

Capacity Limits of Full-Duplex Cellular Network

Kaiming Shen, *Member, IEEE*, Reza K. Farsani, *Student Member, IEEE*, and Wei Yu, *Fellow, IEEE*

Abstract—This paper aims to characterize the capacity limits of a wireless cellular network with a full-duplex (FD) base-station (BS) and half-duplex user terminals, in which three independent messages are communicated: the uplink message m_1 from the uplink user to the BS, the downlink message m_2 from the BS to the downlink user, and the device-to-device (D2D) message m_3 from the uplink user to the downlink user. From an information theoretical perspective, the overall network can be viewed as a generalization of the FD relay broadcast channel with a side message transmitted from the relay to the destination. We begin with a simpler case that involves the uplink and downlink transmissions of (m_1, m_2) only, and propose an achievable rate region based on a novel strategy that uses the BS as a FD relay to facilitate the interference cancellation at the downlink user. We also prove a new converse, which is strictly tighter than the cut-set bound, and characterize the capacity region of the scalar Gaussian FD network without a D2D message to within a constant gap. This paper further studies a general setup wherein (m_1, m_2, m_3) are communicated simultaneously. To account for the D2D message, we incorporate Marton's broadcast coding into the previous scheme to obtain a larger achievable rate region than the existing ones in the literature. We also improve the cut-set bound by means of genie and show that by using one of the two simple rate-splitting schemes, the capacity region of the scalar Gaussian FD network with a D2D message can already be reached to within a constant gap. Finally, a generalization to the vector Gaussian channel case is discussed. Simulation results demonstrate the advantage of using the BS as relay in enhancing the throughput of the FD cellular network.

Index Terms—Approximate capacity, device-to-device, cellular network, full-duplex, relay broadcast channel with side message.

I. INTRODUCTION

TRADITIONAL wireless cellular systems separate uplink and downlink signals by using either time division duplex (TDD) or frequency division duplex (FDD), because at a conventional analog front-end, the echo due to transmitting in one direction can overwhelm the receiver in the other direction. Recent progress in analog and digital echo cancellation [1]–[3] is now opening up the possibility of realizing bi-directional

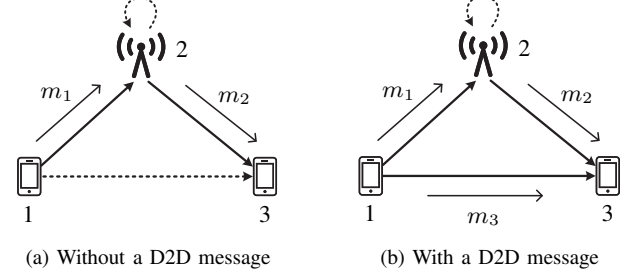


Fig. 1. Cellular network with uplink user (node 1), full-duplex BS (node 2), and downlink user (node 3). Here, m_1 is the uplink message, m_2 is the downlink message, and m_3 is the D2D side-message. The loops represent self-interference. The straight dashed line in (a) represents the uplink-to-downlink cross-channel interference.

communication in a full-duplex (FD) fashion. This paper considers the capacity limits of the FD cellular network.

In a FD cellular network, the base-station (BS) is capable of transmitting and receiving signals in FD mode [4]; but it is often the case that the uplink and downlink user terminals still operate in half-duplex mode. In such a system as depicted in Fig. 1(a), although the uplink transmission of m_1 and the downlink transmission of m_2 occupy the same spectrum simultaneously thereby doubling the frequency-reuse factor as compared to TDD or FDD, the cross-channel interference from the uplink user (node 1) to the downlink user (node 3) is still a major source of impairment. Such cross-channel interference is in fact the performance bottleneck in FD networks as pointed out in [5], [6], especially when the uplink and downlink user terminals are in close proximity to each other.

This paper aims to show that this cross-channel interference can potentially be cancelled or significantly suppressed with the aid from the BS. This is because the BS can act as a relay, as it already needs to decode the uplink message m_1 , so it can help the downlink user cancel the cross-channel interference due to m_1 . Under a scalar Gaussian channel model for the setup depicted in Fig. 1(a) with (m_1, m_2) only, this paper shows that the proposed interference cancellation scheme can achieve the capacity region of this channel to within a constant gap.

This paper also considers a scenario in which in addition to the uplink and downlink messages, the uplink user also wishes to directly send a separate message m_3 to the downlink user via the device-to-device (D2D) link. For this new channel model as shown in Fig. 1(b), we propose to incorporate Marton's broadcast coding [7] of m_1 and m_3 into the previous transmission scheme to derive a general achievable rate region. We further propose two simple rate-splitting schemes and a new converse, and show that using one of the two rate-splitting schemes (depending on the channel condition) already suffices to attain the capacity region to within a constant gap for the

Manuscript received May 2, 2019; revised February 19, 2020; to appear in IEEE Transactions on Information Theory. The work of K. Shen was supported in part by National Natural Science Foundation of China (NSFC) 62001411, in part by Natural Sciences and Engineering Research Council (NSERC) of Canada, and in part by Huawei Technologies Canada. The work of R. K. Farsani and W. Yu was supported in part by NSERC of Canada, and in part by Huawei Technologies Canada. The materials in this paper have been presented in part in IEEE Information Theory Workshop (ITW), December 2018, Guangzhou, China and in part in IEEE International Symposium on Information Theory (ISIT), July 2019, Paris, France.

K. Shen is with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China (e-mail: shenkaiming@cuhk.edu.cn).

R. K. Farsani and W. Yu are with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mails: {rkfarsani, weiyu}@ece.utoronto.ca).

TABLE I
MAIN RESULTS OF THE PAPER

	Achievability	Converse	Capacity Region
Discrete Memoryless Channel (DMC)	Theorem 1	Theorem 2	–
Gaussian Channel	Proposition 2	Proposition 5	Theorems 5, 7 and 8
DMC with D2D	Theorem 3, Corollaries 1 and 2	Theorem 4	–
Gaussian Channel with D2D	Propositions 3 and 4	Proposition 6 and Corollary 3	Theorem 6

scalar Gaussian FD cellular network with a D2D message.

The FD cellular network with only the uplink and the downlink transmissions has been extensively studied in the existing literature. Most of the prior works propose to alleviate the cross-channel interference by optimizing resource allocation, e.g., [8] schedules the uplink and downlink users in accordance with the distance between them, [9] uses power control to combat cross-channel interference and self-interference, and [10], [11] further consider joint power control and user scheduling. Moreover, [12] shows empirically that the gain of FD mode over half-duplex increases with the number of users. For the multiple-input multiple-output (MIMO) setup, [13] exploits spatial diversity by scheduling users in an opportunistic manner. These optimization-based works always treat interference as noise. In contrast, this paper employs more sophisticated coding techniques to try to cancel the interference, while aiming to provide insight into the fundamental capacity limits of the FD cellular network. In particular, the present work determines the capacity region to within a constant gap in the scalar Gaussian channel case, as opposed to the existing theoretical studies in [14]–[16] that only characterize the sum rate in an asymptotic regime as the signal-to-noise ratio (SNR) tends to infinity. Furthermore, the capacity analyses are extended to the FD cellular model with a D2D message.

We point out that the FD cellular network with D2D is equivalent to the relay broadcast channel with side message (or with “private” message [17]). The authors of [17] propose a decode-and-forward scheme and a compress-and-forward scheme for this channel. Our scheme is a further development of the decode-and-forward scheme [17] by incorporating multiple new techniques (including rate splitting, joint decoding, and Marton’s broadcast coding [7]). With respect to the converse, [17] derives an outer bound based on the genie-aided method, but as indicated by the authors, the outer bound of [17] is not computable. This paper develops better use of the auxiliary “genie” variables to improve upon the cut-set bound, and further comes up with a new sum-rate upper bound that plays a key role in characterizing the capacity region for the scalar Gaussian case to within a constant gap.

The FD cellular network with D2D is also a generalization of the *partially-cooperative relay broadcast channel* [18], [19] for which a modified Marton’s broadcast coding scheme has already been proposed. The achievability part of our paper can be thought of as a generalization of [18], [19] in incorporating the transmission of the relay-to-destination side message m_2 into the modified Marton’s coding. Thus, the contribution of

the present paper can also be thought of as the characterization of the capacity of the Gaussian relay broadcast channel (with side message) to within a constant gap.

For ease of reference, we categorize the main results of the paper in Table I as displayed at the top of the page. Specifically, the main contributions of this paper are as follows:

- *Achievability*: For the FD cellular network without a D2D message, we propose a relaying strategy to improve upon the existing achievable uplink and downlink rate region. When the D2D message is included, we extend the scheme by incorporating Marton’s broadcast coding.
- *Converse*: We derive new upper bounds on the sum rate for both the cases with and without D2D. Further, we use different genies to provide tighter converses.
- *Scalar Gaussian Channel*: We characterize the capacity region of the scalar Gaussian FD cellular network (both with and without D2D) to within 1 bit in general. For the case without D2D, (i) a smaller constant gap of approximately 0.6358 bits is established in the *strong interference* regime; (ii) the exact capacity is determined in the *very strong interference* regime.
- *Vector Gaussian Channel*: We discuss the generalization of the achievability results to the MIMO case that includes spatial multiplexing and dirty-paper coding.

Notation: Let $[1 : n]$ be the set $\{1, 2, \dots, n\}$, $\mathbb{C}(x)$ the function $\log_2(1 + x)$ for $x \geq 0$, \mathbb{R}_+ the set of nonnegative real numbers, and \mathbb{C} the set of complex numbers. We use a superscripted letter to denote a sequence of variables, e.g., $X^N = (X_1, \dots, X_N)$, use \mathbf{I} to denote the identity matrix, and use \mathbf{A}^H to denote the Hermitian transpose of matrix \mathbf{A} . For a random variable X , use $\mathbb{E}[X]$ to denote the expected value, and $\text{Var}(X)$ the variance. For two random variables X_1 and X_2 , use $\text{Cov}(X_1, X_2)$ to denote their covariance, and use $X_1 \perp\!\!\!\perp X_2$ to indicate that they are independent.

The rest of the paper is organized as follows. Section II formally defines the various channel models. Section III discusses the discrete memoryless channel. Section IV discusses the scalar Gaussian channel. Section V discusses the generalization to the vector Gaussian channel. Numerical results are presented in Section VI. Finally, we conclude this work in Section VII.

II. FULL-DUPLEX CELLULAR NETWORK MODELS

This work examines two different FD cellular network setups: one has only uplink and downlink transmissions, the other includes D2D transmission in addition. We consider both the discrete memoryless channel case and the Gaussian channel case.

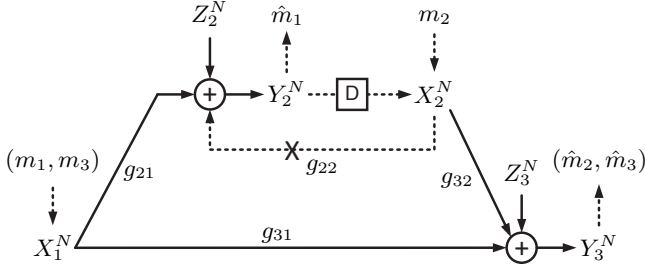


Fig. 2. Gaussian full-duplex relay broadcast channel with side message m_2 . The block “D” represents a one-codeword delay.

A. Without the D2D Message

We first consider the FD cellular network without the D2D message, as shown in Fig. 1(a).

1) *Discrete Memoryless Channel Model*: Let $X_{in} \in \mathcal{X}_i$ be the transmitted signal of node $i \in \{1, 2\}$ and $Y_{jn} \in \mathcal{Y}_j$ be the received signal at node $j \in \{2, 3\}$, at the n th channel use, over the alphabet sets $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_2, \mathcal{Y}_3)$. The discrete memoryless channel model is defined by the channel transition probability mass function (pmf) $p(y_{2n}, y_{3n} | x_{1n}, x_{2n})$, which captures the self-interference from X_{1n} to Y_{2n} . Over a total of N channel uses, node 1 wishes to send $m_1 \in [1 : 2^{NR_1}]$ to node 2, while node 2 wishes to send $m_2 \in [1 : 2^{NR_2}]$ to node 3, where R_1 and R_2 are referred to as the uplink rate and the downlink rate, respectively. The encoding of X_{1n} solely depends on m_1 . In comparison, since the transmitter of X_{2n} and the receiver of Y_{2n} are co-located at node 2, the encoding of X_{2n} can depend on the past received signals Y_2^{n-1} :

$$X_{1n} = \mathcal{E}_1(m_1, n) \text{ and } X_{2n} = \mathcal{E}_2(m_2, Y_2^{n-1}, n), \quad (1)$$

for $n \in [1 : N]$. After N channel uses, node 3 decodes m_2 based on Y_3^N . Because node 2 itself is the downlink transmitter, it can make use of m_2 in addition to Y_2^N in decoding m_1 , i.e.,

$$\hat{m}_1 = \mathcal{D}_1(Y_2^N, m_2) \text{ and } \hat{m}_2 = \mathcal{D}_2(Y_3^N). \quad (2)$$

An uplink-downlink rate pair (R_1, R_2) is said to be achievable if there exists a set of deterministic functions $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D}_2, \mathcal{D}_3)$ such that the error probability $\Pr\{(\hat{m}_1, \hat{m}_2) \neq (m_1, m_2)\}$ tends to zero as $N \rightarrow \infty$.

2) *Gaussian Channel Model*: For the scalar Gaussian channel model, we have $X_{in}, Y_{jn} \in \mathbb{C}$. We impose power constraints on X_{in} , i.e., $\sum_{n=1}^N |X_{in}|^2 \leq NP_i$, $i \in \{1, 2\}$, and have

$$Y_{2n} = g_{21}X_{1n} + Z_{2n}, \quad (3)$$

$$Y_{3n} = g_{31}X_{1n} + g_{32}X_{2n} + Z_{3n}, \quad (4)$$

for $n \in [1 : N]$, where $g_{ji} \in \mathbb{C}$ is the channel gain from the transmitter node i to the receiver node j , and $Z_{jn} \sim \mathcal{CN}(0, \sigma^2)$ for the fixed $\sigma^2 > 0$ is the additive white Gaussian noise at node j in the n th channel use.

For the vector Gaussian channel model, we assume that node 1 has L_1^+ transmit antennas, node 2 has L_2^- receive antennas and L_2^+ transmit antennas, and node 3 has L_3^- receive antennas. The generalizations of g_{ji} , X_{in} , Y_{jn} , and

Z_{jn} to this vector case are $\mathbf{G}_{ji} \in \mathbb{C}^{L_j^- \times L_i^+}$, $\mathbf{X}_{in} \in \mathbb{C}^{L_i^+}$, $\mathbf{Y}_{jn} \in \mathbb{C}^{L_j^-}$, and $\mathbf{Z}_{jn} \in \mathbb{C}^{L_j^-}$, respectively. We remark that \mathbf{Z}_{jn} is an i.i.d. vector Gaussian random variable drawn from $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$. In particular, the power constraints now become $\sum_{n=1}^N \|\mathbf{X}_{in}\|^2 \leq NP_i$, $i \in \{1, 2\}$. Thus, we have

$$\mathbf{Y}_{2n} = \mathbf{G}_{21}\mathbf{X}_{1n} + \mathbf{Z}_{2n}, \quad (5)$$

$$\mathbf{Y}_{3n} = \mathbf{G}_{31}\mathbf{X}_{1n} + \mathbf{G}_{32}\mathbf{X}_{2n} + \mathbf{Z}_{3n}. \quad (6)$$

In both the scalar and the vector Gaussian cases, we assume that the channel state information (CSI), i.e., $\{g_{ji} \text{ or } \mathbf{G}_{ji}, \forall (i, j)\}$, is available everywhere. Due to the fact that the BS (i.e., node 2) operates in a full-duplex mode, the self-interference at the relay has been removed implicitly, as illustrated in Fig. 2. Thus, we make an idealized assumption that the self-interference can be fully removed.

B. With the D2D Message

Next, we consider the FD cellular network with the D2D message, as shown in Fig. 1(b).

1) *Discrete Memoryless Channel Model*: We now include a direct transmission of $m_3 \in [1 : 2^{NR_3}]$ from node 1 to node 3 in the discrete memoryless channel model as described in Section II-A; R_3 is referred to as the D2D rate. The channel setup, i.e., the alphabet sets $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_2, \mathcal{Y}_3)$ and the channel transition probability $p(y_{2n}, y_{3n} | x_{1n}, x_{2n})$, remains the same as before. Because m_1 and m_3 are both transmitted from node 1, the encoding of X_1 now depends on (m_1, m_3) , i.e.,

$$X_{1n} = \mathcal{E}_1(m_1, m_3, n) \text{ and } X_{2n} = \mathcal{E}_2(m_2, Y_2^{n-1}, n). \quad (7)$$

Moreover, since m_2 and m_3 are both intended for node 3, we define the decoding functions differently:

$$\hat{m}_1 = \mathcal{D}_1(Y_2^N, m_2) \text{ and } (\hat{m}_2, \hat{m}_3) = \mathcal{D}_2(Y_3^N). \quad (8)$$

Similarly, a rate triple (R_1, R_2, R_3) is said to be achievable if there exists a set of deterministic functions $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D}_1, \mathcal{D}_2)$ such that the probability of error, $\Pr\{(\hat{m}_1, \hat{m}_2, \hat{m}_3) \neq (m_1, m_2, m_3)\}$, tends to zero as $N \rightarrow \infty$.

2) *Gaussian Channel Model*: The scalar Gaussian channel model follows a similar fashion as in the without D2D case, except the extra message m_3 . The channel outputs (Y_{2n}, Y_{3n}) corresponding to the inputs (X_{1n}, X_{2n}) are still given by (3) and (4). Again, the encoding functions in (7) must satisfy the power constraints $\sum_{n=1}^N |X_{in}|^2 \leq NP_i$, $i \in \{1, 2\}$. The vector Gaussian channel model can be extended to the D2D case similarly.

III. DISCRETE MEMORYLESS CHANNEL MODEL

A. Achievability for Discrete Memoryless Model without D2D

As mentioned earlier, the cross-channel interference from node 1 to node 3 is the main bottleneck [4]. To address this issue, we use the BS (i.e., node 2) as a relay to facilitate cancelling the interfering signal at node 3. We further propose to split message m_1 (which causes the interference) so that node 3 can at least cancel a portion of the interference. The resulting achievable rate region is stated below.

TABLE II
PROPOSED CODING SCHEME FOR THE WITHOUT D2D CASE

t	1	2	\dots	$T-1$	T
X_1	$\mathbf{x}_1^N(m_{11}^1 m_{10}^1, 1)$	$\mathbf{x}_1^N(m_{11}^2 m_{10}^2, m_{10}^1)$	\rightarrow	$\mathbf{x}_1^N(m_{11}^{T-1} m_{10}^{T-1}, m_{10}^{T-2})$	$\mathbf{x}_1^N(1 1, m_{10}^{T-1})$
Y_2	$(\hat{m}_{10}^1, \hat{m}_{11}^1)$	$(\hat{m}_{10}^2, \hat{m}_{11}^2)$	\rightarrow	$(\hat{m}_{10}^{T-1}, \hat{m}_{11}^{T-1})$	\emptyset
X_2	$\mathbf{x}_2^N(m_2^1 1)$	$\mathbf{x}_2^N(m_2^2 \hat{m}_{10}^1)$	\rightarrow	$\mathbf{x}_2^N(m_2^{T-1} \hat{m}_{10}^{T-2})$	$\mathbf{x}_2^N(m_2^T \hat{m}_{10}^{T-1})$
Y_3	$(1, \hat{m}_2^1)$	$(\hat{m}_{10}^1, \hat{m}_2^2)$	\leftarrow	$(\hat{m}_{10}^{T-2}, \hat{m}_2^{T-1})$	$(\hat{m}_{10}^{T-1}, \hat{m}_2^T)$

Theorem 1: For the discrete memoryless FD cellular network without D2D, a rate pair (R_1, R_2) is achievable if it is in the convex hull of the rate regions

$$R_1 \leq I(X_1; Y_2|U, X_2), \quad (9a)$$

$$R_2 \leq I(X_2; Y_3|U, V), \quad (9b)$$

$$R_1 + R_2 \leq I(X_1; Y_2|U, V, X_2) + I(U, V, X_2; Y_3), \quad (9c)$$

over the joint pmf $p(u)p(v, x_1|u)p(x_2|u)$, where the cardinalities of the auxiliary variables U and V can be bounded by $|\mathcal{U}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 2$ and $|\mathcal{V}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1$.

Proof: Consider a total of T blocks in order to carry out the block Markov coding. We use the superscript $t \in [1 : T]$ to denote the variables associated with block t .

For each block t , split m_t^t into a common-private message pair $(m_{10}^t, m_{11}^t) \in [1 : 2^{NR_{10}}] \times [1 : 2^{NR_{11}}]$ with $R_{10} + R_{11} = R_1$; the common message m_{10}^t is decoded at both receivers (i.e., node 2 and node 3), while the private message m_{11}^t is decoded at the intended receiver (i.e., node 2). For each block $t \in [1 : T]$, generate the following codebooks in an i.i.d. fashion according to their respective distributions as in $p(u)p(v, x_1|u)p(x_2|u)$:

- Relay codebook $\mathbf{u}^N(m_{10}^{t-1})$;
- Uplink common message codebook $\mathbf{v}^N(m_{10}^t|m_{10}^{t-1})$;
- Uplink private message codebook $\mathbf{x}_1^N(m_{11}^t|m_{10}^t, m_{10}^{t-1})$;
- Downlink message codebook $\mathbf{x}_2^N(m_2^t|m_{10}^{t-1})$.

Node 1 uses $(\mathbf{u}^N, \mathbf{v}^N, \mathbf{x}_1^N)$ for encoding messages; node 2 uses $(\mathbf{u}^N, \mathbf{x}_2^N)$ for encoding, and uses $(\mathbf{u}^N, \mathbf{v}^N, \mathbf{x}_1^N)$ for decoding; node 3 uses $(\mathbf{u}^N, \mathbf{v}^N, \mathbf{x}_2^N)$ for decoding.

In block t , knowing its past common message m_{10}^{t-1} , node 1 transmits $\mathbf{x}_1^N(m_{11}^t|m_{10}^t, m_{10}^{t-1})$. Here, we set $m_{10}^0 = 0$ by convention. Also, no message is transmitted in the last block, i.e., we set $m_{10}^T = m_{11}^T = 0$.

In block t , after obtaining \hat{m}_{10}^{t-1} from the previous block $t-1$, node 2 transmits $\mathbf{x}_2^N(m_2^t|\hat{m}_{10}^{t-1})$, and recovers $(\hat{m}_{10}^t, \hat{m}_{11}^t)$ from the received signal Y_2^N according to a jointly ϵ -strongly-typical set $\mathcal{T}_\epsilon^{(N)}$. Specifically, node 2 seeks a pair of $(\hat{m}_{10}^t, \hat{m}_{11}^t)$ such that the corresponding codeword $\mathbf{x}_1^N(\hat{m}_{10}^t, \hat{m}_{11}^t)$ produces an empirical pmf $\pi(x_1, y_2)$ with $|\pi(x_1, y_2) - p(x_1, y_2)| \leq \epsilon p(x_1, y_2)$ for all $(x_1, y_2) \in \mathcal{X}_1 \times \mathcal{Y}_2$, namely *strong typicality* [20].

By the packing lemma [21], the error probability $\Pr\{(\hat{m}_{10}^t, \hat{m}_{11}^t) \neq (m_{10}^t, m_{11}^t)\}$ tends to zero as $N \rightarrow \infty$ provided that

$$R_{11} \leq I(X_1; Y_2|U, V, X_2), \quad (10)$$

$$R_{10} + R_{11} \leq I(X_1; Y_2|U, X_2). \quad (11)$$

Node 3 decodes the blocks in a *backward* direction, i.e., block t prior to block $t-1$. In block t , after obtaining m_{10}^t from the previous block $t+1$, node 3 recovers $(\hat{m}_{10}^{t-1}, \hat{m}_2^t)$ jointly from the received signal Y_3^N ; the error probability $\Pr\{(\hat{m}_{10}^{t-1}, \hat{m}_2^t) \neq (m_{10}^{t-1}, m_2^t) | \hat{m}_{10}^{t-1} = m_{10}^{t-1}\}$ tends to zero as $N \rightarrow \infty$ if

$$R_2 \leq I(X_2; Y_3|U, V), \quad (12)$$

$$R_{10} + R_2 \leq I(U, V, X_2; Y_3). \quad (13)$$

The overall error probability P_e , i.e., $\Pr\{(\hat{m}_1, \hat{m}_2) \neq (m_1, m_2)\}$, can be upper bounded as

$$P_e \leq \frac{1}{T} \cdot \sum_{t=1}^T \left[\Pr\{(\hat{m}_{10}^t, \hat{m}_{11}^t) \neq (m_{10}^t, m_{11}^t)\} + \Pr\{(\hat{m}_{10}^{t-1}, \hat{m}_2^t) \neq (m_{10}^{t-1}, m_2^t) | \hat{m}_{10}^{t-1} = m_{10}^{t-1}\} \right], \quad (14)$$

so P_e tends to zero as $N \rightarrow \infty$ if (10)–(13) are satisfied. Furthermore, combining (10)–(13) with $R_{10}, R_{11} \geq 0$ and $R_1 = R_{10} + R_{11}$ by using the Fourier-Motzkin elimination, we obtain the inner bound in (9). Note that the effective uplink rate equals to $(T-1)/T \cdot R_1$. The achievability of (9) is established by letting $T \rightarrow \infty$. Finally, the cardinality bounds on U and V are due to the property of convex set. ■

Table II summarizes the above coding scheme, in which the arrows show the orderings of blocks for encoding or decoding.

In Theorem 1, the auxiliary variable U enables relaying at node 2 to assist node 3 in cancelling the cross-channel interference. The resulting achievable rate region can be strictly larger than that of the no relaying scheme (with $U = \emptyset$).

B. Converse for Discrete Memoryless Model without D2D

The best previous converse is due to [17], which proposes an outer bound for a more general model with D2D with (R_1, R_2, R_3) . When specialized to the without D2D case, i.e., when $R_3 = 0$, their converse amounts to the cut-set bound which consists of two individual upper bounds on R_1 or R_2 . In this section, we propose a new upper bound on $R_1 + R_2$ that improves the cut-set bound. This new bound is stated in the following theorem.

Theorem 2: For the discrete memoryless FD cellular network without D2D, any achievable rate pair (R_1, R_2) must satisfy

$$R_1 \leq I(X_1; Y_2|X_2), \quad (15a)$$

$$R_2 \leq I(X_2; Y_3|X_1), \quad (15b)$$

$$R_1 + R_2 \leq I(X_1; Y_2, Y_3|X_2) + I(X_2; Y_3), \quad (15c)$$

for some joint pmf $p(x_1, x_2)$.

Proof: Let M_i be the random variable denoting the message m_i . Observe that (15a) and (15b) are simply the cut-set bounds. In deriving the sum-rate bound (15c), the main idea is to introduce variable X_1^N into the mutual information term $I(M_1; Y_2^N, X_2^N)$, as in the following:

$$\begin{aligned}
& N(R_1 + R_2 - \epsilon_N) \\
& \leq I(M_1; Y_2^N, X_2^N) + I(M_2; Y_3^N) \\
& \stackrel{(a)}{\leq} I(M_1; Y_2^N, X_2^N, Y_3^N | M_2) + I(M_2; Y_3^N) \\
& \stackrel{(b)}{=} I(M_1; Y_2^N, Y_3^N | M_2) + I(M_2; Y_3^N) \\
& \leq \sum_{n=1}^N \left[I(M_1; Y_{2n}, Y_{3n} | M_2, Y_2^{n-1}, Y_3^{n-1}) \right. \\
& \quad \left. + I(M_2, Y_2^{n-1}, Y_3^{n-1}; Y_{3n}) \right] \\
& \stackrel{(c)}{=} \sum_{n=1}^N \left[I(M_1; Y_{2n}, Y_{3n} | M_2, Y_2^{n-1}, Y_3^{n-1}, X_{2n}) \right. \\
& \quad \left. + I(M_2, Y_2^{n-1}, Y_3^{n-1}, X_{2n}; Y_{3n}) \right] \\
& \leq \sum_{n=1}^N \left[I(M_1, M_2, Y_2^{n-1}, Y_3^{n-1}; Y_{2n}, Y_{3n} | X_{2n}) \right. \\
& \quad \left. + I(X_{2n}; Y_{3n}) \right] \\
& \stackrel{(d)}{=} \sum_{n=1}^N \left[I(X_{1n}; Y_{2n}, Y_{3n} | X_{2n}) + I(X_{2n}; Y_{3n}) \right] \\
& \leq NI(X_1; Y_2, Y_3 | X_2) + NI(X_2; Y_3), \tag{16}
\end{aligned}$$

where ϵ_N tends to zero as $N \rightarrow \infty$ by Fano's inequality, (a) follows as $M_1 \perp\!\!\!\perp M_2$, (b) and (c) both follow as X_{2n} is a deterministic function of (M_2, Y_2^{n-1}) , (d) follows as $(M_1, M_2, Y_2^{n-1}, Y_3^{n-1}) \rightarrow X_{1n} \rightarrow (Y_{2n}, Y_{3n})$ form a Markov chain conditioned on X_{2n} . The converse is then verified. ■

Note that we restrict the relay operation to be deterministic in the channel model and in deriving the above outer bound. We remark that (15c) is not contained in the cut-set bound and yet is critical to characterizing the capacity region to within a constant gap for the scalar Gaussian channel in Section IV-C.

C. Achievability for Discrete Memoryless Model with D2D

Recall that the FD cellular network with D2D is a generalization of the FD relay broadcast channel [18], [19], so we use the previous studies in [18], [19] as a starting point. The works [18], [19] propose to modify the classic Marton's coding [7] for the broadcast channel to the case where one receiver further helps the other receiver via a relay link. The channel model considered in this paper is a further generalization in which the extra side message m_2 is carried in this relay link. The coding strategy proposed below incorporates m_2 in Marton's broadcast coding.

The coding strategy of [18], [19] splits each message (i.e., m_1 and m_3) into the private and common parts which are dealt with differently. The common part is decoded by both node 2 and node 3; node 2 further acts as a relay to assist node 3 in decoding the common message. In contrast, the private parts

are decoded only by the intended node through the broadcast channel without using node 2 as relay, so Marton's coding can be applied. This paper makes two modifications to this strategy in order to enable an extra transmission of m_2 . First, we let the encoding of X_2 be based on both m_1 and m_2 . Second, we let node 3 decode the original common and private message jointly with the new message m_2 . The resulting achievable rate region is stated as follows.

Theorem 3: For the discrete memoryless FD cellular network with D2D, a rate triple (R_1, R_2, R_3) is achievable if it is in the convex hull of the rate regions

$$R_1 \leq \mu_3, \tag{17a}$$

$$R_2 \leq \min\{\mu_5, \mu_2 + \mu_6 - \mu_1\}, \tag{17b}$$

$$R_1 + R_3 \leq \mu_3 + \mu_4 - \mu_1, \tag{17c}$$

$$R_2 + R_3 \leq \mu_7, \tag{17d}$$

$$R_1 + R_2 + R_3 \leq \min\{\mu_2 + \mu_7 - \mu_1, \mu_3 + \mu_6 - \mu_1\}, \tag{17e}$$

over the joint pmf $p(u)p(v, w_1, w_3, x_1|u)p(x_2|u)$ under the constraint that $\mu_1 \leq \mu_2 + \mu_4$, where

$$\mu_1 = I(W_1; W_3 | U, V), \tag{18a}$$

$$\mu_2 = I(W_1; Y_2 | U, V, X_2), \tag{18b}$$

$$\mu_3 = I(V, W_1; Y_2 | U, X_2), \tag{18c}$$

$$\mu_4 = I(W_3; Y_3 | U, V, X_2), \tag{18d}$$

$$\mu_5 = I(X_2; Y_3 | U, V, W_3), \tag{18e}$$

$$\mu_6 = I(W_3; X_2; Y_3 | U, V), \tag{18f}$$

$$\mu_7 = I(U, V, W_3, X_2; Y_3). \tag{18g}$$

Proof: Again, we consider a total of T blocks and use $t \in [1 : T]$ to index the block. For each block t , split m_i^t into the common-private message pair $(m_{i0}^t, m_{ii}^t) \in [1 : 2^{nR_{i0}}] \times [1 : 2^{nR_{ii}}]$ for $i \in \{1, 3\}$. For each block t , in an i.i.d. manner according to their respective distributions, generate a common codebook

- Relay codebook $\mathbf{u}^N(m_{10}^{t-1}, m_{30}^{t-1})$;
- Common codebook $\mathbf{v}^N(m_{10}^t, m_{30}^t | m_{10}^{t-1}, m_{30}^{t-1})$;
- Separate binning codebooks $\mathbf{w}_1^N(\ell_{11}^t)$ and $\mathbf{w}_3^N(\ell_{33}^t)$;
- Joint binning codebook $\mathbf{x}_1^N(\ell_{11}^t, \ell_{33}^t | m_{10}^{t-1}, m_{30}^{t-1})$;
- Downlink codebook $\mathbf{x}_2^N(m_{20}^t | m_{10}^{t-1}, m_{30}^{t-1})$,

where the codebook pair $(\mathbf{w}_1^N(\ell_{11}^t), \mathbf{w}_3^N(\ell_{33}^t))$ is generated for each $(\ell_{11}, \ell_{33}) \in [1 : 2^{NR'_{11}}] \times [1 : 2^{NR'_{33}}]$, with $R'_{ii} \geq R_{ii}$, $i \in \{1, 3\}$, and with each ℓ_{ii}^t uniformly mapped to the bin of m_{ii}^t , namely $\mathcal{B}_i(m_{ii}^t)$. Node 1 uses $(\mathbf{u}^N, \mathbf{v}^N, \mathbf{w}_1^N, \mathbf{w}_3^N, \mathbf{x}_1^N)$ for encoding; node 2 uses $(\mathbf{u}^N, \mathbf{x}_2^N)$ for encoding, and uses $(\mathbf{u}^N, \mathbf{v}^N, \mathbf{w}_1^N, \mathbf{x}_1^N)$ for decoding; node 3 uses $(\mathbf{u}^N, \mathbf{w}_3^N, \mathbf{x}_2^N)$ for decoding.

In block t , node 1 finds a pair of $(\ell_{11}^t, \ell_{33}^t) \in \mathcal{B}_1(m_{11}^t) \times \mathcal{B}_3(m_{33}^t)$ such that $(\mathbf{w}_1^N(\ell_{11}^t), \mathbf{w}_3^N(\ell_{33}^t))$ is in a strongly typical set $\mathcal{T}_{\epsilon'}^{(N)}$, then transmits $\mathbf{x}_1^N(\ell_{11}^t, \ell_{33}^t | m_{10}^{t-1}, m_{30}^{t-1})$. The above typicality criterion $\mathcal{T}_{\epsilon'}^{(N)}$ needs to be stricter than the typicality criterion $\mathcal{T}_{\epsilon}^{(N)}$ used for decoding in the sense that $0 < \epsilon' < \epsilon$. This encoding is guaranteed to be successful provided that

$$R'_{11} + R'_{33} - R_{11} - R_{33} \geq I(W_1; W_3 | U, V). \tag{19}$$

In block t , after obtaining $(\hat{m}_{10}^{t-1}, \hat{m}_{30}^{t-1})$ from the previous

TABLE III
PROPOSED CODING SCHEME FOR THE D2D CASE WITH D2D RATE SPLITTING

t	1	2	\dots	$T-1$	T
X_1	$\mathbf{x}_1^N(m_1^1, m_{30}^1, m_{33}^1 1, 1)$	$\mathbf{x}_1^N(m_1^2, m_{30}^2, m_{33}^2 m_1^1, m_{30}^1)$	\rightarrow	$\mathbf{x}_1^N(m_1^{T-1}, m_{30}^{T-1}, m_{33}^{T-1} m_1^{T-2}, m_{30}^{T-2})$	$\mathbf{x}_1^N(1, 1, 1 m_1^{T-1}, m_{30}^{T-1})$
Y_2	$(\hat{m}_1^1, \hat{m}_{30}^1, \hat{m}_{33}^1)$	$(\hat{m}_1^2, \hat{m}_{30}^2, \hat{m}_{33}^2)$	\rightarrow	$(\hat{m}_1^{T-1}, \hat{m}_{30}^{T-1}, \hat{m}_{33}^{T-1})$	\emptyset
X_2	$\mathbf{x}_2^N(m_2^1 1, 1)$	$\mathbf{x}_2^N(m_2^2 \hat{m}_1^1, \hat{m}_{30}^1)$	\rightarrow	$\mathbf{x}_2^N(m_2^{T-1} \hat{m}_1^{T-2}, \hat{m}_{30}^{T-2})$	$\mathbf{x}_2^N(m_2^T \hat{m}_1^{T-1}, \hat{m}_{30}^{T-1})$
Y_3	$(1, \hat{m}_2^1, 1, \hat{m}_{33}^1)$	$(\hat{m}_1^2, \hat{m}_2^2, \hat{m}_{30}^2, \hat{m}_{33}^2)$	\leftarrow	$(\hat{m}_1^{T-2}, \hat{m}_2^{T-1}, \hat{m}_{30}^{T-2}, \hat{m}_{33}^{T-1})$	$(\hat{m}_1^{T-1}, \hat{m}_2^T, \hat{m}_{30}^{T-1}, 1)$

block $t-1$, node 2 transmits $\mathbf{x}_2^N(m_2^t | \hat{m}_{10}^{t-1}, \hat{m}_{30}^{t-1})$, and recovers $(\hat{m}_{10}^t, \hat{m}_{30}^t)$ jointly from the received signal Y_2^N ; this decoding is successful if

$$R'_{11} \leq I(W_1; Y_2 | U, V, X_2), \quad (20)$$

$$R_{10} + R_{30} + R'_{11} \leq I(V, W_1; Y_2 | U, X_2). \quad (21)$$

Node 3 decodes the blocks in a backward direction (unlike the sliding window decoding scheme of [19]), i.e., block t prior to block $t-1$. In block t , after obtaining $(\hat{m}_{10}^t, \hat{m}_{30}^t)$ from the previous block $t+1$, node 1 recovers $(\hat{m}_{10}^t, \hat{m}_{30}^t, \hat{m}_{33}^t, \hat{m}_2^t)$ jointly; the following conditions guarantee successful decoding:

$$R'_{33} \leq I(W_3; Y_3 | U, V, X_2), \quad (22)$$

$$R_2 \leq I(X_2; Y_3 | U, V, W_3), \quad (23)$$

$$R'_{33} + R_2 \leq I(W_3, X_2; Y_3 | U, V), \quad (24)$$

$$R_{10} + R_{30} + R'_{33} + R_2 \leq I(U, V, W_3, X_2; Y_3). \quad (25)$$

Combining (19)–(25) with $R_{11} \leq R'_{11}$, $R_{33} \leq R'_{33}$, $R_1 = R_{10} + R_{11}$, $R_3 = R_{30} + R_{33}$, and a nonnegative constraint on all the rate variables, and letting $T \rightarrow \infty$, we establish the proposed inner bound, including the constraint $\mu_1 \leq \mu_2 + \mu_4$, via the Fourier-Motzkin elimination. ■

Remark 1: The inner bound in Theorem 3 reduces to that of [18], [19] for the FD relay broadcast channel when $U = X_2$, reduces to a decode-and-forward inner bound (9) of [4] for the D2D case when $W_1 = W_3 = \emptyset$, and reduces to the inner bound in Theorem 1 for the without D2D case when $W_3 = \emptyset$.

Remark 2: The earlier works [18], [19] also use the decode-and-forward relaying but with sliding window decoding. This work uses backward decoding.

In Theorem 3, the term μ_1 is due to Marton's coding [7], reflecting the extent to which the encodings of the private messages m_{11} and m_{33} are coordinated through broadcasting. The following proposition further shows that the constraint $\mu_1 \leq \mu_2 + \mu_4$ must be satisfied automatically if $p(u)p(v, w_1, w_3, x_1 | u)p(x_2 | u)$ is optimally chosen for maximizing the rate region (17).

Proposition 1: The achievable rate region of Theorem 3 remains the same if the constraint $\mu_1 \leq \mu_2 + \mu_4$ is removed.

Proof: Let $\mathcal{R}_1 \subseteq \mathbb{R}_+^3$ be the achievable rate region defined by the set of inequalities in Theorem 3, and let \mathcal{R}_2 be the version without the constraint $\mu_1 \leq \mu_2 + \mu_4$. Clearly, $\mathcal{R}_1 \subseteq \mathcal{R}_2$, so it suffices to show $\mathcal{R}_2 \subseteq \mathcal{R}_1$ in order to prove $\mathcal{R}_1 = \mathcal{R}_2$. Consider some $p(u)p(v, w_1, w_3, x_1 | u)p(x_2 | u)$ such that $\mu_1 > \mu_2 + \mu_4$. Under this pmf, it can be shown that $\mathcal{R}_2 \subseteq \mathcal{R}'_2$

where \mathcal{R}'_2 is another rate region defined by

$$R_2 \leq \min\{\mu_5, I(X_2; Y_3 | U, V)\}, \quad (26a)$$

$$R_1 + R_3 \leq I(V; Y_2 | U, X_2), \quad (26b)$$

$$R_1 + R_2 + R_3 \leq I(U, V, X_2; Y_3). \quad (26c)$$

In the meanwhile, \mathcal{R}'_2 can be attained by setting $W_1 = \emptyset$ in Theorem 3. Thus, $\mathcal{R}_2 \subseteq \mathcal{R}_1$. ■

The inner bound in Theorem 3 involves rate splitting for both m_1 and m_3 . The following two corollaries present the special cases in which only one of (m_1, m_3) has rate splitting and Marton's coding is replaced with the superposition coding.

Corollary 1 (D2D Rate Splitting): For the discrete memoryless FD cellular network with D2D, a rate triple (R_1, R_2, R_3) is achievable if it is in the convex hull of

$$R_1 \leq I(V; Y_2 | U, X_2), \quad (27a)$$

$$R_2 \leq I(X_2; Y_3 | U, X_1), \quad (27b)$$

$$R_1 + R_3 \leq I(V; Y_2 | U, X_2) + I(X_1; Y_3 | U, V, X_2), \quad (27c)$$

$$R_1 + R_2 + R_3 \leq \min\{I(V; Y_2 | U, X_2) + I(X_1, X_2; Y_3 | U, V), I(X_1, X_2; Y_3)\}, \quad (27d)$$

over the joint pmf $p(u)p(v, x_1 | u)p(x_2 | u)$, where the cardinalities of auxiliary variables can be bounded by $|\mathcal{U}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 3$ and $|\mathcal{V}| \leq |\mathcal{X}_1| + 2$.

Proof: This inner bound is obtained by setting $W_1 = \emptyset$ and $W_3 = X_1$ in (17) of Theorem 3. The corresponding encoding and decoding procedure is illustrated in Table III. ■

Corollary 2 (Uplink Rate Splitting): For the discrete memoryless FD cellular network with D2D, a rate triple (R_1, R_2, R_3) is achievable if it is in the convex hull of

$$R_2 \leq I(X_2; Y_3 | U, V), \quad (28a)$$

$$R_1 + R_3 \leq I(X_1; Y_2 | U, X_2), \quad (28b)$$

$$R_2 + R_3 \leq I(U, V, X_2; Y_3), \quad (28c)$$

$$R_1 + R_2 + R_3 \leq I(X_1; Y_2 | U, V, X_2) + I(U, V, X_2; Y_3), \quad (28d)$$

over the joint pmf $p(u)p(v, x_1 | u)p(x_2 | u)$, where the cardinalities of the auxiliary variables can be bounded by $|\mathcal{U}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 3$ and $|\mathcal{V}| \leq |\mathcal{X}_1| + 2$.

Proof: This inner bound is obtained by setting $W_3 = \emptyset$ and $W_1 = X_1$ in (17) of Theorem 3. The corresponding encoding and decoding procedure is illustrated in Table IV. ■

Remark 3: The inner bound (28) reduces to the previous

TABLE IV
PROPOSED CODING SCHEME FOR THE D2D CASE WITH UPLINK RATE SPLITTING

t	1	2	\dots	$T-1$	T
X_1	$\mathbf{x}_1^N(m_{10}^1, m_{11}^1, m_3^1 1, 1)$	$\mathbf{x}_1^N(m_{10}^2, m_{11}^2, m_3^2 m_{10}^1, m_3^1)$	\rightarrow	$\mathbf{x}_1^N(m_{10}^{T-1}, m_{11}^{T-1}, m_3^{T-1} m_{10}^{T-2}, m_3^{T-2})$	$\mathbf{x}_1^N(1, 1, 1 m_{10}^{T-1}, m_3^{T-1})$
Y_2	$(\hat{m}_{10}^1, \hat{m}_{11}^1, \hat{m}_3^1)$	$(\hat{m}_{10}^2, \hat{m}_{11}^2, \hat{m}_3^2)$	\rightarrow	$(\hat{m}_{10}^{T-1}, \hat{m}_{11}^{T-1}, \hat{m}_3^{T-1})$	\emptyset
X_2	$\mathbf{x}_2^N(m_2^1 1, 1)$	$\mathbf{x}_2^N(m_2^2 \hat{m}_{10}^1, \hat{m}_3^1)$	\rightarrow	$\mathbf{x}_2^N(m_2^{T-1} \hat{m}_{10}^{T-2}, \hat{m}_3^{T-2})$	$\mathbf{x}_2^N(m_2^T \hat{m}_{10}^{T-1}, \hat{m}_3^{T-1})$
Y_3	$(1, \hat{m}_2^1, 1)$	$(\hat{m}_{10}^1, \hat{m}_2^2, \hat{m}_3^1)$	\leftarrow	$(\hat{m}_{10}^{T-2}, \hat{m}_2^{T-1}, \hat{m}_3^{T-2})$	$(\hat{m}_{10}^{T-1}, \hat{m}_2^T, \hat{m}_3^{T-1})$

inner bound (9) for the without D2D case when $R_3 = 0$.

It turns out that using one of the two special cases according to the channel condition can already achieve the capacity to within a constant gap for the scalar Gaussian case, as shown in Section III-D

D. Converse for Discrete Memoryless Model with D2D

The existing works [18], [19] on the FD relay broadcast channel use auxiliary “genie” variables to improve the cut-set bound. Similarly, with the aid of genie, [17] enhances the cut-set bound for the case with relay-to-destination side message. As compared to [17], we provide two improvements. First, we further tighten the genie-aided bound by using more suitable auxiliary variables. Second, we propose a new upper bound on $R_1 + R_2 + R_3$ that improves the cut-set bound. Our converse is specified in the following.

Theorem 4: For the discrete memoryless FD cellular network with D2D, any achievable rate triple (R_1, R_2, R_3) must satisfy

$$R_1 \leq \min\{I(U; Y_2|X_2), I(X_1; Y_2, Y_3|V, X_2)\}, \quad (29a)$$

$$R_2 \leq I(X_2; Y_3|X_1), \quad (29b)$$

$$R_3 \leq \min\{I(X_1; Y_2, Y_3|U, X_2), I(V; Y_2, Y_3|X_2)\}, \quad (29c)$$

$$R_1 + R_3 \leq I(X_1; Y_2, Y_3|X_2), \quad (29d)$$

$$R_2 + R_3 \leq I(X_1, X_2; Y_3), \quad (29e)$$

$$R_1 + R_2 + R_3 \leq I(X_1; Y_2, Y_3|X_2) + I(X_2; Y_3), \quad (29f)$$

for some joint pmf $p(u, v, x_1, x_2)$, where the cardinalities of the auxiliary variables can be bounded by $|\mathcal{U}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_1| + 1$ and $|\mathcal{V}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_1| + 1$.

Proof: Observe that those bounds in (29a)–(29e) without U or V are directly from the cut-set bound. The existing work [17] assumes a genie that provides $U'_n = (Y_2^{n-1}, Y_3^{n-1})$ and $V'_n = M_3$ to node 1 and node 2. In contrast, we propose a different genie that provides $U_n = (M_1, M_2, Y_2^{n-1}, Y_3^{n-1})$ to node 1 and node 3, and provides $V_n = (M_2, M_3, Y_2^{n-1}, Y_3^{n-1})$ to node 1 and node 2. This new use of genie yields a tighter outer bound.

The upper bound (29f) on the sum rate is obtained as follows. Considering node 2 and node 3 as two receivers, we follow Sato’s approach in [22] and assume that they could fully coordinate in their decoding with the aid of genie. Considering node 2 as the transmitter of m_2 , we introduce a genie that provides feedback Y_3^{n-1} to it to improve encoding.

The converse is then established by letting $N \rightarrow \infty$. The complete proof is shown in Appendix A. ■

Remark 4: In contrast to the previous converse in [17], which is not computable, the converse of Theorem 4 can be evaluated. We do so for the Gaussian case in the next section.

IV. SCALAR GAUSSIAN CHANNEL MODEL

The mutual information bounds for the discrete memoryless channel model can be carried over to the Gaussian case. We now evaluate the achievability and converse for the scalar Gaussian channel model under power constraints.

A. Achievability for Scalar Gaussian Model

We first compute the mutual information inner bound (9).

Proposition 2: For the scalar Gaussian FD cellular network without D2D, a rate pair (R_1, R_2) is achievable if it is in the convex hull of

$$R_1 \leq C\left(\frac{(b+c)|g_{21}|^2 P_1}{\sigma^2}\right), \quad (30a)$$

$$R_2 \leq C\left(\frac{e|g_{32}|^2 P_2}{\sigma^2 + c|g_{31}|^2 P_1}\right), \quad (30b)$$

$$R_1 + R_2 \leq C\left(\frac{(a+b)|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\sqrt{ad}}{\sigma^2 + c|g_{31}|^2 P_1}\right) + C\left(\frac{c|g_{21}|^2 P_1}{\sigma^2}\right), \quad (30c)$$

over the nonnegative parameters (a, b, c, d, e) with $a+b+c \leq 1$ and $d+e \leq 1$, where

$$J = 2|g_{31}g_{32}|\sqrt{P_1 P_2}. \quad (31)$$

Proof: Generate the codebooks $\mathbf{u}^N(m_{10}^{t-1})$, $\tilde{\mathbf{v}}^N(m_{11}^t)$, $\mathbf{w}_1^N(m_{11}^t)$, and $\mathbf{w}_2^N(m_2^t)$, all according to $\mathcal{CN}(0, 1)$ in an i.i.d. fashion. In block $t \in [1 : T]$, node 1 transmits

$$\mathbf{x}_1^N(t) = \mathbf{v}^N(t) + \sqrt{cP_1}\mathbf{w}_1^N(m_{11}^t), \quad (32)$$

where

$$\mathbf{v}^N(t) = \sqrt{aP_1}\mathbf{u}^N(m_{10}^{t-1}) + \sqrt{bP_1}\tilde{\mathbf{v}}^N(m_{11}^t). \quad (33)$$

In block t , upon learning \hat{m}_{10}^{t-1} from the previous block $t-1$, node 2 transmits

$$\mathbf{x}_2^N(t) = \sqrt{dP_2}\mathbf{u}^N(\hat{m}_{10}^{t-1}) + \sqrt{eP_2}\mathbf{w}_2^N(m_2^t). \quad (34)$$

Plugging the above setting into (9) gives (30). Furthermore, a convex hull is obtained by time sharing across different choices of (a, b, c, d, e) . ■

Remark 5: We remark that an alternative way to achieve the inner bound (30) is by using the binning strategy, i.e., by partitioning m_{10}^t into bins. The bin index is transmitted by node 1, then relayed by node 2, and used by node 3 in the decoding of m_{10}^t . This binning scheme can be realized in practice by means of hybrid automatic repeat request (HARQ) [23] in which the bin index is basically the parity bits of m_{10}^t .

We show that the above achievability coincides with the capacity under a *very strong interference* regime.

Theorem 5: For the scalar Gaussian FD cellular network without D2D, in the *very strong interference* regime, i.e., when $|g_{31}|^2 \geq |g_{21}|^2(1 + |g_{32}|^2)$, the capacity region of the rate pair (R_1, R_2) is

$$R_1 \leq C\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right), \quad (35a)$$

$$R_2 \leq C\left(\frac{|g_{32}|^2 P_2}{\sigma^2}\right). \quad (35b)$$

Proof: The achievability is verified directly by setting $a = c = d = 0$ and $b = e = 1$ in (30); note that (30c) becomes redundant if $|g_{31}|^2 \geq |g_{21}|^2(1 + |g_{32}|^2)$. The converse directly follows from the cut-set bound. ■

We now consider the D2D case. Instead of the full set of mutual information bounds in Theorem 3, we evaluate the two simpler inner bounds in (27) and (28), as in the following.

Proposition 3 (D2D Rate Splitting): For the scalar Gaussian FD cellular network with D2D, a rate triple (R_1, R_2, R_3) is achievable if it satisfies

$$R_1 \leq C\left(\frac{b|g_{21}|^2 P_1}{\sigma^2 + c|g_{21}|^2 P_1}\right), \quad (36a)$$

$$R_2 \leq C(e|g_{32}|^2 P_2 / \sigma^2), \quad (36b)$$

$$R_1 + R_3 \leq C\left(\frac{b|g_{21}|^2 P_1}{\sigma^2 + c|g_{21}|^2 P_1}\right) + C\left(\frac{c|g_{31}|^2 P_1}{\sigma^2}\right), \quad (36c)$$

$$R_1 + R_2 + R_3 \leq \min \left\{ C\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\sqrt{ad}}{\sigma^2}\right), \right. \\ \left. C\left(\frac{c|g_{31}|^2 P_1 + e|g_{32}|^2 P_2}{\sigma^2}\right) \right. \\ \left. + C\left(\frac{b|g_{21}|^2 P_1}{\sigma^2 + c|g_{21}|^2 P_1}\right) \right\}, \quad (36d)$$

for some nonnegative parameters (a, b, c, d, e) with $a + b + c \leq 1$ and $d + e \leq 1$, where J is defined in (31).

Proof: Generate the codebooks $\mathbf{w}_2^N(m_2^t)$, $\mathbf{w}_3^N(m_{33}^t)$, $\tilde{\mathbf{v}}^N(m_{11}^t, m_{30}^t)$, and $\mathbf{u}^N(m_{11}^{t-1}, m_{30}^t)$, all according to $\mathcal{CN}(0, 1)$ in an i.i.d. fashion. In block $t \in [1 : T]$, node 1 transmits

$$\mathbf{x}_1^N(t) = \mathbf{v}^N(t) + \sqrt{cP_1}\mathbf{w}_3^N(m_{33}^t), \quad (37)$$

where

$$\mathbf{v}^N(t) = \sqrt{aP_1}\mathbf{u}^N(m_{11}^{t-1}, m_{30}^{t-1}) + \sqrt{bP_1}\tilde{\mathbf{v}}^N(m_{11}^t, m_{30}^t). \quad (38)$$

In block t , with $(\hat{m}_{11}^{t-1}, \hat{m}_{30}^{t-1})$ obtained from the previous block $t - 1$, node 2 transmits

$$\mathbf{x}_2^N(t) = \sqrt{dP_2}\mathbf{u}^N(\hat{m}_{11}^{t-1}, \hat{m}_{30}^{t-1}) + \sqrt{eP_2}\mathbf{w}_2^N(m_2^t). \quad (39)$$

Substituting the above setting in (27) gives (36). ■

The mutual information inner bound (28) can be computed similarly, as stated below without proof.

Proposition 4 (Uplink Rate Splitting): For the scalar Gaussian FD cellular network with D2D, a rate triple (R_1, R_2, R_3) is achievable if it satisfies

$$R_2 \leq C\left(\frac{e|g_{32}|^2 P_2}{\sigma^2 + c|g_{31}|^2 P_1}\right), \quad (40a)$$

$$R_1 + R_3 \leq C\left(\frac{(b+c)|g_{21}|^2 P_1}{\sigma^2}\right), \quad (40b)$$

$$R_2 + R_3 \leq C\left(\frac{(a+b)|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\sqrt{ad}}{\sigma^2 + c|g_{31}|^2 P_1}\right), \quad (40c)$$

$$R_1 + R_2 + R_3 \leq C\left(\frac{(a+b)|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\sqrt{ad}}{\sigma^2 + c|g_{31}|^2 P_1}\right) \\ + C\left(\frac{c|g_{21}|^2 P_1}{\sigma^2}\right), \quad (40d)$$

for some nonnegative parameters (a, b, c, d, e) with $a + b + c \leq 1$ and $d + e \leq 1$, where J is defined in (31).

Although the above two inner bounds are two special cases of Theorem 3, it turns out that they suffice to attain the capacity of the scalar Gaussian channel case to within one bit, as shown in Section IV-C.

B. Converse for Scalar Gaussian Model

We now compute the outer bounds for the scalar Gaussian FD cellular network. The following proposition shows an evaluation of the outer bound in Theorem 2 for the case without D2D.

Proposition 5: For the Gaussian FD cellular network without D2D, any achievable rate pair (R_1, R_2) must satisfy

$$R_1 \leq C\left(\frac{(1-\rho^2)|g_{21}|^2 P_1}{\sigma^2}\right), \quad (41a)$$

$$R_2 \leq C\left(\frac{(1-\rho^2)|g_{32}|^2 P_2}{\sigma^2}\right), \quad (41b)$$

$$R_1 + R_2 \leq C\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\rho}{\sigma^2}\right) \\ + C\left(\frac{(1-\rho^2)|g_{21}|^2 P_1}{\sigma^2 + (1-\rho^2)|g_{31}|^2 P_1}\right), \quad (41c)$$

for some parameter $\rho \in [-1, 1]$, where J is defined in (31).

Proof: Let $\rho = \frac{1}{\sqrt{P_1 P_2}} \mathbb{E}[X_1 X_2]$ and observe that the correlation coefficient $\rho \in [-1, 1]$. We first evaluate the upper bound (15a) on R_1 :

$$R_1 \leq I(X_1; Y_2 | X_2) \\ = h(g_{21}X_1 + Z_2 | X_2) - h(Z_2) \\ \leq \log_2 \left(\frac{1}{\sigma^2} \text{Var}(g_{21}X_1 | X_2) \right) \\ \stackrel{(a)}{\leq} C\left(\frac{1}{\sigma^2} \left(\mathbb{E}[|g_{21}X_1|^2] - \frac{\mathbb{E}^2[|g_{21}X_1 X_2|]}{\mathbb{E}[|X_2|^2]} \right) \right) \\ = C\left(\frac{(1-\rho^2)|g_{21}|^2 P_1}{\sigma^2}\right). \quad (42)$$

$$R_1 \leq \min \left\{ \mathcal{C} \left(\frac{(1-\rho^2)|g_{21}|^2 P_1}{\sigma^2 + \alpha(1-\rho^2)|g_{21}|^2 P_1} \right), \mathcal{C} \left(\frac{\beta(1-\rho^2)(|g_{21}|^2 + |g_{31}|^2) P_1}{\sigma^2} \right) \right\}, \quad (46a)$$

$$R_2 \leq \mathcal{C} \left(\frac{(1-\rho^2)|g_{32}|^2 P_2}{\sigma^2} \right), \quad (46b)$$

$$R_3 \leq \min \left\{ \mathcal{C} \left(\frac{\alpha(1-\rho^2)(|g_{21}|^2 + |g_{31}|^2) P_1}{\sigma^2} \right), \mathcal{C} \left(\frac{(1-\rho^2)(|g_{21}|^2 + |g_{31}|^2) P_1}{\sigma^2 + \beta(1-\rho^2)(|g_{21}|^2 + |g_{31}|^2) P_1} \right) \right\}, \quad (46c)$$

$$R_1 + R_3 \leq \mathcal{C} \left(\frac{(1-\rho^2)(|g_{21}|^2 + |g_{31}|^2) P_1}{\sigma^2} \right), \quad (46d)$$

$$R_2 + R_3 \leq \mathcal{C} \left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\rho}{\sigma^2} \right), \quad (46e)$$

$$R_1 + R_2 + R_3 \leq \mathcal{C} \left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\rho}{\sigma^2} \right) + \mathcal{C} \left(\frac{(1-\rho^2)|g_{21}|^2 P_1}{\sigma^2 + (1-\rho^2)|g_{31}|^2 P_1} \right). \quad (46f)$$

The proof of step (a) in (42) is as follows. Suppose that we wish to estimate an unknown quantity $S = g_{21}X_1$ based on the observation $Z = X_2$ under a minimum mean squared-error (MMSE) criteria. The optimal MMSE estimator is $\hat{S} = \mathbb{E}[S|Z]$, and the minimum MMSE equals to $\text{Var}(g_{21}X_1|X_2)$. If we further restrict the estimator of S to be a linear function of Z , then the corresponding linear MMSE must be greater than or equal to $\text{Var}(g_{21}X_1|X_2)$. Since the linear MMSE can be computed analytically as

$$\begin{aligned} \text{LMMSE} &= \text{Var}(S) - \text{Cov}(S, Z) \cdot \text{Var}(Z)^{-1} \cdot \text{Cov}(S, Z)^H \\ &= \mathbb{E}[|g_{21}X_1|^2] - \frac{\mathbb{E}^2[|g_{21}X_1X_2|]}{\mathbb{E}[|X_2|^2]}. \end{aligned} \quad (43)$$

Thus, step (a) in (42) is established by using the above LMMSE as an upper bound on $\text{Var}(g_{21}X_1|X_2)$.

Likewise, let $S = Y_2$ and $Z = (X_2, Y_3)$. The linear MMSE of this case is

$$\begin{aligned} \text{LMMSE} &= \text{Var}(Y_2) - \text{Cov}(Y_2, [X_2, Y_3]) \cdot \text{Var}^{-1}([X_2, Y_3]) \\ &\quad \cdot \text{Cov}^H(Y_2, [X_2, Y_3]), \end{aligned} \quad (44)$$

with which we can evaluate (15c) as follows:

$$\begin{aligned} R_1 + R_2 &\leq I(X_1; Y_2, Y_3|X_2) + I(X_2; Y_3) \\ &= h(Y_2|Y_3, X_2) - h(Z_2, Z_3) + h(Y_3) \\ &\leq \log_2 \left(\frac{1}{\sigma^2} \text{Var}(Y_2|Y_3, X_2) \right) + \log_2 \left(\frac{1}{\sigma^2} \text{Var}(Y_3) \right) \\ &\stackrel{(b)}{\leq} \log_2 \left(\frac{\text{LMMSE}}{\sigma^2} \right) + \log_2 \left(\frac{1}{\sigma^2} \text{Var}(Y_3) \right) \\ &= \mathcal{C} \left(\frac{(1-\rho^2)|g_{21}|^2 P_1}{\sigma^2(1-\rho^2)|g_{31}|^2 P_1} \right) \\ &\quad + \mathcal{C} \left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\rho}{\sigma^2} \right), \end{aligned} \quad (45)$$

where (b) follows by the aforementioned property of MMSE. The upper bound on R_2 can be obtained similarly. ■

The following proposition evaluates the outer bound (15) for the scalar Gaussian FD cellular network with D2D.

Proposition 6: For the scalar Gaussian FD cellular network with D2D, any achievable rate triple (R_1, R_2, R_3) must satisfy (46) as displayed at the top of the page, for some parameters

$\alpha, \beta \in [0, 1]$ and $\rho \in [-1, 1]$, where J is defined in (31).

Proof: Please see Appendix B. ■

By setting $\alpha = 0$ and $\beta = 1$ in (46a), $\alpha = 1$ and $\beta = 0$ in (46b), and $\rho = 1$ throughout, also ignoring the bound (46c) on R_3 , we obtain a simpler outer bound:

Corollary 3: For the scalar Gaussian FD cellular network with D2D, any achievable rate triple (R_1, R_2, R_3) must satisfy

$$R_1 \leq \mathcal{C} \left(\frac{|g_{21}|^2 P_1}{\sigma^2} \right), \quad (47a)$$

$$R_2 \leq \mathcal{C} \left(\frac{|g_{32}|^2 P_2}{\sigma^2} \right), \quad (47b)$$

$$R_1 + R_3 \leq \mathcal{C} \left(\frac{(|g_{21}|^2 + |g_{31}|^2) P_1}{\sigma^2} \right), \quad (47c)$$

$$R_2 + R_3 \leq \mathcal{C} \left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J}{\sigma^2} \right), \quad (47d)$$

$$\begin{aligned} R_1 + R_2 + R_3 &\leq \mathcal{C} \left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J}{\sigma^2} \right) \\ &\quad + \mathcal{C} \left(\frac{|g_{21}|^2 P_1}{\sigma^2 + |g_{31}|^2 P_1} \right), \end{aligned} \quad (47e)$$

where J is defined in (31).

We see in the next section that the above simpler outer bound suffices to establish an approximate capacity region of the scalar Gaussian FD cellular network.

C. Capacity Region to Within One Bit

First we define the notion of the capacity region to within a constant gap.

Definition 1: An achievable rate region $\mathcal{R} \subseteq \mathbb{R}_+^k$ is within a constant gap $\delta \geq 0$ of the capacity region \mathcal{C} if $((C_1 - \delta)^+, \dots, (C_k - \delta)^+) \in \mathcal{R}$ for any $(C_1, \dots, C_k) \in \mathcal{C}$, where $(\cdot)^+ = \max\{\cdot, 0\}$.

Theorem 6: For the scalar Gaussian FD cellular network with D2D, the achievable rate region of Proposition 3 is within 1 bit of the capacity region under the condition $|g_{31}| \geq |g_{21}|$; the achievable rate region of Proposition 4 is within 1 bit of the capacity region under the condition $|g_{31}| < |g_{21}|$. Hence, the achievable rate region of Theorem 3 is within 1 bit of the capacity region.

Proof: The key is to set the parameters of the inner bounds properly. We set $a = d = 0$, $b = 1 - c$, $e = 1$, and $c = \min\{1, \sigma^2/(|g_{21}|^2 P_1)\}$ in Proposition 3, and set $a = d = 0$, $b = 1 - c$, $e = 1$, and $c = \min\{1, \sigma^2/(|g_{31}|^2 P_1)\}$ in Proposition 4. The details are provided in Appendix C. ■

Remark 6: The constant-gap optimality stated in the above theorem can be carried over to the *partially cooperative relay broadcast channel* of [18], [19] since it is a special case of our channel model when $R_2 = 0$.

Remark 7: In proving the constant-gap optimality, the newly introduced upper bound on $R_1 + R_2 + R_3$ plays a key role, whereas the cut-set bound used in [17] is not tight enough to determine the approximate capacity. We demonstrate this point numerically in Fig. 8 in Section VI.

The use of different rate-splitting strategies depending on the channel condition is crucial in proving the above result; using either of the two strategies alone does give a bounded gap to the capacity region. This is because suppose the D2D channel g_{31} is much stronger than the uplink channel g_{21} , we would want to let node 3 decode the entire m_1 so it will benefit from interference cancellation; in this case, m_1 ought not to be split. Likewise, we would not want to split m_3 if g_{21} is much stronger. This is why we should apply the D2D rate splitting strategy if $|g_{31}| \geq |g_{21}|$ and the uplink rate splitting strategy otherwise. This approach turns out to be approximately optimal.

We now specialize the result to the without D2D case.

Theorem 7: For the scalar Gaussian FD cellular network without D2D, the achievable rate region (30) of Proposition 2 is within 1 bit of the capacity region.

Proof: The achievable rate region (30) in Proposition 2 is equivalent to (40) when $R_3 = 0$ (see Remark 3), which is itself obtained by uplink rate splitting and is within a constant gap of 1 bit of the capacity region when $|g_{31}| < |g_{21}|$. So, when $|g_{31}| < |g_{21}|$, (30) must be within 1 bit of the capacity region for the case without D2D.

When $|g_{31}| \geq |g_{21}|$, (36) achieves the capacity region to within 1 bit using the D2D rate splitting strategy. But when $R_3 = 0$, there is no D2D rate to split. In fact, (36) reduces to treating interference as noise, so (30) is in fact larger than (36). Thus, when $|g_{31}| \geq |g_{21}|$, (30) must also be within 1 bit of the capacity region for the case without D2D. ■

Remark 8: It turns out that for the without D2D case, it is actually possible to achieve the capacity region to within a constant gap using a simpler strategy without even using the BS as a relay. However, when there is D2D transmission, relaying plays a crucial role in enhancing R_3 and is necessary for achieving the capacity region to within a constant gap.

Furthermore, we show that the value of the constant gap δ can be further reduced for the without D2D case under a strong interference condition.

Theorem 8: For the scalar Gaussian FD cellular network without D2D, in the *strong interference* regime, defined as the regime in which $|g_{31}| \geq |g_{21}|$, the achievable rate region defined by the set of inequalities in Proposition 2 is within $\frac{1}{2} + \frac{1}{2} \log_2(\frac{\sqrt{2}+1}{2}) \approx 0.6358$ bits of the capacity region.

Proof: For each $\rho \in [-1, 1]$ in (41), we correspondingly let $a = d = \rho^2$, $b = e = 1 - \rho^2$, and $c = 0$ in (30).

Contrasting the resulting achievable rate region (30) with the converse bound (41), we find that the gap is determined by the inner and outer bounds of $R_1 + R_2$, namely (30c) and (41c). Consequently, using the condition $|g_{31}| \geq |g_{21}|$, we obtain the following upper bound on the gap δ to the capacity region:

$$\begin{aligned} 2\delta &\leq 1 + C \left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\rho}{\sigma^2} \right) \\ &\quad - C \left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J\rho^2}{\sigma^2} \right) \\ &= 1 + \log_2 \left(\frac{\lambda + \rho}{\lambda + \rho^2} \right), \end{aligned} \quad (48)$$

where

$$\lambda = \frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + \sigma^2}{J}. \quad (49)$$

By (31), it is clear that $\lambda \geq 1$. Thus, the solution to the following optimization problem is an upper bound on δ :

$$\text{maximize}_{\lambda, \rho} \quad \frac{\lambda + \rho}{\lambda + \rho^2} \quad (50a)$$

$$\text{subject to} \quad \lambda \geq 1, \quad (50b)$$

$$-1 \leq \rho \leq 1. \quad (50c)$$

This problem is quasi-convex and thus can be optimally solved by considering its first-order condition, which give rise to the optimal $\lambda^* = 1$ and $\rho^* = \sqrt{2} - 1$. Substituting (λ^*, ρ^*) back in (48) establishes the theorem. ■

V. VECTOR GAUSSIAN CHANNEL MODEL

The achievable rate region of the FD cellular network can be significantly enlarged by spatial multiplexing and Marton's broadcast coding if the BS and the uplink and downlink users are equipped with multiple antennas. For example, by adding just one more antenna to node 1, it is already possible to transmit the D2D message and the uplink message in orthogonal spatial dimensions, thereby achieving a degree-of-freedom (DoF) gain. Furthermore, while splitting either the uplink message or the D2D message alone already suffices to achieve the approximate capacity of the scalar Gaussian case, this is no longer true for the vector Gaussian case.

This section aims to extend the previous results of the scalar Gaussian channel to the vector Gaussian channel case. For the vector Gaussian case, the main challenge in evaluating its achievable rate region is to decide how to set the auxiliary random variables in Marton's broadcast coding scheme optimally in Theorem 3. In the what follows, we provide three possible achievable rate regions all based on the dirty-paper coding. It is likely that none of these is optimal, but they give achievable rates that can be easily evaluated and implemented using beamforming and dirty-paper coding. Here, m_1 is split into (m_{10}, m_{11}) and m_3 split into (m_{30}, m_{33}) . The three achievable rate regions have the same form as the inner bound in (17), but differ in what is treated as dirt and what is treated as noise, and so differ in the values of (μ_1, \dots, μ_7) in (18).

1) *Treating m_{33} as Dirt:* We treat m_{33} as the dirt in the encoding of m_{11} so that node 2 can decode m_{11} as if the interference from m_{33} does not exist; the uplink transmission

$$\Psi = (\mathbf{G}_{31}\Sigma_d + \mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H\mathbf{Q}^H)(\Sigma_d + \mathbf{Q}\mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H\mathbf{Q}^H)^{-1}(\mathbf{G}_{31}\Sigma_d + \mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H\mathbf{Q}^H)^H. \quad (55)$$

has priority in this scheme. The resulting mutual information terms in (18) are computed as

$$\mu_1 = \log \frac{|\Sigma_c + \mathbf{Q}\mathbf{G}_{21}\Sigma_d\mathbf{G}_{21}^H\mathbf{Q}^H|}{|\Sigma_c|}, \quad (51a)$$

$$\mu_2 = \log |I + 1/\sigma^2 \cdot \mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H| + \mu_1, \quad (51b)$$

$$\mu_3 = \log \frac{|\sigma^2 I + \mathbf{G}_{21}(\Sigma_b + \Sigma_c + \Sigma_d)\mathbf{G}_{21}^H|}{|\sigma^2 I + \mathbf{G}_{21}(\Sigma_c + \Sigma_d)\mathbf{G}_{21}^H|} + \mu_2, \quad (51c)$$

$$\mu_4 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|}{|\sigma^2 I + \mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H|}, \quad (51d)$$

$$\mu_5 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H|}{|\sigma^2 I + \mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H|}, \quad (51e)$$

$$\mu_6 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H|}{|\sigma^2 I + \mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H|}, \quad (51f)$$

$$\mu_7 = \log \frac{|\sigma^2 I + \Phi|}{|\sigma^2 I + \mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H|}, \quad (51g)$$

for some $\Lambda_a \in \mathbb{C}^{L_1^+ \times \min\{L_1^+, L_2^+\}}$, $\Lambda_b \in \mathbb{C}^{L_1^+ \times L_1^+}$, $\Lambda_c \in \mathbb{C}^{L_1^+ \times L_1^+}$, $\Lambda_d \in \mathbb{C}^{L_1^+ \times L_1^+}$, $\Lambda_e \in \mathbb{C}^{L_2^+ \times \min\{L_1^+, L_2^+\}}$, and $\Lambda_f \in \mathbb{C}^{L_2^+ \times L_2^+}$ under the power constraints $\text{tr}(\Sigma_a + \Sigma_b + \Sigma_c + \Sigma_d) \leq P_1$ and $\text{tr}(\Sigma_e + \Sigma_f) \leq P_2$, with

$$\mathbf{Q} = \Sigma_c\mathbf{G}_{21}^H(\sigma^2 I + \mathbf{G}_{21}\Sigma_c\mathbf{G}_{21}^H)^{-1}, \quad (52)$$

and

$$\Phi = \mathbf{G}_{31}(\Sigma_a + \Sigma_b + \Sigma_c + \Sigma_d)\mathbf{G}_{31}^H + \mathbf{G}_{32}(\Sigma_e + \Sigma_f)\mathbf{G}_{32}^H + \mathbf{G}_{31}\Lambda_a\Lambda_e^H\mathbf{G}_{32}^H + \mathbf{G}_{32}\Lambda_e\Lambda_a^H\mathbf{G}_{31}^H, \quad (53)$$

where $\Sigma_i = \Lambda_i\Lambda_i^H$, $\forall i \in \{a, b, c, d, e, f\}$. Node 1 uses Λ_a , Λ_b , Λ_c , and Λ_d as the beamformers for the relay message $(m_{10}^{t-1}, m_{10}^{t-1})$, the common message (m_{10}^t, m_{10}^t) , the uplink private message m_{11}^t , and the D2D private message m_{33}^t , respectively, while node 2 uses Λ_e and Λ_f as the beamformers for the common message $(m_{10}^{t-1}, m_{10}^{t-1})$ and the downlink message m_2^t , respectively, where $t \in [1 : T]$ represents the block index in the block Markov coding.

2) *Treating m_{11} as Dirt and Subtracting m_2 First at Node 3:* We now treat m_{11} as the dirt in the encoding of m_{33} so that node 3 can decode m_{33} as if the interference from m_{11} does not exist. We further assume that Node 3 decodes m_2 then subtract it first. Thus, the D2D transmission has priority in this scheme. Using the same variables $(\Lambda_a, \Lambda_b, \Lambda_c, \Lambda_d, \Lambda_e, \Lambda_f)$ as in the previous scheme, we evaluate the mutual information terms in (18) as

$$\mu_1 = \log \frac{|\Sigma_d + \mathbf{Q}\mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H\mathbf{Q}^H|}{|\Sigma_d|}, \quad (54a)$$

$$\mu_2 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|}{|\sigma^2 I + \mathbf{G}_{31}\Sigma_d\mathbf{G}_{31}^H|}, \quad (54b)$$

$$\mu_3 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_b + \Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|}{|\sigma^2 I + \mathbf{G}_{31}\Sigma_d\mathbf{G}_{31}^H|}, \quad (54c)$$

$$\mu_4 = \log |I + 1/\sigma^2 \cdot \mathbf{G}_{31}\Sigma_d\mathbf{G}_{31}^H| + \mu_1, \quad (54d)$$

$$\mu_5 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H|}{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|} + \mu_4 - \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H - \Psi|}{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H - \Psi|}, \quad (54e)$$

$$\mu_6 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H|}{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|} + \mu_4, \quad (54f)$$

$$\mu_7 = \log \frac{|\sigma^2 I + \Phi|}{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|} + \mu_4 - \mu_1, \quad (54f)$$

where a new variable Ψ is defined in (55) as displayed at the top of the page.

3) *Treating m_{11} as Dirt and Treating m_2 as Noise at Node 3:* We still treat m_{11} as the dirt but treat m_2 as the noise in the encoding of m_{33} , so the downlink transmission has priority in this scheme. The mutual information terms in (18) then become

$$\mu_1 = \log \frac{|\Sigma_d + \mathbf{Q}\mathbf{G}_{31}\Sigma_c\mathbf{G}_{31}^H\mathbf{Q}^H|}{|\Sigma_d|}, \quad (55a)$$

$$\mu_2 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|}{|\sigma^2 I + \mathbf{G}_{31}\Sigma_d\mathbf{G}_{31}^H|}, \quad (55b)$$

$$\mu_3 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_b + \Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|}{|\sigma^2 I + \mathbf{G}_{31}\Sigma_d\mathbf{G}_{31}^H|}, \quad (55c)$$

$$\mu_4 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|}{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H - \Psi|}, \quad (55d)$$

$$\mu_5 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H|}{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|} + \mu_4$$

$$- \log \frac{|\sigma^2 I + \mathbf{G}_{31}\Sigma_d\mathbf{G}_{31}^H + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H|}{|\sigma^2 I + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H|} - \mu_1, \quad (55e)$$

$$\mu_6 = \log \frac{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H + \mathbf{G}_{32}\Sigma_f\mathbf{G}_{32}^H|}{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|} + \mu_4, \quad (55f)$$

$$\mu_7 = \log \frac{|\sigma^2 I + \Phi|}{|\sigma^2 I + \mathbf{G}_{31}(\Sigma_c + \Sigma_d)\mathbf{G}_{31}^H|} + \mu_4 - \mu_1, \quad (55f)$$

where the variables are all defined as before.

It is challenging to prove the constant-gap optimality for the above achievable rate regions. But, in the without D2D case, treating m_{33} as dirt method reduces to a vector generalization of Proposition 2, and it can be shown that the resulting achievable rate region reaches the capacity to within $\max\{\min\{L_1^+, L_2^+\}, \min\{L_2^+, L_3^+\}, \frac{1}{2}\min\{L_1^+, L_3^+\}\}$ bits for the vector Gaussian channel model without D2D by setting $\Sigma_d = \sigma^2(\frac{\sigma^2}{P_1}I + \mathbf{G}_{31}^H\mathbf{G}_{31})^{-1}$, $\Sigma_c = P_1I - \Sigma_d$, and $\Sigma_e = \sqrt{P_2}I$, along with $(\Lambda_a, \Lambda_c, \Lambda_e)$ being zero matrices.

VI. NUMERICAL EXAMPLES

To demonstrate the above capacity analyses, we provide some numerical examples of the scalar Gaussian FD cellular

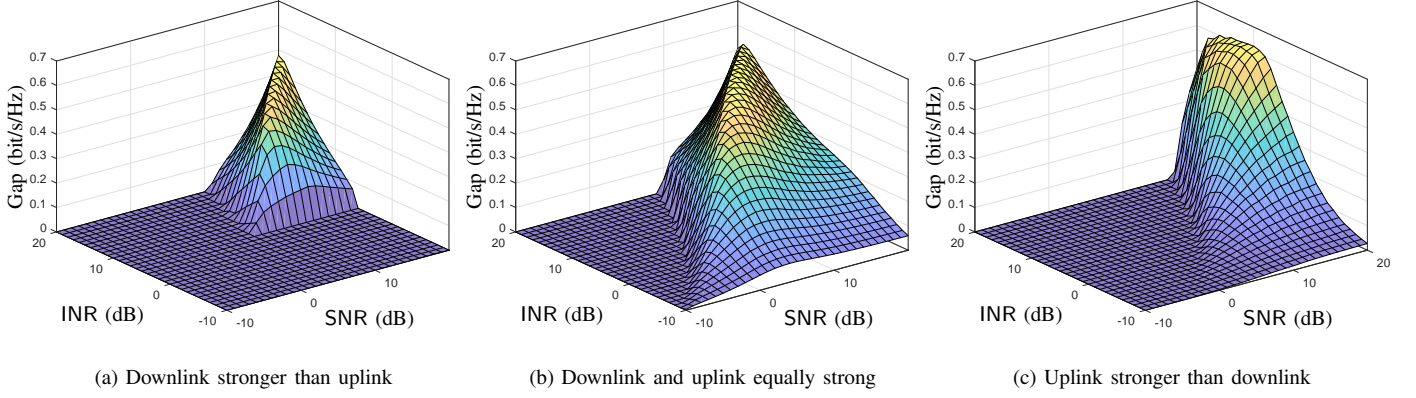


Fig. 3. The gap between the achievable rate region (30) and the converse (41) for the scalar Gaussian FD cellular network without D2D, where the parameters of the bounds are globally optimized by exhaustive search. We consider three different settings: (a) $|g_{32}|^2 P_2 = 5|g_{21}|^2 P_1$; (b) $|g_{32}|^2 P_2 = |g_{21}|^2 P_1$; (c) $5|g_{32}|^2 P_2 = |g_{21}|^2 P_1$.

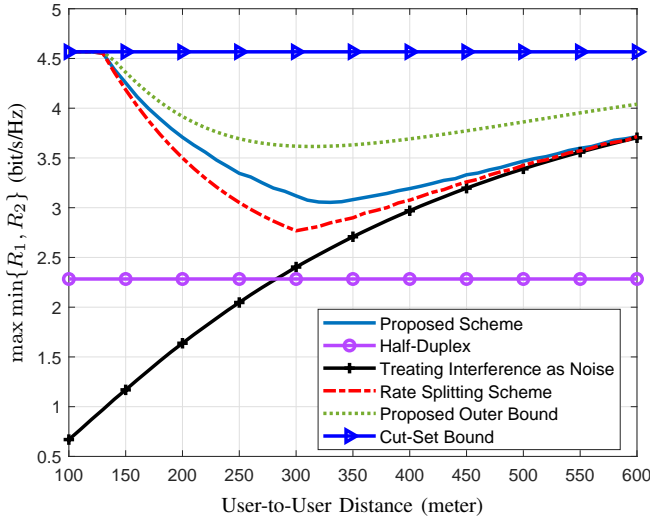


Fig. 4. Symmetric rate $\min\{R_1, R_2\}$ in a Gaussian channel without D2D where the user-to-BS distance is 300 meters.

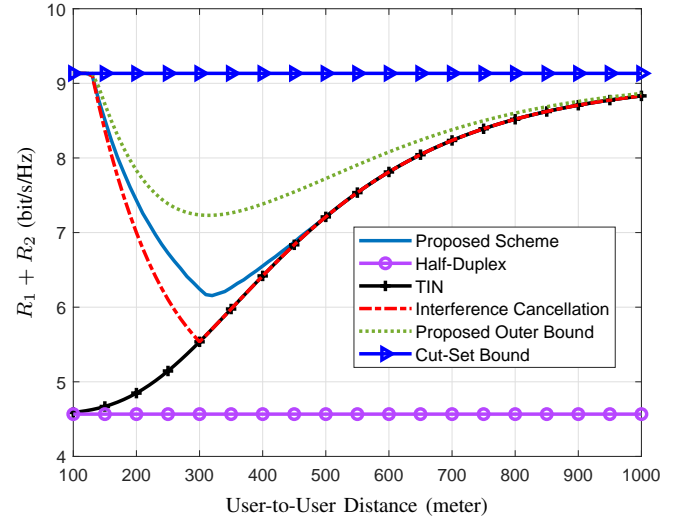


Fig. 5. Sum rate $R_1 + R_2$ in a Gaussian channel without D2D where the user-to-BS distance is 300 meters.

network. Throughout this section, define $\text{SNR} = |g_{21}|^2 P_1 / \sigma^2$ and $\text{INR} = |g_{31}|^2 P_1 / \sigma^2$.

A. Without D2D Case

For the scalar Gaussian channel model without D2D, Theorem 7 shows that the gap δ between the inner bound (30) and the outer bound (41) is at most 1 bit/s/Hz under a particular choice of the parameters (a, b, c, d, e, ρ) . Fig. 3 shows the minimum value of δ under various channel conditions with the parameters (a, b, c, d, e, ρ) globally optimized by exhaustive search. Observe that the gap is equal to zero if INR is sufficiently greater than SNR because of Theorem 5. In all cases, the gap is less than 1 bit as expected.

We further evaluate the data rates in an FD cellular network in the following topology. The BS is at the center of the cell; the uplink/downlink user-to-BS distance is fixed at 300m, the user-to-user distance is set to different values. Let the maximum transmit power spectral density (PSD) be -47

dBm/Hz for both uplink and downlink; we set the PSD of the background noise be -169 dBm/Hz. The channel magnitude is modeled as $-128.1 - 37.6 \log_{10}(\text{dist})$ in dB scale, dist representing the distance (in km). In addition to the inner bound (30), the outer bound (41), and the cut-set bound, we consider the following baseline methods:

- Half-duplex with separate uplink and downlink;
- Treating interference as noise with FD uplink/downlink;
- Rate splitting scheme: The uplink message is split for interference cancellation, but without relaying by the BS.

Fig. 4 compares the symmetric uplink-downlink rate $\min\{R_1, R_2\}$ in the without D2D case. It shows that the proposed BS-aided scheme is more effective than simple interference cancellation, gaining up to about 0.5 bits/s/Hz. Observe that the proposed scheme shows the most benefit when the user-to-user distance is not too close or too far. Moreover, Fig. 5 compares the sum rate of various schemes. It shows that the half-duplex scheme is inferior to all the FD schemes in terms of the total throughput.

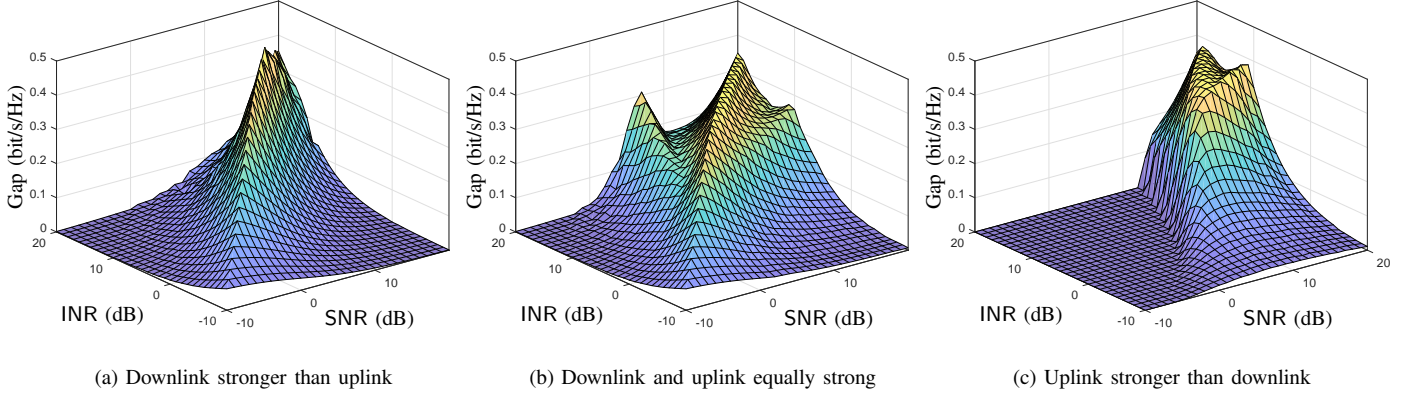


Fig. 6. The gap between the achievable rate region, i.e., the union of (36) and (40), and the converse (46) for the scalar Gaussian FD cellular network with D2D, where the parameters of the bounds are globally optimized by exhaustive search. We consider three different settings: (a) $|g_{32}|^2 P_2 = 5|g_{21}|^2 P_1$; (b) $|g_{32}|^2 P_2 = |g_{21}|^2 P_1$; (c) $5|g_{32}|^2 P_2 = |g_{21}|^2 P_1$.

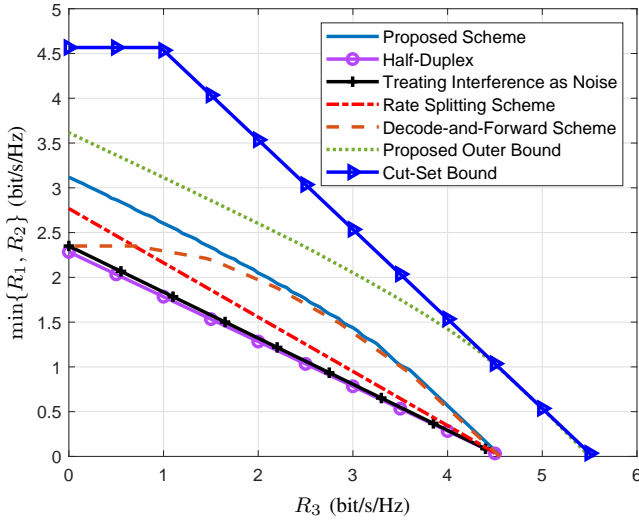


Fig. 7. Trade-off between R_3 and $\min\{R_1, R_2\}$ in the D2D case.

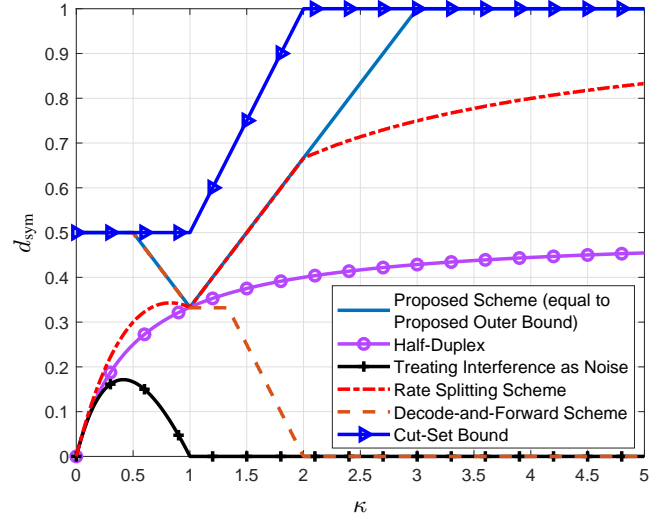


Fig. 8. Symmetric GDoF d_{sym} versus $\kappa = \frac{\log \text{INR}}{\log \text{SNR}}$ in the D2D case.

B. D2D Case

For the scalar Gaussian channel model with D2D, we compare the inner bound, i.e., the union of (36) and (40), with the outer bound (46) to find the minimum gap between them. Again, the parameters of these bounds are optimized via exhaustive search. The minimum gap with respect to different (INR, SNR) pairs is plotted in Fig. 6. We again observe that the gap is always less than 1. Moreover, although we have not determined the exact capacity region for the D2D case, the plot of case (c) shows that the symmetric rate achieved by the proposed scheme is close to the capacity when the cross-channel is sufficiently stronger than the uplink channel and when the uplink channel is stronger than the downlink channel.

We further evaluate the data rates with D2D under the same setting as considered for Fig. 4 except that the user-to-user distance is now fixed at 300 meters. The three baselines are extended to the D2D case by time or frequency division multiplex of cellular and D2D traffic. Moreover, we introduce

a new decode-and-forward benchmark corresponding to the scheme of [17]. Fig. 7 shows the trade-off between the D2D rate and the symmetric uplink-downlink rate. The proposed scheme achieves a larger rate region than the baseline schemes. Moreover, Fig. 7 shows that at user-to-user distance of 300 meters, the D2D rate R_3 can be up to 4.5 bits/s/Hz by using the proposed scheme, whereas the transmission rate from node 1 to node 3 through the BS is only $\min\{R_1, R_2\} = 3.2$ bits/s/Hz in the without D2D case as shown in Fig. 4.

Finally, we compare the asymptotic behavior of the different schemes in terms of the *generalized degree of freedom* (GDoF) [24]. We restrict to the channel parameters for which $|g_{21}|P_1 = |g_{32}|P_2$, then let both INR and SNR go to infinity, while keeping their ratio in the log-domain constant, i.e., fixing $\kappa = \frac{\log \text{INR}}{\log \text{SNR}}$. The symmetric GDoF is defined as

$$d_{\text{sym}} \triangleq \lim_{\substack{\text{SNR} \rightarrow \infty \\ \text{INR} \rightarrow \infty}} \frac{\min\{R_1, R_2, R_3\}}{\log \text{SNR}}. \quad (57)$$

This definition is motivated by regarding the FD cellular

channel model as a Z-interference channel that has a one-way interference from uplink to downlink. Here, SNR and INR both tend to infinity, while $\text{INR} = \text{SNR}^\kappa$ for some fixed $\kappa \in [0, \infty)$.

Fig. 8 shows the GDoF achieved by the different schemes. The GDoF of the proposed scheme is optimal since the proposed scheme attains the capacity of scalar Gaussian FD network to within a constant gap. Now, because there exists a gap between the optimal GDoF curve and the cut-set bound, this means cut-set bound can be arbitrarily loose.

Observe that the optimal GDoF curve consists of four segments, i.e., $[0, 0.5)$, $[0.5, 1)$, $[1, 3)$, and $[3, \infty)$ with respect to κ . When $\kappa \in [0, 0.5)$, because the channel between node 1 and node 3 is very weak, it is useless in terms of GDoF, thus the optimal GDoF does not change with κ in this interval. When $\kappa \in [0.5, 1)$, the uplink rate splitting scheme of Proposition 4 attains the optimal GDoF. When $\kappa \in [1, 3)$, the D2D rate splitting scheme of Proposition 3 attains the optimal GDoF. Furthermore, when $\kappa \in [3, \infty)$, the channel between node 1 and node 3 is strong enough to let node 3 sequentially recover (m_1, m_2, m_3) by successive cancellation without relaying at the BS. In this situation, node 1 allocates power $(1 - \text{SNR}^{1-\kappa})P_1$ to the transmission of m_1 and allocates power $\text{SNR}^{1-\kappa}P_1$ to the transmission of m_3 , while node 2 transmits m_2 at the full power P_2 . After subtracting the self-interference, node 2 removes m_3 and then decodes m_1 via the successive cancellation. Node 3 first decodes m_1 , then m_2 , and finally m_3 via the successive cancellation. As a result, each transmission attains a GDoF of 1. Fig. 8 shows that all the other schemes are suboptimal in terms of GDoF.

VII. CONCLUSION

This paper analyzes the maximum transmission rates of the uplink message, the downlink message, and the D2D message in a wireless cellular network with the FD BS and two half-duplex user terminals. This FD cellular network can be modeled as a relay broadcast channel with side message. We propose new strategies to enlarge the existing achievable rate regions. A crucial component of the proposed new strategies is to use the BS as a relay to facilitate cancelling interference. We further provide novel converse results that are strictly tighter than the cut-set bound and the state-of-the-art outer bound. For the without D2D case, our proposed schemes achieve the capacity regions to within a constant gap for the Gaussian channel case, even when the BS and users are deployed with multiple antennas. For the D2D case, a constant-gap optimality is established for the scalar Gaussian channel case.

APPENDIX A PROOF OF THEOREM 4

We first verify the upper bound (29f) on $R_1 + R_2 + R_3$:

$$\begin{aligned}
 & N(R_1 + R_2 + R_3 - \epsilon_N) \\
 & \leq I(M_1; Y_2^N) + I(M_2; Y_3^N) + I(M_3; Y_3^N) \\
 & \leq I(M_1, M_3; Y_2^N, Y_3^N) + I(M_2; Y_3^N) \\
 & \stackrel{(a)}{=} I(M'_1; Y_2^N, Y_3^N) + I(M_2; Y_3^N) \\
 & \stackrel{(b)}{\leq} NI(X_1; Y_2, Y_3 | X_2) + NI(X_2; Y_3), \tag{58}
 \end{aligned}$$

where we use M'_1 to denote (M_1, M_3) and thus (a) holds; (b) follows by treating M'_1 as M_1 in (16).

We deal with those inequalities having U or V in the rest of the proof. The cut-set bound on R_1 is $R_1 \leq I(X_1; Y_2 | X_2)$, but this must be loose because some portion of X_1 carrying the D2D message m_3 is independent of m_1 . (We do not have this issue in the without D2D scenario.) Therefore, we introduce an auxiliary variable U_n to represent the part of X_1 “related” to R_1 , as follows:

$$U_n = (M_1, M_2, Y_2^{n-1}, Y_3^{n-1}), \quad n \in [1 : N], \tag{59}$$

where M_2 , Y_2^{n-1} , and Y_3^{n-1} are due to the cut-set bounds on R_1 and R_3 . Note that M_3 is excluded from U_n . Another auxiliary variable V_n is similarly motivated:

$$V_n = (M_2, M_3, Y_2^{n-1}, Y_3^{n-1}), \quad n \in [1 : N]. \tag{60}$$

We then apply U and V to improve the cut-set bound. First, consider the bound on R_1 with U_n :

$$\begin{aligned}
 N(R_1 - \epsilon_N) & \leq I(M_1; Y_2^N, X_2^N) \\
 & \leq I(M_1; Y_2^N, X_2^N | M_2) \\
 & \stackrel{(c)}{=} I(M_1; Y_2^N | M_2) \\
 & = \sum_{n=1}^N I(M_1; Y_{2n} | M_2, Y_2^{n-1}) \\
 & \stackrel{(d)}{=} \sum_{n=1}^N I(M_1; Y_{2n} | M_2, Y_2^{n-1}, X_{2n}) \\
 & \leq \sum_{n=1}^N I(M_1, M_2, Y_2^{n-1}, Y_3^{n-1}; Y_{2n} | X_{2n}) \\
 & = \sum_{n=1}^N I(U_n; Y_{2n} | X_{2n}) \\
 & \leq NI(U; Y_2 | X_2), \tag{61}
 \end{aligned}$$

where both of (c) and (d) follow since X_{2n} is a function of (M_2, Y_2^{n-1}) . We then use V_n to give another bound on R_1 :

$$\begin{aligned}
 N(R_1 - \epsilon_N) & \leq I(M_1; Y_2^N, X_2^N) \\
 & \leq I(M_1; Y_2^N, X_2^N | M_2, M_3) \\
 & \stackrel{(e)}{=} I(M_1; Y_2^N, Y_3^N | M_2, M_3) \\
 & = \sum_{n=1}^N I(M_1; Y_{2n}, Y_{3n} | M_2, M_3, Y_2^{n-1}, Y_3^{n-1}) \\
 & \stackrel{(f)}{=} \sum_{n=1}^N I(X_{1n}; Y_{2n}, Y_{3n} | M_2, M_3, Y_2^{n-1}, Y_3^{n-1}) \\
 & \stackrel{(g)}{=} \sum_{n=1}^N I(X_{1n}; Y_{2n}, Y_{3n} | X_{2n}, V_n) \\
 & \leq NI(X_1; Y_2, Y_3 | X_2, V), \tag{62}
 \end{aligned}$$

where (e) and (g) follow since X_{2n} is a function of (M_2, Y_2^{n-1}) , (f) is due to the facts that X_{1n} is a function of (M_1, M_3) and that $M_1 \rightarrow X_{1n} \rightarrow (Y_{2n}, Y_{3n})$ form a Markov chain conditioned on $(M_2, M_3, Y_2^{n-1}, Y_3^{n-1})$.

Further, by using U_n , we establish an outer bound on R_3 :

$$\begin{aligned}
N(R_3 - \epsilon_N) &\leq I(M_3; Y_3^N) \\
&\leq I(M_3; Y_2^N, Y_3^N | M_1, M_2) \\
&= \sum_{n=1}^N I(M_3; Y_{2n}, Y_{3n} | M_1, M_2, Y_2^{n-1}, Y_3^{n-1}) \\
&\stackrel{(h)}{=} \sum_{n=1}^N I(X_{1n}, Y_{2n}, Y_{3n} | M_1, M_2, Y_2^{n-1}, Y_3^{n-1}) \\
&= \sum_{n=1}^N I(X_{1n}, Y_{2n}, Y_{3n} | U_n, X_{2n}) \\
&\leq NI(X_1; Y_2, Y_3 | U, X_2), \tag{63}
\end{aligned}$$

where (h) follows by the same reason as for step (f) in (62). Finally, the auxiliary variable V_n gives rise to the following outer bound on R_3 :

$$\begin{aligned}
N(R_3 - \epsilon_N) &\leq I(M_3; Y_3^N) \\
&\leq I(M_3; Y_2^N, Y_3^N | M_2) \\
&= \sum_{n=1}^N I(M_3; Y_{2n}, Y_{3n} | M_2, Y_2^{n-1}, Y_3^{n-1}) \\
&= \sum_{n=1}^N I(M_3; Y_{2n}, Y_{3n} | M_2, Y_2^{n-1}, Y_3^{n-1}, X_{2n}) \\
&\leq \sum_{n=1}^N I(M_2, M_3, Y_2^{n-1}, Y_3^{n-1}; Y_{2n}, Y_{3n} | X_{2n}) \\
&= \sum_{n=1}^N I(V_n; Y_{2n}, Y_{3n} | X_{2n}) \\
&\leq NI(V; Y_2, Y_3 | X_2). \tag{64}
\end{aligned}$$

Summarizing the above results gives the outer bound in this theorem.

APPENDIX B PROOF OF PROPOSITION 6

We first introduce a useful lemma:

Lemma 1: Letting

$$Y' = \frac{g_{21}Y_2 + g_{31}Y_3}{\sqrt{|g_{21}|^2 + |g_{31}|^2}}, \tag{65}$$

we have $I(X_1; Y_2, Y_3 | U, X_2) = I(X_1; Y' | U, X_2)$.

Proof: Observe that

$$\begin{aligned}
&I(X_1; Y_2, Y_3 | U, X_2) \\
&= I(X_1, Y_2, Y' | U, X_2) \\
&= I(X_1; Y' | U, X_2) + I(X_1; Y_2 | U, X_2, Y'). \tag{66}
\end{aligned}$$

Also, with the shorthand

$$Z'_2 = \frac{|g_{31}|^2 Z_2 - g_{21}g_{31}Z_3}{|g_{21}|^2 + |g_{31}|^2}, \tag{67}$$

we show that

$$\begin{aligned}
I(X_1; Y_2 | U, X_2, Y') &\stackrel{(a)}{\leq} I(X_1; Y_2 | X_2, Y') \\
&= I(X_1; Z'_2 | X_2, Y') \\
&\stackrel{(b)}{=} 0, \tag{68}
\end{aligned}$$

where step (a) follows since $U \rightarrow X_1 \rightarrow Y_2$ form a Markov chain conditioned on (X_2, Y') , step (b) follows since Z'_2 is independent of any of (X_1, X_2, Y') . By the squeeze theorem, we must have $I(X_1; Y_2 | U, X_2, Y') = 0$. Substituting this result back in (66) verifies the lemma. ■

Equipped with the above lemma, we continue to prove Proposition 6. We focus on the inequalities (46a) and (46c) which involve the use of auxiliary variables U and V from Theorem 4. Again, use $\rho \in [-1, 1]$ to denote the correlation coefficient $\frac{1}{\sqrt{P_1 P_2}} \mathbb{E}[X_1 X_2]$. It can be shown that

$$\begin{aligned}
&\log_2(2\pi e \sigma^2) \\
&\leq h(Y' | U, X_1, X_2) \\
&\leq h(Y' | U, X_2) \\
&\leq h(Y' | X_2) \\
&= \log_2 \left(2\pi e (\sigma^2 + (1 - \rho^2)(|g_{21}|^2 + |g_{31}|^2)P_1) \right), \tag{69}
\end{aligned}$$

so there must exist a constant $\alpha \in [0, 1]$ such that

$$\begin{aligned}
&h(Y' | U, X_2) = \\
&\log_2 \left(2\pi e (\sigma^2 + \alpha(1 - \rho^2)(|g_{21}|^2 + |g_{31}|^2)P_1) \right). \tag{70}
\end{aligned}$$

The term with α in the (46c) is obtained as follows:

$$\begin{aligned}
R_3 &\leq I(X_1; Y_2, Y_3 | U, X_2) \\
&\stackrel{(c)}{=} I(X_1; Y' | U, X_2) \\
&= h(Y' | U, X_2) - h(Y' | U, X_1, X_2) \\
&= h(Y' | U, X_2) - h \left(\frac{g_{21}Z_2 + g_{31}Z_3}{\sqrt{|g_{21}|^2 + |g_{31}|^2}} \right) \\
&\stackrel{(d)}{=} C \left(\frac{\alpha(1 - \rho^2)(|g_{21}|^2 + |g_{31}|^2)P_1}{\sigma^2} \right), \tag{71}
\end{aligned}$$

where step (c) follows by Lemma 1 and step (d) follows by (70).

Moreover, we derive that

$$\begin{aligned}
h(Y_2 | U, X_2) &= h(\omega Y' + Z'_2 | U, X_2) \\
&\stackrel{(e)}{\geq} \log_2 \left(2^{h(\omega Y' | U, X_2)} + 2^{h(Z'_2 | U, X_2)} \right) \\
&= \log_2 (1 + \alpha(1 - \rho^2)|g_{21}|^2 P_1), \tag{72}
\end{aligned}$$

where $\omega = \frac{g_{21}}{\sqrt{|g_{21}|^2 + |g_{31}|^2}}$ and Z'_2 is previously defined in (67); step (e) follows by the entropy power inequality (EPI) since $\omega Y' \perp\!\!\!\perp Z'_2$ given (U, X_2) . Consequently, we have

$$\begin{aligned}
R_1 &\leq I(U; Y_2 | X_2) \\
&= h(Y_2 | X_2) - h(Y_2 | U, X_2) \\
&\stackrel{(f)}{\leq} C \left(\frac{(1 - \rho^2)|g_{21}|^2 P_1}{\sigma^2 + \alpha(1 - \rho^2)|g_{21}|^2 P_1} \right), \tag{73}
\end{aligned}$$

where step (f) is due to (72).

Next, we show the upper bounds on R_1 or R_3 that involve the auxiliary variable V . First, we derive the following chain of inequalities:

$$\begin{aligned} I(X_1; Y_2, Y_3 | V, X_2) &\stackrel{(g)}{\leq} I(X_1; Y_2, Y_3 | X_2) \\ &= \mathcal{C}\left(\frac{(1 - \rho^2)(|g_{21}|^2 + |g_{31}|^2)P_1}{\sigma^2}\right), \end{aligned} \quad (74)$$

where step (g) follows since $V \rightarrow X_1 \rightarrow (Y_2, Y_3)$ form a Markov chain conditioned on X_2 . Thus, we are guaranteed to find a constant $\beta \in [0, 1]$ such that

$$\begin{aligned} R_1 &\leq I(X_1; Y_2, Y_3 | V, X_2) \\ &= \mathcal{C}\left(\frac{\beta(1 - \rho^2)(|g_{21}|^2 + |g_{31}|^2)P_1}{\sigma^2}\right). \end{aligned} \quad (75)$$

Finally, we can compute the second mutual information term in (29c) as

$$\begin{aligned} R_3 &\leq I(V; Y_2, Y_3 | X_2) \\ &= I(X_1; Y_2, Y_3 | X_2) - I(X_1; Y_2, Y_3 | V, X_2) \\ &\stackrel{(h)}{=} I(X_1; Y_2, Y_3 | X_2) - \mathcal{C}\left(\frac{\beta(1 - \rho^2)(|g_{21}|^2 + |g_{31}|^2)P_1}{\sigma^2}\right) \\ &\leq \mathcal{C}\left(\frac{(1 - \rho^2)(|g_{21}|^2 + |g_{31}|^2)P_1}{\sigma^2 + \beta(1 - \rho^2)(|g_{21}|^2 + |g_{31}|^2)P_1}\right), \end{aligned} \quad (76)$$

where step (h) is due to the identity in (75). We have established the set of inequalities (46a) and (46c). The verification of the remaining inequalities in (46) is similar to the proof of Proposition 5.

APPENDIX C PROOF OF THEOREM 6

This proof consists of two parts: we first show that the D2D rate splitting method attains the capacity region to within 1 bit when $|g_{21}| < |g_{31}|$, then show that the uplink rate splitting method achieves the same constant-gap when $|g_{21}| \geq |g_{31}|$. For ease of notation, we use $(\varrho_1, \varrho_2, \varrho_3, \varrho_4, \varrho_5)$ to denote each of the terms on the right-hand sides of the inequalities in (47), which correspond to the upper bounds on R_1 , R_2 , $R_1 + R_3$, $R_2 + R_3$, and $R_1 + R_2 + R_3$, respectively.

A. Constant-Gap Optimality of D2D Rate Splitting

We aim to show that the achievable rate region of Proposition 3 is within 1 bit of the capacity region of the scalar Gaussian FD cellular network with D2D when $|g_{21}| < |g_{31}|$. The main idea is to show that the gap is less than or equal to 1 bit between the inner bound (36) and the outer bound (47).

We begin with a special case where $|g_{21}|^2 P_1 \leq \sigma^2$. Since the SNR of the uplink channel from node 1 to node 2 is upper bounded by 1, removing this channel (i.e., setting g_{21} to zero) would cause at most 1 bit loss to each of (R_1, R_2, R_3) . Actually, without the uplink channel, our channel model reduces to a multiple access channel, for which the inner bound (36) coincides with the capacity region if we set $a = b = d = 0$

and $c = e = 1$. Thus, (36) must be within 1 bit of the capacity of the original channel model.

We now assume that $|g_{21}|^2 P_1 > \sigma^2$. When $a = d = 0$, $e = 1$, $c = \frac{\sigma^2}{|g_{21}|^2 P_1}$, and $b = 1 - c$, the inner bound (36) is

$$R_1 \leq \mathcal{C}\left(\frac{|g_{21}|^2 P_1 - \sigma^2}{2\sigma^2}\right), \quad (77a)$$

$$R_2 \leq \mathcal{C}\left(\frac{|g_{32}|^2 P_2}{\sigma^2}\right), \quad (77b)$$

$$R_1 + R_3 \leq \mathcal{C}\left(\frac{|g_{21}|^2 P_1 - \sigma^2}{2\sigma^2}\right) + \mathcal{C}\left(\frac{|g_{31}|^2}{|g_{21}|^2}\right), \quad (77c)$$

$$\begin{aligned} R_1 + R_2 + R_3 &\leq \mathcal{C}\left(\frac{|g_{21}|^2 P_1 - \sigma^2}{2\sigma^2}\right) \\ &\quad + \mathcal{C}\left(\frac{\sigma^2 |g_{31}|^2 / |g_{21}|^2 + |g_{32}|^2 P_2}{\sigma^2}\right), \end{aligned} \quad (77d)$$

$$R_1 + R_2 + R_3 \leq \mathcal{C}\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2}{\sigma^2}\right). \quad (77e)$$

Use $(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ to denote respectively the terms on the right-hand side of (77a)–(77e). We can bound the constant gap δ between (36) and (47) by comparing $(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ with $(\varrho_1, \varrho_2, \varrho_3, \varrho_4, \varrho_5)$. Furthermore, it can be shown that

$$\eta_1 \geq \mathcal{C}\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right) - 1, \quad (78a)$$

$$\eta_2 \geq \mathcal{C}\left(\frac{|g_{32}|^2 P_2}{\sigma^2}\right), \quad (78b)$$

$$\eta_3 \geq \mathcal{C}\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right) + \mathcal{C}\left(\frac{|g_{31}|^2}{|g_{21}|^2}\right) - 1, \quad (78c)$$

$$\min\{\eta_4, \eta_5\} \geq \mathcal{C}\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2}{\sigma^2}\right) - 1. \quad (78d)$$

Thus, with respect to $R_1 + R_3$, the constant gap δ can be bounded by analyzing the difference between (78c) and (47c):

$$\begin{aligned} 2\delta &\leq \varrho_3 - \eta_3 \\ &\leq \mathcal{C}\left(\frac{(|g_{21}|^2 + |g_{31}|^2)P_1}{\sigma^2}\right) - \mathcal{C}\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right) \\ &\quad - \mathcal{C}\left(\frac{|g_{31}|^2}{|g_{21}|^2}\right) + 1 \\ &\stackrel{(a)}{\leq} \mathcal{C}\left(\frac{2|g_{31}|^2 P_1}{\sigma^2}\right) - \mathcal{C}\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right) - \mathcal{C}\left(\frac{|g_{31}|^2}{|g_{21}|^2}\right) + 1 \\ &= \log_2\left(\frac{(|g_{21}|^2 P_1)(\sigma^2 + 2|g_{31}|^2 P_1)}{(\sigma^2 + |g_{21}|^2 P_1)(|g_{21}|^2 + |g_{31}|^2)P_1}\right) + 1 \\ &\stackrel{(b)}{\leq} \log_2(2) + 1 \\ &\leq 2, \end{aligned} \quad (79)$$

where step (a) follows by $|g_{21}| < |g_{31}|$ and step (b) follows by $|g_{21}|^2 P_1 > \sigma^2$. With respect to $R_1 + R_2 + R_3$, we have

$$\begin{aligned} 3\delta &\leq \varrho_5 - \min\{\eta_4, \eta_5\} \\ &\leq \mathcal{C}\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J}{\sigma^2}\right) + \mathcal{C}\left(\frac{|g_{21}|^2 P_1}{\sigma^2 + |g_{31}|^2 P_1}\right) \\ &\quad - \mathcal{C}\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2}{\sigma^2}\right) + 1 \end{aligned}$$

$$\begin{aligned}
&\stackrel{(c)}{\leq} C\left(\frac{2(|g_{31}|^2 P_1 + |g_{32}|^2 P_2)}{\sigma^2}\right) + C(1) \\
&\quad - C\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2}{\sigma^2}\right) + 1 \\
&\leq 3,
\end{aligned} \tag{80}$$

where step (c) follows as $|g_{21}| \leq |g_{31}|$.

Likewise, the constant gap can be established with respect to R_1 by showing that $\varrho_1 - \mu_1 \leq 1$, and with respect to R_2 by showing that $\varrho_2 - \mu_2 \leq 1$. Note that ϱ_4 is not used in this proof because there is no inner bound on $R_2 + R_3$ in (77). Summarizing the above results gives a constant gap of 1 bit.

B. Constant-Gap Optimality of Uplink Rate Splitting

We further show that the achievable rate region of Proposition 4 is within 1 bit of the capacity region of the scalar Gaussian FD cellular network with D2D when $|g_{21}| \geq |g_{31}|$.

First consider a special case where $|g_{31}|^2 P_1 \leq \sigma^2$. Use $(\nu_1, \nu_2, \nu_3, \nu_4)$ to denote respectively the terms on the right-hand side of the inequalities in (40) when $a = b = d = 0$ and $c = e = 1$, which correspond to the inner bounds on R_2 , $R_1 + R_3$, $R_2 + R_3$, and $R_1 + R_2 + R_3$, respectively. Because $|g_{31}|^2 P_1 \leq \sigma^2$, we have

$$\nu_1 \geq C\left(\frac{|g_{32}|^2 P_2}{2\sigma^2}\right), \tag{81a}$$

$$\nu_2 = C\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right), \tag{81b}$$

$$\nu_3 \geq C\left(\frac{|g_{32}|^2 P_2}{2\sigma^2}\right), \tag{81c}$$

$$\nu_4 \geq C\left(\frac{|g_{32}|^2 P_2}{2\sigma^2}\right) + C\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right). \tag{81d}$$

It can be immediately seen that $\varrho_1 - \nu_1 \leq 1$ and thus $\delta \leq 1$ holds with respect to R_1 . Further, we bound δ with respect to $R_1 + R_3$ as follows:

$$\begin{aligned}
2\delta &\leq \varrho_3 - \nu_2 \\
&= C\left(\frac{(|g_{21}|^2 + |g_{31}|^2)P_1}{\sigma^2}\right) - C\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right) \\
&\stackrel{(d)}{\leq} C\left(\frac{|g_{21}|^2 P_1 + \sigma^2}{\sigma^2}\right) - C\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right) \\
&\leq 1,
\end{aligned} \tag{82}$$

where step (d) follows by $|g_{31}|^2 P_1 \leq \sigma^2$. Next, we bound δ with respect to $R_2 + R_3$:

$$\begin{aligned}
2\delta &\leq \varrho_4 - \nu_3 \\
&\leq C\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J}{\sigma^2}\right) - C\left(\frac{|g_{32}|^2 P_2}{2\sigma^2}\right) \\
&\leq C\left(\frac{2(|g_{31}|^2 P_1 + |g_{32}|^2 P_2)}{\sigma^2}\right) - C\left(\frac{|g_{32}|^2 P_2}{2\sigma^2}\right) \\
&\stackrel{(e)}{\leq} C\left(\frac{2(\sigma^2 + |g_{32}|^2 P_2)}{\sigma^2}\right) - C\left(\frac{|g_{32}|^2 P_2}{2\sigma^2}\right) \\
&\leq 2,
\end{aligned} \tag{83}$$

where step (e) is due to $|g_{31}| < |g_{21}|$. Finally, with respect to

$R_1 + R_2 + R_3$, the constant gap δ is upper bounded as

$$\begin{aligned}
3\delta &\leq \varrho_5 - \nu_4 \\
&\leq C\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 + J}{\sigma^2}\right) + C\left(\frac{|g_{21}|^2 P_1}{\sigma^2 + |g_{31}|^2 P_1}\right) \\
&\quad - C\left(\frac{|g_{32}|^2 P_2}{2\sigma^2}\right) - C\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right) \\
&\stackrel{(f)}{\leq} C\left(\frac{2(|g_{31}|^2 P_1 + |g_{32}|^2 P_2)}{\sigma^2}\right) - C\left(\frac{|g_{32}|^2 P_2}{2\sigma^2}\right) \\
&\leq 2.
\end{aligned} \tag{84}$$

Thus, we verify that $\delta \leq 1$ when $|g_{31}|^2 P_1 \leq \sigma^2$.

The remainder of the proof assumes that $|g_{31}|^2 P_1 \geq \sigma^2$. We now set $a = d = 0$, $c = \frac{\sigma^2}{|g_{31}|^2 P_1}$, $e = 1$, and $b = 1 - c$ in (40), thus arriving at this inner bound:

$$R_2 \leq C\left(\frac{|g_{32}|^2 P_2}{2\sigma^2}\right), \tag{85a}$$

$$R_1 + R_3 \leq C\left(\frac{|g_{21}|^2 P_1}{\sigma^2}\right), \tag{85b}$$

$$R_2 + R_3 \leq C\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 - \sigma^2}{2\sigma^2}\right), \tag{85c}$$

$$\begin{aligned}
R_1 + R_2 + R_3 &\leq C\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 - \sigma^2}{2\sigma^2}\right) \\
&\quad + C\left(\frac{|g_{21}|^2}{|g_{31}|^2}\right).
\end{aligned} \tag{85d}$$

Let $(\varphi_1, \varphi_2, \varphi_3, \varphi_4)$ be respectively the terms on the right-hand side of the above inequalities. With respect to $R_1 + R_2 + R_3$, we bound the constant gap δ as follows:

$$\begin{aligned}
3\delta &\leq \varrho_5 - \varphi_4 \\
&\leq C\left(\frac{2(|g_{31}|^2 P_1 + |g_{32}|^2 P_2)}{\sigma^2}\right) + C\left(\frac{|g_{21}|^2 P_1}{\sigma^2 + |g_{31}|^2 P_1}\right) \\
&\quad - C\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2 - \sigma^2}{2\sigma^2}\right) - C\left(\frac{|g_{21}|^2}{|g_{31}|^2}\right) \\
&= C\left(\frac{2(|g_{31}|^2 P_1 + |g_{32}|^2 P_2)}{\sigma^2}\right) + C\left(\frac{|g_{21}|^2 P_1}{\sigma^2 + |g_{31}|^2 P_1}\right) \\
&\quad - C\left(\frac{|g_{31}|^2 P_1 + |g_{32}|^2 P_2}{\sigma^2}\right) - C\left(\frac{|g_{21}|^2}{|g_{31}|^2}\right) + 1 \\
&\leq 2.
\end{aligned} \tag{86}$$

It can be also shown that $\delta \leq \varrho_2 - \varphi_1 \leq 1$, $2\delta \leq \varrho_3 - \varphi_2 \leq 2$, and $2\delta \leq \varrho_4 - \varphi_3 \leq 2$. The proof is then complete.

REFERENCES

- [1] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *ACM Int. Conf. Mobile Comput. Netw. (MobiCom)*, Aug. 2010, pp. 375–386.
- [2] E. Aryafar, M. Khojastepour, K. Sundaresan, S. Rangarajan, and M. Chiang, "MIDU: Enabling MIMO full duplex," in *ACM Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2012, pp. 257–268.
- [3] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, Feb. 2014.
- [4] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sept. 2014.

- [5] X. Xie and X. Zhang, "Does full-duplex double the capacity of wireless networks?," in *IEEE INFOCOM*, Apr. 2014, pp. 253–261.
- [6] S. Goyal, P. Liu, S. S. Panwar, R. A. Difazio, R. Yang, and E. Bala, "Full duplex cellular systems: will doubling interference prevent doubling capacity?," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 121–127, May 2015.
- [7] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.
- [8] S. Goyal, P. Liu, S. Panwar, R. A. DiFazio, R. Yang, J. Li, and E. Bala, "Improving small cell capacity with common-carrier full duplex radios," in *IEEE Int. Conf. Commun. (ICC)*, June 2014.
- [9] J. Maršević, J. Zhou, H. Krishnaswamy, Y. Zhong, and G. Zussman, "Resource allocation and rate gains in practical full-duplex systems," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 292–305, Feb. 2017.
- [10] J.-H. Yun, "Intra and inter-cell resource management in full-duplex heterogeneous cellular networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 2, pp. 392–405, Feb. 2016.
- [11] K. Shen and W. Yu, "Interference management in full-duplex wireless cellular networks via fractional programming," in *IEEE Veh. Technol. Conf. (VTC) Spring*, June 2018.
- [12] S. Shao, D. Liu, K. Deng, Z. Pan, and Y. Tang, "Analysis of carrier utilization in full-duplex cellular networks by dividing the co-channel interference region," *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 1043–1046, June 2014.
- [13] C. Karakus and S. Diggavi, "Opportunistic scheduling for full-duplex uplink-downlink networks," in *IEEE Int. Symp. Inf. Theory (ISIT)*, June 2015, pp. 1019–1023.
- [14] S. H. Chae, S. H. Lim, and S.-W. Jeon, "Degrees of freedom of full-duplex multi-antenna cellular networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 982–995, Feb. 2018.
- [15] S. H. Chae, S.-W. Jeon, and S. H. Lim, "Fundamental limits of spectrum sharing full-duplex multicell networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 3048–3061, Nov. 2016.
- [16] M. A. Khojastepour, K. Sundaresan, and S. Rangarajan, "Scaling wireless full-duplex in multi-cell networks," in *IEEE INFOCOM*, May 2015, pp. 1751–1759.
- [17] R. Tannious and A. Nosratinia, "Relay channel with private messages," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3777–3785, Oct. 2007.
- [18] Y. Liang and V. V. Veeravalli, "Cooperative relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 900–928, Mar. 2007.
- [19] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3517–3535, Oct. 2007.
- [20] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Second Edition)*, John Wiley & Sons, Inc., 2006.
- [21] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge University Press, 2011.
- [22] H. Sato, "An outer bound to the capacity region of broadcast channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 374–377, May 1978.
- [23] P. Razaghi and W. Yu, "Bilayer low-density parity-check codes for decode-and-forward in relay channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3723–3739, Oct. 2007.
- [24] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Nov. 2008.

Kaiming Shen (S'13-M'20) received the B.Eng. degree in information security and the B.S. degree in mathematics from Shanghai Jiao Tong University, Shanghai, China in 2011, then the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Toronto, Ontario, Canada in 2013 and 2020, respectively.

Since 2020, he has been an Assistant Professor with the School of Science and Engineering at the Chinese University of Hong Kong (Shenzhen), China. His main research interests include optimization, wireless communications, data science, and information theory.

Reza K. Farsani was born in Farsan, Iran, in 1986. He received the double major B.Sc. degrees in electrical engineering and pure mathematics from Sharif University of Technology, Tehran, Iran, in 2010, and the M.S. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2016. He is currently pursuing the Ph.D. degree at the University of Toronto, Toronto, Canada.

From 2010 to 2014, he was a Research Scientist with the Institute for Research in Fundamental Sciences (IPM), Tehran, where he accomplished a research project on fundamental limits of communications in interference networks. From 2016 to 2017, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include information and communication theory, machine learning, and statistics.

Mr. Farsani received the Silver Medal from the Iranian Students Mathematical Olympiad in 2003.

Wei Yu (S'97-M'02-SM'08-F'14) received the B.A.Sc. degree in Computer Engineering and Mathematics from the University of Waterloo, Waterloo, Ontario, Canada in 1997 and M.S. and Ph.D. degrees in Electrical Engineering from Stanford University, Stanford, CA, in 1998 and 2002, respectively. Since 2002, he has been with the Electrical and Computer Engineering Department at the University of Toronto, Toronto, Ontario, Canada, where he is now Professor and holds a Canada Research Chair (Tier 1) in Information Theory and Wireless Communications.

Prof. Wei Yu serves as the President of the IEEE Information Theory Society in 2021, and has served on its Board of Governors since 2015. He served as the Chair of the Signal Processing for Communications and Networking Technical Committee of the IEEE Signal Processing Society in 2017–18. Prof. Wei Yu was an IEEE Communications Society Distinguished Lecturer in 2015–16. He is currently an Area Editor for the IEEE Transactions on Wireless Communications, and in the past served as an Associate Editor for IEEE Transactions on Information Theory, IEEE Transactions on Communications, and IEEE Transactions on Wireless Communications.

Prof. Wei Yu is a Fellow of the Canadian Academy of Engineering, and a member of the College of New Scholars, Artists and Scientists of the Royal Society of Canada. He received the Steacie Memorial Fellowship in 2015, the IEEE Marconi Prize Paper Award in Wireless Communications in 2019, the IEEE Communications Society Award for Advances in Communication in 2019, the IEEE Signal Processing Society Best Paper Award in 2017 and 2008, the Journal of Communications and Networks Best Paper Award in 2017, and the IEEE Communications Society Best Tutorial Paper Award in 2015.