New Achievable Sum Degrees of Freedom in Half-duplex Single-antenna Multi-user Multi-hop Networks

Chao Wang, Ping Wang, Fuqiang Liu, and Geyong Min

Abstract

We investigate the achievable sum degrees of freedom (DoF) in a class of single-antenna multi-user multi-hop relay networks. The networks consist of multiple information sources and destinations, without direct signal propagation link between them, so that multiple layers of relays are deployed to assist in information delivery. We consider the situation that relays are unable to shield their receptions from the harmful self-interference and from the interference generated by other relays. Hence ideal full-duplex relaying is not applicable. Utilizing half-duplex decode-and-forward relays, a cluster successive relaying (CSR) transmission scheme is adopted to conduct message transmission. The CSR scheme divides each layer of relays into two successively activated relay clusters to compensate the extra channel consumption demanded by the half-duplex operation. We propose two interference alignment strategies to deal with the interference issues. By properly clustering the relays in each layer, we find the asymptotically achievable sum DoF, subject to time-varying and frequency-selective fading respectively. These results can lead to new lower bounds for the available DoF in the considered class of multi-user multi-hop networks.

Index Terms

Degrees of freedom (DoF), multi-user multi-hop networks, half-duplex relaying, interference alignment.

I. Introduction

Thanks to the elegant concept of interference alignment (IA) [1]–[4], our understanding of the performance limits in large wireless communication networks has been greatly improved over the last few years. Through evaluating the sum *degrees of freedom (DoF)*, how the total transmission data rate in a wireless network may scale with varying signal-to-noise ratio (SNR) can be roughly quantified. A number of discoveries have revealed that the achievable sum DoF can be related to the network scale. This exhibits great advantages of potential network-level transmission designs over the conventional orthogonal link-level designs.

The basic idea behind IA is to force multi-user interference signals to align in a compact receive subspace so that more radio resources can be allocated for desired information transmission. For multiple-input multiple-output (MIMO) systems the multi-dimensional receive space is readily generated in the spatial domain (see e.g., [5]–[8]). But for single-antenna networks it can be created through the *channel-extension technique* [2], [9] in time-varying or frequency-selective fading environments. For instance, it has been shown that the asymptotically achievable sum DoF in a single-antenna M-user interference channel (i.e., a network with M source-destination pairs) can be as high as $\frac{M}{2}$ [2], and that in a single-antenna $M_s \times M_d$

This work was funded in part by the National Natural Science Foundation of China (61771343 and 61331009), the EU Horizon 2020 Programme Marie Sklodowska-Curie Individual Fellowship (H2020-MSCA-IF-2016-752979), and the EU PF7 QUICK project (PIRESES-GA-2013-612652). The material in this paper was presented in part at the 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, April 2018. Ping Wang is the corresponding author.

C. Wang, P. Wang, and F. Liu are with the College of Electronics and Information Engineering, Tongji University, Shanghai, 201804, China (email: chaowang@tongji.edu.cn, pwang@tongji.edu.cn, liufuqiang@tongji.edu.cn). C. Wang is also with the Department of Computer Science, College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, EX4 4QF, U.K.

G. Min is with the Department of Computer Science, College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, EX4 4QF, U.K. (email: g.min@exeter.ac.uk).

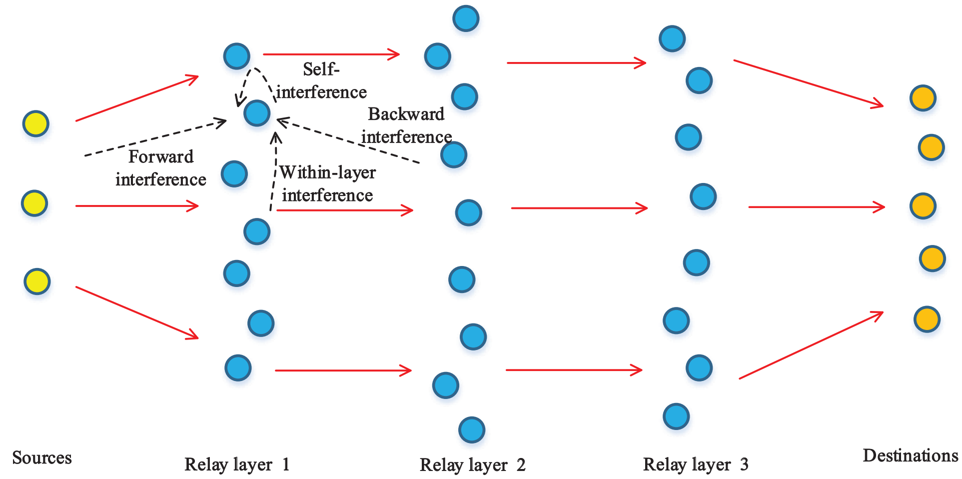


Fig. 1. A single-antenna multi-user multi-hop network. Four types of interference would occur if relays operate in the full-duplex fashion. Hence we consider applying half-duplex relays to conduct message transmissions.

X channel (i.e., a network with M_s sources and M_d destinations) can be $\frac{M_s M_d}{M_s+M_d-1}$ [9]. Both results are much larger than one, which is the sum DoF achieved by classic orthogonal transmissions.

Recently the research attention has been naturally extended from single-hop networks to multi-hop relay networks. When the sources' signals can reach all destinations, the *optimally achievable sum DoF* (referred to as *available DoF* throughout the paper) would be limited by the result in the direct single-hop source-destination network [10]. From the DoF perspective, relays can be discarded without causing performance loss. When there is no direct communication link between the sources and destinations, however, an interesting discovery has been made. Consider a single-antenna M-user two-hop network with M dedicated source-destination pairs. When the number of relays is sufficiently large, the achievable sum DoF can reach M, which is twice the result attained in an M-user interference channel [11]–[13]. The fact holds even when the network has more than two hops [14], [15]. This means that relays can bring not only information delivery paths between the sources and destinations, but also DoF gains over single-hop networks.

But realizing the above results demands an ideal full-duplex relaying assumption. Consider a multi-user multi-hop network as shown in Fig. 1. If all the relays transmit and receive signals simultaneously using the same frequency band, in addition to desired signals coming from the preceding layer, each relay in fact also experiences *four types of interference*: the unwanted signals generated by terminals located respectively in the preceding layer, in the same layer, and in the following layer, as well as the signals generated by its own transmit antenna. We term these interference signals *forward interference*, *within-layer interference*, *backward interference*, and *self-interference* respectively. Most existing investigations on DoF analysis in wireless multi-hop networks take only the forward interference into account, and ignore the other three types. This consideration simplifies the transmission design problem and facilitates reaching insightful discoveries. But the results may not reflect the actual performance in more general conditions where those interference signals cannot be directly neglected. Certainly, half-duplex relaying serves as a natural solution to tackling these interference issues. After all, in a two-hop network, permitting the relays to use orthogonal channels to receive and forward source messages can completely eliminate the self-interference and within-layer interference. Nevertheless, the extra channel usage demanded by half-duplex operation has the potential to reduce spectral efficiency. For example, utilizing half-duplex relays in the schemes proposed by [11]–[15] would halve the sum DoF results achieved in ideal full-duplex systems, as shown in [9], [16]. If source messages have to go through more hops to reach the destinations, the result would decrease further.

To address this problem, our previous work [17] proposed a spectrally-efficient half-duplex *cluster successive relaying (CSR)* transmission scheme for a single-antenna M-user two-hop network. The system model was later extended to multi-user multi-hop networks in [18]. Different from most conventional relaying schemes that demand all relays in the same layer to operate in the same mode (either listening or forwarding) together, we divide each layer of relays into *two identical clusters* and alternatively activate them.[1] It was shown that, the achievable sum DoF can be much larger than the results attained by directly adopting half-duplex relays in the schemes proposed by [11]–[15]. Interestingly, if the number of relays in each layer approaches infinity, the achievable sum DoF approaches an upper bound for the available DoF. The exact available DoF is thus identified. The result is actually the same as that obtained in ideal full-duplex systems. This implies that the aforementioned interference issues may not necessarily reduce the system's DoF. Properly designed half-duplex relaying schemes can still serve to identify multi-hop networks' performance limits.

However, when the number of relays in each layer is limited, the optimally achievable sum DoF is currently still unknown. Pinpointing the exact available DoF in a general multi-user multi-hop relay network is extremely involved. The highest sum DoF presented in existing works would provide an achievable lower bound and help identify the region that the available DoF reside. In this paper, we will show that the known lower bound can be further increased, through finding new achievable sum DoF for the CSR scheme. Actually most existing works on the concept of successive relaying demands a symmetric setup such that at any time instant the number of listening relays and that of forwarding relays should be the same. Consequently, in [17], [18] if a relay layer contains an odd number of terminals, one relay would be discarded from the system. This requirement, which was believed to cause no performance loss, simplifies the IA transceiver filters design and performance analysis. But we argue that it may lead to an inefficient use of the system's hardware resources. In this paper we consider permitting the two clusters to have different sizes so that all relays can participate in the message transmission process. By careful relay clustering and IA construction, we show that improved sum DoF can be attained. The novelties of this paper can be summarized as follows.

1) We consider communications in a class of single-antenna wireless relay networks, which consist of multiple sources, multiple destinations, and multiple layers of intermediate relays. Different from most related works (e.g., [11]–[15]), we take all the four types of interference signals shown in Fig. 1 into account. Ideal full-duplex relaying is not applicable.[2] We consider deploying half-duplex decode-and-forward (DF) relays to carry out message transmission and adopt the CSR scheme to combat potential spectral efficiency loss induced by the half-duplex operation. The remaining interference issues are handled by IA construction at the distributed terminals. We show that properly designed half-duplex relaying strategy can provide better DoF results than directly adopting half-duplex relays in those schemes proposed for ideal full-duplex scenarios.
2) To better utilize the network's hardware resources, we relax the requirement of having two identical clusters in each relay layer, as posed by the original CSR scheme design in [17], [18]. We first fix the relay clustering strategy and show that applying the CSR scheme converts the considered multi-hop network into an equivalent single-hop network. For time-varying fading, the IA transceiver filters for the equivalent network are constructed. Then the asymptotically achievable sum DoF in the original multi-hop network are derived. Finally, we study how to divide each layer of relays into clusters in order to maximize the CSR scheme's achievable sum DoF. It is shown that, by appropriately involving all available relays, a higher sum DoF than those presented in prior literatures can be obtained.

[1]Using two relays to take turns forwarding source messages is widely referred to as successive relaying or alternative relaying. It has attracted considerable attentions in the past decade as an effective way to compensate the half-duplex multiplexing gain loss in single-user relay networks (see e.g., [19]–[22]). In addition to our works on systems with arbitrary sources, arbitrary destinations and arbitrary layers of relays, recently the concept has also been adopted in M-user two-hop single-antenna [23], [24] and MIMO [25], [26] systems to identify the DoF performance in large relay networks.

[2]Although the full-duplex technology has advanced rapidly in recent years, the capability of self-interference mitigation usually demands complex algorithms and costly hardware [27], and thus may not be widely applicable to small, cheap, and power-limited devices in the near future. In addition, even if the relay self-interference can be significantly reduced, the within-layer and backward interference can still severely harm the system performance if terminals operate in the full-duplex fashion.

3) In addition, when the channel fading is frequency-selective, based on channel reciprocity we show that IA design for a dual equivalent network can serve as an extra option to carry out message delivery. This may lead to a further enhancement of the achievable sum DoF, compared with the time-varying fading case. Although the basic idea behind the channel-extension based IA filters construction is similar to those presented in [2], [9], [18], the design has its own constraints and solutions. The DoF results cannot be directly conjectured from existing works. Using these results, we can identify new lower bounds for the available DoF. To the best of our knowledge, these are the best results so far in the considered class of single-antenna multi-user multi-hop networks. Therefore they can help provide new knowledge regarding the performance limits in large relay networks.

The remainder of the paper is organized as follows. In Section II we describe our system model, the CSR transmission process, and the main results of the paper. The achievability of these DoF results is established in Sections III and IV. Specifically, Section III elaborates two IA construction strategies for the equivalent single-hop network created by the CSR scheme. Section IV provides the achievable sum DoF analysis in the original multi-hop network. Finally, Section V concludes the paper.

Notations: $|\mathcal{A}|$ denotes the cardinality of set $\mathcal{A}$. $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ represent the floor and ceiling functions respectively. The transpose and rank of matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$ and $\text{rank}(\mathbf{A})$. $\text{span}(\mathbf{A})$ denotes the space spanned by the column vectors of $\mathbf{A}$. $\mathbf{O}$ denotes an all-zero matrix.

II. System Model and Main Results

A. Network and Channel Models

In this paper we consider a class of wireless single-antenna multi-user multi-hop communication networks. The networks consist of $N+2$ ($N \geq 1$) layers of terminals, i.e., a source layer with $M_s \geq 2$ independent sources, a destination layer with $M_d \geq 2$ independent destinations, and N layers of relays between them. The ith ($i \in \{1, 2, \cdots, N\}$) relay layer deploys $K_i \geq 4$ distributed DF relay nodes. Every source attempts to send one independent message to every destination. In other words, the information delivery demand is similar to that in a single-hop $M_s \times M_d$ X channel: A total of $M_s M_d$ messages generated from the M_s sources should be delivered to the M_d destinations, through the relays. Such $(N+1)$-hop networks are denoted by a general form $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ throughout the paper. A simple four-hop $\{3, 7, 9, 8, 5\}^{(4)}$ example network is illustrated in Fig. 1.

The considered networks exhibit a *layered structure*. Non-adjacent layers do not have direct signal propagation links. But if a terminal is operating in the receiving mode, it always overhears concurrent transmissions of other terminals located in the same layer and of those in adjacent layers using the same frequency band. The channel fading coefficients between these terminals are modelled by independent random variables generated from a continuous distribution, with absolute values bounded away from zero and infinity [10]. In addition, due to hardware limitation, the relays cannot shield their reception from their own transmission, if they operate in a full-duplex fashion. Therefore, we consider using *half-duplex relaying* to avoid the self-interference and keep the remaining three types of interference shown in Fig. 1 to be controllable.

The message transmission process is conducted via a slotted fashion, in either a narrow-band time-varying or a wide-band frequency-selective block-fading environment: The source messages are delivered to the destinations using multiple (unit-length) time slots. Under time-varying fading, channel fading coefficients would change independently across different time slots. Under frequency-selective fading, the available frequency band can be converted into multiple parallel independently-faded frequency-flat unit frequency channels, and the fading coefficients remain unchanged for the entire transmission duration. Channel knowledge regarding the whole network is causally available at all terminals.[3] Perfect

[3]Essentially the channel-extension based IA design does not distinguish these two fading setups. Most previous investigations on the DoF in single-antenna networks attain the same result in both cases. However, for frequency-selective fading, terminals can have the channel knowledge regarding the whole transmission duration since fading coefficients do not change, while for time-varying fading future channel knowledge is not attainable. As will be discussed later, armed with the former level of channel knowledge, we can have an extra option to construct IA design and potentially further improve the system's achievable sum DoF performance in frequency-selective fading environments.

synchronization among terminals is also assumed.

Use ρ to denote the maximum transmit power of a terminal. The achievable sum DoF, d_Σ, represent the scaling behaviour of the sum transmission data rate, R_Σ, regarding ρ as:

$$d_\Sigma = \lim_{\rho\to\infty} \frac{R_\Sigma}{\log_2 \rho}. \tag{1}$$

Certainly, if the available DoF (i.e., the optimally achievable sum DoF), d_Σ^*, can be identified, then we can characterize the network's sum capacity C_Σ by $C_\Sigma(\rho) = d_\Sigma^* \log_2 \rho + o(\log_2 \rho)$ when $\rho \to \infty$. For an $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network, it is easy to use the cut-set bound analysis to upper bound d_Σ^* by $d_\Sigma^* \leq \min\{M_s, K_1, \cdots, K_N, M_d\}$. But so far the exact value of d_Σ^* is known only for a few special cases. For instance, if the within-layer interference, backward interference and self-interference shown in Fig. 1 do not exist, then one can use ideal full-duplex relays and apply the transmission scheme proposed in [14] to obtain $d_\Sigma^* = \min\{M_s, K_1, \cdots, K_N, M_d\}$. If these interference signals cannot be neglected, [18] proves that, when the number of relays in each layer approaches infinity, then using half-duplex relays can still attain $d_\Sigma^* = \min\{M_s, K_1, \cdots, K_N, M_d\} = \min\{M_s, M_d\}$. This means that these interference issues may not necessarily affect network capacity's scaling behaviour, no matter how many layers of relays have to be used to conduct transmissions.

However, for network structures with finite relays in each layer, identifying d_Σ^* is extremely involved. In general one can only obtain a bounded characteristic of d_Σ^* as

$$d_\Sigma \leq d_\Sigma^* \leq \min\{M_s, K_1, \cdots, K_N, M_d\}. \tag{2}$$

Fining a higher value of the achievable sum DoF d_Σ would reduce the gap between the lower and upper bounds, and reduce the uncertainty regarding d_Σ^*. This is the objective of our paper.

In fact, to deliver information in the considered $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network, we can again adopt the scheme proposed by [14] but use half-duplex relays. Due to the half-duplex operation, this approach may cause the available channel resources to be inefficiently utilized, especially when the numbers of hops and relays are large. To handle this issue, our previous works [17], [18] proposed a CSR scheme that divides each layer of relays into two identical clusters and alternatively activates them to mimic full-duplex relays. When the number of relays in each layer is relatively large, the achievable sum DoF can be greater than the former method. However, the original CSR scheme design requires the two relay clusters to have the same size. If a relay layer has an odd number of terminals, one relay will be discarded. In this paper, we will show that this requirement does not efficiently use the available hardware resources. By allowing the two clusters to have different sizes, we can properly involve all the relays in the message delivery process and achieve higher sum DoF. Thus a new lower bound for d_Σ^* can be provided. In the following subsections, we will first describe the CSR transmission process and then present the main results of the paper. The achievability will be proved in Sections III and IV.

B. The CSR Scheme

In this subsection, we describe how the $M_s M_d$ source messages are transmitted to the destinations, by applying the CSR scheme. Divide each layer of relays into two clusters. We denote the sets of sources and destinations by $\mathcal{S}$ and $\mathcal{D}$, respectively. The two relay clusters in the ith ($i \in \{1, 2, \cdots, N\}$) relay layer are denoted by $\mathcal{R}_{i,1}$ and $\mathcal{R}_{i,2}$. The first cluster contains $|\mathcal{R}_{i,1}| = r_i$ ($2 \leq r_i \leq K_i - 2$) relays, and the second cluster has $|\mathcal{R}_{i,2}| = K_i - r_i$ relays.

Recall that every source intends to send an independent message to every destination. We divide each of these source messages into L ($L \geq 1$) *sub-message sets*. The L sub-message sets (i.e., one message) generated by the sth ($s \in \{1, 2, \cdots, M_s\}$) source and desired by the dth ($d \in \{1, 2, \cdots, M_d\}$) destination are denoted by $\mathcal{I}_{d,s}^{[1]}$, $\mathcal{I}_{d,s}^{[2]}$, $\cdots$, $\mathcal{I}_{d,s}^{[L]}$. Every sub-message set contains the same number of I independent *unit-DoF sub-messages*, which means each sub-message can be encoded into a Gaussian codeword with data rate $\log_2 \rho + o(\log_2 \rho)$ for large ρ. Hence every source has a total of LM_d sub-message sets (i.e.,

$\mathcal{I}_{1,s}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,s}^{[1]}$, $\mathcal{I}_{1,s}^{[2]}$, $\cdots$, $\mathcal{I}_{M_d,s}^{[2]}$, $\cdots$, $\mathcal{I}_{1,s}^{[L]}$, $\cdots$, $\mathcal{I}_{M_d,s}^{[L]}$) to transmit and every destination expects LM_s sub-message sets (i.e, $\mathcal{I}_{d,1}^{[1]}$, $\cdots$, $\mathcal{I}_{d,M_s}^{[1]}$, $\mathcal{I}_{d,1}^{[2]}$, $\cdots$, $\mathcal{I}_{d,M_s}^{[2]}$, $\cdots$, $\mathcal{I}_{d,1}^{[L]}$, $\cdots$, $\mathcal{I}_{d,M_s}^{[L]}$).

The message transmission process is conducted using $L+N$ consecutive *time intervals*. For time-varying fading, the nth ($n \in \{1, 2, \cdots, L+N\}$) time interval contains $F_{[n]}$ unit time slots. For frequency-selective fading, every time interval is a unit time slot, but the available frequency band contains $F_{[n]}$ independent unit frequency channels. During each of the first L time intervals, every source broadcasts M_d sub-message sets, one dedicated for each destination, to the first layer of relays. That is, during the nth ($n \in \{1, 2, \cdots, L\}$) time interval, the sth source transmits $\mathcal{I}_{1,s}^{[n]}$, $\mathcal{I}_{2,s}^{[n]}$, $\cdots$, $\mathcal{I}_{M_d,s}^{[n]}$. The two clusters in each relay layer are activated alternatively. At one time interval, the relays within one cluster listen to the transmissions of the terminals in the preceding layer, and the relays within the other cluster forward their previously received sub-messages to the following layer. At the next interval, the two clusters exchange their functioning. In each hop, every transmitter evenly divides its transmit sub-messages and send them to the receivers respectively.

Specifically, during the first time interval, $\mathcal{S}$ transmits M_sM_d sub-message sets $\mathcal{I}_{1,1}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,1}^{[1]}$, $\mathcal{I}_{1,2}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,2}^{[1]}$, $\cdots$, $\mathcal{I}_{1,M_s}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,M_s}^{[1]}$ to the relays in $\mathcal{R}_{1,1}$. During the second time interval, $\mathcal{S}$ transmits M_sM_d new sub-message sets $\mathcal{I}_{1,1}^{[2]}$, $\cdots$, $\mathcal{I}_{M_d,1}^{[2]}$, $\mathcal{I}_{1,2}^{[2]}$, $\cdots$, $\mathcal{I}_{M_d,2}^{[2]}$, $\cdots$, $\mathcal{I}_{1,M_s}^{[2]}$, $\cdots$, $\mathcal{I}_{M_d,M_s}^{[2]}$ to the other relay cluster $\mathcal{R}_{1,2}$. The relays in $\mathcal{R}_{1,1}$ now forward the sub-messages they received in the first time interval, i.e., $\mathcal{I}_{1,1}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,1}^{[1]}$, $\mathcal{I}_{1,2}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,2}^{[1]}$, $\cdots$, $\mathcal{I}_{1,M_s}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,M_s}^{[1]}$, to the relays in $\mathcal{R}_{2,2}$. Certainly, the reception of $\mathcal{R}_{1,2}$ is interfered by the transmission of $\mathcal{R}_{1,1}$. This is the within-layer interference.

During the third time interval, $\mathcal{S}$ transmits M_sM_d new sub-message sets $\mathcal{I}_{1,1}^{[3]}$, $\cdots$, $\mathcal{I}_{M_d,1}^{[3]}$, $\mathcal{I}_{1,2}^{[3]}$, $\cdots$, $\mathcal{I}_{M_d,2}^{[3]}$, $\cdots$, $\mathcal{I}_{1,M_s}^{[3]}$, $\cdots$, $\mathcal{I}_{M_d,M_s}^{[3]}$ again to the relays in $\mathcal{R}_{1,1}$. At the same time, $\mathcal{R}_{1,2}$ forwards $\mathcal{I}_{1,1}^{[2]}$, $\cdots$, $\mathcal{I}_{M_d,1}^{[2]}$, $\mathcal{I}_{1,2}^{[2]}$, $\cdots$, $\mathcal{I}_{M_d,2}^{[2]}$, $\cdots$, $\mathcal{I}_{1,M_s}^{[2]}$, $\cdots$, $\mathcal{I}_{M_d,M_s}^{[2]}$ to the relays in $\mathcal{R}_{2,1}$, and $\mathcal{R}_{2,2}$ forwards $\mathcal{I}_{1,1}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,1}^{[1]}$, $\mathcal{I}_{1,2}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,2}^{[1]}$, $\cdots$, $\mathcal{I}_{1,M_s}^{[1]}$, $\cdots$, $\mathcal{I}_{M_d,M_s}^{[1]}$ to the relays in $\mathcal{R}_{3,1}$. Now the reception of $\mathcal{R}_{1,1}$ experiences within-layer interference generated by $\mathcal{R}_{1,2}$ and backward interference generated by $\mathcal{R}_{2,2}$. The reception of $\mathcal{R}_{2,1}$ also experiences within-layer interference from $\mathcal{R}_{2,2}$.

The process continues similarly. Choose L to be large. Then during most time intervals, we can summarize the operations of all the clusters as follows, and illustrate them in Fig. 2(a).

- During an *odd* time interval: $\mathcal{S}$, $\mathcal{R}_{1,2}$, $\mathcal{R}_{2,2}$, $\cdots$, $\mathcal{R}_{N,2}$ operate in the transmitting mode. $\mathcal{R}_{1,1}$, $\mathcal{R}_{2,1}$, $\cdots$, $\mathcal{R}_{N,1}$, $\mathcal{D}$ operate in the receiving mode. In other words, terminals within $\mathcal{R}_{i,1}$ ($i \in \{1, 2, \cdots, N\}$) intend to receive a total of M_sM_d sub-message sets from the terminals within $\mathcal{R}_{i-1,2}$ ($\mathcal{S}$ for the case $i=1$). But their reception is interfered by the transmission of $\mathcal{R}_{i,2}$ (i.e., the within-layer interference) and $\mathcal{R}_{i+1,2}$ (i.e., the backward interference).
- During an *even* time interval: $\mathcal{S}$, $\mathcal{R}_{1,1}$, $\mathcal{R}_{2,1}$, $\cdots$, $\mathcal{R}_{N,1}$ operate in the transmitting mode. $\mathcal{R}_{1,2}$, $\mathcal{R}_{2,2}$, $\cdots$, $\mathcal{R}_{N,2}$, $\mathcal{D}$ operate in the receiving mode. Terminals within $\mathcal{R}_{i,2}$ desire to receive a total of M_sM_d sub-message sets from the terminals within $\mathcal{R}_{i-1,1}$ ($\mathcal{S}$ for the case $i=1$), while being interfered by $\mathcal{R}_{i,1}$ (within-layer interference) and $\mathcal{R}_{i+1,1}$ (backward interference).

Since the network has $N+1$ hops, the last M_sM_d sub-message sets $\mathcal{I}_{1,1}^{[L]}$, $\cdots$, $\mathcal{I}_{M_d,1}^{[L]}$, $\mathcal{I}_{1,2}^{[L]}$, $\cdots$, $\mathcal{I}_{M_d,2}^{[L]}$, $\cdots$, $\mathcal{I}_{1,M_s}^{[L]}$, $\cdots$, $\mathcal{I}_{M_d,M_s}^{[L]}$ would be delivered by the last layer of relays to the destinations during the $(L+N)$th time interval. Therefore, if all the source messages can be *reliably* delivered to the destinations, then a total of $\sum_{n=1}^{L+N} F_{[n]}$ unit channels are used to complete the transmission of LM_sM_dI unit-DoF sub-messages. The achievable sum DoF, d_{CSR}, is calculated as

$$d_{\text{CSR}} = \lim_{\rho\to\infty} \frac{LM_sM_dI\left(\log_2\rho + o(\log_2\rho)\right)}{\left(\sum_{n=1}^{L+N} F_{[n]}\right)\log_2\rho} = \frac{LM_sM_dI}{\sum_{n=1}^{L+N} F_{[n]}}. \tag{3}$$

A high d_{CSR} can be achieved by signal transmission strategies that produce a large value of I with a small value of $\sum_{n=1}^{L+N} F_{[n]}$. To find such solutions, we inspect the CSR transmission process and Fig. 2(a). It can be clearly seen that during *any* individual time interval, the transmission in the considered

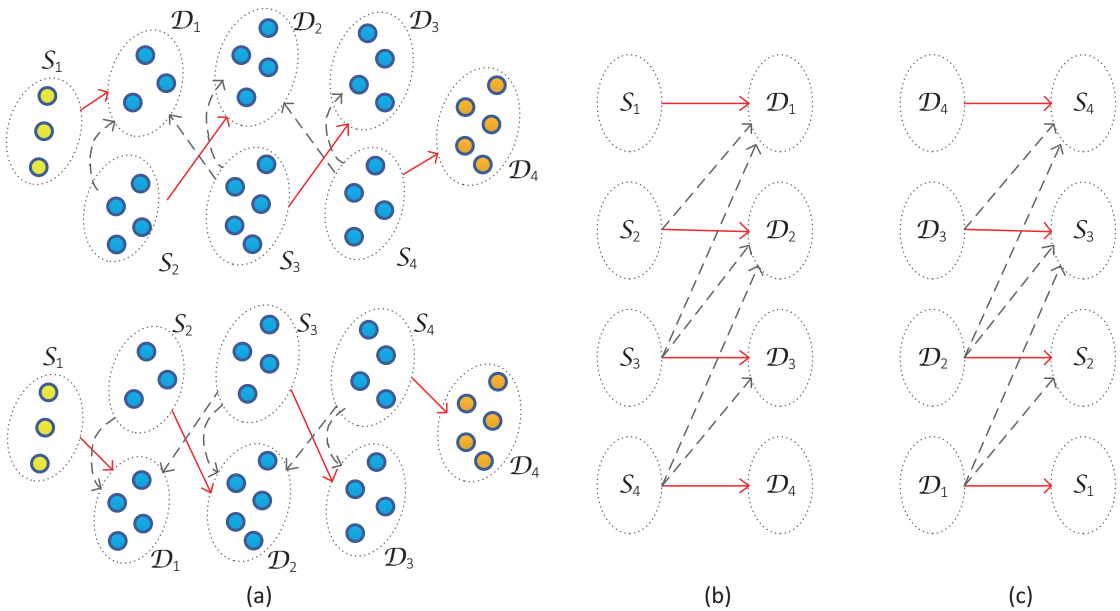


Fig. 2. (a) The CSR transmission process in two consecutive time intervals. (b) The primary equivalent network, and (c) the dual equivalent network. Solid arrow lines represent intended transmission directions. Dashed arrow lines represent inter-cluster interference.

$(N+1)$-hop network is similar to that in a *single-hop* network as displayed in Fig. 2(b). This single-hop network contains $N+1$ *pairs of transmitter-receiver clusters*, denoted by $\mathcal{S}_i$ and $\mathcal{D}_i$ $(i = 1, 2, \cdots, N+1)$ respectively. Every transmitter cluster $\mathcal{S}_i$ intends to send the same number of $M_s M_d I$ sub-messages to $\mathcal{D}_i$. For $i \in \{1, 2, \cdots, N-1\}$, in addition to the signals coming from $\mathcal{S}_i$, the reception of $\mathcal{D}_i$ experiences two types of *inter-cluster interference*, generated by $\mathcal{S}_{i+1}$ and $\mathcal{S}_{i+2}$ respectively. The reception of $\mathcal{D}_N$ is interfered by the transmission of one unintended cluster $\mathcal{S}_{N+1}$. Receivers in $\mathcal{D}_{N+1}$ do not experience inter-cluster interference. In the remainder of the paper, we term this single-hop network the *primary equivalent network.*

Clearly, during an odd time interval, the transmitter clusters $\mathcal{S}_1$, $\mathcal{S}_2$, $\cdots$, $\mathcal{S}_{N+1}$ in the primary equivalent network represent the clusters $\mathcal{S}$, $\mathcal{R}_{1,2}$, $\cdots$, $\mathcal{R}_{N,2}$ operating in the transmitting mode in the original multi-hop network. The receiver clusters $\mathcal{D}_1$, $\mathcal{D}_2$, $\cdots$, $\mathcal{D}_{N+1}$ represent $\mathcal{R}_{1,1}$, $\mathcal{R}_{2,1}$, $\cdots$, $\mathcal{D}$, the clusters that operate in the receiving mode. Similarly, during an even time interval, $\mathcal{S}_1$, $\mathcal{S}_2$, $\cdots$, $\mathcal{S}_{N+1}$ represent $\mathcal{S}$, $\mathcal{R}_{1,1}$, $\cdots$, $\mathcal{R}_{N,1}$, and $\mathcal{D}_1$, $\mathcal{D}_2$, $\cdots$, $\mathcal{D}_{N+1}$ represent $\mathcal{R}_{1,2}$, $\mathcal{R}_{2,2}$, $\cdots$, $\mathcal{D}$. The inter-cluster interference between $\mathcal{S}_{i+1}$ and $\mathcal{D}_i$ is the within-layer interference, and that between $\mathcal{S}_{i+2}$ and $\mathcal{D}_i$ is the backward interference in the $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network. These are displayed in Fig. 2(a).

Since the CSR scheme can convert the original multi-hop network into an equivalent single-hop network, we can actually focus on each individual time interval and find signal transmission solutions for the equivalent network, as long as the fact that a relay can only forward the sub-messages it received in the past time interval is taken into account. In this paper we utilize the channel-extension based IA technique to efficiently tackle the interference in the primary equivalent network. By this mean, we can decide the values of I, $F_{[n]}$, and also the number of relays in each cluster r_i to maximize d_{CSR} expressed in (3). In the next subsection, we present the main findings of the paper. The detailed IA design and DoF analysis are provided afterwards.

C. Main Results

To facilitate presentation, we define a function $\eta(m_s, k_1, \cdots, k_n, m_d)$ regarding integer parameters $m_s \geq 2$, $m_d \geq 2$, $n \geq 1$, and $k_1, \cdots, k_n \geq 4$ as:

$$\begin{aligned}\eta(m_s, k_1, \cdots, k_n, m_d) \quad = \quad \min \Bigg\{ & \frac{2}{\frac{2}{m_s} + \frac{1}{\lfloor \frac{k_1}{2} \rfloor} + \frac{1}{\lceil \frac{k_1}{2} \rceil}}, \frac{2}{\frac{1}{m_s} + \frac{1}{\lfloor \frac{k_1}{2} \rfloor} + \frac{1}{\lfloor \frac{\min\{k_1, \cdots, k_n\}}{2} \rfloor} + \frac{1}{\lceil \frac{\min\{k_1, \cdots, k_n\}}{2} \rceil}}, \\ & \frac{2}{\frac{1}{m_s} + \frac{1}{\lfloor \frac{k_1}{2} \rfloor} + \frac{\lfloor \frac{k_n}{2} \rfloor + m_d - 1}{\lfloor \frac{k_n}{2} \rfloor m_d}}, \frac{2}{\frac{\lfloor \frac{k_n}{2} \rfloor + m_d - 1}{\lfloor \frac{k_n}{2} \rfloor m_d} + \frac{1}{\lfloor \frac{\min\{k_1, \cdots, k_n\}}{2} \rfloor} + \frac{1}{\lceil \frac{\min\{k_1, \cdots, k_n\}}{2} \rceil}}, \\ & \frac{1}{\frac{1}{\lfloor \frac{\min\{k_1, \cdots, k_n\}}{2} \rfloor} + \frac{1}{\lceil \frac{\min\{k_1, \cdots, k_n\}}{2} \rceil}}, \frac{2}{\frac{\lfloor \frac{k_n}{2} \rfloor + m_d - 1}{\lfloor \frac{k_n}{2} \rfloor m_d} + \frac{\lceil \frac{k_n}{2} \rceil + m_d - 1}{\lceil \frac{k_n}{2} \rceil m_d}} \Bigg\}. \end{aligned} \tag{4}$$

We will use $\eta(m_s, k_1, \cdots, k_n, m_d)$ to express the achievable sum DoF of our CSR scheme, d_{CSR}. Let us start from the time-varying fading environment. The result is summarized as follows.

Proposition 1. *Under time-varying fading, applying the CSR scheme in the considered single-antenna $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network can asymptotically achieve the sum DoF*

$$d_{\mathit{CSR}} = \eta(M_s, K_1, \cdots, K_N, M_d). \tag{5}$$

This result is obtained by setting $|\mathcal{R}_{i,1}| = r_i = \lfloor \frac{K_i}{2} \rfloor$ and $|\mathcal{R}_{i,2}| = K_i - r_i = \lceil \frac{K_i}{2} \rceil$ for all $i \in \{1, 2, \cdots, N\}$.

In our previous work [18], by demanding both relay clusters in each layer to have the same size of $\lfloor \frac{K_i}{2} \rfloor$, the following sum DoF can be asymptotically achieved:

$$d_{\text{CSR-old}} = \min \left\{ \frac{M_s \lfloor \frac{K_1}{2} \rfloor}{M_s + \lfloor \frac{K_1}{2} \rfloor}, \frac{\left\lfloor \frac{\min\{K_1, \cdots, K_N\}}{2} \right\rfloor}{2}, \frac{\lfloor \frac{K_N}{2} \rfloor M_d}{\lfloor \frac{K_N}{2} \rfloor + M_d - 1} \right\}. \tag{6}$$

From Proposition 1, it is not difficult to show that when K_1, K_2, $\cdots$, K_N are all even numbers, the expression for the new achievable sum DoF (5) is identical to (6). However, if some relay layers contain odd numbers of terminals, the result in (5) can be larger. We will use a few simple examples to explicitly demonstrate such a result.

Firstly, consider a three-hop $\{3, 8, 8, 5\}^{(3)}$ example network. The result provided in [18], i.e. equation (6), shows that, by dividing each relay layer into two 4-relay clusters, the achievable sum DoF is $d_{\text{CSR-old}} = \frac{12}{7}$. This result is much larger than 1, the sum DoF achieved by adopting the half-duplex version of the scheme proposed by [14]. Now consider that each relay layer has 9 terminals. Using equation (6) the result remains to be $\frac{12}{7}$, because although an extra relay exists in each layer it will be discarded. Proposition 1, however, shows that a larger sum DoF $d_{\text{CSR}} = \frac{120}{67}$ can actually be attained (a 4.5% improvement). In fact $d_{\text{CSR}} = \frac{120}{67}$ remains the same in $\{3, 9, 9, \cdots, 9, 5\}^{(N+1)}$ networks for any value of $N \geq 1$. The number of layers of half-duplex relays that the source messages have to go through does not affect the achievable sum DoF.

To more explicitly exhibit the advantage of properly involving all relays in the message transmission process, we display the d_{CSR} achieved in three-hop $\{3, 2K+1, 2K+1, 5\}^{(3)}$ networks (i.e., both relay layers contain an odd number of $2K+1$ relays) versus the values of K in Fig. 3. The DoF gain over $d_{\text{CSR-old}}$ can be clearly observed. When the number of relays increases, d_{CSR} also increases to approach 3, the upper bound $\min\{M_s, K_1, K_2, M_d\}$ for the available DoF. This is in line with the result presented in [18].

If we swap the numbers of sources and destinations in the above three-hop network and consider a $\{5, 9, 9, 3\}^{(3)}$ network, the achievable sum DoF calculated in Proposition 1 is $d_{\text{CSR}} = \frac{60}{29}$. It is interesting to observe different values of d_{CSR} in the $\{3, 9, 9, 5\}^{(3)}$ and $\{5, 9, 9, 3\}^{(3)}$ networks. This is partly because,

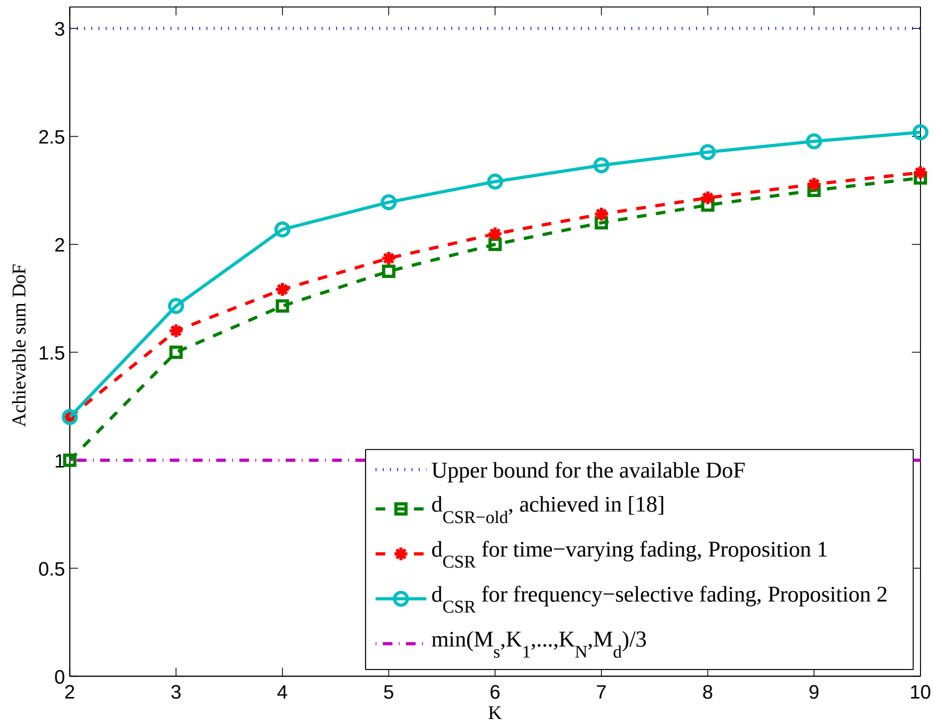


Fig. 3. Achievable sum DoF of the CSR scheme in $\{3, 2K+1, 2K+1, 5\}^{(3)}$ networks.

the CSR scheme converts the considered multi-hop network into a single-hop equivalent network with an asymmetric structure. Those two multi-hop systems thus have different equivalent networks. The other reason is that these results are only the lower bounds for d_Σ^*. They are obtained by different IA strategies and it is not yet known whether they accurately reflect the actual behaviour of d_Σ^*. Moreover, although this observation holds in time-varying fading environments, in this paper we will show that under frequency-selective fading the achievable sum DoF in the $\{3, 9, 9, 5\}^{(3)}$ network can be improved from $\frac{120}{67}$ by 15.5% to $d_{\text{CSR}} = \frac{60}{29}$. Certainly, the sizes of different relay layers do not need to be identical. Proposition 1 is applicable to general $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ networks. For instance, considering a four-hop $\{3, 13, 18, 25, 5\}^{(4)}$ network, the new achievable sum DoF is $d_{\text{CSR}} = \frac{84}{41}$, which is 2.4% larger than $d_{\text{CSR-old}} = 2$ obtained using equation (6).

Now, let us present the achievable sum DoF result when the fading is frequency-selective.

Proposition 2. *Under frequency-selective fading, applying the CSR scheme in the considered single-antenna $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network can asymptotically achieve the sum DoF*

$$d_{CSR} = \max\Big\{\eta(M_s, K_1, \cdots, K_N, M_d), \eta(M_d, K_N, \cdots, K_1, M_s)\Big\}. \tag{7}$$

This result is obtained by setting $|\mathcal{R}_{i,1}| = r_i = \left\lfloor \frac{K_i}{2} \right\rfloor$ and $|\mathcal{R}_{i,2}| = K_i - r_i = \left\lceil \frac{K_i}{2} \right\rceil$ for all $i \in \{1, 2, \cdots, N\}$.

The achievable sum DoF result presented in Proposition 2 is different from that in Proposition 1. This is because, as mentioned earlier, the channel knowledge available at terminals can be different in these two fading scenarios. We will show in Section III-B that with the channel state information regarding the entire transmission duration, for frequency-selective fading, in addition to the method that achieves $\eta(M_s, K_1, \cdots, K_N, M_d)$ shown in (5), an extra option for IA design based on channel reciprocity can be realized. Since the CSR scheme converts the original $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network into an equivalent network with asymmetric structure, the sum DoF achieved based on this new

IA construction method, $\eta(M_d, K_N, \cdots, K_1, M_s)$, can be different from $\eta(M_s, K_1, \cdots, K_N, M_d)$. The strategy that provides higher sum DoF will be chosen to conduct message transmission, which leads to (7).

Let us see some examples. Consider again the three-hop $\{3,8,8,5\}^{(3)}$ network. Even though under time-varying fading the value of d_{CSR} presented in Proposition 1 remains to be the same as $d_{\text{CSR-old}} = \frac{12}{7}$, in a frequency-selective fading environment the result can be improved by 16.7% to $d_{\text{CSR}} = 2$. In addition, for the three-hop $\{3,9,9,5\}^{(3)}$ network, the achievable sum DoF can be further improved to $d_{\text{CSR}} = \frac{60}{29}$, larger than the result shown in Proposition 1. The dependency of d_{CSR} regarding K in the $\{3, 2K+1, 2K+1, 5\}^{(3)}$ networks is also displayed in Fig. 3. A notable improvement over $d_{\text{CSR-old}}$ can be observed. Further, for the aforementioned four-hop $\{3,13,18,25,5\}^{(4)}$ network, in the frequency-selective fading environment, the achievable sum DoF is $d_{\text{CSR}} = \frac{126}{55}$, which is 11.8% larger than the result in a time-varying fading environment.

It should be noted that the results presented in Propositions 1 and 2 may not always serve as the highest achievable lower bound (i.e., d_Σ in (2)) for the available DoF in an arbitrary $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network. The idea behind the concept of successive relaying is to use two alternatively activated relays to mimic one full-duplex relay. For our CSR scheme, when the number of relays in each layer is large, the relays would introduce sufficient degrees of freedom to efficiently align and cancel interference. In general the CSR scheme provides high achievable sum DoF when each relay layer contains a relatively large number of terminals. If this is not the case, we may choose using other half-duplex relaying schemes to deliver information.

Specifically, as mentioned earlier, one can adopt the elegant transmission scheme proposed in [14] (which can achieve the available DoF $d_\Sigma^* = \min\{M_s, K_1, \cdots, K_N, M_d\}$ in an ideal full-duplex system) but replace the full-duplex relays with half-duplex relays. This scheme is referred to as *aligned diagonalization relaying (ADR)*. Directly orthogonalizing the transmissions of all the layers would cause the achievable sum DoF to be very small when the number of hops is large. To avoid such a problem, when $N > 1$ we can allow the ith layer and $(i+3)$th layer of terminals to transmit signals non-orthogonally. Then the whole network would not experience within-layer and backward interference, and at any time instant at most three consecutive hops (two hops for the case $N = 1$) are orthogonalized. The achievable sum DoF thus is

$$d_{\text{ADR}} = \begin{cases} \frac{\min\{M_s, K_1, M_d\}}{2}, & \text{for } N = 1 \\ \frac{\min\{M_s, K_1, \cdots, K_N, M_d\}}{3}, & \text{for } N \geq 2 \end{cases}. \tag{8}$$

In addition, we can also follow the method proposed in [9] and treat the transmission in each single hop as that in an individual X channel. We term this scheme *cascaded X channel relaying (CXR)*. It is known that the optimally achievable sum DoF in an $m_s \times m_d$ X channel is $d_{\text{X}} = \frac{m_s m_d}{m_s + m_d - 1}$. This means that, averagely, delivering a unit-DoF sub-message (a Gaussian codeword with rate $\log_2 \rho + o(\log_2 \rho)$) demands $\frac{1}{d_{\text{X}}} = \frac{m_s + m_d - 1}{m_s m_d}$ unit time/frequency channels. Again, when $N \geq 2$ the ith and the $(i+3)$th layers of terminals can be activated together (i.e., the three consecutive layers, the ith, $(i+1)$th, $(i+2)$th layers are always orthogonalized to avoid interference). The achievable sum DoF, d_{CXR}, can be calculated as

$$d_{\text{CXR}} = \begin{cases} \frac{1}{\frac{M_s + K_1 - 1}{M_s K_1} + \frac{K_1 + M_d - 1}{K_1 M_d}}, & \text{for } N = 1 \\ \min\limits_{i \in \{0,1,\cdots,N-2\}} \left\{ \frac{1}{\frac{K_i + K_{i+1} - 1}{K_i K_{i+1}} + \frac{K_{i+1} + K_{i+2} - 1}{K_{i+1} K_{i+2}} + \frac{K_{i+2} + K_{i+3} - 1}{K_{i+2} K_{i+3}}} \right\}, & \text{for } N \geq 2 \end{cases} \tag{9}$$

where $K_0 = M_s$ and $K_{N+1} = M_d$ [18].

Finally, the *cascaded block relaying (CBR)* scheme proposed in [18] treats the considered $(N+1)$-hop $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network as two orthogonally-activated consecutive sub-networks (blocks), and properly chooses a transmission scheme in each block to deliver information. In particular, one can use the ϵth ($\epsilon \in \{1, 2, \cdots, N\}$) relay layer to divide the network. Then the first block is an ϵ-hop $\{M_s, K_1, \cdots, K_{\epsilon-1}, K_\epsilon\}^{(\epsilon)}$ network and the second is a $\{K_\epsilon, K_{\epsilon+1}, \cdots, K_N, M_d\}^{(K+1-\epsilon)}$ network with

TABLE I
SOME EXAMPLE NETWORKS AND THEIR ACHIEVABLE SUM DOF (ACCURACY IS WITHIN 0.1%).

Networks	Achieved in [18] (time-varying/frequency-selective fading)	New achievable result (time-varying fading)	New achievable result (frequency-selective fading)
$\{5,11,30\}^{(2)}$	2.538 (CXR)	$d_\Sigma = 2.609$ (CSR) ↑2.8%	$d_\Sigma = 2.727$ (CSR) ↑7.4%
$\{5,7,7,7\}^{(3)}$	1.667 (ADR)	$d_\Sigma = 1.714$ (CSR) ↑2.9%	$d_\Sigma = 1.714$ (CSR) ↑2.9%
$\{4,9,10,6\}^{(3)}$	2.000 (CSR)	$d_\Sigma = 2.105$ (CSR) ↑5.3%	$d_\Sigma = 2.222$ (CSR) ↑11.1%
$\{3,17,9,9\}^{(3)}$	2.000 (CSR)	$d_\Sigma = 2.202$ (CSR) ↑10.1%	$d_\Sigma = 2.222$ (CSR) ↑11.1%
$\{9,15,14,19,5\}^{(4)}$	3.462 (CSR)	$d_\Sigma = 3.481$ (CSR) ↑0.6%	$d_\Sigma = 3.481$ (CSR) ↑0.6%
$\{3,19,4,25,5\}^{(4)}$	1.286 (CBR)	$d_\Sigma = 1.300$ (CBR) ↑1.1%	$d_\Sigma = 1.401$ (CBR) ↑9.0%
$\{3,19,25,4,5\}^{(4)}$	1.184 (CBR)	$d_\Sigma = 1.192$ (CBR) ↑0.7%	$d_\Sigma = 1.244$ (CBR) ↑5.1%

$K+1-\epsilon$ hops. Let us use $d_{\text{CBR},\epsilon,1}$ and $d_{\text{CBR},\epsilon,2}$ to denote the maximally achievable sum DoF in these two blocks respectively. These two results can be obtained following the above discussions. For example, if a block contains more than one hop, then we can choose among the CSR, ADR, and CXR schemes and the maximally achievable sum DoF would be $\max\{d_{\text{CSR}}, d_{\text{ADR}}, d_{\text{CXR}}\}$. If a block contains only one hop, it can be considered as an X channel with achievable sum DoF calculated by d_{X}. Now, in order to reach the destination, averagely, each unit-DoF sub-message would require $\frac{1}{d_{\text{CBR},\epsilon,1}}$ unit channels to pass the first block and $\frac{1}{d_{\text{CBR},\epsilon,2}}$ unit channels to pass the second. As a result, the achievable sum DoF of such a CBR scheme is

$$d_{\text{CBR}} = \frac{1}{\frac{1}{d_{\text{CBR},\epsilon,1}} + \frac{1}{d_{\text{CBR},\epsilon,2}}} = \frac{d_{\text{CBR},\epsilon,1} d_{\text{CBR},\epsilon,2}}{d_{\text{CBR},\epsilon,1} + d_{\text{CBR},\epsilon,2}}. \tag{10}$$

For instance, consider a four-hop $\{3,19,4,25,5\}^{(4)}$ network. We can use the second layer of relays to divide the network into a two-hop $\{3,19,4\}^{(2)}$ network and a two-hop $\{4,25,5\}^{(2)}$ network. Applying the CSR scheme to each block leads to $d_{\text{CBR},2,1} = \frac{540}{237}$ and $d_{\text{CBR},2,2} = \frac{312}{103}$ in a time-varying fading environment. The achievable sum DoF is $d_{\text{CBR}} = \frac{4680}{3599}$, much better than the result of 1 achieved by directly applying the CSR scheme to the whole network.

All of these four schemes (i.e., the CSR, ADR, CXR, and CBR schemes) can be applied in the considered $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network. The method that provides the highest sum DoF would be chosen to deliver information. Therefore, based on Propositions 1 and 2 we can have the following corollary regarding the highest achievable sum DoF known so far.

Corollary 1. *A lower bound for the available DoF in the considered single-antenna multi-user $(N+1)$-hop $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network is expressed as*

$$d_\Sigma = \max\left\{d_{ADR}, d_{CXR}, d_{CSR}, d_{CBR}\right\}. \tag{11}$$

Table I lists the values of d_Σ derived using Corollary 1 in a few more example networks. But the second column uses (6) to obtain d_{CSR}, i.e., the result achieved previously in [18]. The scheme that leads to the highest achievable sum DoF is also placed after the DoF result (e.g., for the first $\{5,11,30\}^{(2)}$ example network, the CXR scheme contributes to the calculation of d_Σ). The third and fourth columns use Propositions 1 and 2 to derive d_{CSR}. The notation $\uparrow a\%$ denotes that the new result is an $a\%$ improvement compared with the result shown in the second column. We can clearly see that by involving all the relays in the message transmission process, it is possible to notably improve the DoF performance. To the best of our knowledge, this paper presents the highest achievable sum DoF known by far in single-antenna multi-user multi-hop networks when the within-layer interference, backward interference, and relay self-interference are not negligible so that half-duplex relaying has to be applied. Therefore, the results will be able to serve as new lower bounds for the networks' available DoF. In what follows, we will elaborate the IA design and DoF analysis for the CSR scheme, to prove Propositions 1 and 2.

III. Interference Alignment Construction for the Equivalent Network

Consider the case that the ith layer ($i \in \{1, 2, \cdots, N\}$) of relays is divided into a cluster $\mathcal{R}_{i,1}$ with $|\mathcal{R}_{i,1}| = r_i$ terminals ($2 \le r_i \le K_i - 2$) and a cluster $\mathcal{R}_{i,2}$ with $|\mathcal{R}_{i,2}| = K_i - r_i$ terminals. From Section II-B it can be seen that, during any of the $L + N$ time intervals the CSR transmission process converts the original multi-hop network into the single-hop primary equivalent network displayed in Fig. 2(b). In this section, we focus only on *a single time interval* and present the method to efficiently deliver information in this equivalent network. The actual achievable sum DoF in the original $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network will be analyzed in Section IV. The transmission strategy considered here is based on the channel-extension based IA technique. The linear IA filters construction is in principle similar to those approaches presented in [9] and [18]. But the differences are also notable. First, [9] considers an X channel with a single pair of transmitter and receiver clusters. No interference is generated from external terminals. But in the network shown in Fig. 2(b), there are multiple pairs of transmitter and receiver clusters. And the reception of each receiver is potentially affected by the signals from one or two unintended transmitter clusters. This fact leads to much more interference signals to be dealt with.

More importantly, the network shown in Fig. 2(b) is in fact generated from the original $(N + 1)$-hop $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network. The transmitters in the clusters $\mathcal{S}_2$, $\mathcal{S}_3$, $\cdots$, $\mathcal{S}_{N+1}$ are actually *relay terminals*. Considering a particular transmitter, during any time interval, the sub-messages that it can transmit must be those it received in the past time interval. In addition, since the sources always send out $M_s M_d I$ sub-messages to the first layer of relays in each time interval, the number of sub-messages that can be forwarded by all the relays in any transmitting relay cluster must be identical to this value. This means, in the equivalent network, the number of sub-messages to be delivered between every transmitter-receiver cluster pair must be the same as $M_s M_d I$. The IA design would be constructed subject to these constraints. This is again different from the methods considered in typical interference channels and X channels.

Further, in [18] the two relay clusters in each relay layer have the same size. Now this requirement is withdrawn. Certain adaptations of the IA construction will be adopted. The number of sub-messages that can be successfully delivered and the required channel usage must be carefully decided. These cause the performance analysis to be much more involved. It is not possible to directly conjecture the new achievable sum DoF results from [18].

In the following parts, we will present two approaches to carry out the IA design for the primary equivalent network. For the first method, we construct beamforming matrices at the terminals within $\mathcal{S}_1$, $\mathcal{S}_2$, $\cdots$, $\mathcal{S}_{N+1}$. The objective is that at each terminal within $\mathcal{D}_1$, $\mathcal{D}_2$, $\cdots$, $\mathcal{D}_{N+1}$, all the interference signals are aligned together in a compact receive subspace, so that they can be eliminated together by proper zero-forcing filter matrices. This strategy is termed *primary equivalent network IA construction* and is elaborated as follows.

A. Primary Equivalent Network IA Construction

We treat each cluster pair $\mathcal{S}_i$ and $\mathcal{D}_i$ ($i \in \{1, 2, \cdots, N + 1\}$) in the primary equivalent network as a wireless X channel: Every transmitter in $\mathcal{S}_i$ evenly divides its sub-messages into $|\mathcal{D}_i|$ fractions and sends one fraction to every receiver in $\mathcal{D}_i$. In order to guarantee that each transmitter transmits an integer number of sub-messages to each receiver, we set the total number of sub-messages to be delivered between each transmitter-receiver cluster pair as $\prod_{j=1}^{N+1} |\mathcal{S}_j||\mathcal{D}_j| c^{\Gamma}$, in which c and Γ are integer constants to be decided later. If we define integer $B_i = \frac{\prod_{j=1}^{N+1} |\mathcal{S}_j||\mathcal{D}_j|}{|\mathcal{S}_i||\mathcal{D}_i|}$, then every transmitter in $\mathcal{S}_i$ intends to send $B_i c^{\Gamma}$

sub-messages to every receiver in $\mathcal{D}_i$.[4] That is, a transmitter in $\mathcal{S}_i$ intends to send a total of $B_i|\mathcal{D}_i|c^\Gamma$ independent sub-messages, and each receiver in $\mathcal{D}_i$ expects a total of $|\mathcal{S}_i|B_ic^\Gamma$ independent sub-messages. The message transmission of these transmitter-receiver clusters may interfere each other. As illustrated in Fig. 2(b), for $i \in \{1, 2, \cdots, N-1\}$, the reception of terminals in $\mathcal{D}_i$ experiences *inter-cluster interference* generated by $\mathcal{S}_{i+1}$ and $\mathcal{S}_{i+2}$. The reception of terminals in $\mathcal{D}_N$ is interfered by the transmission of one unintended cluster $\mathcal{S}_{N+1}$, and receivers in $\mathcal{D}_{N+1}$ do not experience inter-cluster interference.

Let F denote the number of unit channels that is required to complete such message transmission. We can use the $F \times F$ diagonal matrix $\mathbf{H}_{q,p}^{[i_2,i_1]}$ ($i_1, i_2 \in \{1, 2, \cdots, N+1\}$, $p \in \{1, 2, \cdots, |\mathcal{S}_{i_1}|\}$, $q \in \{1, 2, \cdots, |\mathcal{D}_{i_2}|\}$) to denote the channel matrix between the pth transmitter in $\mathcal{S}_{i_1}$ and the qth receiver in $\mathcal{D}_{i_2}$. The B_ic^Γ unit-DoF sub-messages transmitted by the pth transmitter in $\mathcal{S}_i$ to the qth receiver in $\mathcal{D}_i$ are represented by B_i different $c^\Gamma \times 1$ vectors $\mathbf{x}_{p,[(q-1)B_i+1]}^{[i]}$, $\mathbf{x}_{p,[(q-1)B_i+2]}^{[i]}$, $\cdots$, $\mathbf{x}_{p,[qB_i]}^{[i]}$, each element of which represents a Gaussian codeword with rate $\log_2\rho + o(\log_2\rho)$. Their transmitter-side beamforming matrices are respectively denoted by $F \times c^\Gamma$ matrices $\mathbf{V}_{p,[(q-1)B_i+1]}^{[i]}$, $\mathbf{V}_{p,[(q-1)B_i+2]}^{[i]}$, $\cdots$, $\mathbf{V}_{p,[qB_i]}^{[i]}$, and receiver-side zero-forcing filter matrices are denoted by $c^\Gamma \times F$ matrices $\mathbf{U}_{p,[(q-1)B_i+1]}^{[i]}$, $\mathbf{U}_{p,[(q-1)B_i+2]}^{[i]}$, $\cdots$, $\mathbf{U}_{p,[qB_i]}^{[i]}$.

Let $\tilde{B} = \max\{B_1|\mathcal{D}_1|, B_2|\mathcal{D}_2|, \cdots, B_{N+1}|\mathcal{D}_{N+1}|\}$. To facilitate presentation, we create a *virtual* transmitter cluster $\mathcal{S}_0$ with only one terminal. This terminal is assumed to broadcast $\tilde{B}(c+1)^\Gamma$ *dummy codewords*, denoted by $\tilde{B}$ different $(c+1)^\Gamma \times 1$ vectors $\mathbf{x}_{0,[1]}^{[0]}$, $\mathbf{x}_{0,[2]}^{[0]}$, $\cdots$, $\mathbf{x}_{0,[\tilde{B}]}^{[0]}$, using $F \times (c+1)^\Gamma$ beamforming matrices $\mathbf{V}_{0,[1]}^{[0]}$, $\mathbf{V}_{0,[2]}^{[0]}$, $\cdots$, $\mathbf{V}_{0,[\tilde{B}]}^{[0]}$, respectively. The channel matrix between this terminal and the qth receiver in $\mathcal{D}_i$ is represented by an $F \times F$ diagonal matrix $\mathbf{H}_q^{[i,0]}$, the diagonal elements of which are generated following the same distribution as those in $\mathbf{H}_{q,p}^{[i_2,i_1]}$. Our IA design would first take the dummy codewords sent from this virtual terminal into account. In the end, the receivers can cancel them so that having any of these codewords in the desired receive subspace would not affect the actual capability of decoding the desired signals.

We denote the received signal at the qth receiver in $\mathcal{D}_i$ ($i \in \{1, 2, \cdots, N+1\}$, $q \in \{1, 2, \cdots, |\mathcal{D}_i|\}$) by an $F \times 1$ vector $\mathbf{y}_q^{[i]}$. It can be expressed as a general form:

$$\begin{aligned}
\mathbf{y}_q^{[i]} = & \sum_{p=1}^{|\mathcal{S}_i|}\mathbf{H}_{q,p}^{[i,i]}\left(\sum_{\kappa=(q-1)B_i+1}^{qB_i}\mathbf{V}_{p,[\kappa]}^{[i]}\mathbf{x}_{p,[\kappa]}^{[i]}\right) + \sum_{p=1}^{|\mathcal{S}_i|}\mathbf{H}_{q,p}^{[i,i]}\left(\sum_{\kappa=1}^{(q-1)B_i}\mathbf{V}_{p,[\kappa]}^{[i]}\mathbf{x}_{p,[\kappa]}^{[i]} + \sum_{\kappa=qB_i+1}^{B_i|\mathcal{D}_i|}\mathbf{V}_{p,[\kappa]}^{[i]}\mathbf{x}_{p,[\kappa]}^{[i]}\right) \\
& + \sum_{p=1}^{|\mathcal{S}_{i+1}|}\mathbf{H}_{q,p}^{[i,i+1]}\left(\sum_{\kappa=1}^{B_{i+1}|\mathcal{D}_{i+1}|}\mathbf{V}_{p,[\kappa]}^{[i+1]}\mathbf{x}_{p,[\kappa]}^{[i+1]}\right) + \sum_{p=1}^{|\mathcal{S}_{i+2}|}\mathbf{H}_{q,p}^{[i,i+2]}\left(\sum_{\kappa=1}^{B_{i+2}|\mathcal{D}_{i+2}|}\mathbf{V}_{p,[\kappa]}^{[i+2]}\mathbf{x}_{p,[\kappa]}^{[i+2]}\right) \\
& + \mathbf{H}_q^{[i,0]}\left(\sum_{\kappa=1}^{\max_{s\in\{0,1,2\}}\{B_{i+s}|\mathcal{D}_{i+s}|\}}\mathbf{V}_{0,[\kappa]}^{[0]}\mathbf{x}_{0,[\kappa]}^{[0]}\right) + \mathbf{z}_q^{[i]}.
\end{aligned} \tag{12}$$

The first term on the right hand side (RHS) is the $|\mathcal{S}_i|B_ic^\Gamma$ desired sub-messages, the second term denotes the undesired interference signals transmitted from terminals within $\mathcal{S}_i$ (i.e., the forward interference in the original network), the third term is the inter-cluster interference from $\mathcal{S}_{i+1}$ (i.e., the within-layer interference, present only at terminals in $\mathcal{D}_1$, $\mathcal{D}_2$, $\cdots$, $\mathcal{D}_N$), the fourth term is the inter-cluster interference

[4] This setup means that, in the original $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network, during any time interval the M_s sources transmit a total of M_sM_dI independent sub-messages to the first layer of relays, where $I = \frac{\prod_{j=1}^{N+1}|\mathcal{S}_j||\mathcal{D}_j|c^\Gamma}{M_sM_d} = \prod_{j=1}^{N} r_j(K_j - r_j)c^\Gamma$ and r_j is the number of relays in $\mathcal{R}_{j,1}$. Assume that in one particular time interval, $\mathcal{R}_{i,1}$ is operating in the listening mode so that it is represented by $\mathcal{D}_i$ in the primary equivalent network. Then a relay within $\mathcal{R}_{i,1}$ would desire a total of $\frac{\prod_{j=1}^{N+1}|\mathcal{S}_j||\mathcal{D}_j|c^\Gamma}{|\mathcal{D}_i|} = \frac{M_sM_d\prod_{j=1}^{N} r_j(K_j-r_j)c^\Gamma}{|\mathcal{R}_{i,1}|}$ sub-messages from $\mathcal{R}_{i-1,2}$ at its preceding layer (i.e., $\mathcal{S}_i$ in the primary equivalent network). In the next time interval, $\mathcal{R}_{i,1}$ becomes a forwarding cluster (i.e., $\mathcal{S}_{i+1}$ in the equivalent network). As mentioned above, any terminal within $\mathcal{S}_{i+1}$ intends to send a total of $\frac{\prod_{j=1}^{N+1}|\mathcal{S}_j||\mathcal{D}_j|c^\Gamma}{|\mathcal{S}_{i+1}|}$ sub-messages to $\mathcal{D}_{i+1}$. Since now $\frac{\prod_{j=1}^{N+1}|\mathcal{S}_j||\mathcal{D}_j|c^\Gamma}{|\mathcal{S}_{i+1}|} = \frac{M_sM_d\prod_{j=1}^{N} r_j(K_j-r_j)c^\Gamma}{|\mathcal{R}_{i,1}|}$, the number of sub-messages a relay would forward is exactly the same as that it received previously.

from $\mathcal{S}_{i+2}$ (i.e., the backward interference, present only at terminals in $\mathcal{D}_1$, $\mathcal{D}_2$, $\cdots$, $\mathcal{D}_{N-1}$), the fifth term is the dummy interference sent from $\mathcal{S}_0$, and $\mathbf{z}_q^{[i]}$ is unit-power additive white Gaussian noise (AWGN). Our IA construction targets aligning all the undesired interference signals at each receiver to the subspace decided by the dummy interference.

First, for every $\kappa \in \{1, 2, \cdots, (q-1)B_i, qB_i+1, qB_i+2, \cdots, |\mathcal{D}_i|B_i\}$, we intend to align the undesired signals $\mathbf{x}_{1,[\kappa]}^{[i]}$, $\mathbf{x}_{2,[\kappa]}^{[i]}$, $\cdots$, $\mathbf{x}_{|\mathcal{S}_i|,[\kappa]}^{[i]}$, appeared in the second term of (12), in the $(c+1)^\Gamma$-dimensional subspace decided by $\mathbf{H}_q^{[i,0]}\mathbf{V}_{0,[\kappa]}^{[0]}$. This means that, for any $i \in \{1, 2, \cdots, N+1\}$:

$$\text{span}\left(\mathbf{H}_{q,p}^{[i,i]}\mathbf{V}_{p,[\kappa]}^{[i]}\right) \subset \text{span}\left(\mathbf{H}_q^{[i,0]}\mathbf{V}_{0,[\kappa]}^{[0]}\right), \quad \forall p \in \{1, 2, \cdots, |\mathcal{S}_i|\}. \tag{13}$$

In addition, for every $\kappa \in \{1, 2, \cdots, |\mathcal{D}_{i+1}|B_{i+1}\}$, we aim to align the inter-cluster interference signals $\mathbf{x}_{1,[\kappa]}^{[i+1]}$, $\mathbf{x}_{2,[\kappa]}^{[i+1]}$, $\cdots$, $\mathbf{x}_{|\mathcal{S}_{i+1}|,[\kappa]}^{[i+1]}$, appeared in the third term of (12), to the $(c+1)^\Gamma$-dimensional subspace decided by $\mathbf{H}_q^{[i,0]}\mathbf{V}_{0,[\kappa]}^{[0]}$. Equivalently, for any $i \in \{1, 2, \cdots, N\}$,

$$\text{span}\left(\mathbf{H}_{q,p}^{[i,i+1]}\mathbf{V}_{p,[\kappa]}^{[i+1]}\right) \subset \text{span}\left(\mathbf{H}_q^{[i,0]}\mathbf{V}_{0,[\kappa]}^{[0]}\right), \quad \forall p \in \{1, 2, \cdots, |\mathcal{S}_{i+1}|\}. \tag{14}$$

Finally, for every $\kappa \in \{1, 2, \cdots, |\mathcal{D}_{i+2}|B_{i+2}\}$, we align the inter-cluster interference signals $\mathbf{x}_{1,[\kappa]}^{[i+2]}$, $\mathbf{x}_{2,[\kappa]}^{[i+2]}$, $\cdots$, $\mathbf{x}_{|\mathcal{S}_{i+2}|,[\kappa]}^{[i+2]}$, appeared in the fourth term of (12), to the $(c+1)^\Gamma$-dimensional subspace decided by $\mathbf{H}_q^{[i,0]}\mathbf{V}_{0,[\kappa]}^{[0]}$. That is, for any $i \in \{1, 2, \cdots, N-1\}$, let

$$\text{span}\left(\mathbf{H}_{q,p}^{[i,i+2]}\mathbf{V}_{p,[\kappa]}^{[i+2]}\right) \subset \text{span}\left(\mathbf{H}_q^{[i,0]}\mathbf{V}_{0,[\kappa]}^{[0]}\right), \quad \forall p \in \{1, 2, \cdots, |\mathcal{S}_{i+2}|\}. \tag{15}$$

It can be seen that, for each receiver in $\mathcal{D}_i$ ($i \in \{1, 2, \cdots, N-1\}$), in order to eliminate all interference and fully recover the $|\mathcal{S}_i|B_ic^\Gamma$ desired Gaussian codewords, the receive space should have no less than $|\mathcal{S}_i|B_ic^\Gamma + \max\{|\mathcal{D}_i|B_i, |\mathcal{D}_{i+1}|B_{i+1}, |\mathcal{D}_{i+2}|B_{i+2}\}(c+1)^\Gamma$ dimensions. For each receiver in $\mathcal{D}_N$, the receive space should have dimensions no less than $|\mathcal{S}_N|B_Nc^\Gamma + \max\{|\mathcal{D}_N|B_N, |\mathcal{D}_{N+1}|B_{N+1}\}(c+1)^\Gamma$. And for each receiver in $\mathcal{D}_{N+1}$, the number of receive space dimensions should be no less than $|\mathcal{S}_{N+1}|B_{N+1}c^\Gamma + (|\mathcal{D}_{N+1}|-1)B_{N+1}(c+1)^\Gamma$. This means that F must be chosen to satisfy all these conditions. The minimum value of F is thus

$$\begin{aligned} F \;=\; \max\Big\{ & |\mathcal{S}_1|B_1c^\Gamma + \max\big\{|\mathcal{D}_1|B_1, |\mathcal{D}_2|B_2, |\mathcal{D}_3|B_3\big\}(c+1)^\Gamma, \cdots, \\ & |\mathcal{S}_{N-1}|B_{N-1}c^\Gamma + \max\big\{|\mathcal{D}_{N-1}|B_{N-1}, |\mathcal{D}_N|B_N, |\mathcal{D}_{N+1}|B_{N+1}\big\}(c+1)^\Gamma, \\ & |\mathcal{S}_N|B_Nc^\Gamma + \max\big\{|\mathcal{D}_N|B_N, |\mathcal{D}_{N+1}|B_{N+1}\big\}(c+1)^\Gamma, \\ & |\mathcal{S}_{N+1}|B_{N+1}c^\Gamma + (|\mathcal{D}_{N+1}|-1)B_{N+1}(c+1)^\Gamma\Big\}. \end{aligned} \tag{16}$$

To construct the beamforming matrices at the transmitters to satisfy the conditions (13)-(15), for all $i \in \{1, 2, \cdots, N+1\}$ and every $\kappa \in \{1, 2, \cdots, |\mathcal{D}_i|B_i\}$, we can choose

$$\mathbf{V}_{1,[\kappa]}^{[i]} = \mathbf{V}_{2,[\kappa]}^{[i]} = \cdots = \mathbf{V}_{|\mathcal{S}_i|,[\kappa]}^{[i]} = \mathbf{V}_{[\kappa]}. \tag{17}$$

Following [18], it can be shown that, for each κ we can use (13)-(15) to formulate a total of $\Gamma = \sum_{i=1}^{N+1}|\mathcal{S}_i|\left(|\mathcal{D}_i|-1\right) + \sum_{i=1}^{N}|\mathcal{S}_{i+1}||\mathcal{D}_i| + \sum_{i=1}^{N-1}|\mathcal{S}_{i+2}||\mathcal{D}_i|$ different relations to be satisfied:

$$\text{span}\left(\mathbf{G}_{\gamma,[\kappa]}\mathbf{V}_{[\kappa]}\right) \subset \text{span}\left(\mathbf{V}_{0,[\kappa]}^{[0]}\right), \ \gamma \in \{1, 2, \cdots, \Gamma\}, \tag{18}$$

where each $\mathbf{G}_{\gamma,[\kappa]}$ is the product of two $F \times F$ diagonal matrices with independently-distributed and bounded diagonal elements. Randomly generate an $F \times 1$ vector $\mathbf{w}_{[\kappa]}$. Let [9]

$$\mathbf{V}_{[\kappa]} = \bigg\{\Big(\prod_{\gamma=1}^{\Gamma}\left(\mathbf{G}_{\gamma,[\kappa]}\right)^{a_\gamma}\Big)\mathbf{w}_{[\kappa]} : (a_1, \cdots, a_\Gamma) \in \{1, \cdots, c\}^\Gamma\bigg\},$$

$$\mathbf{V}_{0,[\kappa]}^{[0]} = \left\{ \Big(\prod_{\gamma=1}^{\Gamma} \left(\mathbf{G}_{\gamma,[\kappa]}\right)^{a_\gamma} \Big) \mathbf{w}_{[\kappa]} : (a_1, \cdots, a_\Gamma) \in \{1, \cdots, c+1\}^\Gamma \right\},$$

where the RHS represents the set of column vectors that form the corresponding left hand side (LHS) matrices. One can show that these matrices can satisfy the conditions (13)-(15). Then the beamforming matrices design is completed. Note that $\mathbf{V}_{[\kappa]}$ is calculated via the product of a diagonal matrix $\prod_{\gamma=1}^{\Gamma} \left(\mathbf{G}_{\gamma,[\kappa]}\right)^{a_\gamma}$ and $\mathbf{w}_{[\kappa]}$. Since $\prod_{\gamma=1}^{\Gamma} \left(\mathbf{G}_{\gamma,[\kappa]}\right)^{a_\gamma}$ is generated from the diagonal channel matrices, after the vectors $\mathbf{w}_{[\kappa]}$ are decided, the fth ($f \in \{1, 2, \cdots, F\}$) element of each column of $\mathbf{V}_{[\kappa]}$ is related to only the fth diagonal elements of the network's channel matrices. The same fact holds for $\mathbf{V}_{0,[\kappa]}^{[0]}$. Causal channel knowledge suffices to guarantee the transmitter-side beamforming design, in both time-varying and frequency-selective fading environments [9].

Furthermore, using a similar approach as that presented in [9], [18], we can prove that at the qth receiver in $\mathcal{D}_i$, with probability one, the $|\mathcal{S}_i|$ different $B_i c^\Gamma$-dimensional subspaces for the $B_i|\mathcal{S}_i|c^\Gamma$ desired codewords are independent to that for the aligned interference signals and also independent to each other. A linear zero-forcing filter suffices to eliminate interference and recover the desired sub-messages. Mathematically, these mean

$$\begin{aligned} \text{rank}\left(\mathbf{U}_{p,[\kappa]}^{[i]} \mathbf{H}_{q,p}^{[i,i]} \mathbf{V}_{p,[\kappa]}^{[i]}\right) &= c^\Gamma, && (19)\\ \mathbf{U}_{p,[\kappa]}^{[i]} \mathbf{H}_{q,p}^{[i,i]} \mathbf{V}_{p,[\kappa']}^{[i]} &= \mathbf{O}, && (20)\\ \mathbf{U}_{p,[\kappa]}^{[i]} \mathbf{H}_{q,p_\pi}^{[i,\pi]} \mathbf{V}_{p_\pi,[\kappa_\pi]}^{[\pi]} &= \mathbf{O}, && (21) \end{aligned}$$

$\forall p \in \{1, 2, \cdots, |\mathcal{S}_i|\}$, $\kappa \in \{(q-1)B_i+1, (q-1)B_i+2, \cdots, qB_i\}$, $\kappa' \in \{1, 2, \cdots, \kappa-1, \kappa+1, \cdots, B_i|\mathcal{D}_i|\}$, $\pi \in \{i, i+1, i+2\}$ when $i \in \{1, 2, \cdots, N-1\}$, $\pi \in \{i, i+1\}$ when $i = N$, $\pi = i$ when $i = N+1$, $p_\pi \in \{1, 2, \cdots, |\mathcal{S}_\pi|\}$, $p_i \neq p$, and $\kappa_\pi \in \{1, 2, \cdots, B_\pi|\mathcal{D}_\pi|\}$.

In summary, for the single-hop primary equivalent network displayed in Fig. 2(b), across F time or frequency slots (which can be calculated by equation (16)), each transmitter cluster $\mathcal{S}_i$ can successfully deliver a total of $M_s M_d I = \prod_{j=1}^{N+1} |\mathcal{S}_j||\mathcal{D}_j|c^\Gamma = M_s M_d \prod_{j=1}^{N} r_j(K_j - r_j)c^\Gamma$ unit-DoF sub-messages to its receiver cluster $\mathcal{D}_i$. This result will be used in Section IV to analyze the achievable sum DoF of the CSR scheme applied in the $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network.

B. Dual Equivalent Network IA Construction

In this subsection, we show that a different IA design can be carried out in the frequency-selective fading environment. Let us for now assume that terminals within each cluster $\mathcal{D}_i$ ($i \in \{1, 2, \cdots, N+1\}$) attempt to send $\prod_{j=1}^{N+1} |\mathcal{S}_j||\mathcal{D}_j|c^\Gamma$ sub-messages to terminals within $\mathcal{S}_i$. In other words, the transmission direction of the primary equivalent network is *reversed*. The resulting network is termed *dual equivalent network*. Fig. 2(c) illustrates the dual equivalent network corresponding to Fig. 2(b). The idea behind this new IA construction method is to first construct beamforming matrices at the terminals within $\mathcal{D}_1$, $\mathcal{D}_2$, $\cdots$, $\mathcal{D}_{N+1}$ such that zero-forcing filter matrices at terminals within $\mathcal{S}_1$, $\mathcal{S}_2$, $\cdots$, $\mathcal{S}_{N+1}$ can be established to eliminate the aligned interference signals. Then we make use of the channel reciprocal property so that actual data transmission in the primary equivalent network allows the transmitters within $\mathcal{S}_1$, $\mathcal{S}_2$, $\cdots$, $\mathcal{S}_{N+1}$ to use these zero-forcing filter matrices to beamform their signals. At the receivers in $\mathcal{D}_1$, $\mathcal{D}_2$, $\cdots$, $\mathcal{D}_{N+1}$, the beamforming matrices designed in the dual equivalent network would be applied to eliminate interference. This method is termed *dual equivalent network IA construction*. Since it requires channel knowledge of all the $L+N$ time intervals to design the transceiver filters, it is more applicable in systems that channel extension is realized in the frequency domain, i.e., a frequency-selective fading environment.

To simplify presentation, we use notation $\overleftarrow{\mathcal{S}}_i$ to replace $\mathcal{D}_{N+2-i}$, and use $\overleftarrow{\mathcal{D}}_i$ to replace $\mathcal{S}_{N+2-i}$ for all $i \in \{1, 2, \cdots, N+1\}$. It is easy to see from Fig. 2(b) and Fig. 2(c) that the dual equivalent network exhibits exactly the same structure as the primary equivalent network. More specifically, it contains $N+1$

pairs of transmitter-receiver clusters. Every transmitter within $\overleftarrow{\mathcal{S}}_i$ intends to transmit $\frac{\prod_{j=1}^{N+1}|\overleftarrow{\mathcal{S}}_j||\overleftarrow{\mathcal{D}}_j|c^{\Gamma}}{|\overleftarrow{\mathcal{S}}_i||\overleftarrow{\mathcal{D}}_i|}$ unit-DoF sub-messages to every receiver in $\overleftarrow{\mathcal{D}}_i$. The network structure is asymmetric: The reception of $\overleftarrow{\mathcal{D}}_i$ ($i \in \{1, 2, \cdots, N-1\}$) is affected by inter-cluster interference from two unintended clusters $\overleftarrow{\mathcal{S}}_{i+1}$ and $\overleftarrow{\mathcal{S}}_{i+2}$, the reception of $\overleftarrow{\mathcal{D}}_N$ is interfered by one unintended cluster $\overleftarrow{\mathcal{S}}_{N+1}$, but $\overleftarrow{\mathcal{D}}_{N+1}$ does not experience inter-cluster interference.

We can follow the IA design strategy presented in the above subsection to construct the transmitter-side beamforming and receiver-side zero-forcing matrices. Let $\overleftarrow{F}$ be the number of unit channels required to complete transmission. Use the $\overleftarrow{F} \times \overleftarrow{F}$ diagonal matrix $\overleftarrow{\mathbf{H}}_{p,q}^{[i_1,i_2]}$ to denote the channel matrix between the qth transmitter in $\overleftarrow{\mathcal{S}}_{i_2}$ and the pth receiver in $\overleftarrow{\mathcal{D}}_{i_1}$. Clearly, due to the channel reciprocity, we have $\overleftarrow{\mathbf{H}}_{p,q}^{[i_1,i_2]} = \left(\mathbf{H}_{q,p}^{[N+2-i_2,N+2-i_1]}\right)^T$, where $\mathbf{H}_{q,p}^{[N+2-i_2,N+2-i_1]}$ is the channel matrix between the pth transmitter in $\mathcal{S}_{N+2-i_1}$ and the qth receiver in $\mathcal{D}_{N+2-i_2}$ in the primary equivalent network. Set $\overleftarrow{B}_i = \frac{\prod_{j=1}^{N+1}|\overleftarrow{\mathcal{S}}_j||\overleftarrow{\mathcal{D}}_j|}{|\overleftarrow{\mathcal{S}}_i||\overleftarrow{\mathcal{D}}_i|}$ for $i \in \{1, 2, \cdots, N+1\}$. We can denote the $\overleftarrow{B}_i c^{\Gamma}$ Gaussian codewords that the qth transmitter in $\overleftarrow{\mathcal{S}}_i$ desires to send to the pth receiver in $\overleftarrow{\mathcal{D}}_i$ by $\overleftarrow{B}_i$ different $c^{\Gamma} \times 1$ vectors $\overleftarrow{\mathbf{x}}_{q,[(p-1)\overleftarrow{B}_i+1]}^{[i]}$, $\overleftarrow{\mathbf{x}}_{q,[(p-1)\overleftarrow{B}_i+2]}^{[i]}$, $\cdots$, $\overleftarrow{\mathbf{x}}_{q,[p\overleftarrow{B}_i]}^{[i]}$. Their beamforming matrices are denoted by $\overleftarrow{F} \times c^{\Gamma}$ matrices $\overleftarrow{\mathbf{V}}_{q,[(p-1)\overleftarrow{B}_i+1]}^{[i]}$, $\overleftarrow{\mathbf{V}}_{q,[(p-1)\overleftarrow{B}_i+2]}^{[i]}$, $\cdots$, $\overleftarrow{\mathbf{V}}_{q,[p\overleftarrow{B}_i]}^{[i]}$ respectively. Similarly, the zero-forcing filter matrices applied at the pth receiver in $\overleftarrow{\mathcal{D}}_i$ are denoted by $c^{\Gamma} \times \overleftarrow{F}$ matrices $\overleftarrow{\mathbf{U}}_{q,[(p-1)\overleftarrow{B}_i+1]}^{[i]}$, $\overleftarrow{\mathbf{U}}_{q,[(p-1)\overleftarrow{B}_i+2]}^{[i]}$, $\cdots$, $\overleftarrow{\mathbf{U}}_{q,[p\overleftarrow{B}_i]}^{[i]}$ respectively.

Analogous to equation (16), let

$$\begin{aligned}
\overleftarrow{F} \;=\; \max\Big\{ & \left|\overleftarrow{\mathcal{S}}_1\right| \overleftarrow{B}_1 c^{\Gamma} + \max\left\{\left|\overleftarrow{\mathcal{D}}_1\right| \overleftarrow{B}_1, \left|\overleftarrow{\mathcal{D}}_2\right| \overleftarrow{B}_2, \left|\overleftarrow{\mathcal{D}}_3\right| \overleftarrow{B}_3\right\}(c+1)^{\Gamma}, \cdots, \\
& \left|\overleftarrow{\mathcal{S}}_{N-1}\right| \overleftarrow{B}_{N-1} c^{\Gamma} + \max\left\{\left|\overleftarrow{\mathcal{D}}_{N-1}\right| \overleftarrow{B}_{N-1}, \left|\overleftarrow{\mathcal{D}}_N\right| \overleftarrow{B}_N, \left|\overleftarrow{\mathcal{D}}_{N+1}\right| \overleftarrow{B}_{N+1}\right\}(c+1)^{\Gamma}, \\
& \left|\overleftarrow{\mathcal{S}}_N\right| \overleftarrow{B}_N c^{\Gamma} + \max\left\{\left|\overleftarrow{\mathcal{D}}_N\right| \overleftarrow{B}_N, \left|\overleftarrow{\mathcal{D}}_{N+1}\right| \overleftarrow{B}_{N+1}\right\}(c+1)^{\Gamma}, \\
& \left|\overleftarrow{\mathcal{S}}_{N+1}\right| \overleftarrow{B}_{N+1} c^{\Gamma} + \left(\left|\overleftarrow{\mathcal{D}}_{N+1}\right| - 1\right) \overleftarrow{B}_{N+1}(c+1)^{\Gamma}\Big\},
\end{aligned} \tag{22}$$

and $\Gamma = \sum_{i=1}^{N+1}\left|\overleftarrow{\mathcal{S}}_i\right|\left(\left|\overleftarrow{\mathcal{D}}_i\right| - 1\right) + \sum_{i=1}^{N}\left|\overleftarrow{\mathcal{S}}_{i+1}\right|\left|\overleftarrow{\mathcal{D}}_i\right| + \sum_{i=1}^{N-1}\left|\overleftarrow{\mathcal{S}}_{i+2}\right|\left|\overleftarrow{\mathcal{D}}_i\right|$. Again, with probability one we can construct the beamforming and zero-forcing matrices to satisfy

$$\begin{aligned}
\operatorname{rank}\left(\overleftarrow{\mathbf{U}}_{q,[\kappa]}^{[i]} \overleftarrow{\mathbf{H}}_{p,q}^{[i,i]} \overleftarrow{\mathbf{V}}_{q,[\kappa]}^{[i]}\right) &= c^{\Gamma}, & (23)\\
\overleftarrow{\mathbf{U}}_{q,[\kappa]}^{[i]} \overleftarrow{\mathbf{H}}_{p,q}^{[i,i]} \overleftarrow{\mathbf{V}}_{q,[\kappa']}^{[i]} &= \mathbf{O}, & (24)\\
\overleftarrow{\mathbf{U}}_{q,[\kappa]}^{[i]} \overleftarrow{\mathbf{H}}_{p,q_\pi}^{[i,\pi]} \overleftarrow{\mathbf{V}}_{q_\pi,[\kappa_\pi]}^{[\pi]} &= \mathbf{O}. & (25)
\end{aligned}$$

Given the beamforming and zero-forcing matrices designed for the dual equivalent network, we focus back on the transmission in the primary equivalent network. Set $\mathbf{V}_{p,[(q-1)B_i+j]}^{[i]} = \left(\overleftarrow{\mathbf{U}}_{q,[(p-1)\overleftarrow{B}_{N+2-i}+j]}^{[N+2-i]}\right)^T$ and $\mathbf{U}_{p,[(q-1)B_i+j]}^{[i]} = \left(\overleftarrow{\mathbf{V}}_{q,[(p-1)\overleftarrow{B}_{N+2-i}+j]}^{[N+2-i]}\right)^T$, $\forall i \in \{1, 2, \cdots, N+1\}$, $p \in \{1, 2, \cdots, |\mathcal{S}_i|\}$, $q \in \{1, 2, \cdots, |\mathcal{D}_i|\}$, $j \in \{1, 2, \cdots, B_i\}$. Since $\mathbf{H}_{q,p}^{[i_2,i_1]} = \left(\overleftarrow{\mathbf{H}}_{p,q}^{[N+2-i_1,N+2-i_2]}\right)^T$, it can be directly seen that the conditions (19)-(21) are also satisfied. Similar to the observation made in [9], IA is realized simultaneously for both directions in a reciprocal bi-directional network, with the same transceiver structure. Different from the case we discussed in Section III-A, now the design of each element of the primary equivalent network's transmitter-side beamforming matrices needs to know all the channel fading coefficients for the entire transmission duration. This implies non-causal channel knowledge for time-varying fading. The solution presented in this subsection is applicable in only the frequency-selective fading environment. In summary, in the primary equivalent network the transmission of $M_sM_dI = \prod_{j=1}^{N+1} |\mathcal{S}_j||\mathcal{D}_j|c^{\Gamma} = M_sM_d\prod_{j=1}^{N} r_j(K_j - r_j)c^{\Gamma}$ unit-DoF sub-messages between every pair of transmitter-receiver clusters can be finished across $\overleftarrow{F}$ unit frequency channels, where $\overleftarrow{F}$ is derived using (22).

IV. System achievable sum DoF analysis

Armed with the above results, we are ready to analyze the achievable sum DoF of applying the CSR scheme in the considered $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network.

A. Sum DoF in Time-Varying Fading Environment

Let us start from the time-varying fading environment, by using the results presented in Section III-A. Apply the CSR scheme in the $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network and choose the sizes of the clusters $\mathcal{R}_{1,1}$, $\cdots$, $\mathcal{R}_{N,1}$ to be r_1, $\cdots$, r_N, respectively. For an odd time interval, the primary equivalent network has $|\mathcal{D}_i| = r_i$ and $|\mathcal{S}_{i+1}| = K_i - r_i$ for $i \in \{1, 2, \cdots, N\}$. The transmitting terminals in each layer are able to successfully deliver $M_sM_dI = \prod_{j=1}^{N+1} |\mathcal{S}_j||\mathcal{D}_j|c^{\Gamma} = M_sM_d\prod_{j=1}^{N} r_j\left(K_j - r_j\right)c^{\Gamma}$ unit-DoF sub-messages to the receiving terminals in the next layer. The number of unit time slots required to complete the transmission in this interval is denoted by F_o, which can be calculated by substituting $|\mathcal{S}_1| = M_s$, $|\mathcal{D}_{N+1}| = M_d$, $|\mathcal{D}_i| = r_i$, and $|\mathcal{S}_{i+1}| = K_i - r_i$ into (16). Similarly, during an even time interval, substituting $|\mathcal{S}_1| = M_s$, $|\mathcal{D}_{N+1}| = M_d$, $|\mathcal{D}_i| = K_i - r_i$ and $|\mathcal{S}_{i+1}| = r_i$ into equation (16) leads to the necessary number of unit time slots to guarantee delivering the same number of $M_sM_dI = M_sM_d\prod_{j=1}^{N} r_j\left(K_j - r_j\right)c^{\Gamma}$ unit-DoF sub-messages in every hop. This value is denoted by F_e.

Therefore, following Section II-B, the total number of unit time slots consumed by the CSR scheme to complete the transmission of all the LM_sM_dI sub-messages from the sources to the destinations can be calculated as $\sum_{n=1}^{L+N} F_{[n]} = \lceil\frac{L+N}{2}\rceil F_o + \lfloor\frac{L+N}{2}\rfloor F_e$. The sum DoF that can be asymptotically achieved can be found by using equation (3) and letting $L \to \infty$ and $c \to \infty$. We denote this value by $d_{\text{CSR}(r_1,\cdots,r_N)}$, as a function of r_1, $\cdots$, r_N, the numbers of relays chosen in $\mathcal{R}_{1,1}$, $\cdots$, $\mathcal{R}_{N,1}$. Thus we have

$$\begin{aligned} d_{\text{CSR}(r_1,\cdots,r_N)} &= \frac{LM_sM_d\prod_{j=1}^{N} r_j\left(K_j - r_j\right)c^{\Gamma}}{\lceil\frac{L+N}{2}\rceil F_o + \lfloor\frac{L+N}{2}\rfloor F_e} \\ &\overset{(a)}{\approx} \frac{2M_sM_d\prod_{j=1}^{N} r_j\left(K_j - r_j\right)c^{\Gamma}}{F_o + F_e} \\ &\overset{(b)}{\approx} \frac{2}{F_1 + F_2}, \end{aligned} \tag{26}$$

where (a) follows from $\frac{\lceil\frac{L+N}{2}\rceil}{L} \approx \frac{\lfloor\frac{L+N}{2}\rfloor}{L} \approx \frac{1}{2}$ when $L \to \infty$, (b) follows from $\frac{c^{\Gamma}}{(c+1)^{\Gamma}} \approx 1$ when $c \to \infty$, and

$$\begin{aligned} F_1 = \max\Bigg\{ & \frac{1}{r_1} + \max\left\{\frac{1}{M_s}, \frac{1}{K_1 - r_1}, \frac{1}{K_2 - r_2}\right\}, \frac{1}{r_2} + \max\left\{\frac{1}{K_1 - r_1}, \frac{1}{K_2 - r_2}, \frac{1}{K_3 - r_3}\right\}, \\ & \cdots, \frac{1}{r_{N-1}} + \max\left\{\frac{1}{K_{N-2} - r_{N-2}}, \frac{1}{K_{N-1} - r_{N-1}}, \frac{1}{K_N - r_N}\right\}, \\ & \frac{1}{r_N} + \max\left\{\frac{1}{K_{N-1} - r_{N-1}}, \frac{1}{K_N - r_N}\right\}, \frac{K_N - r_N + M_d - 1}{(K_N - r_N)M_d}\Bigg\}, \end{aligned} \tag{27}$$

$$\begin{aligned} F_2 = \max\Bigg\{ & \frac{1}{K_1 - r_1} + \max\left\{\frac{1}{M_s}, \frac{1}{r_1}, \frac{1}{r_2}\right\}, \frac{1}{K_2 - r_2} + \max\left\{\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r_3}\right\}, \\ & \cdots, \frac{1}{K_{N-1} - r_{N-1}} + \max\left\{\frac{1}{r_{N-2}}, \frac{1}{r_{N-1}}, \frac{1}{r_N}\right\}, \\ & \frac{1}{K_N - r_N} + \max\left\{\frac{1}{r_{N-1}}, \frac{1}{r_N}\right\}, \frac{r_N + M_d - 1}{r_N M_d}\Bigg\}. \end{aligned} \tag{28}$$

To fully take advantage of the network's hardware resources (i.e., to make the best use of all relays), one should properly divide each relay layer in such a way that the achievable sum DoF is maximized. It means solving the following problem:

$$\begin{aligned} \underset{r_1,r_2,\cdots,r_N}{\text{maximize}} \quad & d_{\text{CSR}(r_1,\cdots,r_N)} = \frac{2}{F_1+F_2} \\ \text{s.t.} \quad & 2 \le r_i \le K_i - 2, \ \forall i \in \{1,2,\cdots,N\}. \end{aligned} \tag{29}$$

We use d_{CSR} to denote the result derived using (29). The analysis in the Appendix shows

$$d_{\text{CSR}} = \eta(M_s, K_1, \cdots, K_N, M_d), \tag{30}$$

where the function $\eta(\cdot)$ is defined by (4). This is the result presented in Proposition 1 and is achieved when we set $|\mathcal{R}_{i,1}| = r_i = \lfloor \frac{K_i}{2} \rfloor$ and $|\mathcal{R}_{i,2}| = K_i - r_i = \lceil \frac{K_i}{2} \rceil$ (or $|\mathcal{R}_{i,1}| = \lceil \frac{K_i}{2} \rceil$ and $|\mathcal{R}_{i,2}| = \lfloor \frac{K_i}{2} \rfloor$) for all $i \in \{1,2,\cdots,N\}$.

B. Sum DoF in Frequency-Selective Fading Environment

If the fading is frequency-selective, the results shown in the above subsection are directly applicable. Equation (30) can be adopted to calculate the achievable sum DoF of the CSR scheme. In addition, the dual equivalent network IA construction described in Section III-B provides an extra option for designing the beamforming and zero-forcing filter matrices. When the cluster sizes are chosen as $|\mathcal{R}_{i,1}| = r_i$ for $i \in \{1,2\cdots,N\}$, a total of $M_sM_dI = M_sM_d\prod_{j=1}^{N} r_j(K_j-r_j)c^{\Gamma}$ unit-DoF sub-messages can be delivered between two adjacent layers (i.e., in each hop) in each time interval. The number of unit frequency channels that such transmission requires can be calculated using (22). During odd time intervals, we have $\left|\overleftarrow{\mathcal{S}}_1\right| = M_d$, $\left|\overleftarrow{\mathcal{D}}_{N+1}\right| = M_s$, $\left|\overleftarrow{\mathcal{S}}_i\right| = r_{N+2-i}$, and $\left|\overleftarrow{\mathcal{D}}_{i-1}\right| = K_{N+2-i} - r_{N+2-i}$ for $i \in \{2,3,\cdots,N+1\}$. During even time intervals, $\left|\overleftarrow{\mathcal{S}}_1\right| = M_d$, $\left|\overleftarrow{\mathcal{D}}_{N+1}\right| = M_s$, $\left|\overleftarrow{\mathcal{S}}_i\right| = K_{N+2-i} - r_{N+2-i}$, and $\left|\overleftarrow{\mathcal{D}}_{i-1}\right| = r_{N+2-i}$. Substituting these values into (22) and letting $L \to \infty$ and $c \to \infty$, the asymptotically achievable sum DoF, $\overleftarrow{d}_{\text{CSR}(r_1,\cdots,r_N)}$, is

$$\begin{aligned} \overleftarrow{d}_{\text{CSR}(r_1,\cdots,r_N)} &= \frac{LM_sM_d\prod_{j=1}^{N} r_j\,(K_j-r_j)\,c^{\Gamma}}{\sum_{n=1}^{L+N} F_{[n]}} \\ &\approx \frac{2}{\overleftarrow{F}_1 + \overleftarrow{F}_2}, \end{aligned} \tag{31}$$

where

$$\begin{aligned} \overleftarrow{F}_1 = \max\Bigg\{ & \frac{1}{K_N - r_N} + \max\left\{\frac{1}{M_d}, \frac{1}{r_N}, \frac{1}{r_{N-1}}\right\}, \frac{1}{K_{N-1} - r_{N-1}} + \max\left\{\frac{1}{r_N}, \frac{1}{r_{N-1}}, \frac{1}{r_{N-2}}\right\}, \\ & \cdots, \frac{1}{K_2 - r_2} + \max\left\{\frac{1}{r_3}, \frac{1}{r_2}, \frac{1}{r_1}\right\}, \frac{1}{K_1 - r_1} + \max\left\{\frac{1}{r_2}, \frac{1}{r_1}\right\}, \frac{r_1 + M_s - 1}{r_1 M_s} \Bigg\}, \end{aligned} \tag{32}$$

$$\begin{aligned} \overleftarrow{F}_2 = \max\Bigg\{ & \frac{1}{r_N} + \max\left\{\frac{1}{M_d}, \frac{1}{K_N - r_N}, \frac{1}{K_{N-1} - r_{N-1}}\right\}, \\ & \frac{1}{r_{N-1}} + \max\left\{\frac{1}{K_N - r_N}, \frac{1}{K_{N-1} - r_{N-1}}, \frac{1}{K_{N-2} - r_{N-2}}\right\}, \\ & \cdots, \frac{1}{r_2} + \max\left\{\frac{1}{K_3 - r_3}, \frac{1}{K_2 - r_2}, \frac{1}{K_1 - r_1}\right\}, \\ & \frac{1}{r_1} + \max\left\{\frac{1}{K_2 - r_2}, \frac{1}{K_1 - r_1}\right\}, \frac{K_1 - r_1 + M_s - 1}{(K_1 - r_1)M_s} \Bigg\}. \end{aligned} \tag{33}$$

Again, to attain the highest achievable sum DoF, we solve:

$$\begin{aligned} \underset{r_1, r_2, \cdots, r_N}{\text{maximize}} \quad & \overleftarrow{d}_{\text{CSR}(r_1, \cdots, r_N)} = \frac{2}{\overleftarrow{F}_1 + \overleftarrow{F}_2} \\ \text{s.t.} \quad & 2 \leq r_i \leq K_i - 2, \ \forall i \in \{1, 2, \cdots, N\}. \end{aligned} \tag{34}$$

Following the analysis in the Appendix, we can see that the value of $\eta(M_d, K_N, \cdots, K_1, M_s)$ serves as the solution to (34). To achieve this result one should also choose $|\mathcal{R}_{i,1}| = r_i = \lfloor \frac{K_i}{2} \rfloor$ and $|\mathcal{R}_{i,2}| = K_i - r_i = \lceil \frac{K_i}{2} \rceil$ (or $|\mathcal{R}_{i,1}| = \lceil \frac{K_i}{2} \rceil$ and $|\mathcal{R}_{i,2}| = \lfloor \frac{K_i}{2} \rfloor$) for all $i \in \{1, 2, \cdots, N\}$.

Now we have two transmission strategies that can be adopted in an $\{M_s, K_1, \cdots, K_N, M_d\}^{(N+1)}$ network. Their achievable sum DoF are $\eta(M_s, K_1, \cdots, K_N, M_d)$ and $\eta(M_d, K_N, \cdots, K_1, M_s)$ respectively. Depending on the network topology, the scheme that provides better performance can be selected to carry out message transmission. This leads to the result in Proposition 2:

$$d_{\text{CSR}} = \max\Big\{\eta(M_s, K_1, \cdots, K_N, M_d), \eta(M_d, K_N, \cdots, K_1, M_s)\Big\}.$$

V. Conclusions

We have studied the potential communication performance limits in a class of single-antenna multi-user multi-hop networks, where the interference between relays located in the same layer and in adjacent layers cannot be neglected. In addition, self-interference cannot be mitigated if relays operate in the full-duplex fashion. We have considered deploying half-duplex DF relays in these networks and adopted a CSR transmission scheme to efficiently deliver information. Allowing the two relay clusters in each relay layer to contain different numbers of terminals, all available relays can be involved and properly clustered to participate in the message transmission process. Two IA construction solutions have been proposed to handle the multi-user interference issues. It has been shown that in the time-varying fading environment, the sum DoF achieved by previous works can be notably improved. When the fading is frequency-selective, the result can be even higher. These results thus can produce new lower bounds for the available DoF in the considered class of relay networks.

It is worth mentioning that the tightness of the lower bounds obtained in this paper is unknown. This is mainly because when all interference issues present in a general multi-user multi-hop network are taken into account, how to optimally carry out transmission design to eliminate their negative impacts is far from an easy task. The advantages of our CSR scheme stems from using multiple relays to create high-dimensional transmission space between the sources and destinations to align and cancel interference. Increasing the number of relays in each layer potentially improves the ability of combating interference and closes the gap between the lower bound and upper bound for the available DoF. However, if only a limited number of relays are deployed, at this point of time it is difficult to draw any conclusion regarding the optimality of the results presented in this paper. Whether these results can be further improved will be considered as interesting and meaningful future works.

Acknowledgment

This result is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 752797. It reflects only the authors' view and the Research Executive Agency and the Commission are not responsible for any use that may be made of the information it contains.

Appendix

A. *Proof of Equation (30)*

Solving (29) is equivalent to finding solutions to minimize $F_1 + F_2$. Define

$$\tilde{F} \;=\; \max\left\{ \max_{1 \leq i \leq N} \left\{ \frac{1}{r_i} + \frac{1}{K_i - r_i} \right\}, \max_{1 \leq i \leq N-1} \left\{ \frac{1}{r_i} + \frac{1}{K_{i+1} - r_{i+1}} \right\}, \max_{2 \leq i \leq N} \left\{ \frac{1}{r_i} + \frac{1}{K_{i-1} - r_{i-1}} \right\} \right\}.$$

It is not difficult to see that

$$\begin{aligned}\tilde{F} &\geq \max\left\{\frac{1}{\lfloor\frac{K_1}{2}\rfloor}+\frac{1}{\lceil\frac{K_1}{2}\rceil},\cdots,\frac{1}{\lfloor\frac{K_N}{2}\rfloor}+\frac{1}{\lceil\frac{K_N}{2}\rceil}\right\}\\ &= \frac{1}{\left\lfloor\frac{\min\{K_1,\cdots,K_N\}}{2}\right\rfloor}+\frac{1}{\left\lceil\frac{\min\{K_1,\cdots,K_N\}}{2}\right\rceil}.\end{aligned}$$

The lower bound is achievable when we choose $r_i = \lfloor\frac{K_i}{2}\rfloor$ for all $i \in \{1,2,\cdots,N\}$.

Consider equations (27) and (28). After some mathematical manipulations we can express F_1 and F_2 as $F_1 = \max\left\{\frac{1}{r_1}+\frac{1}{M_s},\tilde{F},\frac{K_N-r_N+M_d-1}{(K_N-r_N)M_d}\right\}$ and $F_2 = \max\left\{\frac{1}{K_1-r_1}+\frac{1}{M_s},\tilde{F},\frac{r_N+M_d-1}{r_N M_d}\right\}$. Therefore, we have

$$\begin{aligned}F_1+F_2 &= \max\Bigg\{\frac{1}{K_1-r_1}+\frac{1}{r_1}+\frac{2}{M_s},\max\left\{\frac{1}{r_1}+\frac{1}{M_s},\frac{1}{K_1-r_1}+\frac{1}{M_s}\right\}+\tilde{F},\\ &\quad \frac{1}{r_1}+\frac{1}{M_s}+\frac{r_N+M_d-1}{r_N M_d},2\tilde{F},\frac{1}{K_1-r_1}+\frac{1}{M_s}+\frac{K_N-r_N+M_d-1}{(K_N-r_N)M_d},\\ &\quad \max\left\{\frac{r_N+M_d-1}{r_N M_d},\frac{K_N-r_N+M_d-1}{(K_N-r_N)M_d}\right\}+\tilde{F},\\ &\quad \frac{K_N-r_N+M_d-1}{(K_N-r_N)M_d}+\frac{r_N+M_d-1}{r_N M_d}\Bigg\}.\end{aligned}$$

Since

$$\max\left\{\frac{r_i+M_d-1}{r_i M_d},\frac{K_i-r_i+M_d-1}{(K_i-r_i)M_d}\right\} \geq \frac{\lfloor\frac{K_i}{2}\rfloor+M_d-1}{\lfloor\frac{K_i}{2}\rfloor M_d},$$

i

$$\frac{r_i+M_d-1}{r_i M_d}+\frac{K_i-r_i+M_d-1}{(K_i-r_i)M_d} \geq \frac{\lfloor\frac{K_i}{2}\rfloor+M_d-1}{\lfloor\frac{K_i}{2}\rfloor M_d}+\frac{\lceil\frac{K_i}{2}\rceil+M_d-1}{\lceil\frac{K_i}{2}\rceil M_d},$$

$\frac{1}{K_i-r_i}+\frac{1}{r_i} \geq \frac{1}{\lfloor\frac{K_i}{2}\rfloor}+\frac{1}{\lceil\frac{K_i}{2}\rceil}$, and $\max\left\{\frac{1}{r_i},\frac{1}{K_i-r_i}\right\} \geq \frac{1}{\lfloor\frac{K_i}{2}\rfloor}$, it can be seen that the value

$$\begin{aligned}\max\Bigg\{&\frac{1}{\lfloor\frac{K_1}{2}\rfloor}+\frac{1}{\lceil\frac{K_1}{2}\rceil}+\frac{2}{M_s},\frac{1}{\lfloor\frac{K_1}{2}\rfloor}+\frac{1}{M_s}+\frac{\lfloor\frac{K_N}{2}\rfloor+M_d-1}{\lfloor\frac{K_N}{2}\rfloor M_d},\\ &\frac{1}{\lfloor\frac{K_1}{2}\rfloor}+\frac{1}{M_s}+\frac{1}{\left\lfloor\frac{\min\{K_1,\cdots,K_N\}}{2}\right\rfloor}+\frac{1}{\left\lceil\frac{\min\{K_1,\cdots,K_N\}}{2}\right\rceil},\\ &\frac{\lfloor\frac{K_N}{2}\rfloor+M_d-1}{\lfloor\frac{K_N}{2}\rfloor M_d}+\frac{1}{\left\lfloor\frac{\min\{K_1,\cdots,K_N\}}{2}\right\rfloor}+\frac{1}{\left\lceil\frac{\min\{K_1,\cdots,K_N\}}{2}\right\rceil},\\ &\frac{2}{\left\lfloor\frac{\min\{K_1,\cdots,K_N\}}{2}\right\rfloor}+\frac{2}{\left\lceil\frac{\min\{K_1,\cdots,K_N\}}{2}\right\rceil},\frac{\lfloor\frac{K_N}{2}\rfloor+M_d-1}{\lfloor\frac{K_N}{2}\rfloor M_d}+\frac{\lceil\frac{K_N}{2}\rceil+M_d-1}{\lceil\frac{K_N}{2}\rceil M_d}\Bigg\}\end{aligned}$$

serves as an achievable lower bound for F_1+F_2. The achievability is established when we set $r_i = \lfloor\frac{K_i}{2}\rfloor$ (or $r_i = \lceil\frac{K_i}{2}\rceil$) for all $i \in \{1,2,\cdots,N\}$. Using this lower bound, we can see that the maximized value of $d_{\text{CSR}(r_1,\cdots,r_N)}$, i.e., d_{CSR}, is expressed as (30). The analysis in Section IV-B is similar.

REFERENCES

[1] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[2] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[3] S. A. Jafar and S. Shamai, "Degrees of freedom region of the MIMO X channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 151–170, Jan. 2008.

[4] S. A. Jafar, *Interference Alignment - A New Look at Signal Dimensions in a Communication Network*. Now Publications, 2011.

[5] T. Gou and S. A. Jafar, "Degrees of freedom of the K user $M \times N$ MIMO interfrence channel," *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6040–6057, 2010.

[6] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Processing*, vol. 58, no. 9, pp. 4771–4782, 2010.

[7] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inform. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[8] H. Farhadi, C. Wang, and M. Skoglund, "Distributed transceiver design and power control for wireless MIMO interference networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1199–1212, Mar. 2015.

[9] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom of wireless X networks," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 3893–3908, Sep. 2009.

[10] ——, "Degrees of freedom of wireless networks with relays, feedback, cooperation, and full duplex operation," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2334–2344, May 2009.

[11] S.-W. Jeon, S.-Y. Chung, and S. A. Jafar, "Degrees of freedom region of a class of multisource Gaussian relay networks," *IEEE Trans. Inform. Theory*, vol. 57, no. 5, pp. 3032–3044, May 2011.

[12] T. Gou, S. A. Jafar, C. Wang, S.-W. Jeon, and S.-Y. Chung, "Aligned interference neutralization and the degrees of freedom of the $2 \times 2 \times 2$ interference channel," *IEEE Trans. Inform. Theory*, vol. 58, no. 7, pp. 4381–4395, Jul. 2012.

[13] S. Mohajer, S. N. Diggavi, C. Fragouli, and D. N. C. Tse, "Approximate capacity of a class of Gaussian interference-relay networks," *IEEE Trans. Inform. Theory*, vol. 57, no. 5, pp. 2837–2864, May 2011.

[14] I. Shomorony and A. S. Avestimehr, "Degrees of freedom of two-hop wireless networks: 'everyone gets the entire cake'," *IEEE Trans. Inform. Theory*, vol. 60, no. 5, pp. 2417–2431, May 2014.

[15] T. Gou, C. Wang, and S. A. Jafar, "Toward full-duplex multihop multiflow - a study of non-layered two unicast wireless networks," *IEEE J. Select. Areas Commun.*, vol. 32, no. 9, pp. 1738–1751, 2014.

[16] H. Bölcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[17] C. Wang, H. Farhadi, and M. Skoglund, "Achieving the degrees of freedom of wireless multi-user relay networks," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2612–2622, Sep. 2012.

[18] C. Wang and M. Skoglund, "Multi-user multi-hop relay networks: Transmission schemes and degrees of freedom," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4582–4596, Aug. 2015.

[19] Y. Fan, C. Wang, J. Thompson, and H. V. Poor, "Recovering multiplexing loss through successive relaying using repetition coding," *IEEE Trans. Wireless Commun.*, vol. 6, no. 12, pp. 4484–4493, Dec. 2007.

[20] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Select. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.

[21] S. Yang and J.-C. Belfiore, "Towards the optimal amplify-and-forward cooperative diversity scheme," *IEEE Trans. Inform. Theory*, vol. 53, no. 9, pp. 3114–3126, Sep. 2007.

[22] S. O. Gharan, A. Bayesteh, and A. K. Khandani, "On the diversity-multiplexing tradeoff in multiple-relay network," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5423–5444, Dec. 2009.

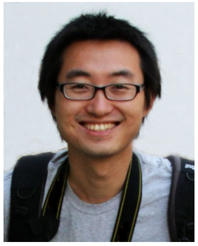
[23] Y. Zhang, Z. Zhang, L. Ping, X. Chen, and C. Zhong, "Capacity scaling of relay networks with successive relaying," in *IEEE ISIT*, Hongkong, China, 14-19 Jun. 2015.

[24] ——, "On the capacity scaling of large multipair relay networks with successive relaying protocol," *IEEE ACCESS*, vol. 5, pp. 5882–5895, Apr. 2017.

[25] N. Lee and R. Heath, Jr., "Degrees of freedom for the two-cell two-hop MIMO interference channel: Interference-free relay transmission and spectrally efficient relaying protocol," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 2882–2896, May 2013.

[26] A. S. Zamzam, A. El-Keyi, M. Nafie, and Y. Mohasseb, "On the degrees of freedom of the two-cell two-hop MIMO network with dedicated and shared relays," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6738–6751, Dec. 2015.

[27] S. K. Sharma, T. E. Bogale, L. B. Le, S. Chatzinotas, X. Wang, and B. Ottersten, "Dynamic spectrum sharing in 5G wireless networks with full-duplex technology: Recent advances and research challenges," *IEEE Commun. Surv. Tut.*, vol. PP, no. 99, pp. 1–1, Nov. 2017.



Chao Wang (S'07-M'09) received his B.E. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2003, his M.Sc. and Ph.D. degrees from The University of Edinburgh, Edinburgh, UK, in 2005 and 2009, respectively. He was a Visiting Student Research Collaborator at Princeton University, Princeton, USA, in 2008. From 2009 to 2012, he was a Postdoctoral Research Associate at KTH-Royal Institute of Technology, Stockholm, Sweden. Since 2013, he has been with Tongji University, Shanghai, China, where he is an Associate Professor. He is currently taking a Marie Sklodowska-Curie Individual Fellowship at the University of Exeter, Exeter, UK. His research interests mainly lie in information theory and signal processing for wireless communication networks, as well as data-driven research and applications for smart city and intelligent transportation systems.

Ping Wang graduated from the Department of Computer Science and Engineering at Shanghai Jiaotong University and received Ph.D. degree in the major of Computer Applied Technology in 2007. Since 2007, he has been working in the Department of Information and Communication Engineering at Tongji University, and now he is an Associate Professor. His research interests include the MAC and routing algorithms for VANETs and the next generation wireless networks. He has been building prototype systems for connected vehicles and intelligent connected vehicles. He has undertaken the design and construction of V2X system for SAIC-Tongji Testbed of intelligent connected vehicles. He has published over 50 scientific papers and 3 books. He has applied 1 international patent and over 20 national patents, and submitted 7 standardization drafts.

Fuqiang Liu received the bachelor's degree from Tianjin University, Tianjin, China, in 1987 and the Ph.D. degree from China University of Mining and Technology, Beijing, China, in 1996. He is currently a Professor at the College of Electronics and Information Engineering, Tongji University, Shanghai, China, and Director of the Broadband Wireless Communication and Multimedia Laboratory. He is also a Guest Professor in the National Institute of Informatics, Tokyo, Japan. He is the author of more than 300 scientific papers and nine books. He has participated in numerous national research projects in China and has received research funding from the U.S., Finland, the European Union, and Japan. His research interests include the theory and technology of broadband wireless communication (the fifth generation mobile communication, vehicular communication/WAVE) and their applications in automotive and intelligent transportation systems.

Geyong Min is a Professor of High Performance Computing and Networking in the Department of Computer Science within the College of Engineering, Mathematics and Physical Sciences at the University of Exeter, United Kingdom. He received the PhD degree in Computing Science from the University of Glasgow, United Kingdom, in 2003, and the B.Sc. degree in Computer Science from Huazhong University of Science and Technology, China, in 1995. His research interests include Future Internet, Computer Networks, Wireless Communications, Multimedia Systems, Information Security, High Performance Computing, Ubiquitous Computing, Modelling and Performance Engineering.