A Unified Relay Framework with both D-F and C-F Relay Nodes

Xiugang Wu and Liang-Liang Xie

University of Waterloo, Waterloo, ON, Canada N2L 3G1

Email: x23wu@uwaterloo.ca, llxie@uwaterloo.ca

Abstract

Decode-and-forward (D-F) and compress-and-forward (C-F) are two fundamentally different relay strategies proposed by (Cover and El Gamal, 1979). Individually, either of them has been successfully generalized to multi-relay channels. In this paper, to allow each relay node the freedom of choosing either of the two strategies, we propose a unified framework, where both the D-F and C-F strategies can be employed simultaneously in the network. It turns out that, to incorporate in full the advantages of both the best known D-F and C-F strategies into a unified framework, the major challenge arises as follows: For the D-F relay nodes to fully utilize the help of the C-F relay nodes, decoding at the D-F relay nodes should not be conducted until all the blocks have been finished; however, in the multi-level D-F strategy, the upstream nodes have to decode prior to the downstream nodes in order to help, which makes simultaneous decoding at all the D-F relay nodes after all the blocks have been finished inapplicable. To tackle this problem, nested blocks combined with backward decoding are used in our framework, so that the D-F relay nodes at different levels can perform backward decoding at different frequencies. As such, the upstream D-F relay nodes can decode before the downstream D-F relay nodes, and the use of backward decoding at each D-F relay node ensures the full exploitation of the help of both the other D-F relay nodes and the C-F relay nodes. The achievable rates under our unified relay framework are found to combine both the best known D-F and C-F achievable rates and include them as special cases. It is also demonstrated through a Gaussian network example that our achievable rates are generally better than the rates obtained with existing unified schemes and with D-F or C-F alone.

I. INTRODUCTION

The relay channel, originally proposed in [1], models a communication scenario where there is one or more relay nodes that can help the information transmission between the source and the destination. The simplest one-relay channel is depicted in Fig. 1, where nodes 0, 1, and 2 are the source, the relay, and the destination, respectively. Two fundamentally different relay strategies have been developed in [2] for such channels, which, depending on whether the relay decodes the information or not, are generally known as *decode-and-forward* (D-F) and *compress-and-forward* (C-F) respectively.
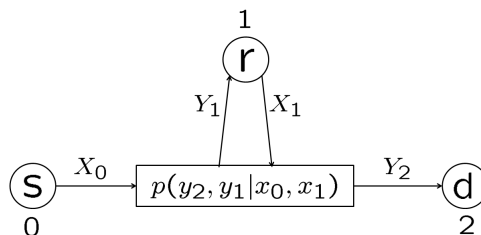

Fig. 1. The single-relay channel.

A. D-F and C-F for single-relay channels

In the D-F relay strategy, the relay first decodes the message sent by the source and then forwards it to the destination, and the destination decodes the message taking into account the inputs of both the source

and the relay. With the D-F relay strategy, the following rate is achievable:

$$R < \max_{p(x_0, x_1)} \min\{I(X_0; Y_1|X_1), I(X_0, X_1; Y_2)\} \tag{1}$$

where, the first condition $R < I(X_0; Y_1|X_1)$ makes node 1 able to decode the message based on the signal transmitted by node 0, and the second condition $R < I(X_0, X_1; Y_2)$ makes node 2 able to decode the message based on the signals transmitted by node 0 and node 1 together. Notably, the maximization in (1) is over $p(x_0, x_1)$, rather than $p(x_0)p(x_1)$, which suggests that (1) can only be achieved by node 0 and node 1 cooperating with each other when transmitting signals. To accomplish such cooperation, an essential technique called block Markov coding was employed in the D-F coding scheme developed in [2]. Besides, the scheme in [2] also used irregular encoding with codebooks of different sizes at the source and at the relay, random partitioning (binning), and successive decoding. Subsequently, some other D-F coding schemes also achieving (1) were found in [3]-[4].

In contrast, the C-F relay strategy is used when the relay cannot decode the message sent by the source, but still can help by compressing its observation Y_1 into $\hat{Y}_1$, and forwarding this compressed version to the destination. The destination then either successively or jointly decodes the compression of the relay's observation and the original message of the source. In the original C-F scheme of [2], the decoder performs successive compression-message decoding, i.e., it first decodes the compression of the relay's observation, and then decodes the original message of the source, leading to the following achievable rate:

$$R < \max_{p(x_0)p(x_1)p(\hat{y}_1|y_1, x_1)} I(X_0; \hat{Y}_1, Y_2|X_1) \tag{2}$$

$$\text{such that} \quad I(Y_1; \hat{Y}_1|X_1, Y_2) \leq I(X_1; Y_2), \tag{3}$$

where (3) ensures that the compression $\hat{Y}_1$ can be first recovered at the destination, and (2) ensures that the destination can decode the original message X_0 based on $\hat{Y}_1$ and Y_2 together.

The two-step compression-message successive decoding process in [2] requires $\hat{Y}_1$ to be decoded first, which facilitates the decoding of X_0, but is not a requirement of the original problem. Recognizing this, a joint compression-message decoding process was proposed in [5], where, instead of successively, the destination decodes $\hat{Y}_1$ and X_0 together. It turns out that the decoding of X_0 can be helped even if $\hat{Y}_1$ cannot be decoded first. In fact, with joint decoding, the constraint (3) is not necessary, and instead of (2), the achievable rate is expressed as

$$R < \max_{p(x_0)p(x_1)p(\hat{y}_1|y_1, x_1)} I(X_0; \hat{Y}_1, Y_2|X_1) - \max\{0, I(Y_1; \hat{Y}_1|X_1, Y_2) - I(X_1; Y_2)\}. \tag{4}$$

Similar formulas as (4) have been derived with different arguments in [6]-[8]. Therefore, compared to successive decoding, joint compression-message decoding provides more freedom in choosing the compression $\hat{Y}_1$. However, the question remains whether joint decoding achieves strictly higher rates for the original message than successive decoding. For the single relay case, it was proved in [8] that the answer is negative, and any rate achievable by either of them can always be achieved by the other, i.e., the achievable rates in (2)-(3) and (4) are essentially the same. In fact, as we will see later in the Introduction, when C-F is generalized to the case of multiple relays, there is no improvement on the achievable rate by joint decoding either.

Combining the D-F and C-F together, one can further consider the hybrid scheme, where the relay partially decodes the message and compresses the rest of its received signals; see, e.g., [2, Thm 7] for the single-relay case and its extension to the multiple-relay case in [9]. However, such hybrid schemes generally involve superposition coding that induces auxiliary random variables, making the expression and evaluation of the achievable rates rather complicated especially in the case of multiple relays that we will consider in the sequel. Thus, in this paper, our discussion focuses on the "pure" D-F or C-F strategies only, i.e., the strategies where the relay either completely decodes the message, or does not decode at all but simply compresses and forwards its observation.

B. D-F and C-F for multi-relay channels

A natural extension of the single-relay channel in Fig. 1 is to the case of multiple relays depicted in Fig. 2, where nodes 0 and $n+1$ are the source and the destination respectively, and nodes $1, 2, \ldots, n$ are the n relay nodes that constitute the relay nodes set, denoted by $\mathcal{N}$. Both the D-F and C-F relay strategies have been separately generalized to such multi-relay channels in [10]-[20], among which, [13]-[16] and [18]-[20] provide the best achievable rates for D-F and C-F respectively.

Specifically, in generalizing D-F to the multi-relay channel, [13]-[14] modified the original irregular encoding/successive decoding scheme of [2] to a regular encoding/sliding window decoding scheme to realize the "multi-level" D-F relay strategy. For any fixed permutation π on $\{0, 1, \ldots, n+1\}$ with $\pi(1) = 0$ and $\pi(n+2) = n+1$, i.e., any specific ordering of the relay nodes as $\pi(2), \pi(3), \ldots, \pi(n+1)$, their multi-level D-F scheme [13]-[14] achieves the following rate:

$$R < \max_{p(x_0, x_1, \ldots, x_n)} \min_{2 \leq k \leq n+2} I(X_{\pi(1:k-1)}; Y_{\pi(k)} | X_{\pi(k:n+1)}), \tag{5}$$

where $\pi(k_1 : k_2) := \{\pi(k_1), \pi(k_1 + 1), \ldots, \pi(k_2)\}$. Later on, it was found in [15]-[16] that (5) can also be achieved with backward decoding.

The formula (5) has a similar interpretation as (1). For each node $\pi(k), k = 2, 3, \ldots, n+2$, the corresponding rate constraint is

$$R < I(X_{\pi(1:k-1)}; Y_{\pi(k)} | X_{\pi(k:n+1)}), \tag{6}$$

which implies that for the decoding at node $\pi(k)$, the signals transmitted by nodes $\pi(k+1 : n+1)$ are known *a priori*, and the signals transmitted by nodes $\pi(1 : k-1)$ are cooperating in providing the information. A simple explanation of this feasibility is the following. In the multi-level D-F relay strategy, information is passed along the route $\pi(1) \to \pi(2) \to \cdots \to \pi(n+2)$, so that i) any information obtained by the downstream nodes of $\pi(k)$, i.e., nodes $\pi(k+1 : n+1)$, has already been obtained by node $\pi(k)$, and therefore their inputs are predictable by node $\pi(k)$, and ii) by the time the information reaches node $\pi(k)$, all its upstream nodes $\pi(1 : k-1)$ have already obtained the same information and can therefore cooperate with the technique of block Markov coding. The formula (5) also demonstrates a remarkable feature of the multi-level D-F relay strategy in [13]-[16], i.e., it completely eliminates the interference in the network: To any node, the signal transmitted by any other node is either a "real" signal that can be used for decoding, or a *priori* known signal that can be subtracted completely.
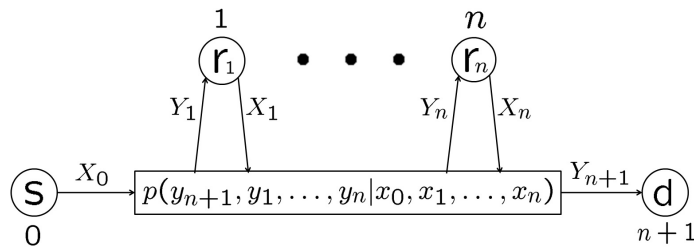

Fig. 2. The multiple-relay channel.

In the line of generalizing C-F to multi-relay channels, substantial advances have been recently made in [18]-[20]. First, in [18], a new C-F scheme termed noisy network coding was proposed. Different from the original C-F scheme of [2], where cumulative encoding/block-by-block forward decoding was used, this noisy network coding scheme employed repetitive encoding/all blocks united decoding. Besides, it also used compression-message joint decoding without uniquely decoding the relays' compressions, instead of compression-message successive decoding as in the original C-F scheme. It turns out [18] that the noisy network coding scheme achieves the same rate as the original C-F scheme for the single-relay channel, but improves the original C-F scheme in the case of multiple relays to achieve higher rates as follows:

$$R < \max_{p(x_0) \prod_{i=1}^{n} p(x_i) p(\hat{y}_i | y_i, x_i)} \min_{\mathcal{S} \subseteq \mathcal{N}} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{N} \setminus \mathcal{S}}, Y_{n+1} | X_{\mathcal{N} \setminus \mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{N} \setminus \mathcal{S}}). \tag{7}$$

However, soon in [19]-[20], it was discovered that neither repetitive encoding/all blocks united decoding nor compression-message joint decoding used in [18] is necessary to achieve the rate (7); in particular, a cumulative encoding/block-by-block backward decoding/compression-message successive decoding scheme was developed, and its corresponding achievable rate was shown to be the same as (7), with the following form:

$$R < \max_{p(x_0)\prod_{i=1}^{n} p(x_i)p(\hat{y}_i|x_i,y_i)} I(X_0; \hat{Y}_{\mathcal{N}}, Y_{n+1}|X_{\mathcal{N}}) \tag{8}$$

$$\text{such that } I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{N}\setminus\mathcal{S}}, Y_{n+1}|X_{\mathcal{N}\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{N}\setminus\mathcal{S}}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{N}, \tag{9}$$

where (8)-(9) can be similarly interpreted as (2)-(3) for the single-relay case, i.e., (9) ensures that the relays' compressions $\hat{Y}_{\mathcal{N}}$ can be first recovered at the destination, and (8) ensures that the destination can decode the original message of the source based on Y_{n+1} and $\hat{Y}_{\mathcal{N}}$ together. Note that the rate equivalence between (7) and (8)-(9) also demonstrates that in the case of multiple relays, there is no improvement on the achievable rate by joint compression-message decoding either, which is consistent with the conclusion made in the single-relay case. More interestingly, in proving such a rate equivalence, [19]-[20] found that the the R.H.S. (right-hand-side) of (7) is maximized *only when* the compressions $\hat{Y}_{\mathcal{N}}$ are chosen to satisfy (9), i.e., to maximize the achievable rate for the original message, the compressions should always be chosen to support successive decoding, and any compressions not supporting successive decoding will actually lead to strictly lower achievable rates for the original message.

Since block-by-block backward decoding and compression-message successive decoding are relatively easier to implement than all blocks united decoding and compression-message joint decoding respectively, the cumulative encoding/block-by-block backward decoding/compression-message successive decoding scheme of [19]-[20] becomes the simplest choice in achieving the highest C-F rate in the case of multiple relays. Moreover, the fact that this scheme achieves the same rate as noisy network coding also reveals the essential reason for the improvement of the achievable rate—not repetitive encoding/all blocks united decoding, not joint compression-message decoding, but delayed decoding until all the blocks have been finished. This delayed decoding is generally necessary because the multiple-relay case differs from the single-relay case in that it may take multiple blocks for the relays to help each other before their compressions can finally reach the destination. Hence, the block-by-block forward decoding scheme, which is sufficient for the single-relay case, may not work satisfactorily for multiple relays in general [20].

It is worth noting that although the optimal C-F rate is achieved only when the compressions are chosen to support successive decoding in single-destination networks, in a network with multiple destinations ([18], [20]), the compressions may not be chosen to support successive decoding at all the destinations, and joint decoding might have to be used. For this, a more general scheme of cumulative encoding/block-by-block backward decoding/compression-message joint decoding was developed in [20]. For any given distribution $p(x_0)\prod_{i=1}^{n} p(x_i)p(\hat{y}_i|x_i,y_i)$, this scheme achieves the following rate:

$$R < \min_{\mathcal{S}\subseteq\mathcal{D}} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}\setminus\mathcal{S}}, Y_{n+1}|X_{\mathcal{D}\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_{\mathcal{D}}, Y_{n+1}, \hat{Y}_{\mathcal{D}\setminus\mathcal{S}}), \tag{10}$$

where $\mathcal{D}$ is the unique largest subset of $\mathcal{N}$ satisfying

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}\setminus\mathcal{S}}, Y_{n+1}|X_0, X_{\mathcal{D}\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_{\mathcal{D}}, Y_{n+1}, \hat{Y}_{\mathcal{D}\setminus\mathcal{S}}) > 0, \forall \mathcal{S} \subseteq \mathcal{D}, \mathcal{S} \neq \emptyset, \tag{11}$$

and $\hat{Y}_{\mathcal{D}}$ can be decoded jointly with X_0. Here, $\mathcal{D}$ can be interpreted as the "jointly decodable" relay nodes set such that the compressions of the relays in this set are decodable jointly with the original message X_0. In contrast, the compression of any relay node in $\mathcal{N} \setminus \mathcal{D}'$ is not decodable even jointly with X_0, where $\mathcal{D}'$ is the unique largest subset of $\mathcal{N}$ satisfying

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}'\setminus\mathcal{S}}, Y_{n+1}|X_0, X_{\mathcal{D}'\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_{\mathcal{D}'}, Y_{n+1}, \hat{Y}_{\mathcal{D}'\setminus\mathcal{S}}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}'. \tag{12}$$

On the other hand, for any given distribution $p(x_0) \prod_{i=1}^{n} p(x_i) p(\hat{y}_i | x_i, y_i)$, the achievable rate (7) can be more generally expressed as

$$R < \min_{\mathcal{S} \subseteq \mathcal{T}} I(X_0, X_\mathcal{S}; \hat{Y}_{\mathcal{T} \backslash \mathcal{S}}, Y_{n+1} | X_{\mathcal{T} \backslash \mathcal{S}}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | X_0, X_\mathcal{T}, Y_{n+1}, \hat{Y}_{\mathcal{T} \backslash \mathcal{S}}) \tag{13}$$

if we only consider a subset of relays $\mathcal{T} \subseteq \mathcal{N}$ for the decoding, while treating the other inputs as purely noise. Interestingly, it was found in [20] that among all the choices of $\mathcal{T} \subseteq \mathcal{N}$, the R.H.S. of (13) is maximized when $\mathcal{T} = \mathcal{D}$ or $\mathcal{T} = \mathcal{D}'$, but is strictly less than the maximum when $\mathcal{T} \nsubseteq \mathcal{D}'$. Therefore, only those relays whose compressions are jointly decodable are helpful to the decoding of the original message, and including the jointly un-decodable compressions in the formula (13), i.e., choosing $\mathcal{T} \nsubseteq \mathcal{D}'$, will even strictly lower the achievable rate.

By comparing (10) and (13) with $\mathcal{T} = \mathcal{D}$, it can be concluded that for any compressions chosen at the relays, the cumulative encoding/block-by-block backward decoding/compression-message joint decoding scheme of [20] achieves the same rate as the noisy network coding scheme [18].[1]

C. A unified relay framework with both D-F and C-F relay nodes

In the above discussions, all the relay nodes in the network perform only one type of relay strategy, either D-F or C-F. However, to obtain higher achievable rate, it might be better to let each relay node choose from D-F and C-F its relay strategy depending on the channel condition, e.g., let the relay node close to the source perform D-F while let the relay node close to the destination perform C-F. This invokes a unified relay framework that includes both the D-F and C-F relay nodes in the network. In developing such a framework, one naturally wants to combine the advantages of both the best known D-F and C-F schemes, i.e., the multi-level D-F schemes in [13]-[16] and the recent advances on C-F schemes in [18]-[20].

Some attempts towards this unified relay framework have been made in [16, Thm 4], [24]. The work [16], however, had been done before noisy network coding was proposed, and thus the recent progress on C-F schemes was not reflected in it. Moreover, in the scheme of [16, Thm 4], the relay nodes are not fully cooperating in the sense that the D-F relay nodes do not utilize the help of the C-F relay nodes. In the recent work [24], the authors incorporate noisy network coding to their scheme and let the D-F relay nodes exploit the help of the C-F relay nodes. Nevertheless, [24] does not use the multi-level D-F schemes as in [13]-[16]. Instead, all the D-F relay nodes in [24] are at the same level, and thus the decoding at each D-F relay node cannot exploit the help of other D-F relay nodes. Besides, in [24], although the destination performs backward decoding to fully exploit the help of the C-F relay nodes, the decoding at each D-F relay node is based on two consecutive blocks only and thus does not fully utilize the help of the C-F relay nodes as in [18]-[20].

Indeed, it turns out that, to incorporate in full the advantages of both the best known D-F and C-F relay strategies into a unified framework is nontrivial due to the following major challenge: For the D-F relay nodes to fully utilize the help of the C-F relay nodes as in [18]-[20], decoding at the D-F relay nodes should not be conducted until all the blocks have been finished; however, to perform the multi-level D-F strategy as in [13]-[16], the upstream nodes have to decode prior to the downstream nodes in order to help, which makes simultaneous decoding at all the D-F relay nodes after all the blocks have been finished inapplicable.

To tackle this problem, nested blocks ([15]-[16], [25]) combined with backward decoding are used in our framework, so that the D-F relay nodes at different levels can perform backward decoding at different frequencies: the closer to the source in the information passing route, the higher decoding frequency. As such, the upstream D-F relay nodes can decode before the downstream D-F relay nodes and the destination, and the use of backward decoding at each D-F relay node ensures the full exploitation of the help of both the other D-F relay nodes and the C-F relay nodes.

---

[1]Part of the results in [19]-[20] have also been recognized in [21]-[23].

Specifically, we partition the relay nodes set $\mathcal{N}$ into two sets, $\mathcal{M}$ with $|\mathcal{M}| = M$ and $\mathcal{N} \backslash \mathcal{M}$, as depicted in Fig. 3, and fix some permutation π on $\{0\} \bigcup \mathcal{M} \bigcup \{n+1\}$ with $\pi(1) = 0$ and $\pi(M+2) = n+1$. Let the relay nodes in $\mathcal{M}$ perform the multi-level D-F cooperatively along the route $\pi(1) \to \pi(2) \to \cdots \to \pi(M+2)$, while let each node $i \in \mathcal{N} \backslash \mathcal{M}$ perform C-F as in [18]-[20] independently. Then, a total of B^{M+1} blocks will be used and the length of a "virtual" block for node $\pi(k), k = 2, 3, \ldots, M+2$, will be B^{k-2} blocks. The backward decoding at the destination, i.e., node $\pi(M+2)$, will happen at the end of all B^{M+1} blocks, while the backward decoding at the D-F relay node $\pi(k), k = 2, 3, \ldots, M+1$, will happen whenever it has received B new "virtual" blocks, i.e., at the end of each block $b = vB^{k-1}, v \in [1 : B^{M+1}/B^{k-1}]$. Also, both the D-F relay nodes and the destination will perform compression-message joint decoding, which is in general necessary since the compressions of the C-F relay nodes may not be chosen to support successive decoding at all the D-F relay nodes and the destination.

Under the above described framework, for any given distribution $p(x_0)p(x_{\mathcal{M}}|x_0) \prod_{i \in \mathcal{N} \backslash \mathcal{M}} p(x_i)p(\hat{y}_i|y_i, x_i)$, the following rate is achievable:

$$R < \min_{2 \leq k \leq M+2} \min_{\mathcal{S} \subseteq \mathcal{D}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_k \backslash \mathcal{S}}, Y_{\pi(k)} | X_{\mathcal{D}_k \backslash \mathcal{S}}, X_{\pi(k:M+1)}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k \backslash \mathcal{S}}),$$
$$(14)$$

where $\mathcal{D}_k$ is the unique largest subset of $\mathcal{N} \backslash \mathcal{M}$ satisfying

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_k \backslash \mathcal{S}}, Y_{\pi(k)} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k \backslash \mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k \backslash \mathcal{S}}) > 0, \qquad (15)$$

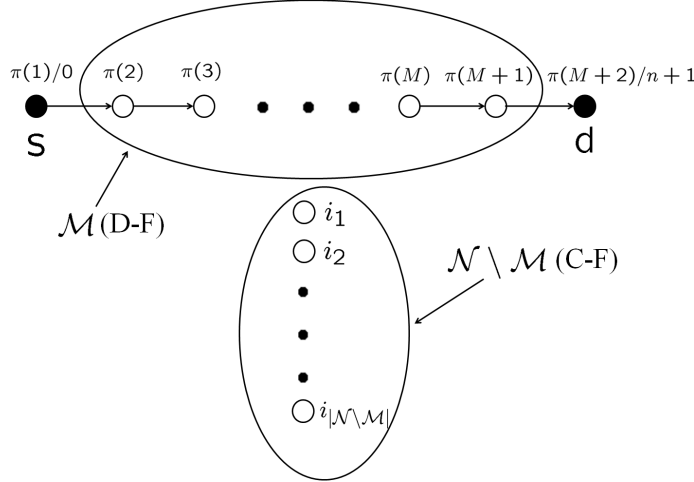for any nonempty $\mathcal{S} \subseteq \mathcal{D}_k$.

Fig. 3. A unified relay framework with both the D-F and C-F relay nodes.

(14) has the flavors of both (5) and (10). Specifically, for each node $\pi(k), k = 2, 3, \ldots, M+2$, the corresponding rate constraint is

$$R < \min_{\mathcal{S} \subseteq \mathcal{D}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_k \backslash \mathcal{S}}, Y_{\pi(k)} | X_{\mathcal{D}_k \backslash \mathcal{S}}, X_{\pi(k:M+1)}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k \backslash \mathcal{S}}), \quad (16)$$

which is in a form similar to (10) but with the appearance of $X_{\pi(1:k-1)}$, $X_{\pi(k:M+1)}$ and $X_{\pi(1:M+1)}$. (16) has the similar form as (10) since node $\pi(k)$ uses the help of the C-F relay nodes as in [19]-[20]. $X_{\pi(1:k-1)}$, $X_{\pi(k:M+1)}$ and $X_{\pi(1:M+1)}$ appear in (16) because node $\pi(k)$ also utilizes the help of other D-F relay nodes as in [13]-[16] so that the signals of its upstream nodes, i.e., $X_{\pi(1:k-1)}$, are cooperatively providing the information while the signals of its downstream nodes and itself $X_{\pi(k:M+1)}$ are known at $\pi(k)$. Also, the set $\mathcal{D}_k$ defined in (15) has a similar interpretation as the set $\mathcal{D}$ defined in (11), i.e., the "jointly decodable" C-F relay nodes set at node $\pi(k)$ such that the compressions of the relays in this set are decodable jointly with $X_{\pi(1:k-1)}$ given that $X_{\pi(k:M+1)}$ are known at node $\pi(k)$.

It can be easily seen that (14) includes the achievable rates in (5) and (10) as special cases: When $\mathcal{M} = \mathcal{N}$, i.e., all the relays perform D-F, $\mathcal{D}_k = \emptyset$ and (14) reduces to (5); When $\mathcal{M} = \emptyset$, i.e., all the relays perform C-F, (14) reduces to (10).

Finally, it should be noted that, the achievable rate (14) is proved by using the block-by-block backward decoding scheme in [20]. We can also modify the all blocks united decoding scheme in [18] to a B-blocks-by-B-blocks backward decoding scheme, to fit it into our unified relay framework and prove the following achievable rate:

$$R < \min_{2 \leq k \leq M+2} \max_{\mathcal{T}_k \subseteq \mathcal{N} \backslash \mathcal{M}} \min_{\mathcal{S} \subseteq \mathcal{T}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_k \backslash \mathcal{S}}, Y_{\pi(k)} | X_{\mathcal{T}_k \backslash \mathcal{S}}, X_{\pi(k:M+1)}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{T}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{T}_k \backslash \mathcal{S}}).$$
(17)

Similarly to the equivalence between (10) and (13), here (14) and (17) are also equivalent. One can also easily check that (17) includes the achievable rates in (5) and (13) as special cases by letting $\mathcal{M} = \mathcal{N}$ and $\mathcal{M} = \emptyset$ respectively. Notably, again, in terms of complexity, block-by-block backward decoding is relatively easier to implement since B-blocks-by-B-blocks backward decoding involves B blocks united decoding.

The remainder of the paper is organized as the following. In Section II, we formally state our problem setup and summarize the main results. Then, in Section III and Section IV, our unified relay framework with block-by-block backward decoding and with B-blocks-by-B-blocks backward decoding will be presented in detail respectively. Finally, some concluding remarks are included in Section V.

II. MAIN RESULTS

Consider a multiple-relay channel consisting of $n + 2$ nodes, as depicted in Fig. 2, where nodes 0 and $n+1$ are the source and the destination respectively, and nodes $1, 2, \ldots, n$ are the n relay nodes. Formally, this channel can be denoted by

$$(\mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_n, \, p(y_{n+1}, y_1, \ldots, y_n | x_0, x_1, \ldots, x_n), \, \mathcal{Y}_{n+1} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n)$$

where, $\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_n$ are the transmitter alphabets of the source and the relays respectively, $\mathcal{Y}_{n+1}, \mathcal{Y}_1, \ldots, \mathcal{Y}_n$ are the receiver alphabets of the destination and the relays respectively, and a collection of probability distributions $p(\cdot, \cdot, \ldots, \cdot | x_0, x_1, \ldots, x_n)$ on $\mathcal{Y}_{n+1} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n$, one for each $(x_0, x_1, \ldots, x_n) \in \mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. The interpretation is that x_0 is the input to the channel from the source, y_{n+1} is the output of the channel to the destination, and y_i is the output received by the i-th relay. The i-th relay sends an input x_i based on what it has received:

$$x_i(t) = r_{i,t}(y_i(t-1), y_i(t-2), \ldots), \quad \text{for every time } t, \tag{18}$$

where $r_{i,t}(\cdot)$ can be any causal function.

Before presenting the main results, we introduce some simplified notations. Denote the set $\mathcal{N} = \{1, 2, \ldots, n\}$. For any subset $\mathcal{S} \subseteq \{0, 1, \ldots, n+1\}$, let $X_{\mathcal{S}} = \{X_i, i \in \mathcal{S}\}$, and use similar notations for other variables. For any $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$, let $\pi(\{0, \mathcal{M}, n+1\})$ be a permutation on $\{0\} \bigcup \mathcal{M} \bigcup \{n+1\}$ with $\pi(1) = 0$ and $\pi(M+2) = n+1$, and let $\pi(k_1 : k_2) = \{\pi(k_1), \pi(k_1+1), \ldots, \pi(k_2)\}$.

Under our unified relay framework as described in the Introduction, the following Theorems 2.1 and 2.2 present the achievable rates by block-by-block backward decoding and B-blocks-by-B-blocks backward decoding respectively. The coding schemes used to prove these theorems constitute the key contributions of our paper, and will be presented in detail in Sections III and IV respectively.

Theorem 2.1: For the multiple-relay channel, a rate R is achievable if for some $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$, there exists a permutation $\pi(\{0, \mathcal{M}, n+1\})$ and some

$$p(q)p(x_0|q)p(x_{\mathcal{M}}|x_0, q) \prod_{i \in \mathcal{N} \backslash \mathcal{M}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

such that for any $k = 2, 3, \ldots, M + 2$,

$$R < \min_{\mathcal{S} \subseteq \mathcal{D}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}, Y_{\pi(k)} | X_{\mathcal{D}_k \setminus \mathcal{S}}, X_{\pi(k:M+1)}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}, Q),$$

(19)

where $\mathcal{D}_k$ is the unique largest subset of $\mathcal{N} \setminus \mathcal{M}$ satisfying

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}, Y_{\pi(k)} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k \setminus \mathcal{S}}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}, Q) > 0, \qquad (20)$$

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_k$.

Theorem 2.2: For the multiple-relay channel, a rate R is achievable if for some $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$, there exists a permutation $\pi(\{0, \mathcal{M}, n+1\})$ and some

$$p(q)p(x_0|q)p(x_{\mathcal{M}}|x_0, q) \prod_{i \in \mathcal{N} \setminus \mathcal{M}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

such that for any $k = 2, 3, \ldots, M + 2$,

$$R < \max_{\mathcal{T}_k \subseteq \mathcal{N} \setminus \mathcal{M}} \min_{\mathcal{S} \subseteq \mathcal{T}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_k \setminus \mathcal{S}}, Y_{\pi(k)} | X_{\mathcal{T}_k \setminus \mathcal{S}}, X_{\pi(k:M+1)}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{T}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{T}_k \setminus \mathcal{S}}, Q).$$

(21)

The following theorem establishes the equivalence between the achievable rates in Theorems 2.1 and 2.2. The proof of this theorem can be immediately obtained by analogy to the proof of [20, Thm 2.8] and will be omitted in this paper.

Theorem 2.3: For any $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$, any permutation $\pi(\{0, \mathcal{M}, n+1\})$, any distribution

$$p(q)p(x_0|q)p(x_{\mathcal{M}}|x_0, q) \prod_{i \in \mathcal{N} \setminus \mathcal{M}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

and any $k = 2, 3, \ldots, M + 2$, the maximum in the R.H.S. of (21) is attained when $\mathcal{T}_k = \mathcal{D}_k$, where $\mathcal{D}_k$ is as defined in (20).

Remark 2.1: Finally, we point out that Theorems 2.1 and 2.2 can also be applied to multiple-destination problems, by choosing the D-F relay nodes set $\mathcal{M}$ to include the other destinations.

A. An example of Gaussian networks

We now apply the above results to additive white Gaussian noise (AWGN) networks and compare the achievable rates under our unified framework to the rates with D-F or C-F alone and with other unified schemes [16], [24]. In particular, we consider an AWGN two-relay channel with the geometry shown in Fig. 4, where nodes 0 and 3 are the source and destination respectively, and nodes 1 and 2 are the relay nodes. This channel is described by

$$Y_1 = g_{01}X_0 + g_{21}X_2 + Z_1,$$
$$Y_2 = g_{02}X_0 + g_{12}X_1 + Z_2,$$
$$Y_3 = g_{03}X_0 + g_{13}X_1 + g_{23}X_2 + Z_3,$$

in which, the channel gains are $g_{03} = 1$, $g_{01} = g_{23} = d^{-\alpha/2}$, $g_{02} = g_{13} = (1-d)^{-\alpha/2}$, $g_{12} = g_{21} = (1-2d)^{-\alpha/2}$, where α is the path-loss exponent; the noises Z_1, Z_2, Z_3 are zero-mean unit-variance Gaussian random variables that are independent of each other and the channel inputs. We further assume uniform average power constraint P on each transmitter $X_i, i = 0, 1, 2$, and noncoherent transmission among the transmitters, i.e., X_0, X_1, X_2 are independent.[2] The achievable rates with our unified schemes and various other schemes are evaluated for this two-relay channel in Appendix A. Fig. 5 plots these rates for $P = 10$, $\alpha = 2$, and $d \in (0, 0.5)$. As can be seen, the rates with our unified schemes are generally better than the rates with existing unified schemes and with D-F or C-F alone. Detailed analyses based on Fig. 5 are as follows.
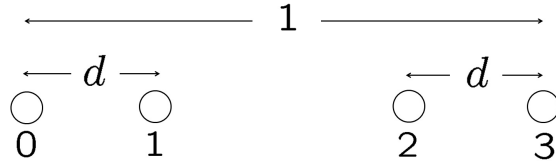
Fig. 4.   A two-relay channel.

(a) Our schemes vs. D-F or C-F alone

(b) Our schemes vs. Unified scheme in [16]
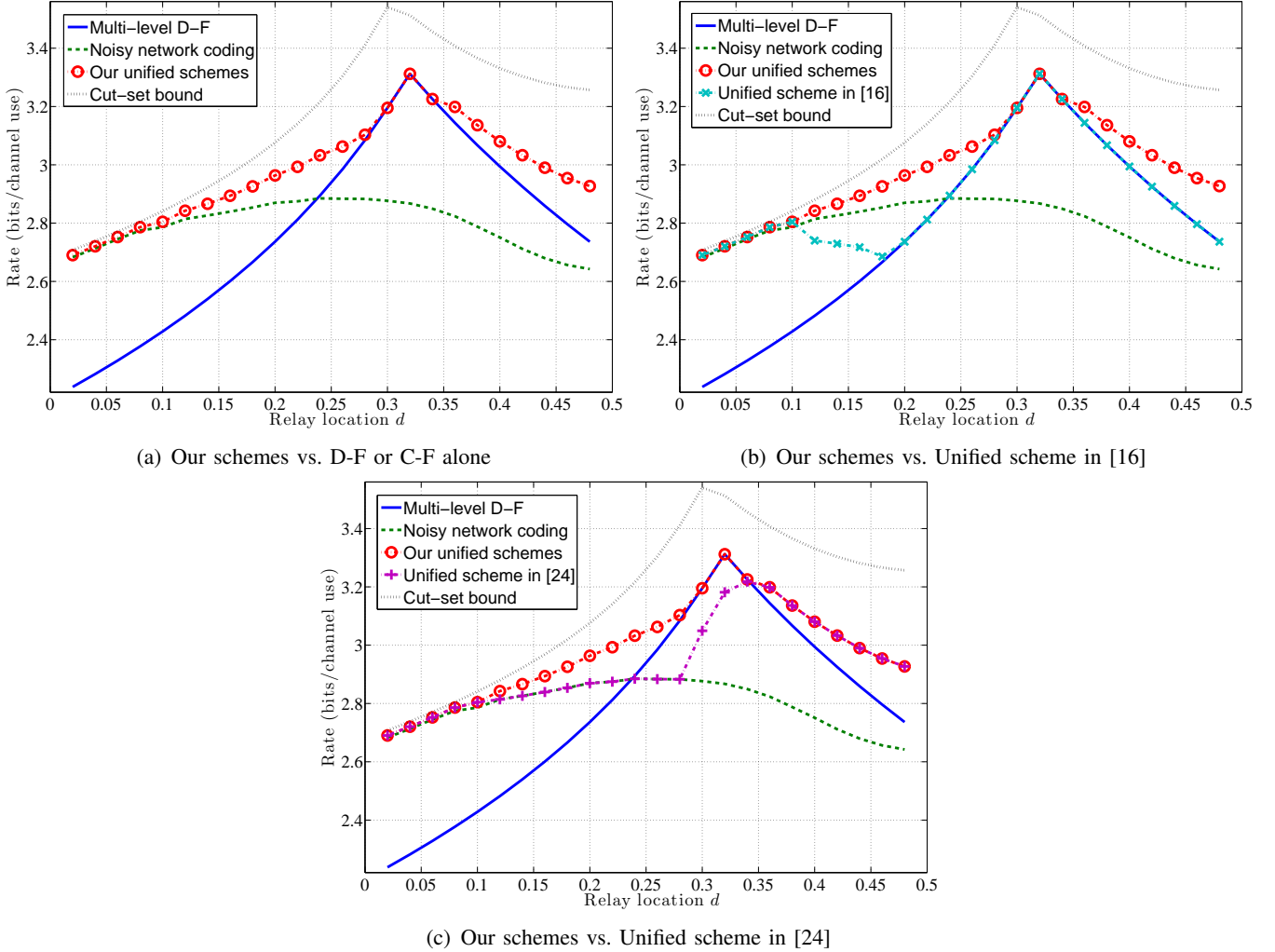
(c) Our schemes vs. Unified scheme in [24]

Fig. 5.   Achievable rates for the two-relay channel.

D-F vs. C-F: The D-F and C-F achievable rates plotted in Fig. 5 are the multi-level D-F rates [13]-[16] and noisy network coding rates [18]-[20] respectively. Roughly speaking, when d is relatively large, D-F performs better than C-F; when d is relatively small, C-F outperforms D-F. This is because, relatively large d means that both relay nodes are not far from the midpoint, which facilitates the use of multi-level D-F; while when d is small, i.e., when relay nodes 1 and 2 are respectively close to the source and destination, the D-F scheme still requires node 2 to be able to decode the message, thus limiting the achievable rate for the channel compared to the C-F scheme which has no requirement of decoding at either relay node. Also note here the highest D-F rate is achieved when d is around $1/3$, i.e., when relay nodes 1 and 2 are approximately evenly distributed along the line.

[2]The assumption of noncoherent transmission is made mainly to simplify the evaluation of the achievable rates; nevertheless, it does also reflect some realistic wireless communication scenarios, e.g., the phase-fading channels with the phase information unknown to the transmitter, so that coherent beamforming cannot be achieved [16].

Our unified schemes vs. D-F or C-F alone: With multi-level D-F and noisy network coding included as their special cases, our unified schemes achieve the same rate as multi-level D-F when d is around $1/3$, and provide considerable gains on the achievable rates against using D-F or C-F alone as d decreases or increases from around $1/3$. The interpretation is that, in our unified schemes, for d around $1/3$, nodes 1 and 2 should perform multi-level D-F; otherwise, a better choice is to let node 1 perform D-F while let node 2 perform C-F. It is not surprising that as d decreases from around $1/3$, i.e., as nodes 1 and 2 move towards the source and destination respectively, introducing the freedom of choosing nodes 1 and 2 to be the D-F and C-F node respectively can improve the achievable rates, compared to performing D-F or C-F alone at both relay nodes. However, it may not be so obvious why this choice of the D-F and C-F node also leads to better achievable rates than multi-level D-F as d increases from around $1/3$, i.e., as relay nodes 1 and 2 move towards each other. For this, one has to realize that in our schemes, D-F node 1, when decoding the message from source node 0, also exploits the help of C-F node 2, which could result in a looser rate constraint for decoding at D-F node 1 than in multi-level D-F, thus potentially improving the achievable rate for the channel. In contrast, if the D-F node does not utilize the help of the C-F node, which is the case in the unified scheme of [16], then, as d increases from around $1/3$, this gain will not happen.

Our unified schemes vs. unified scheme in [16]: Generally, the unified scheme in [16] also improves D-F alone by introducing the flexibility of choosing node 2 to be C-F node, but cannot achieve the same rates as our unified schemes. The reasons are two-fold as mentioned in Subsection I-C, i.e., i) it incorporates an inferior C-F scheme instead of noisy network coding, and ii) it does not allow the D-F node to utilize the help of the C-F node. For the first reason, it cannot achieve the noisy network coding rates for d roughly between 0.1 and 0.24. Due to the second reason, it cannot provide any gain against using multi-level D-F as d increases from around $1/3$, in contrast to our unified schemes.

Our unified schemes vs. unified scheme in [24]: Our unified schemes also outperform the unified scheme in [24]. Recall that there are two drawbacks in the scheme of [24]. First, in the scheme of [24], all the D-F relay nodes are at the same level, and thus it cannot achieve the same rates as multi-level D-F for $d \in (0.24, 0.34)$. Second, it does not allow the D-F node to perform backward decoding to fully utilize the help of the C-F node as in our schemes, which, together with the first drawback, leads to lower achievable rates than our schemes for $d \in (0.1, 0.36)$.

III. UNIFIED RELAY FRAMEWORK WITH BLOCK-BY-BLOCK BACKWARD DECODING

To prove Theorem 2.1, we incorporate the multi-level D-F scheme in [13]-[16] and the cumulative encoding/block-by-block backward decoding/comression-message joint decoding C-F scheme in [20] into the unified relay framework described in the Introduction.

Specifically, we divide the relay set $\mathcal{N}$ into two sets, $\mathcal{M}$ with $|\mathcal{M}| = M$ and $\mathcal{N} \setminus \mathcal{M}$, as shown in Fig. 3, and fix some permutation $\pi(\{0, \mathcal{M}, n+1\})$ with $\pi(1) = 0$ and $\pi(M+2) = n+1$. The source performs cumulative encoding, in the sense that a new message is encoded at the source in each new block; the nodes in $\mathcal{M}$ perform the multi-level D-F cooperatively, along the route $\pi(1) \to \pi(2) \to \cdots \to \pi(M+2)$, in a similar manner with [13]-[16]; each node $i \in \mathcal{N} \setminus \mathcal{M}$ performs C-F independently in the same way as [18]-[20]; both the D-F relay nodes and the destination node, i.e., nodes $\pi(2 : M+2)$, perform compression-message joint decoding in a block-by-block backward manner. (Note here, the nodes $\pi(2 : M + 2)$ will be treated as multiple destinations with respect to the C-F relay nodes, and thus compression-message joint decoding is generally necessary at the these nodes, as mentioned in the Introduction.) A total of B^{M+1} blocks will be used and the length of a "virtual" block for node $\pi(k), k = 2, 3 \ldots, M + 2$, will be B^{k-2} blocks. The backward decoding at the destination, i.e., node $\pi(M + 2)$, will happen at the end of all B^{M+1} blocks, while the backward decoding at the D-F relay node $\pi(k), k = 2, 3, \ldots, M + 1$, will happen at the end of every B^{k-1} blocks, i.e., at the end of block $b = vB^{k-1}, v \in [1 : B^{M+1}/B^{k-1}]$.

To make the presentation of the detailed coding scheme easier to follow, we first consider the case of single D-F relay node, i.e., when $M = 1$, and then present the extension to the general case of multiple D-F relay nodes, i.e., when $M \geq 2$.

A. Single D-F relay node (M = 1)

Assume that, among the relay nodes set $\mathcal{N}$, only node 1 is the D-F relay node, and all other relay nodes are the C-F relay nodes. Denote $\tilde{\mathcal{N}} = \mathcal{N} \setminus \{1\}$. Specializing Theorem 2.1 to this case, we have that a rate R is achievable, if there exists some

$$p(q)p(x_0|q)p(x_1|x_0, q) \prod_{i \in \tilde{\mathcal{N}}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

such that

$$R < \min \left\{ \begin{array}{l} \min\limits_{\mathcal{S} \subseteq \mathcal{D}_1} I(X_0, X_\mathcal{S}; \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}}, Y_1 | X_1, X_{\mathcal{D}_1 \setminus \mathcal{S}}, Q) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | X_0, X_1, X_{\mathcal{D}_1}, Y_1, \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}}, Q) \\ \min\limits_{\mathcal{S} \subseteq \mathcal{D}_{n+2}} I(X_0, X_1, X_\mathcal{S}; \hat{Y}_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Y_{n+2} | X_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Q) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | X_0, X_1, X_{\mathcal{D}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Q) \end{array} \right\}$$
(22)

where $\mathcal{D}_1$ is the unique largest subset of $\tilde{\mathcal{N}}$ satisfying

$$I(X_\mathcal{S}; \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}}, Y_1 | X_0, X_1, X_{\mathcal{D}_1 \setminus \mathcal{S}}, Q) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | X_0, X_1, X_{\mathcal{D}_1}, Y_1, \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}}, Q) > 0,$$
(23)

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_1$, and $\mathcal{D}_{n+2}$ is the unique largest subset of $\tilde{\mathcal{N}}$ satisfying

$$I(X_\mathcal{S}; \hat{Y}_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Y_{n+2} | X_0, X_1, X_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Q) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | X_0, X_1, X_{\mathcal{D}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Q) > 0,$$
(24)

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_{n+2}$.

The uniqueness of $\mathcal{D}_1$ and $\mathcal{D}_{n+2}$ can be immediately obtained by analogy to the proof of [20, Thm 2.7]. Below, we focus on proving the achievablity of the rate in (22). For simplicity of notation, we only prove the achievability for the case $Q = \emptyset$. Achievability for an arbitrary time-sharing random variable Q can be obtained by using the standard technique of time sharing [26], [8]. The same consideration on Q applies throughout all the proofs of this paper.

In the case of single D-F relay node, a total of B^2 blocks will be used; the backward decoding at the destination node $n + 2$ will happen at the end of all B^2 blocks, while the backward decoding at the D-F relay node 1 will happen at the end of every B blocks, i.e., at the end of block $b = vB, v \in [1 : B]$. See Fig. 6 for an illustration. Note here, in order to fully utilize the help of the C-F nodes as in [18]-[20], even the only D-F relay node 1, has to perform backward decoding, which is different from the situation arising in [15]-[16] and [25], where there is no issue of exploiting the help of the C-F nodes and node 1 can decode at the end of every block. The detailed codebook generation and encoding/decoding process are as follows, which can be understood with the help of Table I.
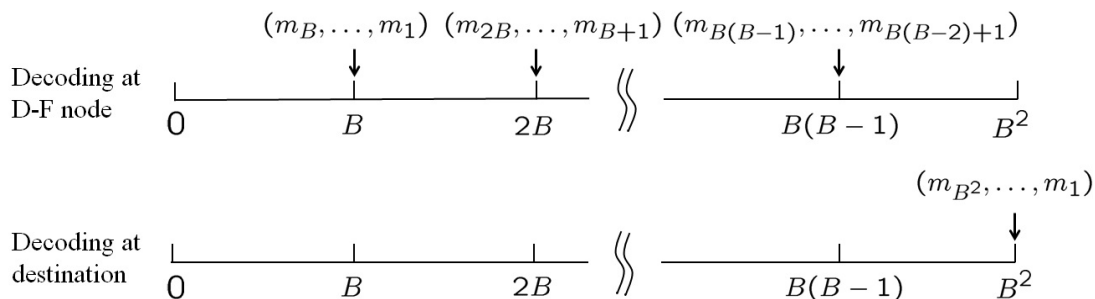

Fig. 6. An illustration of nested blocks with backward decoding.

Codebook Generation: Fix $p(x_0)p(x_1|x_0) \prod_{i \in \tilde{\mathcal{N}}} p(x_i)p(\hat{y}_i|y_i, x_i)$. We randomly and independently generate a codebook for each block.

i) First consider the codebook generation for the source node 0 and the D-F relay node 1. A joint codebook for these two nodes will be generated in a backward manner similar to [14] for each block. Specifically, for each block $b \in [1 : B^2]$, randomly generate 2^{TR} independent sequences $\mathbf{x}_{1,b}(m_{b-B})$ for

node 1, and randomly generate 2^{TR} conditionally independent sequences $\mathbf{x}_{0,b}(m_b|m_{b-B})$ for node 0, where $m_b, m_{b-B} \in [1 : 2^{TR}]$. As in [14], the codebook is generated in the backward manner because the source node 0 knows what the D-F relay node 1 is going to transmit, and therefore can adjust its own transmission accordingly, but not the converse. The difference from [14] is that here the delay between the messages transmitted by node 1 and node 0 is B blocks, instead of 1 block in [14], since in our framework node 1 has to wait for every B blocks to perform backward decoding for exploiting the help of the C-F relay nodes.

ii) Then we generate the codebooks for the C-F relay nodes in the same way as in [18]-[20]. For each block $b \in [1 : B^2]$ and each relay node $i \in \tilde{\mathcal{N}}$, randomly and independently generate $2^{T\hat{R}_i}$ sequences $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, where $\hat{R}_i = I(Y_i; \hat{Y}_i|X_i) + \epsilon$; for each relay node $i \in \tilde{\mathcal{N}}$ and each $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, randomly and conditionally independently generate $2^{T\hat{R}_i}$ sequences $\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})$, $l_{i,b} \in [1 : 2^{T\hat{R}_i}]$.

The combination of i) and ii) defines the codebook for any block $b \in [1 : B^2]$,

$$\mathcal{C}_b = \Big\{ \mathbf{x}_{1,b}(m_{b-B}), \mathbf{x}_{0,b}(m_b|m_{b-B}) : m_b, m_{b-B} \in [1 : 2^{TR}];$$

$$\mathbf{x}_{i,b}(l_{i,b-1}), \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}) : l_{i,b}, l_{i,b-1} \in [1 : 2^{T\hat{R}_i}], i \in \tilde{\mathcal{N}} \Big\}. \tag{25}$$

Encoding: Let $\mathbf{m} = (m_1, m_2, \ldots, m_{B^2})$ be the message vector to be sent and let $m_b = 1$ be the dummy message for any

$$b \in \cup_{w=1}^{B}[wB - L + 1 : wB] \bigcup [(B-1)B + 1 : B^2] \tag{26}$$

and for any $b \leq 0$. As we will see, these dummy messages are inserted to ensure the start of block-by-block backward decoding. Due to these dummy messages, the actually achievable rate becomes $\frac{(B-L)(B-1)}{B^2}R$, which, however, can be made arbitrarily close to R by choosing $L \ll B$, i.e., the rate loss $\frac{B(L+1)-L}{B^2}R$ can always be made arbitrarily small.

TABLE I
BLOCK-BY-BLOCK BACKWARD DECODING FOR THE SINGLE D-F RELAY NODE CASE

Block	1	2	$\cdots$	$B-L$	$B-L+1$	$\cdots$	B	$\cdots$
X_0	$\mathbf{x}_{0,1}(m_1|1)$	$\mathbf{x}_{0,2}(m_2|1)$	$\cdots$	$\mathbf{x}_{0,B-L}(m_{B-L}|1)$	$\mathbf{x}_{0,B-L+1}(1|1)$	$\cdots$	$\mathbf{x}_{0,B}(1|1)$	$\cdots$
Y_1	$\emptyset$	$\emptyset$	$\cdots$	$\emptyset$	$\emptyset$	$\cdots$	$(m_1, m_2, \ldots, m_B)$	$\cdots$
X_1	$\mathbf{x}_{1,1}(1)$	$\mathbf{x}_{1,2}(1)$	$\cdots$	$\mathbf{x}_{1,B-L}(1)$	$\mathbf{x}_{1,B-L+1}(1)$	$\cdots$	$\mathbf{x}_{1,B}(1)$	$\cdots$
$Y_{\tilde{\mathcal{N}}}$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},1}(l_{\tilde{\mathcal{N}},1}|1)$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},2}(l_{\tilde{\mathcal{N}},2}|l_{\tilde{\mathcal{N}},1})$	$\cdots$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B-L}(l_{\tilde{\mathcal{N}},B-L}|l_{\tilde{\mathcal{N}},B-L-1})$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B-L+1}(l_{\tilde{\mathcal{N}},B-L+1}|l_{\tilde{\mathcal{N}},B-L})$	$\cdots$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B}(l_{\tilde{\mathcal{N}},B}|l_{\tilde{\mathcal{N}},B-1})$	$\cdots$
$X_{\tilde{\mathcal{N}}}$	$\mathbf{x}_{\tilde{\mathcal{N}},1}(\mathbf{1})$	$\mathbf{x}_{\tilde{\mathcal{N}},2}(l_{\tilde{\mathcal{N}},1})$	$\cdots$	$\mathbf{x}_{\tilde{\mathcal{N}},B-L}(l_{\tilde{\mathcal{N}},B-L-1})$	$\mathbf{x}_{\tilde{\mathcal{N}},B-L+1}(l_{\tilde{\mathcal{N}},B-L})$	$\cdots$	$\mathbf{x}_{\tilde{\mathcal{N}},B}(l_{\tilde{\mathcal{N}},B-1})$	$\cdots$
Y_{n+2}	$\emptyset$	$\emptyset$	$\cdots$	$\emptyset$	$\emptyset$	$\cdots$	$\emptyset$	$\cdots$

Block	B^2-B+1	$\cdots$	B^2-L	B^2-L+1	$\cdots$	B^2
X_0	$\mathbf{x}_{0,B^2-B+1}(1|m_{B^2-2B+1})$	$\cdots$	$\mathbf{x}_{0,B^2-L}(1|m_{B^2-B-L})$	$\mathbf{x}_{0,B^2-L+1}(1|1)$	$\cdots$	$\mathbf{x}_{0,B^2}(1|1)$
Y_1	$\emptyset$	$\cdots$	$\emptyset$	$\emptyset$	$\cdots$	$(m_{B^2-B+1}, \ldots, m_{B^2})$
X_1	$\mathbf{x}_{1,B^2-B+1}(m_{B^2-2B+1})$	$\cdots$	$\mathbf{x}_{1,B^2-L}(m_{B^2-B-L})$	$\mathbf{x}_{1,B^2-L+1}(1)$	$\cdots$	$\mathbf{x}_{1,B^2}(1)$
$Y_{\tilde{\mathcal{N}}}$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-B+1}(l_{\tilde{\mathcal{N}},B^2-B+1}|l_{\tilde{\mathcal{N}},B^2-B})$	$\cdots$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-L}(l_{\tilde{\mathcal{N}},B^2-L}|l_{\tilde{\mathcal{N}},B^2-L-1})$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-L+1}(l_{\tilde{\mathcal{N}},B^2-L+1}|l_{\tilde{\mathcal{N}},B^2-L})$	$\cdots$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2}(l_{\tilde{\mathcal{N}},B^2}|l_{\tilde{\mathcal{N}},B^2-1})$
$X_{\tilde{\mathcal{N}}}$	$\mathbf{x}_{\tilde{\mathcal{N}},B^2-B+1}(l_{\tilde{\mathcal{N}},B^2-B})$	$\cdots$	$\mathbf{x}_{\tilde{\mathcal{N}},B^2-L}(l_{\tilde{\mathcal{N}},B^2-L-1})$	$\mathbf{x}_{\tilde{\mathcal{N}},B^2-L+1}(l_{\tilde{\mathcal{N}},B^2-L})$	$\cdots$	$\mathbf{x}_{\tilde{\mathcal{N}},B^2}(l_{\tilde{\mathcal{N}},B^2-1})$
Y_{n+2}	$\emptyset$	$\cdots$	$\emptyset$	$\emptyset$	$\cdots$	$(m_1, m_2, \ldots, m_{B^2})$

i) First consider the encoding process for nodes 0 and 1.
- In block $b \in [1 : B^2]$, the source node 0 transmits $\mathbf{x}_{0,b}(m_b|m_{b-B})$.
- At the end of block vB, $v \in [1 : B-1]$, the D-F relay node 1 has decoded messages

$$(m_{vB-B+1}, m_{vB-B+2}, \ldots, m_{vB})$$

using backward decoding (see the decoding part). In the next B blocks, i.e., in block $b \in [vB + 1 : (v + 1)B]$, the relay node 1 transmits $\mathbf{x}_{1,b}(m_{b-B})$, where m_{b-B} for any $b \in [vB + 1 : (v+1)B])$ has been decoded by block vB.

ii) For any block $b \in [1 : B^2]$, each relay node $i \in \tilde{\mathcal{N}}$, upon receiving $\mathbf{y}_{i,b}$ at the end of block b, finds an index $l_{i,b}$ such that

$$(\mathbf{x}_{i,b}(l_{i,b-1}), \mathbf{y}_{i,b}, \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})) \in A_\epsilon(X_i, Y_i, \hat{Y}_i),$$

where $l_{i,0} = 1$ by convention. In block $b \in [1 : B^2]$, the relay node $i \in \tilde{\mathcal{N}}$ transmits $\mathbf{x}_{i,b}(l_{i,b-1})$.

Decoding: We present the decoding process at the D-F relay node 1 and at the destination node $n + 2$ separately.

i) At the end of block $b = vB, v \in [1 : B]$, the D-F relay node 1 decodes messages

$$(m_{b-B+1}, m_{b-B+2}, \ldots, m_b)$$

using block-by-block backward decoding. In fact, among these messages, $(m_{b-L+1}, m_{b-L+2}, \ldots, m_b)$ are dummy messages according to (26) and only $(m_{b-B+1}, m_{b-B+2}, \ldots, m_{b-L})$ need decoding.

- a) Node 1 first finds the unique $l_{\mathcal{D}_1,b-L} = \{l_{i,b-L}, i \in \mathcal{D}_1\}$ such that there exists some $l^b_{\mathcal{D}_1,b-L+1}$ satisfying that for any block $j = b - L + 1, b - L + 2, \ldots, b$,

$$(\mathbf{X}_{0,j}(m_j|m_{j-B}), \mathbf{X}_{1,j}(m_{j-B}), \{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_1\}, \mathbf{Y}_{1,j}) \in A_\epsilon(X_0, X_1, X_{\mathcal{D}_1}, \hat{Y}_{\mathcal{D}_1}, Y_1).$$
$$(27)$$

Note in (27), for any $j = b - L + 1, b - L + 2, \ldots, b$, m_j and m_{j-B} are both dummy messages according to (26), and both $\mathbf{X}_{0,j}(m_j|m_{j-B})$ and $\mathbf{X}_{1,j}(m_{j-B})$ are known at node 1. Then, it follows from the proof of [20, Thm 2.7] that $l_{\mathcal{D}_1,b-L}$ can be decoded if

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_1 \backslash \mathcal{S}}, Y_1 | X_0, X_1, X_{\mathcal{D}_1 \backslash \mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{D}_1}, Y_1, \hat{Y}_{\mathcal{D}_1 \backslash \mathcal{S}}) > 0, \quad (28)$$

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_1$. For self-containedness, we also include the details on the derivation of (28) in Appendix B.

- b) Backwardly and sequentially from block $j = b - L$ to $j = b - B + 1$, node 1 finds the unique pair $(m_j, l_{\mathcal{D}_1,j-1})$ satisfying (27), where $l_{\mathcal{D}_1,j}$ has already been recovered due to the backward property of decoding, and m_{j-B} has been decoded by block $b - B$.

  At each block $j = b - L, b - L - 1, \ldots, b - B + 1$, error occurs with m_j if the true m_j does not satisfy (27) with any $l_{\mathcal{D}_1,j-1}$, or a false m_j satisfies (27) with some $l_{\mathcal{D}_1,j-1}$. According to the properties of typical sequences, the true $(m_j, l_{\mathcal{D}_1,j-1})$ satisfies (27) with high probability.

  For a false m_j and a $l_{\mathcal{D}_1,j-1}$ with false $\{l_{i,j-1}, i \in \mathcal{S}\}$ but true $\{l_{i,j-1}, i \in \mathcal{D}_1 \backslash \mathcal{S}\}$, $\mathbf{X}_{0,j}(m_j|m_{j-B})$ is conditionally independent of $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_1\}$ and $\mathbf{Y}_{1,j}$ given $\mathbf{X}_{1,j}(m_{j-B})$; and $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{S}\}$ are independent of $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_1 \backslash \mathcal{S}\}$, $\mathbf{X}_{1,j}(m_{j-B})$ and $\mathbf{Y}_{1,j}$.

  Therefore, the probability that such false $(m_j, l_{\mathcal{D}_1,j-1})$ satisfies (27) can be upper bounded by

$$2^{T(H(X_0, X_1, X_{\mathcal{D}_1}, \hat{Y}_{\mathcal{D}_1}, Y_1) + \epsilon)} 2^{-T(H(X_1, X_{\mathcal{D}_1 \backslash \mathcal{S}}, \hat{Y}_{\mathcal{D}_1 \backslash \mathcal{S}}, Y_1) - \epsilon)}$$

$$\times 2^{-T(H(X_0|X_1) - \epsilon)} 2^{-T(H(X_{\mathcal{S}}) - \epsilon)} 2^{-T(\sum_{i \in \mathcal{S}} (H(\hat{Y}_i|X_i) - \epsilon))}.$$

Since the number of such false $(m_j, l_{\mathcal{D}_1,j-1})$ is upper bounded by $2^{TR} \prod_{i \in \mathcal{S}} 2^{T(I(Y_i; \hat{Y}_i|X_i) + \epsilon)}$, with the union bound, it is easy to check that the probability of finding a false m_j goes to zero as $T \to \infty$, if

$$R < \min_{\mathcal{S} \subseteq \mathcal{D}_1} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_1 \backslash \mathcal{S}}, Y_1 | X_1, X_{\mathcal{D}_1 \backslash \mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{D}_1}, Y_1, \hat{Y}_{\mathcal{D}_1 \backslash \mathcal{S}}). \quad (29)$$

Then, based on the recovered m_j, m_{j-B}, and $l_{\mathcal{D}_1,j}$, again from the proof of [20, Thm 2.7], it follows that $l_{\mathcal{D}_1,j-1}$ can be decoded if (28) holds.

By a) and b) together, at the end of block $b = vB, v \in [1 : B]$, the D-F relay node 1 can decode messages $(m_{b-B+1}, m_{b-B+2}, \ldots, m_b)$ if both (28) and (29) hold.

ii) At the end of all B^2 block, the destination node $n + 2$ decodes messages $(m_1, m_2, \ldots, m_{B^2})$ using block-by-block backward decoding. Similarly, we only consider the decoding of $(m_1, m_2, \ldots, m_{B^2-B-L})$, since $(m_{B^2-B-L+1}, m_{B^2-B-L+2}, \ldots, m_{B^2})$ are all dummy messages according to (26).

- a) Node $n + 2$ first finds the unique $l_{\mathcal{D}_{n+2}, B^2-L} = \{l_{i,B^2-L}, i \in \mathcal{D}_{n+2}\}$ such that there exists some $l_{\mathcal{D}_{n+2}, B^2-L+1}^{B^2}$ satisfying that for any block $j = B^2 - L + 1, B^2 - L + 2, \ldots, B^2$,

$$(\mathbf{X}_{0,j}(m_j|m_{j-B}), \mathbf{X}_{1,j}(m_{j-B}), \{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_{n+2}\}, \mathbf{Y}_{n+2,j})$$
$$\in A_\epsilon(X_0, X_1, X_{\mathcal{D}_{n+2}}, \hat{Y}_{\mathcal{D}_{n+2}}, Y_{n+2}), \tag{30}$$

where, similarly, m_j and m_{j-B} are both dummy messages according to (26), and $\mathbf{X}_{0,j}(m_j|m_{j-B})$ and $\mathbf{X}_{1,j}(m_{j-B})$ are both known at node $n + 2$. Still, from the proof of [20, Thm 2.7], $l_{\mathcal{D}_{n+2}, B^2-L}$ can be decoded if

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_{n+2}\setminus\mathcal{S}}, Y_{n+2}|X_0, X_1, X_{\mathcal{D}_{n+2}\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_1, X_{\mathcal{D}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{D}_{n+2}\setminus\mathcal{S}}) > 0, \tag{31}$$

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_{n+2}$.

- b) Backwardly and sequentially from block $j = B^2 - L$ to $j = 1$, node $n + 2$ finds the unique pair $(m_{j-B}, l_{\mathcal{D}_{n+2}, j-1})$ satisfying (30), where $l_{\mathcal{D}_{n+2}, j}$ has already been recovered due to the backward property of decoding, and m_j either is a dummy message (for $j = B^2 - L, B^2 - L - 1, \ldots, B^2 - B - L + 1$) or has been decoded due to the backward property of decoding (for $j = B^2 - B - L, B^2 - B - L - 1, \ldots, 1$).

  At each block $j = B^2 - L, B^2 - L - 1, \ldots, 1$, error occurs with m_{j-B} if the true m_{j-B} does not satisfy (30) with any $l_{\mathcal{D}_{n+2}, j-1}$, or a false m_{j-B} satisfies (30) with some $l_{\mathcal{D}_{n+2}, j-1}$. According to the properties of typical sequences, the true $(m_{j-B}, l_{\mathcal{D}_{n+2}, j-1})$ satisfies (30) with high probability.

  For a false m_{j-B} and a $l_{\mathcal{D}_{n+2}, j-1}$ with false $\{l_{i,j-1}, i \in \mathcal{S}\}$ but true $\{l_{i,j-1}, i \in \mathcal{D}_{n+2} \setminus \mathcal{S}\}$, $\mathbf{X}_{0,j}(m_j|m_{j-B})$ and $\mathbf{X}_{1,j}(m_{j-B})$ are independent of $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_{n+2}\}$ and $\mathbf{Y}_{n+2,j}$; and $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{S}\}$ are independent of $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_{n+2} \setminus \mathcal{S}\}$ and $\mathbf{Y}_{n+2,j}$.

  Therefore, the probability that such false $(m_j, l_{\mathcal{D}_{n+2}, j-1})$ satisfies (30) can be upper bounded by

$$2^{T(H(X_0, X_1, X_{\mathcal{D}_{n+2}}, \hat{Y}_{\mathcal{D}_{n+2}}, Y_{n+2}) + \epsilon)} 2^{-T(H(X_{\mathcal{D}_{n+2}\setminus\mathcal{S}}, \hat{Y}_{\mathcal{D}_{n+2}\setminus\mathcal{S}}, Y_{n+2}) - \epsilon)}$$
$$\times 2^{-T(H(X_0, X_1) - \epsilon)} 2^{-T(H(X_{\mathcal{S}}) - \epsilon)} 2^{-T(\sum_{i \in \mathcal{S}}(H(\hat{Y}_i|X_i) - \epsilon))}.$$

Since the number of such false $(m_j, l_{\mathcal{D}_{n+2}, j-1})$ is upper bounded by $2^{TR} \prod_{i \in \mathcal{S}} 2^{T(I(Y_i; \hat{Y}_i|X_i) + \epsilon)}$, with the union bound, it is easy to check that the probability of finding a false m_j goes to zero as $T \to \infty$, if

$$R < \min_{\mathcal{S} \subseteq \mathcal{D}_1} I(X_0, X_1, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_{n+2}\setminus\mathcal{S}}, Y_{n+2}|X_{\mathcal{D}_{n+2}\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_1, X_{\mathcal{D}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{D}_{n+2}\setminus\mathcal{S}}). \tag{32}$$

  Then, similarly, based on the recovered m_j, m_{j-B}, and $l_{\mathcal{D}_{n+2}, j}$, $l_{\mathcal{D}_{n+2}, j-1}$ can be decoded if (31) holds.

  By a) and b) together, at the end of all B^2 block, the destination node $n + 2$ can decode messages $(m_1, m_2, \ldots, m_{B^2})$ if both (31) and (32) hold.

Combining i) and ii), and using the standard technique of time sharing, we conclude that the rate described in (22)-(24) is achievable.

B. Multiple D-F relay nodes ($M \geq 2$)

When there are multiple D-F relay nodes, i.e., $M \geq 2$, a total of B^{M+1} blocks will be used. The detailed codebook generation and encoding/decoding process are as follows.

Codebook Generation: Fix $p(x_0)p(x_{\mathcal{M}}|x_0) \prod_{i \in \mathcal{N} \setminus \mathcal{M}} p(x_i)p(\hat{y}_i|y_i, x_i)$. We randomly and independently generate a codebook for each block.

i) First consider the codebook generation for nodes $\pi(1 : M + 1)$.

- For each block $b \in [1 : B^{M+1}]$, backwardly and sequentially for each relay node $\pi(k), k = M + 1, M, \ldots, 2$, randomly generate 2^{TR} conditionally independent sequences

$$\mathbf{x}_{\pi(k),b}(m_{b-B^{k-1}}|m_{b-B^k}, \ldots, m_{b-B^M}),$$

where $m_{b-B^{k-1}}, m_{b-B^k}, \ldots, m_{b-B^M} \in [1 : 2^{TR}]$;

- For each block $b \in [1 : B^{M+1}]$ and node $\pi(1)$, i.e., the source node 0, randomly generate 2^{TR} conditionally independent sequences $\mathbf{x}_{0,b}(m_b|m_{b-B}, \ldots, m_{b-B^M})$, where

$$m_b, m_{b-B}, \ldots, m_{b-B^M} \in [1 : 2^{TR}].$$

ii) The codebook generation for the nodes in $\mathcal{N} \setminus \mathcal{M}$ is the same as that in the case of $M = 1$. For each block $b \in [1 : B^{M+1}]$ and each relay node $i \in \mathcal{N} \setminus \mathcal{M}$, randomly and independently generate $2^{T\hat{R}_i}$ sequences $\mathbf{x}_{i,b}(l_{i,b-1}), l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, where $\hat{R}_i = I(Y_i; \hat{Y}_i|X_i) + \epsilon$; for each relay node $i \in \mathcal{N} \setminus \mathcal{M}$ and each $\mathbf{x}_{i,b}(l_{i,b-1}), l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, randomly and conditionally independently generate $2^{T\hat{R}_i}$ sequences $\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}), l_{i,b} \in [1 : 2^{T\hat{R}_i}]$.

The combination of i) and ii) defines the codebook for any block $b \in [1 : B^{M+1}]$,

$$\mathcal{C}_b = \Big\{ \mathbf{x}_{\pi(k),b}(m_{b-B^{k-1}}|m_{b-B^k}, \ldots, m_{b-B^M}) : m_{b-B^{k-1}}, m_{b-B^k}, \ldots, m_{b-B^M} \in [1 : 2^{TR}], k = M+1, M, \ldots, 2;$$

$$\mathbf{x}_{0,b}(m_b|m_{b-B}, \ldots, m_{b-B^M}) : m_b, m_{b-B}, \ldots, m_{b-B^M} \in [1 : 2^{TR}];$$

$$\mathbf{x}_{i,b}(l_{i,b-1}), \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}) : l_{i,b}, l_{i,b-1} \in [1 : 2^{T\hat{R}_i}], i \in \mathcal{N} \setminus \mathcal{M} \Big\}. \tag{33}$$

Encoding: Let $\mathbf{m} = (m_1, m_2, \ldots, m_{B^{M+1}})$ be the message vector to be sent and let $m_b = 1$ be the dummy message for any

$$b \in \cup_{w=1}^{B^M}[wB - L + 1 : wB] \bigcup \cup_{u=1}^M \cup_{v=1}^{B^{M-u}} [v(B-1)B^u + 1 : vB^{u+1}], \tag{34}$$

and for any $b \leq 0$. Now, the actually achievable rate is $\frac{B-L}{B}(\frac{B-1}{B})^M R$ due to the dummy messages, which can still be made arbitrarily close to R by choosing $L \ll B$, similarly as in the proof for the single D-F relay node case in Subsection III-A.

i) We still first consider the encoding process for nodes $\pi(1 : M + 1)$.

- In block $b \in [1 : B^{M+1}]$, node $\pi(1)$, i.e., the source node 0, transmits $\mathbf{x}_{0,b}(m_b|m_{b-B}, \ldots, m_{b-B^M})$.
- By the end of block $vB^{k-1}, v \in [1 : \frac{B^{M+1}}{B^{k-1}} - 1]$, the D-F relay node $\pi(k), k = 2, \ldots, M + 1$, has decoded messages $(m_1, m_2, \ldots, m_{vB^{k-1}})$ using backward decoding (see the decoding part). In the next B^{k-1} blocks, i.e., in block $b \in [vB^{k-1}+1 : (v+1)B^{k-1}]$, node $\pi(k), k = 2, \ldots, M+1$, transmits $\mathbf{x}_{\pi(k),b}(m_{b-B^{k-1}}|m_{b-B^k}, \ldots, m_{b-B^M})$, where

$$(m_{b-B^{k-1}}, m_{b-B^k}, \ldots, m_{b-B^M}), b \in [vB^{k-1} + 1 : (v+1)B^{k-1}]$$

have all been decoded by block vB^{k-1}.

ii) The encoding process for the nodes in $\mathcal{N} \setminus \mathcal{M}$ is still the same as that in the case of $M = 1$. For any block $b \in [1 : B^{M+1}]$, each relay node $i \in \mathcal{N} \setminus \mathcal{M}$, upon receiving $\mathbf{y}_{i,b}$ at the end of block b, finds an index $l_{i,b}$ such that

$$(\mathbf{x}_{i,b}(l_{i,b-1}), \mathbf{y}_{i,b}, \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})) \in A_\epsilon(X_i, Y_i, \hat{Y}_i),$$

where $l_{i,0} = 1$ by convention. In block $b \in [1 : B^{M+1}]$, the relay node $i \in \mathcal{N} \setminus \mathcal{M}$ transmits $\mathbf{x}_{i,b}(l_{i,b-1})$.

Decoding: At the end of block $b = vB^{k-1}, v \in [1 : B^{M+1}/B^{k-1}]$, the node $\pi(k), k = 2, \ldots, M+2$, decodes messages $(m_{b-B^{k-1}+1}, \ldots, m_b)$ using block-by-block backward decoding as follows.

i) The node $\pi(k), k = 2, \ldots, M+2$, first finds the unique $l_{\mathcal{D}_k, b-L} = \{l_{i,b-L}, i \in \mathcal{D}_k\}$ such that there exists some $l_{\mathcal{D}_k, b-L+1}^b$ satisfying that for any block $j = b-L+1, b-L+2, \ldots, b$,

$$(\mathbf{X}_{0,j}(m_j|m_{j-B}, \ldots, m_{j-B^M}),$$
$$\{\mathbf{X}_{\pi(s),j}(m_{j-B^{s-1}}|m_{j-B^s}, \ldots, m_{j-B^M}), s = 2, \ldots, k-1, k, k+1, \ldots, M+1\},$$
$$\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_k\}, \mathbf{Y}_{\pi(k),j}) \in A_\epsilon(X_0, X_\mathcal{M}, X_{\mathcal{D}_k}, \hat{Y}_{\mathcal{D}_k}, Y_{\pi(k)}). \tag{35}$$

Note in (35), $(m_j, m_{j-B}, \ldots, m_{j-B^M}), j = b-L+1, b-L+2, \ldots, b$ are all dummy messages according to (34), and thus $\mathbf{X}_{\pi(s),j}, s = 1, \ldots, M+1$ are all known at node $\pi(k)$. Then, it follows from the proof of [20, Thm 2.7] that $l_{\mathcal{D}_k, b-L}$ can be decoded if

$$I(X_\mathcal{S}; \hat{Y}_{\mathcal{D}_k\setminus\mathcal{S}}, Y_{\pi(k)}|X_{\pi(1:M+1)}, X_{\mathcal{D}_k\setminus\mathcal{S}}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k\setminus\mathcal{S}}) > 0, \tag{36}$$

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_k$.

ii) Backwardly and sequentially from block $j = b-L$ to $j = b-B^{k-1}+1$, node $\pi(k), k = 2, 3, \ldots, M+2$, jointly decodes the message transmitted by its immediate upstream node $\pi(k-1)$, and the compressions of the C-F relay nodes. Specifically, node $\pi(k), k = 3, 4, \ldots, M+2$ finds the unique pair $(m_{j-B^{k-2}}, l_{\mathcal{D}_k, j-1})$ satisfying (35), and node $\pi(k), k = 2$, finds the unique pair $(m_j, l_{\mathcal{D}_k, j-1})$ satisfying (35). Here the exception for node $\pi(2)$ arises because the source node $\pi(1)$ transmits m_j rather than m_{j-1} in block j, but the ideas of the decoding processes at all $\pi(k), k = 2, 3, \ldots, M+2$, are exactly the same. Thus, below, we only present the decoding at node $\pi(k), k = 3, 4, \ldots, M+2$, while the decoding at node $\pi(2)$ can be easily obtained by analogy. The same consideration also applies to the proof in IV-B.

In (35), $l_{\mathcal{D}_k, j}$ has already been recovered due to the backward property of decoding, and among the messages $(m_j, m_{j-B}, \ldots, m_{j-B^M})$, only $m_{j-B^{k-2}}$ is the unknown message at node $\pi(k)$ that needs to be decoded in block j. In fact, $(m_{j-B^{k-1}}, \ldots, m_{j-B^M})$ have been decoded by block $b - B^{k-1}$, while $(m_j, \ldots, m_{j-B^{k-3}})$ either are dummy messages according to (34) (for block $j = b-L, b-L-1 \ldots, b-B^{k-2}-L+1$) or have been decoded due to the backward property of decoding (for block $j = b-B^{k-2}-L, b-B^{k-2}-L-1, \ldots, b-B^{k-1}+1$).

At each block $j = b-L, b-L-1, \ldots, b-B^{k-1}+1$, error occurs with $m_{j-B^{k-2}}$ if the true $m_{j-B^{k-2}}$ does not satisfy (35) with any $l_{\mathcal{D}_k, j-1}$, or a false $m_{j-B^{k-2}}$ satisfies (35) with some $l_{\mathcal{D}_k, j-1}$. According to the properties of typical sequences, the true $(m_{j-B^{k-2}}, l_{\mathcal{D}_k, j-1})$ satisfies (35) with high probability.

For a false $m_{j-B^{k-2}}$ and a $l_{\mathcal{D}_k, j-1}$ with false $\{l_{i,b-1}, i \in \mathcal{S}\}$ but true $\{l_{i,b-1}, i \in \mathcal{D}_k \setminus \mathcal{S}\}$,

$$\{\mathbf{X}_{\pi(1),j}(m_j|m_{j-B}, \ldots, m_{j-B^M}), \mathbf{X}_{\pi(s),j}(m_{j-B^{s-1}}|m_{j-B^s}, \ldots, m_{j-B^M}), s = 2, \ldots, k-1\}$$

are conditionally independent of $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_k\}$ and $\mathbf{Y}_{\pi(k),j}$ given

$$\{\mathbf{X}_{\pi(s),j}(m_{j-B^{s-1}}|m_{j-B^s}, \ldots, m_{j-B^M}), s = k, \ldots, M+1\};$$

and $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{S}\}$ are independent of

$$\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_k \setminus \mathcal{S}\}, \{\mathbf{X}_{\pi(s),j}(m_{j-B^{s-1}}|m_{j-B^s}, \ldots, m_{j-B^M}), s \in [k : M+1]\}, \mathbf{Y}_{\pi(k),j}.$$

Therefore, the probability that such false $(m_{j-B^{k-2}}, l_{\mathcal{D}_k, j-1})$ satisfies (35) can be upper bounded by

$$2^{T(H(X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, \hat{Y}_{\mathcal{D}_k}, Y_{\pi(k)})+\epsilon)} 2^{-T(H(X_{\pi(k:M+1)}, X_{\mathcal{D}_k\setminus\mathcal{S}}, \hat{Y}_{\mathcal{D}_k\setminus\mathcal{S}}, Y_{\pi(k)})-\epsilon)}$$

$$\times 2^{-T(H(X_{\pi(1:k-1)}|X_{\pi(k:M+1)})-\epsilon)} 2^{-T(H(X_\mathcal{S})-\epsilon)} 2^{-T(\sum_{i\in\mathcal{S}}(H(\hat{Y}_i|X_i)-\epsilon))}.$$

Since the number of such false $(m_{j-B^{k-2}}, l_{\mathcal{D}_k, j-1})$ is upper bounded by $2^{TR} \prod_{i\in\mathcal{S}} 2^{T(I(Y_i; \hat{Y}_i|X_i)+\epsilon)}$, with the union bound, it is easy to check that the probability of finding a false $m_{j-B^{k-2}}$ goes to zero as $T \to \infty$, if

$$R < \min_{\mathcal{S}\subseteq\mathcal{D}_k} I(X_{\pi(1:k-1)}, X_\mathcal{S}; \hat{Y}_{\mathcal{D}_k\setminus\mathcal{S}}, Y_{\pi(k)}|X_{\mathcal{D}_k\setminus\mathcal{S}}, X_{\pi(k:M+1)}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k\setminus\mathcal{S}}). \tag{37}$$

Then, based on the recovered $(m_j, m_{j-B}, \ldots, m_{j-B^M})$ and $l_{\mathcal{D}_k,j}$, from the proof of [20, Thm 2.7], it follows that $l_{\mathcal{D}_k,j-1}$ can be decoded if (36) holds.

Combining i) and ii), using the technique of time sharing, we obtain the achievable rate (19)-(20).

IV. UNIFIED RELAY FRAMEWORK WITH B-BLOCKS-BY-B-BLOCKS BACKWARD DECODING

Under the unified relay framework using nested blocks and backward decoding, we can also consider combining the noisy network coding scheme [18] with the multi-level D-F scheme. However, since noisy network coding uses repetitive encoding/all blocks united decoding, to make it fit into our framework, a modification is needed. Specifically, assume some fixed $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$ and $\pi(\{0, \mathcal{M}, n + 1\})$, and a total of B^{M+1} blocks are used. The source can still repetitively encode intra-B-blocks as in [18], but inter-B-blocks, the source has to cumulatively encode to allow for the operation of D-F strategy; Correspondingly, both the D-F relay nodes and the destination will perform B-Blocks-By-B-Blocks backward decoding, which is essentially a combination of backward decoding and B blocks united decoding. Same as in Section III, the backward decoding at node $\pi(k), k = 2, 3, \ldots, M + 2$, will happen at the end of every B^{k-1} blocks, i.e., at the end of block $b = vB^{k-1}, v \in [1 : B^{M+1}/B^{k-1}]$, and both the D-F relay nodes and the destination node perform compression-message joint decoding. Below, we still first consider the case of single D-F relay node ($M = 1$) to illustrate the main idea, and then extend it to the general case of multiple D-F relay nodes ($M \geq 2$).

A. Single D-F relay node ($M = 1$)

Still assume that only node 1 is the D-F relay node, and all other relay nodes are the C-F relay nodes, and let $\tilde{\mathcal{N}} := \mathcal{N} \setminus \{1\}$. Specializing Theorem 2.2 to this case, we have that a rate R is achievable, if there exists some

$$p(q)p(x_0|q)p(x_1|x_0, q) \prod_{i \in \tilde{\mathcal{N}}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

such that

$$R < \min \left\{ \begin{array}{l} \max\limits_{\mathcal{T}_1 \subseteq \tilde{\mathcal{N}}} \min\limits_{\mathcal{S} \subseteq \mathcal{T}_1} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_1 \setminus \mathcal{S}}, Y_1 | X_1, X_{\mathcal{T}_1 \setminus \mathcal{S}}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{T}_1}, Y_1, \hat{Y}_{\mathcal{T}_1 \setminus \mathcal{S}}, Q), \\ \max\limits_{\mathcal{T}_{n+2} \subseteq \tilde{\mathcal{N}}} \min\limits_{\mathcal{S} \subseteq \mathcal{T}_{n+2}} I(X_0, X_1, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_{n+2} \setminus \mathcal{S}}, Y_{n+2} | X_{\mathcal{T}_{n+2} \setminus \mathcal{S}}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{T}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{T}_{n+2} \setminus \mathcal{S}}, Q). \end{array} \right\}$$
(38)

Still, a total of B^2 blocks will be used. The detailed codebook generation and encoding/decoding process are as follows, which can be understood with the help of Table II.

Codebook Generation: Fix $p(x_0)p(x_1|x_0) \prod_{i \in \tilde{\mathcal{N}}} p(x_i)p(\hat{y}_i|y_i, x_i)$. We randomly and independently generate a codebook for each block.

i) First consider the codebook generation for the source node 0 and the D-F relay node 1. Denote $f(b) := \lceil \frac{b}{B} \rceil$, i.e., the smallest integer greater than or equal to $\frac{b}{B}$. For each block $b \in [1 : B^2]$, randomly generate 2^{TBR} independent sequences $\mathbf{x}_{1,b}(m_{f(b-B)})$ for node 1, and randomly generate 2^{TBR} conditionally independent sequences $\mathbf{x}_{0,b}(m_{f(b)}|m_{f(b-B)})$ for node 0, where $m_{f(b)}, m_{f(b-B)} \in [1 : 2^{TBR}]$.

ii) The codebook generation for the C-F relay nodes is exactly the same as that in Section III. For each block $b \in [1 : B^2]$ and each relay node $i \in \tilde{\mathcal{N}}$, randomly and independently generate $2^{T\hat{R}_i}$ sequences $\mathbf{x}_{i,b}(l_{i,b-1}), l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, where $\hat{R}_i = I(Y_i; \hat{Y}_i|X_i) + \epsilon$; for each relay node $i \in \tilde{\mathcal{N}}$ and each $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, randomly and conditionally independently generate $2^{T\hat{R}_i}$ sequences $\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})$, $l_{i,b} \in [1 : 2^{T\hat{R}_i}]$.

The combination of i) and ii) defines the codebook for any block $b \in [1 : B^2]$,

$$\mathcal{C}_b = \Big\{ \mathbf{x}_{1,b}(m_{f(b-B)}), \mathbf{x}_{0,b}(m_{f(b)}|m_{f(b-B)}) : m_{f(b)}, m_{f(b-B)} \in [1 : 2^{TBR}];$$

$$\mathbf{x}_{i,b}(l_{i,b-1}), \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}) : l_{i,b}, l_{i,b-1} \in [1 : 2^{T\hat{R}_i}], i \in \tilde{\mathcal{N}} \Big\}.$$
(39)

TABLE II
B-BLOCKS-BY-B-BLOCKS BACKWARD DECODING FOR THE SINGLE D-F RELAY NODE CASE

Block	1	2	$\cdots$	$B-1$	B	$\cdots$
X_0	$\mathbf{x}_{0,1}(m_1\|1)$	$\mathbf{x}_{0,2}(m_1\|1)$	$\cdots$	$\mathbf{x}_{0,B-1}(m_1\|1)$	$\mathbf{x}_{0,B}(m_1\|1)$	$\cdots$
Y_1	$\emptyset$	$\emptyset$	$\cdots$	$\emptyset$	m_1	$\cdots$
X_1	$\mathbf{x}_{1,1}(1)$	$\mathbf{x}_{1,2}(1)$	$\cdots$	$\mathbf{x}_{1,B-1}(1)$	$\mathbf{x}_{1,B}(1)$	$\cdots$
$Y_{\tilde{\mathcal{N}}}$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},1}(l_{\tilde{\mathcal{N}},1}\|\mathbf{1})$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},2}(l_{\tilde{\mathcal{N}},2}\|l_{\tilde{\mathcal{N}},1})$	$\cdots$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B-1}(l_{\tilde{\mathcal{N}},B-1}\|l_{\tilde{\mathcal{N}},B-2})$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B}(l_{\tilde{\mathcal{N}},B}\|l_{\tilde{\mathcal{N}},B-1})$	$\cdots$
$X_{\tilde{\mathcal{N}}}$	$\mathbf{x}_{\tilde{\mathcal{N}},1}(\mathbf{1})$	$\mathbf{x}_{\tilde{\mathcal{N}},2}(l_{\tilde{\mathcal{N}},1})$	$\cdots$	$\mathbf{x}_{\tilde{\mathcal{N}},B-1}(l_{\tilde{\mathcal{N}},B-2})$	$\mathbf{x}_{\tilde{\mathcal{N}},B}(l_{\tilde{\mathcal{N}},B-1})$	$\cdots$
Y_{n+2}	$\emptyset$	$\emptyset$	$\cdots$	$\emptyset$	$\emptyset$	$\cdots$

Block	B^2-B+1	B^2-B+2	$\cdots$	B^2-1	B^2
X_0	$\mathbf{x}_{0,B^2-B+1}(1\|m_{B-1})$	$\mathbf{x}_{0,B^2-B+2}(1\|m_{B-1})$	$\cdots$	$\mathbf{x}_{0,B^2-1}(1\|m_{B-1})$	$\mathbf{x}_{0,B^2}(1\|m_{B-1})$
Y_1	$\emptyset$	$\emptyset$	$\cdots$	$\emptyset$	m_B
X_1	$\mathbf{x}_{1,B^2-B+1}(m_{B-1})$	$\mathbf{x}_{1,B^2-B+2}(m_{B-1})$	$\cdots$	$\mathbf{x}_{1,B^2-1}(m_{B-1})$	$\mathbf{x}_{1,B^2}(m_{B-1})$
$Y_{\tilde{\mathcal{N}}}$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-B+1}(l_{\tilde{\mathcal{N}},B^2-B+1}\|l_{\tilde{\mathcal{N}},B^2-B})$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-B+2}(l_{\tilde{\mathcal{N}},B^2-B+2}\|l_{\tilde{\mathcal{N}},B^2-B+1})$	$\cdots$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-1}(l_{\tilde{\mathcal{N}},B^2-1}\|l_{\tilde{\mathcal{N}},B^2-2})$	$\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2}(l_{\tilde{\mathcal{N}},B^2}\|l_{\tilde{\mathcal{N}},B^2-1})$
$X_{\tilde{\mathcal{N}}}$	$\mathbf{x}_{\tilde{\mathcal{N}},B^2-B+1}(l_{\tilde{\mathcal{N}},B^2-B})$	$\mathbf{x}_{\tilde{\mathcal{N}},B^2-B+2}(l_{\tilde{\mathcal{N}},B^2-B+1})$	$\cdots$	$\mathbf{x}_{\tilde{\mathcal{N}},B^2-1}(l_{\tilde{\mathcal{N}},B^2-2})$	$\mathbf{x}_{\tilde{\mathcal{N}},B^2}(l_{\tilde{\mathcal{N}},B^2-1})$
Y_{n+2}	$\emptyset$	$\emptyset$	$\cdots$	$\emptyset$	$(m_1,m_2,\ldots,m_B)$

Encoding: Let the message vector to be sent be

$$\mathbf{m} = (\underbrace{m_1, m_1, \ldots, m_1}_{B}, \underbrace{m_2, m_2, \ldots, m_2}_{B}, \ldots, \underbrace{m_B, m_B, \ldots, m_B}_{B}).$$

Let $m_B = 1$ be the dummy message, i.e., $m_{f(b)} = 1$ for any

$$b \in [(B-1)B+1 : B^2], \tag{40}$$

and for any $b \leq 0$. The actually achievable rate is $\frac{B-1}{B}R$ due to the dummy messages, which, however, can be made arbitrarily close to R by letting $B \to \infty$.

i) First consider the encoding process for nodes 0 and 1.

- In block $b \in [1 : B^2]$, the source node 0 transmits $\mathbf{x}_{0,f(b)}(m_{f(b)}|m_{f(b-B)})$.
- At the end of block $vB, v \in [1 : B-1]$, the D-F relay node 1 has decoded message m_v using B blocks united decoding (see the decoding part). In the next B blocks, i.e., in block $b \in [vB+1 : (v+1)B]$, the relay node 1 transmits $\mathbf{x}_{1,b}(m_{f(b-B)})$, where $m_{f(b-B)}$ for any $b \in [vB+1 : (v+1)B]$ is corresponding to m_v that has been decoded by block vB.

ii) For any block $b \in [1 : B^2]$, each relay node $i \in \tilde{\mathcal{N}}$, upon receiving $\mathbf{y}_{i,b}$ at the end of block b, finds an index $l_{i,b}$ such that

$$(\mathbf{x}_{i,b}(l_{i,b-1}), \mathbf{y}_{i,b}, \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})) \in A_\epsilon(X_i, Y_i, \hat{Y}_i),$$

where $l_{i,0} = 1$ by convention. In block $b \in [1 : B^2]$, the relay node $i \in \tilde{\mathcal{N}}$ transmits $\mathbf{x}_{i,b}(l_{i,b-1})$.

Decoding: We present the decoding process at the D-F relay node 1 and at the destination node $n+2$ separately.

i) At the end of block $b = vB, v \in [1 : B]$, the D-F relay node 1 decodes messages m_v using B blocks united decoding, i.e., it finds the unique m_v, such that there exists some $l^{vB}_{\tilde{\mathcal{N}},(v-1)B+1}$ satisfying that for any block $j = (v-1)B+1, (v-1)B+2, \ldots, vB$,

$$(\mathbf{X}_{0,j}(m_{f(j)}|m_{f(j-B)}), \mathbf{X}_{1,j}(m_{f(j-B)})$$
$$\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \tilde{\mathcal{N}}\}, \mathbf{Y}_{1,j}) \in A_\epsilon(X_0, X_1, X_{\tilde{\mathcal{N}}}, \hat{Y}_{\tilde{\mathcal{N}}}, Y_1), \tag{41}$$

where $m_{f(j-B)}$ is corresponding to m_{v-1} and has been decoded by the end of block $(v-1)B$, and $m_{f(j)}$ is corresponding to m_v. From [18, Thm 1] and its proof (see also [20, Thm 2.4]), we have that m_v can be decoded if

$$R < \min_{\mathcal{S} \subseteq \tilde{\mathcal{N}}} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\tilde{\mathcal{N}} \backslash \mathcal{S}}, Y_1 | X_{\tilde{\mathcal{N}} \backslash \mathcal{S}}, X_1) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\tilde{\mathcal{N}}}, Y_1, \hat{Y}_{\tilde{\mathcal{N}} \backslash \mathcal{S}}). \tag{42}$$

Note, (42) can be improved by considering only a subset $\mathcal{T}_1 \subseteq \tilde{\mathcal{N}}$ for the decoding while treating the inputs of other C-F relay nodes as purely noise, leading to following more general rate constraint:

$$R < \max_{\mathcal{T}_1 \subseteq \tilde{\mathcal{N}}} \min_{\mathcal{S} \subseteq \mathcal{T}_1} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_1 \backslash \mathcal{S}}, Y_1 | X_{\mathcal{T}_1 \backslash \mathcal{S}}, X_1) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{T}_1}, Y_1, \hat{Y}_{\mathcal{T}_1 \backslash \mathcal{S}}). \tag{43}$$

ii) At the end of all B^2 block, the destination node decodes all messages $(m_1, m_2, \ldots, m_B)$ using B-Blocks-By-B-Blocks backward decoding. In fact, since $m_B = 1$ is dummy message, only $(m_1, m_2, \ldots, m_{B-1})$ need to be decoded. For this, backwardly and sequentially for $g = B-1, B-2, \ldots, 1$, node $n+2$ finds the unique m_g such that there exists some $l_{\tilde{\mathcal{N}}, gB+1}^{gB+B}$ satisfying that for any block $j = gB+1, gB+2, \ldots, gB+B$,

$$(\mathbf{X}_{0,j}(m_{f(j)} | m_{f(j-B)}), \mathbf{X}_{1,j}(m_{f(j-B)})$$
$$\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j} | l_{i,j-1})) : i \in \tilde{\mathcal{N}}\}, \mathbf{Y}_{n+2,j}) \in A_\epsilon(X_0, X_1, X_{\tilde{\mathcal{N}}}, \hat{Y}_{\tilde{\mathcal{N}}}, Y_{n+2}). \tag{44}$$

Note in (44), for $j = gB + 1, gB + 2, \ldots, gB + B$, only $m_{f(j-B)}$, corresponding to m_g, needs decoding; and $m_{f(j)}$, corresponding to m_{g+1}, either is a dummy message (for $g = B-1$, i.e., $j = (B-1)B+1, (B-1)B+2, \ldots, B^2$), or has been decoded due to the backward property of decoding (for $g = B-2, \ldots, 1$). Thus, X_0 and X_1 are cooperatively transmitting the message m_g, and similarly as above, m_g can be decoded if

$$R < \min_{\mathcal{S} \subseteq \tilde{\mathcal{N}}} I(X_0, X_1, X_{\mathcal{S}}; \hat{Y}_{\tilde{\mathcal{N}} \backslash \mathcal{S}}, Y_{n+2} | X_{\tilde{\mathcal{N}} \backslash \mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\tilde{\mathcal{N}}}, Y_{n+2}, \hat{Y}_{\tilde{\mathcal{N}} \backslash \mathcal{S}}). \tag{45}$$

Also, (45) can be improved by considering only a subset $\mathcal{T}_{n+2}$ for the decoding, leading to the following rate constraint:

$$R < \max_{\mathcal{T}_{n+2} \subseteq \tilde{\mathcal{N}}} \min_{\mathcal{S} \subseteq \mathcal{T}_{n+2}} I(X_0, X_1, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_{n+2} \backslash \mathcal{S}}, Y_{n+2} | X_{\mathcal{T}_{n+2} \backslash \mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{T}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{T}_{n+2} \backslash \mathcal{S}}).$$
$$\tag{46}$$

Combining (43) and (46) and using the technique of time sharing, we have that the rate in (38) is achievable.

B. Multiple D-F relay nodes ($M \geq 2$)

Codebook Generation: Fix $p(x_0) p(x_{\mathcal{M}} | x_0) \prod_{i \in \mathcal{N} \backslash \mathcal{M}} p(x_i) p(\hat{y}_i | y_i, x_i)$. We randomly and independently generate a codebook for each block. The codebook generation for the C-F relay nodes is exactly the same as that in III and IV-A, and hence omitted. We only present the codebook generation for nodes $\pi(1 : M + 1)$. Still, denote $f(b) := \lceil \frac{b}{B} \rceil$, i.e., the smallest integer greater than or equal to $\frac{b}{B}$.

- For each block $b \in [1 : B^{M+1}]$, backwardly and sequentially for each relay node $\pi(k), k = M + 1, M, \ldots, 2$, randomly generate 2^{TBR} conditionally independent sequences

$$\mathbf{x}_{\pi(k),b}(m_{f(b-B^{k-1})} | m_{f(b-B^k)}, \ldots, m_{f(b-B^M)}),$$

where $m_{f(b-B^{k-1})}, m_{f(b-B^k)}, \ldots, m_{f(b-B^M)} \in [1 : 2^{TBR}]$.

- For each block $b \in [1 : B^{M+1}]$ and node $\pi(1)$, i.e., the source node 0, randomly generate 2^{TBR} conditionally independent sequences

$$\mathbf{x}_{0,b}(m_{f(b)} | m_{f(b-B)}, \ldots, m_{f(b-B^M)}),$$

where $m_{f(b)}, m_{f(b-B)}, \ldots, m_{f(b-B^M)} \in [1 : 2^{TBR}]$.

The above, together with the codebook generation for the C-F relay nodes, defines the codebook for any block $b \in [1 : B^{M+1}]$,

$$
\begin{aligned}
\mathcal{C}_b = \{ &\mathbf{x}_{\pi(k),b}(m_{f(b-B^{k-1})}|m_{f(b-B^k)}, \ldots, m_{f(b-B^M)}) : m_{f(b-B^{k-1})}, \ldots, m_{f(b-B^M)} \in [1 : 2^{TBR}], k = M+1, M, \ldots, 2; \\
&\mathbf{x}_{0,b}(m_{f(b)}|m_{f(b-B)}, \ldots, m_{f(b-B^M)}) : m_{f(b)}, m_{f(b-B)}, \ldots, m_{f(b-B^M)} \in [1 : 2^{TBR}]; \\
&\mathbf{x}_{i,b}(l_{i,b-1}), \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}) : l_{i,b}, l_{i,b-1} \in [1 : 2^{T\hat{R}_i}], i \in \mathcal{N} \setminus \mathcal{M} \}.
\end{aligned}
$$

Encoding: Let the message vector to be sent be

$$
\mathbf{m} = (\underbrace{m_1, m_1, \ldots, m_1}_{B}, \underbrace{m_2, m_2, \ldots, m_2}_{B}, \ldots, \underbrace{m_{B^M}, m_{B^M}, \ldots, m_{B^M}}_{B}).
$$

Let $m_{f(b)} = 1$ be the dummy message for any

$$
b \in \cup_{u=1}^{M} \cup_{v=1}^{B^{M-u}} [v(B-1)B^u + 1 : vB^{u+1}], \tag{47}
$$

and for any $b \le 0$. The actually achievable rate is $(\frac{B-1}{B})^M R$ due to the dummy messages, which can still be made arbitrarily close to R by letting $B \to \infty$ for any M.

The encoding process for the C-F relay nodes is still exactly the same as that in III and IV-A, and hence omitted. We only present the encoding process for nodes $\pi(1 : M + 1)$.

- In block $b \in [1 : B^{M+1}]$, the source node 0 transmits $\mathbf{x}_{0,b}(m_{f(b)}|m_{f(b-B)}, \ldots, m_{f(b-B^M)})$.
- At the end of block $vB^{k-1}, v \in [1 : \frac{B^{M+1}}{B^{k-1}} - 1]$, the relay node $\pi(k), k = 2, \ldots, M + 1$, has decoded messages $(m_1, m_2, \ldots, m_{vB^{k-2}})$ using backward decoding (see the decoding part). In the next B^{k-1} blocks, i.e., in block $b \in [vB^{k-1} + 1 : (v+1)B^{k-1}]$, the relay node $\pi(k), k = 2, \ldots, M + 1$, transmits

$$
\mathbf{x}_{\pi(k),b}(m_{f(b-B^{k-1})}|m_{f(b-B^k)}, \ldots, m_{f(b-B^M)}),
$$

  where $(m_{f(b-B^{k-1})}, m_{f(b-B^k)}, \ldots, m_{f(b-B^M)})$ for any $b \in [vB^{k-1} + 1 : (v+1)B^{k-1}]$ have all been decoded by block vB^{k-1}.

Decoding: At the end of every B^{k-1} blocks, the node $\pi(k), k = 2, \ldots, M + 2$ decodes B^{k-2} messages using B-Blocks-By-B-Blocks backward decoding. (Note every B^{k-1} blocks carry B^{k-2} messages.) Specifically, at the end of block $b = vB^{k-1}, v \in [1 : B^{M+1}/B^{k-1}]$, the node $\pi(k), k = 2, \ldots, M + 2$, decodes messages $(m_{(v-1)B^{k-2}+1}, \ldots, m_{vB^{k-2}})$. In fact, $(m_{vB^{k-2}-B^{k-3}+1}, \ldots, m_{vB^{k-2}})$ are dummy messages according to (47), and only $(m_{(v-1)B^{k-2}+1}, \ldots, m_{vB^{k-2}-B^{k-3}})$ need decoding. For this, backwardly and sequentially for $g = vB^{k-2} - B^{k-3}, vB^{k-2} - B^{k-3} - 1, \ldots, (v-1)B^{k-2} + 1$, node $\pi(k)$ finds the unique m_g such that there exists some $l_{\mathcal{N} \setminus \mathcal{M}, (g-1)B+B^{k-2}+1}^{gB+B^{k-2}}$ satisfying that for any block $j = (g-1)B + B^{k-2} + 1, (g-1)B + B^{k-2} + 2, \ldots, gB + B^{k-2}$,

$$
\begin{aligned}
&(\mathbf{X}_{0,j}(m_{f(j)}|m_{f(j-B)}, \ldots, m_{f(j-B^M)}), \\
&\{ \mathbf{X}_{\pi(s),j}(m_{f(j-B^{s-1})}|m_{f(j-B^s)}, \ldots, m_{f(j-B^M)}), s = 2, \ldots, k-1, k, k+1, \ldots, M+1 \}, \\
&\{ (\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{N} \setminus \mathcal{M} \}, \mathbf{Y}_{\pi(k),j}) \in A_\epsilon(X_0, X_\mathcal{M}, X_{\mathcal{N} \setminus \mathcal{M}}, \hat{Y}_{\mathcal{N} \setminus \mathcal{M}}, Y_{\pi(k)}), \tag{48}
\end{aligned}
$$

where $(m_{f(j)}, m_{f(j-B)}, \ldots, m_{f(j-B^{k-3})}, m_{f(j-B^{k-2})}, m_{f(j-B^{k-1})}, \ldots, m_{f(j-B^M)})$ are corresponding to

$$
(m_{g+B^{k-3}}, m_{g+B^{k-3}-1}, \ldots, m_{g+B^{k-3}-B^{k-4}}, m_g, m_{g+B^{k-3}-B^{k-2}}, \ldots, m_{g+B^{k-3}-B^{M-1}}) \tag{49}
$$

Among the messages in (49), only m_g, corresponding to $m_{f(j-B^{k-2})}$, is the unknown message at node $\pi(k)$ that needs to be decoded. In fact,

$$
(m_{g+B^{k-3}-B^{k-2}}, \ldots, m_{g+B^{k-3}-B^{M-1}}), \text{corresponding to } (m_{f(j-B^{k-1})}, \ldots, m_{f(j-B^M)}),
$$

have been decoded by block $b - B^{k-1}$, while

$$
(m_{g+B^{k-3}}, m_{g+B^{k-3}-1}, \ldots, m_{g+B^{k-3}-B^{k-4}}), \text{corresponding to } (m_{f(j)}, m_{f(j-B)}, \ldots, m_{f(j-B^{k-3})}),
$$

either are dummy messages according to (47) (for $g = vB^{k-2} - B^{k-3}, \ldots, vB^{k-2} - 2B^{k-3} + 1$) or have been decoded due to the backward property of decoding (for $g = vB^{k-2} - 2B^{k-3}, vB^{k-2} - 2B^{k-3} - 1, \ldots, (v-1)B^{k-2} + 1$). Therefore, in (48),

$$\{\mathbf{X}_{\pi(s),j}, s = k, k+1, \ldots, M+1\}$$

are known at node $\pi(k)$, while

$$\{\mathbf{X}_{\pi(s),j}, s = 1, \ldots, k-1\}$$

are cooperatively transmitting the message m_g. Having noted this fact, from [18, Thm 1] and its proof (see also [20, Thm 2.4]), we have that m_g can be decoded if

$$R < \min_{\mathcal{S} \subseteq \mathcal{N} \setminus \mathcal{M}} I(X_{(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{(\mathcal{N} \setminus \mathcal{M}) \setminus \mathcal{S}}, Y_{\pi(k)} | X_{(\mathcal{N} \setminus \mathcal{M}) \setminus \mathcal{S}}, X_{(k:M+1)}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{(1:M+1)}, X_{\mathcal{N} \setminus \mathcal{M}}, Y_{\pi(k)}, \hat{Y}_{(\mathcal{N} \setminus \mathcal{M}) \setminus \mathcal{S}}). \tag{50}$$

By considering only a subset $\mathcal{T}_k \subseteq \mathcal{N} \setminus \mathcal{M}$ for the decoding at node $\pi(k)$ while treating the inputs of other C-F relay nodes as purely noise, and using the technique of time sharing, (50) can be improved to

$$R < \max_{\mathcal{T}_k \subseteq \mathcal{N} \setminus \mathcal{M}} \min_{\mathcal{S} \subseteq \mathcal{T}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_k \setminus \mathcal{S}}, Y_{\pi(k)} | X_{\mathcal{T}_k \setminus \mathcal{S}}, X_{\pi(k:M+1)}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{T}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{T}_k \setminus \mathcal{S}}, Q), \tag{51}$$

which proves Theorem 2.2.

V. CONCLUSION

We have proposed a unified relay framework with both the D-F and C-F relay nodes for multiple-relay channels. This framework employs nested blocks combined with backward decoding to allow for the full incorporation of the best known D-F and C-F relay strategies. The achievable rates obtained under such a framework turn out to combine both the best known D-F and C-F achievable rates and include them as special cases.

APPENDIX A
RATES FOR THE AWGN TWO-RELAY CHANNEL

With $C(x) := \frac{1}{2} \log_2(1 + x)$, various rates are evaluated for the AWGN two-relay channel as follows.

A. Multi-level D-F

The best achievable D-F rates are the multi-level D-F rates [13]-[16]

$$R_a = \max\{R_{a_1}, R_{a_2}\}$$

where

$$R_{a_1} = \min\{a_{11}, a_{12}, a_{13}\}$$
$$R_{a_2} = \min\{a_{21}, a_{22}, a_{23}\}$$

with

$$a_{11} = I(X_0; Y_1 | X_1, X_2)$$
$$= C(g_{01}^2 P)$$
$$a_{12} = I(X_0, X_1; Y_2 | X_2)$$
$$= C((g_{02}^2 + g_{12}^2)P)$$
$$a_{13} = I(X_0, X_1, X_2; Y_3)$$

$$= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P)$$
$$a_{21} = I(X_0; Y_2 | X_1, X_2)$$
$$= C(g_{02}^2 P)$$
$$a_{22} = I(X_0, X_2; Y_1 | X_1)$$
$$= C((g_{01}^2 + g_{21}^2)P)$$
$$a_{23} = I(X_0, X_1, X_2; Y_3)$$
$$= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P).$$

B. Noisy network coding

The best achievable C-F rates are the noisy network coding rates [18]-[20]

$$R_b = \min\{b_1, b_2, b_3, b_4\}$$

with

$$b_1 = I(X_0; \hat{Y}_1, \hat{Y}_2, Y_3 | X_1, X_2)$$
$$= C\left(\frac{g_{01}^2 P}{1 + \hat{\sigma}_1^2} + \frac{g_{02}^2 P}{1 + \hat{\sigma}_2^2} + g_{03}^2 P\right)$$
$$b_2 = I(X_0, X_1; \hat{Y}_2, Y_3 | X_2) - I(Y_1; \hat{Y}_1 | X_0, X_1, X_2, \hat{Y}_2, Y_3)$$
$$= C\left((g_{03}^2 + g_{13}^2)P + \frac{(g_{02}^2 + g_{12}^2)P}{1 + \hat{\sigma}_2^2} + \frac{(g_{02}g_{13} - g_{12}g_{03})^2 P^2}{1 + \hat{\sigma}_2^2}\right) - C\left(\frac{1}{\hat{\sigma}_1^2}\right)$$
$$b_3 = I(X_0, X_2; \hat{Y}_1, Y_3 | X_1) - I(Y_2; \hat{Y}_2 | X_0, X_1, X_2, \hat{Y}_1, Y_3)$$
$$= C\left((g_{03}^2 + g_{23}^2)P + \frac{(g_{01}^2 + g_{21}^2)P}{1 + \hat{\sigma}_1^2} + \frac{(g_{01}g_{23} - g_{21}g_{03})^2 P^2}{1 + \hat{\sigma}_1^2}\right) - C\left(\frac{1}{\hat{\sigma}_2^2}\right)$$
$$b_4 = I(X_0, X_1, X_2; Y_3) - I(Y_1, Y_2; \hat{Y}_1, \hat{Y}_2 | X_0, X_1, X_2, Y_3)$$
$$= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P) - C\left(\frac{1 + \hat{\sigma}_1^2 + \hat{\sigma}_2^2}{\hat{\sigma}_1^2 \hat{\sigma}_2^2}\right)$$

where the optimal $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined numerically.

C. Our unified schemes

The rates under our unified framework are

$$R_c = \max\{R_a, R_b, R_{c_1}, R_{c_2}\}$$

where

$$R_{c_1} = \min\{\max\{c_{11}, \min\{c_{12}, c_{13}\}\}, \max\{c_{14}, \min\{c_{15}, c_{16}\}\}\}$$
$$R_{c_2} = \min\{\max\{c_{21}, \min\{c_{22}, c_{23}\}\}, \max\{c_{24}, \min\{c_{25}, c_{26}\}\}\}$$

with

$$c_{11} = I(X_0; Y_1 | X_1)$$
$$= C\left(\frac{g_{01}^2 P}{1 + g_{21}^2 P}\right)$$
$$c_{12} = I(X_0; \hat{Y}_2, Y_1 | X_1)$$
$$= C\left(g_{01}^2 P + \frac{g_{02}^2 P}{1 + \hat{\sigma}_2^2}\right)$$

$$c_{13} = I(X_0, X_2; Y_1 | X_1) - I(Y_2; \hat{Y}_2 | X_0, X_1, X_2, Y_1)$$
$$= C((g_{01}^2 + g_{21}^2)P) - C\left(\frac{1}{\hat{\sigma}_2^2}\right)$$

$$c_{14} = I(X_0, X_1; Y_3)$$
$$= C\left(\frac{(g_{03}^2 + g_{13}^2)P}{1 + g_{23}^2 P}\right)$$

$$c_{15} = I(X_0, X_1; \hat{Y}_2, Y_3 | X_2)$$
$$= C\left((g_{03}^2 + g_{13}^2)P + \frac{(g_{02}^2 + g_{12}^2)P}{1 + \hat{\sigma}_2^2} + \frac{(g_{02}g_{13} - g_{12}g_{03})^2 P^2}{1 + \hat{\sigma}_2^2}\right)$$

$$c_{16} = I(X_0, X_1, X_2; Y_3) - I(Y_2; \hat{Y}_2 | X_0, X_1, X_2, Y_3)$$
$$= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P) - C\left(\frac{1}{\hat{\sigma}_2^2}\right)$$

$$c_{21} = I(X_0; Y_2 | X_2)$$
$$= C\left(\frac{g_{02}^2 P}{1 + g_{12}^2 P}\right)$$

$$c_{22} = I(X_0; \hat{Y}_1, Y_2 | X_2)$$
$$= C\left(g_{02}^2 P + \frac{g_{01}^2 P}{1 + \hat{\sigma}_1^2}\right)$$

$$c_{23} = I(X_0, X_1; Y_2 | X_2) - I(Y_1; \hat{Y}_1 | X_0, X_1, X_2, Y_2)$$
$$= C((g_{02}^2 + g_{12}^2)P) - C\left(\frac{1}{\hat{\sigma}_1^2}\right)$$

$$c_{24} = I(X_0, X_2; Y_3)$$
$$= C\left(\frac{(g_{03}^2 + g_{23}^2)P}{1 + g_{13}^2 P}\right)$$

$$c_{25} = I(X_0, X_2; \hat{Y}_1, Y_3 | X_1)$$
$$= C\left((g_{03}^2 + g_{23}^2)P + \frac{(g_{01}^2 + g_{21}^2)P}{1 + \hat{\sigma}_1^2} + \frac{(g_{01}g_{23} - g_{21}g_{03})^2 P^2}{1 + \hat{\sigma}_1^2}\right)$$

$$c_{26} = I(X_0, X_1, X_2; Y_3) - I(Y_1; \hat{Y}_1 | X_0, X_1, X_2, Y_3)$$
$$= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P) - C\left(\frac{1}{\hat{\sigma}_1^2}\right).$$

D. Unified scheme in [16]

The rates in [16, Thm 4] (with $U_i = \emptyset, i = 1, 2$) are

$$R_d = \max\{R_a, R_{d_1}, R_{d_2}, R_{d_3}\}$$

with

$$R_{d_1} = d_{11} \quad \text{s.t. } d_{12} \leq d'_{12}, d_{13} \leq d'_{13}, d_{14} \leq d'_{14}$$
$$R_{d_2} = \min\{d_{21}, d_{22}\} \quad \text{s.t. } d_{23} \leq d'_{23}$$
$$R_{d_3} = \min\{d_{31}, d_{32}\} \quad \text{s.t. } d_{33} \leq d'_{33}$$

where

$$d_{11} = I(X_0; \hat{Y}_1, \hat{Y}_2, Y_3 | X_1, X_2)$$

$$= b_1$$

$$d_{12} = I(Y_1; \hat{Y}_1|X_1, X_2, \hat{Y}_2, Y_3) + I(\hat{Y}_1; X_2|X_1)$$

$$= C\left(\frac{1}{\hat{\sigma}_1^2} + \frac{g_{01}^2 P}{\hat{\sigma}_1^2(1 + \frac{g_{02}^2 P}{1+\hat{\sigma}_2^2} + g_{03}^2 P)}\right) + C\left(\frac{g_{21}^2 P}{g_{01}^2 P + 1 + \hat{\sigma}_1^2}\right)$$

$$d'_{12} = I(X_1; Y_3|X_2)$$

$$= C\left(\frac{g_{13}^2 P}{1 + g_{03}^2 P}\right)$$

$$d_{13} = I(Y_2; \hat{Y}_2|X_1, X_2, \hat{Y}_1, Y_3) + I(\hat{Y}_2; X_1|X_2)$$

$$= C\left(\frac{1}{\hat{\sigma}_2^2} + \frac{g_{02}^2 P}{\hat{\sigma}_2^2(1 + \frac{g_{01}^2 P}{1+\hat{\sigma}_1^2} + g_{03}^2 P)}\right) + C\left(\frac{g_{12}^2 P}{g_{02}^2 P + 1 + \hat{\sigma}_2^2}\right)$$

$$d'_{13} = I(X_2; Y_3|X_1)$$

$$= C\left(\frac{g_{23}^2 P}{1 + g_{03}^2 P}\right)$$

$$d_{14} = I(Y_1, Y_2; \hat{Y}_1, \hat{Y}_2|X_1, X_2, Y_3) + I(\hat{Y}_1; X_2|X_1) + I(\hat{Y}_2; X_1|X_2)$$

$$= C\left(\frac{1 + \hat{\sigma}_1^2 + \hat{\sigma}_2^2}{\hat{\sigma}_1^2 \hat{\sigma}_2^2} + \frac{g_{01}^2(1 + \hat{\sigma}_2^2)P + g_{02}^2(1 + \hat{\sigma}_1^2)P}{\hat{\sigma}_1^2 \hat{\sigma}_2^2(1 + g_{03}^2 P)}\right)$$

$$+ C\left(\frac{g_{21}^2 P}{g_{01}^2 P + 1 + \hat{\sigma}_1^2}\right) + C\left(\frac{g_{12}^2 P}{g_{02}^2 P + 1 + \hat{\sigma}_2^2}\right)$$

$$d'_{14} = I(X_1, X_2; Y_3)$$

$$= C\left(\frac{(g_{13}^2 + g_{23}^2)P}{1 + g_{03}^2 P}\right)$$

$$d_{21} = I(X_0; Y_1|X_1)$$

$$= c_{11}$$

$$d_{22} = I(X_0, X_1; \hat{Y}_2, Y_3|X_2)$$

$$= c_{15}$$

$$d_{23} = I(Y_2; \hat{Y}_2|X_2, Y_3)$$

$$= C\left(\frac{1}{\hat{\sigma}_2^2} + \frac{(g_{02}g_{13} - g_{12}g_{03})^2 P^2}{\hat{\sigma}_2^2((g_{03}^2 + g_{13}^2)P + 1)} + \frac{(g_{02}^2 + g_{12}^2)P}{\hat{\sigma}_2^2((g_{03}^2 + g_{13}^2)P + 1)}\right)$$

$$d'_{23} = I(X_2; Y_3)$$

$$= C\left(\frac{g_{23}^2 P}{1 + (g_{03}^2 + g_{13}^2)P}\right)$$

$$d_{31} = I(X_0; Y_2|X_2)$$

$$= c_{21}$$

$$d_{32} = I(X_0, X_2; \hat{Y}_1, Y_3|X_1)$$

$$= c_{25}$$

$$d_{33} = I(Y_1; \hat{Y}_1|X_1, Y_3)$$

$$= C\left(\frac{1}{\hat{\sigma}_1^2} + \frac{(g_{01}g_{03} - g_{21}g_{23})^2 P^2}{\hat{\sigma}_1^2((g_{03}^2 + g_{23}^2)P + 1)} + \frac{(g_{01}^2 + g_{21}^2)P}{\hat{\sigma}_1^2((g_{03}^2 + g_{23}^2)P + 1)}\right)$$

$$d'_{33} = I(X_1; Y_3)$$

$$= C\left(\frac{g_{13}^2 P}{1 + (g_{03}^2 + g_{23}^2)P}\right)$$

E. Unified scheme in [24]

The rates in [24] are

$$R_e = \max\{R_{e_1}, R_b, R_{e_2}, R_{e_3}\}$$

where

$$R_{e_1} = \min\{e_{11}, e_{12}, e_{13}\}$$
$$R_{e_2} = \min\{e_{21}, \max\{e_{22}, \min\{e_{23}, e_{24}\}\}\}$$
$$R_{e_3} = \min\{e_{31}, \max\{e_{32}, \min\{e_{33}, e_{34}\}\}\}$$

with

$$e_{11} = a_{11}$$
$$e_{12} = a_{21}$$
$$e_{13} = a_{13}$$
$$e_{21} = \begin{cases} \min\{c_{12}, c_{13}\} & \text{if} \quad C\left(\frac{1}{\hat{\sigma}_2^2}\right) \leq C\left(\frac{g_{21}^2 P}{1+g_{01}^2 P}\right) \\ c_{11} & \text{otherwise} \end{cases}$$
$$e_{22} = c_{14}$$
$$e_{23} = c_{15}$$
$$e_{24} = c_{16}$$
$$e_{31} = \begin{cases} \min\{c_{22}, c_{23}\} & \text{if} \quad C\left(\frac{1}{\hat{\sigma}_1^2}\right) \leq C\left(\frac{g_{12}^2 P}{1+g_{02}^2 P}\right) \\ c_{21} & \text{otherwise} \end{cases}$$
$$e_{32} = c_{24}$$
$$e_{33} = c_{25}$$
$$e_{34} = c_{26}.$$

F. Cut-set bound

Finally, the cut-set bound is given by

$$R_f = \min\{f_1, f_2, f_3, f_4\}$$

where

$$\begin{aligned} f_1 &= I(X_0; Y_1, Y_2, Y_3 | X_1, X_2) \\ &= C((g_{01}^2 + g_{02}^2 + g_{03}^2)P) \\ f_2 &= I(X_0, X_1; Y_2, Y_3 | X_2) \\ &= C((g_{02}g_{13} - g_{12}g_{03})^2 P^2 + (g_{02}^2 + g_{12}^2 + g_{03}^2 + g_{13}^2)P) \\ f_3 &= I(X_0, X_2; Y_1, Y_3 | X_1) \\ &= C((g_{03}g_{21} - g_{23}g_{01})^2 P^2 + (g_{01}^2 + g_{21}^2 + g_{03}^2 + g_{23}^2)P) \\ f_4 &= I(X_0, X_1, X_2; Y_3) \\ &= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P). \end{aligned}$$

APPENDIX B
DERIVATION OF (28)

For notational convenience, denote $\{l_{i,j}, i \in \mathcal{D}_1\}$ by a $|\mathcal{D}_1|$-dimensional vector $\mathbf{l}_j$ for any j. Recall that node 1 finds the unique $\mathbf{l}_{b-L}$ such that there exists some $\mathbf{l}_{b-L+1}^b$ satisfying (27) for any block $j = b - L + 1, b - L + 2, \ldots, b$, where both $\mathbf{X}_{0,j}(m_j|m_{j-B})$ and $\mathbf{X}_{1,j}(m_{j-B})$ are known at node 1.

Assume the true $\mathbf{l}_{b-L}^b = \mathbf{1}^{L+1}$, where $\mathbf{1} := (1, \ldots, 1)$ is a $|\mathcal{D}_1|$-dimensional all-ones vector. Then, error occurs if $\mathbf{l}_{b-L} = \mathbf{1}$ does not satisfy (27) with any $\mathbf{l}_{b-L+1}^b$ for any $j = b - L + 1, b - L + 2, \ldots, b$, or a false $\mathbf{l}_{b-L} \neq \mathbf{1}$ satisfies (27) with some $\mathbf{l}_{b-L+1}^b$ for any $j = b - L + 1, b - L + 2, \ldots, b$. Since $\mathbf{l}_{b-L}^b = \mathbf{1}^{L+1}$ satisfies (27) for any $j = b - L + 1, b - L + 2, \ldots, b$ with high probability according to the properties of typical sequences, we only need to bound $\Pr(\bigcup_{\mathbf{l}_{b-L} \neq \mathbf{1}} \mathcal{E}_{\mathbf{l}_{b-L}})$, where $\mathcal{E}_{\mathbf{l}_{b-L}}$ is defined as the event that $\mathbf{l}_{b-L}$ satisfies (27) with some $\mathbf{l}_{b-L+1}^b$ for any $j = b - L + 1, b - L + 2, \ldots, b$. For any $(\mathbf{l}_{j-1}, \mathbf{l}_j)$, define $\mathcal{A}_j(\mathbf{l}_{j-1}, \mathbf{l}_j)$ as the event that $(\mathbf{l}_{j-1}, \mathbf{l}_j)$ satisfies (27). Then, we have

$$
\Pr(\bigcup_{\mathbf{l}_{b-L} \neq \mathbf{1}} \mathcal{E}_{\mathbf{l}_{b-L}}) = \Pr(\bigcup_{\mathbf{l}_{b-L+1}^b} \bigcup_{\mathbf{l}_{b-L} \neq \mathbf{1}} \bigcap_{b=b-L+1}^b \mathcal{A}_j(\mathbf{l}_{j-1}, \mathbf{l}_j))
$$

$$
= \Pr(\bigcup_{a=1}^{L-1} \bigcup_{\mathbf{l}_{b-L+1}^b : \mathbf{l}_{b-L+a} = \mathbf{1}} \bigcup_{\substack{\mathbf{l}_{b-L+1}^b \, : \, \mathbf{l}_{b-L+a} \neq \mathbf{1}, \\ \forall a \in [1:L-1]}} \bigcup_{\mathbf{l}_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^b \mathcal{A}_j(\mathbf{l}_{j-1}, \mathbf{l}_j))
$$

$$
\leq \sum_{a=1}^{L-1} \Pr(\bigcup_{\mathbf{l}_{b-L+1}^b : \mathbf{l}_{b-L+a} = \mathbf{1}} \bigcup_{\mathbf{l}_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^b \mathcal{A}_j(\mathbf{l}_{j-1}, \mathbf{l}_j)) + \Pr(\bigcup_{\substack{\mathbf{l}_{b-L+1}^b \, : \, \mathbf{l}_{b-L+a} \neq \mathbf{1}, \\ \forall a \in [1:L-1]}} \bigcup_{\mathbf{l}_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^b \mathcal{A}_j(\mathbf{l}_{j-1}, \mathbf{l}_j)).
$$

$$(52)$$

Let us first consider the second term in (52). For any $\mathbf{l}_{b-L}^b$, let $\mathcal{S}_j(\mathbf{l}_{b-L}^b) = \{i \in \mathcal{D}_1 : l_{i,j-1} \neq 1\}$. Note $\mathcal{S}_j(\mathbf{l}_{b-L}^b)$ only depends on $\mathbf{l}_{j-1}$, so we also write it as $\mathcal{S}_j(\mathbf{l}_{j-1})$. Define $\mathbf{X}_j(\mathcal{S}_j(\mathbf{l}_{j-1}))$ as $\{\mathbf{X}_{i,j}(l_{i,j-1}), i \in \mathcal{S}_j(\mathbf{l}_{j-1})\}$, and similarly define $\mathbf{Y}_j(\mathcal{S}_j(\mathbf{l}_{j-1}))$ and $\hat{\mathbf{Y}}_j(\mathcal{S}_j(\mathbf{l}_{j-1}))$. Then, $(\mathbf{X}_j(\mathcal{S}_j(\mathbf{l}_{j-1})), \hat{\mathbf{Y}}_j(\mathcal{S}_j(\mathbf{l}_{j-1})))$ are independent of $(\mathbf{X}_{0,j}, \mathbf{X}_{1,j}, \mathbf{X}_j(\mathcal{D}_1 \setminus \mathcal{S}_j(\mathbf{l}_{b-L}^b)), \hat{\mathbf{Y}}_j(\mathcal{D}_1 \setminus \mathcal{S}_j(\mathbf{l}_{b-L}^b)), \mathbf{Y}_{1,j})$, and $\Pr(\mathcal{A}_j(\mathbf{l}_{j-1}, \mathbf{l}_j))$ can be upper bounded by

$$
2^{T(H(X_0, X_1, X_{\mathcal{D}_1}, \hat{Y}_{\mathcal{D}_1}, Y_1) + \epsilon)} 2^{-T(H(X_0, X_1, X_{\mathcal{D}_1 \setminus \mathcal{S}_j(\mathbf{l}_{b-L}^b)}, \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}_j(\mathbf{l}_{b-L}^b)}, Y_1) - \epsilon)}
$$

$$
\times 2^{-T(H(X_{\mathcal{S}_j(\mathbf{l}_{j-1})}) - \epsilon)} 2^{-T(\sum_{i \in \mathcal{S}_j(\mathbf{l}_{j-1})}(H(\hat{Y}_i|X_i) - \epsilon))}
$$

$$
=: 2^{-T(\mathcal{I}(\mathcal{S}_j(\mathbf{l}_{j-1})) - \epsilon')}
$$

where

$$
\mathcal{I}(\mathcal{S}_j(\mathbf{l}_{j-1})) = I(X_{\mathcal{S}_j(\mathbf{l}_{j-1})}; \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}_j(\mathbf{l}_{b-L}^b)}, Y_1 | X_0, X_1, X_{\mathcal{D}_1 \setminus \mathcal{S}_j(\mathbf{l}_{b-L}^b)})
$$

$$
- H(\hat{Y}_{\mathcal{S}_j(\mathbf{l}_{j-1})} | X_0, X_1, X_{\mathcal{D}_1}, \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}_j(\mathbf{l}_{b-L}^b)}, Y_1) + \sum_{i \in \mathcal{S}_j(\mathbf{l}_{j-1})} H(\hat{Y}_i | X_i)
$$

and $\epsilon' \to 0$ as $\epsilon \to 0$. Then, we have

$$
\Pr(\bigcup_{\substack{\mathbf{l}_{b-L+1}^b \, : \, \mathbf{l}_{b-L+a} \neq \mathbf{1}, \\ \forall a \in [1:L-1]}} \bigcup_{\mathbf{l}_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^b \mathcal{A}_j(\mathbf{l}_{j-1}, \mathbf{l}_j))
$$

$$
\leq \sum_{\substack{\mathbf{l}_{b-L+1}^b \, : \, \mathbf{l}_{b-L+a} \neq \mathbf{1}, \\ \forall a \in [1:L-1]}} \sum_{\mathbf{l}_{b-L} \neq \mathbf{1}} \prod_{j=b-L+1}^b \Pr(\mathcal{A}_j(\mathbf{l}_{j-1}, \mathbf{l}_j))
$$

$$\leq \sum_{l_b} \sum_{\substack{l_{b-L+1}^{b-1} : l_{b-L+a} \neq \mathbf{1}, \\ \forall a \in [1:L-1]}} \sum_{l_{b-L} \neq \mathbf{1}} \prod_{j=b-L+1}^{b} 2^{-T(\mathcal{I}(\mathcal{S}_j(l_{j-1}))-\epsilon')}$$

$$= \sum_{l_b} \sum_{\substack{\mathcal{S}_{b-L+1},\ldots,\mathcal{S}_b : \\ \mathcal{S}_{b-L+a} \neq \emptyset, \forall a \in [1:L]}} \sum_{\substack{l_{b-L}^{b-1} : \mathcal{S}_j(l_{b-L}^{b-1}) = \mathcal{S}_j, \ j=b-L+1 \\ \forall j \in [b-L+1:b]}} \prod_{j=b-L+1}^{b} 2^{-T(\mathcal{I}(\mathcal{S}_j(l_{j-1}))-\epsilon')}$$

$$\leq \sum_{l_b} \sum_{\substack{\mathcal{S}_{b-L+1},\ldots,\mathcal{S}_b : \\ \mathcal{S}_{b-L+a} \neq \emptyset, \forall a \in [1:L]}} \prod_{j=b-L+1}^{b} 2^{T(\sum_{i \in \mathcal{S}_j} (I(Y_i;\hat{Y}_i|X_i)+\epsilon))} \prod_{j=b-L+1}^{b} 2^{-T(\mathcal{I}(\mathcal{S}_j)-\epsilon')}$$

$$\leq \sum_{l_b} \sum_{\substack{\mathcal{S}_{b-L+1},\ldots,\mathcal{S}_b : \\ \mathcal{S}_{b-L+a} \neq \emptyset, \forall a \in [1:L]}} \prod_{j=b-L+1}^{b} 2^{-T(I(X_{\mathcal{S}_j};\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}_j},Y_1|X_0,X_1,X_{\mathcal{D}_1 \setminus \mathcal{S}_j})-I(Y_{\mathcal{S}_j};\hat{Y}_{\mathcal{S}_j}|X_0,X_1,X_{\mathcal{D}_1},Y_1,\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}_j})-\epsilon'')}$$

$$= \sum_{l_b} \sum_{\substack{\mathcal{S}_{b-L+1},\ldots,\mathcal{S}_b : \\ \mathcal{S}_{b-L+a} \neq \emptyset, \forall a \in [1:L]}} 2^{-T \sum_{j=b-L+1}^{b} (I(X_{\mathcal{S}_j};\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}_j},Y_1|X_0,X_1,X_{\mathcal{D}_1 \setminus \mathcal{S}_j})-I(Y_{\mathcal{S}_j};\hat{Y}_{\mathcal{S}_j}|X_0,X_1,X_{\mathcal{D}_1},Y_1,\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}_j})-\epsilon'')}$$

$$\leq \sum_{l_b} (2^n)^L 2^{-TL(\min_{\mathcal{S} \subseteq \mathcal{D}_1 : \mathcal{S} \neq \emptyset} \{I(X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}},Y_1|X_0,X_1,X_{\mathcal{D}_1 \setminus \mathcal{S}})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_0,X_1,X_{\mathcal{D}_1},Y_1,\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}})-\epsilon''\})}$$

$$\leq 2^{T(\sum_{i \in \mathcal{N}} (I(\hat{Y}_i;Y_i|X_i)+\epsilon))} 2^{nL} 2^{-TL(\min_{\mathcal{S} \subseteq \mathcal{D}_1 : \mathcal{S} \neq \emptyset} \{I(X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}},Y_1|X_0,X_1,X_{\mathcal{D}_1 \setminus \mathcal{S}})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_0,X_1,X_{\mathcal{D}_1},Y_1,\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}})-\epsilon''\})}$$

where $\epsilon'' \to 0$ as $\epsilon \to 0$. Thus, as both T and L go to infinity, the second term in (52) goes to 0, if (28) holds.

Now consider the first term in (52). For any $a \in [1:L-1]$, we have

$$\Pr\left(\bigcup_{l_{b-L+1}^b : l_{b-L+a}=\mathbf{1}} \bigcup_{l_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^{b} \mathcal{A}_j(l_{j-1},l_j) \right) \leq \Pr\left(\bigcup_{l_{b-L+1}^{b-L+a} : l_{b-L+a}=\mathbf{1}} \bigcup_{l_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^{b-L+a} \mathcal{A}_j(l_{j-1},l_j) \right).$$

Note $\Pr\left(\bigcup_{l_{b-L+1}^{b-L+a} : l_{b-L+a}=\mathbf{1}} \bigcup_{l_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^{b-L+a} \mathcal{A}_j(l_{j-1},l_j) \right)$ is the probability that there exists a false $l_{b-L} \neq \mathbf{1}$ satisfying (27) with some l_{b-L+1}^{b-L+a} for any block $j \in [b-L+1:b-L+a]$, where $l_{b-L+a} = \mathbf{1}$ is true. Below, we show that this probability goes to 0 with the idea of backward decoding.

For any $c \in [1:a]$, $a \in [1:L-1]$, denote

$$p_{b-L+c} := \Pr\left(\bigcup_{l_{b-L+1}^{b-L+c} : l_{b-L+c}=\mathbf{1}} \bigcup_{l_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^{b-L+c} \mathcal{A}_j(l_{j-1},l_j) \right).$$

Then, we have

$$p_{b-L+c} = \Pr\left(\bigcup_{l_{b-L+1}^{b-L+c} : l_{b-L+c}=\mathbf{1}} \bigcup_{l_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^{b-L+c} \mathcal{A}_j(l_{j-1},l_j) \right)$$

$$\leq \Pr\left(\bigcup_{\substack{l_{b-L+1}^{b-L+c} : l_{b-L+c}=\mathbf{1}, \\ l_{b-L+c-1}=\mathbf{1}}} \bigcup_{l_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^{b-L+c} \mathcal{A}_j(l_{j-1},l_j) \right) + \Pr\left(\bigcup_{\substack{l_{b-L+1}^{b-L+c} : l_{b-L+c}=\mathbf{1}, \\ l_{b-L+c-1} \neq \mathbf{1}}} \bigcup_{l_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^{b-L+c} \mathcal{A}_j(l_{j-1},l_j) \right)$$

$$\leq \Pr\left(\bigcup_{l_{b-L+1}^{b-L+c-1} : l_{b-L+c-1}=\mathbf{1}} \bigcup_{l_{b-L} \neq \mathbf{1}} \bigcap_{j=b-L+1}^{b-L+c-1} \mathcal{A}_j(l_{j-1},l_j) \right) + \Pr\left(\bigcup_{l_{b-L+c}=\mathbf{1},l_{b-L+c-1} \neq \mathbf{1}} \mathcal{A}_{b-L+c}(l_{b-L+c-1},l_{b-L+c}) \right)$$

$$=: p_{b-L+c-1} + p'_{b-L+c},$$

where

$$p'_{b-L+c} := \Pr\left(\bigcup_{\mathbf{l}_{b-L+c}=\mathbf{1}, \mathbf{l}_{b-L+c-1} \neq \mathbf{1}} \mathcal{A}_{b-L+c}(\mathbf{l}_{b-L+c-1}, \mathbf{l}_{b-L+c}) \right)$$

and especially

$$p_{b-L+1} = p'_{b-L+1} = \Pr\left(\bigcup_{\mathbf{l}_{b-L+1}=\mathbf{1}, \mathbf{l}_{b-L} \neq \mathbf{1}} \mathcal{A}_{b-L+1}(\mathbf{l}_{b-L}, \mathbf{l}_{b-L+1}) \right).$$

Recursively, for any $a \in [1 : L-1]$,

$$p_{b-L+a} \leq p_{b-L+a-1} + p'_{b-L+a} \leq p_{b-L+a-2} + p'_{b-L+a-1} + p'_{b-L+a} \leq \cdots \leq \sum_{c=1}^{c=a} p'_{b-L+c}.$$

For any $c \in [1 : a]$, $a \in [1 : L-1]$, with $\mathcal{S}_{b-L+c}(\mathbf{l}_{b-L+c-1}) := \{i \in \mathcal{D}_1 : l_{i,b-L+c-1} \neq 1\}$, we have

$$p'_{b-L+c} = \Pr\left(\bigcup_{\mathbf{l}_{b-L+c}=\mathbf{1}, \mathbf{l}_{b-L+c-1} \neq \mathbf{1}} \mathcal{A}_{b-L+c}(\mathbf{l}_{b-L+c-1}, \mathbf{l}_{b-L+c}) \right)$$

$$= \Pr\left(\bigcup_{\mathbf{l}_{b-L+c}=\mathbf{1}, \mathbf{l}_{b-L+c-1}:\mathcal{S}_{b-L+c}(\mathbf{l}_{b-L+c-1}) \neq \emptyset} \mathcal{A}_{b-L+c}(\mathbf{l}_{b-L+c-1}, \mathbf{l}_{b-L+c}) \right)$$

$$\leq \sum_{\mathcal{S}_{b-L+c} \neq \emptyset} \sum_{\substack{\mathbf{l}_{b-L+c} = \mathbf{1}, \\ \mathbf{l}_{b-L+c-1} : \mathcal{S}_{b-L+c}(\mathbf{l}_{b-L+c-1}) = \mathcal{S}_{b-L+c}}} \Pr(\mathcal{A}_{b-L+c}(\mathbf{l}_{b-L+c-1}, \mathbf{l}_{b-L+c}))$$

$$\leq \sum_{\mathcal{S}_{b-L+c} \neq \emptyset} 2^{T(\sum_{i \in \mathcal{S}_{b-L+c}} (I(Y_i;\hat{Y}_i|X_i)+\epsilon))} 2^{-T(\mathcal{I}(\mathcal{S}_{b-L+c})-\epsilon')}$$

$$\leq 2^n 2^{-T(\min_{\mathcal{S} \subseteq \mathcal{D}_1 : \mathcal{S} \neq \emptyset} \{I(X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}}, Y_1|X_0,X_1,X_{\mathcal{D}_1 \setminus \mathcal{S}}) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_0,X_1,X_{\mathcal{D}_1},Y_1,\hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}})-\epsilon''\})},$$

and thus $p_{b-L+a} \to 0$ as $T \to \infty$ if (28) holds. Therefore, if (28) holds, the first term in (52) also goes to 0 as $T \to \infty$, and $\mathbf{l}_{b-L}$, i.e., $\{l_{i,b-L}, i \in \mathcal{D}_1\}$ can be decoded.

ACKNOWLEDGMENT

REFERENCES

[1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[2] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, 1979.

[3] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 6, pp. 841–850, Nov. 1982.

[4] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.

[5] L.-L. Xie, "An improvement of Cover/El Gamal's compress-and-forward relay scheme," August 2009, available online at http://arxiv.org/abs/0908.0163.

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

[7] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sep. 2007.

[8] A. El Gamal and Y.-H. Kim, "Lecture notes on network information theory," January 2010, available online at http://arxiv.org/abs/1001.3404.

[9] P. Rost and G. Fettweis, "Analysis of a mixed strategy for multiple relay networks," *IEEE Trans. Inform. Theory*, vol. 55, pp. 174–189, Jan. 2009.

[10] M. R. Aref, "Information Flow in Relay Networks," Ph.D. dissertation, Stanford University, Stanford, CA, 1980.

[11] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1877–1894, Aug. 2003.

[12] A. Reznik, S. R. Kulkarni, and S. Verdú, "Degraded Gaussian multirelay channel: Capacity and optimal power allocation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3037–3046, Dec. 2004.

[13] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 748–767, May 2004.

[14] L.-L. Xie and P. R. Kumar, " An achievable rate for the multiple-level relay channel," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1348–1358, April 2005.

[15] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proc. 41th Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Oct. 2003.

[16] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3037–3063, September 2005.

[17] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.

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

[19] X. Wu and L.-L. Xie, "On the optimality of successive decoding in compress-and-forward relay schemes," in *Proc. 48th Annual Allerton Conf. Commun., Contr. Comput.*, pp. 534–541, Monticello, IL, Sep. 29-Oct. 1, 2010.

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

[21] G. Kramer and J. Hou, "Short-message quantize-forward network coding," in *Proc. of 8th Int. Workshop on Multi-Carrier Systems & Solutions*, Herrsching, Germany, 2011.

[22] G. Kramer and J. Hou, "On message lengths for noisy network coding," in *Proc. IEEE Inf. Theory Workshop*, Paraty, Brazil, 2011.

[23] A. Raja and P. Viswanath, "Compress-and-forward scheme for relay networks: backward decoding and connection to bisubmodular flows," available online at http://arxiv.org/abs/1012.0416v3.

[24] A. Behboodi and P. Piantanida, "Selective coding strategy for unicast composite networks," in *Proc. of the IEEE International Symposium on Information Theory*, Cambridge, MA, USA, July 2012.

[25] L.-L. Xie and P. R. Kumar, "Multi-source, multi-destination, multi-relay wireless networks," *IEEE Transactions on Information Theory, Special Issue on Models, Theory and Codes for Relaying and Cooperation in Communication Networks*, vol. 53, pp. 3586–3595, October 2007.

[26] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.