practical COPE protocol [20]. COPE protocol contains three major components: (i) Opportunistic listening: Each destination is in a promiscuous monitoring mode and stores all the overheard packets, (ii) Opportunistic coding: The relay node decides which packets to be coded together opportunistically, based on the overhearing patterns of its neighbors, and (iii) Learning the states of the neighbors: Although reception reports are periodically sent to advertise the overhearing patterns of the next-hop neighbors of the relay, the relay node still needs to extrapolate the overhearing status of its neighbors since there is always a time lag due to the infrequent periodic feedback.

Our 2-stage scheme closely captures the opportunistic listening component of COPE by modeling the wireless packet transmission as a random broadcast packet erasure channel. In (3), the channel status vector is used to make the coding decision, which captures the opportunistic coding component of COPE. By allowing sufficiently large block length n, we also circumvent the problem of learning the states of the neighbors since we simply wait for a sufficiently large $M \cdot n$ time slots and then use the channel status vectors to convey the *exact* overhearing pattern to the relay r. There is thus no need for extrapolating the overhearing status.

In COPE, the reception reports are broadcast periodically. In contrast, in our 2-staged setting the relay is only allowed to use the channel status vectors of the source-to-destination output $\mathbf{Z}_{i\to j}$ (see (3)) but not the status vectors of the relay-to-destination output $\mathbf{Z}_{r\to j}$. Since channel output feedback could potentially increase the broadcast capacity [17], [31], [45], the periodic broadcast of reception reports of COPE could lead to a throughput advantage over a 2-staged scheme. On the other hand, empirically, most throughput gain of the COPE protocol follows from the MSI instead of the channel out-

[4]Since we focus only on 1-hop INC at relay r, we required that each s_i can only transmit coded symbols generated from its own information packets. That is, no INC is allowed at s_i even if s_i may overhear some packets from s_j or when s_i and s_j are situated at the same physical node. On the other hand, we allow two destinations d_i and d_j to be situated at the same physical node, which is modelled by choosing the $p_{s_i\to\cdot}$ and $p_{r\to\cdot}$ parameters so that the success events for d_i and for d_j are perfectly correlated with each other.

put feedback. Analytically, our preliminary feedback-oriented study [?] shows that in average the difference of the capacity region with $\mathbf{1}_{\{Z_{r\to j}(\cdot)\neq *\}}$ feedback and that of our 2-staged setting is less than 3% for $M=2$. In sum, the contribution of the 2-staged setting is two fold: Firstly, it fully characterizes the benefits of MSI by decoupling it from the convoluted relationship with channel output feedback. Secondly, it serves as a good approximation of the existing COPE protocol. A unified treatment of MSI and feedback is likely to require a different set of analytical approaches (cf. [40] for the broadcast PEC capacity with feedback), which is beyond the scope of this work.

D. Non-Canonical Settings of 1-Hop Wireless Relay Networks

In a wireless network, it is possible that a packet transmitted wirelessly by s_i may jump two hops and directly be heard by d_i although this two-hop jump generally happens with a smaller probability than a 1-hop jump. See Fig. 1(c) for illustration. Recently, new *opportunistic routing* (OR) protocols, such as the MORE protocol [5], [23], take advantage of this spatial diversity and show that throughput can be drastically improved even without INC [5], [30].

The second commonly used technique to enhance the throughput of wireless networks is to perform *cross-layer resource allocation* [29] that allots different numbers of time slots n_i and n_r, for the sources s_i and the relay r, respectively, in order to maximize the achievable throughput (or sometimes maximize the utility instead).

Although our main results in Section III are based on the canonical setting of Section II-B, they can also be generalized to take into account both OR and cross-layer optimization. It turns out that introducing cross-layer optimization does not affect the tightness of the proposed outer and inner bounds (they still match for $M\leq 3$). However, incorporating OR does widen the gap, 1–2% numerically [?], between the outer and inner bounds. Detailed discussion of the non-canonical setting is relegated to Appendix H.

III. Main Results

A. The Capacity Outer Bound For General M Values

We first introduce a critical definition and a central proposition for deriving the capacity outer bound.

Definition 3: We use the convention that any function that is not decreasing is called *an increasing function*. With this convention, a function $f:[0,1]\mapsto\mathbb{R}^+$ is a ZPPLCIF if $f(\cdot)$ is a Zero-Passing (i.e., $f(0)=0$), Piecewise Linear, Concave, Increasing Function. Note that for the following, whenever we say a function is a ZPPLCIF, we automatically imply that its domain is $[0,1]$ and its range is non-negative reals $\mathbb{R}^+$.

Consider a 1-to-3 PEC from a source s with the input symbol denoted by $W\in\mathsf{GF}(q)$ and the output symbols denoted by Z_1, Z_2, and Z_3, respectively. Recall that $p_{s;i}=\mathsf{Prob}(Z_i=W)$, $i=1,2,3$, is the marginal success probability and we use p_1, p_2, and p_3 as shorthand. Without loss of generality, assume that $p_1>p_2>p_3$, which can be achieved by relabeling the receivers.

Proposition 1 (Concavity of Mutual Information): For any n, suppose we use the memoryless and stationary PEC n times and let $\mathbf{W}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ denote the n-dimensional vectors of the input and output symbols, respectively. Let U denote an auxiliary random variable such that we transmit $\mathbf{W}=g(U)$ for some encoding function $g(\cdot)$ and $U\to\mathbf{W}\to(\mathbf{Z}_1,\mathbf{Z}_2,\mathbf{Z}_3)$ form a Markov chain. For any distribution P_U and any encoding function $g(\cdot)$, we have

$$I(U;\mathbf{Z}_1)\geq I(U;\mathbf{Z}_2)\geq I(U;\mathbf{Z}_3) \tag{5}$$

$$I(U;\mathbf{Z}_2)\geq \frac{p_2-p_3}{p_1-p_3}I(U;\mathbf{Z}_1)+\frac{p_1-p_2}{p_1-p_3}I(U;\mathbf{Z}_3). \tag{6}$$

By repeatedly using the above proposition, one can also derive the following corollary:

Corollary 1: Consider a 1-to-M PEC with M different receivers $\mathbf{Z}_1$ to $\mathbf{Z}_M$. Without loss of generality, we add two auxiliary receivers $\mathbf{Z}_0$ and $\mathbf{Z}_{M+1}$ and assume that their marginals satisfy $p_0=1>p_1>\cdots>p_M>p_{M+1}=0$. Let U denote an auxiliary random variable such that we transmit $\mathbf{W}=g(U)$ for some encoding function $g(\cdot)$ and $U\to\mathbf{W}\to(\mathbf{Z}_0,\cdots,\mathbf{Z}_{M+1})$ form a Markov chain. For any distribution P_U and any encoding function $g(\cdot)$, if we define $f:[0,1]\mapsto\mathbb{R}^+$ by $f(p_i)\stackrel{\Delta}{=}I(U;\mathbf{Z}_i),\forall i\in\{0,\cdots,M+1\}$ and linearly interpolating the $f(p)$ value for those $p\in(p_{i-1},p_i),\forall i\in[M+1]$, then $f(\cdot)$ is a ZPPLCIF.

The proof of Proposition 1 is relegated to Appendix A.

Recall that $p_{r;i}$ denotes the marginal success probability from r to d_i. Also recall that $p_{s_i;j}$ (resp. $p_{s_i;r}$) is the marginal probability that d_j (resp. r) receives a packet sent by s_i, and $p_{s_i;\cup T}$ is the probability that a packet sent by s_i is received by *at least* one node in T. The capacity outer bound can then be stated based on the concept of ZPPLCIFs.

Proposition 2: [Capacity Outer Bound] Without loss of generality, assume that $p_{r;1}>p_{r;2}>\cdots>p_{r;M}>0$. A rate vector $(R_1,\cdots,R_M)$ is feasible only if there exist 2^M ZPPLCIFs $\left\{f_S(\cdot):\forall S\in 2^{[M]}\right\}$ indexed by $S\in 2^{[M]}$ such that the following $(M+M^2 2^{M-1}+1)$ inequalities are satisfied:

$$\forall i\in[M],\ R_i\leq p_{s_i;r} \tag{7}$$

$$\forall i,k\in[M], S,T\in 2^{[M]} \text{ satisfying } k\notin S \text{ and } T=S\cup\{k\},$$
$$f_S(p_{r;i})\geq f_T(p_{r;i})+1_{\{i\leq k\}}\left(R_k-p_{s_k;\cup S}\right)^+ \tag{8}$$
$$\text{and } f_\emptyset(p_{r;M})\leq p_{r;M}, \tag{9}$$

where $1_{\{\cdot\}}$ is the indicator function and $(\cdot)^+\stackrel{\Delta}{=}\max(\cdot,0)$ is the projection to the non-negative reals.

We first provide a sketch of the proof, which sheds some insights on the above outer bound. A detailed proof of Proposition 2 is relegated to Appendix B.

Sketch of the proof: Following the 2-stage interpretation in Section II-C, let us focus on the 1-to-M broadcast PEC used by r in the second stage with MSI already created in the first stage. The key step is that for any given wireless 1-hop INC scheme, we consider the following normalized mutual information function with the the base of the logarithm being

q. That is, we choose the function $f_S(\cdot)$ by setting

$$f_S(p_{r;i}) \triangleq \frac{I(U;\mathbf{Z}_{r\to i}|\{\mathbf{Z}_{l\to j}\mathbf{X}_j : \forall j \in S, \forall l \in [M]\backslash j\})}{n}, \tag{10}$$

for all $p_{r;i}$ and interpolating the $f_S(p)$ value for all $p \neq p_{r;i}$. The random variable U in (10) is defined by

$$U = \{\mathbf{X}_i, \mathbf{Z}_{i\to r}, \mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}} : \forall i, j \in [M], i \neq j\}.$$

Namely, U contains all information packets $\mathbf{X}_i$, all the packets that arrive at the relay r, and all the channel status indicators that are fed back to r. The $\mathbf{Z}_{r\to i}$ in (10) denotes the n-dimensional (r, d_i) PEC output vector, $\{\mathbf{Z}_{l\to j} : \forall j \in S, \forall l \in [M]\backslash j\}$ is all the side information received by $\{d_j : \forall j \in S\}$ in the first stage, and $\{\mathbf{X}_j : \forall j \in S\}$ is the information messages for session d_j, $j \in S$. By Proposition 1, $f_S(\cdot)$ must be a ZPPLCIF.

With the above construction of $f_S(\cdot)$, for any $k \notin S$ and $T = S \cup \{k\}$, the following simplified version of (8)

$$f_S(p_{r;k}) \geq f_T(p_{r;k}) + (R_k - p_{s_k;S})^+ \tag{11}$$

can be interpreted as follows.

Let us temporarily allow a genie to grant d_k some additional information: $\{\mathbf{Z}_{l\to j}\mathbf{X}_j : \forall j \in S, \forall l \in [M]\backslash j\}$.

The message information desired by d_k: The term $(R_k - p_{s_k;S})^+$ quantifies the (minimum) amount of information symbols $X_{k,l}$, $l \in [nR_k]$ that are not received by any d_j, $j \in S$. Since $k \notin S$, none of the above information symbols are included in the additional information $\{\mathbf{Z}_{l\to j}\mathbf{X}_j : \forall j \in S, \forall l \in [M]\backslash j\}$ given to d_k. As a result, d_k still desires this amount of information.

The total information observable by d_k: By (10), the mutual information expression $f_S(p_{r;k})$ corresponds to the total amount of additional information that can still reach d_k even after d_k receiving the additional information $\{\mathbf{Z}_{l\to j}\mathbf{X}_j : \forall j \in S, \forall l \in [M]\backslash j\}$.

The undesired interference from the perspective of d_k: We also observe that

$$f_T(p_{r;k}) \triangleq \frac{I(U;\mathbf{Z}_{r\to k}|\{\mathbf{Z}_{l\to j}\mathbf{X}_j : \forall j \in S\cup\{k\}, \forall l \in [M]\backslash j\})}{n}$$

quantifies the amount of information that is not intended for d_k (since we condition on $\mathbf{X}_k$) and cannot be canceled by the side information that is available at d_k, which includes the overheard $\{\mathbf{Z}_{l\to k} : \forall l \in [M]\backslash k\}$ and the information given by the genie $\{\mathbf{Z}_{l\to j}\mathbf{X}_j : \forall j \in S, \forall l \in [M]\backslash j\}$. Therefore, $f_T(p_{r;k})$ can be viewed as the amount of *interference* that cannot be resolved by the side information available to d_k.

Combining the above three observations: (11) is thus a restatement that the total amount of available mutual information must be no smaller than the amount of interference plus the amount of information that needs to be decoded.

Since by definition (10) $f_\emptyset(p_{r;M}) = \frac{I(U;\mathbf{Z}_{r\to M})}{n}$, equation (9) is simply a restatement that the total mutual information at d_M cannot exceed the total number of successfully received time slots. Eq. (7) quantifies the maximal rate that s_i can covey to the relay node r. A detailed proof can be completed following the above reasonings for (7) to (9). ■

Remark: The central idea of Proposition 2 is the interpretation of (11), a similar version of which was also used in the index-coding outer bound [Theorem 6 of [3]]. Proposition 2 can thus be viewed as a generalization of the index-coding outer bound for the wireless setting.

B. The Capacity Region For $M \leq 3$

The tightness result of the outer bound in Proposition 2 is provided as follows.

Proposition 3: The capacity outer bound in Proposition 2 is indeed the capacity region for the cases of $M \leq 3$.

A detailed proof of Proposition 3 is provided in Sections V-A and V-B for $M = 2$ and 3, respectively.

C. The Capacity Inner Bound For General M Values

To quantify the tightness of the proposed outer bound for general $M \geq 4$, we provide the following capacity inner bound, which needs some new notation.

For any arbitrary permutation π on $[M]$, $\pi : [M] \mapsto [M]$, we define a monotonically decreasing sequence of $(M+1)$ sets S_j^π for $j \in [M+1]$ such that for each j, we have $S_j^\pi \triangleq \{\pi(j') : j' = j, \cdots, M\}$. By definition, S_1^π and S_{M+1}^π are $[M]$ and $\emptyset$, respectively. Note that there are $M!$ different distinct permutations π on $[M]$. For the capacity inner bound, we consider $(M+1)!$ ZPPLCIFs $\left\{f_{\pi,S_j^\pi}(\cdot) : \forall \pi, \forall j \in [M+1]\right\}$, which are indexed by π and S_j^π jointly.

For example, if $M = 2$, we have two permutations $\pi_1 = (1,2)$ and $\pi_2 = (2,1)$. Thus $S_1^{\pi_1} = \{1,2\}$, $S_2^{\pi_1} = \{2\}$, $S_3^{\pi_1} = \emptyset$; $S_1^{\pi_2} = \{1,2\}$, $S_2^{\pi_2} = \{1\}$, $S_3^{\pi_2} = \emptyset$. We thus consider 6 ZPPLCIFs $f_{\pi_1,\{1,2\}}$, $f_{\pi_1,\{2\}}$, $f_{\pi_1,\emptyset}$, $f_{\pi_2,\{1,2\}}$, $f_{\pi_2,\{1\}}$, and $f_{\pi_2,\emptyset}$.

We then describe an achievable rate region as follows.

Proposition 4 (Capacity Inner Bound): Following the same settings and definitions of Proposition 2, a rate vector $(R_1, \cdots, R_M)$ can be achieved by a *linear network code* if there exist $M!(2^{M+1} - M - 2)$ non-negative variables:

$$\Big\{y_{k,T,(\pi,S_j^\pi)} \geq 0 : \forall k \in [M], \forall T \in 2^{[M]}, \forall \pi, \forall j \in [M+1], \text{ satisfying } k \notin S_j^\pi \text{ and } T \subseteq S_j^\pi\Big\},$$

and $M2^{M-1}$ non-negative variables:

$$\Big\{z_{k,T} \geq 0 : \forall T \in 2^{[M]}, \forall k \in [M]\backslash T\Big\},$$

and $(M+1)!$ ZPPLCIFs $\left\{f_{\pi,S_j^\pi} : \forall \pi, \forall j \in [M+1]\right\}$, such that jointly the following $(M + M!M^3 + M2^{M-1} + M2^{M-1} + M + 1)$ inequalities are satisfied:

$$\forall i \in [M],\ R_i < p_{s_i;r} \tag{12}$$

$$\forall \pi, \forall i, j, k \in [M], \quad f_{\pi,S_{j+1}^\pi}(p_{r;i}) \geq f_{\pi,S_j^\pi}(p_{r;i}) + 1_{\{i\leq k, k\notin S_{j+1}^\pi\}} \sum_{\{\forall T\in 2^{[M]}: T\subseteq S_{j+1}^\pi\}} y_{k,T,(\pi,S_{j+1}^\pi)} \tag{13}$$

$$\forall k \in [M], \forall T \in 2^{[M]}, k \notin T,$$
$$\sum_{\{\forall \pi, \forall j \in [M+1]: S_j^\pi \supseteq T, k \notin S_j^\pi\}} y_{k,T,(\pi,S_j^\pi)} \geq z_{k,T} \quad (14)$$
$$\forall S \in 2^{[M]}, \forall k \in [M]\backslash S,\ z_{k,S} \leq p_{s_k \to ([M]\backslash(S\cup\{k\})\overline{S}} \quad (15)$$
$$\forall k \in [M], \quad \sum_{\forall S \in 2^{[M]}: k \notin S} z_{k,S} \geq R_k \quad (16)$$
$$\sum_{\forall \pi} f_{\pi,\emptyset}(p_{r;M}) < p_{r;M}. \quad (17)$$

A detailed proof of Proposition 4 is provided in Section VI.

It is worth noting the similarity between Propositions 2 and 4. The main differences are: (i) In the inner bound, the newly introduced permutation π in (17) is used to perform time multiplexing when compared to (9); (ii) In the inner bound, the quantities $(R_k - p_{s_k;\cup S})^+$ are replaced by the *redistribution* inequalities (14) to (16), which are based on the y and z variables and on the probability mass function $p_{s_k \to T_1\overline{T_2}}$; (iii) In the inner bound, when considering a fixed pair of S_{j+1}^π versus S_j^π, multiple k values are considered in (13) for all $k \notin S_{j+1}^\pi$. For comparison, for a fixed pair of S versus T, only a single value k is considered. This is due to that for the achievability result, one needs to take care of decoding the information for all destinations d_k. But for the outer bound result, one only needs to consider a single receiver for which decoding may fail.

IV. Alternative Forms of The Outer Bound

In this section, we present two equivalent forms of the outer bound of Proposition 2, the former of which is based on the rate vectors $(R_1, \cdots, R_M)$ only, while the latter of which is based on the feasibility of a linear programming problem.

A. A Rate-Based Outer Bound For $M \leq 3$

We first consider the simplest case for which $M = 2$. In this case, we have four functions $f_S(\cdot)$ to consider: $f_{\{1,2\}}(\cdot)$, $f_{\{1\}}(\cdot)$, $f_{\{2\}}(\cdot)$, and $f_\emptyset(\cdot)$. We discuss each function separately as follows.

Case 1: $S = \{1, 2\}$. Since $f_{\{1,2\}}(\cdot)$ is always on the smaller side of the inequalities in (8) and (9), we can simply set $f_{\{1,2\}}(p) \triangleq 0$ for all $p \in [0, 1]$.

Case 2: $S = \{1\}$. (8) implies that

$$f_{\{1\}}(p_{r;1}) \geq f_{\{1,2\}}(p_{r;1}) + (R_2 - p_{s_2;1})^+ = (R_2 - p_{s_2;1})^+$$
$$f_{\{1\}}(p_{r;2}) \geq f_{\{1,2\}}(p_{r;2}) + (R_2 - p_{s_2;1})^+ = (R_2 - p_{s_2;1})^+.$$

Case 3: $S = \{2\}$. (8) implies that

$$\begin{aligned} f_{\{2\}}(p_{r;1}) &\geq f_{\{1,2\}}(p_{r;1}) + (R_1 - p_{s_1;2})^+ \\ &= (R_1 - p_{s_1;2})^+ \end{aligned} \quad (18)$$
$$f_{\{2\}}(p_{r;2}) \geq f_{\{1,2\}}(p_{r;2}) + 0 = 0.$$

Note that since $f_{\{2\}}$ is a ZPPLCIF, (18) implies that we must also have

$$f_{\{2\}}(p_{r;2}) \geq \frac{p_{r;2}}{p_{r;1}} f_{\{2\}}(p_{r;1}) \geq \frac{p_{r;2}}{p_{r;1}} (R_1 - p_{s_1;2})^+.$$

Case 4: $S = \emptyset$. Using $S = \emptyset$ and $T = \{1\}$, the discussion in Case 2 and (8) jointly imply that

$$f_\emptyset(p_{r;1}) \geq f_{\{1\}}(p_{r;1}) + R_1 \geq R_1 + (R_2 - p_{s_2;1})^+ \quad (19)$$
$$f_\emptyset(p_{r;2}) \geq f_{\{1\}}(p_{r;2}) + 0 \geq (R_2 - p_{s_2;1})^+. \quad (20)$$

Using $S = \emptyset$ and $T = \{2\}$, the discussion in Case 3 and (8) jointly imply that

$$f_\emptyset(p_{r;1}) \geq f_{\{2\}}(p_{r;1}) + R_2 \geq R_2 + (R_1 - p_{s_1;2})^+ \quad (21)$$
$$\begin{aligned} f_\emptyset(p_{r;2}) &\geq f_{\{2\}}(p_{r;2}) + R_2 \\ &\geq R_2 + \frac{p_{r;2}}{p_{r;1}} (R_1 - p_{s_1;2})^+. \end{aligned} \quad (22)$$

Together (19) to (22) and the concavity of $f_\emptyset(\cdot)$ imply that

$$\begin{aligned} f_\emptyset(p_{r;2}) &\geq \max\left(\frac{p_{r;2}(19)}{p_{r;1}}, (20), \frac{p_{r;2}(21)}{p_{r;1}}, (22)\right) \\ &\geq \max\left(\frac{p_{r;2}}{p_{r;1}}\left(R_1 + (R_2 - p_{s_2;1})^+\right),\right. \\ &\qquad \left. R_2 + \frac{p_{r;2}}{p_{r;1}} (R_1 - p_{s_1;2})^+\right). \end{aligned} \quad (23)$$

Combining (23) and (9), the outer bound becomes

$$\begin{aligned} p_{r;2} \geq \max\left(\frac{p_{r;2}}{p_{r;1}}\left(R_1 + (R_2 - p_{s_2;1})^+\right),\right. \\ \left. R_2 + \frac{p_{r;2}}{p_{r;1}} (R_1 - p_{s_1;2})^+\right). \end{aligned}$$

By further taking into account of (7), we have

Corollary 2: For the case of $M = 2$, the outer bound of Proposition 2 has the following equivalent form:

$$R_1 \leq \min\left(p_{s_1;r}, p_{r;1} - (R_2 - p_{s_2;1})^+\right)$$
$$R_2 \leq \min\left(p_{s_2;r}, p_{r;2} - \frac{p_{r;2}}{p_{r;1}} (R_1 - p_{s_1;2})^+\right).$$

Similar steps can be carried out for the case of $M = 3$. We then have the following alternative form of the outer bound for $M = 3$, the proof of which is provided in Appendix C.

Corollary 3: For $M = 3$, the outer bound of Proposition 2 has the equivalent form of (24) to (27) in Fig. 3 (on the top portion of the next page).

Remark: Corollaries 2 and 3 admit a more conventional form of capacity regions, which depends only on the rate vectors $(R_1, \cdots, R_M)$ and the given PEC parameters $p_{s_i \to \cdot}$ and $p_{r \to \cdot}$.

B. A Linear-Programming-Based Outer Bound For Arbitrary M Values

Although we can repeat the steps of the previous subsection and convert the outer bound in Proposition 2 to its rate-based counterpart for arbitrary M values, the description of the outer bound grows exponentially. On the other hand, a unique feature of the proposed outer bound is that it can be checked by any linear programming (LP) solver. To that end, we first describe the following lemma.

Lemma 1: For any ZPPLCIF $f(\cdot)$ that changes its slope only at M points: $(p_i, f(p_i))$ for $i \in [M]$ and $p_1 > \cdots >$

$$R_i \le p_{s_i;r},\ \forall i \in \{1,2,3\} \tag{24}$$

$$R_1 \le p_{r;1} - \max\left(\left(R_2 - p_{s_2;1}\right)^+ + \left(R_3 - p_{s_3;\{1,2\}}\right)^+, \left(R_2 - p_{s_2;\{1,3\}}\right)^+ + \left(R_3 - p_{s_3;1}\right)^+\right) \tag{25}$$

$$\begin{aligned} R_2 \le p_{r;2} - \max\Bigg(& \frac{p_{r;2}}{p_{r;1}}\left(R_1 - p_{s_1;\{2,3\}}\right)^+ + \left(R_3 - p_{s_3;2}\right)^+, \frac{p_{r;2}}{p_{r;1}}\left(\left(R_1 - p_{s_1;2}\right)^+ + \left(R_3 - p_{s_3;\{1,2\}}\right)^+\right), \\ & \frac{p_{r;1}-p_{r;2}}{p_{r;1}-p_{r;3}}\left(\frac{p_{r;3}}{p_{r;1}}\left(R_1 - p_{s_1;\{2,3\}}\right)^+ + \left(R_3 - p_{s_3;2}\right)^+\right) + \frac{p_{r;2}-p_{r;3}}{p_{r;1}-p_{r;3}}\left(\left(R_1 - p_{s_1;2}\right)^+ + \left(R_3 - p_{s_3;\{1,2\}}\right)^+\right)\Bigg) \end{aligned} \tag{26}$$

$$R_3 \le p_{r;3} - \max\left(\frac{p_{r;3}}{p_{r;1}}\left(\left(R_1 - p_{s_1;3}\right)^+ + \left(R_2 - p_{s_2;\{1,3\}}\right)^+\right), \frac{p_{r;3}}{p_{r;1}}\left(R_1 - p_{s_1;\{2,3\}}\right)^+ + \frac{p_{r;3}}{p_{r;2}}\left(R_2 - p_{s_2;3}\right)^+\right). \tag{27}$$

Fig. 3. The capacity description for the case of $M = 3$.

$p_M \ge 0$, there exist M non-negative values x_h for all $h \in [M]$ such that

$$\forall p \in [0,1],\ f(p) = \sum_{h=1}^{M} x_h \cdot \min(p, p_h). \tag{28}$$

The converse is also true. That is, for any M non-negative values x_h, $\forall h \in [M]$, the corresponding function $f(\cdot)$ satisfying (28) must be a ZPPLCIF that changes its slope only at $(p_i, f(p_i))$ for $i \in [M]$.

The above lemma enables the following equivalent form of Proposition 2.

Proposition 5: A rate vector $(R_1, \cdots, R_M)$ is feasible only if there exist $M2^M$ non-negative variables $\left\{x_{S,h} \ge 0 : \forall S \in 2^{[M]}, \forall h \in [M]\right\}$ indexed by $S \in 2^{[M]}$ and $h \in [M]$ such that the following $(M + M^2 2^{M-1} + 1)$ inequalities are satisfied:

$\forall i \in [M],\ R_i \le p_{s_i;r}$

$\forall i,k \in [M], \forall S, T \in 2^{[M]}$ satisfying $k \notin S$ and $T = S \cup \{k\}$,

$$\sum_{h=1}^{M} x_{S,h} \min(p_{r;i}, p_{r;h}) \ge \sum_{h=1}^{M} x_{T,h} \min(p_{r;i}, p_{r;h}) + 1_{\{i \le k\}}\left(R_k - p_{s_k;\cup S}\right)^+$$

and $\displaystyle\sum_{h=1}^{M} x_{\emptyset,h} \min(p_{r;M}, p_{r;h}) \le p_{r;M} \iff \sum_{h=1}^{M} x_{\emptyset,h} \le 1.$

Therefore, checking whether the outer bound is satisfied is equivalent to checking the feasibility of the corresponding LP problem, which can be achieved by any LP solver. For example, when $M = 6$, the LP solver simply checks the feasibility of a LP problem with 384 non-negative variables and 1159 linear inequalities. Such LP formulation also can be used for solving a utility maximization problem. Namely, given any concave utility function $U(R_1, \cdots, R_M)$, we can maximize the utility over the linear-inequality-based capacity outer-bound region of Proposition 5 by any standard convex optimization solver, the result of which is an upper bound on the optimal utility of any 2-staged INC scheme.

V. The Capacity Region for The Cases of $M \le 3$

In this section, we prove that the outer bound in Proposition 2 is indeed the capacity region for the cases of $M \le 3$ by explicitly constructing capacity-achieving linear network codes. Note that since our constructions focus on a sufficiently large block length n and a sufficient large finite field size $\mathsf{GF}(q)$, we can, without loss of generality, resort to the first-order analysis (the law of large numbers) and assume that random linear network coding (RLNC) [18] generates Maximum Distance Separable (MDS) codes with arbitrarily close-to-one probability.

A. The Capacity of $M = 2$

For any (R_1, R_2) in the interior of the outer bound in Corollary 2, we use the following linear INC scheme that has three major steps.

Step 1: For each time slot $t \in [n]$, each source s_i, $i \in \{1,2\}$, sends out the coded symbol $W_i(t)$ by RLNC, i.e.,

$$W_i(t) = \sum_{k=1}^{nR_i} c_{i,k}(t) X_{i,k}, \tag{29}$$

for which we choose the mixing coefficients $c_{i,k}(t)$ independently and uniformly randomly from $\mathsf{GF}(q)$. Oftentimes, we use $\mathbf{c}_i(t) = (c_{i,1}(t), \cdots, c_{i,nR_i}(t))$ to denote the (nR_i)-dimensional coding vector used in time t. (29) can thus be simplified as $W_i(t) = \mathbf{c}_i(t)\mathbf{X}_i^{\mathrm{T}}$ where $\mathbf{X}_i^{\mathrm{T}}$ is the transpose of the row vector $\mathbf{X}_i$. We also use $\Omega_i = (\mathsf{GF}(q))^{nR_i}$ to denote the full (nR_i)-dimensional message space for the i-th user.

After Step 1, relay r receives $np_{s_i;r}$ coded packets, which is larger than nR_i, the number of information packets. By the MDS property of RLNC, r can successfully decode $X_{i,1}$ to X_{i,nR_i} for all $i \in \{1,2\}$.

Step 2: Each destination d_j sends to r its own overhearing channel status vector $\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}}$ for all distinct $i,j \in \{1,2\}$. Assuming that the seed used by s_i to generate the coded packets is also known to the relay r, r thus knows the $\mathbf{c}_i(t)$ values for all $i \in \{1,2\}$ and $t \in [n]$. Based on the knowledge of the channel status vectors $\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}}$, the relay r also derives

$$\mathcal{V}_{1\to 2} \triangleq \mathsf{span}\left(\{\mathbf{c}_1(t) : \forall t \text{ s.t. } Z_{1\to 2}(t) \neq *\}\right)$$

That is, $\mathcal{V}_{1\to 2}$ is the linear subspace spanned by all coding vectors of the user-1 packets that are overheard by d_2. Again by the MDS property of RLNC, the corresponding normalized rank of $\mathcal{V}_{1\to 2}$ is

$$\mathsf{Rk}_{1\to 2} \triangleq \frac{1}{n}\,\mathrm{Rank}(\mathcal{V}_{1\to 2}) = \min(R_1, p_{s_1;2}). \tag{30}$$

Before proceeding, we define the *complement space* operator "$\ominus$."

Definition 4: For any three linear subspaces $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$, we say $\mathcal{W} = \mathcal{U} \ominus \mathcal{V}$ is a complement space[5] of $\mathcal{V}$ with respect to $\mathcal{U}$ if

$$\begin{aligned} &\mathsf{span}(\mathcal{V}, \mathcal{W}) = \mathcal{U} \\ \text{and } &\mathrm{Rank}(\mathcal{V}) + \mathrm{Rank}(\mathcal{W}) = \mathrm{Rank}(\mathcal{U}). \end{aligned}$$

The relay r then constructs a complement space $\mathcal{V}_{1\to\bar{2}} = \Omega_1 \ominus \mathcal{V}_{1\to 2}$, i.e., $\mathcal{V}_{1\to\bar{2}}$ is the linear subspace that is not overheard by d_2. By the property of complement spaces, the normalized rank of $\mathcal{V}_{1\to\bar{2}}$ becomes

$$\begin{aligned} \mathsf{Rk}_{1\to\bar{2}} \triangleq \frac{\mathrm{Rank}(\mathcal{V}_{1\to\bar{2}})}{n} &= R_i - \frac{\mathrm{Rank}(\mathcal{V}_{1\to 2})}{n} \\ &= (R_1 - p_{s_1;2})^+ \,, \end{aligned} \tag{31}$$

Let $\mathbf{v}_k^{[1\to 2]}$, $k = 1\cdots, n\mathsf{Rk}_{1\to 2}$ and $\mathbf{v}_l^{[1\to\bar{2}]}$, $l = 1\cdots, n\mathsf{Rk}_{1\to\bar{2}}$ denote the basis vectors of $\mathcal{V}_{1\to 2}$ and $\mathcal{V}_{1\to\bar{2}}$, respectively. We also define $\mathcal{V}_{2\to 1}$, $\mathcal{V}_{2\to\bar{1}}$, $\mathbf{v}_k^{[2\to 1]}$, and $\mathbf{v}_l^{[2\to\bar{1}]}$ by symmetry.

Step 3: The goal of this step is for the relay r to construct the corresponding intersession-coded packet $W_r(t)$ for $t \in [n]$. More explicitly, the intersession-coded packet is constructed by

$$\begin{aligned} &W_r(t) = \\ &\left(\sum_{k_1=1}^{n\mathsf{Rk}_{1\to 2}} c_{k_1}^{[1\to 2]}(t)\mathbf{v}_{k_1}^{[1\to 2]} + \sum_{k_2=1}^{n\mathsf{Rk}_{1\to\bar{2}}} c_{k_2}^{[1\to\bar{2}]}(t)\mathbf{v}_{k_2}^{[1\to\bar{2}]}\right)\mathbf{X}_1^{\mathrm{T}} \\ &+\left(\sum_{k_3=1}^{n\mathsf{Rk}_{2\to 1}} c_{k_3}^{[2\to 1]}(t)\mathbf{v}_{k_3}^{[2\to 1]} + \sum_{k_4=1}^{n\mathsf{Rk}_{2\to\bar{1}}} c_{k_4}^{[2\to\bar{1}]}(t)\mathbf{v}_{k_4}^{[2\to\bar{1}]}\right)\mathbf{X}_2^{\mathrm{T}}, \end{aligned}$$

and our task is to design the coefficients $c_{k_1}^{[1\to 2]}(t)$, $c_{k_2}^{[1\to\bar{2}]}(t)$, $c_{k_3}^{[2\to 1]}(t)$, and $c_{k_4}^{[2\to\bar{1}]}(t)$. Namely, instead of working on the elementary basis, we now work on the new basis $\mathbf{v}_{k_1}^{[1\to 2]}$, $\mathbf{v}_{k_2}^{[1\to\bar{2}]}$, $\mathbf{v}_{k_3}^{[2\to 1]}$, and $\mathbf{v}_{k_4}^{[2\to\bar{1}]}$.

The design of the coefficients contains two phases. **Phase 1** lasts for $\frac{n\mathsf{Rk}_{1\to\bar{2}}}{p_{r;1}}$ time slots. Namely, for all t satisfying $0 < t \leq \frac{n\mathsf{Rk}_{1\to\bar{2}}}{p_{r;1}}$, we choose the coefficients $c_{k_2}^{[1\to\bar{2}]}(t)$ independently and uniformly randomly from $\mathsf{GF}(q)$ for all $k_2 \in [n\mathsf{Rk}_{1\to\bar{2}}]$. For all other coefficients $c_{k_1}^{[1\to 2]}(t)$, $c_{k_3}^{[2\to 1]}(t)$, and $c_{k_4}^{[2\to\bar{1}]}(t)$, we set them to zero.

Phase 2 lasts for

$$\max\left(\frac{n\mathsf{Rk}_{1\to 2} + n\mathsf{Rk}_{2\to\bar{1}}}{p_{r;1}}, \frac{n\mathsf{Rk}_{2\to 1} + n\mathsf{Rk}_{2\to\bar{1}}}{p_{r;2}}\right) \tag{32}$$

[5]The complement subspace $\mathcal{U} \ominus \mathcal{V}$ is not unique, i.e., we may have $\mathcal{W}_1 = \mathcal{U} \ominus \mathcal{V}$ and $\mathcal{W}_2 = \mathcal{U} \ominus \mathcal{V}$ for two distinct subspaces $\mathcal{W}_1$ and $\mathcal{W}_2$. The correctness of our network code construction holds for any choice of the complement subspace.

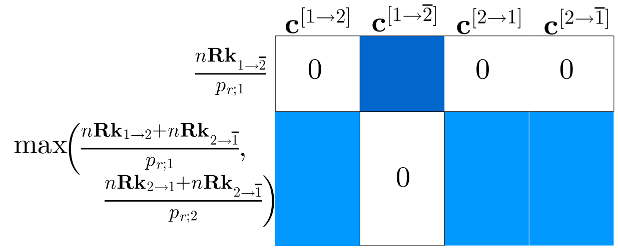


Fig. 4. Illustration of Step 3 of the capacity-achieving scheme for the case of $M = 2$ sessions. The shaded areas indicate that the corresponding coefficients are chosen independently and uniformly randomly.

time slots. Namely, for all t satisfying $\frac{n\mathsf{Rk}_{1\to\bar{2}}}{p_{r;1}} < t \leq \frac{n\mathsf{Rk}_{1\to\bar{2}}}{p_{r;1}} + (32)$, we choose the coefficients $c_{k_1}^{[1\to 2]}(t)$, $c_{k_3}^{[2\to 1]}(t)$, and $c_{k_4}^{[2\to\bar{1}]}(t)$ independently and uniformly randomly from $\mathsf{GF}(q)$ for all valid k_1, k_3, and k_4 values. For the remaining coefficients $c_{k_2}^{[1\to\bar{2}]}(t)$, we set them to zero.

For any $t \in [n]$, if we group all the $c_{k_1}^{[1\to 2]}(t)$, $c_{k_2}^{[1\to\bar{2}]}(t)$, $c_{k_3}^{[2\to 1]}(t)$, and $c_{k_4}^{[2\to\bar{1}]}(t)$ as a row vector and vertically concatenate all the vectors to form an $n \times (nR_1 + nR_2)$ matrix, the above 2-phased scheme can be illustrated in Fig. 4. The shaded areas in Fig. 4 indicate that entries in the corresponding submatrices are chosen independently and uniformly randomly from $\mathsf{GF}(q)$.

To prove that such a 2-phased scheme guarantees decodability, we need the following definition.

Definition 5: A network node can *decode* a vector $\mathbf{v}$ if it can compute the value of the inner product $\mathbf{v}\mathbf{X}^{\mathrm{T}}$, where $\mathbf{X}^{\mathrm{T}}$ is the transpose of the row vector of the information symbols $\mathbf{X}$. A network node can *decode* a subspace $\mathcal{V}$ if it can decode all $\mathbf{v} \in V$.

We then have the following self-exploratory lemmas

Lemma 2: If a network node can decode two subspaces $\mathcal{V}_1$ and $\mathcal{V}_2$, then it can decode the corresponding linear span $\mathsf{span}(\mathcal{V}_1, \mathcal{V}_2)$.

Lemma 3: A network node can decode all the information symbols $\mathbf{X}$ if and only if it can decode the entire message space Ω.

We then notice that with the 2-phased scheme d_1 will receive $p_{r;1} \cdot \frac{n\mathsf{Rk}_{1\to\bar{2}}}{p_{r;1}}$ Phase-1 packets. Since Phase 1 packets are constructed by the $n\mathsf{Rk}_{1\to\bar{2}}$ basis vectors of $\mathcal{V}_{1\to\bar{2}}$, by the MDS property of RLNC, d_1 can decode $\mathcal{V}_{1\to\bar{2}}$. By similar arguments, d_1 will receive no less than $n\,(\mathsf{Rk}_{1\to 2} + \mathsf{Rk}_{2\to\bar{1}})$ packets in Phase 2. Since d_1 has already had all the side information of $\mathcal{V}_{2\to 1}$, from the perspective of decoding, d_1 is facing only two unknown subspaces $\mathcal{V}_{1\to 2}$ and $\mathcal{V}_{2\to\bar{1}}$. Since the number of observed packets is no smaller than the overall degrees of freedom $n\,(\mathsf{Rk}_{1\to 2} + \mathsf{Rk}_{2\to\bar{1}})$, by the MDS property of RLNC, it is guaranteed that d_1 can decode both subspaces $\mathcal{V}_{1\to 2}$ and $\mathcal{V}_{2\to\bar{1}}$. Since d_1 decodes $\mathcal{V}_{1\to\bar{2}}$ and $\mathcal{V}_{1\to 2}$ in Phases 1 and 2, respectively, Lemmas 2 and 3 ensure that d_1 can decode its information symbols $\mathbf{X}_1$.

From d_2's perspective, it will receive no less than $p_{r;2}\frac{n\mathsf{Rk}_{2\to 1}+n\mathsf{Rk}_{2\to\bar{1}}}{p_{r;2}}$ packets in Phase 2. Since the $\mathcal{V}_{1\to 2}$ subspace is known to d_2 as side information, d_2 is facing two unknown subspaces $\mathcal{V}_{2\to 1}$ and $\mathcal{V}_{2\to\bar{1}}$ with the total degrees of freedom $n(\mathsf{Rk}_{2\to 1} + \mathsf{Rk}_{2\to\overline{1}})$. By the MDS property of RLNC, it is guaranteed that d_2 can decode both subspaces $\mathcal{V}_{2\to 1}$ and $\mathcal{V}_{2\to\overline{1}}$, and thus decode its information symbols $\mathbf{X}_2$. The above discussion proves the decodability of the proposed 2-phased scheme.

The remaining task is to show that the 2-phased scheme does not exceed the block length budget n. That is, we need to show that

$$n \geq \frac{n\mathsf{Rk}_{1\to\overline{2}}}{p_{r;1}} + \max\left(\frac{n\mathsf{Rk}_{1\to 2} + n\mathsf{Rk}_{2\to\overline{1}}}{p_{r;1}}, \frac{n\mathsf{Rk}_{2\to 1} + n\mathsf{Rk}_{2\to\overline{1}}}{p_{r;2}}\right). \tag{33}$$

By substituting the Rk variables in (33) with (30) and (31) and by noticing that the resulting (33) is a restatement of the outer bound in Corollary 2, the proof is thus complete.

Remark: In the decoding step, we assume that for any t, the *global encoding kernel* [21] of $W_r(t)$, also known as the *global coding vector*, is available at the destinations. This assumption generally is not true. However, we can use the same idea as in the generation-based network coding protocol [8] to circumvent this difficulty. Namely, we raise the finite field size from $\mathsf{GF}(q)$ to $\mathsf{GF}(q^K)$, and treat each $\mathsf{GF}(q^K)$ symbol as a vector of K symbols. Then we can use the first $n(R_1+R_2)$ symbols of each vector to carry the $(n(R_1+R_2))$-dimensional global encoding kernel and repeatedly use the same kernel to encode the remaining $(K-n(R_1+R_2))$ $\mathsf{GF}(q)$-symbols. In this way, the cost of transmitting the kernel can be fully absorbed when we choose a sufficiently large K. For simplicity, we will assume that the global encoding kernel is known to the destinations in our subsequent discussion of achievability.

B. The Capacity of $M = 3$

For any (R_1, R_2, R_3) in the interior of the outer bound of Corollary 3, perform the following three steps.

Step 1: For each time slot $t \in [n]$, each source s_i, $i \in \{1,2,3\}$, sends out the coded symbol $W_i(t)$ by RLNC, i.e.,

$$W_i(t) = \sum_{k=1}^{nR_i} c_{i,k}(t) X_{i,k},$$

for which we choose the mixing coefficients $c_{i,k}(t)$ independently and uniformly randomly from $\mathsf{GF}(q)$. $\mathbf{c}_i(t) \triangleq (c_{i,1}(t), \cdots, c_{i,nR_i}(t))$ denotes the corresponding (nR_i)-dimensional coding vector used in time t.

After Step 1, relay r receives $np_{s_i;r}$ coded packets, which is larger than nR_i, the number of information packets. By the MDS property of RLNC, r can successfully decode $X_{i,1}$ to X_{i,nR_i} for all $i \in \{1,2,3\}$.

Step 2: Each destination d_j, $j \in \{1,2,3\}$, sends to r its own overhearing channel status vector $\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}}$ for $i \neq j$. Based on the knowledge about the channel status vectors $\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}}$, the relay r derives the following linear subspaces $\mathcal{V}_{i;j}$ for all $i, j \in \{1,2,3\}$, $i \neq j$:

$$\mathcal{V}_{i;j} \triangleq \mathsf{span}\left(\{\mathbf{c}_i(t) : \forall t \text{ s.t. } Z_{i\to j}(t) \neq *\}\right)$$

That is, $\mathcal{V}_{i;j}$ is the linear subspace spanned by all the coding vectors of the user-i packets that are overheard by d_j. For any distinct $i, j_1, j_2 \in \{1,2,3\}$, define

$$\begin{aligned}
\mathcal{V}_{i\to j_1 j_2} &\triangleq \mathcal{V}_{i;j_1} \cap \mathcal{V}_{i;j_2} \\
\mathcal{V}_{i\to j_1\overline{j_2}} &\triangleq \mathcal{V}_{i;j_1} \ominus \mathcal{V}_{i\to j_1 j_2} \\
\mathcal{V}_{i\to \overline{j_1} j_2} &\triangleq \mathcal{V}_{i;j_2} \ominus \mathcal{V}_{i\to j_1 j_2} \\
\mathcal{V}_{i\to \overline{j_1 j_2}} &\triangleq \Omega_i \ominus \mathsf{span}\left(\mathcal{V}_{i;j_1}, \mathcal{V}_{i;j_2}\right),
\end{aligned}$$

which correspond to the linear subspaces of user-i that are overheard by both j_1 and j_2, by j_1 only, by j_2 only, and by neither j_1 nor j_2, respectively. By the definition of complement spaces and by the MDS property of RLNC, the corresponding normalized ranks $\mathsf{Rk}_{i\to\cdot}$ must satisfy:

$$\begin{aligned}
&\mathsf{Rk}_{i\to j_1 j_2} + \mathsf{Rk}_{i\to j_1\overline{j_2}} = \frac{\mathrm{Rank}(V_{i;j_1})}{n} = \min(R_i, p_{s_i;j_1}) \\
&\mathsf{Rk}_{i\to j_1 j_2} + \mathsf{Rk}_{i\to \overline{j_1} j_2} = \frac{\mathrm{Rank}(V_{i;j_2})}{n} = \min(R_i, p_{s_i;j_2}) \\
&\mathsf{Rk}_{i\to j_1 j_2} + \mathsf{Rk}_{i\to j_1\overline{j_2}} + \mathsf{Rk}_{i\to \overline{j_1} j_2} \\
&\quad = \frac{\mathrm{Rank}(\mathsf{span}(V_{i;j_1}, V_{i;j_2}))}{n} = \min(R_i, p_{s_i;\cup\{j_1,j_2\}}) \\
&\mathsf{Rk}_{i\to j_1 j_2} + \mathsf{Rk}_{i\to j_1\overline{j_2}} + \mathsf{Rk}_{i\to \overline{j_1} j_2} + \mathsf{Rk}_{i\to \overline{j_1 j_2}} = R_i.
\end{aligned}$$

Solving the above equations and using the following notation

$$\forall T \in 2^{[M]}, \forall i \in [M]\backslash T,\ \overline{R}_{i;T} \triangleq \left(R_i - p_{s_i;\cup T}\right)^+, \tag{34}$$

we thus have

$$\mathsf{Rk}_{i\to \overline{j_1 j_2}} = \left(R_i - p_{r;\cup\{j_1,j_2\}}\right)^+ = \overline{R}_{i;\{j_1,j_2\}} \tag{35}$$

$$\begin{aligned}
\mathsf{Rk}_{i\to j_1\overline{j_2}} &= \left(R_i - p_{r;j_2}\right)^+ - \left(R_i - p_{r;\cup\{j_1,j_2\}}\right)^+ \\
&= \overline{R}_{i;j_2} - \overline{R}_{i;\{j_1,j_2\}}
\end{aligned} \tag{36}$$

$$\begin{aligned}
\mathsf{Rk}_{i\to \overline{j_1} j_2} &= \left(R_i - p_{r;j_1}\right)^+ - \left(R_i - p_{r;\cup\{j_1,j_2\}}\right)^+ \\
&= \overline{R}_{i;j_1} - \overline{R}_{i;\{j_1,j_2\}}
\end{aligned} \tag{37}$$

$$\mathsf{Rk}_{i\to j_1 j_2} = R_i - \overline{R}_{i;j_1} - \overline{R}_{i;j_2} + \overline{R}_{i;\{j_1,j_2\}}. \tag{38}$$

With the above definition of the subspaces, let $\mathbf{v}_k^{[i\to T\overline{\{j_1,j_2\}\backslash T}]}$, $k = 1\cdots, n\mathsf{Rk}_{i\to T\overline{\{j_1,j_2\}\backslash T}}$ denote the basis vectors of $\mathcal{V}_{i\to T\overline{\{j_1,j_2\}\backslash T}}$ for all $T \subseteq \{j_1, j_2\}$.

Step 3: Relay r constructs the following intersession-coded packet $W_r(t)$ for all $t \in [n]$:

$$W_r(t) = \sum_{i=1}^{3} \left(\sum_{\substack{\forall T_1 \subseteq \{1,2,3\}\backslash i, \\ T_2 = \{1,2,3\}\backslash(T_1 \cup \{i\})}} \sum_{k=1}^{n\mathsf{Rk}_{i\to T_1\overline{T_2}}} c_k^{[i\to T_1\overline{T_2}]}(t) \mathbf{v}_k^{[i\to T_1\overline{T_2}]} \right) \mathbf{X}_i^{\mathrm{T}},$$

based on the new basis vectors $\mathbf{v}_k^{[i\to T_1\overline{T_2}]}$. Our task is to design the corresponding coefficients $c_k^{[i\to T_1\overline{T_2}]}(t)$ for all $t \in [n]$.

The design of the coefficients contains four phases and each of them is of duration L_1 to L_4, respectively, such that $L_1 + L_2 + L_3 + L_4 = n$ is the total number available time slots. The "l-th phase" corresponds to the t values satisfying

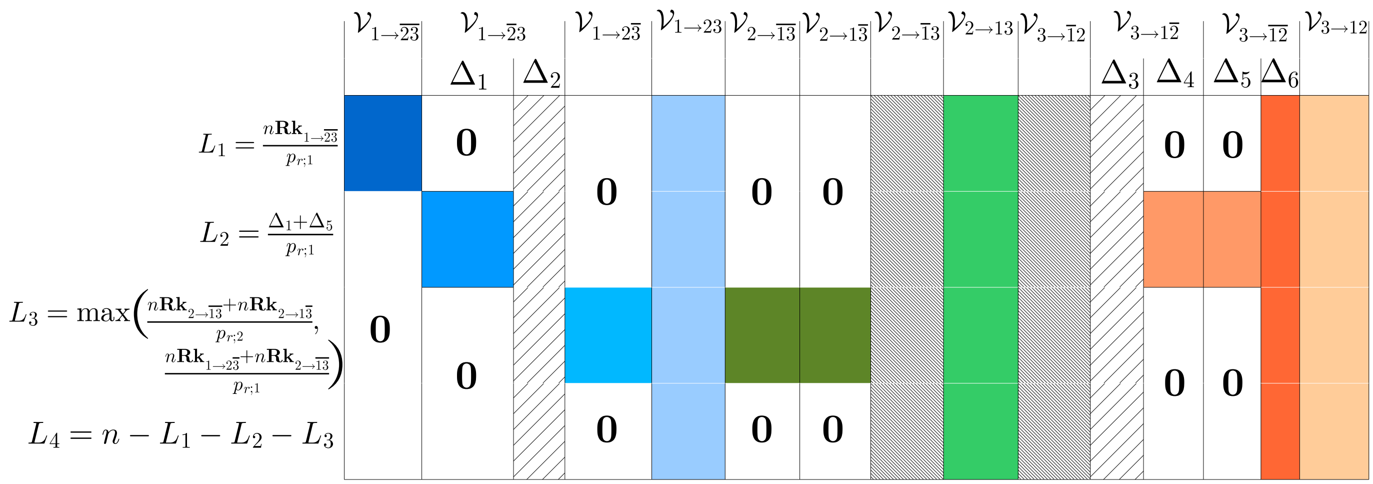


Fig. 5. Illustration of Step 3 of the capacity-achieving scheme for the case of $M = 3$ sessions.

$\sum_{k=1}^{l-1} L_k < t \le \sum_{k=1}^{l} L_k$. With different choices of i, $T_1 \subseteq \{1,2,3\}\backslash i$, and $T_2 = \{1,2,3\}\backslash(T_1 \cup \{i\})$, we have 12 different types of $\mathcal{V}_{i\to T_1\overline{T_2}}$ to consider. Similar to the discussion of the case $M = 2$, we can describe the coefficients by their corresponding $n\times\left(n\sum_{i=1}^{3} R_i\right)$ matrix form. Namely, for any $t \in [n]$, we treat the corresponding coefficients $\{c_k^{[i\to T_1\overline{T_2}]}(t) : \forall i, T_1, T_2, k\}$ as a row vector and vertically concatenate all n row vectors to form a coding matrix. Fig. 5 illustrates the high-level matrix-based description of a capacity-achieving network coding solution. For the following, we will explain in details how to construct the coding matrix in Fig. 5.

We first notice that in Fig. 5 the $n\mathsf{Rk}_{1\to\bar{2}3}$ columns corresponding to $\mathcal{V}_{1\to\bar{2}3}$ are further partitioned into two groups of Δ_1 and Δ_2 columns, respectively. The values of Δ_1 and Δ_2 will be described shortly after. Similarly, the columns corresponding to $\mathcal{V}_{3\to1\bar{2}}$ (resp. $\mathcal{V}_{3\to\overline{12}}$) are partitioned into two groups of sizes (Δ_3, Δ_4) (resp. (Δ_5, Δ_6)).

We then set the durations $L_1 = \frac{n\mathsf{Rk}_{1\to\overline{23}}}{p_{r;1}}$, $L_2 = \frac{\Delta_1+\Delta_5}{p_{r;1}}$,

$$L_3 = n\cdot\max\left(\frac{\mathsf{Rk}_{2\to\overline{13}}+\mathsf{Rk}_{2\to1\bar{3}}}{p_{r;2}}, \frac{\mathsf{Rk}_{1\to2\bar{3}}+\mathsf{Rk}_{2\to\overline{13}}}{p_{r;1}}\right),$$

and $L_4 = n - L_1 - L_2 - L_3$. As a result, to fully describe the structure of the coding matrix, we only need to describe three parameters Δ_1, Δ_4, and Δ_5 since $\Delta_2 = n\mathsf{Rk}_{1\to\bar{2}3} - \Delta_1$, $\Delta_3 = n\mathsf{Rk}_{3\to1\bar{2}} - \Delta_4$, and $\Delta_6 = n\mathsf{Rk}_{3\to\overline{12}} - \Delta_5$. The choices of Δ_1, Δ_4, and Δ_5 depend on the values of $\mathsf{Rk}_{1\to\bar{2}3}$, $\mathsf{Rk}_{3\to1\bar{2}}$, and $\mathsf{Rk}_{3\to\overline{12}}$. Totally, we have 4 different cases when deciding the values of Δ_1, Δ_4, and Δ_5.

Case 1: $\mathsf{Rk}_{3\to1\bar{2}} \ge \mathsf{Rk}_{1\to\bar{2}3}$. In this case, we set $\Delta_1 = \Delta_4 = \Delta_5 = 0$.

Case 2: $\mathsf{Rk}_{1\to\bar{2}3} \ge \mathsf{Rk}_{3\to1\bar{2}} \ge \frac{p_{r;3}}{p_{r;1}}\mathsf{Rk}_{1\to\bar{2}3}$. In this case, we set

$$\Delta_1 = n\frac{p_{r;1}\left(\mathsf{Rk}_{1\to\bar{2}3} - \mathsf{Rk}_{3\to1\bar{2}}\right)}{p_{r;1} - p_{r;3}}$$
$$\Delta_4 = n\frac{p_{r;3}\left(\mathsf{Rk}_{1\to\bar{2}3} - \mathsf{Rk}_{3\to1\bar{2}}\right)}{p_{r;1} - p_{r;3}}, \text{ and } \Delta_5 = 0. \tag{39}$$

Case 3: $\frac{p_{r;3}}{p_{r;1}}\mathsf{Rk}_{1\to\bar{2}3} \ge \mathsf{Rk}_{3\to1\bar{2}}$ and

$$\mathsf{Rk}_{3\to1\bar{2}} + \mathsf{Rk}_{3\to\overline{12}} \ge \frac{p_{r;3}}{p_{r;1}}\left(\mathsf{Rk}_{1\to\bar{2}3} + \mathsf{Rk}_{3\to\overline{12}}\right). \tag{40}$$

In this case, we set

$$\Delta_1 = n\mathsf{Rk}_{1\to\bar{2}3}, \quad \Delta_4 = n\mathsf{Rk}_{3\to1\bar{2}},$$
$$\text{and } \Delta_5 = n\frac{p_{r;3}\mathsf{Rk}_{1\to\bar{2}3} - p_{r;1}\mathsf{Rk}_{3\to1\bar{2}}}{p_{r;1} - p_{r;3}}. \tag{41}$$

Case 4: $\frac{p_{r;3}}{p_{r;1}}\left(\mathsf{Rk}_{1\to\bar{2}3} + \mathsf{Rk}_{3\to\overline{12}}\right) \ge \mathsf{Rk}_{3\to1\bar{2}} + \mathsf{Rk}_{3\to\overline{12}}$. In this case, we set $\Delta_1 = n\mathsf{Rk}_{1\to\bar{2}3}$, $\Delta_4 = n\mathsf{Rk}_{3\to1\bar{2}}$ and $\Delta_5 = n\mathsf{Rk}_{3\to\overline{12}}$.

The following lemma ensures the legitimacy of the above choices of Δ_1, Δ_4, and Δ_5.

Lemma 4: For any (R_1, R_2, R_3) in the interior of the outer bound of Corollary 3, all Δ_1 to Δ_6 are non-negative and L_4 is also non-negative. Moreover, the following inequality always holds.

$$\frac{\Delta_1 + \Delta_5}{p_{r;1}} \ge \frac{\Delta_4 + \Delta_5}{p_{r;3}}. \tag{42}$$

To prove this lemma, we need to use the assumption that (R_1, R_2, R_3) are in the interior of (25) to (27). A detailed proof is relegated to Appendix D.

With the structure of the coding matrix fixed, we now describe how to choose the entries of the submatrices. In Fig. 5, we use "**0**" to denote the zero submatrix in the overall coding matrix. For example, in Phases 2 to 4, all $c_k^{[1\to\overline{23}]}(t)$, $k = 1,\cdots,n\mathsf{Rk}_{1\to\overline{23}}$ are set to zero. Similarly, in Phases 1, 3, and 4, the coefficients $c_k^{[1\to\bar{2}3]}(t)$ of the first Δ_1 columns of $\mathcal{V}_{1\to\bar{2}3}$, i.e., $k = 1,\cdots,\Delta_1$, are set to zero.

We now consider the shaded boxes in Fig. 5 that are not in the columns corresponding to $\mathcal{V}_{2\to\bar{1}3}$, $\mathcal{V}_{3\to\bar{1}2}$, the Δ_2 portion of $\mathcal{V}_{1\to\bar{2}3}$, and the Δ_3 portion of $\mathcal{V}_{3\to1\bar{2}}$. Namely, consider those boxes in solid colors. For those submatrices, we choose the corresponding coefficient $c_k^{[i\to T_1\overline{T_2}]}(t)$ independently and uniformly randomly from $\mathsf{GF}(q)$. For example, in Phase 3, we choose $c_k^{[1\to2\bar{3}]}(t)$, $k = 1,\cdots,n\mathsf{Rk}_{1\to2\bar{3}}$ independently and uniformly randomly from $\mathsf{GF}(q)$. Similarly, in Phase 2, we choose the coefficients $c_k^{[3\to 1\overline{2}]}(t)$ of the last Δ_4 columns of $\mathcal{V}_{3\to 1\overline{2}}$, i.e., $k = (n\mathsf{Rk}_{3\to 1\overline{2}} - \Delta_4) + 1, \cdots, n\mathsf{Rk}_{3\to 1\overline{2}}$, independently and uniformly randomly from $\mathsf{GF}(q)$.

The construction of the four groups of columns (i) the last Δ_2 columns in $\mathcal{V}_{1\to \overline{2}3}$, (ii) $\mathcal{V}_{2\to \overline{1}3}$, (iii) $\mathcal{V}_{3\to \overline{1}2}$, and (iv) the first Δ_3 columns in $\mathcal{V}_{3\to 1\overline{2}}$ is special, and we use the *code alignment* technique, which was also used in [38], [40], [43] and is in parallel with the interference alignment technique for additive Gaussian channels [4].

Using the code alignment technique, we first construct an $n\times\max(\Delta_2, \Delta_3)$ matrix $\mathbf{A}$ by randomly generating each entry of $\mathbf{A}$ independently and uniformly randomly from $\mathsf{GF}(q)$. We then copy the first Δ_2 columns of $\mathbf{A}$ to the last Δ_2 columns of $\mathcal{V}_{1\to \overline{2}3}$, and copy the first Δ_3 columns of $\mathbf{A}$ to the first Δ_3 columns of $\mathcal{V}_{3\to 1\overline{2}}$. Note that with this construction, the last Δ_2 columns of $\mathcal{V}_{1\to \overline{2}3}$ are independently and uniformly randomly distributed, and so are the first Δ_3 columns of $\mathcal{V}_{3\to 1\overline{2}}$. On the other hand, jointly these columns are *highly correlated* as they are the overlapped copies of the common matrix $\mathbf{A}$. More explicitly, for all $t \in [n]$ and for all $1 \leq k \leq \min(\Delta_2, \Delta_3)$, we have

$$c_{k+\Delta_1}^{[1\to \overline{2}3]}(t) = c_k^{[3\to 1\overline{2}]}(t). \tag{43}$$

Since the entries of these two groups of columns are perfectly aligned, it gives the name *code alignment*.

Similarly, code alignment is also applied to columns of $\mathcal{V}_{2\to \overline{1}3}$ and $\mathcal{V}_{3\to \overline{1}2}$. Namely, we construct an $n \times (n\cdot\max(\mathsf{Rk}_{2\to \overline{1}3}, \mathsf{Rk}_{3\to \overline{1}2}))$ matrix $\mathbf{B}$ by randomly generating each entry of $\mathbf{B}$ independently and uniformly randomly from $\mathsf{GF}(q)$. We then copy the first $n\mathsf{Rk}_{2\to \overline{1}3}$ columns of $\mathbf{B}$ to the $n\mathsf{Rk}_{2\to \overline{1}3}$ columns of $\mathcal{V}_{2\to \overline{1}3}$, and copy the first $n\mathsf{Rk}_{3\to \overline{1}2}$ columns of $\mathbf{B}$ to the $n\mathsf{Rk}_{3\to \overline{1}2}$ columns of $\mathcal{V}_{3\to \overline{1}2}$. For all $t \in [n]$ and for all $1 \leq k \leq n\cdot\min(\mathsf{Rk}_{2\to \overline{1}3}, \mathsf{Rk}_{3\to \overline{1}2})$, we thus have

$$c_k^{[2\to \overline{1}3]}(t) = c_k^{[3\to \overline{1}2]}(t). \tag{44}$$

In Fig. 5, the code aligned submatrices are illustrated with linear hatching. The description of the entire network code is thus complete.

The remaining task is to prove that with the linear network code in Fig. 5, each destination d_j can decode the desired information packets $\mathbf{X}_j$.

The case of d_1: We first consider the Phase-3 and Phase-4 packets received by d_1, which form a $(p_{r;1}(L_3 + L_4)) \times \left(n\sum_{i=1}^3 R_i\right)$ matrix. We rearrange the columns so that the columns corresponding to $\mathcal{V}_{1\to 2\overline{3}}$ and $\mathcal{V}_{2\to \overline{13}}$ are the first $n\left(\mathsf{Rk}_{1\to 2\overline{3}} + \mathsf{Rk}_{2\to \overline{13}}\right)$ columns. We call these columns the Group-1 columns and use $\mathsf{nc}_{\mathrm{G1}} \triangleq n\left(\mathsf{Rk}_{1\to 2\overline{3}} + \mathsf{Rk}_{2\to \overline{13}}\right)$ to denote the number of Group-1 columns. We also rearrange the rows so that the Phase-3 packets correspond to the first $p_{r;1}L_3$ rows while the Phase-4 packets correspond to the last $p_{r;1}L_4$ rows. After rearrangement, the coding matrix becomes

$$\left(\begin{array}{c|c} \mathbf{A}_1 & \mathbf{A}_2 \\ \hline \mathbf{0} & \mathbf{A}_3 \end{array}\right), \tag{45}$$

where the four corners correspond to the intersections of the Phase-3/Phase-4 rows (packets) and the Group-1/non-Group-1 columns.

We now consider the "degrees of freedom" in terms of the ranks of unknown spaces. For Group-1 columns, the corresponding linear spaces $\mathcal{V}_{1\to 2\overline{3}}$ and $\mathcal{V}_{2\to \overline{13}}$ are unknown spaces from the perspective of d_1. Totally, they contribute to $\mathsf{nc}_{\mathrm{G1}}$ degrees of freedom. We then turn our focus to the non-Group-1 columns. We first notice that in our construction of Fig. 5, the columns corresponding to $\mathcal{V}_{1\to \overline{23}}$, the Δ_1 portion of $\mathcal{V}_{1\to \overline{2}3}$, the Δ_4 portion of $\mathcal{V}_{3\to 1\overline{2}}$, and the Δ_5 portion of $\mathcal{V}_{3\to \overline{12}}$ must be all-zero. Moreover, from d_1's perspective, the subspaces $\mathcal{V}_{2\to 1\overline{3}}$, $\mathcal{V}_{2\to 13}$, $\mathcal{V}_{3\to 1\overline{2}}$, and $\mathcal{V}_{3\to 12}$ are transparent since d_1 can decode these subspaces by the side information $\mathbf{Z}_{2\to 1}$ and $\mathbf{Z}_{3\to 1}$. As a result, from a decoding's perspective, among all non-Group-1 columns, only the following subspaces are unknown to d_1 and participate in the submatrices $\mathbf{A}_2$ and $\mathbf{A}_3$: $\mathcal{V}_{1\to 23}$, $\mathcal{V}_{2\to \overline{1}3}$, $\mathcal{V}_{3\to \overline{1}2}$, the Δ_2 portion of $\mathcal{V}_{1\to \overline{2}3}$, and the Δ_6 portion of $\mathcal{V}_{3\to \overline{12}}$. Totally, there are thus

$$n\mathsf{Rk}_{1\to 23} + n\mathsf{Rk}_{2\to \overline{1}3} + n\mathsf{Rk}_{3\to \overline{1}2} + \Delta_2 + \Delta_6 \tag{46}$$

degrees of freedom for the unknown subspaces corresponding to the non-Group-1 columns.

On the other hand, since we use the code alignment technique to construct the coding matrix, see (44), the contribution of the $\mathbf{v}_k^{[2\to \overline{1}3]}$ and $\mathbf{v}_k^{[3\to \overline{1}2]}$ basis vectors must satisfy

$$\begin{aligned}
&\sum_{k_1=1}^{n\mathsf{Rk}_{2\to \overline{1}3}} c_{k_1}^{[2\to \overline{1}3]}(t)\mathbf{v}_{k_1}^{[2\to \overline{1}3]}\mathbf{X}_2^{\mathrm{T}} + \sum_{k_2=1}^{n\mathsf{Rk}_{3\to \overline{1}2}} c_{k_2}^{[3\to \overline{1}2]}(t)\mathbf{v}_{k_2}^{[3\to \overline{1}2]}\mathbf{X}_3^{\mathrm{T}} \\
&= \sum_{k=1}^{n\cdot\min(\mathsf{Rk}_{2\to \overline{1}3}, \mathsf{Rk}_{3\to \overline{1}2})} c_k^{[2\to \overline{1}3]}(t)\left(\mathbf{v}_k^{[2\to \overline{1}3]}\mathbf{X}_2^{\mathrm{T}} + \mathbf{v}_k^{[3\to \overline{1}2]}\mathbf{X}_3^{\mathrm{T}}\right) \qquad (47) \\
&\quad + \sum_{k_1=n\cdot\min(\mathsf{Rk}_{2\to \overline{1}3}, \mathsf{Rk}_{3\to \overline{1}2})+1}^{n\mathsf{Rk}_{2\to \overline{1}3}} c_{k_1}^{[2\to \overline{1}3]}(t)\mathbf{v}_{k_1}^{[2\to \overline{1}3]}\mathbf{X}_2^{\mathrm{T}} \\
&\quad + \sum_{k_2=n\cdot\min(\mathsf{Rk}_{2\to \overline{1}3}, \mathsf{Rk}_{3\to \overline{1}2})+1}^{n\mathsf{Rk}_{3\to \overline{1}2}} c_{k_2}^{[3\to \overline{1}2]}(t)\mathbf{v}_{k_2}^{[3\to \overline{1}2]}\mathbf{X}_3^{\mathrm{T}}.
\end{aligned}$$

Therefore, from the perspective of decoding at d_1, the *effective degrees of freedom* of the *undesired spaces* $\mathcal{V}_{2\to \overline{1}3}$ and $\mathcal{V}_{3\to \overline{1}2}$ is reduced by the code alignment technique from the original $n\mathsf{Rk}_{2\to \overline{1}3} + n\mathsf{Rk}_{3\to \overline{1}2}$ to $n\cdot\max(\mathsf{Rk}_{2\to \overline{1}3}, \mathsf{Rk}_{3\to \overline{1}2})$. Namely, instead of decoding $\mathbf{v}_k^{[2\to \overline{1}3]}\mathbf{X}_2^{\mathrm{T}}$ and $\mathbf{v}_k^{[3\to \overline{1}2]}\mathbf{X}_3^{\mathrm{T}}$ individually, we simply need to decode the "aligned interference" $\left(\mathbf{v}_k^{[2\to \overline{1}3]}\mathbf{X}_2^{\mathrm{T}} + \mathbf{v}_k^{[3\to \overline{1}2]}\mathbf{X}_3^{\mathrm{T}}\right)$ for $k = 1, \cdots, n\cdot\min(\mathsf{Rk}_{2\to \overline{1}3}, \mathsf{Rk}_{3\to \overline{1}2})$, as expressed in (47). Based on the above reasoning, the effective degrees of freedom of all non-Group-1 columns are reduced from (46) to

$$n\mathsf{Rk}_{1\to 23} + n\cdot\max(\mathsf{Rk}_{2\to \overline{1}3}, \mathsf{Rk}_{3\to \overline{1}2}) + \Delta_2 + \Delta_6. \tag{48}$$

As a result, by the MDS property of RLNC, if we can prove

$$\mathsf{nc}_{\mathrm{G1}} \leq p_{r;1}L_3 \tag{49}$$

$$\text{and } \mathsf{nc}_{\mathrm{G1}} + (48) \leq p_{r;1}(L_3 + L_4), \tag{50}$$

where $p_{r;1}L_3$ and $p_{r;1}L_4$ are the numbers of Phase-3 and Phase-4 packets received by d_1, respectively, then we can solve the linear equations described in (45) and decode the spaces of $\mathcal{V}_{1\to 2\overline{3}}$, $\mathcal{V}_{2\to \overline{13}}$, $\mathcal{V}_{1\to 23}$, the Δ_2 portion of $\mathcal{V}_{1\to \overline{2}3}$, the Δ_6

portion of $\mathcal{V}_{3\to\overline{12}}$, and the *aligned interfering space* of $\mathcal{V}_{2\to\overline{1}3}$ and $\mathcal{V}_{3\to\overline{1}2}$. To that end, we first notice that (49) is satisfied by the construction of L_3 (see Fig. 5). Eq. (50) can be proven by substituting (35) to (38) into (48) and by (25) of Corollary 3.

We now turn our attention to the Phase-1 and Phase-2 packets received by d_1. There are totally $p_{r;1}(L_1+L_2)$ such packets. We observe that (i) The side information subspaces $\mathcal{V}_{2\to 1\overline{3}}$, $\mathcal{V}_{2\to 13}$, $\mathcal{V}_{3\to 1\overline{2}}$, and $\mathcal{V}_{3\to 12}$ are transparent to d_1; (ii) According to the construction in Fig. 5, for Phases-1 and -2 packets, the columns corresponding to $\mathcal{V}_{1\to 2\overline{3}}$, $\mathcal{V}_{2\to\overline{13}}$, and $\mathcal{V}_{2\to 1\overline{3}}$ must be all-zero; (iii) The following subspaces have been decoded by d_1 using the Phases-3 and -4 packets and thus are transparent to d_1 as well, including $\mathcal{V}_{1\to 23}$, the code aligned space $(\mathcal{V}_{2\to\overline{1}3}, \mathcal{V}_{3\to\overline{1}2})$, the Δ_2 portion of $\mathcal{V}_{1\to\overline{2}3}$, and the Δ_6 portion of $\mathcal{V}_{3\to\overline{12}}$. By omitting the above three types of columns that are transparent to d_1, the remaining coding matrix for the Phases-1 and -2 packets can be expressed as

$$\left(\begin{array}{c|c|c} \mathbf{A}_4 & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_5 & \mathbf{A}_6 \end{array}\right), \tag{51}$$

where $\mathbf{A}_4$ corresponds to the intersection of the Phase-1 packets and the columns of $\mathcal{V}_{1\to\overline{23}}$, $\mathbf{A}_5$ corresponds to the intersection of the Phase-2 packets and the columns of the Δ_1 portion of $\mathcal{V}_{1\to\overline{2}3}$, and $\mathbf{A}_6$ corresponds to the intersection of the Phase-2 packets and the columns of the Δ_5 portion of $\mathcal{V}_{3\to\overline{12}}$. Note that all the entries of $\mathbf{A}_4$ to $\mathbf{A}_6$ are chosen independently and uniformly randomly from $\mathsf{GF}(q)$. By the MDS property of RLNC, if we can prove

$$n \cdot \mathsf{Rk}_{1\to\overline{23}} \leq p_{r;1} L_1 \tag{52}$$

$$\text{and } \Delta_1 + \Delta_5 \leq p_{r;1} L_2, \tag{53}$$

then we can solve the linear equations described in (51) and decode the spaces of $\mathcal{V}_{1\to\overline{23}}$, the Δ_1 portion of $\mathcal{V}_{1\to\overline{2}3}$, and the Δ_5 portion of $\mathcal{V}_{3\to\overline{12}}$. We then notice that (52) and (53) are satisfied by the construction of L_1 and L_2, respectively, see Fig. 5.

Based on the above discussion, d_1 can decode all its desired information subspaces.

The case of d_2: Decoding at d_2 only uses the Phases-3 and -4 packets received by d_2, which form a $(p_{r;2}(L_3+L_4)) \times \left(n\sum_{i=1}^{3} R_i\right)$ matrix. We rearrange the columns so that the columns corresponding to $\mathcal{V}_{2\to\overline{13}}$ and $\mathcal{V}_{2\to 1\overline{3}}$ are the first $n\left(\mathsf{Rk}_{2\to\overline{13}} + \mathsf{Rk}_{2\to 1\overline{3}}\right)$ columns. We call these columns the Group-2 columns and use $\mathsf{nc}_{\mathrm{G2}} \triangleq n\left(\mathsf{Rk}_{2\to\overline{13}} + \mathsf{Rk}_{2\to 1\overline{3}}\right)$ to denote the number of Group-2 columns. We also rearrange the rows so that the Phase-3 packets correspond to the first $p_{r;2}L_3$ rows while the Phase-4 packets correspond to the last $p_{r;2}L_4$ rows. After rearrangement, the coding matrix becomes

$$\left(\begin{array}{c|c} \mathbf{A}_7 & \mathbf{A}_8 \\ \hline \mathbf{0} & \mathbf{A}_9 \end{array}\right), \tag{54}$$

where the four corners correspond to the intersections of the Phase-3/Phase-4 rows (packets) and the Group-2/non-Group-2 columns.

Similar to the case of d_1, we consider the degrees of freedom of unknown spaces. For Group-2 columns, the corresponding linear spaces $\mathcal{V}_{2\to\overline{13}}$ and $\mathcal{V}_{2\to 1\overline{3}}$ contribute to $\mathsf{nc}_{\mathrm{G2}}$ degrees of freedom. We then turn our focus to the non-Group-2 columns. Again we can exclude from our consideration the columns that do not participate in Phases 3 and 4 and those *transparent* columns corresponding to the overheard side information $\mathbf{Z}_{1\to 2}$ and $\mathbf{Z}_{3\to 2}$. As a result, from a decoding's perspective, among all non-Group-2 columns, only the following subspaces are unknown to d_2 and participate in the submatrices $\mathbf{A}_8$ and $\mathbf{A}_9$: $\mathcal{V}_{2\to\overline{1}3}$, $\mathcal{V}_{2\to 13}$, the Δ_2 portion of $\mathcal{V}_{1\to\overline{2}3}$, the Δ_3 portion of $\mathcal{V}_{3\to 1\overline{2}}$, and the Δ_6 portion of $\mathcal{V}_{3\to\overline{12}}$. Totally, there are thus

$$n\mathsf{Rk}_{2\to\overline{1}3} + n\mathsf{Rk}_{2\to 13} + \Delta_2 + \Delta_3 + \Delta_6 \tag{55}$$

degrees of freedom for the subspaces corresponding to the non-Group-2 columns.

Since we use the code alignment technique to construct the coding matrix, see (43), the last Δ_2 basis vectors of $\mathcal{V}_{1\to\overline{2}3}$: $\mathbf{v}^{[1\to\overline{2}3]}_{k+\Delta_1}$ and the first Δ_3 basis vectors of $\mathcal{V}_{3\to 1\overline{2}}$: $\mathbf{v}^{[3\to 1\overline{2}]}_{k}$ are aligned. The effective degrees of freedom of all non-Group-2 columns are thus reduced from (55) to

$$n\mathsf{Rk}_{2\to\overline{1}3} + n\mathsf{Rk}_{2\to 13} + \max(\Delta_2, \Delta_3) + \Delta_6. \tag{56}$$

As a result, by the MDS property of RLNC, if we can prove

$$\mathsf{nc}_{\mathrm{G2}} \leq p_{r;2} L_3 \tag{57}$$

$$\text{and } \mathsf{nc}_{\mathrm{G2}} + (56) \leq p_{r;2}(L_3 + L_4), \tag{58}$$

then we can solve the linear equations described in (54) and d_2 can successfully decode its desired subspaces: $\mathcal{V}_{2\to\overline{13}}$, $\mathcal{V}_{2\to 1\overline{3}}$, $\mathcal{V}_{2\to\overline{1}3}$ and $\mathcal{V}_{2\to 13}$, plus the interfering subspaces: the code-aligned subspace $(\Delta_2 \text{ of } \mathcal{V}_{1\to\overline{2}3}, \Delta_3 \text{ of } \mathcal{V}_{3\to 1\overline{2}})$, and the Δ_6 portion of $\mathcal{V}_{3\to\overline{12}}$.

To that end, we first notice that (57) is satisfied by the construction of L_3 (see Fig. 5). A detailed proof of (58) is provided in Appendix E, which uses the assumption that (26) is satisfied.

Based on the above discussion, d_2 can decode all its desired information subspaces.

The case of d_3: Decoding at d_3 only uses the Phase-2 and Phase-4 packets received by d_3, which form a $(p_{r;3}(L_2+L_4)) \times \left(n\sum_{i=1}^{3} R_i\right)$ matrix. We rearrange the columns so that the columns corresponding to the last Δ_4 columns of $\mathcal{V}_{3\to 1\overline{2}}$ and the first Δ_5 columns of $\mathcal{V}_{3\to\overline{12}}$ are moved to the first $\Delta_4 + \Delta_5$ columns. We call these columns the Group-3 columns and use $\mathsf{nc}_{\mathrm{G3}} \triangleq \Delta_4 + \Delta_5$ to denote the number of Group-3 columns. We also rearrange the rows so that the Phase-2 packets correspond to the first $p_{r;3}L_2$ rows while the Phase-4 packets correspond to the last $p_{r;3}L_4$ rows. After rearrangement, the coding matrix becomes

$$\left(\begin{array}{c|c} \mathbf{A}_{10} & \mathbf{A}_{11} \\ \hline \mathbf{0} & \mathbf{A}_{12} \end{array}\right), \tag{59}$$

where the four corners correspond to the intersections of the Phase-2/Phase-4 rows (packets) and the Group-3/non-Group-3 columns.

Similar to the cases of d_1 and d_2, we consider the degrees of freedom in terms of the ranks of unknown spaces. Specifically, Group-3 columns contribute to $\mathsf{nc}_{\mathrm{G3}}$ degrees of freedom. We then turn our focus to the non-Group-3 columns. By similar arguments as used in the cases of d_1 and d_2, among all non-Group-3 columns, only the following subspaces are unknown to d_3 and participate in the submatrices $\mathbf{A}_{11}$ and $\mathbf{A}_{12}$: $\mathcal{V}_{3\to\overline{1}2}$, $\mathcal{V}_{3\to 12}$, the Δ_3 portion of $\mathcal{V}_{3\to 1\overline{2}}$, and the Δ_6 portion of $\mathcal{V}_{3\to\overline{12}}$. Totally, there are thus

$$n\mathsf{Rk}_{3\to\overline{1}2} + n\mathsf{Rk}_{3\to 12} + \Delta_3 + \Delta_6 \tag{60}$$

degrees of freedom for the subspaces corresponding to the non-Group-3 columns.

As a result, by the MDS property of RLNC, if we can prove

$$\mathsf{nc}_{\mathrm{G3}} \le p_{r;3}L_2 \tag{61}$$

$$\text{and } \mathsf{nc}_{\mathrm{G3}} + (60) \le p_{r;3}(L_2+L_4), \tag{62}$$

then we can solve the linear equations described in (59) and d_3 can successfully decode its desired subspaces: $\mathcal{V}_{3\to 1\overline{2}}$, $\mathcal{V}_{3\to\overline{12}}$, $\mathcal{V}_{3\to\overline{1}2}$, and $\mathcal{V}_{3\to 12}$.

To that end, we first notice that Lemma 4 implies (61). Eq. (62) can be proven by substituting (35) to (38) into (60) and by (27) of Corollary 3. Based on the above discussion, d_3 can decode all its desired information subspaces.

The proof of Proposition 3 is thus complete.

VI. The Achievability Results For Arbitrary M Values

The capacity-achieving network code in the previous section is deliberately designed for the $M=3$ case. Such an optimal design is difficult to generalize for the cases of $M\ge 4$. In this section, we provide a slightly suboptimal but systematic code design that attains the capacity inner bound of Proposition 4.

A. High-Level Description of The Achievability Scheme

In this subsection, we first present the skeleton of our code construction while the detailed description of a central step is relegated to Section VI-B.

Before proceeding, we first convert Proposition 4 to an equivalent LP problem by the alternative formulation of a ZPPLCIF in Lemma 1.

Proposition 6: Following the same settings and definitions of Proposition 4, a rate vector $(R_1,\cdots,R_M)$ can be achieved by a *linear network code* if there exist $M(M+1)!$ non-negative variables $\{x_{\pi,S^\pi_j,h} : \forall\pi, \forall j\in[M+1], \forall h\in[M]\}$ and $M!(2^{M+1}-M-2)$ non-negative variables:

$$\Big\{y_{k,T,(\pi,S^\pi_j)} \ge 0 : \forall k\in[M], \forall T\in 2^{[M]}, \forall\pi, \forall j\in[M+1], \\ \text{satisfying } k\notin S^\pi_j \text{ and } T\subseteq S^\pi_j\Big\},$$

and $M2^{M-1}$ non-negative variables:

$$\Big\{z_{k,T}\ge 0 : \forall T\in 2^{[M]}, \forall k\in[M]\backslash T\Big\},$$

such that jointly the rate R_i and the x, y, z variables satisfy (12), (14), (15), (16), and the following inequalities:

$$\forall\pi, \forall i,j,k\in[M], \\ \sum_{h=1}^{M} x_{\pi,S^\pi_{j+1},h}\cdot\min(p_{r;i},p_{r;h}) \ge \sum_{h=1}^{M} x_{\pi,S^\pi_j,h}\cdot\min(p_{r;i},p_{r;h}) \\ + 1_{\{i\le k, k\notin S^\pi_{j+1}\}} \sum_{\{\forall T\in 2^{[M]}: T\subseteq S^\pi_{j+1}\}} y_{k,T,(\pi,S^\pi_{j+1})} \tag{63}$$

$$\sum_{\forall\pi}\sum_{h=1}^{M} x_{\pi,\emptyset,h}\cdot p_{r;M} < p_{r;M} \Longleftrightarrow \sum_{\forall\pi}\sum_{h=1}^{M} x_{\pi,\emptyset,h} < 1. \tag{64}$$

Our construction scheme is based on the above alternative formulation of the inner bound.

For any rate vector $(R_1,\cdots,R_M)$ satisfies the inner bound of Proposition 6, perform the following four steps.

Step 1: For each time slot $t\in[n]$, each source s_i, $i\in[M]$, sends out the coded symbol $W_i(t)$ by RLNC, i.e.,

$$W_i(t) = \sum_{k=1}^{nR_i} c_{i,k}(t) X_{i,k},$$

for which we choose the mixing coefficients $c_{i,k}(t)$ independently and uniformly randomly from $\mathsf{GF}(q)$. $\mathbf{c}_i(t) \triangleq (c_{i,1}(t),\cdots,c_{i,nR_i}(t))$ denotes the corresponding (nR_i)-dimensional coding vector used in time t.

After Step 1, relay r receives $np_{s_i;r}$ coded packets, which is larger than nR_i, the number of information packets. By the MDS property of RLNC, r can successfully decode $X_{i,1}$ to X_{i,nR_i} for all $i\in[M]$.

Step 2: Each destination d_j, $j\in[M]$, sends to r its own overhearing channel status vector $\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}}$ for $i\neq j$. Based on $\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}}$, the relay r derives the following linear subspaces $\mathcal{V}_{i;j}$ for all $i,j\in[M]$, $i\neq j$:

$$\mathcal{V}_{i;j} \triangleq \mathsf{span}\left(\{\mathbf{c}_i(t) : \forall t \text{ s.t. } Z_{i\to j}(t)\neq *\}\right)$$

That is, $\mathcal{V}_{i;j}$ is the linear subspace spanned by all the coding vectors of the user-i packets that are overheard by d_j. Based on $\{\mathcal{V}_{i;j} : \forall i\neq j\}$, the relay r further derives multiple subspaces $\mathcal{V}_{i\to T_1\overline{T_2}}$ for all $i\in[M]$, $T_1\subseteq[M]\backslash i$, and $T_2=[M]\backslash(T_1\cup\{i\})$ such that the following three properties are satisfied:

(i) $\forall j\in T_1, \quad \mathcal{V}_{i\to T_1\overline{T_2}} \subseteq \mathcal{V}_{i;j}$,

(ii) $\mathsf{span}\{\mathcal{V}_{i\to T_1\overline{T_2}} : \forall T_1\subseteq[M]\backslash i, T_2=[M]\backslash(T_1\cup\{i\})\} = \Omega_i$,

(iii) $\dfrac{\mathrm{Rank}(\mathcal{V}_{i\to T_1\overline{T_2}})}{n} = z_{i,T_2}$.

Intuitively, the message space Ω_i of user-i is partitioned into multiple subspaces $\mathcal{V}_{i\to T_1\overline{T_2}}$ such that each subspace $\mathcal{V}_{i\to T_1\overline{T_2}}$ is *known* to all the users $j\in T_1$, and the normalized rank of $\mathcal{V}_{i\to T_1\overline{T_2}}$ equals to the feasible solution z_{i,T_2} satisfying (15) and (16) of Propositions 4 and 6. Eqs. (15) and (16) can thus be viewed as the *space partitioning* inequalities.

Lemma 5: Construction of $\{\mathcal{V}_{i\to T_1\overline{T_2}}\}$ is always possible.

A proof of Lemma 5 is provided in Appendix F.

With the above construction, let $\mathbf{v}_k^{[i\to T_1\overline{T_2}]}$, $k=1\cdots,n\cdot z_{i,T_2}$ denote the basis vectors of $\mathcal{V}_{i\to T_1\overline{T_2}}$.

Step 3: We then partition the n time slots into $M!$ subgroups. Each group, indexed by a permutation π, contains $n\cdot\left(\sum_{h=1}^{M} x_{\pi,\emptyset,h}\right)$ time slots. Inequality (64) ensures that such time slot partition is always possible. We will perform network coding for each subgroup separately.

Note that partitioning in time is equivalent to partitioning the rows of the coding matrix. We now turn our focus to partitioning the basis vectors (the columns) $\mathbf{v}_k^{[i\to T_1\overline{T_2}]}$, $k=1\cdots,n\cdot z_{i,T_2}$. To that end, we rely on new z' variables such that

$$\forall T_2 \text{ satisfying } i\notin T_2,\ \sum_{\forall\pi} z'_{i,T_2,\pi}=z_{i,T_2}. \tag{65}$$

On the other hand, any vector $\mathbf{v}\in\mathcal{V}_{i\to T_1\overline{T_2}}$ can also be encoded as if it has only been overheard by users in a subset $S_1\subseteq T_1$, i.e., we forgo the fact that $\mathbf{v}$ is also known to all $j\in T_1\backslash S_1$. As a result, by defining $S_2=[M]\backslash(S_1\cup\{i\})$, (65) can be revised as follows.

$$\forall T_2 \text{ satisfying } i\notin T_2,\quad \sum_{\forall\pi,\forall S_2\supseteq T_2, i\notin S_2} y'_{i,T_2,(\pi,S_2)}=z_{i,T_2}, \tag{66}$$

where $y'_{i,T_2,(\pi,S_2)}$ takes into account the redistribution of $\mathbf{v}_k^{[i\to T_1\overline{T_2}]}$ vectors into partitions indexed by (π,S_2).

To facilitate the actual encoding operation in the subsequent Step 4, we do not allow full degrees of freedom when redistributing the $\mathbf{v}_k^{[i\to T_1\overline{T_2}]}$ vectors. More explicitly, we only allows $y'_{i,T_2,(\pi,S_2)}$ to be non-zero when S_2 is of the form $S_2=S_j^\pi$ for some $j\in[M+1]$ and $i\notin S_j^\pi$. Combining this restriction with (66), the final form of the redistribution equation becomes

$$\forall T_2 \text{ satisfying } i\notin T_2,\quad \sum_{\{\forall\pi,\forall j\in[M+1]: S_j^\pi\supseteq T_2, i\notin S_j^\pi\}} y_{i,T_2,(\pi,S_j^\pi)}=z_{i,T_2}, \tag{67}$$

which is a restatement[6] of (14). Eq. (14) is thus a *subspace redistribution* inequality.

Step 4: Following Step 3, for any permutation π, the goal of Step 4 is to encode all the partitioned basis vectors, totally

$$\sum_{\{\forall i\in[M], j\in[M+1], T\in 2^{[M]}: i\notin S_j^\pi, T\subseteq S_j^\pi\}} n\cdot y_{i,T,(\pi,S_j^\pi)}$$

of them, using the allocated time slots $n\cdot\left(\sum_{h=1}^{M} x_{\pi,\emptyset,h}\right)$.

Step 4 is the central step of the construction, which will be elaborated in Section VI-B. We conclude this subsection by introducing the concept of "chunks of time slots" such that for each π, we can partition the group of totally $\sum_{h=1}^{M} x_{\pi,\emptyset,h}$ time slots into multiple chunks of time slots. Each chunk has $n\cdot\chi_{\pi,l}$ time slots for a distinct l. The detailed properties of the "chunks" are described by the following lemma and corollary.

Lemma 6: Without loss of generality, for any x, y, and z values satisfying (63), (64), (12), (14), (15), and (16) of Proposition 6, we can assume that

- $\forall\pi,\forall h\in[M]$, $x_{\pi,[M],h}=0$.
- $\forall\pi,\forall k\in[M],\forall j\in[M]$,

$$\sum_{h=1}^{k} x_{\pi,S_{j+1}^\pi,h}\geq\sum_{h=1}^{k} x_{\pi,S_j^\pi,h}. \tag{68}$$

The following is a corollary of this lemma.

Corollary 4: Following Lemma 6, we can further assume that there exists a finite set of (variable, function) pairs $\{(\chi_{\pi,l},\theta_{\pi,l}(\cdot)):\forall\pi,\forall l\}$, such that each variable $\chi_{\pi,l}$ is in the range $[0,1]$, each function $\theta_{\pi,l}(\cdot)$ is a non-increasing map from $[M]$ to $[M+1]$, and jointly they satisfy (i) $\theta_{\pi,l}(M)\leq M$, and (ii)

$$\forall\pi,\forall j,h\in[M],\ x_{\pi,S_{j+1}^\pi,h}=\sum_{\forall l:\theta_{\pi,l}(j)=h}\chi_{\pi,l}. \tag{69}$$

We term $\chi_{\pi,l}$ and $\theta_{\pi,l}(\cdot)$ as the (normalized) size and the label function of the l-th chunk.

The proofs of Lemma 6 and Corollary 4 are relegated to Appendix G.

Remark 1: We first note that since the label function satisfies $\theta_{\pi,l}(M)\leq M$, we must have

$$\begin{aligned}&\{\chi_{\pi,l}:\forall\pi,\forall l\}\\&=\{\chi_{\pi,l}:\forall\pi,\forall h\in[M],\forall l,\ \text{s.t. } \theta_{\pi,l}(M)=h\}.\end{aligned}$$

By choosing $j=M$ in (69), we have

$$\forall\pi,\forall h\in[M],\ x_{\pi,\emptyset,h}=\sum_{\forall l\in[L]:\theta_{\pi,l}(M)=h}\chi_{\pi,l}.$$

As a result, the summation of all $\chi_{\pi,l}$ equals to $\left(\sum_{h=1}^{M} x_{\pi,\emptyset,h}\right)$. Therefore, the term $\chi_{\pi,l}$ can indeed be viewed as the normalized size of the l-th chunk.

Remark 2: One special property of a chunk is the associated label function $\theta_{\pi,l}(\cdot)$ that is non-increasing. Such a property is critical to our code construction.

B. The Central Step (Step 4) of The Achievability Scheme

In this subsection, we concentrate on how to perform network coding over the partition $\{\chi_{\pi,l}:\forall\pi,\forall l\}$. For the l-th chunk, we will design the corresponding $(n\chi_{\pi,l})\times\left(n\sum_{i=1}^{M} R_i\right)$ coding matrix $\mathbf{A}_{\pi,l}$. After finishing the creation for each chunk, the overall coding matrix is an $n\times\left(n\sum_{i=1}^{M} R_i\right)$ matrix that vertically concatenates all the coding matrices of the partition $\{\chi_{\pi,l}:\forall\pi,\forall l\}$.

More explicitly, we define

$$\tilde{y}_{i,(\pi,S_j^\pi)}\triangleq\sum_{\forall T\in 2^{[M]}:T\subseteq S_j^\pi} y_{i,T,(\pi,S_j^\pi)}$$

for any given permutation π, $i\in[M]$, and S_j^π satisfying $i\notin S_j^\pi$. Namely, $n\cdot\tilde{y}_{i,(\pi,S_j^\pi)}$ is the number of basis vectors $\mathbf{v}$ that have been reassigned to a special group indexed by (π,S_j^π),

[6] For any y values that satisfy (14) with strict inequality, we can repeatedly reduce the values of any positive y variables until (14) becomes an equality. Note that the newly reduced y variables will still satisfy (13) of Proposition 4 (resp. (63) of Proposition 6). As a result, without loss of generality, we can replace inequality (14) by an equality (67).

through (67). If we define $T_j^\pi \triangleq [M]\backslash(S_j^\pi \cup \{i\})$, and use $\mathbf{v}_k^{\pi,[i\to T_j^\pi \overline{S_j^\pi}]}$, $k = 1, \cdots, n\cdot\tilde{y}_{i,(\pi,S_j^\pi)}$ to denote those $\mathbf{v}$ vectors, then by our discussion in the previous subsection,

$$\left\{\mathbf{v}_k^{\pi,[i\to T_j^\pi \overline{S_j^\pi}]} : \forall\pi, \forall j, \forall k\right\}$$

form the basis vectors of the entire message space Ω_i. Using the new basis vectors, for each time slot t, the relay r transmits a coded packet $W_r(t)$ by

$$W_r(t) = \sum_{i=1}^{M} \left(\sum_{\forall\pi,\forall j\in[M+1], i\notin S_j^\pi} \sum_{k=1}^{n\tilde{y}_{i,(\pi,S_j^\pi)}} c_k^{\pi,[i\to T_j^\pi \overline{S_j^\pi}]}(t)\mathbf{v}_k^{\pi,[i\to T_j^\pi \overline{S_j^\pi}]} \right) \mathbf{X}_i^{\mathrm{T}},$$

Our task is to design the coefficients $c_k^{\pi,[i\to T_j^\pi \overline{S_j^\pi}]}(t)$ for all $t \in [n]$. Similar to the discussion in Sections V-A and V-B, we can express the coefficients $c_k^{\pi,[i\to T_j^\pi \overline{S_j^\pi}]}(t)$ by the equivalent $n \times \left(n\sum_{i=1}^M R_i\right)$ *coding matrix.* Our systematic construction then focuses on the individual $(n\chi_{\pi,l}) \times \left(n\sum_{i=1}^M R_i\right)$ coding sub-matrices $\mathbf{A}_{\pi,l}$. For future reference, we also use $\mathcal{V}^\pi_{i\to T_j^\pi \overline{S_j^\pi}}$ to denote the space spanned from $\left\{\mathbf{v}_k^{\pi,[i\to T_j^\pi \overline{S_j^\pi}]}\right\}$.

Our construction will rely on the code alignment technique first introduced in Section V-B. For each chunk $(\chi_{\pi,l}, \theta_{\pi,l}(\cdot))$, we create an $(n\chi_{\pi,l}) \times (n\chi_{\pi,l})$ matrix $\mathbf{B}_{\pi,l}$ by choosing each entry independently and uniformly randomly from $\mathsf{GF}(q)$. The $\mathbf{B}_{\pi,l}$ matrix will later be used to create the $(n\chi_{\pi,l}) \times \left(n\sum_{i=1}^M R_i\right)$ coding matrix $\mathbf{A}_{\pi,l}$. For each permutation π, the following subroutine will construct all $\{\mathbf{A}_{\pi,l} : \forall l\}$ simultaneously.

1: **Initialization:** All entries of $\mathbf{A}_{\pi,l}$ are set to zero for all l.
2: **for** $j_0 = 1, \cdots, M$ **do**
3:   **for** each i_0 that is not in $S^\pi_{j_0+1}$ **do**
4:     **set** $\mathsf{cnt}_{\text{assigned}} \leftarrow 0$.
5:     **set** the status of all chunks $(\chi_{\pi,l}, \theta_{\pi,l}(\cdot))$ to "not-processed."
6:     **while** $\mathsf{cnt}_{\text{assigned}} < n\cdot\tilde{y}_{i_0,(\pi,S^\pi_{j_0+1})}$ **do**
7:       Arbitrarily choose one "not-processed" chunk and denote it by $(\chi_{\pi,l_0}, \theta_{\pi,l_0}(\cdot))$ for some l_0.
8:       **set**[7] $\Delta_{i_0,\text{curr}} \leftarrow n\cdot\chi_{\pi,l_0}\cdot\min(p_{r;i_0}, p_{r;\theta_{\pi,l_0}(j_0)})$.
9:       **if** $j_0 > 1$ **then**
10:         **set** $\Delta_{i_0,\text{prev}} \leftarrow n\cdot\chi_{\pi,l_0}\cdot\min(p_{r;i_0}, p_{r;\theta_{\pi,l_0}(j_0-1)})$.
11:       **else**
12:         **set** $\Delta_{i_0,\text{prev}} \leftarrow 0$.
13:       **end if**
14:       **set** $\Delta_{i_0,\text{extr}} \leftarrow (\Delta_{i_0,\text{curr}} - \Delta_{i_0,\text{prev}})$.
15:       **set** $\mathsf{cnt}_{\text{new}} \leftarrow \min(\mathsf{cnt}_{\text{assigned}} + \Delta_{i_0,\text{extr}}, n\tilde{y}_{i_0,(\pi,S^\pi_{j_0+1})})$.
16:       **set** $\Delta \leftarrow \mathsf{cnt}_{\text{new}} - \mathsf{cnt}_{\text{assigned}}$.
17:       Consider the $(\Delta_{i_0,\text{prev}}+1)$-th to the $(\Delta_{i_0,\text{prev}}+\Delta)$-th columns of matrix $\mathbf{B}_{\pi,l_0}$. Copy that portion of $\mathbf{B}_{\pi,l_0}$ to the columns of $\mathbf{A}_{\pi,l_0}$ corresponding to $\mathbf{v}_k^{\pi,[i_0\to T^\pi_{j_0+1}\overline{S^\pi_{j_0+1}}]}$, for all $k \in (\mathsf{cnt}_{\text{assigned}}, \mathsf{cnt}_{\text{new}}]$.
18:       **set** the status of the chunk $(\chi_{\pi,l_0}, \theta_{\pi,l_0}(\cdot))$ to "processed."
19:       **set** $\mathsf{cnt}_{\text{assigned}} \leftarrow \mathsf{cnt}_{\text{new}}$.
20:     **end while**
21:   **end for**
22: **end for**

Intuitively, during the FOR loop of the given j_0, the above sub-routine "assigns" the $n\cdot\tilde{y}_{i_0,(\pi,S^\pi_{j_0+1})}$ columns corresponding to vectors $\mathbf{v}_k^{\pi,[i_0\to T^\pi_{j_0+1}\overline{S^\pi_{j_0+1}}]}$ to different chunks $(\chi_{\pi,l}, \theta_{\pi,l}(\cdot))$ for all $i_0 \notin S^\pi_{j_0+1}$. Although the assignment, Lines 3 to 21, is made separately for distinct $i_0 \notin S^\pi_{j_0+1}$, *a common* $\mathbf{B}_{\pi,l}$ *is used in Line 17*, which results in "aligned codeword spaces" that is necessary for approaching the optimal throughput. Lines 4 and 5 initialize the FOR loop. Line 6 stops the inner loop whenever all columns have been assigned, i.e., when $\mathsf{cnt}_{\text{assigned}} = n\cdot\tilde{y}_{i_0,(\pi,S^\pi_{j_0+1})}$.

During the inner loop, Lines 6 to 20, each *unprocessed* chunk $(\chi_{\pi,l_0}, \theta_{\pi,l_0}(\cdot))$ has already been assigned $\Delta_{i_0,\text{prev}}$ columns during the previous j-th steps for those $j < j_0$. On the other hand, for the j_0-th step, each chunk can still carry $\Delta_{i_0,\text{extr}}$ extra columns, which we term the "extra rank capacity" and is computed by Lines 8, 10, and 14 of the sub-routine. It is worth noting that the computation of the extra rank capacity in Lines 8, 10, and 14 depends on i_0, the user of interest. The reason is that each user i_0 is facing different success probability $p_{r;i_0}$, which affects how many columns a chunk can carry for each user i_0.

Now we need to prove that after processing all chunks, we can successfully assign all $n\cdot\tilde{y}_{i_0,(\pi,S^\pi_{j_0+1})}$ columns. To that end, we notice that when $j_0 = 1$, by Corollary 4 with $j = 1$, the first part of Lemma 6, and by (63) with $j = 1$, we have

$$\begin{aligned}
&\sum_{\forall l} \left(\chi_{\pi,l}\cdot\left(\min\left(p_{r;i}, p_{r;\theta_{\pi,l}(1)}\right) - 0\right)\right) && (70)\\
&= \sum_{h=1}^{M} \left(\sum_{\forall l:\theta_{\pi,l}(1)=h} \chi_{\pi,l}\right)\cdot\min(p_{r;i}, p_{r;h}) \\
&= \sum_{h=1}^{M} x_{\pi,S_2^\pi,h}\cdot\min(p_{r;i}, p_{r;h}) \\
&\geq 1_{\{i\leq i_0\}}\cdot\tilde{y}_{i_0,(\pi,S_2^\pi)} && (71)
\end{aligned}$$

for all $i_0 \notin S_2^\pi$. Similarly, for $j_0 \geq 2$, by Corollary 4 and by

[7] We use the convention that $p_{r;M+1} = 0$.

(63) with $j = j_0$, we have

$$\sum_{\forall l} \left(\chi_{\pi,l} \cdot \left(\min\left(p_{r;i}, p_{r;\theta_{\pi,l}(j_0)}\right) - \min\left(p_{r;i}, p_{r;\theta_{\pi,l}(j_0-1)}\right)\right)\right) \tag{72}$$

$$= \sum_{h=1}^{M} \left(\sum_{\forall l:\theta_{\pi,l}(j_0)=h} \chi_{\pi,l} \right) \cdot \min(p_{r;i}, p_{r;h}) - \sum_{h=1}^{M} \left(\sum_{\forall l:\theta_{\pi,l}(j_0-1)=h} \chi_{\pi,l} \right) \cdot \min(p_{r;i}, p_{r;h})$$

$$= \sum_{h=1}^{M} \left(x_{\pi,S^\pi_{j_0+1},h} - x_{\pi,S^\pi_{j_0},h} \right) \cdot \min(p_{r;i}, p_{r;h})$$

$$\geq 1_{\{i \leq i_0\}} \cdot \tilde{y}_{i_0,(\pi,S^\pi_{j_0+1})} \tag{73}$$

for all $i_0 \notin S^\pi_{j_0+1}$. We then notice that by choosing $i = i_0$, each summand of (70) and (72) is the normalized extra rank capacity, $\frac{\Delta_{i_0,\text{extr}}}{n}$, of a given j_0 value (see Line 14). (71) and (73) then guarantee that the summation of the extra rank capacity $\Delta_{i_0,\text{extr}}$ of all chunks is no less than the total number of columns of interest $n \cdot \tilde{y}_{i_0,(\pi,S^\pi_{j_0+1})}$. As a result, collectively all the chunks can carry all the columns of interest. The above sub-routine thus terminates properly.

Fig. 6 illustrates the above code construction for a simple example of $M = 3$, $\pi = (2, 3, 1)$. Suppose there are only four chunks $(\chi_{\pi,1}, \theta_{\pi,1}(\cdot))$ to $(\chi_{\pi,4}, \theta_{\pi,4}(\cdot))$ that are associated with the permutation $\pi = (2, 3, 1)$ of interest. For the first chunk $l = 1$, the corresponding label function $\theta_{\pi,1}(\cdot)$ is assumed to be

$$\theta_{\pi,1}(j) = \begin{cases} 4 & \text{if } j = 1 \\ 3 & \text{if } j = 2 \\ 1 & \text{if } j = 3 \end{cases}.$$

For simplicity, we use $\theta_{\pi,1}(1, 2, 3) = (4, 3, 1)$ as shorthand. For the second to the fourth chunks, the corresponding label functions $\theta_{\pi,l}(\cdot)$ are assumed to be $\theta_{\pi,2}(1, 2, 3) = (2, 2, 1)$, $\theta_{\pi,3}(1, 2, 3) = (3, 2, 1)$, and $\theta_{\pi,4}(1, 2, 3) = (4, 4, 2)$ (see Fig. 6). The subsroutine will design the corresponding sub-matrices $\mathbf{A}_{\pi,1}$ to $\mathbf{A}_{\pi,4}$. For illustration purposes, in Fig. 6 we only plot the $\widetilde{\mathcal{V}}^\pi_{i_0 \to T^\pi_j \overline{S^\pi_j}}$ columns that participate in these four chunks of the given π. In our example, these columns are the ones corresponding to $\widetilde{\mathcal{V}}^\pi_{2\to\overline{13}}$, $\widetilde{\mathcal{V}}^\pi_{2\to\overline{1}3}$, $\widetilde{\mathcal{V}}^\pi_{3\to\overline{1}2}$, $\widetilde{\mathcal{V}}^\pi_{1\to23}$, $\widetilde{\mathcal{V}}^\pi_{2\to13}$, and $\widetilde{\mathcal{V}}^\pi_{3\to12}$.

Before executing the subroutine, we construct independently and uniformly randomly the $(n\chi_{\pi,l}) \times (n\chi_{\pi,l})$ square matrix $\mathbf{B}_{\pi,l}$ for $l = 1$ to 4. Consider $j_0 = 1$ when executing the subroutine. The only i_0 not in $S^\pi_2 = \{1, 3\}$ is $i_0 = 2$. With the specified label functions, the extra rank capacity $\Delta_{2,\text{extr}}$ that can be used for carrying $\widetilde{\mathcal{V}}^\pi_{2\to\overline{13}}$ columns is computed by Line 14:

$$\begin{aligned} &n\chi_{\pi,1} \cdot \min(p_{r;2}, p_{r;4}) = 0, \\ &n\chi_{\pi,2} \cdot \min(p_{r;2}, p_{r;2}) = n\chi_{\pi,2} \cdot p_{r;2}, \\ &n\chi_{\pi,3} \cdot \min(p_{r;2}, p_{r;3}) = n\chi_{\pi,3} \cdot p_{r;3}, \\ &n\chi_{\pi,4} \cdot \min(p_{r;2}, p_{r;4}) = 0, \end{aligned}$$

for the four chunks, respectively. The subroutine for $j_0 = 1$ then copies the first $n\chi_{\pi,2} \cdot p_{r;2}$ columns and the first $n\chi_{\pi,3} \cdot p_{r;3}$ columns of $\mathbf{B}_{\pi,2}$ and $\mathbf{B}_{\pi,3}$ to their corresponding columns of $\widetilde{\mathcal{V}}^\pi_{2\to\overline{13}}$, respectively. See Fig. 6 for illustration. For simplicity of the figure, we actually treat the first $n\chi_{\pi,2} \cdot p_{r;3}$ columns of $\mathbf{B}_{\pi,2}$ and the next $n\chi_{\pi,2} \cdot (p_{r;2} - p_{r;3})$ columns of $\mathbf{B}_{\pi,2}$ as two separate groups of columns. Therefore the subroutine copies two groups of columns of $\mathbf{B}_{\pi,2}$ to their corresponding positions in $\mathbf{A}_{\pi,2}$. In Fig. 6, we use the dashed lines to represent the "copy" operations.

For $j_0 = 2$, since $S^\pi_{2+1} = \{1\}$ we consider $i_0 = 2$ and 3, and the corresponding columns of $\widetilde{\mathcal{V}}^\pi_{2\to\overline{1}3}$ and $\widetilde{\mathcal{V}}^\pi_{3\to\overline{1}2}$ separately. First consider $i_0 = 2$ and the corresponding $\widetilde{\mathcal{V}}^\pi_{2\to\overline{1}3}$. With the specified label functions, the extra rank capacities that can be used for carrying $\widetilde{\mathcal{V}}^\pi_{2\to\overline{1}3}$ columns are computed by Line 14:

$$\begin{aligned} &n\chi_{\pi,1} \cdot (\min(p_{r;2}, p_{r;3}) - \min(p_{r;2}, p_{r;4})) = n\chi_{\pi,1} \cdot p_{r;3}, \\ &n\chi_{\pi,2} \cdot (\min(p_{r;2}, p_{r;2}) - \min(p_{r;2}, p_{r;2})) = 0, \\ &n\chi_{\pi,3} \cdot (\min(p_{r;2}, p_{r;2}) - \min(p_{r;2}, p_{r;3})) \\ &\quad = n\chi_{\pi,3} \cdot (p_{r;2} - p_{r;3}), \\ &n\chi_{\pi,4} \cdot (\min(p_{r;2}, p_{r;4}) - \min(p_{r;2}, p_{r;4})) = 0, \end{aligned}$$

for the four chunks, respectively. The subroutine then copies the first $n\chi_{\pi,1} \cdot p_{r;3}$ columns of $\mathbf{B}_{\pi,1}$ to their corresponding columns of $\widetilde{\mathcal{V}}^\pi_{2\to\overline{1}3}$. If we treat the first $n\chi_{\pi,3} \cdot p_{r;3}$ columns of $\mathbf{B}_{\pi,3}$ as the first group, and the next $n\chi_{\pi,3} \cdot (p_{r;2} - p_{r;3})$ columns of $\mathbf{B}_{\pi,3}$ as the second group, then the subroutine copies the second group of columns of $\mathbf{B}_{\pi,3}$ to their corresponding columns of $\widetilde{\mathcal{V}}^\pi_{2\to\overline{1}3}$. (See Fig. 6.)

The subroutine then turns to the case of $i_0 = 3$ and the corresponding $\widetilde{\mathcal{V}}^\pi_{3\to\overline{1}2}$. The extra rank capacities that can be used for carrying $\widetilde{\mathcal{V}}^\pi_{3\to\overline{1}2}$ columns are computed as follows.

$$\begin{aligned} &n\chi_{\pi,1} \cdot (\min(p_{r;3}, p_{r;3}) - \min(p_{r;3}, p_{r;4})) = n\chi_{\pi,1} \cdot p_{r;3}, \\ &n\chi_{\pi,2} \cdot (\min(p_{r;3}, p_{r;2}) - \min(p_{r;3}, p_{r;2})) = 0, \\ &n\chi_{\pi,3} \cdot (\min(p_{r;3}, p_{r;2}) - \min(p_{r;3}, p_{r;3})) = 0, \\ &n\chi_{\pi,4} \cdot (\min(p_{r;3}, p_{r;4}) - \min(p_{r;3}, p_{r;4})) = 0, \end{aligned}$$

for the four chunks, respectively. The subroutine then copies the first $n\chi_{\pi,1} \cdot p_{r;3}$ columns of $\mathbf{B}_{\pi,1}$ (the first group of columns of $\mathbf{B}_{\pi,1}$) to their corresponding columns of $\widetilde{\mathcal{V}}^\pi_{3\to\overline{1}2}$. It is worth emphasizing that *code alignment is achieved in this construction since two identical copies of the first group of columns of* $\mathbf{B}_{\pi,1}$ *are used by both* $\widetilde{\mathcal{V}}^\pi_{2\to\overline{1}3}$ *and* $\widetilde{\mathcal{V}}^\pi_{3\to\overline{1}2}$. (See Fig. 6.)

When $j_0 = 3$, since $S^\pi_{3+1} = \emptyset$ we consider $i_0 = 1, 2$ and 3, and the corresponding columns $\widetilde{\mathcal{V}}^\pi_{1\to23}$, $\widetilde{\mathcal{V}}^\pi_{2\to13}$, and $\widetilde{\mathcal{V}}^\pi_{3\to12}$ separately. First consider $i_0 = 1$ and the corresponding $\widetilde{\mathcal{V}}^\pi_{1\to23}$. With the specified label functions, the extra rank capacities that can be used for carrying $\widetilde{\mathcal{V}}^\pi_{1\to23}$ columns are computed as

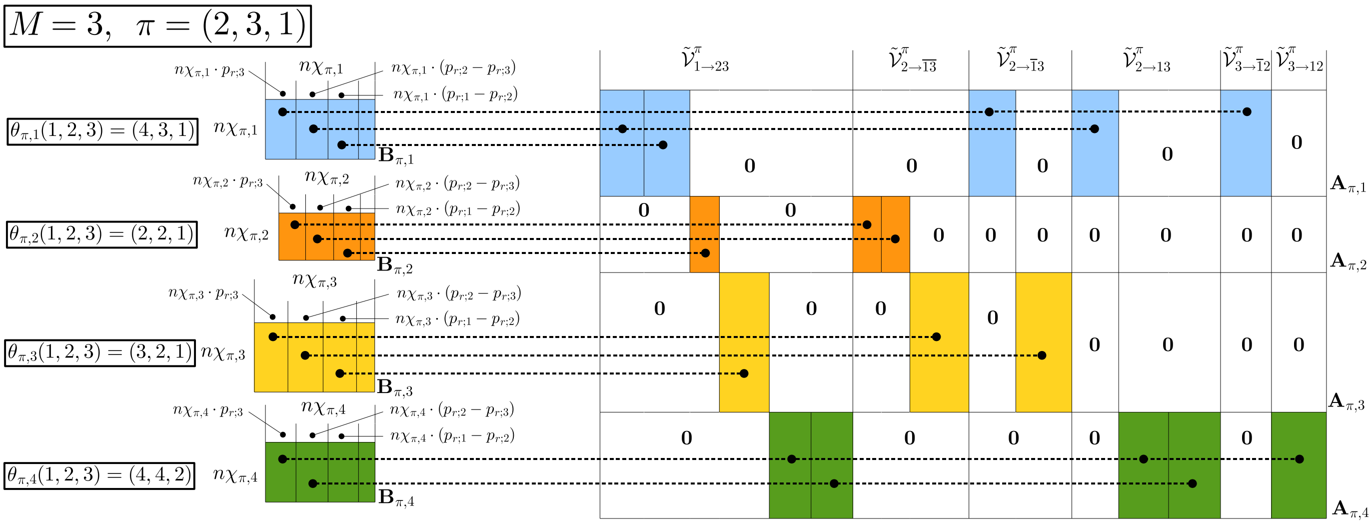


Fig. 6. Illustration of the systematic code construction in Section VI-B.

follows.

$$\begin{aligned}
&n\chi_{\pi,1} \cdot (\min(p_{r;1}, p_{r;1}) - \min(p_{r;1}, p_{r;3}))\\
&\quad = n\chi_{\pi,1} \cdot (p_{r;1} - p_{r;3}),\\
&n\chi_{\pi,2} \cdot (\min(p_{r;1}, p_{r;1}) - \min(p_{r;1}, p_{r;2}))\\
&\quad = n\chi_{\pi,2} \cdot (p_{r;1} - p_{r;2}),\\
&n\chi_{\pi,3} \cdot (\min(p_{r;1}, p_{r;1}) - \min(p_{r;1}, p_{r;2}))\\
&\quad = n\chi_{\pi,3} \cdot (p_{r;1} - p_{r;2}),\\
&n\chi_{\pi,4} \cdot (\min(p_{r;1}, p_{r;2}) - \min(p_{r;1}, p_{r;4})) = n\chi_{\pi,4} \cdot p_{r;2},
\end{aligned}$$

for the four chunks, respectively. Using the same concept of "groups" as discussed in the cases of $j_0 = 1$ and 2, the subroutine copies the second and the third groups of columns of $\mathbf{B}_{\pi,1}$, the third group of columns of $\mathbf{B}_{\pi,2}$, the third group of columns of $\mathbf{B}_{\pi,3}$, and the first and the second groups of columns of $\mathbf{B}_{\pi,4}$ to their corresponding columns of $\widetilde{\mathcal{V}}^{\pi}_{1\to 23}$, respectively. (See Fig. 6.)

For $i_0 = 2$, the extra rank capacities that can be used for carrying $\widetilde{\mathcal{V}}^{\pi}_{2\to 13}$ columns are

$$\begin{aligned}
&n\chi_{\pi,1} \cdot (\min(p_{r;2}, p_{r;1}) - \min(p_{r;2}, p_{r;3}))\\
&\quad = n\chi_{\pi,1} \cdot (p_{r;2} - p_{r;3}),\\
&n\chi_{\pi,2} \cdot (\min(p_{r;2}, p_{r;1}) - \min(p_{r;2}, p_{r;2})) = 0,\\
&n\chi_{\pi,3} \cdot (\min(p_{r;2}, p_{r;1}) - \min(p_{r;2}, p_{r;2})) = 0,\\
&n\chi_{\pi,4} \cdot (\min(p_{r;2}, p_{r;2}) - \min(p_{r;2}, p_{r;4})) = n\chi_{\pi,4} \cdot p_{r;2},
\end{aligned}$$

for the four chunks, respectively. The subroutine then copies the second group of columns of $\mathbf{B}_{\pi,1}$, and the first and the second groups of columns of $\mathbf{B}_{\pi,4}$ to their corresponding columns of $\widetilde{\mathcal{V}}^{\pi}_{2\to 13}$, respectively. (See Fig. 6.)

For $i_0 = 3$, the extra rank capacities that can be used for carrying $\widetilde{\mathcal{V}}^{\pi}_{3\to 12}$ columns are

$$\begin{aligned}
&n\chi_{\pi,1} \cdot (\min(p_{r;3}, p_{r;1}) - \min(p_{r;3}, p_{r;3})) = 0,\\
&n\chi_{\pi,2} \cdot (\min(p_{r;3}, p_{r;1}) - \min(p_{r;3}, p_{r;2})) = 0,\\
&n\chi_{\pi,3} \cdot (\min(p_{r;3}, p_{r;1}) - \min(p_{r;3}, p_{r;2})) = 0,\\
&n\chi_{\pi,4} \cdot (\min(p_{r;3}, p_{r;2}) - \min(p_{r;3}, p_{r;4})) = n\chi_{\pi,4} \cdot p_{r;3},
\end{aligned}$$

for the four chunks, respectively. The subroutine then copies the first group of columns of $\mathbf{B}_{\pi,4}$ to their corresponding columns of $\widetilde{\mathcal{V}}^{\pi}_{3\to 12}$. (See Fig. 6.)

To prove the decodability of the above network code construction, we first notice that in the above construction, each column participates in *one and only one* chunk $(\chi_{\pi,l}, \theta_{\pi,l}(\cdot))$ for some π and l (as can also be seen in Fig. 6). The following lemma then guarantees the decodability of the above network code construction.

Lemma 7: For any chunk index (π, l), any destination d_i can successfully decode all the user-i columns[8] that participate in chunk $(\chi_{\pi,l}, \theta_{\pi,l}(\cdot))$ by the observations of all time slots t that correspond to the coding submatrix $\mathbf{A}_{\pi,l}$.

We conclude this section by the proof of Lemma 7.

Proof of Lemma 7: Consider any chunk with index (π, l_0). For any user i, we use $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^{\pi}_{j_0+1}\overline{S^{\pi}_{j_0+1}}}$ to denote the columns of $\widetilde{\mathcal{V}}^{\pi}_{i\to T^{\pi}_{j_0+1}\overline{S^{\pi}_{j_0+1}}}$ that participate in chunk $(\chi_{\pi,l_0}, \theta_{\pi,l_0}(\cdot))$ during the j_0-th iteration of the outer FOR loop, Lines 2 to 22. Proving Lemma 7 is thus equivalent to proving that any user i_0 can decode the subspace $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i_0\to T^{\pi}_{j_0+1}\overline{S^{\pi}_{j_0+1}}}$ for all $j_0 \in [M]$ satisfying $i_0 \notin S^{\pi}_{j_0+1}$.

For any i_0, let j^* denote the unique index in $[M]$ satisfying $i_0 \notin S^{\pi}_{j^*+1}$ but $i_0 \in S^{\pi}_{j^*}$. By the definition S^{π}_j, we thus have $i_0 = \pi(j^*)$. We now compute an upper bound on the sum of the extra rank capacities of chunk (π, l_0) during the j_0-th

[8] We use "decoding a user-i column" as shorthand for "decoding a vector $\mathbf{v}$ that corresponds to one of the user-i columns."

iteration for $j_0 = j^*$ to M:[9]

$$\sum_{j_0=j^*}^{M} n\chi_{\pi,l_0} \cdot \Big(\min(p_{r;i_0}, p_{r;\theta_{\pi,l_0}(j_0)}) - \min(p_{r;i_0}, p_{r;\theta_{\pi,l_0}(j_0-1)}) \Big)$$
$$= n\chi_{\pi,l_0} \Big(\min(p_{r;i_0}, p_{r;\theta_{\pi,l_0}(M)}) - \min(p_{r;i_0}, p_{r;\theta_{\pi,l_0}(j^*-1)}) \Big)$$
$$\leq n\chi_{\pi,l_0} \cdot \max\Big(0, p_{r;i_0} - p_{r;\theta_{\pi,l_0}(j^*-1)}\Big) \quad (74)$$

Note that from user i_0's perspective, the columns of the *desired information* $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i_0\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ are assigned to $\mathbf{A}_{\pi,l}$ through the concept of extra rank capacity during $j_0 = j$. Therefore the total number of information columns $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i_0\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ for $j = j^*$ to M must be no larger than (74).

We note that since other user i, $i \neq i_0$, also participates in $\mathbf{A}_{\pi,l_0}$, decoding for user-i_0 has to take into account other users $i \neq i_0$. We first consider any $i = \pi(h)$ for some $h > j^*$. The corresponding subspaces of interest $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ must satisfy $j \geq h$ otherwise we will have $i \in S^\pi_{j+1}$. Since $j \geq h > j^*$, we must have $i_0 \in [M]\backslash(S^\pi_{j+1} \cup \{i\}) = T^\pi_{j+1}$. Therefore, from user-$i_0$'s perspective, all the subspaces $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ are side information already known to user i_0. Therefore, the columns corresponding to $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ are transparent from user-i_0's perspective.

We then consider any $i = \pi(h)$ for some $h < j^*$. The corresponding subspaces of interest $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ must satisfy $j \geq h$ otherwise we will have $i \in S^\pi_{j+1}$. By the same argument as in the previous case, those $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ for some $j \geq j^*$ are side information already known to user i_0 and can be omitted in our discussion. We thus focus only on the subspaces $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ for all $j \in [h, j^*-1]$. Recall that all the columns of the subspaces $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ are assigned to $\mathbf{A}_{\pi,l}$ during $j_0 = j$. Therefore, the columns of all those non-transparent subspaces $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ are copied from the k-th column of $\mathbf{B}_{\pi,l_0}$, where k is in the range

$$k \in [n\chi_{\pi,l_0} \cdot \min(p_{r;i}, p_{r;\theta_{\pi,l_0}(j-1)}) + 1, \; n\chi_{\pi,l_0} \cdot \min(p_{r;i}, p_{r;\theta_{\pi,l_0}(j)})]$$
$$\subseteq [1, n\chi_{\pi,l_0} \cdot p_{r;\theta_{\pi,l_0}(j^*-1)}],$$

where the "$\subseteq$" relationship is due to that the range of j for non-transparent subspaces is $[h, j^*-1]$ and $\theta_{\pi,l_0}(\cdot)$ is a non-increasing function. *Due to our code-alignment-based construction, all columns of the non-transparent subspaces* $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ *are copied from (a subset of) the first* $n\chi_{\pi,l_0} \cdot p_{r;\theta_{\pi,l_0}(j^*-1)}$ *columns of the same matrix* $\mathbf{B}_{\pi,l_0}$. Therefore, from the decoding perspective of user i_0, the effective degrees of freedom of the non-transparent columns is no larger than

$$n\chi_{\pi,l_0} \cdot p_{r;\theta_{\pi,l_0}(j^*-1)}. \quad (75)$$

[9]For simplicity, we use the convention that $\min(p_{r;i_0}, p_{r;\theta_{\pi,l_0}(0)}) \triangleq 0$ in (74) and in the following discussion.

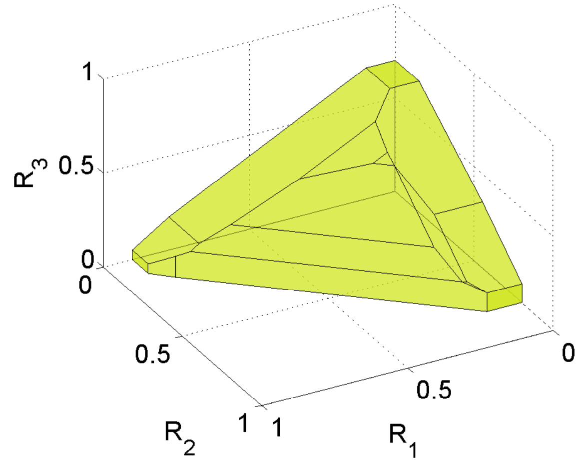


Fig. 7. Illustration of an example capacity region for the case of $M = 3$.

Also note that the total number of successful reception of user i_0 for the chunk $(\chi_{\pi,l_0}, \theta_{\pi,l_0}(\cdot))$ is

$$n\chi_{\pi,l_0} \cdot p_{r;i_0}. \quad (76)$$

For the non-trivial case in which the total number of the to-be-decoded columns of $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i_0\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ is non-zero, we can remove the $\max(0,\cdot)$ operation of (74). Then by comparing (74), (75), and (76), we must have that the "total number of to-be-decoded columns" plus "the effective degrees of freedom of the non-transparent interference columns" is no larger than "the total number of observed time slots." By the MDS property of RLNC, it is thus guaranteed that user i_0 can successfully decode all the desired information columns $\widetilde{\mathcal{V}}^{(\pi,l_0)}_{i_0\to T^\pi_{j+1}\overline{S^\pi_{j+1}}}$ from the observation of the $\mathbf{A}_{\pi,l_0}$ coding sub-matrix. The proof is thus complete. ■

VII. Numerical Experiments And Some Observations

A. Numerical Experiments

1) Capacity Region for $M \leq 3$: For the simplest case of $M = 2$, the capacity region is described by Corollary 2, which can be expressed by six linear inequalities once we expand the min and the projection operators.

For $M = 3$, we plot the 3-D capacity region for an example network. To simplify the illustration, we assume all the PECs are spatially independent. Namely, the joint success probability $p_{s\to S\overline{T}}$ can be expressed as the product of the marginals. We consider an example network with the marginals being

$$\begin{bmatrix} p_{s_1;r} & p_{s_1;2} & p_{s_1;3} \\ p_{s_2;1} & p_{s_2;r} & p_{s_2;3} \\ p_{s_3;1} & p_{s_3;2} & p_{s_3;r} \end{bmatrix} = \begin{bmatrix} 1 & 0.12 & 0.1 \\ 0.1 & 1 & 0.15 \\ 0.05 & 0.1 & 1 \end{bmatrix}$$

and $(p_{r;1}, p_{r;2}, p_{r;3}) = (0.9, 0.8, 0.7)$.

Fig. 7 illustrates the corresponding 3-D capacity region of the above wireless 1-hop network. As seen in Fig. 7, the capacity region of $M = 3$ is non-trivial and contains many facets that are governed by the linear inequalities in Fig. 3.

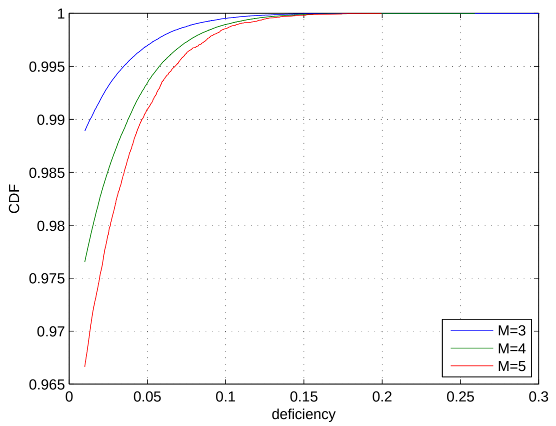


Fig. 8. The cumulative distribution functions (CDFs) of defi for $M = 4$ and 5, respectively.

2) Numerical Results for General $M \geq 4$: For general $M \geq 4$, the outer bound may not match the capacity region. We use numerical experiments to evaluate the tightness of the outer and the inner bounds in the following way. We first assume the spatial independence between receivers so that we can specify the PECs by their marginal success probabilities. We then assume that $p_{s_i;r} = 1$ for all $i \in [M]$ and assume that all other $p_{s_i;j}$ are independently and uniformly distributed over $[0, 1]$. The $p_{r;i}$ are chosen by sorting M independent uniform random variables on $[0, 1]$ in the descending order. To capture the M-dimensional capacity *region*, we first choose a search direction $\vec{u} = (u_1, \cdots, u_M)$ uniformly randomly from an M-dimensional unit ball. With the chosen values of the parameters $p_{\cdot;\cdot}$ and the search direction $\vec{u}$, we use a LP solver to find the largest t_{outer} such that $(R_1, \cdots, R_M) = (u_1 \cdot t_{\text{outer}}, \cdots, u_M \cdot t_{\text{outer}})$ satisfies the capacity outer bound in Proposition 2, and find the largest t_{inner} such that $(R_1, \cdots, R_M) = (u_1 \cdot t_{\text{inner}}, \cdots, u_M \cdot t_{\text{inner}})$ satisfies the capacity inner bound in Proposition 4. The deficiency is then defined as defi $\triangleq \frac{t_{\text{outer}} - t_{\text{inner}}}{t_{\text{outer}}}$. We then repeat the above experiment for 10^6 and 10^5 times, for $M = 4$ and 5, respectively. For both the cases of $M = 4$ and 5, defi is less than 1% for at least 96.7% of the times and is less than 5% for 99% of times. See Fig. 8 for the corresponding cumulative distribution functions (cdfs). The numerical results show the effectiveness of the proposed outer and inner bound pair in terms of bracketing the capacity region, at least from a practical perspective.

B. Observations and Remarks

1) Tightness of The Inner Bound: Regarding the tightness of the capacity outer and inner bounds, the following corollary can be proven.

Corollary 5: The outer and inner bounds in Propositions 2 and 4 meet for the case of $M = 2$.

For $M \geq 3$, the two bounds do not always meet. To investigate the tightness of the systematic inner bound in Proposition 4, we also compute the defi value numerically for 10^6 instances of the case of $M = 3$ based on the same methodology used in the previous subsection. Note that for $M = 3$, the outer bound of Proposition 2 is the capacity region. defi thus quantifies the suboptimality of the inner bound of Proposition 4 when compared to the true capacity region. In our numerical evaluation (see Fig. 8), the difference defi is no larger than 1% for 98.9% of the time.

To explore the causes of the suboptimality of the inner bound, we first notice that the achievability scheme in Section VI consists of two components. The first component is to decide "which basis vector has been overheard by which destination" so that we can partition the message space Ω_i into proper *side information subspaces*, which is achieved in Step 2 of the scheme. The second component is to redistribute the side information subspaces (as in Step 3) and then use the code alignment technique in Step 4 to "pack the side information subspaces" as compactly as possible.

First consider the second component. For each permutation π, the corresponding side information subspaces $\widetilde{\mathcal{V}}^{\pi}_{i\to T^{\pi}_{j+1}\overline{S^{\pi}_{j+1}}}$ have an acyclic relationship in the sense that the corresponding non-overhearing sets S^{π}_{j+1}, $j \in [M]$, have a total ordering $S^{\pi}_2 \supseteq S^{\pi}_3 \supseteq \cdots \supseteq S^{\pi}_{M+1}$. It can be proven that given the to-be-packed subspaces $\widetilde{\mathcal{V}}^{\pi}_{i\to T^{\pi}_{j+1}\overline{S^{\pi}_{j+1}}}$, $\forall j \in [M], \forall i \in [M]\backslash S^{\pi}_{j+1}$, the proposed redistribution plus code-alignment coding scheme is the optimal. Namely, if we are only interested in packing subspaces $\widetilde{\mathcal{V}}^{\pi}_{i\to T^{\pi}_{j+1}\overline{S^{\pi}_{j+1}}}$ that have an acyclic relationship, the proposed redistribution plus code-alignment scheme can pack the subspaces using the minimal number of time slots. However, the suboptimality lies in the approach of performing coding on the partitioned subspaces separately. For comparison, the optimal coding solution in Section V-B performs coding with holistic consideration of all subspaces. That is, the coding matrix in Fig. 5 does not have the clean separation as in the coding matrix of Fig. 6. On the other hand, a holistic network coding scheme may be difficult to design for $M > 3$ due to the exponential number of different types of overhearing patterns for large M.

Further, even the first component that partitions the message space Ω_i into proper side information subspaces turns out to be highly non-trivial for the cases of $M \geq 3$. For example, consider an $(n \cdot R)$-dimensional message vector $\mathbf{X}$, and we use RLNC to mix all nR symbols by randomly generated n coding vectors $\mathbf{c}(1)$ to $\mathbf{c}(n)$. Consider two receivers d_1 and d_2 and the corresponding success probabilities being p_1 and p_2, respectively. By the MDS property of RLNC, the linear subspace received by d_1, denoted by $\mathcal{V}_1$, has rank $\text{Rank}(\mathcal{V}_1) = \min(nR, np_1)$. The linear subspace received by d_2, denoted by $\mathcal{V}_2$, has rank $\text{Rank}(\mathcal{V}_2) = \min(nR, np_2)$. The question is how large is the rank of the intersection, $\text{Rank}(\mathcal{V}_1 \cap \mathcal{V}_2)$. The answer to this question is critical for partitioning the message space since it corresponds to how many basis vectors that are known to both d_1 and d_2. With some algebraic computation, it can be shown that

$$\begin{aligned}&\text{Rank}(\mathcal{V}_1 \cap \mathcal{V}_2)\\&= \min(nR, np_1) + \min(nR, np_2) - \min(nR, np_{1\cup 2}) \quad (77)\end{aligned}$$

where $p_{1\cup 2}$ is the probability that a packet is received by at least one of d_1 and d_2. One can verify that for very general choices of p_1, p_2, and $p_{1\cup 2}$, say $(p_1, p_2, p_{1\cup 2}) = (0.5, 0.5, 0.75)$, (77) is neither concave nor convex with respect to R. This thus presents a substantial challenge when partitioning the message space into its proper side information subspaces. In the broadest sense, the optimal construction in Section V-B directly uses the optimal expression (77) but our systematic inner bound construction in Section VI uses a conservative but more tractable lower bound

$$\mathrm{Rank}(\mathcal{V}_1 \cap \mathcal{V}_2) \geq \min(nR, n \cdot (p_1 + p_2 - p_{1\cup 2})),$$

which is concave with respect to R. However, using the above lower bound induces capacity loss and part of it is reflected in the defi of Fig. 8.

2) Complexity Of The Outer And Inner Bounds: Although checking the feasibility of LP problems is a P problem, the number of variables and linear constraints of Propositions 2 and 4 is exponential (or super exponential) with respect to M. On the other hand, for any fixed M value, the complexity of computing the bounds does not depend on n the block length of the network code, even though there are totally $\sum_{i=1}^{M} nR_i$ symbols, $X_{i,1}, \cdots, X_{1,nR_i}, i \in [M]$, to be coded together. This fractional setting thus presents a big complexity improvement over the integral settings of [27] and of [3], for which the INC complexity is exponential with respect to the number of symbols to be coded together.

Note that like any NP-hardness results, the NP-hardness of finding the index code capacity [1], [3] is a worst-case analysis. As can be seen in the cumulative distribution functions in Fig. 8, the gap between the outer and the inner bounds can be as large as 27% in the worst case, which fits the prediction of the NP-hardness results. On the other hand, the outer and inner bounds meet in the average cases and can thus be used effectively as the substitute of capacity from the practical perspective.

We would also like to emphasize that in a wireless network setting, it is unlikely[10] that more than $M \geq 6$ sessions (s_i, d_i), $i \in [M]$, are using the same relay node r. For small $M \leq 5$, the outer and inner bounds can be easily evaluated by standard LP solvers.

VIII. Conclusion

In this work, we have studied the capacity region of wireless 1-hop relay networks with M coexisting unicast sessions, based on the packet erasure channel model and a 2-staged network coding setting that captures the benefits of message side information. The 2-staged intersession network coding (2-staged INC) capacity region for $M \leq 3$ has been fully characterized. For $M \geq 4$, the 2-staged INC capacity has been bounded by a new pair of outer and inner bounds, the tightness of which is within 1% for at least 96.7% of the times for $M \in \{4, 5\}$. The outer and inner bounds can thus be used effectively as the substitute of 2-staged INC capacity from the practical perspective.

The proposed achievability scheme is based on the concept of *code alignment*, which was also used in [13], [38], [40], [43] and is in parallel with the interference alignment technique in the setting of Gaussian interference channel. The results in this paper can also be viewed as the generalization of index-coding capacity from wireline broadcast with binary alphabets to wireless broadcast with high-order alphabets.

[10]For a satellite communication setting as in [3], we may have more than $M \geq 6$ sessions sharing the same satellite.

Acknowledgment

We would like to thank the anonymous reviewers for their valuable suggestions and help on refining the proofs.

Appendix A
Proof of Proposition 1

Proof of Proposition 1: Inequality (5) can be proven by the statistically degraded channel arguments. To prove inequality (6), we assume the broadcast PEC is physically degraded. Namely, the channel status $(1_{\{Z_{1,t}\neq *\}}, 1_{\{Z_{2,t}\neq *\}}, 1_{\{Z_{3,t}\neq *\}})$ has the following distribution:

$$\begin{aligned}&\mathsf{Prob}\left(1_{\{Z_{1,t}\neq *\}} = a, 1_{\{Z_{2,t}\neq *\}} = b, 1_{\{Z_{3,t}\neq *\}} = c\right)\\ &= \begin{cases} 1 - p_1 & \text{if } a = b = c = 0 \\ p_1 - p_2 & \text{if } a = 1,\, b = c = 0 \\ p_2 - p_3 & \text{if } a = b = 1,\, c = 0 \\ p_3 & \text{if } a = b = c = 1 \\ 0 & \text{otherwise} \end{cases}.\end{aligned}$$

With the above construction, $U \to \mathbf{W} \to \mathbf{Z}_1 \to \mathbf{Z}_2 \to \mathbf{Z}_3$ form a Markov chain.

We will first prove the following inequality:

$$I(U; \mathbf{Z}_1 | \mathbf{Z}_2 \mathbf{Z}_3) \leq \frac{p_1 - p_2}{p_1 - p_3} I(U; \mathbf{Z}_1 | \mathbf{Z}_3). \tag{78}$$

To that end, we notice the following

$$\begin{aligned} &I(U; \mathbf{Z}_1 | \mathbf{Z}_2 \mathbf{Z}_3) \\ &= \sum_{t=1}^{n} I(U; Z_1(t) | [Z_1]_1^{t-1} \mathbf{Z}_2 \mathbf{Z}_3) \\ &\leq \sum_{t=1}^{n} I(U; Z_1(t) | [Z_1]_1^{t-1} Z_2(t) \mathbf{Z}_3), \end{aligned} \tag{79}$$

where $[Z_1]_1^{t-1} \triangleq \{Z_1(\tau) : \forall \tau \in [t-1]\}$. Inequality (79) follows from the facts $H(Z_1(t)|[Z_1]_1^{t-1}\mathbf{Z}_2\mathbf{Z}_3) \leq H(Z_1(t)|[Z_1]_1^{t-1}Z_2(t)\mathbf{Z}_3)$ and $H(Z_1(t)|U[Z_1]_1^{t-1}\mathbf{Z}_2\mathbf{Z}_3) = H(Z_1(t)|U[Z_1]_1^{t-1}Z_2(t)\mathbf{Z}_3)$. Each term of the summation in (79) can be further simplified by

$$\begin{aligned} &I(U; Z_1(t) | [Z_1]_1^{t-1} Z_2(t) \mathbf{Z}_3) \\ &= I(U; Z_1(t) 1_{\{Z_1(t)\neq *\}} | [Z_1]_1^{t-1} Z_2(t) \mathbf{Z}_3) \\ &= I(U; Z_1(t) | [Z_1]_1^{t-1} 1_{\{Z_1(t)\neq *\}} Z_2(t) \mathbf{Z}_3) \\ &= \mathsf{Prob}(Z_1(t) \neq *, Z_2(t) = *) \\ &\quad \cdot I(U; Z_1(t) | [Z_1]_1^{t-1}, Z_1(t) \neq *, Z_2(t) = *, \mathbf{Z}_3) \\ &= (p_1 - p_2) \\ &\quad \cdot I(U; Z_1(t) | [Z_1]_1^{t-1}, Z_1(t) \neq *, Z_2(t) = Z_3(t) = *, [Z_3]_{\backslash t}) \end{aligned} \tag{80}$$

$$= (p_1 - p_2) I(U; Z_1(t) | [Z_1]_1^{t-1}, Z_1(t) \neq *, Z_3(t) = *, [Z_3]_{\backslash t}), \tag{81}$$

where $[Z_3]_{\backslash t} \triangleq \{Z_3(\tau) : \tau \in [n]\backslash t\}$ and (80) follows from that $Z_2(t) = *$ implies $Z_3(t) = *$. Eq. (81) follows from the fact that conditioning on $Z_1(t) \neq *$, whether $Z_2(t) = *$ or not does not change the entropy of $Z_1(t)$. That is,

$$\begin{aligned}
&H(Z_1(t)|[Z_1]_1^{t-1}, Z_1(t) \neq *, Z_2(t) = Z_3(t) = *, [Z_3]_{\backslash t})\\
&= H(Z_1(t)|[Z_1]_1^{t-1}, Z_1(t) \neq *, Z_3(t) = *, [Z_3]_{\backslash t}) \quad \text{and}\\
&H(Z_1(t)|U[Z_1]_1^{t-1}, Z_1(t) \neq *, Z_2(t) = Z_3(t) = *, [Z_3]_{\backslash t})\\
&= H(Z_1(t)|U[Z_1]_1^{t-1}, Z_1(t) \neq *, Z_3(t) = *, [Z_3]_{\backslash t}).
\end{aligned}$$

We also notice that

$$\begin{aligned}
&I(U; Z_1(t)|[Z_1]_1^{t-1}\mathbf{Z}_3)\\
&= I(U; Z_1(t)|[Z_1]_1^{t-1}1_{\{Z_1(t)\neq *\}}\mathbf{Z}_3)\\
&= \mathsf{Prob}(Z_1(t) \neq *, Z_3(t) = *)\\
&\quad \cdot I(U; Z_1(t)|[Z_1]_1^{t-1}, Z_1(t) \neq *, Z_3(t) = *, [Z_3]_{\backslash t})\\
&= (p_1 - p_3)I(U; Z_1(t)|[Z_1]_1^{t-1}, Z_1(t) \neq *, Z_3(t) = *, [Z_3]_{\backslash t}).
\end{aligned}$$

As a result,

$$(81) = \frac{p_1 - p_2}{p_1 - p_3} I(U; Z_1(t)|[Z_1]_1^{t-1}\mathbf{Z}_3). \tag{82}$$

Substituting (82) into (79), we have proven (78).

Since $I(U; \mathbf{Z}_2|\mathbf{Z}_3) + I(U; \mathbf{Z}_1|\mathbf{Z}_2\mathbf{Z}_3) = I(U; \mathbf{Z}_1\mathbf{Z}_2|\mathbf{Z}_3) = I(U; \mathbf{Z}_1|\mathbf{Z}_3)$, Eq. (78) implies

$$I(U; \mathbf{Z}_2|\mathbf{Z}_3) \geq \frac{p_2 - p_3}{p_1 - p_3} I(U; \mathbf{Z}_1|\mathbf{Z}_3). \tag{83}$$

The final step is to notice that $U \to \mathbf{W} \to \mathbf{Z}_1 \to \mathbf{Z}_2 \to \mathbf{Z}_3$ form a Markov chain. We thus have

$$\begin{aligned}
&I(U; \mathbf{Z}_2|\mathbf{Z}_3) = I(U; \mathbf{Z}_2) - I(U; \mathbf{Z}_3)\\
\text{and} \quad &I(U; \mathbf{Z}_1|\mathbf{Z}_3) = I(U; \mathbf{Z}_1) - I(U; \mathbf{Z}_3).
\end{aligned}$$

Substituting the above two equalities into (83), we have proven (6). The proof of Proposition 1 is thus complete. ■

Appendix B
Proof of Proposition 2

In this proof, we assume the (r, d_k) PEC is physically degraded with respect to the (r, d_i) PEC for any $i \leq k$.

By definition, for any $\epsilon > 0$, let n, q, and (g_{s_i}, g_r, g_{d_j}) denote the block length, the underlying finite field size, and the corresponding network code satisfying $\max_{i\in[M]} \mathsf{Prob}(\hat{\mathbf{X}}_i \neq \mathbf{X}_i) < \epsilon$. We first notice that for any $i \in [M]$, $\mathbf{X}_i \to \mathbf{W}_i \to \mathbf{Z}_{i\to r} \to \hat{\mathbf{X}}_i$ form a Markov chain. As a result, we have

$$I(\mathbf{X}_i; \hat{\mathbf{X}}_i) \leq I(\mathbf{W}_i; \mathbf{Z}_{i\to r}) \leq np_{s_i;r}\log(q). \tag{84}$$

On the other hand, by Fano's inequality, we have

$$I(\mathbf{X}_i; \hat{\mathbf{X}}_i) \geq nR_i(1 - \epsilon)\log(q) - H_2(\epsilon), \tag{85}$$

where $H_2(\cdot)$ is the binary entropy function. Jointly (84) and (85) imply that

$$R_i(1 - \epsilon) - \frac{H_2(\epsilon)}{n\log(q)} \leq p_{s_i;r}. \tag{86}$$

By choosing an infinitesimal $\epsilon > 0$, the above inequality implies (7).

For any $\epsilon > 0$ and the corresponding network code, we define 2^M functions $f_S^{[\epsilon]}(\cdot)$ indexed by $S \in 2^{[M]}$ such that for all $i \in [M]$,

$$f_S^{[\epsilon]}(p_{r;i}) = \frac{I(U; \mathbf{Z}_{r\to i}|\{\mathbf{Z}_{l\to j}\mathbf{X}_j : \forall j \in S, \forall l \in [M]\backslash j\})}{n\log(q)} \tag{87}$$

and we interpolate the $f_S^{[\epsilon]}(p)$ value for all $p \neq p_{r;i}$, where

$$U = \{\mathbf{X}_i, \mathbf{Z}_{i\to r}, \mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}} : \forall i, j \in [M], i \neq j\}$$

denotes all the information available at the relay r plus all the information messages at all the sources; $\mathbf{Z}_{r\to i}$ denotes the n-dimensional (r, d_i) PEC output vector; $\{\mathbf{Z}_{l\to j} : \forall j \in S, \forall l \in [M]\backslash j\}$ is all the side information received by $\{d_j : \forall j \in S\}$; and $\{\mathbf{X}_j : \forall j \in S\}$ is the information messages for destination d_j, $j \in S$. By Proposition 1, $f_S^{[\epsilon]}(\cdot)$ must be a ZPPLCIF.

We first notice that when $S = \emptyset$, by definition we have

$$f_\emptyset^{[\epsilon]}(p_{r;M}) = \frac{I(U; \mathbf{Z}_{r\to M})}{n\log(q)} \leq \frac{np_{r;M}\log(q)}{n\log(q)} = p_{r;M}. \tag{88}$$

For notational simplicity, for any $j \in [M]$, we use $\mathbf{Z}_{\mathsf{SI}\to j} \triangleq \{\mathbf{Z}_{l\to j} : \forall l \in [M]\backslash j\}$ to denote all the side information overheard by destination d_j. We then notice that for all $i, k \in [M], S, T \in 2^{[M]}$ satisfying $k \notin S$ and $T = S \cup \{k\}$, we have

$$\begin{aligned}
&(n\log(q)) \cdot f_S^{[\epsilon]}(p_{r;i})\\
&= I(U; \mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) &(89)\\
&= I(U\mathbf{X}_k\mathbf{Z}_{\mathsf{SI}\to k}; \mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) &(90)\\
&= I(U; \mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S \cup \{k\}\})\\
&\quad + I(\mathbf{X}_k\mathbf{Z}_{\mathsf{SI}\to k}; \mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) &(91)\\
&= (n\log(q)) \cdot f_T^{[\epsilon]}(p_{r;i})\\
&\quad + I(\mathbf{X}_k\mathbf{Z}_{\mathsf{SI}\to k}; \mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}), &(92)
\end{aligned}$$

where (89) follows from definition, (90) follows from that $(\mathbf{X}_k, \mathbf{Z}_{\mathsf{SI}\to k})$ is a function of U, (91) follows from the chain rule, and (92) follows from definition. We can further simplify the second term of (92) as follows.

$$\begin{aligned}
&I(\mathbf{X}_k\mathbf{Z}_{\mathsf{SI}\to k}; \mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) &(93)\\
&= I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\})\\
&\quad + I(\mathbf{Z}_{\mathsf{SI}\to k}; \mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\})\\
&\quad - I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) &(94)\\
&= I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\})\\
&\quad + I(\mathbf{Z}_{\mathsf{SI}\to k}; \mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) &(95)\\
&\geq I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to i}|\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}), &(96)
\end{aligned}$$

where (94) follows from the information equality $I(XY; Z) = I(X; YZ) + I(Y; Z) - I(X; Y)$; (95) follows from that conditioning on $\{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}$, $\mathbf{X}_k$, the information at source s_k, is independent from the side information $\mathbf{Z}_{\mathsf{SI}\to k} = \{\mathbf{Z}_{l\to k} : \forall l \in [M]\backslash k\}$ at destination d_k; and (96) follows from the non-negativeness of mutual information.

We then notice that for any $i \leq k$,

$$
\begin{aligned}
&I(\mathbf{X}_k; \{\mathbf{Z}_{k\to j} : \forall j \in S\}) \\
&\quad + I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to i} | \{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) \\
&= I(\mathbf{X}_k; \{\mathbf{Z}_{k\to j} : \forall j \in S\} | \{\mathbf{X}_j : \forall j \in S\}) \\
&\quad + I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to i} | \{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) \qquad (97)\\
&= I(\mathbf{X}_k; \{\mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\} | \{\mathbf{X}_j : \forall j \in S\}) \\
&\quad + I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to i} | \{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) \qquad (98)\\
&= I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to i}\{\mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\} | \{\mathbf{X}_j : \forall j \in S\}) \qquad (99)\\
&\geq I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to k}\{\mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\} | \{\mathbf{X}_j : \forall j \in S\}) \qquad (100)\\
&\geq I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to k} | \{\mathbf{X}_j : \forall j \in S\}) \qquad (101)\\
&\geq I(\mathbf{X}_k; \mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to k}) \qquad (102)\\
&\geq I(\mathbf{X}_k; \hat{\mathbf{X}}_k) \geq n(R_k - \epsilon)\log(q) - H_2(\epsilon), \qquad (103)
\end{aligned}
$$

where (97) follows from that the joint distribution of $(\mathbf{X}_k, \{\mathbf{Z}_{k\to j} : \forall j \in S\})$ is independent from the value of $\{\mathbf{X}_j : \forall j \in S\}$; (98) follows from that conditioning on $\{\mathbf{X}_j : \forall j \in S\}$, both $\mathbf{X}_k \to \{\mathbf{Z}_{k\to j} : \forall j \in S\} \to \{\mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}$ and $\mathbf{X}_k \to \{\mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\} \to \{\mathbf{Z}_{k\to j} : \forall j \in S\}$ are Markov chains; (99) follows from the chain rule; (100) follows from that the (r, d_k) channel is physically degraded with respect to the (r, d_i) channel; (101) follows from removing one extra observation; (102) follows from that the distribution $\mathbf{X}_k$ is independent from that of $\{\mathbf{X}_j : \forall j \in S\}$; and (103) follows from that $\hat{\mathbf{X}}_k = g_{d_k}(\mathbf{Z}_{\mathsf{SI}\to k}\mathbf{Z}_{r\to k})$ and Fano's inequality.

Combining (96), (103), and the non-negativeness of mutual information, we have

$$
\begin{aligned}
&I(\mathbf{X}_k\mathbf{Z}_{\mathsf{SI}\to k}; \mathbf{Z}_{r\to i} | \{\mathbf{X}_j, \mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}) \\
&\geq \left(n\log(q)(R_k - \epsilon) - H_2(\epsilon) - I(\mathbf{X}_k; \{\mathbf{Z}_{k\to j} : \forall j \in S\})\right)^+, \qquad (104)
\end{aligned}
$$

where $(\cdot)^+$ is the projection to non-negative reals. Since

$$I(\mathbf{X}_k; \{\mathbf{Z}_{k\to j} : \forall j \in S\}) \leq n\log(q)p_{s_k;\cup S}, \qquad (105)$$

combining (92), (104), and (105) we thus have that $\forall i, k \in [M], \forall S, T \in 2^{[M]}$ satisfying $k \notin S$ and $T = S \cup \{k\}$,

$$
\begin{aligned}
f_S^{[\epsilon]}(p_{r;i}) &\geq f_T^{[\epsilon]}(p_{r;i}) \\
&\quad + 1_{\{i\leq k\}}\left(R_k - \epsilon - \frac{H_2(\epsilon)}{n\log(q)} - p_{s_k;\cup S}\right)^+. \qquad (106)
\end{aligned}
$$

For any strictly positive, infinite, decreasing sequence $(\epsilon_1, \epsilon_2, \cdots)$ satisfying $\lim_{t\to\infty}\epsilon_t = 0$, we construct

$$f_S(p) \triangleq \lim_{t\to\infty} \inf\{f_S^{[\epsilon_\tau]}(p) : \forall \tau \geq t\}.$$

for any $S \in 2^{[M]}$ as the point-wise infimum limit of $\{f_S^{[\epsilon_t]}(p)\}_t$. Obviously, $f_S(\cdot) \geq 0$. Since each $f_S^{[\epsilon_t]}(\cdot)$ is ZPPLCIF, it can be shown that $f_S(\cdot)$ must be a zero-passing, concave, increasing function. Moreover, since $\lim_{t\to\infty}\epsilon_t = 0$, (88) and (106) ensure that

$$
\begin{aligned}
&f_\emptyset(p_{r;M}) \leq p_{r;M} \text{ and} \\
&f_S(p_{r;i}) \geq f_T(p_{r;i}) + 1_{\{i\leq k\}}\left(R_k - p_{s_k;\cup S}\right)^+.
\end{aligned}
$$

Such $f_S(\cdot), \forall S \in 2^{[M]}$ thus satisfy (8) and (9) in Proposition 2. Note that $f_S(\cdot)$ may not be piece-wise linear. On the other hand, we can "linearize" $f_S(\cdot)$ by constructing another $f_{S,\text{PL}}(\cdot)$ from $f_S(\cdot)$ satisfying

$$f_{S,\text{PL}}(p_{r;i}) \triangleq f_S(p_{r;i})$$

and by interpolating the $f_{S,\text{PL}}(p)$ value for all $p \notin \{p_{r;i} : \forall i \in [M]\}$. The resulting $f_{S,\text{PL}}(\cdot)$ will still satisfy (8) and (9) and is guaranteed to be a ZPPLCIF. The proof of Proposition 2 is thus complete.

Appendix C
Proof of Corollary 3

In this appendix, we prove the alternative form of Proposition 2 for the case of $M = 3$. In this case, we have eight functions $f_S(\cdot)$ to consider for $S = \{1,2,3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1\}$, $\{2\}$, $\{3\}$, and $\emptyset$. Before discussing each function separately. we introduce the following self-explanatory lemma.

Lemma 8: Consider p_1, p_2, and p_3 satisfying $0 \leq p_3 < p_2 < p_1 \leq 1$. Consider any ZPPLCIF $f(\cdot)$ that satisfies

$$f(p_1) \geq a, \quad f(p_2) \geq b, \text{ and } f(p_3) \geq c,$$

for some $a, b, c \geq 0$. Then due to the concavity and the monotonicity of $f(\cdot)$, we must also have

$$
\begin{aligned}
&f(p_1) \geq \max(a, b, c), \quad f(p_3) \geq \max\left(\frac{p_3}{p_1}a, \frac{p_3}{p_2}b, c\right) \\
&\text{and } f(p_2) \geq \max\left(\frac{p_2}{p_1}a, b, c, \frac{p_2 - p_3}{p_1 - p_3}a + \frac{p_1 - p_2}{p_1 - p_3}c\right).
\end{aligned}
$$

For the following, we discuss the eight cases separately.

Case 1: $S = \{1,2,3\}$. Similar to the $M = 2$ case, we can simply choose $f_{\{1,2,3\}}(p) \triangleq 0$ for all $p \in [0,1]$.

Case 2: $S = \{1,2\}$. (8) implies that

$$
\begin{aligned}
&f_{\{1,2\}}(p_{r;1}) \geq 0 + \left(R_3 - p_{s_3;\cup\{1,2\}}\right)^+ \\
&f_{\{1,2\}}(p_{r;2}) \geq 0 + \left(R_3 - p_{s_3;\cup\{1,2\}}\right)^+ \\
&f_{\{1,2\}}(p_{r;3}) \geq 0 + \left(R_3 - p_{s_3;\cup\{1,2\}}\right)^+.
\end{aligned}
$$

Case 3: $S = \{1,3\}$. (8) implies that

$$
\begin{aligned}
&f_{\{1,3\}}(p_{r;1}) \geq 0 + \left(R_2 - p_{s_2;\cup\{1,3\}}\right)^+ \\
&f_{\{1,3\}}(p_{r;2}) \geq 0 + \left(R_2 - p_{s_2;\cup\{1,3\}}\right)^+ \\
&f_{\{1,3\}}(p_{r;3}) \geq 0. \qquad (107)
\end{aligned}
$$

Lemma 8 then implies that (107) can be strengthened and replaced by the following inequality

$$f_{\{1,3\}}(p_{r;3}) \geq \frac{p_{r;3}}{p_{r;2}}\left(R_2 - p_{s_2;\cup\{1,3\}}\right)^+$$

Case 4: $S = \{2,3\}$. (8) implies that

$$
\begin{aligned}
&f_{\{2,3\}}(p_{r;1}) \geq 0 + \left(R_1 - p_{s_1;\cup\{2,3\}}\right)^+ \\
&f_{\{2,3\}}(p_{r;2}) \geq 0 \qquad (108)\\
&f_{\{2,3\}}(p_{r;3}) \geq 0. \qquad (109)
\end{aligned}
$$

Lemma 8 then implies that (108) and (109) can be strengthened and replaced by the following inequalities

$$f_{\{2,3\}}(p_{r;2}) \geq \frac{p_{r;2}}{p_{r;1}} \left(R_1 - p_{s_1;\cup\{2,3\}}\right)^+$$
$$f_{\{2,3\}}(p_{r;3}) \geq \frac{p_{r;3}}{p_{r;1}} \left(R_1 - p_{s_1;\cup\{2,3\}}\right)^+ .$$

The discussion of Cases 5 to 8 will use the notation $\overline{R}_{i;T}$, which was first introduced in (34).

Case 5: $S = \{1\}$. We now have two different choices of T: $T = \{1,2\}$ and $T = \{1,3\}$. By applying (8) for $T = \{1,2\}$ and for $T = \{1,3\}$, respectively, and by the discussion of Cases 2 and 3, we have

$$\begin{aligned} f_{\{1\}}(p_{r;1}) &\geq \max\left(\overline{R}_{2;1} + \overline{R}_{3;\{1,2\}}, \overline{R}_{2;\{1,3\}} + \overline{R}_{3;1}\right) \\ f_{\{1\}}(p_{r;2}) &\geq \max\left(\overline{R}_{2;1} + \overline{R}_{3;\{1,2\}}, \overline{R}_{2;\{1,3\}} + \overline{R}_{3;1}\right) \\ f_{\{1\}}(p_{r;3}) &\geq \max\left(\overline{R}_{3;\{1,2\}}, \frac{p_{r;3}}{p_{r;2}}\overline{R}_{2;\{1,3\}} + \overline{R}_{3;1}\right), \end{aligned} \tag{110}$$

where each term of the maximum operation corresponds to $T = \{1,2\}$ or $T = \{1,3\}$, respectively. Lemma 8 then implies that (110) can be strictly strengthened and replaced by the following inequality

$$f_{\{1\}}(p_{r;3}) \geq \max\left(\frac{p_{r;3}}{p_{r;2}}\left(\overline{R}_{2;1} + \overline{R}_{3;\{1,2\}}\right), \frac{p_{r;3}}{p_{r;2}}\overline{R}_{2;\{1,3\}} + \overline{R}_{3;1}\right).$$

Case 6: $S = \{2\}$. We now have two different choices of T: $T = \{1,2\}$ and $T = \{2,3\}$. By applying (8) for $T = \{1,2\}$ and for $T = \{2,3\}$, respectively, and by the discussion of Cases 2 and 4, we have

$$f_{\{2\}}(p_{r;1}) \geq \max\left(\overline{R}_{1;2} + \overline{R}_{3;\{1,2\}}, \overline{R}_{1;\{2,3\}} + \overline{R}_{3;2}\right)$$
$$f_{\{2\}}(p_{r;2}) \geq \max\left(\overline{R}_{3;\{1,2\}}, \frac{p_{r;2}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \overline{R}_{3;2}\right) \tag{111}$$
$$f_{\{2\}}(p_{r;3}) \geq \max\left(\overline{R}_{3;\{1,2\}}, \frac{p_{r;3}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \overline{R}_{3;2}\right), \tag{112}$$

where each term of the maximum operation corresponds to either $T = \{1,2\}$ or $T = \{2,3\}$, respectively. Lemma 8 then implies that (111) and (112) can be strictly strengthened and replaced by the following inequalities

$$\begin{aligned} f_{\{2\}}(p_{r;2}) \geq \\ \max\Bigg(&\frac{p_{r;2}}{p_{r;1}}\left(\overline{R}_{1;2} + \overline{R}_{3;\{1,2\}}\right), \frac{p_{r;2}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \overline{R}_{3;2}, \\ &\frac{p_{r;2} - p_{r;3}}{p_{r;1} - p_{r;3}}\left(\overline{R}_{1;2} + \overline{R}_{3;\{1,2\}}\right) \\ &\qquad + \frac{p_{r;1} - p_{r;2}}{p_{r;1} - p_{r;3}}\left(\frac{p_{r;3}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \overline{R}_{3;2}\right)\Bigg) \end{aligned}$$

$$f_{\{2\}}(p_{r;3}) \geq \max\left(\frac{p_{r;3}}{p_{r;1}}\left(\overline{R}_{1;2} + \overline{R}_{3;\{1,2\}}\right), \frac{p_{r;3}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \overline{R}_{3;2}\right).$$

Case 7: $S = \{3\}$. We now have two different choices of T: $T = \{1,3\}$ and $T = \{2,3\}$. By applying (8) for $T = \{1,3\}$ and for $T = \{2,3\}$, respectively, and by the discussion of Cases 3 and 4, we have

$$f_{\{3\}}(p_{r;1}) \geq \max\left(\overline{R}_{1;3} + \overline{R}_{2;\{1,3\}}, \overline{R}_{1;\{2,3\}} + \overline{R}_{2;3}\right)$$
$$f_{\{3\}}(p_{r;2}) \geq \max\left(\overline{R}_{2;\{1,3\}}, \frac{p_{r;2}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \overline{R}_{2;3}\right) \tag{113}$$
$$f_{\{3\}}(p_{r;3}) \geq \max\left(\frac{p_{r;3}}{p_{r;2}}\overline{R}_{2;\{1,3\}}, \frac{p_{r;3}}{p_{r;1}}\overline{R}_{1;\{2,3\}}\right). \tag{114}$$

where each term of the maximum operation corresponds to either $T = \{1,3\}$ or $T = \{2,3\}$, respectively. Lemma 8 then implies that (113) and (114) can be strictly strengthened and replaced by the following inequalities

$$f_{\{3\}}(p_{r;2}) \geq \max\left(\frac{p_{r;2}}{p_{r;1}}\left(\overline{R}_{1;3} + \overline{R}_{2;\{1,3\}}\right), \frac{p_{r;2}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \overline{R}_{2;3}\right)$$
$$f_{\{3\}}(p_{r;3}) \geq \max\left(\frac{p_{r;3}}{p_{r;1}}\left(\overline{R}_{1;3} + \overline{R}_{2;\{1,3\}}\right), \frac{p_{r;3}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \frac{p_{r;3}}{p_{r;2}}\overline{R}_{2;3}\right).$$

Case 8: $S = \emptyset$. We now have three different choices of T: $T = \{1\}$, $\{2\}$, and $\{3\}$. By applying (8) for all three choices, respectively, we have

$$\begin{aligned} f_{\emptyset}(p_{r;1}) &\geq \max\big(f_{\{1\}}(p_{r;1}) + R_1, f_{\{2\}}(p_{r;1}) + R_2, \\ &\qquad f_{\{3\}}(p_{r;1}) + R_3\big) \\ f_{\emptyset}(p_{r;2}) &\geq \max\big(f_{\{1\}}(p_{r;2}), f_{\{2\}}(p_{r;2}) + R_2, \\ &\qquad f_{\{3\}}(p_{r;2}) + R_3\big) \\ f_{\emptyset}(p_{r;3}) &\geq \max\left(f_{\{1\}}(p_{r;3}), f_{\{2\}}(p_{r;3}), f_{\{3\}}(p_{r;3}) + R_3\right). \end{aligned} \tag{115}$$

By the discussion of Cases 5 to 7 and by Lemma 8, (115) can be strictly strengthened and replaced by the following inequality

$$\begin{aligned} f_{\emptyset}(p_{r;3}) \geq \\ \max\Bigg(&\frac{p_{r;3}}{p_{r;1}}\left(R_1 + \overline{R}_{2;1} + \overline{R}_{3;\{1,2\}}\right), \\ &\frac{p_{r;3}}{p_{r;1}}\left(R_1 + \overline{R}_{2;\{1,3\}} + \overline{R}_{3;1}\right), \\ &\frac{p_{r;3}}{p_{r;2}}\left(R_2 + \frac{p_{r;2}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \overline{R}_{3;2}\right), \\ &\frac{p_{r;3}}{p_{r;2}}\left(R_2 + \frac{p_{r;2}}{p_{r;1}}\left(\overline{R}_{1;2} + \overline{R}_{3;\{1,2\}}\right)\right), \\ &\frac{p_{r;3}}{p_{r;2}}\Bigg(R_2 + \frac{p_{r;1} - p_{r;2}}{p_{r;1} - p_{r;3}}\left(\frac{p_{r;3}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \overline{R}_{3;2}\right) + \\ &\qquad \frac{p_{r;2} - p_{r;3}}{p_{r;1} - p_{r;3}}\left(\overline{R}_{1;2} + \overline{R}_{3;\{1,2\}}\right)\Bigg), \\ &\left(R_3 + \frac{p_{r;3}}{p_{r;1}}\left(\overline{R}_{1;3} + \overline{R}_{2;\{1,3\}}\right)\right), \\ &\left(R_3 + \frac{p_{r;3}}{p_{r;1}}\overline{R}_{1;\{2,3\}} + \frac{p_{r;3}}{p_{r;2}}\overline{R}_{2;3}\right)\Bigg). \end{aligned}$$

After further incorporating (7) and the inequality that $f_{\emptyset}(p_{r;3}) \leq p_{r;3}$, the proof of Corollary 3 is complete.

Appendix D
Proof of Lemma 4

In this appendix, we first prove that all Δ_1 to Δ_6 are non-negative for all four cases, respectively. We then prove that the inequalities $L_4 \geq 0$ and $\frac{\Delta_1+\Delta_5}{p_{r;1}} \geq \frac{\Delta_4+\Delta_5}{p_{r;3}}$ always hold.

Case 1: In this case, obviously all Δ_1 to Δ_6 are non-negative.

Case 2: In this case, since $\mathsf{Rk}_{1\to\bar{2}3} \geq \mathsf{Rk}_{3\to 1\bar{2}}$, obviously Δ_1, Δ_4, Δ_5, and Δ_6 are non-negative. By computing $\Delta_2 = n\mathsf{Rk}_{1\to\bar{2}3} - \Delta_1$ and $\Delta_3 = n\mathsf{Rk}_{3\to 1\bar{2}} - \Delta_4$, we also have

$$\Delta_2 = n\frac{p_{r;1}\mathsf{Rk}_{3\to 1\bar{2}} - p_{r;3}\mathsf{Rk}_{1\to\bar{2}3}}{p_{r;1} - p_{r;3}} = \Delta_3. \tag{116}$$

Since $\mathsf{Rk}_{3\to 1\bar{2}} \geq \frac{p_{r;3}}{p_{r;1}}\mathsf{Rk}_{1\to\bar{2}3}$, we must have $\Delta_2 = \Delta_3 \geq 0$.

Case 3: In this case, since $\frac{p_{r;3}}{p_{r;1}}\mathsf{Rk}_{1\to\bar{2}3} \geq \mathsf{Rk}_{3\to 1\bar{2}}$, obviously Δ_1, Δ_2, Δ_3, Δ_4, and Δ_5 are non-negative. By computing $\Delta_6 = n\mathsf{Rk}_{3\to\overline{12}} - \Delta_5$, we also have

$$\Delta_6 = n\frac{p_{r;1}\left(\mathsf{Rk}_{3\to 1\bar{2}} + \mathsf{Rk}_{3\to\overline{12}}\right) - p_{r;3}\left(\mathsf{Rk}_{1\to\bar{2}3} + \mathsf{Rk}_{3\to\overline{12}}\right)}{p_{r;1} - p_{r;3}}. \tag{117}$$

By condition (40) of Case 3, we must have $\Delta_6 \geq 0$.

Case 4: In this case, obviously Δ_1 to Δ_6 are non-negative.

To prove $L_4 \geq 0$, we will prove that $L_1 + L_2 + L_3 \leq n$. Note that $\Delta_2 \geq 0$ implies that $\Delta_1 \leq n\mathsf{Rk}_{1\to\bar{2}3}$, and $\Delta_6 \geq 0$ implies that $\Delta_5 \leq n\mathsf{Rk}_{3\to\overline{12}}$. We thus will prove that

$$\frac{L_1 + L_2 + L_3}{n} \leq \frac{\mathsf{Rk}_{1\to\overline{23}}}{p_{r;1}} + \frac{\mathsf{Rk}_{1\to\bar{2}3} + \mathsf{Rk}_{3\to\overline{12}}}{p_{r;1}} + \max\left(\frac{\mathsf{Rk}_{2\to\overline{13}} + \mathsf{Rk}_{2\to 1\bar{3}}}{p_{r;2}}, \frac{\mathsf{Rk}_{1\to 2\bar{3}} + \mathsf{Rk}_{2\to\overline{13}}}{p_{r;1}}\right) \leq 1.$$

Equivalently, we will prove that the following two inequalities hold simultaneously.

$$\frac{\mathsf{Rk}_{1\to\overline{23}} + \mathsf{Rk}_{1\to\bar{2}3} + \mathsf{Rk}_{3\to\overline{12}}}{p_{r;1}} + \frac{\mathsf{Rk}_{2\to\overline{13}} + \mathsf{Rk}_{2\to 1\bar{3}}}{p_{r;2}} \leq 1 \tag{118}$$

$$\frac{\mathsf{Rk}_{1\to\overline{23}} + \mathsf{Rk}_{1\to\bar{2}3} + \mathsf{Rk}_{3\to\overline{12}} + \mathsf{Rk}_{1\to 2\bar{3}} + \mathsf{Rk}_{2\to\overline{13}}}{p_{r;1}} \leq 1. \tag{119}$$

Using (35) to (38), (118) is a straightforward result of (26) (using the second term of the max operation), and (119) is a straightforward result of (25) (using any of the two terms of the max operation).

Inequality (42) can be proven by plugging the values of Δ_1, Δ_4, Δ_5 for different cases. In Case 1, $\Delta_1 = \Delta_4 = \Delta_5 = 0$ and (42) is satisfied. In Cases 2 and 3, plugging the values gives us

$$\frac{\Delta_1 + \Delta_5}{p_{r;1}} = \frac{\Delta_4 + \Delta_5}{p_{r;3}}.$$

In Case 4, (42) is guaranteed due to the condition of Case 4.

The proof of Lemma 4 is thus complete.

Appendix E
Decoding at d_2 for the Case of $M = 3$

By noticing $L_1+L_2+L_3+L_4 = n$, we can rewrite $p_{r;2}(L_3+L_4) - \mathsf{nc}_{\mathrm{G2}}$ by

$$\begin{aligned} &p_{r;2}(L_3 + L_4) - \mathsf{nc}_{\mathrm{G2}} \\ &= n \cdot \left(p_{r;2} - \frac{p_{r;2}}{p_{r;1}}\mathsf{Rk}_{1\to\overline{23}} - \mathsf{Rk}_{2\to\overline{13}} - \mathsf{Rk}_{2\to 1\bar{3}}\right) \\ &\quad - \frac{p_{r;2}}{p_{r;1}}(\Delta_1 + \Delta_5). \end{aligned} \tag{120}$$

To prove (58), we will prove an equivalent statement that (120) is no less than (56). Since the values of Δ_1 to Δ_6 are chosen depending on the values of $\mathsf{Rk}_{1\to\bar{2}3}$, $\mathsf{Rk}_{3\to 1\bar{2}}$, and $\mathsf{Rk}_{3\to\overline{12}}$ (Cases 1 to 4), we will establish the inequality for each case separately.

Case 1: $\mathsf{Rk}_{3\to 1\bar{2}} \geq \mathsf{Rk}_{1\to\bar{2}3}$. In this case, we have $\Delta_1 = \Delta_4 = \Delta_5 = 0$ and $\Delta_2 = n\mathsf{Rk}_{1\to\bar{2}3}$, $\Delta_3 = n\mathsf{Rk}_{3\to 1\bar{2}}$, and $\Delta_6 = n\mathsf{Rk}_{3\to\overline{12}}$. Therefore (120) becomes

$$n \cdot \left(p_{r;2} - \frac{p_{r;2}}{p_{r;1}}\mathsf{Rk}_{1\to\overline{23}} - \mathsf{Rk}_{2\to\overline{13}} - \mathsf{Rk}_{2\to 1\bar{3}}\right),$$

and (56) becomes

$$n \cdot \left(\mathsf{Rk}_{2\to\bar{1}3} + \mathsf{Rk}_{2\to 13} + \mathsf{Rk}_{3\to 1\bar{2}} + \mathsf{Rk}_{3\to\overline{12}}\right).$$

The inequality (120) $\geq$ (56) is thus established by (35) to (38) and by the first term of the max operation in (26) of Corollary 3.

Case 2: $\mathsf{Rk}_{1\to\bar{2}3} \geq \mathsf{Rk}_{3\to 1\bar{2}} \geq \frac{p_{r;3}}{p_{r;1}}\mathsf{Rk}_{1\to\bar{2}3}$. In this case, the Δ_1 to Δ_5 values are decided in (39) and (116), and $\Delta_6 = \mathsf{Rk}_{3\to\overline{12}}$. Therefore (120) becomes

$$n \cdot \left(p_{r;2} - \frac{p_{r;2}}{p_{r;1}}\mathsf{Rk}_{1\to\overline{23}} - \mathsf{Rk}_{2\to\overline{13}} - \mathsf{Rk}_{2\to 1\bar{3}} - \frac{p_{r;2}(\mathsf{Rk}_{1\to\bar{2}3} - \mathsf{Rk}_{3\to 1\bar{2}})}{p_{r;1} - p_{r;3}}\right),$$

and (56) becomes

$$n \cdot \left(\mathsf{Rk}_{2\to\bar{1}3} + \mathsf{Rk}_{2\to 13} + \frac{p_{r;1}\mathsf{Rk}_{3\to 1\bar{2}} - p_{r;3}\mathsf{Rk}_{1\to\bar{2}3}}{p_{r;1} - p_{r;3}} + \mathsf{Rk}_{3\to\overline{12}}\right).$$

The inequality (120) $\geq$ (56) is thus established by (35) to (38) and by the third term of the max operation in (26) of Corollary 3.

Case 3: $\frac{p_{r;3}}{p_{r;1}}\mathsf{Rk}_{1\to\bar{2}3} \geq \mathsf{Rk}_{3\to 1\bar{2}}$ and

$$\mathsf{Rk}_{3\to 1\bar{2}} + \mathsf{Rk}_{3\to\overline{12}} \geq \frac{p_{r;3}}{p_{r;1}}\left(\mathsf{Rk}_{1\to\bar{2}3} + \mathsf{Rk}_{3\to\overline{12}}\right).$$

In this case, the values of Δ_1, Δ_4, Δ_5, and Δ_6 values are decided in (41) and (117), and $\Delta_2 = \Delta_3 = 0$. Therefore (120) becomes

$$n \cdot \left(p_{r;2} - \frac{p_{r;2}}{p_{r;1}}\mathsf{Rk}_{1\to\overline{23}} - \mathsf{Rk}_{2\to\overline{13}} - \mathsf{Rk}_{2\to 1\bar{3}} - \frac{p_{r;2}}{p_{r;1}}\left(\mathsf{Rk}_{1\to\bar{2}3} + \frac{p_{r;3}\mathsf{Rk}_{1\to\bar{2}3} - p_{r;1}\mathsf{Rk}_{3\to 1\bar{2}}}{p_{r;1} - p_{r;3}}\right)\right),$$

and (56) becomes

$$n \cdot \left(\mathsf{Rk}_{2\to\overline{1}3} + \mathsf{Rk}_{2\to 13} + \frac{p_{r;1}(\mathsf{Rk}_{3\to 1\overline{2}} + \mathsf{Rk}_{3\to\overline{12}}) - p_{r;3}(\mathsf{Rk}_{1\to\overline{2}3} + \mathsf{Rk}_{3\to\overline{12}})}{p_{r;1} - p_{r;3}}\right).$$

The inequality (120) $\geq$ (56) is thus established by (35) to (38) and by the third term of the max operation in (26) of Corollary 3.

Case 4: $\frac{p_{r;3}}{p_{r;1}}(\mathsf{Rk}_{1\to\overline{2}3} + \mathsf{Rk}_{3\to\overline{12}}) \geq \mathsf{Rk}_{3\to 1\overline{2}} + \mathsf{Rk}_{3\to\overline{12}}$. In this case, we have $\Delta_1 = n\mathsf{Rk}_{1\to\overline{2}3}$, $\Delta_4 = n\mathsf{Rk}_{3\to 1\overline{2}}$, $\Delta_5 = n\mathsf{Rk}_{3\to\overline{12}}$, and $\Delta_2 = \Delta_3 = \Delta_6 = 0$. Therefore (120) becomes

$$n \cdot \left(p_{r;2} - \frac{p_{r;2}}{p_{r;1}}\mathsf{Rk}_{1\to\overline{23}} - \mathsf{Rk}_{2\to\overline{13}} - \mathsf{Rk}_{2\to 1\overline{3}} - \frac{p_{r;2}}{p_{r;1}}(\mathsf{Rk}_{1\to\overline{2}3} + \mathsf{Rk}_{3\to\overline{12}})\right),$$

and (56) becomes $n \cdot (\mathsf{Rk}_{2\to\overline{1}3} + \mathsf{Rk}_{2\to 13})$. The inequality (120) $\geq$ (56) is thus established by (35) to (38) and by the second term of the max operation in (26) of Corollary 3.

APPENDIX F
PROOF OF LEMMA 5

Proof: In Step 2, we first notice that for any z values that satisfy (16) with strict inequality, we can repeatedly reduce the values of any positive z variables until (16) becomes an equality. The newly reduced z variables will still satisfy (14) and (15). Therefore, we can replace inequality (16) by an equality without loss of generality.

For any $i \in [M], T_1 \subseteq [M]\backslash i, T_2 = [M]\backslash(T_1 \cup \{i\})$, consider the following set of coding vectors:

$$\left\{\mathbf{c}_i(t) : \forall t \text{ satisfying } \forall j \in [M]\backslash i, 1_{\{Z_{i\to j}(t)\neq *\}} = 1_{\{j\in T_1\}}\right\},$$

i.e., the set of coding vectors that are overheard *by and only by* users in T_1. Since z_{i,T_2} satisfies (15), we can construct the subspace $\mathcal{V}_{i\to T_1\overline{T_2}}$ as the linear span of the first $(n\cdot z_{i,T_2})$ vectors from the above set of coding vectors. By this construction, the MDS property of RLNC, and by (16), properties (i), (ii), and (iii) of $\mathcal{V}_{i\to T_1\overline{T_2}}$ are satisfied. The proof is complete. ■

APPENDIX G
PROOFS OF LEMMA 6 AND COROLLARY 4

Proof of Lemma 6: Since $S_1^\pi = [M]$ for any permutation π, we have $x_{\pi,[M],h} = x_{\pi,S_1^\pi,h}$. We also notice that $x_{\pi,S_1^\pi,h}$ only participates in (63) when $j = 1$. Since $x_{\pi,S_1^\pi,h}$ is on the right-hand side of (63) when $j = 1$, for any $x_{\pi,S_1^\pi,h}$ that is strictly positive, we can reset its value to zero while the new x values still satisfy (63). Therefore, we can set any strictly positive $x_{\pi,[M],h}$ value to zero without loss of generality. The first part of Lemma 6 is proven.

To prove the second part of the lemma, we use the results in Lemma 1. Recall that we assume $p_{r;1} > p_{r;2} > \cdots > p_{r;M} > 0$. For any ZPPLCIF $f_{\pi,S_j^\pi}(\cdot)$ that corresponds to $\{x_{\pi,S_j^\pi,h} : \forall h \in [M]\}$ through (28), we first notice that $\sum_{h=1}^k x_{\pi,S_j^\pi,h}$ is the first order derivative $\frac{d}{dp}f_{\pi,S_j^\pi}(p)$ for all $p \in (p_{r;k+1}, p_{r;k})$, using the convention $p_{r;M+1} = 0$. We then prove the second part by contradiction.

Assume that (68) does not hold for some (π, k_0, j). Then for such (π, k_0, j), we have

$$\forall p \in (p_{r;k_0+1}, p_{r;k_0}),\ \frac{d}{dp}f_{\pi,S_{j+1}^\pi}(p) < \frac{d}{dp}f_{\pi,S_j^\pi}(p). \tag{121}$$

Define

$$k^* = \inf\{i \in \{k_0+1, k_0+2, \cdots, M\} : \frac{d}{dp}f_{\pi,S_{j+1}^\pi}(p_{r;i}^+) \neq \frac{d}{dp}f_{\pi,S_{j+1}^\pi}(p_{r;i}^-)\},$$

where $p_{r;i}^- \triangleq p_{r;i} - \epsilon$ and $p_{r;i}^+ \triangleq p_{r;i} + \epsilon$ for some sufficiently small $\epsilon > 0$ are the left and the right limits of $p_{r;i}$, respectively. Namely, $p_{r;k^*}$ is the largest $p \in (0, p_{r;k_0})$ at which the slope of $f_{\pi,S_{j+1}^\pi}(\cdot)$ changes. There are two cases to discuss depending on whether such k^* exists or not.

Case 1: k^* exists. In this case, since $f_{\pi,S_{j+1}^\pi}(\cdot)$ is a ZPPLCIF, we must have

$$\frac{d}{dp}f_{\pi,S_{j+1}^\pi}(p_{r;k^*}^+) < \frac{d}{dp}f_{\pi,S_{j+1}^\pi}(p_{r;k^*}^-). \tag{122}$$

Without loss of generality, we can also assume that

$$f_{\pi,S_{j+1}^\pi}(p_{r;k^*}) = f_{\pi,S_j^\pi}(p_{r;k^*}) + \max_{\forall k\in[M]:k\geq k^*, k\notin S_{j+1}^\pi}\left\{\sum_{\{\forall T\in 2^{[M]}:T\subseteq S_{j+1}^\pi\}} y_{k,T,(\pi,S_{j+1}^\pi)}\right\}, \tag{123}$$

i.e., (13) holds with equality (using one of the k values) when evaluated at $p^* \triangleq p_{r;k^*}$. The reason behind this assumption is as follows. For any $\epsilon > 0$, we define a function $g_\epsilon(\cdot)$ by setting $g_\epsilon(p) = f_{\pi,S_{j+1}^\pi}(p)$ for all $p \in \{1, p_{r;1}, \cdots, p_{r;M}, 0\}\backslash p_{r;k^*}$ and setting $g_\epsilon(p_{r;k^*}) = f_{\pi,S_{j+1}^\pi}(p_{r;k^*}) - \epsilon$, and by interpolating $g_\epsilon(p)$ values for all p values not in $\{1, p_{r;1}, \cdots, p_{r;M}, 0\}$. Suppose the equality (123) does not hold. Then by (122), there exists a sufficiently small $\epsilon^* > 0$ such that

$$\begin{aligned} g_{\epsilon^*}(p_{r;k^*}) &= f_{\pi,S_{j+1}^\pi}(p_{r;k^*}) - \epsilon^* \\ &\geq f_{\pi,S_j^\pi}(p_{r;k^*}) + \max_{\forall k\in[M]:k\geq k^*, k\notin S_{j+1}^\pi}\left\{\sum_{\{\forall T\in 2^{[M]}:T\subseteq S_{j+1}^\pi\}} y_{k,T,(\pi,S_{j+1}^\pi)}\right\}, \end{aligned}$$

and

$$\begin{aligned} \frac{d}{dp}f_{\pi,S_{j+1}^\pi}(p_{r;k^*}^+) &< \frac{d}{dp}g_{\epsilon^*}(p_{r;k^*}^+) \\ &\leq \frac{d}{dp}g_{\epsilon^*}(p_{r;k^*}^-) < \frac{d}{dp}f_{\pi,S_{j+1}^\pi}(p_{r;k^*}^-). \end{aligned}$$

Therefore $g_{\epsilon^*}(\cdot)$ is also a ZPPLCIF. Note that $g_{\epsilon^*}(\cdot)$ is point-wise no larger than $f_{\pi,S_{j+1}^\pi}(\cdot)$ for all $p \in [0,1]$; is strictly smaller than $f_{\pi,S_{j+1}^\pi}(\cdot)$ when evaluated at $p_{r;k^*}$; and satisfies the inequality (13). We can thus replace $f_{\pi,S_{j+1}^\pi}(\cdot)$ by the newly constructed $g_{\epsilon^*}(\cdot)$ while still satisfying all the inequalities (13). By repeatedly applying the above construction of replacing $f_{\pi,S_{j+1}^\pi}(\cdot)$ with a point-wise smaller $g_{\epsilon^*}(\cdot)$, we have thus shown that (123) must hold when taking the point-wise limit of the functions of the above repeated construction. As a result, we can assume that (123) holds without loss of generality.

Now we compare the two terms $(f_{\pi,S^\pi_{j+1}}(p_{r;k^*}) - f_{\pi,S^\pi_j}(p_{r;k^*}))$ versus $(f_{\pi,S^\pi_{j+1}}(p_{r;k_0}) - f_{\pi,S^\pi_j}(p_{r;k_0}))$. We first note that by our construction, the slope of $f_{\pi,S^\pi_{j+1}}(\cdot)$ remains unchanged for all $p \in (p_{r;k^*}, p_{r;k_0})$. Since $f_{\pi,S^\pi_j}(\cdot)$ is a ZPPLCIF, we must also have

$$\begin{aligned}\frac{d}{dp} f_{\pi,S^\pi_j}(p) &\geq \frac{d}{dp} f_{\pi,S^\pi_j}(p^-_{r;k_0}) \\ &> \frac{d}{dp} f_{\pi,S^\pi_{j+1}}(p^-_{r;k_0}) = \frac{d}{dp} f_{\pi,S^\pi_{j+1}}(p)\end{aligned}$$

for all $p \in (p_{r;k^*}, p_{r;k_0})$, where the strict inequality follows from the assumption (121). As a result, we must have

$$\begin{aligned}& f_{\pi,S^\pi_{j+1}}(p_{r;k_0}) - f_{\pi,S^\pi_{j+1}}(p_{r;k^*}) \\ &\qquad < f_{\pi,S^\pi_j}(p_{r;k_0}) - f_{\pi,S^\pi_j}(p_{r;k^*}) \\ \Leftrightarrow & f_{\pi,S^\pi_{j+1}}(p_{r;k^*}) - f_{\pi,S^\pi_j}(p_{r;k^*}) \\ &\qquad > f_{\pi,S^\pi_{j+1}}(p_{r;k_0}) - f_{\pi,S^\pi_j}(p_{r;k_0}). \quad (124)\end{aligned}$$

On the other hand, by (123), we also have

$$\begin{aligned}& f_{\pi,S^\pi_{j+1}}(p_{r;k^*}) - f_{\pi,S^\pi_j}(p_{r;k^*}) \\ &= \max_{\forall k\in[M]: k\geq k^*, k\notin S^\pi_{j+1}} \left\{ \sum_{\{\forall T\in 2^{[M]}: T\subseteq S^\pi_{j+1}\}} y_{k,T,(\pi,S^\pi_{j+1})} \right\} \\ &\leq \max_{\forall k\in[M]: k\geq k_0, k\notin S^\pi_{j+1}} \left\{ \sum_{\{\forall T\in 2^{[M]}: T\subseteq S^\pi_{j+1}\}} y_{k,T,(\pi,S^\pi_{j+1})} \right\} \\ &\leq f_{\pi,S^\pi_{j+1}}(p_{r;k_0}) - f_{\pi,S^\pi_j}(p_{r;k_0}), \qquad (125)\end{aligned}$$

where the first inequality follows from changing the domain of the max operation from $k \geq k^*$ to $k \geq k_0$ and from the fact that $k^* > k_0$, and the second inequality follows from (13). The contradiction between (124) and (125) thus proves that (68) holds in Case 1.

Case 2: k^* does not exist. In this case, we have $\frac{d}{dp} f_{\pi,S^\pi_{j+1}}(p^-_{r;k_0}) = \frac{d}{dp} f_{\pi,S^\pi_{j+1}}(0^+)$. By noting that both $f_{\pi,S^\pi_{j+1}}(\cdot)$ and $f_{\pi,S^\pi_j}(\cdot)$ are ZPPLCIFs and $f_{\pi,S^\pi_{j+1}}(\cdot)$ is point-wise no smaller than $f_{\pi,S^\pi_j}(\cdot)$, we must also have

$$\frac{d}{dp} f_{\pi,S^\pi_{j+1}}(0^+) \geq \frac{d}{dp} f_{\pi,S^\pi_j}(0^+) \geq \frac{d}{dp} f_{\pi,S^\pi_j}(p^-_{r;k_0}).$$

Jointly we thus have

$$\frac{d}{dp} f_{\pi,S^\pi_{j+1}}(p^-_{r;k_0}) \geq \frac{d}{dp} f_{\pi,S^\pi_j}(p^-_{r;k_0}),$$

which contradicts the assumption (121). As a result, (68) holds in Case 2 as well. The proof of Lemma 6 is complete. ■

Proof of Corollary 4: We prove this corollary by explicit construction for each permutation π.

Consider a permutation π that is fixed throughout the proof. We will first construct a partition $\{\chi_{\pi,l} : \forall l\}$ and the corresponding label function $\{\theta_{\pi,l}(\cdot) : \forall l\}$ that satisfies (69) for $j = 1$. Then we will iteratively refine the partition until (69) is satisfied for all $j \in [M]$. More explicitly, we perform the following steps sequentially.

Step 1: We construct a partition of $(M+1)$ chunks in this step. For $l = 1$ to M, choose $(\chi_{\pi,l}, \theta_{\pi,l}(\cdot))$ as follows:

$$\begin{aligned}\chi_{\pi,l} &= x_{\pi,S^\pi_2,l} \qquad (126) \\ \theta_{\pi,l}(k) &= \begin{cases} l & \text{if } k = 1 \\ -1 & \text{if } k \in [M]\backslash 1 \end{cases}\end{aligned}$$

where the value -1 indicates that the value of $\theta_{\pi,l}(k)$ has not been assigned for $k \geq 2$. For $l = M+1$, choose

$$\begin{aligned}\chi_{\pi,l} &= \left(\sum_{h=1}^{M} x_{\pi,\emptyset,h}\right) - \left(\sum_{h=1}^{M} x_{\pi,S^\pi_2,h}\right) \qquad (127) \\ \theta_{\pi,l}(k) &= \begin{cases} M+1 & \text{if } k = 1 \\ -1 & \text{if } k \in [M]\backslash 1 \end{cases}.\end{aligned}$$

Eq. (68) of Lemma 6 guarantees the non-negativeness of $\chi_{\pi,M+1}$. After Step 1, the partition contains $(M+1)$ different chunks. It can be easily verified that (69) is satisfied for $j = 1$ and $\theta_{\pi,l}(\cdot)$ is non-increasing.

We perform the following "Step j_0" sequentially for $j_0 = 2, \cdots, M$. We use the induction assumption that after finishing the $(j_0 - 1)$-th step, the latest partition $\{\theta_{\pi,l}(\cdot) : \forall l\}$ satisfies (69) for all $j \leq j_0 - 1$. For Step j_0, perform the following routine:

Set $k \leftarrow 1$.
for $h = 1$ to M **do**
  Set $\Delta \leftarrow x_{\pi,S^\pi_{j_0+1},h}$.
  while $\Delta > 0$ **do**
    if there exists a chunk $(\chi_{\pi,l}, \theta_{\pi,l}(\cdot))$ satisfying $\theta_{\pi,l}(j_0 - 1) = k$ and $\theta_{\pi,l}(j_0) = -1$ **then**
      Choose one such chunk $(\chi_{\pi,l_0}, \theta_{\pi,l_0}(\cdot))$.
      if $\Delta \geq \chi_{\pi,l_0}$ **then**
        Set $\theta_{\pi,l_0}(j_0) \leftarrow h$ and $\Delta \leftarrow (\Delta - \chi_{\pi,l_0})$.
      else
        Replace the chosen chunk $(\chi_{\pi,l_0}, \theta_{\pi,l_0}(\cdot))$ by two new sub-chunks $(\chi_{\pi,l_1}, \theta_{\pi,l_1}(\cdot))$ and $(\chi_{\pi,l_2}, \theta_{\pi,l_2}(\cdot))$ such that

$$\begin{aligned}\chi_{\pi,l_1} &= \Delta \\ \theta_{\pi,l_1}(k) &= \theta_{\pi,l_0}(k),\ \forall k \in [M]\end{aligned}$$

        and

$$\begin{aligned}\chi_{\pi,l_2} &= \chi_{\pi,l_0} - \Delta \\ \theta_{\pi,l_2}(k) &= \theta_{\pi,l_0}(k),\ \forall k \in [M].\end{aligned}$$

        Set $\theta_{\pi,l_1}(j_0) \leftarrow h$ and $\Delta \leftarrow 0$.
      end if
    else
      Set $k \leftarrow k + 1$.
    end if
  end while
end for
for all chunks $(\chi_{\pi,l}, \theta_{\pi,l}(\cdot))$ satisfying $\theta_{\pi,l}(j_0) = -1$ **do**
  Set $\theta_{\pi,l}(j_0) \leftarrow M + 1$.
end for

The correctness of the above algorithm is proven as follows. In the beginning when $h = 1$, we notice that all chunks have been assigned in Step (j_0-1) and thus have $\theta_{\pi,l}(j_0-1) \neq -1$. Therefore, all chunks have $\theta_{\pi,l}(j_0-1) \geq 1$. When assigned during the case of $h = 1$, and their newly assigned value is $\theta_{\pi,l}(j_0) = h = 1$. Therefore, the label function $\theta_{\pi,l}(\cdot)$ remains non-increasing when $h = 1$. Note that Lines 7 to 12 also guarantee that when the WHILE loop moves to $h = 2$ for the first time, the equality of (69) is satisfied for $(j, h) = (j_0, 1)$.

Consider the scenario in which we move to $h = h_0$ for some $h_0 \in [2, M]$ for the first time and assume by induction that the equality of (69) holds for all (j, h) satisfying either (i) $j = j_0$ and $h \leq h_0 - 1$, or (ii) $j \leq j_0 - 1$ and $h \in [M]$. Eq. (68) of Lemma 6 and the induction assumption then guarantees that

$$\sum_{k'=1}^{h_0-1}\left(\sum_{\forall l:\theta_{\pi,l}(j_0-1)=k'}\chi_{\pi,l}\right) = \sum_{k'=1}^{h_0-1} x_{\pi,S^{\pi}_{j_0},k'} \leq \sum_{k'=1}^{h_0-1} x_{\pi,S^{\pi}_{j_0+1},k'}.$$

Therefore, when we move to $h = h_0$ for the first time, we must have exhaustively processed all chunks with $\theta_{\pi,l}(j_0-1) = k'$, $k' = 1, \cdots, h_0 - 1$. Hence we must have $k \geq h_0$. As a result, all the $\theta_{\pi,l}$ functions that are assigned during the period of $h = h_0$ must have $\theta_{\pi,l}(j_0-1) = k \geq h_0$ and thus remain non-increasing since their newly assigned value is $\theta_{\pi,l}(j_0) = h_0$. Suppose $h_0 < M$. Lines 7 to 12 also guarantee that when we move to $h = h_0 + 1$ for the first time, (69) is satisfied for $j = j_0$ and $h = h_0$.

Consider the scenario when we finish the last iteration of the FOR loop with $h = M$. By (126) and (127), the total sum of all $\chi_{\pi,l}$ is

$$\sum_{k'=1}^{M+1}\left(\sum_{\forall l:\theta_{\pi,l}(j_0-1)=k'}\chi_{\pi,l}\right) = \sum_{k'=1}^{M+1}\left(\sum_{\forall l:\theta_{\pi,l}(1)=k'}\chi_{\pi,l}\right) = \sum_{k'=1}^{M} x_{\pi,\emptyset,k'} \geq \sum_{k'=1}^{M} x_{\pi,S^{\pi}_{j_0+1},k'}, \tag{128}$$

where the inequality follows from repeatedly using (68). Therefore, when $\Delta = 0$ for the iteration of $h = M$, the k value is still bounded within $[M + 1]$. On the other hand, by the induction assumption, we also have

$$\sum_{k'=1}^{M}\left(\sum_{\forall l:\theta_{\pi,l}(j_0-1)=k'}\chi_{\pi,l}\right) = \sum_{k'=1}^{M} x_{\pi,S^{\pi}_{j_0},k'} \leq \sum_{k'=1}^{M} x_{\pi,S^{\pi}_{j_0+1},k'}.$$

As a result, when we finish the last iteration (with $h = M$ and $\Delta = 0$), we have processed all chunks $(\chi_{\pi,l}, \theta_{\pi,l}(\cdot))$ with $\theta_{\pi,l}(j_0-1) \leq M$. Therefore, when we move to the "clean-up" stage: Lines 18 to 20, all the remaining chunks have $\theta_{\pi,l}(j_0 - 1) = M + 1$. Therefore the non-increasing property of $\theta_{\pi,l}(\cdot)$ still holds after the clean-up stage. Lines 7 to 12 then guarantee that (69) is satisfied for all $j \leq j_0$ and $h \in [M]$.

When $j_0 = M$, (128) becomes an equality. Therefore, when we have finished the last iteration (with $h = M$ and $\Delta = 0$) of $j_0 = M$, we must have processed all the chunks. Line 19 will not be executed. As a result, $\theta_{\pi,l}(M) \leq M$ for all π and l. The above construction of $\{\chi_{\pi,l} : \forall l\}$ completes the proof of Corollary 4. ■

Appendix H
A More General Non-Canonical Setting

In this appendix, we generalize the canonical setting in Section II to take into account both opportunistic routing and cross-layer optimization, which further enlarge the capacity of erasure-channel-based 2-staged INC.

We consider the following 1-hop relay network, which consists of M source/destination pairs (s_i, d_i), $\forall i \in [M]$ and a common relay node r. Source s_i would like to send nR_i information packets $X_{i,1}$ to X_{i,nR_i} to destination d_i through the common relay r for some positive integer n and fractional rate R_i. Each s_i is associated with a 1-to-$(M + 1)$ PEC that sends a symbol $W_i \in \mathsf{GF}(q)$ from s_i to $\{d_j : \forall j \in [M]\} \cup \{r\}$. Namely, we allow direct communication between s_i and d_i. For each s_i, the outputs of its 1-to-$(M + 1)$ PEC are denoted by $Z_{i\to j}$, $j \in [M]$ and by $Z_{i\to r}$, which refer to the outputs arriving at destination d_j and the relay r, respectively. The common relay r is associated with a 1-to-M PEC that sends a symbol $W_r \in \mathsf{GF}(q)$ from r to $\{d_j : \forall j \in [M]\}$, and the corresponding outputs are denoted by $Z_{r\to j}$ for all $j \in [M]$.

We also assume that the sources s_i and the relay r share the time resources (potentially due to underlying interference constraints). More specifically, consider a set of $(M + 1)$ time allocation variables $(\beta_1, \cdots, \beta_M, \beta_r)$ that satisfy $\beta_r + \sum_{i=1}^{M}\beta_i = M+1$. Each of the sources s_1 to s_M and the relay r can use their associated PECs for exactly $n\beta_i$ and $n\beta_r$ times. We use the following vector notation to denote the $(M+1)n$ transmitted symbols and their corresponding outputs. For all $i \in [M]$, we define

$$\begin{aligned}\mathbf{W}_i &\triangleq (W_i(1), W_i(2), \cdots, W_i(n\beta_i)).\\ \mathbf{Z}_{i\to r} &\triangleq (Z_{i\to r}(1), Z_{i\to r}(2), \cdots, Z_{i\to r}(n\beta_i))\\ \mathbf{W}_r &\triangleq (W_r(1), W_r(2), \cdots, W_r(n\beta_r)).\\ \mathbf{Z}_{r\to i} &\triangleq (Z_{r\to i}(1), Z_{r\to i}(2), \cdots, Z_{r\to i}(n\beta_r)),\end{aligned}$$

and for all $i, j \in [M]$, we define

$$\mathbf{Z}_{i\to j} \triangleq (Z_{i\to j}(1), Z_{i\to j}(2), \cdots, Z_{i\to j}(n\beta_i)),$$

for which we use the input argument "(t)", to distinguish each channel usage. All the PECs are memoryless and stationary and described by the joint success probabilities $p_{r\to\cdot}$ and $p_{s_i\to\cdot}$ for all $i \in [M]$ as discussed in Section II-A. The channel statistics $p_{r\to\cdot}$ and $p_{s_i\to\cdot}$ are assumed to be known to all network nodes. An illustration of this general 1-hop relay network with direct path is illustrated in Fig. 9 for the case of $M = 2$.

The *channel status vector* of the PEC outputs $\mathbf{Z}_{i\to j}$ is defined by

$$\mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}} \triangleq (1_{\{Z_{i\to j}(1)\neq *\}}, \cdots, 1_{\{Z_{i\to j}(n\beta_i)\neq *\}}).$$

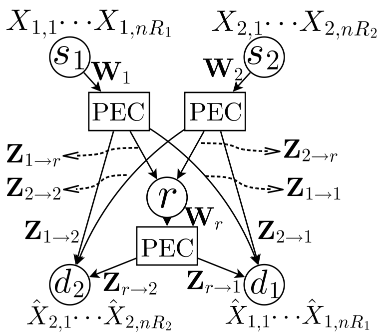


Fig. 9. Illustration of the general cross-layer-optimization-enabled 1-hop relay network with direct communication between the source and destination (s_i, d_i) for the case of $M = 2$.

A 2-staged 1-hop intersession network code is then defined by $(M+1)$ encoding functions $g_{s_i}(\cdot)$ and $g_r(\cdot)$:

$$\forall i \in [M],\ \mathbf{W}_i = g_{s_i}(\mathbf{X}_i)$$
$$\mathbf{W}_r = g_r(\{\mathbf{Z}_{i\to r}, \mathbf{1}_{\{Z_{i\to j}(\cdot)\neq *\}} : \forall i, j \in [M]\}),$$

and M decoding functions $g_{d_j}(\cdot)$:

$$\forall j \in [M],\ \hat{\mathbf{X}}_j = g_{d_j}(\{\mathbf{Z}_{i\to j}, \mathbf{Z}_{r\to j} : \forall i \in [M]\}).$$

The encoding and decoding functions can be linear or nonlinear.

Assume that the information symbols $\mathbf{X}_i = (X_{i,1}, \cdots, X_{i,nR_i})$ are uniformly and independently distributed over $\mathsf{GF}(q)$ for all $i \in [M]$. We define the achievability of any 2-staged 1-hop intersession network code as follows.

Definition 6: Given the values of the time allocation variables $(\beta_1, \cdots, \beta_M, \beta_r)$, a rate vector $(R_1, \cdots, R_M)$ is achievable if for any $\epsilon > 0$, there exists a 1-hop intersession network code satisfying

$$\forall i \in [M],\ \mathsf{P}\left(\hat{\mathbf{X}}_i \neq \mathbf{X}_i\right) < \epsilon,$$

for sufficiently large n and sufficiently large underlying finite field $\mathsf{GF}(q)$.

Definition 7: The capacity region of a 2-staged 1-hop relay network of M sessions is defined as the closure of all achievable rate vectors $(R_1, \cdots, R_M)$ of all possible assignment of the time allocation variables $(\beta_1, \cdots, \beta_M, \beta_r)$.

Note that the previous setting in Section II is a special case of the above new setting by choosing the success probabilities $p_{s_i \to \cdot}$ with zero marginal probabilities $p_{s_i;d_i} = 0$ for all $i \in [M]$, and by considering only uniform time allocation variables $(\beta_1, \cdots, \beta_M, \beta_r) = (1, \cdots, 1)$. The outer and inner bounds in Section III can then be generalized as follows, for which we assume that $p_{r;1} > p_{r;2} > \cdots > p_{r;M} > 0$.

Proposition 7: A rate vector $(R_1, \cdots, R_M)$ is in the capacity region only if there exist $(M+1)$ time allocation variables $(\beta_1, \cdots, \beta_M, \beta_r)$ with the total sum being $(M+1)$, and 2^M ZPPLCIFs $\left\{f_S(\cdot) : \forall S \in 2^{[M]}\right\}$ indexed by $S \in 2^{[M]}$ such that the following $(M + M^2 2^{M-1} + 1)$ inequalities are satisfied:

$$\forall i \in [M],\ R_i \leq \beta_i \cdot p_{s_i;\cup\{d_i,r\}} \tag{129}$$

$\forall i, k \in [M], \forall S, T \in 2^{[M]}$ satisfying $k \notin S$ and $T = S \cup \{k\}$,

$$f_S(p_{r;i}) \geq f_T(p_{r;i}) + 1_{\{i\leq k\}}\left(R_k - \beta_k p_{s_k;\cup T}\right)^+ \tag{130}$$

and $f_\emptyset(p_{r;M}) \leq \beta_r \cdot p_{r;M}$. $\quad$ (131)

Proposition 8: A rate vector $(R_1, \cdots, R_M)$ can be achieved by a *linear network code* if there exist $(M+1)$ time allocation variables $(\beta_1, \cdots, \beta_M, \beta_r)$ with the total sum being $(M+1)$, $M!(2^{M+1} - M - 2)$ non-negative variables:

$$\Big\{y_{k,T,(\pi,S_j^\pi)} \geq 0 : \forall k \in [M], \forall T \in 2^{[M]}, \forall \pi, \forall j \in [M+1],$$
$$\text{satisfying } k \notin S_j^\pi \text{ and } T \subseteq S_j^\pi\Big\},$$

and $M2^{M-1}$ non-negative variables:

$$\left\{z_{k,T} \geq 0 : \forall T \in 2^{[M]}, \forall k \in [M]\backslash T\right\},$$

and $(M+1)!$ ZPPLCIFs $\left\{f_{\pi,S_j^\pi} : \forall \pi, \forall j \in [M+1]\right\}$, such that jointly the following $(M + M!M^3 + M2^{M-1} + M2^{M-1} + M + 1)$ inequalities are satisfied:

$$\forall i \in [M],\ R_i < \beta_i \cdot p_{s_i;\cup\{d_i,r\}} \tag{132}$$

$$\forall \pi, \forall i, j, k \in [M],\quad f_{\pi,S_{j+1}^\pi}(p_{r;i}) \geq f_{\pi,S_j^\pi}(p_{r;i}) + 1_{\{i\leq k, k\notin S_{j+1}^\pi\}} \sum_{\{\forall T \in 2^{[M]} : T \subseteq S_{j+1}^\pi\}} y_{k,T,(\pi,S_{j+1}^\pi)} \tag{133}$$

$\forall k \in [M], \forall T \in 2^{[M]}, k \notin T,$

$$\sum_{\{\forall \pi, \forall j \in [M+1] : S_j^\pi \supseteq T, k \notin S_j^\pi\}} y_{k,T,(\pi,S_j^\pi)} \geq z_{k,T} \tag{134}$$

$\forall S \in 2^{[M]}, \forall k \in [M]\backslash S, T = S \cup \{k\},$

$$z_{k,S} \leq \beta_k \cdot p_{s_k \to (([M]\cup\{r\})\backslash T)\overline{T}} \tag{135}$$

$$\forall k \in [M],\quad \sum_{\forall S \in 2^{[M]} : k \notin S} z_{k,S} \geq (R_k - \beta_k \cdot p_{s_k;d_k})^+ \tag{136}$$

$$\sum_{\forall \pi} f_{\pi,\emptyset}(p_{r;M}) < \beta_r \cdot p_{r;M}. \tag{137}$$

Both Propositions 7 and 8 can be expressed as a LP feasibility problem with the help of Lemma 1, in a similar way as in Propositions 5 and 6. The proofs of Propositions 7 and 8 are provided in Appendix I.

Appendix I
Proofs of Propositions 7 and 8

Proof of Proposition 7: This proof follows closely the proof of Proposition 2 in Appendix B.

To prove (129), we first notice that for any $i \in [M]$, $\mathbf{X}_i \to \mathbf{W}_i \to (\mathbf{Z}_{i\to r}, \mathbf{Z}_{i\to i}) \to \hat{\mathbf{X}}_i$ form a Markov chain. As a result, we have

$$I(\mathbf{X}_i; \hat{\mathbf{X}}_i) \leq I(\mathbf{W}_i; \mathbf{Z}_{i\to r}\mathbf{Z}_{i\to r}) \leq n_i \cdot p_{s_i;\cup\{d_i,r\}} \log(q).$$

Following the same remaining steps as in (85) and (86), we can thus prove (129).

The remaining proof again closely follows that of Proposition 2. The main difference is that for any $\epsilon > 0$ and the corresponding network code, we define a new set of 2^M functions $f_S^{[\epsilon]}(\cdot)$ by

$$f_S^{[\epsilon]}(p_{r;i}) = \frac{I(U;\mathbf{Z}_{r\to i}|\{\mathbf{Z}_{l\to j}\mathbf{X}_j : \forall j \in S, \forall l \in [M]\backslash j\}\mathbf{Z}_{\text{self}})}{n\log(q)} \quad (138)$$

and by interpolating the $f_S^{[\epsilon]}(p)$ value for all $p \neq p_{r;i}$. Comparing (87) and (138), the difference is the conditioning of a new set of random vectors $\mathbf{Z}_{\text{self}} \triangleq \{\mathbf{Z}_{i\to i} : \forall i \in [M]\}$. Namely, we consider only the conditional mutual information given what has already been received by d_i through the 2-hop jump from s_i to d_i.

When $S = \emptyset$, by definition we have

$$f_\emptyset^{[\epsilon]}(p_{r;M}) = \frac{I(U;\mathbf{Z}_{r\to M}|\mathbf{Z}_{\text{self}})}{n\log(q)} \leq \beta_r \cdot p_{r;M},$$

which will be used to prove (131).

Using the new definition of $f_S^{[\epsilon]}(p)$, we can derive the following equality by similar steps as in (89) to (92):

$$\begin{aligned}&(n\log(q)) \cdot f_S^{[\epsilon]}(p_{r;i})\\ &= (n\log(q)) \cdot f_T^{[\epsilon]}(p_{r;i})\\ &\quad + I(\mathbf{X}_k\mathbf{Z}_{\mathsf{SI}\to k};\mathbf{Z}_{r\to i}|\{\mathbf{X}_j,\mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}\mathbf{Z}_{\text{self}}). \quad (139)\end{aligned}$$

Define $T = S \cup \{k\}$. Using the following equality,

$$\begin{aligned}&I(\mathbf{X}_k;\{\mathbf{Z}_{k\to j} : \forall j \in T\})\\ &= I(\mathbf{X}_k;\mathbf{Z}_{k\to k}) + I(\mathbf{X}_k;\{Z_{k\to j} : \forall j \in S\}|\mathbf{Z}_{k\to k})\\ &= I(\mathbf{X}_k;\mathbf{Z}_{\text{self}}) + I(\mathbf{X}_k;\{Z_{k\to j} : \forall j \in S\}|\mathbf{Z}_{\text{self}}),\end{aligned}$$

we can carry out similar steps as in (93) to (103) and prove that

$$\begin{aligned}&I(\mathbf{X}_k\mathbf{Z}_{\mathsf{SI}\to k};\mathbf{Z}_{r\to i}|\{\mathbf{X}_j,\mathbf{Z}_{\mathsf{SI}\to j} : \forall j \in S\}\mathbf{Z}_{\text{self}})\\ &\geq (n\log(q)(R_k - \epsilon) - H_2(\epsilon) - I(\mathbf{X}_k;\{\mathbf{Z}_{k\to j} : \forall j \in T\}))^+. \quad (140)\end{aligned}$$

Since

$$I(\mathbf{X}_k : \{\mathbf{Z}_{k\to j} : \forall j \in T\}) \leq n\beta_k \log(q) p_{s_k;\cup T}, \quad (141)$$

combining (139), (140), and (141), we thus have that $\forall i, k \in [M], \forall S, T \in 2^{[M]}$ satisfying $k \notin S$ and $T = S \cup \{k\}$,

$$\begin{aligned}f_S^{[\epsilon]}(p_{r;i}) &\geq f_T^{[\epsilon]}(p_{r;i})\\ &\quad + 1_{\{i\leq k\}}\left(R_k - \epsilon - \frac{H_2(\epsilon)}{n\log(q)} - \beta_k \cdot p_{s_k;\cup T}\right)^+,\end{aligned}$$

which will be used to prove (130). By repeating the point-wise functional limit construction in Appendix B, the proof of Proposition 7 is complete. ■

Proof of Proposition 8: This proof follows closely that of Proposition 4 in Sections VI-A and VI-B. To that end, we first convert the inner bound of Proposition 8 to an equivalent LP problem by the alternative formulation of a ZPPLCIF in Lemma 1.

Proposition 9: Following the same settings and definitions of Proposition 8, a rate vector $(R_1, \cdots, R_M)$ can be achieved by a *linear network code* if there exist $(M+1)$ time allocation variables $(\beta_1, \cdots, \beta_M, \beta_r)$ with the total sum being $(M+1)$, $M(M+1)!$ non-negative variables $\{x_{\pi,S_j^\pi,h} : \forall \pi, \forall j \in [M+1], \forall h \in [M]\}$ and $M!(2^{M+1}-M-2)$ non-negative variables:

$$\Big\{y_{k,T,(\pi,S_j^\pi)} \geq 0 : \forall k \in [M], \forall T \in 2^{[M]}, \forall \pi, \forall j \in [M+1], \text{ satisfying } k \notin S_j^\pi \text{ and } T \subseteq S_j^\pi\Big\},$$

and $M2^{M-1}$ non-negative variables:

$$\Big\{z_{k,T} \geq 0 : \forall T \in 2^{[M]}, \forall k \in [M]\backslash T\Big\},$$

such that jointly the rate R_i and the x, y, z, β variables satisfy (132), (134), (135), (136), (63), and the following inequality:

$$\sum_{\forall \pi}\sum_{h=1}^{M} x_{\pi,\emptyset,h} \cdot p_{r;M} < \beta_r \cdot p_{r;M} \Longleftrightarrow \sum_{\forall \pi}\sum_{h=1}^{M} x_{\pi,\emptyset,h} < \beta_r.$$

For any rate vector $(R_1, \cdots, R_M)$ satisfies the inner bound of Proposition 9, perform the following four steps.

Step 1: For each time slot $t \in [n]$, each source s_i, $i \in [M]$, sends out the coded symbol $W_i(t)$ by RLNC, i.e.,

$$W_i(t) = \sum_{k=1}^{nR_i} c_{i,k}(t) X_{i,k},$$

for which we choose the mixing coefficients $c_{i,k}(t)$ independently and uniformly randomly from $\mathsf{GF}(q)$. $\mathbf{c}_i(t) \triangleq (c_{i,1}(t), \cdots, c_{i,nR_i}(t))$ denotes the corresponding (nR_i)-dimensional coding vector used in time t.

Step 2: The relay r derives the following linear subspaces $\mathcal{V}_{i;i}$ and $\mathcal{V}_{i;j}$ for all $i, j \in [M]$, $i \neq j$:

$$\begin{aligned}\mathcal{V}_{i;i} &\triangleq \mathsf{span}(\{\mathbf{c}_i(t) : \forall t \text{ s.t. } Z_{i\to i}(t) \neq *\})\\ \mathcal{V}_{i;j} &\triangleq \mathsf{span}(\{\mathbf{c}_i(t) : \forall t \text{ s.t.}\\ &\qquad Z_{i\to j}(t) = Z_{i\to r}(t) \neq *, Z_{i\to i}(t) = *\})\end{aligned}$$

Namely, $\mathcal{V}_{i;j}$ is the subspace of user i that is decodable by r, not received by d_i directly, but is overheard by d_j. Based on $\{\mathcal{V}_{i;i}, \mathcal{V}_{i;j} : \forall i \neq j\}$, the relay r further derives multiple subspaces $\mathcal{V}_{i\to T_1\overline{T_2}}$ for all $i \in [M]$, $T_1 \subseteq [M]\backslash i$, and $T_2 = [M]\backslash(T_1 \cup \{i\})$ in a similar way as in Step 2 of Section VI-A, such that the following three properties are satisfied:

(i) $\forall j \in T_1, \quad \mathcal{V}_{i\to T_1\overline{T_2}} \subseteq \mathcal{V}_{i;j}$,

(ii) $\mathsf{span}\{\mathcal{V}_{i;i}, \mathcal{V}_{i\to T_1\overline{T_2}} : \forall T_1 \subseteq [M]\backslash i, T_2 = [M]\backslash(T_1 \cup \{i\})\} = \Omega_i$,

(iii) $\dfrac{\text{Rank}(\mathcal{V}_{i\to T_1\overline{T_2}})}{n} = z_{i,T_2}$.

Namely, the message space Ω_i of user-i is partitioned into multiple subspaces $\mathcal{V}_{i\to T_1\overline{T_2}}$ such that each subspace $\mathcal{V}_{i\to T_1\overline{T_2}}$ is *known* to the relay r and all the users $j \in T_1$, and the normalized rank of $\mathcal{V}_{i\to T_1\overline{T_2}}$ equals to the feasible solution z_{i,T_2} satisfying (135) and (136) of Proposition 8.

With the above construction of the subspaces, let $\mathbf{v}_k^{[i\to T_1\overline{T_2}]}$, $k = 1 \cdots, n\cdot z_{i,T_2}$ denote the basis vectors of $\mathcal{V}_{i\to T_1\overline{T_2}}$. We can then perform the regrouping operation and the code-alignment-based encoding method in the same way as in Steps 3 and 4 of Sections VI-A and VI-B. It is then guaranteed that all the information subspaces $\mathcal{V}_{i\to T_1\overline{T_2}}$ at the relay r can be conveyed to destination d_i. Property (ii) of $\mathcal{V}_{i\to T_1\overline{T_2}}$ then guarantees the decodability at d_i. The proof is thus complete.

■

References

[1] N. Alon, A. Hassidim, E. Lubetzky, U. Stav, and A. Weinstein, "Broadcasting with side information," in *Proc. Annual IEEE Symp. Found. Computer Science*, 2008.

[2] M. Bakshi and M. Effros, "On achievable rates for multicast in the presence of side information," in *Proc. IEEE Int'l Symp. Inform. Theory*. Tornoto, Canada, 2008, pp. 1661–1665.

[3] Z. Bar-Yossev, Y. Birk, T. Jayram, and T. Kol, "Index coding with side information," in *Proc. Annual IEEE Symp. Found. Computer Science*. Berkeley, California, USA, October 2006.

[4] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, August 2008.

[5] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," in *Proc. ACM Special Interest Group on Data Commun. (SIGCOMM)*. Kyoto, Japan, August 2007.

[6] P. Chaporkar and A. Proutiere, "Adaptive network coding and scheduling for maximizing throughput in wireless networks," in *Proc. Annual ACM Int'l. Conf. on Mobile Computing & Networking (MobiCom)*. Montréal, Canada, September 2007, pp. 315–146.

[7] M. Chaudhry and A. Sprintson, "Efficient algorithms for index coding," in *Proc. 27th IEEE Conference on Computer Communications (INFOCOM)*. Pheonix, USA, April 2008, pp. 1–4.

[8] P. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. 41st Annual Allerton Conf. on Comm., Contr., and Computing*. Monticello, IL, October 2003.

[9] A. Cohen, S. Avestimehr, and M. Effros, "On networks with side information," in *Proc. IEEE Int'l Symp. Inform. Theory*. Seoul, Korea, June 2009.

[10] T. Cui, L. Chen, and T. Ho, "Energy efficient opportunistic network coding for wireless networks," in *Proc. 27th IEEE Conference on Computer Communications (INFOCOM)*. Phoenix, USA, April 2008.

[11] A. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 789–804, March 2006.

[12] A. Dana and B. Hassibi, "The capacity region of multiple input erasure broadcast channels," in *Proc. IEEE Int'l Symp. Inform. Theory*. Adelaide, Australia, September 2005.

[13] A. Das, S. Vishwanath, S. Jafar, and A. Markopoulou, "Network coding for multiple unicasts: An interference alignment approach," in *Proc. IEEE Int'l Symp. Inform. Theory*. Austin, Texas, USA, June 2010.

[14] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2745–2759, August 2005.

[15] ——, "Networks, matroids, and non-Shannon information inequality," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 1949–1969, June 2007.

[16] ——, "Linear network codes and systems of polynomial equations," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2303–2316, May 2008.

[17] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback — capacity and algorithms," in *Proc. 5th Workshop on Network Coding, Theory, & Applications (NetCod)*. Lausanne, Switzerland, June 2009, pp. 54–61.

[18] T. Ho, M. Médard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, October 2006.

[19] C. Jiang, B. Smith, B. Hassibi, and S. Vishwanath, "Multicast in wireless erasure networks with feedback," in *IEEE Symp. Computer & Commun.*, July 2008, pp. 562–565.

[20] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: Practical wireless network," in *Proc. ACM Special Interest Group on Data Commun. (SIGCOMM)*, 2006.

[21] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782–795, October 2003.

[22] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 60–64, January 1977.

[23] D. Koutsonikolas, C.-C. Wang, and Y. Hu, "CCACK: Efficient network coding based opportunistic routing through cumulative coded acknowledgments," in *Proc. 29th IEEE Conference on Computer Communications (INFOCOM)*. San Diego, USA, March 2010, pp. 1–9.

[24] G. Kramer, "Communication on line networks with deterministic or erasure broadcast channels," in *Proc. IEEE Inform. Theory Workshop*. Taormina, Italy, October 2009, pp. 404–405.

[25] G. Kramer and S. Shamai, "Capacity for classes of broadcast channels with receiver side information," in *Proc. IEEE Inform. Theory Workshop*. Lake Tahoe, California, USA, September 2007, pp. 313–318.

[26] J. Le, J. Lui, and D. Chiu, "How many packets can we encode? — an analysis of practical wireless network coding," in *Proc. 27th IEEE Conference on Computer Communications (INFOCOM)*. Phoenix, USA, April 2008.

[27] A. Lehman, "Network coding," Ph.D. dissertation, MIT, 2005.

[28] S.-Y. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, February 2003.

[29] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452–1463, August 2006.

[30] D. Lun, N. Ratnakar, M. Médard, R. Koetter, D. Karger, T. Ho, E. Ahmed, and F. Zhao, "Minimum-cost multicast over coded packet networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2608–2623, June 2006.

[31] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, July 1984.

[32] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE Int'l Symp. Inform. Theory*. Seattle, USA, 2006.

[33] S. Rayanchu, S. Sen, J. Wu, S. Banerjee, and S. Sengupta, "Loss-aware network coding for unicast wireless sessions: Design, implementation, and performance evaluation," in *SIGMETRICS*. Annapolis, Maryland, USA, June 2008.

[34] K. Salamatian and R. Khalili, "An information theory for erasure channels," in *Proc. 43rd Annual Allerton Conf. on Comm., Contr., and Computing*. Monticello, Illinois, USA, September 2005.

[35] S. Sengupta, S. Rayanchu, and S. Banerjee, "An analysis of wireless network coding for unicast sessions: the case for coding aware routing," in *Proc. 26th IEEE Conference on Computer Communications (INFOCOM)*. Anchorage, USA, May 2007, pp. 1028–1036.

[36] B. Smith and B. Hassibi, "Wireless erasure networks with feedback," in *Proc. IEEE Int'l Symp. Inform. Theory*, July 2008, pp. 339–343.

[37] A. Steiner and S. Shamai, "Achievable rates with imperfect transmitter side information using a broadcast transmission strategy," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 1043–1051, March 2008.

[38] C. Suh and K. Ramachandran, "Exact-repair MDS code construction using interference alignment," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1425–1442, March 2011.

[39] P. Tarasak, U. Sethakaset, and S. Sun, "Capacity analysis of two-user opportunistic scheduling for wireless network coding," in *Proc. IEEE Int'l Symp. Inform. Theory*. Seoul, Korea, June 2009, pp. 2572–2576.

[40] C.-C. Wang, "Capacity of 1-to-K broadcast packet erasure channels with channel output feedback," in *Proc. 48th Annual Allerton Conf. on Comm., Contr., and Computing*. Monticello, Illinois, USA, September 2010.

[41] C.-C. Wang and N. Shroff, "Beyond the butterfly — a graph-theoretic characterization of the feasibility of network coding with two simple unicast sessions," in *Proc. IEEE Int'l Symp. Inform. Theory*. Nice, France, June 2007, pp. 121–125.

[42] Y. Wu, "Broadcasting when receivers know some messages a priori," in *Proc. IEEE Int'l Symp. Inform. Theory*. Nice, France, June 2007, pp. 1141–1145.

[43] Y. Wu and A. Dimakis, "Reducing repair traffic for erasure coding-based storage via interference alignment," in *Proc. IEEE Int'l Symp. Inform. Theory*. Seoul, Korea, June 2009.

[44] Y. Wu, J.Padhye, R. Chandra, V. Padmanabhan, and P. Chou, "The local mixing problem," in *Proc. Inform. Theory and Applications Workshop*. San Diego, California, USA, February 2006.

[45] F. Xue and X. Yang, "Network coding and packet-erasure broadcast channel," in *Proc. the 5th IEEE Annual Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops (SECON)*. San Francisco, CA, USA, July 2008.

[46] J. Yoo, T. Liu, and F. Xue, "Gaussian broadcast channels with receiver message side information," in *Proc. IEEE Int'l Symp. Inform. Theory*. Seoul, Korea, June 2009.

[47] F. Zhao and M. Médard, "On analyzing and improving COPE performance," in *Proc. Inform. Theory and Applications Workshop*. San Diego, USA, February 2010.