RESEARCH　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

Vector bin-and-cancel for MIMO distributed full-duplex

Jingwen Bai[*] ⓘD and Ashutosh Sabharwal

Abstract

In a multi-input multi-output (MIMO) full-duplex network, where an in-band full-duplex infrastructure node communicates with two half-duplex mobiles supporting simultaneous up- and downlink flows, the inter-mobile interference between the up- and downlink mobiles limits the system performance. We study the impact of leveraging an out-of-band side channel between mobiles in such network under different channel models. For time-invariant channels, we aim to characterize the generalized degrees-of-freedom (GDoF) of the side-channel-assisted MIMO full-duplex network. For slow-fading channels, we focus on the diversity-multiplexing tradeoff (DMT) of the system with various assumptions as to the availability of channel state information at the transmitter (CSIT). The key to the optimal performance is a vector bin-and-cancel strategy leveraging Han-Kobayashi message splitting, which is shown to achieve the system capacity region to within a constant bit. We quantify how the side channel improve the GDoF and DMT compared to a system without the extra orthogonal spectrum. The insights gained from our analysis reveal (i) the tradeoff between spatial resources from multiple antennas at different nodes and spectral resources of the side channel and (ii) the interplay between the channel uncertainty at the transmitter and use of the side channel.

1 Introduction

Increasingly, mobile devices have multiple radios to simultaneously access different parts of the spectrum, e.g., cellular and ISM bands. The ability of simultaneous access to multiple parts of the spectrum provides an opportunity to use multiple bands in new and unique ways. A common method is to use the two bands to access both cellular and ISM band networks (notably WiFi) at the same time, which is now an integral part of cellular provider data strategy to offload cellular traffic to WiFi networks [1]. In this paper, we will consider the use of device-to-device (D2D) wireless channels between mobile devices, to serve as side channels to *aid* main-channel communication with the infrastructure nodes. For example, the main network could be on a cellular band while the wireless side channel could be on an unlicensed ISM band. The conventional use of D2D involves establishing peer-to-peer communication [2], forming virtual MIMO by cooperative communication [3] or offloading cellular traffic [4]. In contrast, we propose to use the D2D side channel for interference management to improve the cellular capacity,
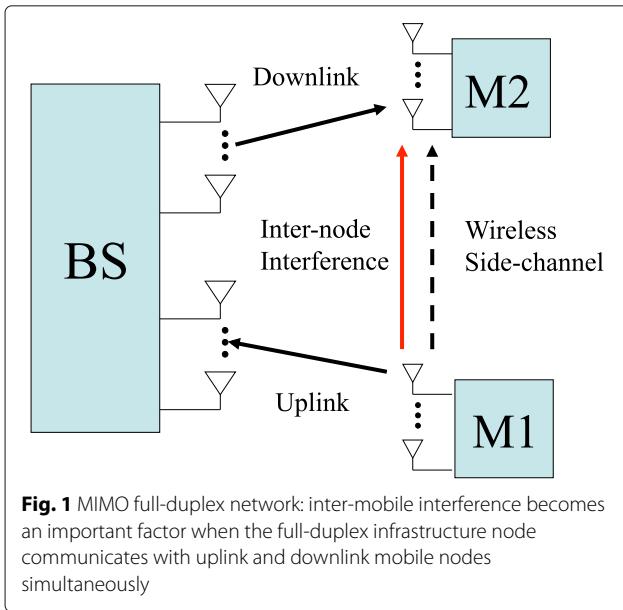
a scenario which we labeled as *ISM-in-cellular* communication [5–7].

In this paper, we will study how the side channel will impact the system performance in a two-user MIMO full-duplex network. In-band full-duplex operation promises to double the spectral efficiency as compared to the half-duplex counterpart which uses either time division or frequency division for transmission and reception. It is in fact feasible to design near-perfect full-duplex base stations owing to the available freedom (bigger size, non-battery-powered operation) in their designs (e.g., see [8–10] and the references therein). And in-band full-duplex has already become part of the ongoing standard both in 3GPP [11] and 802.11-ax [12]. Thus, we envision that the first use of full-duplex capabilities might be in a small cell infrastructure [13], supporting legacy half-duplex mobile nodes.

In Fig. 1, a full-duplex-capable base station (BS) communicates with two half-duplex mobiles simultaneously to support one uplink (UL) and one downlink (DL) flow. A major bottleneck in this network is the inter-mobile interference from uplink mobile (node M1) to downlink mobile node (node M2), because of which the degrees-of-freedom of the network collapse to one when all nodes

*Correspondence: jingwen.bai@intel.com
Department of Electrical and Computer Engineering, Rice University, Houston, USA

Fig. 1 MIMO full-duplex network: inter-mobile interference becomes an important factor when the full-duplex infrastructure node communicates with uplink and downlink mobile nodes simultaneously

are equipped with single antenna (SISO) [6]. As a result, we proposed a *distributed full-duplex architecture* [6] to leverage the wireless side channels to mitigate inter-mobile interference. In the case of MIMO scenario which has been widely studied for half-duplex system [14], one driving question for distributed full-duplex system is if and how the spatial degree-of-freedom, i.e., number of antennas at the base station and mobiles, will be correlated to the spectral degrees-of-freedom offered by the side channel.

In our setup, we assume that uplink node M1 has M_{ul} transmit antennas, the downlink node M2 has N_{dl} receive antennas, the full-duplex BS has M_{dl} and N_{ul} transmit and receive antennas, respectively. The bandwidth of the side channel between the mobiles is W-fold compared to the main-channel. We summarize the main results in this work as follows.

1. In the time-invariant channels, we obtain the capacity region to within a constant bit achieved by a vector bin-and-cancel scheme. We also analyze the role of channel uncertainty at the transmitter and characterize the GDoF as a function of antenna numbers and side-channel bandwidth under different assumptions of CSIT. The insights gained from GDoF reveal the tradeoff between spatial resources from multiple antennas and spectral resources of the side channels as well as the interplay between the channel uncertainty at the transmitter and the use of side channel. In the case when BS has more antennas than mobiles, if there are more downlink receive antennas than uplink transmit

antennas, i.e., $N_{dl} \geq M_{ul}$, there is no benefit to obtain CSIT since with and without CSIT achieve the same degrees-of-freedom. On the other hand, if $M_{ul} > N_{dl}$, having CSIT require less side-channel bandwidth to achieve no-interference performance. Thus, we conclude that having more spatial degree-of-freedom at the interfered downlink receiver or larger side-channel bandwidth can simplify transceiver design by ruling out the necessity to obtain CSIT.

2. In slow-fading channels, we derive the general DMT regarding different assumptions of CSIT. Specifically, we quantify the bandwidth of the side channel required to compensate for *lack of* CSIT such that the DMT without CSIT achieves the optimal DMT with CSIT. Interestingly, in the case when $M_{dl} = N_{ul} = M \geq M_{ul}, N_{dl}$, the required bandwidth is inversely proportional to the number of antennas at the BS, i.e., $W \propto \frac{1}{M}$. The caveat is that the side-channel channel SNR, in the meantime, has to grow with the number of antennas at BS. The result provides guidance towards system design; larger number of BS antennas, e.g., recent discussions on massive MIMO [15], can help reduce the required side-channel bandwidth to combat inter-mobile interference.

   We also observe the dependency of CSIT and the antenna number ratio between the mobiles. For the symmetric DMT, when $M_{ul} > N_{dl}$, without side channel, the lack of CSIT will result in performance loss. However, larger side-channel bandwidth will help bridge the performance gap. On the other hand, when $N_{dl} \geq M_{ul}$, there is no benefit to obtain CSIT to achieve no-interference DMT since, with and without CSIT, one requires the same amount of side-channel bandwidth to completely eliminate the effect of interference. Hence, in the protocol design, the scheduler could possibly group downlink user with more receive antennas to eliminate the overhead of acquiring CSIT.

3. We evaluate the required side-channel bandwidth to achieve the no-interference GDoF and DMT under different channel models such that the effect of inter-mobile interference can be completely eliminated via side channel. The key difference in the findings between the two channel scenarios, for instance, when $M_{dl} = N_{ul} = M \geq M_{ul}, N_{dl}$, is that in GDoF analysis under time-invariant channels, the required W does not depend on the antenna number ratio between the mobiles; while in DMT analysis under slow-fading channels, the required W is a function of the antenna number ratio $A = \frac{\max(M_{ul}, N_{dl})}{\min(M_{ul}, N_{dl})}$ and $W \propto \frac{1}{A}$. The impact on the system design is that we should schedule up- and downlink user pair with

higher antenna ratio to cancel out interference with reduced side-channel bandwidth.

The rest of paper is organized as follows. The Section 2 presents the system model. In the Section 3, we show that a vector bin-and-cancel scheme achieves within a constant gap of the capacity region in time-invariant channels. We give a characterization of GDoF which reveals tradeoff between spatial resources from multiple antennas and spectral resources of the side channels under both CSIT and no-CSIT assumptions. In the Section 4, we derive the general DMT with and without CSIT in slow-fading channels. We also study the spatial and spectral tradeoff between multiple antennas and side channel on the symmetric DMT. The Section 5 concludes the paper.

Notations We use $A^\dagger$ to denote Hermitian of A, and $|A|$ to denote the determinant of A. We use $(x)^+$ to denote $\max(x, 0)$. We use $\mathcal{CN}(0, Q)$ to denote a circularly symmetric complex Gaussian distribution with zero mean and covariance matrix Q. We use I_N to denote identity matrix of rank N. We use $f(\rho) \doteq g(\rho)$ to denote that $\lim_{\rho \to \infty} \frac{\log f(\rho)}{\log g(\rho)} = 1$. We use $A \preceq B$ to denote that matrix $B - A$ is a positive-semidefinite positive (p.s.d) matrix.

2 System model

In this section, we describe the system model to be used for the rest of the paper. We assume the full-duplex BS is equipped with M_{dl} transmit antennas for the downlink and N_{ul} receive antennas for the uplink. The uplink mobile M1 is equipped with M_{ul} transmit antennas, and downlink mobile M2 is equipped with N_{dl} receive antennas. Besides the main-channel which includes uplink, downlink, and interference link, there also exists an out-of-band wireless side channel between the uplink mobile and downlink mobile. We refer to the channel model shown in Fig. 2 as $(M_{\mathrm{dl}}, N_{\mathrm{dl}}, M_{\mathrm{ul}}, N_{\mathrm{ul}})$ side-channel-assisted MIMO full-duplex network. Let W_m and W_s denote the bandwidth of the main channel and side channel, respectively. Parameter $W = \frac{W_s}{W_m}$ represents the bandwidth ratio of the side channel to that of the main-channel.

Since one of the transmitter and receiver is co-located in the same node, the base-station BS, the uplink message received by the BS is causally known to the BS transmitter for downlink transmission. As a result, the side-channel-assisted full-duplex network can be viewed as a Z-interference channel with implicit feedback and an out-of-band side channel.

We assume that the channel parameters in our system model consist of two components: a small-scale fading factor due to multipath and a large-scale fading factor due to path loss. We denote the small-scale fading channels matrix as $\mathcal{H} = \{H_{\mathrm{dl}}, H_{\mathrm{ul}}, H_{\mathrm{I}}, H_{\mathrm{S}}\}$, where each entry in $\mathcal{H}$ represents the small-scale fading channel matrix for the

Fig. 2 Channel model: $(M_{\mathrm{dl}}, N_{\mathrm{dl}}, M_{\mathrm{ul}}, N_{\mathrm{ul}})$ side-channel-assisted MIMO full-duplex network
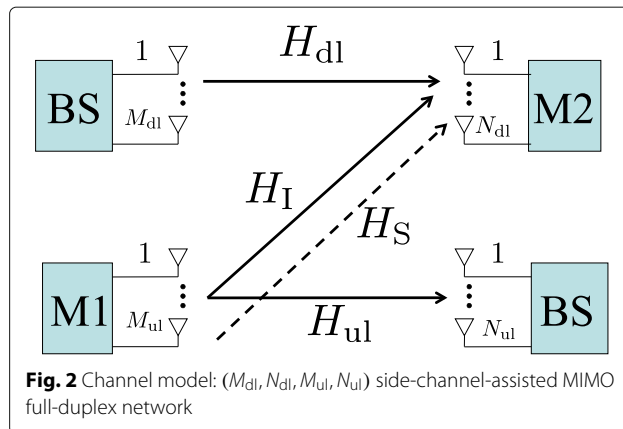
downlink, uplink, inter-mobile interference channel, and the side channel, as shown in Fig. 2. We assume that all entries in H_k, where $k \in \{\mathrm{dl}, \mathrm{ul}, \mathrm{I}, \mathrm{S}\}$, are mutually independent and identically distributed (i.i.d.) according to $\mathcal{CN}(0, 1)$ and all channel matrices are full rank with probability one. We will consider two different scenarios for the small-scale fading.

- *Time-invariant channels*: $\mathcal{H}$ is fixed during the entire communication period.
- *Slow-fading channels*: $\mathcal{H}$ remains unchanged during each fade duration or coherence time and varies i.i.d. between distinct fade periods.

As for the large-scale fading factor, it captures the channel attenuation due to distance. Thus, the channel attenuation between the transmitter and receiver is the same for every transmit-receive antenna pair. Hence, the channel attenuation for each channel is denoted by a scalar γ_k, where $k \in \{\mathrm{dl}, \mathrm{ul}, \mathrm{I}, \mathrm{S}\}$. The transmitter at BS and uplink node M1 have a maximum power budget P_{dl} and P_{ul}, respectively. To simplify the notation, let $\rho_{\mathrm{dl}} = \gamma_{\mathrm{dl}} P_{\mathrm{dl}}$, $\rho_{\mathrm{ul}} = \gamma_{\mathrm{ul}} P_{\mathrm{ul}}$, $\rho_{\mathrm{S}} = \gamma_{\mathrm{S}} P_{\mathrm{ul}}$, and $\rho_{\mathrm{I}} = \gamma_{\mathrm{I}} P_{\mathrm{ul}}$, which denotes the average signal-to-noise ratio and interference-to-noise ratio at each receive antenna with additive Gaussian noise of unit variance.

Next, we describe the channel input-output relationships as follows.

2.1 Uplink

The node M1 will split the transmit power between main-channel and side channel, i.e., $\bar\lambda P_{\mathrm{ul}}$ and λP_{ul} for main-channel and side-channel data transmission, respectively. We define $\bar\lambda = 1 - \lambda, \lambda \in [0, 1]$. Thus, the received uplink signal $Y_{\mathrm{ul}} \in \mathbb{C}^{N_{\mathrm{ul}} \times 1}$ at BS is given by

$$Y_{\mathrm{ul}}(t) = \sqrt{\bar\lambda \rho_{\mathrm{ul}}} H_{\mathrm{ul}} X_{\mathrm{ul}}(t) + Z_{\mathrm{ul}}(t), \qquad (1)$$

where $X_{\text{ul}}(t) \in \mathbb{C}^{M_{\text{ul}} \times 1}$ is the uplink vector signal; $H_{\text{ul}} \in \mathbb{C}^{N_{\text{ul}} \times M_{\text{ul}}}$ represents uplink channel and $Z_{\text{ul}}(t) \in \mathbb{C}^{N_{\text{ul}} \times 1}$ is the receiver additive Gaussian noise which contains i.i.d. $\mathcal{CN}(0, 1)$ entries.

2.2 Downlink

The received downlink signal $Y_{\text{dl}} \in \mathbb{C}^{N_{\text{dl}} \times 1}$ at the node M2 is a combination of the downlink signal and the interfering uplink signal and is given by

$$Y_{\text{dl}}(t) = \sqrt{\rho_{\text{dl}}} H_{\text{dl}} X_{\text{dl}}(t) + \sqrt{\bar{\lambda} \rho_{\text{I}}} H_{\text{I}} X_{\text{ul}}(t) + Z_{\text{dl}}(t), \tag{2}$$

where $X_{\text{dl}}(t) \in \mathbb{C}^{M_{\text{dl}} \times 1}$ is the downlink vector signal, $H_{\text{dl}} \in \mathbb{C}^{N_{\text{dl}} \times M_{\text{dl}}}$ is the downlink channel matrix and $H_{\text{I}} \in \mathbb{C}^{N_{\text{dl}} \times M_{\text{ul}}}$ is the inter-mobile interference channel matrix, and $Z_{\text{dl}}(t) \in \mathbb{C}^{N_{\text{dl}} \times 1}$ is the receiver additive Gaussian noise which contains i.i.d. $\mathcal{CN}(0, 1)$ entries.

2.3 Side channel

We assume that the number of side-channel antennas are the same as those of the main-channel.

Thus, the received signal $Y_{\text{S}} \in \mathbb{C}^{N_{\text{dl}} \times 1}$ at the node M2 is given by

$$Y_{\text{S}}(t) = \sqrt{\lambda \rho_{\text{S}}} H_{\text{S}} X_{\text{S}}(t) + Z_{\text{S}}(t), \tag{3}$$

where $X_{\text{S}}(t) \in \mathbb{C}^{M_{\text{ul}} \times 1}$ is the side-channel vector signal, $H_{\text{S}} \in \mathbb{C}^{N_{\text{dl}} \times M_{\text{ul}}}$ is the channel matrix of the side channel, and $Z_{\text{dl}}(t) \in \mathbb{C}^{N_{\text{dl}} \times 1}$ is the Gaussian noise added to the side channel which contains i.i.d. $\mathcal{CN}(0, W)$ entries. Note that the noise variance of each entry in the side channel is W times larger than that in the main-channel.

The power constraint of the input signals is given as:

$$\frac{1}{L} \sum_{t=1+Lk}^{L(k+1)} \text{Trace} \left(\mathbb{E} \left[X_i(t) X_i(t)^{\dagger} \right] \right) \leq 1, \ k \in \mathbb{N}, i \in \{\text{dl,ul,S}\}, \tag{4}$$

where in time-invariant channels, $k = 0$, and L denotes the entire communication duration; in slow-fading channels, L denotes the coherence time.[1]

We define the strength level of different links with respect to nominal SNR, ρ, in decibels[2]

$$\alpha_i = \frac{\log \rho_i}{\log \rho}, \ i \in \{\text{dl, ul, I, S}\}. \tag{5}$$

Note that the above normalization allows different links to have disparate strength.

3 Vector bin-and-cancel scheme

A full-duplex node can be viewed as "two nodes," with a co-located transmitter and receiver, which are connected by an *infinite* capacity link. Inspired by this interpretation, in [6], we proposed a *distributed full-duplex* architecture

which is enabled by a wireless side channel of *finite* bandwidth when the transmitter and interfered receiver are not co-located. When channel knowledge is known globally, we showed that a bin-and-cancel scheme achieves the capacity region to within 1 bit/s/Hz of the capacity region for all channel parameters in SISO case [6].

In this section, we will study the capacity region in MIMO case under different assumptions of channel uncertainty at the transmitter. CSIT plays a critical role in MIMO interference channels. With CSIT, the transmitter can design the precoding matrix to steer the direction of the transmit signal to achieve higher rate. However, the cost of obtaining CSIT is also prohibitive since the receiver has to feed back the channel knowledge within the coherence time which incurs operational overhead. Thus, it is crucial to explore the role of channel uncertainty at the transmitter in system performance. We assume that the receiver-side-channel information is always available as the receiver can track the instantaneous channel from the training pilots. In what follows, we will study the capacity region in time-invariant channels. Next, we will present how CSIT and the use of side channel is correlated, we also characterize the spatial and spectral tradeoff between multiple antennas at different nodes and spectral resources provided by side channel.

3.1 Capacity region to within a constant gap with CSIT
3.1.1 Outer bound

Lemma 1 *Given the channel realization $\mathcal{H}$, the capacity region $\mathcal{C}(\mathcal{H})$ of the side-channel-assisted MIMO full-duplex network is outer bounded by*

$$R_{\text{dl}} \leq W_m \left(\log \left| I_{N_{\text{dl}}} + \rho_{\text{dl}} H_{\text{dl}} H_{\text{dl}}^{\dagger} \right| \right) \triangleq \overline{C}_{\text{dl}},$$

$$R_{\text{ul}} \leq W_m \left(\log \left| I_{N_{\text{ul}}} + \bar{\lambda} \rho_{\text{ul}} H_{\text{ul}} H_{\text{ul}}^{\dagger} \right| \right) \triangleq \overline{C}_{\text{ul}},$$

$$R_{\text{dl}} + R_{\text{ul}} \leq W_m \left(\log \left| I_{N_{\text{dl}}} + \rho_{\text{dl}} H_{\text{dl}} H_{\text{dl}}^{\dagger} + \bar{\lambda} \rho_{\text{I}} H_{\text{I}} H_{\text{I}}^{\dagger} \right| \right.$$

$$+ W \log \left| I_{N_{\text{dl}}} + \frac{\lambda \rho_{\text{S}}}{W} H_{\text{S}} H_{\text{S}}^{\dagger} \right|$$

$$+ \log \left| I_{N_{\text{ul}}} + \bar{\lambda} \rho_{\text{ul}} H_{\text{ul}} (I_{M_{\text{ul}}} + \bar{\lambda} \rho_{\text{I}} H_{\text{I}}^{\dagger} H_{\text{I}})^{-1} H_{\text{ul}}^{\dagger} \right|$$

$$\left. + N_{\text{dl}} \right) \triangleq \overline{C}_{\text{sum}}, \tag{6}$$

Proof See Appendix 1. Note that if the interference channel (ρ_{I}) or side-channel quality ($W \rho_{\text{S}}$) exceeds certain threshold such that $C_{\text{sum}} \geq C_{\text{dl}} + C_{\text{ul}}$, the capacity is just trivially outer bounded by the first two individual constraints in (6). □

3.1.2 Achievable rate region

A vector bin-and-cancel scheme based on a simple Han-Kobayashi coding strategy achieves the following rate

region when CSIT is available. The scheme will be elucidated later in the Section 3.2.

Lemma 2 *The achievable rate region $\mathcal{R}_{BC}(\mathcal{H})$ of the side-channel-assisted MIMO three-node full-duplex network for time-invariant channels is*

$$
\begin{aligned}
R_{\mathrm{dl}} &\leq \overline{C}_{\mathrm{dl}} - W_m c_1, \\
R_{\mathrm{ul}} &\leq \overline{C}_{\mathrm{ul}} - W_m c_2, \\
R_{\mathrm{dl}} + R_{\mathrm{ul}} &\leq \overline{C}_{\mathrm{sum}} - W_m \left(c_1 + c_2 \right),
\end{aligned}
\tag{7}
$$

where

$$
\begin{aligned}
c_1 &= \min\{M_{\mathrm{dl}} + M_{\mathrm{ul}}, N_{\mathrm{dl}}\} \log \left(\max\{M_{\mathrm{dl}}, M_{\mathrm{ul}}\} \right) + \hat{m}_{\mathrm{I}}, \\
c_2 &= (m_{\mathrm{ul}} + W m_{\mathrm{I}}) \log M_{\mathrm{ul}} + m_X \log(M_{\mathrm{ul}} + 1), \\
\hat{m}_{\mathrm{I}} &= m_{\mathrm{I}} \log \left(1 + \frac{1}{M_{\mathrm{ul}}} \right), \\
m_{\mathrm{dl}} &= \min\{M_{\mathrm{dl}}, N_{\mathrm{dl}}\}, \; m_{\mathrm{ul}} = \min\{M_{\mathrm{ul}}, N_{\mathrm{ul}}\}, \\
m_X &= \max\{M_{\mathrm{ul}}, N_{\mathrm{dl}}\}, \; m_{\mathrm{I}} = \min\{M_{\mathrm{ul}}, N_{\mathrm{dl}}\}.
\end{aligned}
\tag{8}
$$

Proof See the Section 3.2 for the description of the achievability and Appendix 2 for the rate calculation. □

Based on the lemmas above, we will state the result of constant-bit gap to capacity region under time-invariant channels in the following theorem.

Theorem 1 *For the side-channel-assisted two-user MIMO full-duplex network under time-invariant channels, the achievable rate region $\mathcal{R}_{BC}(\mathcal{H})$ is within $\max\{c_1, c_2\}$ bit/s/Hz of the capacity region $\mathcal{C}(\mathcal{H})$, where c_i, $i = 1, 2$ is given in* (8).

Proof The proof is straightforward. From Lemmas 1 and 2, we can calculate the rate difference and divide it by the total bandwidth $W_m + W_s$ of the system.

In other word, for any given rate pair $(R_{\mathrm{dl}}, R_{\mathrm{ul}}) \in \mathcal{C}(\mathcal{H})$ (bit/s), the rate pair $\left((R_{\mathrm{dl}} - (W_m + W_s)c_1)^+, (R_{\mathrm{ul}} - (W_m + W_s)c_2)^+ \right)$ is achievable in $\mathcal{R}_{BC}(\mathcal{H})$. □

In the SISO case, we can easily verify that the vector bin-and-cancel achieves the capacity region to within one bit.

3.2 Achievability

In this section, we will describe the vector bin-and-cancel scheme used to show the achievability in Lemma 2. In vector bin-and-cancel, we use Han-Kobayashi [16] style common-private message splitting with a simple power splitting. The common message can be decoded at both

receivers while the private message can only be decoded at the intended receiver. The downlink message ω_{dl} only consists of private message for the downlink receiver, which is of size $2^{nR_{\mathrm{dl}}}$, and is encoded into codeword X_{dl}. The uplink message is divided into the common part $\omega_{\mathrm{ul},c}$ of size $2^{nR_{\mathrm{ul},c}}$ and the private part $\omega_{\mathrm{ul},p}$ of size $2^{nR_{\mathrm{ul},p}}$. The uplink codeword is then obtained by superposition of the codewords of both $\omega_{\mathrm{ul},c}$ and $\omega_{\mathrm{ul},p}$,

$$
X_{\mathrm{ul}} = S_{\mathrm{ul}} + U_{\mathrm{ul}},
$$

where S_{ul} and U_{ul} are the codewords of uplink common message $\omega_{\mathrm{ul},c}$ and private message $\omega_{\mathrm{ul},p}$, respectively.

Next, we partition the uplink common message $\omega_{\mathrm{ul},c}$: the common message set is divided into 2^{nR_S} equal size bins such that $\mathcal{B}(l) = \left[(l-1)2^{n(R_{\mathrm{ul},c} - R_S)} + 1 : l 2^{n(R_{\mathrm{ul},c} - R_S)} \right], l \in [1 : 2^{nR_S}]$. The total number of bin indices 2^{nR_S} is determined by the strength of the side channel, α_S, and the bandwidth ratio W. The bin index l is then encoded into codeword X_S and sent from the uplink transmit antenna arrays over the side channel, which is shown in Fig. 3.

All the codewords are mutually independent complex Gaussian random vectors with covariance matrices given as follows to satisfy the power constraint given in (4):

$$
\begin{aligned}
\mathbb{E}\left(X_{\mathrm{dl}} X_{\mathrm{dl}}^{\dagger} \right) &= \frac{1}{M_{\mathrm{dl}}} I_{M_{\mathrm{dl}}}, \; \mathbb{E}\left(U_{\mathrm{ul}} U_{\mathrm{ul}}^{\dagger} \right) \\
&= \frac{1}{M_{\mathrm{ul}}} \left(I_{M_{\mathrm{ul}}} + \bar{\lambda} \rho_{\mathrm{I}} H_{\mathrm{I}}^{\dagger} H_{\mathrm{I}} \right)^{-1} \\
\mathbb{E}\left(S_{\mathrm{ul}} S_{\mathrm{ul}}^{\dagger} \right) &= \frac{1}{M_{\mathrm{ul}}} I_{M_{\mathrm{ul}}} - \mathbb{E}\left(U_{\mathrm{ul}} U_{\mathrm{ul}}^{\dagger} \right), \; \mathbb{E}\left(X_{\mathrm{ul}}^{s} X_{\mathrm{ul}}^{s\dagger} \right) \\
&= \frac{1}{M_{\mathrm{ul}}} I_{M_{\mathrm{ul}}},
\end{aligned}
\tag{9}
$$

where $\lambda \in (0, 1), \bar{\lambda} + \lambda = 1$. The parameter λ denotes the fraction of power allocated to the side channel. For the power splitting between the uplink private and common message, we set the power of the private message such
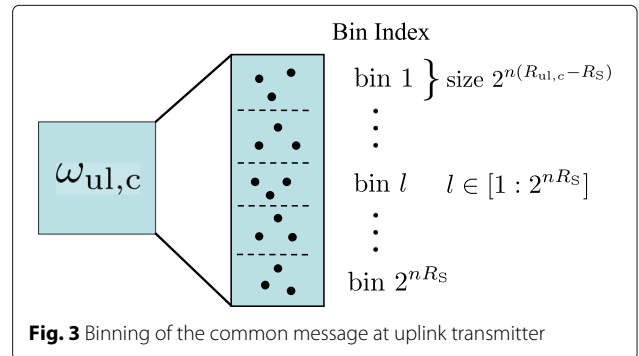
Fig. 3 Binning of the common message at uplink transmitter

that its received signal strength is below the noise floor at each unintended receiver's antenna. And we allocate the power of the codewords equally among the transmit antenna array.

Now, we describe the decoding process. The decoding at the BS is straightforward. Upon receiving Y_{ul}, the BS decodes $(\omega_{ul,c}, \omega_{ul,p})$. The achievable rate region of $(R_{ul,c}, R_{ul,p})$ is the capacity region of multiple-access channel denoted as $\mathcal{C}_1$, where

$$
\begin{aligned}
R_{ul,c} &\leq I(S_{ul}; Y_{dl}|X_{dl}) \\
R_{ul,p} &\leq I(U_{ul}; Y_{ul}|S_{ul}) \\
R_{ul,c} + R_{ul,p} &\leq I(S_{ul}, U_{ul}; Y_{ul})
\end{aligned}
\tag{10}
$$

The decoding at the downlink receiver has two stages as shown in Fig. 4. In stage one, upon receiving Y_S, the downlink receiver first decodes the bin index l from the side channel. In stage two, upon receiving Y_{dl}, the downlink receiver decodes $(\omega_{dl}, \omega_{ul,c})$ with the help of side-channel information while treating uplink private message $\omega_{ul,p}$ as noise.[3] This is a multiple-access channel (MAC) with side channel whose capacity region denoted as $\mathcal{C}_2$ is given in [6]; hence, we have

$$
\begin{aligned}
R_{dl} &\leq I(X_{dl}; Y_{dl}|S_{ul}) \\
R_{ul,c} &\leq I(S_{ul}; Y_{dl}|X_{dl}) + I(X_S; Y_S) \\
R_{dl} + R_{ul,c} &\leq I(X_{dl}, S_{ul}; Y_{dl}) + I(X_S; Y_S).
\end{aligned}
\tag{11}
$$

The achievable rate region of side-channel-assisted full-duplex network is the set of all (R_{dl}, R_{ul}) such that $R_{dl}, R_{ul} = R_{ul,c} + R_{ul,p}$ satisfying that $(R_{ul,c}, R_{ul,p}) \in \mathcal{C}_1$ and $(R_{dl}, R_{ul,c}) \in \mathcal{C}_2$. Using Fourier-Motzkin elimination, the achievable rate pairs (R_{dl}, R_{ul}) are constrained by the following rate region

$$
\begin{aligned}
R_{dl} &\leq I(X_{dl}; Y_{dl}|S_{ul}) \\
R_{ul} &\leq \min \{I(S_{ul}, U_{ul}; Y_{ul}), I(U_{ul}; Y_{ul}|S_{ul}) \\
&\quad + I(S_{ul}; Y_{dl}|X_{dl}) + I(X_S; Y_S)\} \\
R_{dl} + R_{ul} &\leq I(U_{ul}; Y_{ul}|S_{ul}) + I(X_{dl}, S_{ul}; Y_{dl}) \\
&\quad + I(X_S; Y_S).
\end{aligned}
\tag{12}
$$
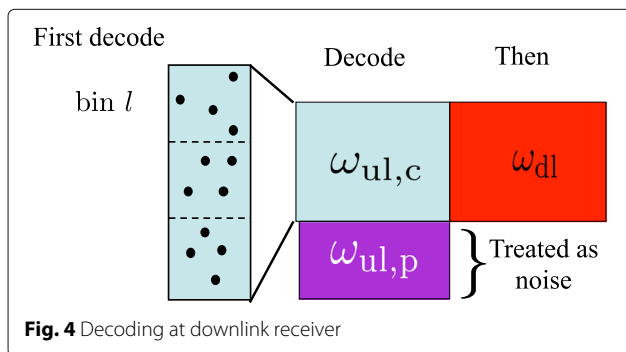
Fig. 4 Decoding at downlink receiver

The achievable rate region given above is calculated in Appendix 2; thus, we can obtain the explicit achievable rate expression in Lemma 2.

3.3 High SNR approximation
From Theorem 1, vector bin-and-cancel scheme achieves the capacity region to within a constant bit for *all* values of channel parameters under time-invariant channels. In the high SNR limit, a constant number of bits (which do not vary with respect to SNR) are insignificant and can be ignored on the scale of interest. Therefore, we can establish the high SNR capacity region approximation to within $\mathcal{O}(1)$ in the following corollary.

Corollary 1 *For a given the channel realization $\mathcal{H}$, vector bin-and-cancel is asymptotically capacity achieving and the asymptotic capacity approximation $\mathcal{C}(\mathcal{H})$ is given by*

$$
\begin{aligned}
\mathcal{C}(\mathcal{H}) \doteq \Big\{ (R_{dl}, R_{ul}) : R_{dl} &\leq W_m \log \left| I_{N_{dl}} + \rho_{dl} H_{dl} H_{dl}^{\dagger} \right| \triangleq C_{dl}, \\
R_{ul} &\leq W_m \log \left| I_{N_{ul}} + \bar{\lambda} \rho_{ul} H_{ul} H_{ul}^{\dagger} \right| \triangleq C_{ul}, \\
R_{dl} + R_{ul} &\leq W_m \left(\log \left| I_{N_{dl}} + \rho_{dl} H_{dl} H_{dl}^{\dagger} + \bar{\lambda} \rho_I H_I H_I^{\dagger} \right| \right. \\
&\quad + W \log \left| I_{N_{dl}} + \frac{\lambda \rho_S}{W} H_S H_S^{\dagger} \right| \\
&\quad + \left. \log \left| I_{N_{ul}} + \bar{\lambda} \rho_{ul} H_{ul} \left(I_{M_{ul}} + \bar{\lambda} \rho_I H_I^{\dagger} H_I \right)^{-1} H_{ul}^{\dagger} \right| \right) \\
&\triangleq C_{sum} \Big\}.
\end{aligned}
\tag{13}
$$

The high SNR capacity approximation can be used to derive the generalized degrees-of-freedom (GDoF). The GDoF captures the asymptotic behavior of the capacity and the corresponding optimal schemes, allowing different links to grow at disparate rates.

The GDoF region is defined as follows [4]

$$
\begin{aligned}
\Big\{ (\mathsf{DoF}_{dl}, \mathsf{DoF}_{ul}) : \mathsf{DoF}_i &= \lim_{\rho \to \infty} \frac{R_i(\rho_i)}{W_m \log \rho}, \ i \in \{dl, ul\} \\
\text{and } (R_{dl}, R_{ul}) &\in \mathcal{C}(\mathcal{H}) \Big\},
\end{aligned}
\tag{14}
$$

where $W_m \log \rho$ is the point-to-point main-channel capacity with nominal SNR in bit/s. DoF_{dl} and DoF_{ul} denote the degrees-of-freedom (DoF) of downlink and uplink, respectively. Using high SNR capacity approximation, we state the GDoF region as follows.

Corollary 2 *Assuming* $\alpha_{dl} = \alpha_{ul} = 1$, *the* GDoF *region of* $(M_{dl}, N_{dl}, M_{ul}, N_{ul})$ *side-channel-assisted MIMO full-duplex network satisfies the following constraints*

$$\mathsf{DoF}_{dl} \leq m_{dl}, \ \ \mathsf{DoF}_{ul} \leq m_{ul},$$
$$\mathsf{DoF}_{dl} + \mathsf{DoF}_{ul} \leq f\left(N_{ul}, \left((1-\alpha_I)^+, m_I\right), \left(1, (M_{ul}-N_{dl})^+\right)\right)$$
$$+ f\left(N_{dl}, (\alpha_I, M_{ul}), (1, M_{dl})\right) + Wf\left(N_{dl}, (\alpha_S, M_{ul})\right),$$
$$(15)$$

where $m_{dl} = \min\{M_{dl}, N_{dl}\}, m_{ul} = \min\{M_{ul}, N_{ul}\}, m_I = \min\{M_{ul}, N_{dl}\}$ *as defined in* (8); *function* $f\left(x, (y_1, x_1), (y_2, x_2)\right) = \min\{x, x_1\}y_1^+ + \min\{(x-x_1)^+, x_2\}y_2^+$ *for* $y_1 \geq y_2$.

Proof The proof is akin to [17] (see Appendix 3), so we will only provide an interpretation of the GDoF result here.

First, the DoF of downlink and uplink is limited by the number of transmit and receive antennas, much like the point-to-point MIMO channel. Next, we will explain the sum GDoF. Let $\mathsf{DoF}_{ul,c}$ and $\mathsf{DoF}_{ul,p}$ denote the DoF of the uplink common message and private message, respectively.

Adopting the singular value decomposition (SVD), we can decompose the interference channel as $H_I = U\Lambda V^\dagger$, where U and V are $N_{dl} \times N_{dl}$ and $M_{ul} \times M_{ul}$ unitary matrices, respectively, Λ is $N_{dl} \times M_{ul}$ diagonal matrix containing singular values of H_I. Thus, H_I is decomposed into m_I parallel channels, leaving $(M_{ul}-m_I)^+ = (M_{ul}-N_{dl})^+$ effective inputs at uplink transmitter that do not cause any interference to the downlink receiver. The uplink transmitter divides the private streams into two parts. The first part is sent along the $(M_{ul}-N_{dl})^+$-dimensional null space of interference channel H_I and reaches BS at an SNR of ρ with N_{ul} receive antennas. In the remaining m_I dimensions, the second part is transmitted at a power level of $\rho^{-\alpha_I}$ such that it reaches the unintended receiver at the noise floor and reaches BS at an SNR of $\rho^{(1-\alpha_I)^+}$. The process can be viewed as a combination of signal space and signal scale interference alignment. Thus, the DoF of the uplink private message is

$$\mathsf{DoF}_{ul,p} = f\left(N_{ul}, \left((1-\alpha_I)^+, m_I\right), \left(1, (M_{ul}-N_{dl})^+\right)\right).$$
$$(16)$$

Since the common message can be decoded at both receivers, the downlink receiver with N_{dl} receive antennas is a side-channel-assisted multiple-access channel receiver. The downlink message ω_{dl} reaches the downlink receiver at an SNR of ρ with M_{dl} transmit antennas. The uplink common message $\omega_{ul,c}$ reaches the downlink receiver through both main-channel at an SNR of ρ^{α_I} and side channel as an orthogonal spectral space at an SNR of $\rho^{W\alpha_S}$ with M_{ul} transmit antennas. Thus, we have

$$\mathsf{DoF}_{dl} + \mathsf{DoF}_{ul,c} = f\left(N_{dl}, (\alpha_I, M_{ul}), (1, M_{dl})\right)$$
$$+ Wf\left(N_{dl}, (\alpha_S, M_{ul})\right).$$
$$(17)$$

Combining (16) and (17) leads to the sum GDoF. □

Remark 1 *When $W = 0$, i.e., there is no side channel, the* GDoF *is the same as that of MIMO Z-interference channel in* [17]; *hence, we conclude that the implicit feedback at the full-duplex capable BS does not help improve* GDoF *regime in the two-user MIMO full-duplex network.*

This is due to the fact that there is only one-sided interference. When $W > 0$, the implicit feedback is still not useful in terms of GDoF, *because our scheme does not rely on any feedback.*

3.4 Special cases
In this section, we give several special cases to illustrate the GDoF results above.

Theorem 2 *(Case A) When $M_{dl} = M_{ul} = M, N_{dl} = N_{ul} = N$, and $\alpha_{ul} = \alpha_{dl} = 1$, the sum* GDoF *per antenna denoted as $\frac{\mathsf{GDoF}_{sum}}{\min(M,N)}$ for the symmetric side-channel-assisted MIMO full-duplex network is given by*

$$\frac{\mathsf{GDoF}_{sum}}{\min(M,N)}$$
$$= \begin{cases} \min\left\{2, 2 - \left(2 - \frac{\max(M,N)}{\min(M,N)}\right)^+ \alpha_I + W\alpha_S\right\} & \alpha_I < 1, \\ \min\left\{2, \alpha_I + \frac{\max(M,N)}{\min(M,N)} - 1 + W\alpha_S\right\} & \alpha_I \geq 1. \end{cases}$$

In this case, one can observe that the sum GDoF per antenna increases linearly with the antenna ratio $\frac{\max(M,N)}{\min(M,N)}$ and side-channel quality $W\alpha_S$.

Another case of interest is when the BS has more antennas than mobile clients, i.e., $M_{dl}, N_{ul} \geq M_{ul}, N_{dl}$. This scenario is almost always true in practical systems and the ongoing trend is that the BS can accommodate many antennas such as in massive MIMO systems [15], while the small-form factor mobiles will have a relatively fewer antennas due to its physical size constraint.

Theorem 3 *(Case B) When BS has more antennas than mobiles, i.e., $M_{dl}, N_{ul} \geq M_{ul}, N_{dl}$ with $\alpha_{ul} = \alpha_{dl} = 1$, the sum* GDoF *per antenna denoted as $\frac{\mathsf{GDoF}_{sum}}{\min(M_{ul}, N_{dl})}$ is given as*

$$\frac{\mathsf{GDoF}_{sum}}{\min(M_{ul}, N_{dl})}$$
$$= \begin{cases} \min\left\{\frac{m_X}{m_I} + 1, \frac{m_X}{m_I} + 1 - \alpha_I + W\alpha_S\right\} & \alpha_I < 1 \\ \min\left\{\frac{m_X}{m_I} + 1, \frac{m_X}{m_I} - 1 + \alpha_I + W\alpha_S\right\} & \alpha_I \geq 1. \end{cases}$$

*where* $m_X = \max(M_{ul}, N_{dl}), m_I = \min(M_{ul}, N_{dl})$.

Figure 5 illustrates how the sum GDoF per antenna varies as the side-channel quality changes when $M_{ul} = N_{dl}$ given an excess of antennas at BS. When $W\alpha_S = 0$, i.e., there is no side channel, the curve maintains "V" shape as in the Z-interference channel. When $W\alpha_S$ increases, the curve gradually becomes a lifted "V" and finally reach the maximum sum GDoF per antenna of 2 for all regimes that one can achieve without interference.

We also give an example to clarify the DoF of vector bin-and-cancel in Case B assuming $\alpha_I = \alpha_S = 1$. Using the standard MIMO SVD of channel matrices, the interference channel and side channel can be converted to $m_I = \min\{N_{dl}, M_{ul}\}$ parallel paths from uplink node Tx_U to downlink node Rx_D. In Fig. 6, the diagonalized interference and side-channel paths are depicted in bold.

In Fig. 6, the base station Tx_B sends N_{dl} independent streams to downlink node Rx_D, which is indicated by the black circles. Uplink node Tx_U sets $(1 - W)m_I$ effective inputs[5] to zero, which is indicated by the white circles; Tx_U then sends $(M_{ul} - N_{dl})^+$-independent private streams in the null space of the signal from Tx_B and Wm_I common message which can be heard at Rx_D. Using vector bin-and-cancel, each transmitter sends Wm_I streams of its common message to the interfering receiver through the side channel, which is indicated by the blue circles. At the downlink receiver Rx_D, Wm_I streams of the interfering message can be canceled out; thus, downlink can achieve N_{dl} DoFs and uplink can achieve $\min\{(M_{ul} - N_{dl})^+ + Wm_I, M_{ul}\}$ DoFs. Thus, in total, we can obtain $\min\{\max\{N_{dl}, M_{ul}\} + Wm_I, N_{dl} + M_{ul}\}$ DoFs.

3.5 GDoF without CSIT
Acquiring the CSIT incurs a large overhead, especially in a MIMO system with many antennas. Hence, it is of practical interest to study the GDoF performance of the system without CSIT.
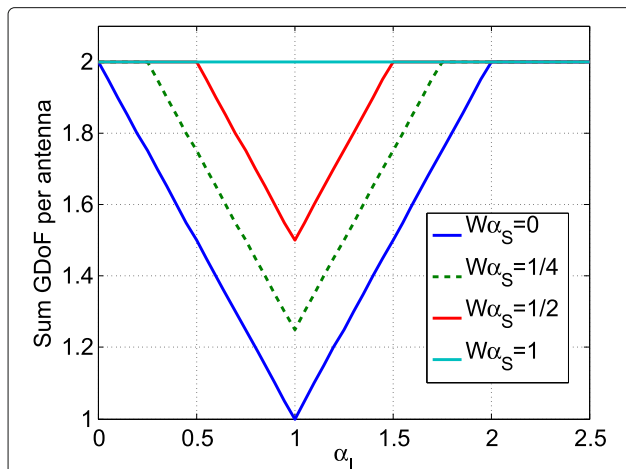
Fig. 5 The sum GDoF per antenna for $M_{ul} = N_{dl}$ when BS has an excess of antennas
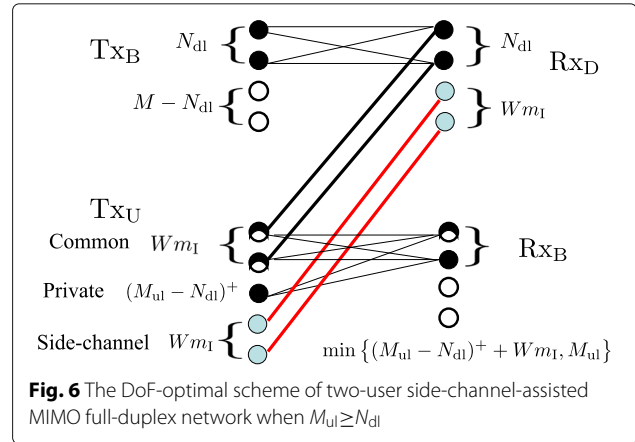
Fig. 6 The DoF-optimal scheme of two-user side-channel-assisted MIMO full-duplex network when $M_{ul} \geq N_{dl}$

We first describe the encoding and decoding strategy under the no-CSIT assumption. Both transmitters encode their messages using independent Gaussian codebooks for the main-channel. The uplink transmitter sends common message only and applies vector bin-and-cancel scheme. The side-channel bins all the uplink message and encodes the bin indices using an independent Gaussian codebook. From the downlink user's perspective, the channel is a MAC with side channel. At the decoding process, the downlink user uses joint maximum likelihood (ML) decoder to decode both downlink message and uplink messages with the help of side channel. Hence, we can obtain the achievable rate region $\mathcal{R}^{\text{No-CSIT}}$ as

$$
\begin{aligned}
\mathcal{R}^{\text{No-CSIT}} = \Bigg\{ &(R_{dl}, R_{ul}) : R_{dl} \leq W_m \log\left|I_{N_{dl}} + \frac{\rho_{dl}}{M_{dl}} H_{dl} H_{dl}^\dagger\right|, \\
&R_{ul} \leq W_m \min\left\{\log\left|I_{N_{ul}} + \frac{\bar{\lambda}\rho_{ul}}{M_{ul}} H_{ul} H_{ul}^\dagger\right|, \log\left|I_{N_{dl}} + \frac{\bar{\lambda}\rho_I}{M_{ul}} H_I H_I^\dagger\right| \right. \\
&\left. + W\log\left|I_{N_{dl}} + \frac{\lambda\rho_S}{WM_{ul}} H_S H_S^\dagger\right|\right\}, \\
&R_{dl} + R_{ul} \leq W_m \left(\log\left|I_{N_{dl}} + \frac{\rho_{dl}}{M_{dl}} H_{dl} H_{dl}^\dagger + \frac{\bar{\lambda}\rho_I}{M_{ul}} H_I H_I^\dagger\right|\right. \\
&\left. + W\log\left|I_{N_{dl}} + \frac{\lambda\rho_S}{WM_{ul}} H_S H_S^\dagger\right|\right)\Bigg\},
\end{aligned}
$$

$$(18)$$

where $\lambda \in (0, 1)$, for instance, we can fix $\lambda = \bar{\lambda} = 0.5$. The achievable rate region given above can be calculated easily from Eq. (12) with uplink private message set to null and equal power allocation among transmit antennas which does not require any CSIT.

Now, we can obtain the lower bound of the GDoF under the no-CSIT assumption.

Corollary 3 *Assuming $\alpha_{dl} = \alpha_{ul} = 1$ and no-CSIT, the achievable GDoF region of $(M_{dl}, N_{dl}, M_{ul}, N_{ul})$*

side-channel-assisted MIMO full-duplex network satisfies the following constraints

$$
\begin{aligned}
\mathsf{DoF}_{\mathrm{dl}} &\leq m_{\mathrm{dl}}, \ \ \mathsf{DoF}_{\mathrm{ul}} \leq \min\{m_{\mathrm{ul}}, \alpha_{\mathrm{I}} m_{\mathrm{I}} + W\alpha_{\mathrm{S}} m_{\mathrm{I}}\}, \\
\mathsf{DoF}_{\mathrm{dl}} + \mathsf{DoF}_{\mathrm{ul}} &\leq f\big(N_{\mathrm{dl}}, (\alpha_{\mathrm{I}}, M_{\mathrm{ul}}), (1, M_{\mathrm{dl}})\big) \\
&\quad + Wf\big(N_{\mathrm{dl}}, (\alpha_{\mathrm{S}}, M_{\mathrm{ul}})\big).
\end{aligned}
\tag{19}
$$

Proof The achievable GDoF region without CSIT can be derived following the same argument as in the case with CSIT. □

Remark 2 *Comparing Corollaries* 2 *and* 3, *we conclude that when* $\alpha_{\mathrm{I}} \geq 1$ *and* $N_{\mathrm{dl}} \geq M_{\mathrm{ul}}$, *acquiring CSIT is of no use as the* GDoF *without CSIT achieves the optimal* GDoF *with CSIT. In the strong interference regime where* INR > SNR, *larger number of receiver antennas is sufficient to null out the interference to achieve the optimal* GDoF *regime.*

3.6 Spatial and spectral tradeoff in GDoF

In this section, we will compare three systems: (i) the side-channel-assisted full-duplex network with CSIT, (ii) the side-channel-assisted full-duplex network without CSIT, and (iii) an idealized full-duplex network without interference, i.e., a parallel uplink and a downlink channel; the last network provides us the benchmark for the best possible performance. By comparing these three systems, we aim to quantify the relationship between the spatial resources of multiple antennas and spectral resources of the side channel. We start by presenting several corollaries to Theorems 2 and 3.

Corollary 4 *(Case A with CSIT) The effect of interference can be completely eliminated if the bandwidth ratio of the side channel to the main channel satisfies the following condition,*

$$
W_{CSIT} = \begin{cases} \dfrac{\alpha_{\mathrm{I}}}{\alpha_{\mathrm{S}}}\left(2 - \dfrac{\max(M,N)}{\min(M,N)}\right)^{+}, & \text{for } \alpha_{\mathrm{I}} < 1, \\[2.5ex] \dfrac{1}{\alpha_{\mathrm{S}}}\left(3 - \dfrac{\max(M,N)}{\min(M,N)} - \alpha_{\mathrm{I}}\right)^{+}, & \text{for } \alpha_{\mathrm{I}} \geq 1. \end{cases}
\tag{20}
$$

From Corollary 4, we can see that the required bandwidth ratio is a linearly decreasing function of the antenna number ratio $\frac{\max(M,N)}{\min(M,N)}$ to achieve the interference-free performance. Therefore, the spatial resources of the number of antennas at transmitters and receivers is interchangeable with the spectral resources of the side-channel bandwidth to eliminate interference. The intuition behind it is that the additional spatial signaling dimension to perform transmit/receive beamforming is equivalent to leveraging the extra spectral signaling dimension of the side channel for interference cancelation.

From Corollary 3, we can also find out the required bandwidth ratio under the no-CSIT assumption in order to achieve the no-interference upper bound. The required bandwidth ratio without CSIT in Case A for $\alpha_{\mathrm{I}} = 1$ is given by

$$
W_{\text{No-CSIT}} = \begin{cases} \dfrac{1}{\alpha_{\mathrm{S}}}\left(2 - \dfrac{N}{M}\right)^{+}, & \text{for } N \geq M, \\[2ex] \dfrac{1}{\alpha_{\mathrm{S}}}, & \text{for } M > N. \end{cases}
\tag{21}
$$

Corollary 5 *(Case B with CSIT) The effect of interference can be completely eliminated if the bandwidth ratio of the side channel to the main channel satisfies the following condition:*

$$
W_{CSIT} = \begin{cases} \dfrac{\alpha_{\mathrm{I}}}{\alpha_{\mathrm{S}}}, & \text{for } \alpha_{\mathrm{I}} < 1, \\[2ex] \dfrac{(2 - \alpha_{\mathrm{I}})^{+}}{\alpha_{\mathrm{S}}}, & \text{for } \alpha_{\mathrm{I}} \geq 1. \end{cases}
$$

We observe that in Case B, the required side-channel bandwidth to achieve the no-interference sum GDoF is not affected by the number of antennas in the system but received interference signal strength and side-channel signal strength levels. For $\alpha_{\mathrm{I}} < 1$, lower interference level requires less side-channel bandwidth while for $\alpha_{\mathrm{I}} \geq 1$, higher interference level leads to smaller side-channel bandwidth requirement.

In Case B, we can also derive the required bandwidth ratio under the no-CSIT assumption from Corollary 3, to achieve the no-interference performance. The required bandwidth ratio without CSIT in Case B for $\alpha_{\mathrm{I}} = 1$ is given by

$$
W_{\text{No-CSIT}} = \begin{cases} \dfrac{1}{\alpha_{\mathrm{S}}}, & \text{for } N_{\mathrm{dl}} \geq M_{\mathrm{ul}}, \\[2ex] \dfrac{M_{\mathrm{ul}}}{N_{\mathrm{dl}}\alpha_{\mathrm{S}}}, & \text{for } M_{\mathrm{ul}} > N_{\mathrm{dl}}. \end{cases}
\tag{22}
$$

In Figs. 7 and 8, we show the spatial and spectral tradeoff in both Case A and Case B when $\alpha_{\mathrm{I}} = 1$. We observe
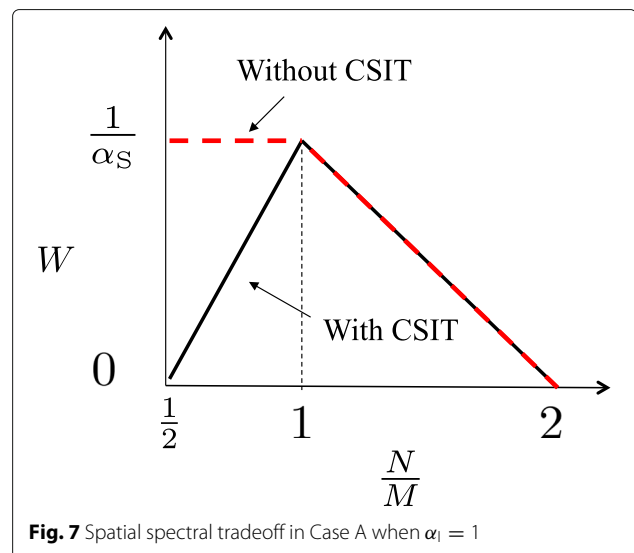
Fig. 7 Spatial spectral tradeoff in Case A when $\alpha_{\mathrm{I}} = 1$
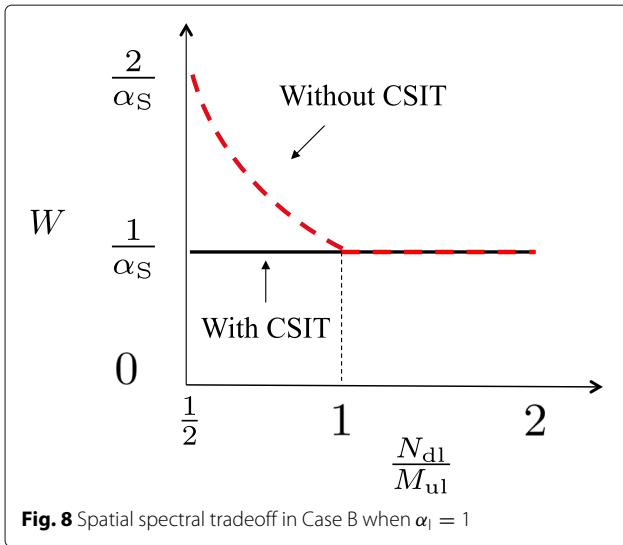
Fig. 8 Spatial spectral tradeoff in Case B when $\alpha_I = 1$

that when there are more downlink receive antennas than uplink transmit antennas, obtaining CSIT is unavailing since with and without CSIT requires the same amount of side-channel bandwidth to completely eliminate interference. However, when we have more uplink transmit antennas, if we do not have CSIT, the extra spatial degrees-of-freedom are wasted and we need more side-channel bandwidth to achieve the no-interference performance.

In Fig. 9, we give an illustration of the comparisons of the three systems in DoF as a function of the side-channel bandwidth when there is an excess of BS antennas.

4 Diversity and multiplexing tradeoff of MIMO-distributed full-duplex

In this section, we consider a slow-fading scenario. When the channel experiences slow fading, an important metric to characterize the MIMO system performance is the diversity and multiplexing tradeoff (DMT), which delineates the asymptotic tradeoff between data rate and reliability in the high SNR limit. The optimal DMT, first introduced in MIMO point-to-point channels [18], represents the optimal diversity gain $d^*(r)$ for each multiplexing gain r among all possible schemes. Similar to

our definition of GDoF, we define the multiplexing gain of both downlink and uplink channels in our system as follows:

$$r_i = \lim_{\rho \to \infty} \frac{R_i(\rho_i)}{W_m \log \rho}, \ i \in \{\mathrm{dl, ul}\}, \quad (23)$$

where R_{dl} and R_{ul} are the achievable rates (bit/s) of downlink and uplink, respectively.

Assuming the overall average error probability is $P_e(r_{\mathrm{dl}}, r_{\mathrm{ul}})$, the DMT is

$$d(r_{\mathrm{dl}}, r_{\mathrm{ul}}) = \lim_{\rho \to \infty} \frac{-\log P_e(r_{\mathrm{dl}}, r_{\mathrm{ul}})}{\log \rho}. \quad (24)$$

We define $d^{\mathrm{opt}}(r_{\mathrm{dl}}, r_{\mathrm{ul}})$ as the supremum of $d(r_{\mathrm{dl}}, r_{\mathrm{ul}})$ computed over all possible schemes. Thus, $d^{\mathrm{opt}}(r_{\mathrm{dl}}, r_{\mathrm{ul}})$ is the optimal DMT of the system.

In this section, we will study the DMT performance under different assumptions regarding the availability of CSIT. We assume that the channel knowledge is known at the receivers. In the following, we will first obtain the optimal DMT with CSIT which can be achieved by vector bin-and-cancel as described in the Section 3. Next, we study the case without CSIT and derive the corresponding achievable DMT. Finally, based on the DMT result, we will investigate the spatial and spectral tradeoff as well as the interplay between CSIT and side channel.

4.1 With CSIT case

In a slow-fading scenario, the channel matrices remain fixed over a fade period with a short-term power constraint given in (4); thus, the capacity region in time-invariant channels can serve as instantaneous capacity region in each fade period. We define the outage event as the target rate pair not contained in the instantaneous capacity region: $\mathfrak{B} \triangleq \{(R_{\mathrm{dl}}, R_{\mathrm{ul}}) \notin \mathcal{C}(\mathcal{H})\}$, where $\mathcal{C}(\mathcal{H})$ is given in Corollary 1. From [18], it can be easily shown that $P_e^*(r_{\mathrm{dl}}, r_{\mathrm{ul}}) \doteq \mathrm{Pr}(\mathfrak{B})$, where $P_e^*(r_{\mathrm{dl}}, r_{\mathrm{ul}})$ is the infimum of the overall average error probability among all possible schemes. In the high SNR limit, we can obtain that

$$\mathrm{Pr}(\mathfrak{B}) \doteq \max_{i \in \{\mathrm{dl,ul,sum}\}} \mathrm{Pr}(C_i < R_i), \implies \rho^{-d^*(r_{\mathrm{dl}}, r_{\mathrm{ul}})}$$

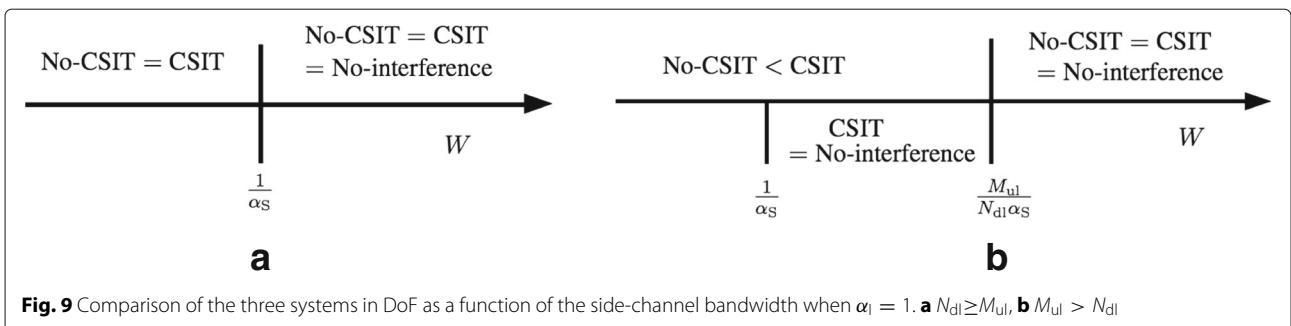$$\doteq \max_{i \in \{\mathrm{dl,ul,sum}\}} \mathrm{Pr}(C_i < R_i),$$

Fig. 9 Comparison of the three systems in DoF as a function of the side-channel bandwidth when $\alpha_I = 1$. **a** $N_{\mathrm{dl}} \geq M_{\mathrm{ul}}$, **b** $M_{\mathrm{ul}} > N_{\mathrm{dl}}$

where C_i is given in (13) and $R_{\text{sum}} = R_{\text{dl}} + R_{\text{ul}}$. Thus the optimal diversity order is

$$d^*(r_{\text{dl}}, r_{\text{ul}}) = \min_{i \in \{\text{dl,ul,sum}\}} d_{\mathfrak{B}_i}(r_i), \text{ where } d_{\mathfrak{B}_i}(r_i)$$
$$= \lim_{\rho \to \infty} -\frac{\log \Pr(C_i < W_m r_i \log \rho)}{\log \rho}, \quad (25)$$

In the Section 3, we showed that vector bin-and-cancel achieves the asymptotic capacity region. Hence, in the asymptotic DMT characterization, the optimal DMT with CSIT can be achieved by vector bin-and-cancel which only requires CSIT of the interference channel between the up- and downlink nodes since the uplink message splitting depends on the interference channel. The derivation of the optimal DMT curve of side-channel-assisted MIMO full-duplex network follows from two steps. In [18], we know that the optimal DMT for MIMO point-to-point channel is $d_{M,N}(r) = (M - r)(N - r)$, which is a piecewise linear curve joining the integer point $r \in [0, \min(M, N)]$. For a general channel level $\alpha_i, i \in \{\text{dl, ul}\}$ of a point-to-point channel, we will invoke Lemma 6 (in Appendix 3) for our calculation. Hence, we first obtain the optimal diversity order of each individual downlink and uplink given as

$$d_{\mathfrak{B}_i}(r_i) = \alpha_i d_{M_i, N_i}\left(\frac{r_i}{\alpha_i}\right), \forall r_i \in [0, \min\{M_i, N_i\}\alpha_i],$$
$$i \in \{\text{dl, ul}\}. \quad (26)$$

Next, we evaluate $d_{\mathfrak{B}_{\text{sum}}}(r_{\text{sum}})$ in the following lemma.

Lemma 3 *The diversity order with CSIT given the sum multiplexing gain of both uplink and downlink is the minimum of the following objective function:*

$$d_{\mathfrak{B}_{\text{sum}}}(r_{\text{sum}}) = \min_{\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu}} \sum_{i=1}^{m_{\text{dl}}} (M_{\text{dl}} + N_{\text{dl}} + 1 - 2i) \mu_i$$
$$+ \sum_{j=1}^{m_{\text{ul}}} (M_{\text{ul}} + N_{\text{ul}} + 1 - 2j) \sigma_j - (M_{\text{dl}} + N_{\text{dl}}) m_1 \alpha_1$$
$$+ \sum_{k=1}^{m_1} (M_{\text{dl}} + N_{\text{ul}} + M_{\text{ul}} + N_{\text{dl}} + 1 - 2k) \theta_k$$
$$+ \sum_{l=1}^{m_1} (M_{\text{ul}} + N_{\text{dl}} + 1 - 2l) \nu_l$$
$$+ \sum_{i=1}^{m_{\text{dl}}} \sum_{k=1}^{\min\{N_{\text{dl}}-i, M_{\text{ul}}\}} (\alpha_1 - \mu_i - \theta_k)^+$$
$$+ \sum_{j=1}^{m_{\text{ul}}} \sum_{k=1}^{\min\{M_{\text{ul}}-j, N_{\text{dl}}\}} (\alpha_1 - \sigma_j - \theta_k)^+;$$

Subject to
$$\sum_{i=1}^{m_{\text{dl}}} (\alpha_1 - \mu_i)^+ + \sum_{j=1}^{m_{\text{ul}}} (\alpha_2 - \sigma_j)^+ + \sum_{k=1}^{m_1} (\alpha_1 - \theta_k)^+$$
$$+ W \sum_{l=1}^{m_1} (\alpha_S - \nu_l)^+ < r_{\text{sum}};$$
$$0 \le \mu_1 \le \cdots \mu_{m_{\text{dl}}}; 0 \le \sigma_1 \le \cdots \sigma_{m_{\text{ul}}}; 0 \le \theta_1$$
$$\le \cdots \theta_{m_1}; 0 \le \nu_1 \le \cdots \nu_{m_1};$$
$$\mu_i + \theta_k \ge \alpha_1, \forall (i + k) \ge N_{\text{dl}} + 1;$$
$$\sigma_j + \theta_k \ge \alpha_1, \forall (j + k) \ge M_{\text{ul}} + 1, \quad (27)$$

where $\bar{\mu} = \{\mu_1, \cdots, \mu_{m_{\text{dl}}}\}, \bar{\sigma} = \{\sigma_1, \cdots, \sigma_{m_{\text{ul}}}\}, \bar{\theta} = \{\theta_1, \cdots, \theta_{m_1}\}, \bar{\nu} = \{\nu_1, \cdots, \nu_{m_1}\}$ *and* $m_{\text{dl}}, m_{\text{ul}}$ *and* m_1 *are defined in* (8).

Proof We provide the proof in Appendix 4. □

With $d_{\mathfrak{B}_i}$ for $i \in \{\text{dl, ul, sum}\}$ derived above, we have the following theorem which gives the optimal DMT in its most general form, allowing different channel parameters and multiplexing gains for uplink and downlink with arbitrary number of antennas at each node.

Theorem 4 *The optimal DMT of* $(M_{\text{dl}}, N_{\text{dl}}, M_{\text{ul}}, N_{\text{ul}})$ *side-channel-assisted MIMO full-duplex network with CSIT denoted as* $d^{\text{CSIT,opt}}$ *is given by*

$$d^{\text{CSIT,opt}}_{(M_{\text{dl}}, N_{\text{dl}}, M_{\text{ul}}, N_{\text{ul}})} (r_{\text{dl}}, r_{\text{ul}}) = \min_{i \in \{\text{dl,ul,sum}\}} d_{\mathfrak{B}_i}(r_i),$$

where $d_{\mathfrak{B}_i}(r_i)$ *is given in* (26) *and Lemma* 3.

The optimization problem in Lemma 3 is a convex optimization problem [19] with linear constraints, which can be solved using linear programming. The general form of the optimal DMT with CSIT in Theorem 4, though can be calculated using numerical methods, does not result in a closed-form solution. In the following corollary, a closed-form DMT result is derived in the case of single-antenna mobiles communicating with multiple-antenna BS with M transmit and receive antennas, i.e., $M_{\text{dl}} = N_{\text{ul}} = M$.

Corollary 6 *In the case of* $(M, 1, 1, M)$ *with symmetric DMT* $r_{\text{ul}} = r_{\text{dl}} = r$ *when* $\alpha_{\text{dl}} = \alpha_{\text{ul}} = \alpha_1 = 1$. *The closed-form optimal DMT with CSIT is given which completely characterizes the optimal DMT under all side-channel conditions:*

- *when* $W \le \frac{1}{2M+1}$ *and* $W\alpha_S < 1$,

$$d^{\text{CSIT,opt}}_{(M,1,1,M)}(r)$$
$$= \begin{cases} M(1-r), & 0 \le r \le \frac{M+1+(2M+1)W\alpha_S}{3M+2} \\ (2M+1)(1+W\alpha_S) - (4M+2)r, & \frac{M+1+(2M+1)W\alpha_S}{3M+2} \le r \le \frac{1+W\alpha_S}{2} \end{cases}$$
$$(28)$$

- when $\frac{1}{2M+1} \leq W < \frac{2}{M}, \alpha_S \geq \frac{M}{2}$, and $W\alpha_S < 1$,

$$d_{(M,1,1,M)}^{CSIT,opt}(r) = \begin{cases} M(1-r), & 0 \leq r \leq \beta^* \\ \alpha_S + \frac{1}{W}(1-2r), & \beta^* \leq r \leq \frac{1+W\alpha_S}{2} \end{cases}$$
(29)

- when $W \geq \frac{1}{2M+1}, \alpha_S < \frac{M}{2}$, and $W\alpha_S < 1$,

$$d_{(M,1,1,M)}^{CSIT,opt}(r)$$
$$= \begin{cases} M(1-r), & 0 \leq r \leq \frac{M+1+\alpha_S}{3M+2} \\ 2M+1+\alpha_S - (4M+2)r, & \frac{M+1+\alpha_S}{3M+2} \leq r \leq \frac{1}{2} \\ \alpha_S + \frac{1}{W}(1-2r), & \frac{1}{2} \leq r \leq \frac{1+W\alpha_S}{2} \end{cases}$$
(30)

- when $W \geq \frac{1}{2M+1}, \alpha_S < \frac{M}{2}$, and $W\alpha_S \geq 1$,

$$d_{(M,1,1,M)}^{CSIT,opt}(r)$$
$$= \begin{cases} M(1-r), & 0 \leq r \leq \frac{M+1+\alpha_S}{3M+2} \\ 2M+1+\alpha_S - (4M+2)r, & \frac{M+1+\alpha_S}{3M+2} \leq r \leq \frac{1}{2} \\ \alpha_S + \frac{1}{W}(1-2r), & \frac{1}{2} \leq r \leq \beta^* \\ M(1-r), & \beta^* \leq r \leq 1 \end{cases}$$
(31)

- when $\alpha_S \geq \frac{M}{2}$ and $W\alpha_S \geq 1$,

$$d_{(M,1,1,M)}^{CSIT,opt}(r) = M(1-r), 0 \leq r \leq 1$$
(32)

where $\beta^* = \frac{\alpha_S + \frac{1}{W} - M}{\frac{2}{W} - M}$.

Proof The DMT of the point-to-point channel is $M(1-r), \forall r \in [0,1]$. Thus, we only need to solve for the optimization problem given sum multiplexing gain. One way to find the minimum of the optimization problem in Lemma 3 is to apply the Karush-Kuhn-Tucker condition. Here, we will provide another approach which is the key to the proof of a general case. The method we adopt is gradient descent which finds the local optimum. Since the optimization problem we have is convex with linear constraints, the local optimum is actually the global optimum in convex optimization [19]. Hence we can obtain the global optimum via gradient descent algorithm.

We first simplify the objective function of the diversity order in Lemma 3 given sum multiplexing gain. By substituting $\nu_l' = W\nu_l$ in (27), we can express the objective function as

$$d_{sum}^{CSIT} = \min M\mu_1 + M\sigma_1 + (2M+1)\theta_1 + \frac{\nu_1'}{W} - 2M,$$

Subject to $(1-\mu_1)^+ + (1-\sigma_1)^+ + (1-\theta_1)^+$
$$+ (W\alpha_S - \nu_1')^+ < r_{sum};$$
$$\mu_1, \sigma_1, \theta_1, \nu_1' \geq 0;$$
$$\mu_1 + \theta_1 \geq 1; \sigma_1 + \theta_1 \geq 1.$$
(33)

Next, we differentiate the objective function in (33) with respect to different variables

$$\frac{\partial d_{sum}^{CSIT}}{\partial \nu_1'} = \frac{1}{W};$$
(34)

$$\frac{\partial d_{sum}^{CSIT}}{\partial \theta_1} = 2M+1;$$
(35)

$$\frac{\partial d_{sum}^{CSIT}}{\partial \mu_1} = \frac{\partial d_{sum}^{CSIT}}{\partial \sigma_1} = M < \frac{\partial d_{sum}^{CSIT}}{\partial \theta_1}.$$
(36)

Comparing the gradient of each variable, when $W \leq \frac{1}{2M+1}$, the steepest descent of the objective function is along the decreasing value of ν_1' with $\theta_1 = \mu_1 = \sigma_1 = 1$, for $r_{sum} \leq W\alpha_S$. Thus, we have $d_{sum}^{CSIT}(r) = 2M+1+\alpha_S - \frac{r_{sum}}{W}, \forall r_{sum} \in [0, W\alpha_S]$. This also implies that for $r_{sum} \geq W\alpha_s$, $\nu_1' = 0$ in the optimal solution. Now, the steepest descent of the objective function in (33) is along the decreasing value of θ_1 with $\mu_1 = \sigma_1 = 1$, and the corresponding minimum is $d_{sum}^{CSIT}(r_{sum}) = (2M+1)(1+W\alpha_S) - (2M+1)r_{sum}, \forall r_{sum} \in [W\alpha_S, 1+W\alpha_S]$.

When $W \geq \frac{1}{2M+1}$, the steepest descent of the objective function is along the decreasing value of θ_1 with $\mu_1 = \sigma_1 = 1, \nu_1' = W\alpha_S$, for $r_{sum} \leq 1$. Thus, we have $d_{sum}^{CSIT}(r_{sum}) = 2M+1+\alpha_S - (2M+1)r_{sum}, \forall r_{sum} \in [0,1]$. Again, for $r_{sum} \geq 1$, the optimal solution has $\theta_1 = 0$. We will rewrite the objective function as

$$d_{sum}^{CSIT} = \min M\mu_1 + M\sigma_1 + \frac{\nu_1'}{W} - 2M,$$

Subject to $(1-\mu_1)^+ + (1-\sigma_1)^+ + (W\alpha_S - \nu_1')^+$
$$\leq r_{sum} - 1;$$
$$\mu_1, \sigma_1, \nu_1' \geq 0;$$
$$\mu_1 \geq 1; \sigma_1 \geq 1.$$
(37)

To minimize the objective function above, we should let $\mu_1 = \sigma_1 = 1$. Hence, the minimum of the objective function is $d_{sum}^{CSIT}(r_{sum}) = \alpha_S + \frac{1}{W}(1-r_{sum}), \forall r_{sum} \in [1, 1+W\alpha_S]$. Now, combining all the results above, we have

$$d_{(M,1,1,M)}^{CSIT,opt}(r) = \min\{M(1-r), d_{sum(M,1,1,M)}^{CSIT}(r)\}$$
for $0 \leq r \leq 1$.
(38)

where $d_{sum(M,1,1,M)}^{CSIT}(r)$ is given as

- when $W \leq \frac{1}{2M+1}$

$$d_{sum(M,1,1,M)}^{CSIT}(r)$$
$$= \begin{cases} 2M+1+\alpha_S - \frac{2r}{W}, & 0 \leq r \leq \frac{W\alpha_S}{2} \\ (2M+1)(1+W\alpha_S) - (4M+2)r, & \frac{W\alpha_S}{2} \leq r \leq \frac{1+W\alpha_S}{2} \end{cases}$$
(39)

- when $W \geq \frac{1}{2M+1}$

$$d_{\text{sum}(M,1,1,M)}^{\text{CSIT}}(r)$$
$$= \begin{cases} 2M + 1 + \alpha_S - (4M + 2)r, & 0 \leq r \leq \frac{1}{2} \\ \alpha_S + \frac{1}{W}(1 - 2r), & \frac{1}{2} \leq r \leq \frac{1+W\alpha_S}{2} \end{cases}$$
$$(40)$$

Further simplification of (38) will lead to the analytical expression in Corollary 6. □

Remark 3 *The optimal DMT with CSIT in the no side-channel case is a special case of Corollary* 6 *when* $W = 0$, *and is given as*

$$d_{(M,1,1,M)}^{\text{No-SC,CSIT,opt}}(r) = \begin{cases} M(1 - r), & 0 \leq r \leq \frac{M+1}{3M+2} \\ (2M + 1)(1 - 2r), & \frac{M+1}{3M+2} \leq r \leq \frac{1}{2} \end{cases}$$
$$(41)$$

From Corollary 6, we can completely quantify the improvement of DMT with side channel under all side-channel conditions. Figure 10 depicts the comparison of DMT with/without (w/wo) side channel when $W = \frac{1}{2M+1}, \alpha_S = \frac{M}{2}$. We define the light loading threshold as the multiplexing gain threshold within which the system error event is dominated by single-user performance. In the case with CSIT, the light loading threshold of the system without side channel is B shown in Fig. 10. When $r > B$, the dominant error event is that all users are in error. With the help of side channel, the light loading threshold

Fig. 10 DMT comparison w/wo side-channel w/wo CSIT when $W = \frac{1}{2M+1}, \alpha_S = \frac{M}{2}$
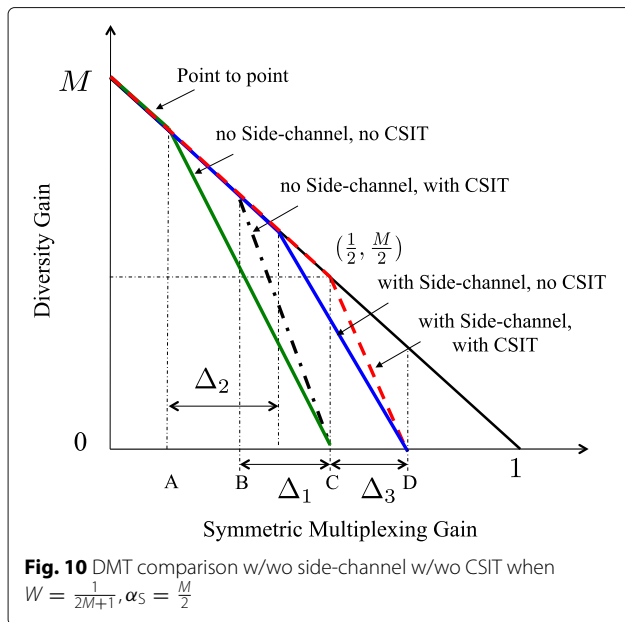
is increased by Δ_1, where $\Delta_1 = \frac{(2M+1)W\alpha_S}{3M+2}$. Moreover, we can see that the side channel also improves system maximum multiplexing gain (when the diversity order is zero) by Δ_3, where $\Delta_3 = \frac{W\alpha_S}{2}$. Both improvement amount Δ_1 and Δ_3 will scale with side-channel quality $W\alpha_S$ (for $W \leq \frac{1}{2M+1}$) till either point C or D reaches the symmetric maximum multiplexing gain of one which corresponds to the no-interference point.

When $W = \frac{1}{2M+1}, \alpha_S = \frac{M}{2}$, we have $\Delta_1 = \frac{M}{6M+4}$ and $\Delta_3 = \frac{M}{8M+4}$. We conclude that in this case, both improvement amount Δ_1 and Δ_3 will scale with the number of antennas at the BS. In the limit of M (as in massive MIMO, BS has unlimited number of antennas), we will have improvement of $\lim_{M \to \infty} \Delta_1 = \frac{1}{6}$ and $\lim_{M \to \infty} \Delta_3 = \frac{1}{8}$.

4.2 Without CSIT case

We define the outage event $\mathbf{O}$ in the case without CSIT as the target rate pair does not lie in the achievable rate region $\mathcal{R}^{\text{No-CSIT}}$: $\mathbf{O} \triangleq \{(R_{\text{dl}}, R_{\text{ul}}) \notin \mathcal{R}\}$, where $\mathcal{R}$ is given (with $\lambda = \bar{\lambda} = 0.5$)

$$\mathcal{R} = \left\{ (R_{\text{dl}}, R_{\text{ul}}) : R_{\text{dl}} \leq W_m \log \left| I_{N_{\text{dl}}} + \frac{\rho_{\text{dl}}}{M_{\text{dl}}} H_{\text{dl}} H_{\text{dl}}^\dagger \right|; \right.$$

$$R_{\text{ul}} \leq W_m \log \left| I_{N_{\text{ul}}} + \frac{\bar{\lambda}\rho_{\text{ul}}}{M_{\text{ul}}} H_{\text{ul}} H_{\text{ul}}^\dagger \right|;$$

$$R_{\text{dl}} + R_{\text{ul}} \leq W_m \left(\log \left| I_{N_{\text{dl}}} + \frac{\rho_{\text{dl}}}{M_{\text{dl}}} H_{\text{dl}} H_{\text{dl}}^\dagger + \frac{\bar{\lambda}\rho_{\text{I}}}{M_{\text{ul}}} H_{\text{I}} H_{\text{I}}^\dagger \right| \right.$$

$$\left. \left. + W \log \left| I_{N_{\text{dl}}} + \frac{\lambda\rho_S}{WM_{\text{ul}}} H_S H_S^\dagger \right| \right) \right\},$$
$$(42)$$

The difference between (42) and the achievable rate region in (18) is that (42) does not have a constraint on R_{ul} for the transmission from up- to downlink mobile. This is because the downlink mobile is not interested in the uplink's message, thereby the failure of decoding uplink's message alone will not be declared as an error event at the downlink receiver.

Under the no-CSIT assumption, the diversity order of each MIMO downlink/uplink channel is still the same as given in (26). As for the diversity order for a given sum multiplexing gain, it is characterized by the following lemma.

Lemma 4 *The diversity order at a given sum multiplexing gain in the case without CSIT is the minimum of the following objective function:*

$$
\begin{aligned}
d_{o_{\text{sum}}}(r_{\text{sum}}) = \min_{\bar{\mu},\bar{\theta},\bar{\nu}} \sum_{i=1}^{m_{\text{dl}}} & (M_{\text{dl}} + N_{\text{dl}} + 1 - 2i)\,\mu_i \\
& + \sum_{k=1}^{m_I} (M_{\text{ul}} + N_{\text{dl}} + M_{\text{dl}} + 1 - 2k)\,\theta_k \\
& + \sum_{l=1}^{m_I} (M_{\text{ul}} + N_{\text{dl}} + 1 - 2l)\,\nu_l - M_{\text{dl}} m_I \alpha_I \\
& + \sum_{i=1}^{m_{\text{dl}}} \sum_{k=1}^{\min\{N_{\text{dl}}-i,M_{\text{ul}}\}} (\alpha_I - \mu_i - \theta_k)^+
\end{aligned}
$$

$$
\begin{aligned}
\text{Subject to} \quad & \sum_{i=1}^{m_{\text{dl}}} (\alpha_{\text{dl}} - \mu_i)^+ + \sum_{k=1}^{m_I} (\alpha_I - \theta_k)^+ \\
& + W \sum_{l=1}^{m_I} (\alpha_S - \nu_l)^+ < r_{\text{sum}}; \\
& 0 \le \mu_1 \le \cdots \le \mu_{m_{\text{dl}}}; \; 0 \le \theta_1 \le \cdots \\
& \le \theta_{m_I}; \; 0 \le \nu_1 \le \cdots \le \nu_{m_I}; \\
& \mu_i + \theta_k \ge \alpha_I, \; \forall (i+k) \ge N_{\text{dl}} + 1;
\end{aligned}
\tag{43}
$$

Proof We provide the proof in Appendix 5. □

Theorem 5 *A lower bound of the DMT of* $(M_{\text{dl}}, N_{\text{dl}}, M_{\text{ul}}, N_{\text{ul}})$ *side-channel-assisted MIMO full-duplex network without CSIT is given as*

$$
d_{(M_{\text{dl}},N_{\text{dl}},M_{\text{ul}},N_{\text{ul}})}^{\textit{No-CSIT}}(r_{\text{dl}}, r_{\text{ul}}) = \min_{i \in \{\text{dl,ul,sum}\}} d_{o_i}(r_i).
$$

where $d_{o_i}(r_i)$ *is given in* (26) *and Lemma* 4.

While the general form of the lower bound of DMT without CSIT is given in Theorem 5, we provide the closed-form no-CSIT DMT in the following corollary when single-antenna mobiles communicate with multiple-antenna BS in line with the analysis in the Section 4.1.

Corollary 7 *In the case of* $(M, 1, 1, M)$ *with symmetric DMT* $r_{\text{ul}} = r_{\text{dl}} = r$ *when* $\alpha_{\text{dl}} = \alpha_{\text{ul}} = \alpha_I = 1$. *The closed-form lower bound of the DMT without CSIT is given that completely characterizes the achievable DMT under all side-channel conditions:*

- *when* $W \le \frac{1}{M+1}$ *and* $W\alpha_S < 1$,

$$
d_{(M,1,1,M)}^{\textit{No-CSIT}}(r)
$$
$$
= \begin{cases} M(1-r), & 0 \le r \le \frac{1+(M+1)W\alpha_S}{M+2} \\ (M+1)(1+W\alpha_S) - (2M+2)r, & \frac{1+(M+1)W\alpha_S}{M+2} \le r \le \frac{1+W\alpha_S}{2} \end{cases}
\tag{44}
$$

- *when* $\frac{1}{M+1} \le W < \frac{2}{M}, \alpha_S \ge \frac{M}{2}$, *and* $W\alpha_S < 1$,

$$
d_{\text{sum}(M,1,1,M)}^{\textit{No-CSIT}}(r) = \begin{cases} M(1-r), & 0 \le r \le \beta^* \\ \alpha_S + \frac{1}{W}(1-2r), & \beta^* \le r \le \frac{1+W\alpha_S}{2} \end{cases}
\tag{45}
$$

- *when* $W \ge \frac{1}{M+1}, \alpha_S < \frac{M}{2}$, *and* $W\alpha_S < 1$,

$$
d_{\text{sum}(M,1,1,M)}^{\textit{No-CSIT}}(r) = \begin{cases} M(1-r), & 0 \le r \le \frac{1+\alpha_S}{M+2} \\ M+1+\alpha_S - (2M+2)r, & \frac{1+\alpha_S}{M+2} \le r \le \frac{1}{2} \\ \alpha_S + \frac{1}{W}(1-2r), & \frac{1}{2} \le r \le \frac{1+W\alpha_S}{2} \end{cases}
\tag{46}
$$

- *when* $W \ge \frac{1}{M+1}, \alpha_S < \frac{M}{2}$, *and* $W\alpha_S \ge 1$,

$$
d_{\text{sum}(M,1,1,M)}^{\textit{No-CSIT}}(r) = \begin{cases} M(1-r), & 0 \le r \le \frac{1+\alpha_S}{M+2} \\ M+1+\alpha_S - (2M+2)r, & \frac{1+\alpha_S}{M+2} \le r \le \frac{1}{2} \\ \alpha_S + \frac{1}{W}(1-2r), & \frac{1}{2} \le r \le \beta^* \\ M(1-r), & \beta^* \le r \le 1 \end{cases}
\tag{47}
$$

- *when* $\alpha_S \ge \frac{M}{2}$ *and* $W\alpha_S \ge 1$,

$$
d_{\text{sum}(M,1,1,M)}^{\textit{No-CSIT}}(r) = M(1-r), 0 \le r \le 1
\tag{48}
$$

where $\beta^* = \frac{\alpha_S + \frac{1}{W} - M}{\frac{2}{W} - M}$.

Proof The proof is similar to that in Corollary 6 which uses gradient descent method. □

Remark 4 *The lower bound of the DMT without CSIT in the no side-channel case is a special case of Corollary* 7 *when* $W = 0$, *and is given by*

$$
d_{(M,1,1,M)}^{\textit{No-SC,No-CSIT}}(r) = \begin{cases} M(1-r), & 0 \le r \le \frac{1}{M+2} \\ (M+1)(1-2r), & \frac{1}{M+2} \le r \le \frac{1}{2} \end{cases}
\tag{49}
$$

Remarks 3 and 4 describe the DMT without side channel under CSIT and no-CSIT assumptions. One can easily verify that the no-side-channel cases in [20, 21] w/wo CSIT are special cases incorporated in our derivation of DMT.

Now we compare the lower bound of the DMT w/wo side channel under the no-CSIT assumption. When $W \le \frac{1}{M+1}$, in the case without CSIT, with the help of side channel, the light loading threshold over the no-side-channel system is increased by Δ_2, where $\Delta_2 = \frac{(M+1)W\alpha_S}{M+2}$. In Fig. 10, the DMT without CSIT w/wo side channel is given when $W = \frac{1}{2M+1}$ and $\alpha_S = \frac{M}{2}$. Compared with the light loading improvement under the CSIT assumption, we can

see that the side channel is more effective in increasing the DMT performance in the lack of CSIT as $\Delta_2 \geq \Delta_1$.

4.3 Spatial and spectral tradeoff in DMT

In this section, we will derive symmetric DMT in closed form for a more general case where the mobiles have multiple antennas communicating with the BS with M transmit and receive antennas. Using the closed-form DMT expressions, again we will compare the three systems: with and without CSIT and the no-interference-idealized full-duplex network. We will characterize the relationship between the spatial degrees-of-freedom of the antenna resources and the extra spectral degrees-of-freedom due to the side channels under slow-fading channels.

We still assume that BS has more antennas, i.e., $M \geq M_{\mathrm{ul}}, N_{\mathrm{dl}}$. The closed-form symmetric DMT of the general $(M, N_{\mathrm{dl}}, M_{\mathrm{ul}}, M)$ system with $\alpha_{\mathrm{dl}} = \alpha_{\mathrm{ul}} = \alpha_{\mathrm{I}} = 1$ and $r_{\mathrm{dl}} = r_{\mathrm{ul}} = r$ are given under CSIT and no-CSIT assumptions in Lemmas 9 and 10 (in Appendix 6), respectively.

First, we ask the question that how much side-channel bandwidth is required to compensate for the lack of CSIT such that the DMT of the system without CSIT achieves that of the system with CSIT. The sufficient condition is given in the following theorem.

Theorem 6 *In case of* $(M, N_{\mathrm{dl}}, M_{\mathrm{ul}}, M)$, *sufficient conditions such that no-CSIT DMT is the same as full CSIT DMT are given by*

1. $W = \min \left\{ \frac{N_{\mathrm{dl}} + M_{\mathrm{ul}} - 1}{M + N_{\mathrm{dl}} - M_{\mathrm{ul}} + 1}, \frac{1}{\alpha_S} \left(2 - \frac{N_{\mathrm{dl}}}{M_{\mathrm{ul}}} \right)^+ \right\}$

    where $\alpha_S \geq \dfrac{d_{M,M_{\mathrm{ul}}}\left(\frac{M_{\mathrm{ul}}}{2}\right) - M(N_{\mathrm{dl}} - M_{\mathrm{ul}})}{M_{\mathrm{ul}} N_{\mathrm{dl}}}$, *when* $N_{\mathrm{dl}} \geq M_{\mathrm{ul}}$,
    $M_{\mathrm{ul}} = 1, 2$;

2. $W = 0$, *when* $N_{\mathrm{dl}} \geq \dfrac{d_{M_{\mathrm{ul}},M}\left(\frac{M_{\mathrm{ul}}}{2}\right)}{M} + M_{\mathrm{ul}}$, $M_{\mathrm{ul}} = 1, 2$.

Proof With the conditions given above, we can verify that the symmetric DMT with CSIT in Lemma 9 is the same as the DMT without CSIT in Lemma 10. □

Corollary 8 *When* $M_{\mathrm{ul}} > N_{\mathrm{dl}}$, *if* $W < \frac{1}{\alpha_S}$, *the DMT without CSIT is strictly smaller than that with CSIT.*

Corollary 8 can be readily obtained by comparing Lemmas 9 and 10. If $M_{\mathrm{ul}} > N_{\mathrm{dl}}$ and $W < \frac{1}{\alpha_S}$, the availability of CSIT is crucial in performing transmit beamforming to yield higher DMT.

The next question we will ask is how much side-channel bandwidth is required to eliminate the effect of interference such that the DMT of the system w/wo CSIT achieves that of a system without interference. The following theorem characterizes the effect of the side-channel

bandwidth on the performance of the symmetric DMT to reach no-interference DMT.

Theorem 7 *In case of* $(M, N_{\mathrm{dl}}, M_{\mathrm{ul}}, M)$, *the sufficient conditions are given under CSIT and no-CIST assumptions, respectively, where the effect of interference can be completely eliminated to achieve the optimal no-interference DMT:*

1. $W_{CSIT} = \frac{1}{\alpha_S} \left(2 - \frac{m_X}{m_{\mathrm{I}}} \right)^+, \ \alpha_S \geq \frac{(2m_{\mathrm{I}} - m_X)(M - m_{\mathrm{I}} + 1)}{m_{\mathrm{I}}(2|N_{\mathrm{dl}} - M_{\mathrm{ul}}| + 2)}$;

2. $W_{No\text{-}CSIT} =$
    $\begin{cases} \frac{1}{\alpha_S}, & \alpha_S \geq \frac{M - N_{\mathrm{dl}} + 1}{2(M_{\mathrm{ul}} - N_{\mathrm{dl}} + 1)}, \ when \ M_{\mathrm{ul}} \geq N_{\mathrm{dl}} \\ \frac{1}{\alpha_S}\left(2 - \frac{m_X}{m_{\mathrm{I}}}\right)^+, & \alpha_S \geq \frac{(2m_{\mathrm{I}} - m_X)(M - m_{\mathrm{I}} + 1)}{m_{\mathrm{I}}(2|N_{\mathrm{dl}} - M_{\mathrm{ul}}| + 2)}, \ when \ N_{\mathrm{dl}} \geq M_{\mathrm{ul}} \end{cases}$

where $m_X = \max(M_{\mathrm{ul}}, N_{\mathrm{dl}}), m_{\mathrm{I}} = \min(M_{\mathrm{ul}}, N_{\mathrm{dl}})$.

Proof We need to show that with the conditions above, the DMT of our system w/wo CSIT is not dominated by the diversity order given sum multiplexing gain $d_{\mathrm{sum}(M,N_{\mathrm{dl}},M_{\mathrm{ul}},M)}^{\mathrm{w/wo\ CSIT}}(r_{\mathrm{sum}}), \ \forall r \in [0, m_{\mathrm{I}}]$. It is sufficient if we show that the conditions above indicate that the decay slope of $d_{\mathrm{sum}(M,N_{\mathrm{dl}},M_{\mathrm{ul}},M)}^{\mathrm{w/wo\ CSIT}}(r_{\mathrm{sum}})$ is larger than that of the PTP channel $d_{M,m_{\mathrm{I}}}(r)\forall r$, and the maximum symmetric multiplexing gain of $d_{\mathrm{sum}(M,N_{\mathrm{dl}},M_{\mathrm{ul}},M)}^{\mathrm{w/wo\ CSIT}}(r_{\mathrm{sum}})$ is larger than m_{I}.

The decay slope of the piecewise linear function $d_{M,N}^k(r)$ is $(M + N - 2k + 1)$ in each interval $r \in [k-1, k]$, where $k \in [1, \min(M, N)]$ is an integer. Thus, the decay slope of $d_{M,N}^k(r)$ decreases as the interval k increases. Also, the decay slope difference between $d_{M,N}^{k-1}(r)$ and $d_{M,N}^k(r)$ is a constant of 2. We know that the DMT performance will be improved as side-channel bandwidth ratio W increases. Therefore, with W large enough, $d_{\mathrm{sum}(M,N_{\mathrm{dl}},M_{\mathrm{ul}},M)}^{\mathrm{w/wo\ CSIT}}(r_{\mathrm{sum}})$ will lastly be dominated by side-channel condition in the last admissible interval. With the special structure of the decay slope, in order to find the conditions where DMT w/wo achieves PTP performance, it suffices to show the following: (A) the decay slope of side channel given sum multiplexing gain is larger than $d_{M,m_{\mathrm{I}}}(r)$ in their last admissible intervals, respectively, and (B) $\max(r_{\mathrm{sum}}) \geq 2m_{\mathrm{I}}$.

Under the CSIT assumption, from Corollary 3, we know the maximum sum multiplexing gain is $m_X + m_{\mathrm{I}} W \alpha_S$. We set $m_X + m_{\mathrm{I}} W \alpha_S = 2m_{\mathrm{I}}$ to meet condition (B); thus, $W = \frac{1}{\alpha_S}\left(2 - \frac{m_X}{m_{\mathrm{I}}}\right)^+$. Next to meet condition (A), the decay slope of the side channel in the last interval $\alpha_S d_{M_{\mathrm{ul}},N_{\mathrm{dl}}}\left(\frac{r_{\mathrm{sum}} - m_X}{W\alpha_S}\right), \ \forall r_{\mathrm{sum}} \in [m_X, m_X + m_{\mathrm{I}} W \alpha_S]$, i.e., $\frac{2}{W}(|M_{\mathrm{ul}} - N_{\mathrm{dl}}| + 1)$, should be larger than the decay slope of $d_{M,m_{\mathrm{I}}}(r), \ \forall r \in [0, m_{\mathrm{I}}]$ in its last interval, i.e., $M - m_{\mathrm{I}} + 1$. Hence, $\frac{2}{W}(|M_{\mathrm{ul}} - N_{\mathrm{dl}}| + 1) \geq (M - m_{\mathrm{I}} + 1)$. By substituting $W = \frac{1}{\alpha_S}\left(2 - \frac{m_X}{m_{\mathrm{I}}}\right)^+$ into the inequality above, we have

$\alpha_S \geq \frac{(2m_1 - m_X)(M - m_1 + 1)}{m_1(2|N_{dl} - M_{ul}| + 2)}$. With the side-channel condition derived above, the DMT with CSIT achieves the PTP DMT.

Under the no-CSIT assumption, when $N_{dl} \geq M_{ul}$, the results can be derived similarly. When $M_{ul} > N_{dl}$, the maximum multiplexing gain is $N_{dl}(1 + W\alpha_S)$ according to Corollary 3. To satisfy condition B, we set $N_{dl}(1 + W\alpha_S) = 2N_{dl}$; hence, we have $W = \frac{1}{\alpha_S}$. To meet condition (A), the decay slope of the side-channel in the last interval $\alpha_S d_{M_{ul}, N_{dl}} \left(\frac{r_{sum} - N_{dl}}{W\alpha_S} \right)$, $\forall r_{sum} \in [N_{dl}, N_{dl}(1 + W\alpha_S)]$, i.e., $\frac{2}{W}(M_{ul} - N_{dl} + 1)$, should be greater than the decay slope of $d_{M, N_{dl}}(r)$ in its last interval, i.e., $(M - N_{dl} + 1)$. By substituting $W = \frac{1}{\alpha_S}$, we obtain that $\alpha_S \geq \frac{M - N_{dl} + 1}{2(M_{ul} - N_{dl} + 1)}$. □

4.4 Discussion of the results

Figure 11 illustrates the comparison of the three systems in DMT as a function of the side-channel bandwidth. When $N_{dl} \geq M_{ul}$, there are three regimes in comparison to DMT. In the first regime, the performance the system without CSIT is worse than that with CSIT. In the second regime, with side-channel bandwidth ratio W greater than a threshold, CSIT is of no use. In the last regime, the use of side-channel helps reduce the probability of outage event where all users are in error such that the dominant error event is a single-user error. On the other hand, when $M_{ul} > N_{dl}$, the availability of CSIT always provide an additional gain in performing transmit beamforming. However, larger side-channel bandwidth aids the no-CSIT system to achieve the no-interference upper bound. Note that the strength of the side-channel level α_S is implicitly incorporated in Theorems 6 and 7, thus is omitted in Fig. 11.

In the following section, we will elaborate the findings in single-antenna mobiles and multiple-antenna mobiles cases, respectively.
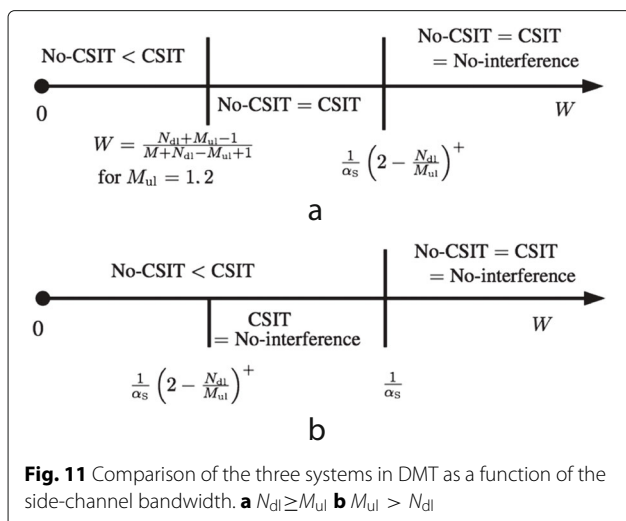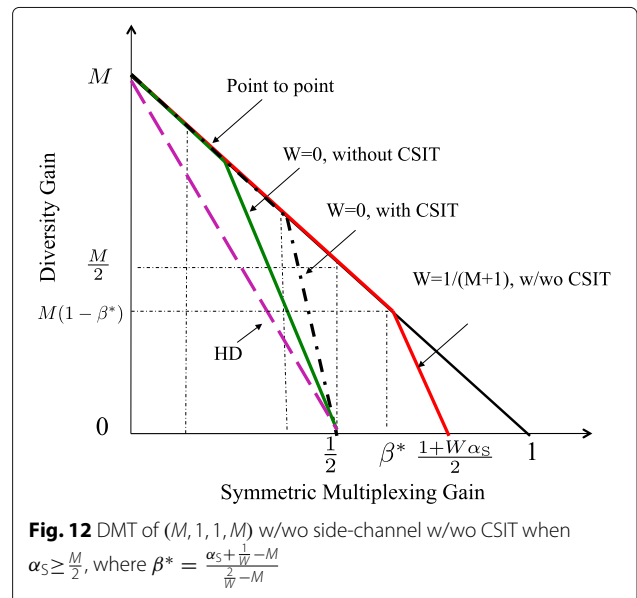
4.4.1 Single-antenna mobiles

We first show the symmetric DMT w/wo side-channel and w/wo CSIT when $\alpha_{dl} = \alpha_{ul} = \alpha_I = 1$. From Fig. 12, we can see that in the two-user uplink and downlink system, the full-duplex capable BS is always superior to its half-duplex (HD) counterpart where the BS adopts either time-division multiplexing (TDM) or frequency-division multiplexing (FDM) for uplink and downlink. In the special case of $W = 0$, i.e., no side-channel, having CSIT always yields a better DMT performance.

However, with the help of side-channel, as shown in Fig. 12, when $W = \frac{1}{M+1}$ and $\alpha_S \geq \frac{M}{2}$, there is no benefit to obtain CSIT as the DMT without CSIT already achieves the optimal DMT with CSIT. Such result indicates that as BS accommodates more antennas (tens or hundreds of BS antennas as in massive MIMO), the required side-channel bandwidth can be reduced superinearly to combat interference.

Figure 13 illustrates the side-channel bandwidth ratio required to compensate for CSIT as stated in Theorem 6 with single-antenna mobiles. The required W is inversely proportional to the antenna resources at the BS. The caveat is that the side-channel level α_S, in the meantime, has to grow with the increasing number of antennas at the BS.

To understand the result above, let us look at the different decay slopes in DMT in the network. From the downlink's viewpoint, the channel is MAC with side channel. The decay slope of MAC without CSIT is $M + 1$, while the the decay slope of the side channel is $\frac{1}{W}$. When the symmetric multiplexing gain $r \leq \frac{1}{2}$, if $W \geq \frac{1}{M+1}$, the users in MAC will first be in error followed by the users' error event in the side channel. Moreover, if $\alpha_S \geq \frac{M}{2}$, the

Fig. 11 Comparison of the three systems in DMT as a function of the side-channel bandwidth. **a** $N_{dl} \geq M_{ul}$ **b** $M_{ul} > N_{dl}$

Fig. 12 DMT of $(M, 1, 1, M)$ w/wo side-channel w/wo CSIT when $\alpha_S \geq \frac{M}{2}$, where $\beta^* = \frac{\alpha_S + \frac{1}{W} - M}{\frac{2}{W} - M}$
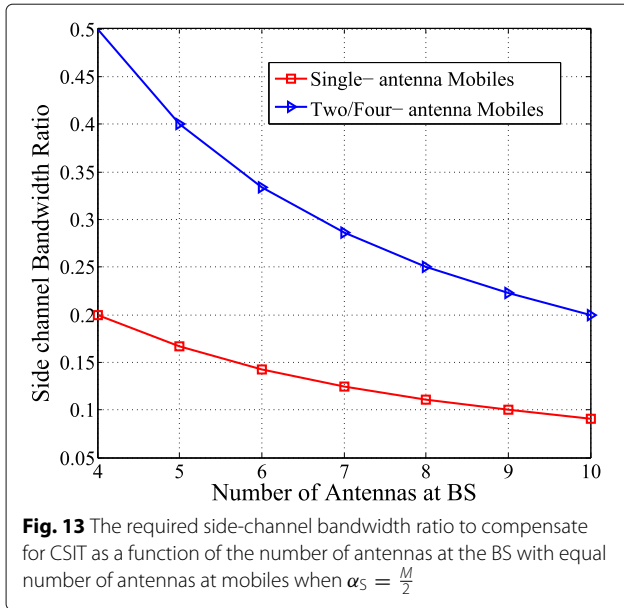
Fig. 13 The required side-channel bandwidth ratio to compensate for CSIT as a function of the number of antennas at the BS with equal number of antennas at mobiles when $\alpha_S = \frac{M}{2}$

error event w/wo CSIT is dominated by single-user performance when $r \leq \frac{1}{2}$. And when $r \geq \frac{1}{2}$, the dominant error event is determined by the side channel, which is the same for both CSIT and no-CSIT cases[6].

In order to eliminate the effect of interference such that the DMT w/wo CSIT achieve no-interference upper bound, it is sufficient if the side-channel condition satisfies that $W\alpha_S \geq 1$ according to Theorem 7. Hence, the required side-channel bandwidth is inversely proportional to the strength of the side channel as to eliminate the effect of interference. The implication of such result is that in a highly clustered urban scenario, when the mobile devices are close to each other indicating higher side-channel strength, less side-channel bandwidth is required to achieve the single-user DMT performance.

4.4.2 Multiple-antenna mobiles

Figure 14 shows the DMT in the absence of the side channel when both the mobiles have multiple antennas. First,
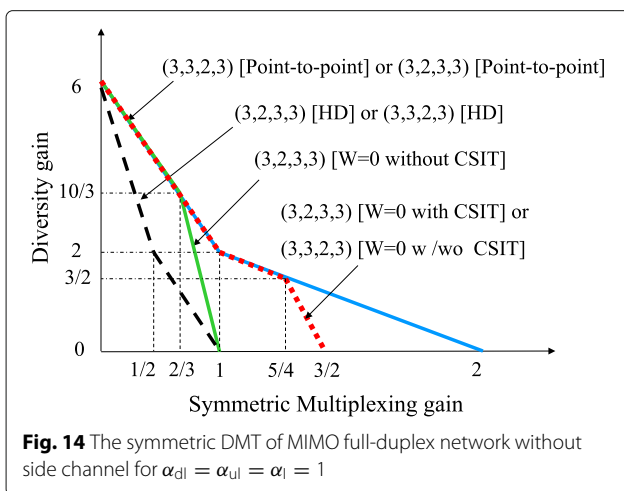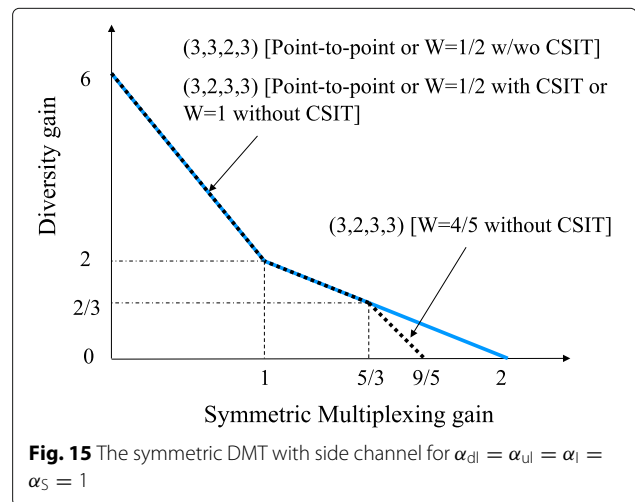
we can find out that the gains due to the full-duplex capable BS over half-duplex BS is particularly larger for MIMO channels. Second, a larger number of downlink receive antennas alone can completely eliminate the effect of CSIT such that the DMT w/wo CSIT have the same performance as stated in Theorem 6. For example, the DMT of $(3,3,2,3)$ without CSIT is the same as that with CSIT. While in the case of $(3,2,3,3)$, the lack of CSIT will result in significant loss.

Comparing Figs. 14 and 15, we can quantify the gains due to the extra side-channel bandwidth, which is significant especially in MIMO. In the case of $(3,2,3,3)$ when the system is lightly loaded, for instance, $r \leq 2/3$, there is no additional gain due to CSIT or $r \leq 5/4$, there is no gain due to the side channel since the error event is dominated by single-user error. Beyond those points, the dominant error event is that all users are in error; thus, leveraging the CSIT for transmit beamforming or side channel to perform vector bin-and-cancel will reduce the probability that such outage event happens.

The required side-channel bandwidth ratio for compensation of CSIT in the case of two-antenna mobiles is also depicted in Fig. 13, which again demonstrates that the required $W \propto \frac{1}{M}$ similar as in the single-antenna-mobile case.

From Theorem 7, we conclude that with CSIT, as the antenna number ratio $\frac{\max\{M_{ul}, N_{dl}\}}{\min\{M_{ul}, N_{dl}\}}$ increases, the side-channel bandwidth required to completely eliminate the effect of interference reduces. Hence, the spatial resources of the multiple antennas at mobiles is interchangeable with the spectral resources of the side-channel bandwidth to reduce the outage probability at a given multiplexing gain such that single-user DMT can be achieved.

We also infer from Theorem 7 that when $M_{ul} > N_{dl}$, the system with CSIT always outperforms that without CSIT by requiring less side-channel bandwidth to reach single-user performance[7]. However, Please check

Fig. 14 The symmetric DMT of MIMO full-duplex network without side channel for $\alpha_{dl} = \alpha_{ul} = \alpha_l = 1$

Fig. 15 The symmetric DMT with side channel for $\alpha_{dl} = \alpha_{ul} = \alpha_l = \alpha_S = 1$

if the edit to the sentence "However, when $N_{dl} \geq M_{ul}$, there is no advantage..." retained its intended meaning. Otherwise, please modify. when $N_{dl} \geq M_{ul}$, there is no advantage due to CSIT achieving the single-user DMT since, with and without CSIT require the same amount of side-channel bandwidth to achieve interference-free performance. Thus, we conclude that having more spatial degree-of-freedom at the interfered downlink receiver or larger side-channel bandwidth can simplify transceiver design by ruling out the necessity of obtaining CSIT to null out the effect of inter-mobile interference.

5 Conclusion

In this paper, we propose the use of wireless side channel to manage inter-mobile interference in MIMO full-duplex network where the BS supports both an up- and downlink flow in the same band simultaneously for the half-duplex mobile nodes. We study if and how the antennas resources at nodes will impact the spectral resource from the side channel under different channel models. For time-invariant channels, we derive a constant-gap capacity region by a vector bin-and-cancel scheme and the corresponding GDoF region. And for slow-fading channels, we obtain DMT w/wo CSIT of the system. Both the GDoF and DMT results reveal various insights about the effect of the side channels and the spatial and spectral tradeoff between antenna resources and bandwidth of the side channels. Our future work will be to develop practical protocols guided by our analysis.

Endnotes

[1] In the rest of the paper, we omit the time-index t in the expressions.

[2] We can set $\rho = \rho_{dl}$ or ρ_{ul} such that either $\alpha_{dl} = 1$ or $\alpha_{ul} = 1$.

[3] With the assistance of the bin index, more uplink common message can be decoded which otherwise is restricted by the interference link.

[4] Notice that our definition deviates slightly from the conventional definition of GDoF in that we account for the asymmetric bandwidths of different links and the rate is calculated as bit/s instead of bit/s/Hz.

[5] The effective input is a product of the unitary matrices by SVD and the initial input vector.

[6] The DMT of MAC channel with CSIT is different from that without CSIT as shown in Fig. 12.

[7] The system with CSIT also has a weaker requirement of the side-channel strength level as compared to that without CSIT.

[8] When $W = 0$, we define $W \log \left(1 + \frac{x}{W}\right) \triangleq 0$.

Appendix 1
Proof of Lemma 1

First, we complete the converse part. Transmitters uniformly and independently generate the downlink and uplink messages ω_{dl} and ω_{ul}, respectively. The messages will be delivered over n time blocks. Since the full-duplex BS has an implicit feedback of infinite capacity link, BS encodes the ω_{dl} by codeword $X_{dl,i}$ which is a function of $(\omega_{dl}, Y_{ul}^{i-1})$, for $i \in [1, n]$. The point-to-point outer bounds on R_{dl} and R_{ul} can be easily obtained following the same argument in Lemma 1 of [22], which are given by

$$
\begin{aligned}
R_{dl} &\leq W_m \left(\log \left| I_{N_{dl}} + \rho_{dl} H_{dl} H_{dl}^{\dagger} \right| \right), \\
R_{ul} &\leq W_m \left(\log \left| I_{N_{ul}} + \bar{\lambda} \rho_{ul} H_{ul} H_{ul}^{\dagger} \right| \right).
\end{aligned}
\tag{50}
$$

Next, we need to prove the sum-capacity upper bound. We define a genie $V_{ul} = \sqrt{\lambda \rho_I} H_I X_{ul} + Z_{dl}$. The sum-capacity upper bound is derived by providing the genie V_{ul}^n to the BS. By Fano's inequality, for any codebook of block length n,

$$
\begin{aligned}
&n(R_{dl} + R_{ul} - \epsilon_n) \\
&\leq I\left(\omega_{dl}; Y_{dl}^n, Y_S^n\right) + I\left(\omega_{ul}; Y_{ul}^n | \omega_{dl}\right) \tag{51} \\
&= I\left(\omega_{dl}; Y_{dl}^n\right) + I\left(\omega_{dl}; Y_S^n | Y_{dl}^n\right) + I\left(\omega_{ul}; Y_{ul}{}^n | \omega_{dl}\right) \tag{52} \\
&= I\left(\omega_{dl}; Y_{dl}^n\right) + h\left(Y_S^n | Y_{dl}^n\right) - h\left(Y_S^n | Y_{dl}^n, \omega_{dl}\right) \\
&\quad + I\left(\omega_{ul}; Y_{ul}^n | \omega_{dl}\right) \tag{53} \\
&\leq I\left(\omega_{dl}; Y_{dl}^n\right) + h\left(Y_S^n\right) - h\left(Y_S^n | X_S^n, Y_{dl}^n, \omega_{dl}\right) \\
&\quad + I\left(\omega_{ul}; Y_{ul}^n | \omega_{dl}\right) \tag{54} \\
&= h\left(Y_{dl}^n\right) + h\left(Y_S^n\right) - h\left(Y_S^n | X_S^n\right) \\
&\quad + \underbrace{h\left(Y_{ul}^n | \omega_{dl}\right) - h\left(Y_{dl}^n | \omega_{dl}\right) - h\left(Y_{ul}^n | \omega_{ul}, \omega_{dl}\right)}_{U} \tag{55}
\end{aligned}
$$

where (51) follows due to the independence of messages; (54) follows because conditioning reduces entropy; (55) follows because $(Y_{dl}^n, \omega_{dl}) \to X_S^n \to Y_S^n$ forms a Markov chain.

We can rewrite $h(Y_{ul}^n | \omega_{ul}, \omega_{dl})$ in (55) in U as

$$
\begin{aligned}
h\left(Y_{ul}^n | \omega_{ul}, \omega_{dl}\right) &= \sum_{i=1}^n h\left(Y_{ul,i} | Y_{ul}^{i-1}, \omega_{ul}, \omega_{dl}\right) \tag{56} \\
&= \sum_{i=1}^n h\left(Y_{ul,i} | X_{ul,i}, Y_{ul}^{i-1}, \omega_{ul}, \omega_{dl}\right) \tag{57} \\
&= \sum_{i=1}^n h(Z_{ul,i}) \tag{58}
\end{aligned}
$$

where (57) follows because $X_{ul,i}$ is a function of ω_{ul} and conditioned on $X_{ul,i}$, $Y_{ul,i}$ is independent of everything else.

We also rewrite $h(Y_{\mathrm{ul}}^n|\omega_{\mathrm{dl}}) - h(Y_{\mathrm{dl}}^n|\omega_{\mathrm{dl}})$ in (55) in U as

$$h\left(Y_{\mathrm{ul}}^n|\omega_{\mathrm{dl}}\right) - h\left(Y_{\mathrm{dl}}^n|\omega_{\mathrm{dl}}\right) \tag{59}$$

$$= h\left(Y_{\mathrm{ul}}^n, Y_{\mathrm{dl}}^n|\omega_{\mathrm{dl}}\right) - h\left(Y_{\mathrm{dl}}^n|Y_{\mathrm{ul}}^n, \omega_{\mathrm{dl}}\right) - \left(h\left(Y_{\mathrm{ul}}^n, Y_{\mathrm{dl}}^n|\omega_{\mathrm{dl}}\right) \right.$$
$$\left. - h\left(Y_{\mathrm{ul}}^n|Y_{\mathrm{dl}}^n, \omega_{\mathrm{dl}}\right)\right) \tag{60}$$

$$= h\left(Y_{\mathrm{ul}}^n|Y_{\mathrm{dl}}^n, \omega_{\mathrm{dl}}\right) - h\left(Y_{\mathrm{dl}}^n|Y_{\mathrm{ul}}^n, \omega_{\mathrm{dl}}\right) \tag{61}$$

$$= \sum_{i=1}^n h\left(Y_{\mathrm{ul},i}|Y_{\mathrm{dl}}^n, Y_{\mathrm{ul}}^{i-1}, \omega_{\mathrm{dl}}\right)$$
$$- \sum_{i=1}^n h\left(Y_{\mathrm{dl},i}|Y_{\mathrm{ul}}^n, Y_{\mathrm{dl}}^{i-1}, \omega_{\mathrm{dl}}\right) \tag{62}$$

$$\leq \sum_{i=1}^n h\left(Y_{\mathrm{ul},i}|Y_{\mathrm{dl}}^n, Y_{\mathrm{ul}}^{i-1}, \omega_{\mathrm{dl}}\right)$$
$$- \sum_{i=1}^n h\left(Y_{\mathrm{dl},i}|X_{\mathrm{ul},i}, X_{\mathrm{dl},i}, Y_{\mathrm{ul}}^n, Y_{\mathrm{dl}}^{i-1}, \omega_{\mathrm{dl}}\right) \tag{63}$$

$$= \sum_{i=1}^n h\left(Y_{\mathrm{ul},i}|X_{\mathrm{dl},i}, V_{\mathrm{ul},i}, Y_{\mathrm{dl}}^n, Y_{\mathrm{ul}}^{i-1}, \omega_{\mathrm{dl}}\right) - \sum_{i=1}^n h\left(Z_{\mathrm{dl},i}\right) \tag{64}$$

$$\leq \sum_{i=1}^n h\left(Y_{\mathrm{ul},i}|V_{\mathrm{ul},i}\right) - \sum_{i=1}^n h\left(Z_{\mathrm{dl},i}\right), \tag{65}$$

where (63) follows because conditioning reduces entropy; (64) follows since $X_{\mathrm{dl},i}$ is a function of $(\omega_{\mathrm{dl}}, Y_{\mathrm{ul}}^{i-1})$ and the genie $V_{\mathrm{ul},i}$ can be determined by $X_{\mathrm{dl},i}$ and $Y_{\mathrm{dl},i}$ as $Y_{\mathrm{dl}} = \sqrt{\rho_{\mathrm{dl}}} H_{\mathrm{dl}} X_{\mathrm{dl}} + V_{\mathrm{ul}}$. Also conditioned on $(X_{\mathrm{ul},i}, X_{\mathrm{dl},i})$, $Y_{\mathrm{dl},i}$ is independent of everything else; (65) follows since removing condition does not reduce entropy.

Thus U can be upper bounded as

$$U \leq \sum_{i=1}^n h\left(Y_{\mathrm{ul},i}|V_{\mathrm{ul},i}\right) - \sum_{i=1}^n \left(h\left(Z_{\mathrm{dl},i}\right) + h\left(Z_{\mathrm{ul},i}\right)\right). \tag{66}$$

Combining the results above and applying the chain rule, we have

$$R_{\mathrm{dl}} + R_{\mathrm{ul}} - \epsilon_n \leq \frac{1}{n} \sum_{i=1}^n \left(h\left(Y_{\mathrm{dl},i}\right) + h\left(Y_{\mathrm{S},i}\right) \right.$$
$$\left. + h\left(Y_{\mathrm{ul},i}|V_{\mathrm{ul},i}\right) - \left[h\left(Z_{\mathrm{dl},i}\right) + h\left(Z_{\mathrm{ul},i}\right) + h\left(Z_{\mathrm{S},i}\right)\right]\right).$$

Now by applying the standard time sharing argument, we can obtain

$$R_{\mathrm{dl}} + R_{\mathrm{ul}} \leq \underbrace{h(Y_{\mathrm{dl}}) - h(Z_{\mathrm{dl}})}_{R_{us,1}} + \underbrace{h(Y_{\mathrm{ul}}|V_{\mathrm{ul}}) - h(Z_{\mathrm{ul}})}_{R_{us,2}}$$
$$+ \underbrace{h(Y_{\mathrm{S}}) - h(Z_{\mathrm{S}})}_{R_{us,3}}. \tag{67}$$

We denote the covariance matrix of Y_{dl} as $K_{Y_{\mathrm{dl}}} = \mathbb{E}\left(Y_{\mathrm{dl}} Y_{\mathrm{dl}}^\dagger\right)$ that is maximized by Gaussian input in the presence of Gaussian noise. It can be easily shown that

$$K_{Y_{\mathrm{dl}}} = I_{N_{\mathrm{dl}}} + \rho_{\mathrm{dl}} H_{\mathrm{dl}} Q_{\mathrm{dl}} H_{\mathrm{dl}}^\dagger + \bar{\lambda} \rho_{\mathrm{I}} H_{\mathrm{I}} Q_{\mathrm{ul}} H_{\mathrm{I}}^\dagger$$
$$+ \sqrt{\bar{\lambda} \rho_{\mathrm{dl}} \rho_{\mathrm{I}}} H_{\mathrm{dl}} Q_{\mathrm{d,u}} H_{\mathrm{I}}^\dagger + \sqrt{\bar{\lambda} \rho_{\mathrm{dl}} \rho_{\mathrm{I}}} H_{\mathrm{I}} Q_{\mathrm{u,d}} H_{\mathrm{dl}}^\dagger, \tag{68}$$

where $Q_{\mathrm{dl}} = \mathbb{E}(X_{\mathrm{dl}} X_{\mathrm{dl}}^\dagger), Q_{\mathrm{ul}} = \mathbb{E}(X_{\mathrm{ul}} X_{\mathrm{ul}}^\dagger), Q_{\mathrm{d,u}} = \mathbb{E}(X_{\mathrm{dl}} X_{\mathrm{ul}}^\dagger), Q_{\mathrm{u,d}} = \mathbb{E}(X_{\mathrm{ul}} X_{\mathrm{dl}}^\dagger).$

Let $J = \begin{bmatrix} V_{\mathrm{ul}} \\ Y_{\mathrm{ul}} \end{bmatrix}$, the covariance matrix of J denoted by K_J can be maximized with Gaussian inputs, it can be verified that

$$K_J = \mathbb{E}(JJ^\dagger) = \begin{bmatrix} I_{N_{\mathrm{dl}}} + \bar{\lambda} \rho_{\mathrm{I}} H_{\mathrm{I}} Q_{\mathrm{ul}} H_{\mathrm{I}}^\dagger & \bar{\lambda} \sqrt{\rho_{\mathrm{ul}} \rho_{\mathrm{I}}} H_{\mathrm{I}} Q_{\mathrm{ul}} H_{\mathrm{ul}}^\dagger \\ \bar{\lambda} \sqrt{\rho_{\mathrm{ul}} \rho_{\mathrm{I}}} H_{\mathrm{ul}} Q_{\mathrm{ul}} H_{\mathrm{I}}^\dagger & I_{N_{\mathrm{ul}}} + \bar{\lambda} \rho_{\mathrm{ul}} H_{\mathrm{ul}} Q_{\mathrm{ul}} H_{\mathrm{ul}}^\dagger \end{bmatrix}. \tag{69}$$

Likewise, the covariance matrix of Y_{S} will be maximized by Gaussian input and computed as

$$K_{Y_{\mathrm{S}}} = \mathbb{E}\left(Y_{\mathrm{S}} Y_{\mathrm{S}}^\dagger\right) = W I_{N_{\mathrm{dl}}} + \lambda \rho_{\mathrm{S}} H_{\mathrm{S}} Q_{\mathrm{S}} H_{\mathrm{S}}^\dagger, \tag{70}$$

where $Q_{\mathrm{S}} = \mathbb{E}(X_{\mathrm{S}} X_{\mathrm{S}}^\dagger)$.

Using the result in (68), we can upper bound the first term $R_{us,1}$ (bit/s) in (67) as

$$\frac{R_{us,1}}{W_m} \leq \log\left|I_{N_{\mathrm{dl}}} + \rho_{\mathrm{dl}} H_{\mathrm{dl}} Q_{\mathrm{dl}} H_{\mathrm{dl}}^\dagger \right.$$
$$+ \bar{\lambda} \rho_{\mathrm{I}} H_{\mathrm{I}} Q_{\mathrm{ul}} H_{\mathrm{I}}^\dagger + \sqrt{\bar{\lambda} \rho_{\mathrm{dl}} \rho_{\mathrm{I}}} H_{\mathrm{dl}} Q_{\mathrm{d,u}} H_{\mathrm{I}}^\dagger$$
$$\left. + \sqrt{\bar{\lambda} \rho_{\mathrm{dl}} \rho_{\mathrm{I}}} H_{\mathrm{I}} Q_{\mathrm{u,d}} H_{\mathrm{dl}}^\dagger\right| \tag{71}$$

$$\leq \log\left|I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}} + G_{\mathrm{ul}}\right| \tag{72}$$

$$= \log\left|(I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})(I_{N_{\mathrm{dl}}} + (I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})^{-1} G_{\mathrm{ul}})\right| \tag{73}$$

$$= \log\left|I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}}\right| + \log\left|I_{N_{\mathrm{dl}}} + (I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})^{-1} G_{\mathrm{ul}}\right| \tag{74}$$

$$\leq \log\left|I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}}\right| + \log\left|2I_{N_{\mathrm{dl}}}\right| \tag{75}$$

$$= \log\left|I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}}\right| + N_{\mathrm{dl}}, \tag{76}$$

where $G_{\mathrm{dl}} = \rho_{\mathrm{dl}} H_{\mathrm{dl}} H_{\mathrm{dl}}^\dagger + \bar{\lambda} \rho_{\mathrm{I}} H_{\mathrm{I}} H_{\mathrm{I}}^\dagger, G_{\mathrm{ul}} = \sqrt{\bar{\lambda} \rho_{\mathrm{dl}} \rho_{\mathrm{I}}} H_{\mathrm{dl}} Q_{\mathrm{d,u}} H_{\mathrm{I}}^\dagger + \sqrt{\bar{\lambda} \rho_{\mathrm{dl}} \rho_{\mathrm{I}}} H_{\mathrm{I}} Q_{\mathrm{u,d}} H_{\mathrm{dl}}^\dagger$; (72) follows because $\mathrm{trace}(Q_i) \leq 1, i \in \{\mathrm{ul}, \mathrm{dl}\}$; thus, $Q_i \preceq I$, and $\log|.|$ is an increasing function on the cone of positive-definite matrices; (75) follows from the following lemma.

Lemma 5 *For p.s.d. matrices G_{dl} and G_{ul}, we have*

$$\log\left|I_{N_{\mathrm{dl}}} + \left(I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}}\right)^{-1} G_{\mathrm{ul}}\right| \leq \log\left|2I_{N_{\mathrm{dl}}}\right|. \tag{77}$$

Proof First, we show that $G_{\mathrm{ul}} \preceq G_{\mathrm{dl}}$. Let $A = \sqrt{\rho_{\mathrm{dl}}} H_{\mathrm{dl}} Q_{\mathrm{d,u}} - \sqrt{\bar{\lambda} \rho_{\mathrm{I}}} H_{\mathrm{I}}$, the product of matrices $AA^\dagger$

is always p.s.d., because for any vector x, $x^\dagger A A^\dagger x = (A^\dagger x)^\dagger (A^\dagger x) \geq 0$. Hence, we have the following

$$\rho_{\mathrm{dl}} H_{\mathrm{dl}} Q_{\mathrm{d,u}} Q_{\mathrm{d,u}}^\dagger H_{\mathrm{dl}}^\dagger + \bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}} H_{\mathrm{I}}^\dagger \geq \sqrt{\bar{\lambda}\rho_{\mathrm{dl}}\rho_{\mathrm{I}}} H_{\mathrm{dl}} Q_{\mathrm{d,u}} H_{\mathrm{I}}^\dagger$$
$$+ \sqrt{\bar{\lambda}\rho_{\mathrm{dl}}\rho_{\mathrm{I}}} H_{\mathrm{I}} Q_{\mathrm{d,u}}^\dagger H_{\mathrm{dl}}^\dagger. \quad (78)$$

Since $Q_{\mathrm{d,u}} Q_{\mathrm{d,u}}^\dagger \preceq I$, and $Q_{\mathrm{d,u}}^\dagger = Q_{\mathrm{u,d}}$, now we can obtain that

$$\rho_{\mathrm{dl}} H_{\mathrm{dl}} H_{\mathrm{dl}}^\dagger + \bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}} H_{\mathrm{I}}^\dagger \geq \sqrt{\bar{\lambda}\rho_{\mathrm{dl}}\rho_{\mathrm{I}}} H_{\mathrm{dl}} Q_{\mathrm{d,u}} H_{\mathrm{I}}^\dagger \\ + \sqrt{\bar{\lambda}\rho_{\mathrm{dl}}\rho_{\mathrm{I}}} H_{\mathrm{I}} Q_{\mathrm{u,d}} H_{\mathrm{dl}}^\dagger. \quad (79)$$

Hence, we have verified that $G_{\mathrm{ul}} \preceq G_{\mathrm{dl}}$, which also leads to the fact that $G_{\mathrm{ul}} \preceq I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}}$. As a result, for any given vector x, we have

$$x^\dagger \left[(I + G_{\mathrm{dl}})^{\frac{1}{2}} \left(I - (I + G_{\mathrm{dl}})^{-\frac{1}{2}} G_{\mathrm{ul}} (I + G_{\mathrm{dl}})^{-\frac{1}{2}} \right) (I + G_{\mathrm{dl}})^{\frac{1}{2}} \right] x$$
$$\geq 0, \text{ or}$$
$$\left((I + G_{\mathrm{dl}})^{\frac{1}{2}} x \right)^\dagger \left[I - (I + G_{\mathrm{dl}})^{-\frac{1}{2}} G_{\mathrm{ul}} (I + G_{\mathrm{dl}})^{-\frac{1}{2}} \right] \left((I + G_{\mathrm{dl}})^{\frac{1}{2}} x \right)$$
$$\geq 0. \quad (80)$$

From the definition of partial order of p.s.d. matrices [23], we have $(I + G_{\mathrm{dl}})^{-\frac{1}{2}} G_{\mathrm{ul}} (I + G_{\mathrm{dl}})^{-\frac{1}{2}} \preceq I$. Hence, we have

$$\log \left| I_{N_{\mathrm{dl}}} + (I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})^{-1} G_{\mathrm{ul}} \right| \quad (81)$$
$$= \log \left| (I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})^{-\frac{1}{2}} \left(I_{N_{\mathrm{dl}}} + (I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})^{-\frac{1}{2}} \right. \right.$$
$$\left. \left. \times G_{\mathrm{ul}} (I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})^{-\frac{1}{2}} \right) (I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})^{\frac{1}{2}} \right|$$
$$= \log \left| I_{N_{\mathrm{dl}}} + (I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})^{-\frac{1}{2}} G_{\mathrm{ul}} (I_{N_{\mathrm{dl}}} + G_{\mathrm{dl}})^{-\frac{1}{2}} \right|$$
$$\leq \log \left| 2 I_{N_{\mathrm{dl}}} \right|.$$

$\square$

The second term in (67) can be written as

$$R_{us,2} = h(Y_{\mathrm{ul}}, V_{\mathrm{ul}}) - h(V_{\mathrm{ul}}) - h(Z_{\mathrm{ul}}). \quad (82)$$

Using the covariance matrix, we derived in (69), and invoking Lemma 8 in [22], we can upper bound (82) as follows:

$$\frac{R_{us,2}}{W_m} \leq \log \left| I_{N_{\mathrm{ul}}} + \bar{\lambda}\rho_{\mathrm{ul}} H_{\mathrm{ul}} \left(I_{M_{\mathrm{ul}}} + \bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}}^\dagger H_{\mathrm{I}} \right)^{-1} H_{\mathrm{ul}}^\dagger \right|. \quad (83)$$

Finally, we can upper bound the third term in (67) with the covariance matrix in (70),

$$\frac{R_{us,3}}{W_s} \leq \log \left| W I_{N_{\mathrm{dl}}} + \lambda \rho_{\mathrm{S}} H_{\mathrm{S}} H_{\mathrm{S}}^\dagger \right| - \log \left| W I_{N_{\mathrm{dl}}} \right| \quad (84)$$
$$= \log \left| I_{N_{\mathrm{dl}}} + \frac{\lambda \rho_{\mathrm{S}}}{W} H_{\mathrm{S}} H_{\mathrm{S}}^\dagger \right|. \quad (85)$$

Combining all the results we derived above, we can prove Lemma 1.

Appendix 2
Rate calculation in Lemma 2
For the Gaussian inputs with the covariance matrices given in (9), the achievable rate in (12) can be calculated as

$$I(X_{\mathrm{dl}}; Y_{\mathrm{dl}} | S_{\mathrm{ul}})$$
$$= W_m \left(\log \left| I_{N_{\mathrm{dl}}} + \frac{\rho_{\mathrm{dl}}}{M_{\mathrm{dl}}} H_{\mathrm{dl}} H_{\mathrm{dl}}^\dagger + \bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}} K_u H_{\mathrm{I}}^\dagger \right| \right.$$
$$\left. - \log \left| I_{N_{\mathrm{dl}}} + \bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}} K_u H_{\mathrm{I}}^\dagger \right| \right) \quad (86)$$
$$\geq W_m \left(\log \left| I_{N_{\mathrm{dl}}} + \frac{\rho_{\mathrm{dl}}}{M_{\mathrm{dl}}} H_{\mathrm{dl}} H_{\mathrm{dl}}^\dagger + \bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}} K_u H_{\mathrm{I}}^\dagger \right| - \hat{m}_{\mathrm{I}} \right) \quad (87)$$
$$\geq W_m \left(\log \left| I_{N_{\mathrm{dl}}} + \rho_{\mathrm{dl}} H_{\mathrm{dl}} H_{\mathrm{dl}}^\dagger \right| - m_{\mathrm{dl}} \log M_{\mathrm{dl}} - \hat{m}_{\mathrm{I}} \right), \quad (88)$$

where $K_u = \frac{1}{M_{\mathrm{ul}}} (I_{M_{\mathrm{ul}}} + \bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}}^\dagger H_{\mathrm{I}})^{-1}, \hat{m}_{\mathrm{I}} = m_{\mathrm{I}} \log \left(1 + \frac{1}{M_{\mathrm{ul}}} \right); m_{\mathrm{I}} = \min\{M_{\mathrm{ul}}, N_{\mathrm{dl}}\}, m_{\mathrm{dl}} = \min\{M_{\mathrm{dl}}, N_{\mathrm{dl}}\}$, which are the rank of H_{I} and H_{dl}, respectively. Step (87) is established because of the following argument, applying the singular value decomposition to H_{I} such that $H_{\mathrm{I}} = U \Lambda V^\dagger$, where U and V are unitary matrices, Λ is $N_{\mathrm{dl}} \times M_{\mathrm{ul}}$ diagonal matrix containing singular values of H_{I}. Now, we can rewrite $\bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}} K_u H_{\mathrm{I}}^\dagger$ as

$$\bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}} K_u H_{\mathrm{I}}^\dagger = \frac{\bar{\lambda}\rho_{\mathrm{I}}}{M_{\mathrm{ul}}} U \Lambda \left(I_{M_{\mathrm{ul}}} + \bar{\lambda}\rho_{\mathrm{I}} \Lambda^\dagger \Lambda \right)^{-1} \Lambda^\dagger U^\dagger. \quad (89)$$

Since $\bar{\lambda}\rho_{\mathrm{I}} \Lambda (I_{M_{\mathrm{ul}}} + \bar{\lambda}\rho_{\mathrm{I}} \Lambda^\dagger \Lambda)^{-1} \Lambda^\dagger \preceq I_{N_{\mathrm{dl}}}$, for p.s.d. matrices, given a vector x, we can show that

$$x^\dagger \left(\bar{\lambda}\rho_{\mathrm{I}} H_{\mathrm{I}} K_u H_{\mathrm{I}}^\dagger \right) x$$
$$= \frac{1}{M_{\mathrm{ul}}} \left(U^\dagger x \right)^\dagger \bar{\lambda}\rho_{\mathrm{I}} \Lambda \left(I_{M_{\mathrm{ul}}} + \bar{\lambda}\rho_{\mathrm{I}} \Lambda^\dagger \Lambda \right)^{-1} \Lambda^\dagger (U^\dagger x) \quad (90)$$
$$\leq \frac{1}{M_{\mathrm{ul}}} \left(U^\dagger x \right)^\dagger I_{N_{\mathrm{dl}}} \left(U^\dagger x \right). \quad (91)$$

Thus, $\bar{\lambda}\rho_I H_I K_u H_I^\dagger \le \frac{1}{M_{ul}} I_{N_{dl}}$, which implies that

$$
\log\left| I_{N_{dl}} + \bar{\lambda}\rho_I H_I K_u H_I^\dagger \right| \le \min\{M_{ul}, N_{dl}\}
$$
$$
\log\left(1 + \frac{1}{M_{ul}} \right) \triangleq \hat{m}_I. \tag{92}
$$

Next, we compute $I(S_{ul}, U_{ul}; Y_{ul})$ as follows:

$$
I(S_{ul}, U_{ul}; Y_{ul}) = W_m \log\left| I_{N_{ul}} + \frac{\bar{\lambda}\rho_{ul}}{M_{ul}} H_{ul} H_{ul}^\dagger \right| \tag{93}
$$
$$
\ge W_m \Big(\log\left| I_{N_{ul}} + \bar{\lambda}\rho_{ul} H_{ul} H_{ul}^\dagger \right|
$$
$$
- m_{ul}\log M_{ul} \Big), \tag{94}
$$

where $m_{ul} = \min\{M_{ul}, N_{ul}\}$, which is the rank of H_{ul}. Similarly,

$$
I(X_S; Y_S) = W_s \log\left| I_{N_{dl}} + \frac{\lambda\rho_S}{M_{ul}W} H_S H_S^\dagger \right| \tag{95}
$$
$$
\ge W_m \Big(W\log\left| I_{N_{dl}} + \frac{\lambda\rho_S}{W} H_S H_S^\dagger \right|
$$
$$
- m_I W\log M_{ul} \Big). \tag{96}
$$

And

$$
I(U_{ul}; Y_{ul}|S_{ul})
$$
$$
= W_m \log\left| I_{N_{ul}} + \bar{\lambda}\rho_{ul} H_{ul} K_u H_{ul}^\dagger \right| \tag{97}
$$
$$
\ge W_m \Big(\log\left| I_{N_{ul}} + \bar{\lambda}\rho_{ul} H_{ul} \left(I_{M_{ul}} + \bar{\lambda}\rho_I H_I^\dagger H_I \right)^{-1} H_{ul}^\dagger \right|
$$
$$
- m_{ul}\log M_{ul} \Big), \tag{98}
$$

$$
I(S_{ul}; Y_{dl}|X_{dl})
$$
$$
= W_m \Big(\log\left| I_{N_{dl}} + \frac{\bar{\lambda}\rho_I}{M_{ul}} H_I H_I^\dagger \right|
$$
$$
- \log\left| I_{N_{dl}} + \bar{\lambda}\rho_I H_I K_u H_I^\dagger \right| \Big) \tag{99}
$$
$$
\ge W_m \left(\log\left| I_{N_{dl}} + \frac{\bar{\lambda}\rho_I}{M_{ul}} H_I H_I^\dagger \right| - \hat{m}_I \right) \tag{100}
$$
$$
\ge W_m \left(\log\left| I_{N_{dl}} + \bar{\lambda}\rho_I H_I H_I^\dagger \right| - m_I\log M_{ul} - \hat{m}_I \right) \tag{101}
$$
$$
= W_m \left(\log\left| I_{N_{dl}} + \bar{\lambda}\rho_I H_I H_I^\dagger \right| - m_I\log(M_{ul}+1) \right) \tag{102}
$$

where (100) follows from step (92).

Now, we can calculate $I(U_{ul}; Y_{ul}|S_{ul}) + I(S_{ul}; Y_{dl}|X_{dl})$ as

$$
I(U_{ul}; Y_{ul}|S_{ul}) + I(S_{ul}; Y_{dl}|X_{dl}) \tag{103}
$$
$$
\ge W_m \Big(\log\left| I_{N_{ul}} + \bar{\lambda}\rho_{ul} H_{ul} \left(I_{M_{ul}} + \bar{\lambda}\rho_I H_I^\dagger H_I \right)^{-1} H_{ul}^\dagger \right|
$$
$$
+ \log\left| I_{N_{dl}} + \bar{\lambda}\rho_I H_I H_I^\dagger \right|
$$
$$
- m_{ul}\log M_{ul} - m_I\log(M_{ul}+1) \Big) \tag{104}
$$
$$
= W_m \Big(\log\left| I_{M_{ul}} + \bar{\lambda}\rho_{ul} H_{ul}^\dagger H_{ul}(I_{M_{ul}} + \bar{\lambda}\rho_I H_I^\dagger H_I)^{-1} \right|
$$
$$
+ \log\left| I_{M_{ul}} + \bar{\lambda}\rho_I H_I^\dagger H_I \right|
$$
$$
- m_{ul}\log M_{ul} - m_I\log(M_{ul}+1) \Big) \tag{105}
$$
$$
= W_m \Big(\log\left| I_{M_{ul}} + \bar{\lambda}\rho_{ul} H_{ul}^\dagger H_{ul} + \bar{\lambda}\rho_I H_I^\dagger H_I \right|
$$
$$
- m_{ul}\log M_{ul} + m_I\log(M_{ul}+1) \Big) \tag{106}
$$
$$
\ge W_m \Big(\log\left| I_{N_{ul}} + \bar{\lambda}\rho_{ul} H_{ul} H_{ul}^\dagger \right| - m_{ul}\log M_{ul}
$$
$$
- m_I\log(M_{ul}+1) \Big), \tag{107}
$$

where (105) and (107) follow from Sylvester's determinant theorem.

Finally, we compute $I(X_{dl}, S_{ul}; Y_{dl})$ as follows:

$$
I(X_{dl}, S_{ul}; Y_{dl})
$$
$$
= W_m \Big(\log\left| I_{N_{dl}} + \frac{\rho_{dl}}{M_{dl}} H_{dl} H_{dl}^\dagger + \frac{\bar{\lambda}\rho_I}{M_{ul}} H_I H_I^\dagger \right|
$$
$$
- \log\left| I_{N_{dl}} + \bar{\lambda}\rho_I H_I K_u H_I^\dagger \right| \Big) \tag{108}
$$
$$
\ge W_m \Big(\log\left| I_{N_{dl}} + \rho_{dl} H_{dl} H_{dl}^\dagger + \bar{\lambda}\rho_I H_I H_I^\dagger \right| - \hat{m}_I
$$
$$
- \min\{M_{dl} + M_{ul}, N_{dl}\}\log(\max\{M_{dl}, M_{ul}\}) \Big), \tag{109}
$$

where (109) holds because the rank of the matrix $I_{N_{dl}} + \frac{\rho_{dl}}{M_{dl}} H_{dl} H_{dl}^\dagger + \frac{\bar{\lambda}\rho_I}{M_{ul}} H_I H_I^\dagger$ is less than the rank of an enhanced multiple-access channel matrix by allowing full-cooperation between transmitters which is $\min\{M_{dl} + M_{ul}, N_{dl}\}$.

Combining all the expressions we derived above, we can obtain the capacity region inner bound as[8] $\mathcal{R}_{BC}(\mathcal{H}) = \left\{ (R_{dl}, R_{ul}) : R_{dl} \le \overline{C}_{dl} - c_1, R_{ul} \le \overline{C}_{ul} - c_2, R_{sum} \le \overline{C}_{sum} - (c_1 + c_2) \right\}$, where c_1 and c_2 are given in (8).

Appendix 3
Useful Lemmas
Random matrix theory plays a critical role in the analysis of MIMO wireless networks. Here, we will restate some

important properties of random matrices in the following lemmas which will be used for our derivation.

Lemma 6 *(Lemma 3 in [21]) For a PTP channel, where $H \in \mathbb{C}^{M \times N}$ with i.i.d, $\mathbb{CN}(0,1)$ entries and the channel level is α, the optimal DMT is equivalent to the minimum of the following optimization problem,*

$$d(r) = \min \sum_{i=1}^{\min(M,N)} (M + N + 1 - 2i)x_i$$

$$s.t \quad \sum_{i=1}^{\min(M,N)} (\alpha - x_i)^+ \leq r \tag{110}$$

$$0 \leq x_1 \leq \cdots \leq x_{\min(M,N)},$$

and the optimal solution is $d(r) = \alpha d_{M,N}(\frac{r}{\alpha})$, for $0 \leq r \leq \min(M,N)\alpha$, where $d_{M,N}(r) = (M - r)(N - r)$ is a piecewise linear curve joining the integer point $r \in [0, \min(M,N)]$.

Lemma 7 *(Theorem 4 in [18]) Let $H \in \mathbb{C}^{M \times N}$ have i.i.d, $\mathbb{CN}(0,1)$ entries. Suppose the nonzero-ordered eigenvalues of $R = HH^\dagger$ are denoted by $\beta_1 \geq \cdots \beta_q > 0$, where $q = \min(M,N)$. Let $\beta_i = \rho^{-\mu_i}, i \in [1,q]$, assuming that all the eigenvalues vary exponentially with SNR. Let $\bar{\mu} = \{\mu_1, \cdots, \mu_q\}$, thus the asymptotic distribution of $\bar{\mu}$ is*

$$p(\bar{\mu}) \doteq \begin{cases} \rho^{-\sum_{i=1}^{q}(M+N+1-2i)\mu_i} & \text{if } 0 \leq \mu_1 \leq \cdots \mu_q \\ 0 & \text{Otherwise,} \end{cases} \tag{111}$$

Lemma 8 *(Theorem 1 and 2 in [24]) Let $H_1 \in \mathbb{C}^{N_2 \times N_1}$ and $H_2 \in \mathbb{C}^{N_2 \times N_3}$ be two mutually independent random matrices with i.i.d, $\mathbb{CN}(0,1)$ entries. Suppose the ordered eigenvalues of $V_1 = H_1^\dagger \left(I_{N_2} + \rho^\alpha H_2 H_2^\dagger \right)^{-1} H_1, V_2 = H_2 H_2^\dagger$ are denoted by $\beta_1 \geq \cdots \beta_q > 0, \eta_1 \geq \cdots \eta_p > 0$ where $q = \min(N_1, N_2), p = \min(N_2, N_3)$. Let $\beta_i = \rho^{-\mu_i}, i \in [1,q]; \eta_k = \rho^{-\theta_k}, k \in [0,p]$, and $\bar{\mu} = \{\mu_1, \cdots, \mu_q\}, \bar{\theta} = \{\theta_1, \cdots, \theta_p\}$. Hence, the conditional distribution of $\bar{\mu}$ given $\bar{\theta}$ is*

$$p(\bar{\mu}|\bar{\theta}) \doteq \begin{cases} \rho^{-E_1(\bar{\mu}, \bar{\theta})} & \text{if } (\bar{\mu}, \bar{\theta}) \in \mathcal{D}_1 \\ 0 & \text{Otherwise,} \end{cases} \tag{112}$$

where

$$E_1(\bar{\mu}, \bar{\theta}) = \sum_{i=1}^{q} (N_1 + N_2 + 1 - 2i)\mu_i + \sum_{i=1}^{q} \sum_{k=1}^{\min(N_2-i, N_3)}$$

$$\times (\alpha - \mu_i - \theta_k)^+ - N_1 \sum_{k=1}^{p} (\alpha - \theta_k)^+$$

$$\mathcal{D}_1 = \Big\{ 0 \leq \mu_1 \leq \cdots \leq \mu_q; 0 \leq \theta_1 \leq \cdots \leq \theta_p; $$

$$\mu_i + \theta_k \geq \alpha, \forall (i+k) \geq N_2 + 1 \Big\}. \tag{113}$$

Appendix 4
Proof of Lemma 3
From Corollary 1, we can express the high SNR asymptotic sum-capacity as

$$C_{\text{sum}}(\mathcal{H}) \doteq \max_{0 \leq \lambda \leq 1} F(\mathcal{H}, \lambda, \bar{\lambda}) \tag{114}$$

where

$$\begin{aligned} F(\mathcal{H}, \lambda, \bar{\lambda}) = &W_m \Bigg(\log \Big| I_{N_{\text{dl}}} + \rho_{\text{dl}} H_{\text{dl}} H_{\text{dl}}^\dagger + \bar{\lambda}\rho_I H_I H_I^\dagger \Big| \\ &+ W \log \Big| I_{N_{\text{dl}}} + \frac{\lambda \rho_S}{W} H_S H_S^\dagger \Big| \\ &+ \log \Big| I_{N_{\text{ul}}} + \bar{\lambda}\rho_{\text{ul}} H_{\text{ul}}(I_{M_{\text{ul}}} + \bar{\lambda}\rho_I H_I^\dagger H_I)^{-1} H_{\text{ul}}^\dagger \Big| \Bigg), \\ = &W_m \Bigg(\log \Big| I_{M_{\text{dl}}} + \rho_{\text{dl}} H_{\text{dl}}^\dagger (I_{N_{\text{dl}}} + \bar{\lambda}\rho_I H_I H_I^\dagger)^{-1} H_{\text{dl}} \Big| \\ &+ \log \Big| I_{N_{\text{dl}}} + \bar{\lambda}\rho_I H_I H_I^\dagger \Big| \\ &+ \log \Big| I_{N_{\text{ul}}} + \bar{\lambda}\rho_{\text{ul}} H_{\text{ul}}(I_{M_{\text{ul}}} + \bar{\lambda}\rho_I H_I^\dagger H_I)^{-1} H_{\text{ul}}^\dagger \Big| \\ &+ W \log \Big| I_{N_{\text{dl}}} + \frac{\lambda \rho_S}{W} H_S H_S^\dagger \Big| \Bigg). \end{aligned}$$

The ordered eigenvalues of $G_1 = H_{\text{dl}}^\dagger \Big(I_{N_{\text{dl}}} + \bar{\lambda}\rho_I H_I H_I^\dagger \Big)^{-1} H_{\text{dl}}, G_2 = H_{\text{ul}} \Big(I_{M_{\text{ul}}} + \bar{\lambda}\rho_I H_I^\dagger H_I \Big)^{-1} H_{\text{ul}}^\dagger, G_3 = H_I H_I^\dagger$ and $G_4 = H_S H_S^\dagger$ are denoted by $\beta_1 \geq \cdots \beta_{m_{\text{dl}}} > 0, \gamma_1 \geq \cdots \gamma_{m_{\text{ul}}} > 0, \eta_1 \geq \cdots \eta_{m_1} > 0$ and $\xi_1 \geq \cdots \xi_{m_1} > 0$. Let $\beta_i = \rho^{-\mu_i}, i \in [1, m_{\text{dl}}]; \gamma_j = \rho^{-\sigma_j}, j \in [1, m_{\text{ul}}]; \eta_k = \rho^{-\theta_k}, k \in [0, m_1]; \xi_l = \rho^{-\nu_l}, l \in [0, m_1]$. When $\rho \rightarrow \infty$, we have

$$\begin{aligned} \rho^{-d_{\mathfrak{B}_{\text{sum}}}(r_{\text{sum}})} &\doteq \Pr \Big(C_{\text{sum}} < W_m r_{\text{sum}} \log \rho \Big) \\ &\doteq \Pr \Bigg(\max_{0 \leq \lambda \leq 1} \prod_{i=1}^{m_{\text{dl}}} (1 + \rho^{\alpha_{\text{dl}}} \beta_i) \prod_{j=1}^{m_{\text{ul}}} (1 + \bar{\lambda}\rho^{\alpha_{\text{ul}}} \gamma_j) \\ &\quad \times \prod_{k=1}^{m_1} (1 + \bar{\lambda}\rho^{\alpha_I} \eta_k) \Bigg(\prod_{l=1}^{m_1} \Big(1 + \frac{\lambda}{W}\rho^{\alpha_S} \xi_l \Big) \Bigg)^W < \rho^{r_{\text{sum}}} \Bigg) \\ &\doteq \Pr \Bigg(\max_{0 \leq \lambda \leq 1} (\bar{\lambda})^{m_{\text{ul}} + m_1} \Big(\frac{\lambda}{W} \Big)^{W m_1} \prod_{i=1}^{m_{\text{dl}}} \rho^{(\alpha_{\text{dl}} - \mu_i)^+} \\ &\quad \times \prod_{j=1}^{m_{\text{ul}}} \rho^{(\alpha_{\text{ul}} - \sigma_j)^+} \prod_{k=1}^{m_1} \rho^{(\alpha_I - \theta_k)^+} \prod_{l=1}^{m_1} \rho^{W(\alpha_S - \nu_l)^+} < \rho^{r_{\text{sum}}} \Bigg) \end{aligned} \tag{115}$$

where $m_{\text{dl}}, m_{\text{ul}}$ and m_1 are defined in (8).

For any channel realization $\mathcal{H}$ in a particular fade period, we have $F(\mathcal{H}, \lambda = \bar{\lambda} = 0.5) \leq C_{\text{sum}}(\mathcal{H}) < F(\mathcal{H}, \lambda = \bar{\lambda} = 1)$; hence, the sum-capacity outage event $\mathfrak{B}_{\text{sum}} \triangleq \{R_{\text{sum}} \notin C_{\text{sum}}(\mathcal{H})\}$ can be bounded as $\{R_{\text{sum}} \notin F(\mathcal{H}, \lambda = \bar{\lambda} = 1)\} \subset \mathfrak{B}_{\text{sum}} \subseteq \{R_{\text{sum}} \notin F(\mathcal{H}, \lambda = \bar{\lambda} = 0.5)\}$. Consequently, we have

$$\begin{aligned} &\Pr \Big(R_{\text{sum}} \notin F(\mathcal{H}, \lambda = \bar{\lambda} = 1) \Big) < \rho^{-d_{\mathfrak{B}_{\text{sum}}}(r_{\text{sum}})} \\ &\leq \Pr \Big(R_{\text{sum}} \notin F(\mathcal{H}, \lambda = \bar{\lambda} = 0.5) \Big). \end{aligned} \tag{116}$$

From (116), we can see that when $\rho \to \infty$, $\rho^{-d_{\mathfrak{B}_{\text{sum}}}(r_{\text{sum}})}$ converges to the following result as $\frac{\lambda}{W}$, $\bar{\lambda}$ do not grow at the same rate as ρ thus can be ignored on the scale of interest

$$\rho^{-d_{\mathfrak{B}_{\text{sum}}}(r_{\text{sum}})} \doteq \Pr\left(\sum_{i=1}^{m_{\text{dl}}}(\alpha_{\text{dl}} - \mu_i)^+ + \sum_{j=1}^{m_{\text{ul}}}(\alpha_{\text{ul}} - \sigma_j)^+ \right.$$
$$\left. + \sum_{k=1}^{m_{\text{I}}}(\alpha_{\text{I}} - \theta_k)^+ + W\sum_{l=1}^{m_{\text{I}}}(\alpha_{\text{S}} - \nu_l)^+ < r_{\text{sum}}\right).$$
$$(117)$$

Let $\bar{\mu} = \{\mu_1, \cdots, \mu_{m_{\text{dl}}}\}$, $\bar{\sigma} = \{\sigma_1, \cdots, \sigma_{m_{\text{ul}}}\}$, $\bar{\theta} = \{\theta_1, \cdots, \theta_{m_{\text{I}}}\}$, and $\bar{\nu} = \{\nu_1, \cdots, \nu_{m_{\text{I}}}\}$. The joint distribution of $\bar{\mu}, \bar{\sigma}, \bar{\theta}$ and $\bar{\nu}$ can be calculated as

$$p(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu}) = p(\bar{\mu}, \bar{\sigma}, \bar{\theta})p(\bar{\nu}) \qquad (118)$$
$$= p(\bar{\mu}\bar{\sigma}|\bar{\theta})p(\bar{\theta})p(\bar{\nu}) \qquad (119)$$
$$= p(\bar{\mu}|\bar{\theta})p(\bar{\sigma}|\bar{\theta})p(\bar{\theta})p(\bar{\nu}) \qquad (120)$$

where (118) follows from the fact that matrix G_4 is independent of other matrices; (120) follows from the random matrix theory that the dependence of G_1 and G_2 is only through the eigenvalues of G_3. Thus, given the eigenvalues of G_3, the eigenvalues of G_1 and G_2 are conditionally independent.

By invoking Lemmas 7 and 8, we have

$$p(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu}) \doteq \begin{cases} \rho^{-E(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu})} & \text{if } (\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu}) \in \mathcal{D} \\ 0 & \text{Otherwise,} \end{cases}$$

where

$$E(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu}) = \left\{\sum_{i=1}^{m_{\text{dl}}}(M_{\text{dl}} + N_{\text{dl}} + 1 - 2i)\mu_i \right.$$
$$+ \sum_{j=1}^{m_{\text{ul}}}(M_{\text{ul}} + N_{\text{ul}} + 1 - 2j)\sigma_j - (M_{\text{dl}} + N_{\text{ul}})m_{\text{I}}\alpha_{\text{I}}$$
$$+ \sum_{k=1}^{m_{\text{I}}}(M_{\text{dl}} + N_{\text{ul}} + M_{\text{ul}} + N_{\text{dl}} + 1 - 2k)\theta_k$$
$$+ \sum_{l=1}^{m_{\text{I}}}(M_{\text{ul}} + N_{\text{dl}} + 1 - 2l)\nu_l \sum_{i=1}^{m_{\text{dl}}}\sum_{k=1}^{\min\{N_{\text{dl}} - i, M_{\text{ul}}\}}$$
$$\times (\alpha_{\text{I}} - \mu_i - \theta_k)^+ + \sum_{j=1}^{m_{\text{ul}}}\sum_{k=1}^{\min\{M_{\text{ul}} - j, N_{\text{dl}}\}}$$
$$\left. \times (\alpha_{\text{I}} - \sigma_j - \theta_k)^+ \right\},$$
$$(121)$$

$$\mathcal{D} = \left\{\sum_{i=1}^{m_{\text{dl}}}(\alpha_{\text{dl}} - \mu_i)^+ + \sum_{j=1}^{m_{\text{ul}}}(\alpha_{\text{ul}} - \sigma_j)^+ + \sum_{k=1}^{m_{\text{I}}}(\alpha_{\text{I}} - \theta_k)^+ \right.$$
$$+ W\sum_{l=1}^{m_{\text{I}}}(\alpha_{\text{S}} - \nu_l)^+ < r_{\text{sum}};$$
$$0 \le \mu_1 \le \cdots \le \mu_{m_{\text{dl}}}; \ 0 \le \sigma_1 \le \cdots \le \sigma_{m_{\text{ul}}};$$
$$0 \le \theta_1 \le \cdots \le \theta_{m_{\text{I}}}; \ 0 \le \nu_1 \le \cdots \le \nu_{m_{\text{I}}};$$
$$\mu_i + \theta_k \ge \alpha_{\text{I}}, \forall (i + k) \ge N_{\text{dl}} + 1; \ \sigma_j + \theta_k \ge \alpha_{\text{I}},$$
$$\left. \forall (j + k) \ge M_{\text{ul}} + 1 \right\}.$$
$$(122)$$

With the joint distribution of $p(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu})$, we have obtained above, the outage probability is as follows:

$$\Pr(\mathfrak{B}_{\text{sum}}) \doteq \int_{\mathcal{D}} p(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu}) \doteq \int_{\mathcal{D}} \rho^{-E(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu})}.$$
$$(123)$$

Using Laplace's principle, step (117) can be calculated by minimizing the SNR exponent $E(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu})$ which has the dominant probability. Thus, we have

$$d_{\mathfrak{B}_{\text{sum}}} = \min_{(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu}) \in \mathcal{D}} E(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{\nu}), \qquad (124)$$

which proves Lemma 3.

Appendix 5
Proof of Lemma 4
We first express the asymptotic achievable sum-rate (with $\lambda = \bar{\lambda} = 0.5$) as follows

$$I_{\text{sum}} \doteq W_m\left(\log\left|I_{M_{\text{dl}}} + \rho_{\text{dl}}H_{\text{dl}}^{\dagger}(I_{N_{\text{dl}}} + \bar{\lambda}\rho_{\text{I}}H_{\text{I}}H_{\text{I}}^{\dagger})^{-1}H_{\text{dl}}\right| \right.$$
$$\left. + \log\left|I_{N_{\text{dl}}} + \bar{\lambda}\rho_{\text{I}}H_{\text{I}}H_{\text{I}}^{\dagger}\right| + W\log\left|I_{N_{\text{dl}}} + \frac{\lambda\rho_{\text{S}}}{W}H_{\text{S}}H_{\text{S}}^{\dagger}\right|\right).$$
$$(125)$$

We still use the same notations defined in Appendix 4 to represent the ordered eigenvalue of $G_1 = H_{\text{dl}}^{\dagger}\left(I_{N_{\text{dl}}} + \bar{\lambda}\rho_{\text{I}}H_{\text{I}}H_{\text{I}}^{\dagger}\right)^{-1}H_{\text{dl}}$, $G_3 = H_{\text{I}}H_{\text{I}}^{\dagger}$ and $G_4 = H_{\text{S}}H_{\text{S}}^{\dagger}$. Thus, we obtain that

$$\rho^{-d_{\mathcal{D}_{\text{sum}}}(r_{\text{sum}})} \doteq \Pr\left(\sum_{i=1}^{m_{\text{dl}}}(\alpha_{\text{dl}} - \mu_i)^+ + \sum_{k=1}^{m_{\text{I}}}(\alpha_{\text{I}} - \theta_k)^+ \right.$$
$$\left. + W\sum_{l=1}^{m_{\text{I}}}(\alpha_{\text{S}} - \nu_l)^+ < r_{\text{sum}}\right).$$

The joint distribution of $(\bar{\mu}, \bar{\theta}, \bar{\nu})$ can be derived by following the same steps in Appendix 4. Likewise, Lemma 4 can be proved and we omit the steps to avoid redundancy.

Appendix 6
DMT calculation of (M, N_{dl}, M_{ul}, M) with and without CSIT

Lemma 9 *For the (M, N_{dl}, M_{ul}, M) side-channel-assisted full-duplex network with $\alpha_{dl} = \alpha_{ul} = \alpha_I = 1$ and with CSIT, the optimal DMT at multiplexing gain pair (r_{dl}, r_{ul}) is*

$$d_{(M,N_{dl},M_{ul},M)}^{CSIT,opt}(r_{dl}, r_{ul}) = \min\left\{ d_{M,N_{dl}}(r_{dl}), d_{M_{ul},M}(r_{ul}), d_{\text{sum}(M,N_{dl},M_{ul},M)}^{CSIT}(r_{\text{sum}}) \right\}. \tag{126}$$

where $d_{\text{sum}(M,N_{dl},M_{ul},M)}^{CSIT}(r_{\text{sum}})$ is given as follows:

- *when $W \leq \frac{|M_{ul}-N_{dl}|+1}{2M+M_{ul}+N_{dl}-1}$,*

$$d_{\text{sum}(M,N_{dl},M_{ul},M)}^{CSIT}(r_{\text{sum}}) = \begin{cases} \alpha_S d_{M_{ul},N_{dl}}\left(\frac{r_{\text{sum}}}{W\alpha_S}\right) + M_{ul}N_{dl} + M(M_{ul}+N_{dl}), \; r_{\text{sum}} \leq m_I W\alpha_S \\ d_{m_I,2M+m_X}(r_{\text{sum}} - m_I W\alpha_S) + M|M_{ul}-N_{dl}|, \; m_I W\alpha_S \leq r_{\text{sum}} \leq m_I(1+W\alpha_S) \\ d_{|M_{ul}-N_{dl}|,M}(r_{\text{sum}} - m_I(1+W\alpha_S)), \; m_I(1+W\alpha_S) \leq r_{\text{sum}} \leq m_X + m_I W\alpha_S \end{cases}$$

- *when $W \in \left[\frac{M_{ul}+N_{dl}-1}{2M+|M_{ul}-N_{dl}|+1}, \frac{|M_{ul}-N_{dl}|+1}{M+|M_{ul}-N_{dl}|-1} \right]$,*

$$d_{\text{sum}(M,N_{dl},M_{ul},M)}^{CSIT}(r_{\text{sum}}) = \begin{cases} d_{m_I,2M+m_X}(r_{\text{sum}}) + M_{ul}N_{dl}\alpha_S + M|M_{ul}-N_{dl}|, \; r_{\text{sum}} \leq m_I \\ \alpha_S d_{M_{ul},N_{dl}}\left(\frac{r_{\text{sum}}-m_I}{W\alpha_S}\right) + M|M_{ul}-N_{dl}|, \; m_I \leq r_{\text{sum}} \leq m_I(1+W\alpha_S) \\ d_{|M_{ul}-N_{dl}|,M}(r_{\text{sum}} - m_I(1+W\alpha_S)), \; m_I(1+W\alpha_S) \leq r_{\text{sum}} \leq m_X + m_I W\alpha_S \end{cases}$$

- *when $W \geq \frac{M_{ul}+N_{dl}-1}{M-|M_{ul}-N_{dl}|+1}$,*

$$d_{\text{sum}(M,1,1,M)}^{CSIT}(r_{\text{sum}}) = \begin{cases} d_{m_I,2M+m_X}(r_{\text{sum}}) + M_{ul}N_{dl}\alpha_S + M|M_{ul}-N_{dl}|, \; r_{\text{sum}} \leq m_I \\ d_{|M_{ul}-N_{dl}|,M}(r_{\text{sum}} - m_I) + M_{ul}N_{dl}\alpha_S, \; m_I \leq r_{\text{sum}} \leq m_X \\ \alpha_S d_{M_{ul},N_{dl}}\left(\frac{r_{\text{sum}}-m_X}{W\alpha_S}\right), \; m_X \leq r_{\text{sum}} \leq m_X + m_I W\alpha_S \end{cases}$$

where $m_I = \min\{M_{ul}, N_{dl}\}, m_X = \max\{M_{ul}, N_{dl}\}$.

Proof The details of the proof can be found in Appendix 7. □

The achievable DMT of (M, N_{dl}, M_{ul}, M) without CSIT is given in the following lemma.

Lemma 10 *Consider the case in Lemma 9 under no-CSIT assumption, the achievable DMT at multiplexing gain pair (r_{dl}, r_{ul}) is*

$$d_{(M,N_{dl},M_{ul},M)}^{No\text{-}CSIT}(r_{dl}, r_{ul}) = \min\left\{ d_{M,N_{dl}}(r_{dl}), d_{M_{ul},M}(r_{ul}), d_{\text{sum}(M,N_{dl},M_{ul},M)}^{No\text{-}CSIT}(r_{\text{sum}}) \right\}.$$

where $d_{\text{sum}(M,N_{dl},M_{ul},M)}^{No\text{-}CSIT}(r_{\text{sum}})$ is given as follows: if $M_{ul} \geq 2(N_{dl}-1)$,

- *when $W \leq \frac{M_{ul}-N_{dl}+1}{M+M_{ul}+N_{dl}-1}$,*

$$d_{\text{sum}(M,N_{dl},M_{ul},M)}^{CSIT}(r_{\text{sum}}) = \begin{cases} \alpha_S d_{M_{ul},N_{dl}}\left(\frac{r_{\text{sum}}}{W\alpha_S}\right) + N_{dl}(M_{ul}+M), \; r_{\text{sum}} \leq N_{dl} W\alpha_S \\ d_{N_{dl},M+M_{ul}}(r_{\text{sum}} - N_{dl} W\alpha_S), \; N_{dl} W\alpha_S \leq r_{\text{sum}} \leq N_{dl}(1+W\alpha_S) \end{cases}$$

- *when $W \geq \frac{M_{ul}+N_{dl}-1}{M+M_{ul}-N_{dl}+1}$,*

$$d_{\text{sum}(M,N_{dl},M_{ul},M)}^{No\text{-}CSIT}(r_{\text{sum}}) = \begin{cases} d_{N_{dl},M+M_{ul}}(r_{\text{sum}}) + M_{ul}N_{dl}\alpha_S, \; r_{\text{sum}} \leq N_{dl} \\ \alpha_S d_{M_{ul},N_{dl}}\left(\frac{r_{\text{sum}}-N_{dl}}{W\alpha_S}\right), \; N_{dl} \leq r_{\text{sum}} \leq N_{dl}(1+W\alpha_S) \end{cases}$$

And if $N_{dl} \geq M_{ul}$, and $M_{ul} \leq 2$:

- *when $W \leq \frac{N_{dl}-M_{ul}+1}{M+M_{ul}+N_{dl}-1}$,*

$$d_{\text{sum}(M,N_{dl},M_{ul},M)}^{No\text{-}CSIT}(r_{\text{sum}}) = \begin{cases} \alpha_S d_{M_{ul},N_{dl}}\left(\frac{r_{\text{sum}}}{W\alpha_S}\right) + N_{dl}(M_{ul}+M), \; r_{\text{sum}} \leq M_{ul} W\alpha_S \\ d_{M_{ul},M+N_{dl}}(r_{\text{sum}} - M_{ul} W\alpha_S) + M(N_{dl}-M_{ul}), \; M_{ul} W\alpha_S \leq r_{\text{sum}} \leq M_{ul}(1+W\alpha_S) \\ d_{N_{dl}-M_{ul},M}(r_{\text{sum}} - M_{ul}(1+W\alpha_S)), \; M_{ul}(1+W\alpha_S) \leq r_{\text{sum}} \leq N_{dl} + M_{ul} W\alpha_S \end{cases}$$

- when $W \in \left[\frac{M_{ul}+N_{dl}-1}{M+N_{dl}-M_{ul}+1}, \frac{N_{dl}-M_{ul}+1}{M+N_{dl}-M_{ul}-1} \right]$,

$$d_{sum(M,N_{dl},M_{ul},M)}^{No\text{-}CSIT}(r_{sum}) = \begin{cases} d_{M_{ul},M+N_{dl}}(r_{sum}) + M_{ul}N_{dl}\alpha_S + M(N_{dl}-M_{ul}), \ r_{sum} \leq M_{ul} \\ \alpha_S d_{M_{ul},N_{dl}}\left(\frac{r_{sum}-M_{ul}}{W\alpha_S}\right) + M(N_{dl}-M_{ul}), \ M_{ul} \leq r_{sum} \leq M_{ul}(1+W\alpha_S) \\ d_{N_{dl}-M_{ul},M}(r_{sum}-M_{ul}(1+W\alpha_S)), \ M_{ul}(1+W\alpha_S) \leq r_{sum} \leq N_{dl}+M_{ul}W\alpha_S \end{cases}$$

- when $W \geq \frac{M_{ul}+N_{dl}-1}{M-N_{dl}+M_{ul}+1}$,

$$d_{sum(M,N_{dl},M_{ul},M)}^{CSIT}(r_{sum}) = \begin{cases} d_{M_{ul},M+N_{dl}}(r_{sum}) + M_{ul}N_{dl}\alpha_S + M(N_{dl}-M_{ul}), \ r_{sum} \leq M_{ul} \\ d_{N_{dl}-M_{ul},M}(r_{sum}-M_{ul})) + M_{ul}N_{dl}\alpha_S, \ M_{ul} \leq r_{sum} \leq N_{dl} \\ \alpha_S d_{M_{ul},N_{dl}}\left(\frac{r_{sum}-N_{dl}}{W\alpha_S}\right), \ N_{dl} \leq r_{sum} \leq N_{dl}+M_{ul}W\alpha_S \end{cases}$$

Proof The results can be derived by following the similar steps in the proof of Lemma 9. □

Appendix 7
Proof of Lemma 9

As demonstrated in the proof of Corollary 6, we use gradient descent method to find the local optimum value for each value of the multiplexing gain which is equivalent to global optimum in the convex optimization problem. This method is also used in [21] to derive the DMT for MIMO Z-interference channel for some special cases. In our setting of (M, N_{dl}, M_{ul}, M) with $\alpha_{dl} = \alpha_{ul} = \alpha_I = 1$ and $r_{dl} = r_{ul} = r$, we can simplify the objective function in Lemma 3 given sum multiplexing gain. We will first give the analysis when $M_{ul} \geq N_{dl}$. By substituting $v_l' = W v_l$ in (27), we can express the objective function as

$$d_{sum}^{CSIT} = \min \sum_{k=1}^{N_{dl}} (2M + M_{ul} + N_{dl} + 1 - 2k)\theta_k + \frac{1}{W} \sum_{l=1}^{N_{dl}} (M_{ul} + N_{dl} + 1 - 2l) v_l'$$

$$+ \sum_{i=1}^{N_{dl}} (M + N_{dl} + 1 - 2i)\mu_i + \sum_{j=1}^{M_{ul}} (M + M_{ul} + 1 - 2j)\sigma_j - 2MN_{dl}$$

$$+ \sum_{i=1}^{N_{dl}} \sum_{k=1}^{N_{dl}-i} (1 - \mu_i - \theta_k)^+ + \sum_{j=1}^{M_{ul}} \sum_{k=1}^{\min\{M_{ul}-j,N_{dl}\}} (1 - \sigma_j - \theta_k)^+, \tag{127}$$

Subject to $\sum_{i=1}^{N_{dl}} (1 - \mu_i)^+ + \sum_{j=1}^{M_{ul}} (1 - \sigma_j)^+ + \sum_{k=1}^{N_{dl}} (1 - \theta_k)^+ + \sum_{l=1}^{N_{dl}} (W\alpha_S - v_l')^+ < r_{sum};$

$0 \leq \mu_1 \leq \cdots \leq \mu_{N_{dl}}; \ 0 \leq \sigma_1 \leq \cdots \leq \sigma_{M_{ul}}; \ 0 \leq \theta_1 \leq \cdots \leq \theta_{N_1}; \ 0 \leq v_1' \leq \cdots \leq v_{N_{dl}}';$

$\mu_i + \theta_k \geq 1, \ \forall (i+k) \geq N_{dl} + 1;$

$\sigma_j + \theta_k \geq 1, \ \forall (j+k) \geq M_{ul} + 1.$

Next, we differentiate the objective function in (127) with respect to different variables,

$$\frac{\partial d_{sum}^{CSIT}}{\partial v_l'} = \frac{1}{W}(M_{ul} + N_{dl} + 1 - 2l), \ l \leq N_{dl}; \tag{128}$$

$$\left. \frac{\partial d_{sum}^{CSIT}}{\partial \theta_k} \right|_{\mu_i=\sigma_j=1,\forall i,j} = 2M + M_{ul} + N_{dl} + 1 - 2k, \ k \leq N_{dl}; \tag{129}$$

$$\left. \frac{\partial d_{sum}^{CSIT}}{\partial \mu_1} \right|_{\theta_k=1,\forall k} = M + N_{dl} - 1 < \frac{\partial d_{sum}^{CSIT}}{\partial \theta_k}, \forall k; \tag{130}$$

$$\left. \frac{\partial d_{sum}^{CSIT}}{\partial \sigma_1} \right|_{\theta_k=1,\forall k} = M + M_{ul} - 1 < \frac{\partial d_{sum}^{CSIT}}{\partial \theta_k}, \forall k. \tag{131}$$

Since the slope of the objective function decreases with the increasing index of μ_i, σ_j, it suffices to only consider the decay of the function with μ_1, σ_1. We can also easily verify that the decay slopes of μ_1 and σ_1 are smaller than that of $\theta_k, \forall k$.

Case 1

In this case, v_l' has the steepest descent, i.e., $\frac{\partial d_{\text{sum}}^{\text{CSIT}}}{\partial v_{N_{\text{dl}}}'} \geq \frac{\partial d_{\text{sum}}^{\text{CSIT}}}{\partial \theta_1}$.

Thus, when $W \leq \frac{M_{\text{ul}} - N_{\text{dl}} + 1}{2M + M_{\text{ul}} + N_{\text{dl}} - 1}$, for $(l-1)W\alpha_S \leq r_{\text{sum}} \leq lW\alpha_S$, $\forall l$, the steepest descent of the objective function is along the decreasing value of v_l' with $\mu_i = \sigma_j = \theta_k = 1, \forall i, j, k$. Now, the optimization problem becomes

$$d_{\text{sum}}^{\text{CSIT}} = \min \frac{1}{W} \sum_{l=1}^{N_{\text{dl}}} (M_{\text{ul}} + N_{\text{dl}} + 1 - 2l)v_l'$$
$$+ M_{\text{ul}}N_{\text{dl}} + M(M_{\text{ul}} + N_{\text{dl}}),$$

$$\text{Subject to} \quad \sum_{l=1}^{N_{\text{dl}}} (W\alpha_S - v_l')^+ \leq r_{\text{sum}};$$
$$0 \leq v_1' \cdots \leq v_{N_{\text{dl}}}'.$$

Invoking Lemma 6, the solution to the optimization problem above is

$$d_{\mathfrak{B}_{\text{sum}}}^{\text{CSIT}} = \alpha_S d_{M_{\text{ul}}, N_{\text{dl}}}\left(\frac{r_{\text{sum}}}{W\alpha_S}\right) + M_{\text{ul}}N_{\text{dl}} + M(M_{\text{ul}} + N_{\text{dl}}),$$
$$\forall r_{\text{sum}} \leq N_{\text{dl}}W\alpha_S.$$

If $r_{\text{sum}} \geq N_{\text{dl}}W\alpha_S$, it can be implied from the solution above that $v_l' = 0$, $\forall l$ are in the optimal solution. We can see that now the steepest descent of the objective function in (127) is along the decreasing value of θ_k with $\mu_i = \sigma_j = 1, \forall i, j$, and the corresponding optimization function becomes

$$d_{\text{sum}}^{\text{CSIT}} = \min \sum_{k=1}^{N_{\text{dl}}} (2M + M_{\text{ul}} + N_{\text{dl}} + 1 - 2k)\theta_k$$
$$+ MM_{\text{ul}} - MN_{\text{dl}}$$

$$\text{Subject to} \quad \sum_{k=1}^{N_{\text{dl}}} (1 - \theta_k)^+ \leq r_{\text{sum}} - N_{\text{dl}}W\alpha_S;$$
$$0 \leq \theta_1 \leq \cdots \leq \theta_{N_{\text{dl}}}.$$
$$(132)$$

Again, invoking Lemma 6, we have

$$d_{\mathfrak{B}_{\text{sum}}}^{\text{CSIT}} = d_{N_{\text{dl}}, 2M + M_{\text{ul}}}(r_{\text{sum}} - N_{\text{dl}}W\alpha_S) + M(M_{\text{ul}} - N_{\text{dl}}),$$
$$N_{\text{dl}}W\alpha_S \leq r_{\text{sum}} \leq N_{\text{dl}}W\alpha_S + N_{\text{dl}}.$$

Likewise, when $r_s \geq N_{\text{dl}}W\alpha_S + N_{\text{dl}}$, $\theta_k = 0$ $\forall k$, the optimization problem is given as

$$d_{\text{sum}}^{\text{CSIT}} = \min \sum_{i=1}^{N_{\text{dl}}} (M + N_{\text{dl}} + 1 - 2i)\mu_i$$
$$+ \sum_{j=1}^{M_{\text{ul}}} (M + M_{\text{ul}} + 1 - 2j)\sigma_j$$
$$- 2MN_{\text{dl}} + \sum_{i=1}^{N_{\text{dl}}} \sum_{k=1}^{N_{\text{dl}}-i} (1 - \mu_i)^+ + \sum_{j=1}^{M_{\text{ul}}} \sum_{k=1}^{\min\{M_{\text{ul}}-j, N_{\text{dl}}\}} (1 - \sigma_j)^+;$$

$$\text{Subject to} \quad \sum_{i=1}^{N_{\text{dl}}} (1 - \mu_i)^+ + \sum_{j=1}^{M_{\text{ul}}} (1 - \sigma_j)^+ \leq r_{\text{sum}}$$
$$- N_{\text{dl}}W\alpha_S - N_{\text{dl}};$$
$$0 \leq \mu_1 \leq \cdots \leq \mu_{N_{\text{dl}}}; \ 0 \leq \sigma_1 \leq \cdots \leq \sigma_{M_{\text{ul}}};$$
$$\mu_i \geq 1, \ \forall i + k \geq N_{\text{dl}} + 1, \forall k$$
$$\sigma_j \geq 1, \ \forall j + k \geq M_{\text{ul}} + 1, \forall k.$$
$$(133)$$

Apparently, to minimize the objective function above, we should let $\mu_i = 1, \forall i$ and $\sigma_j = 1, \forall j \geq M_{\text{ul}} - N_{\text{dl}} + 1$. Hence, the last term in (133) can be rewritten as

$$\sum_{j=1}^{M_{\text{ul}}} \sum_{k=1}^{\min\{M_{\text{ul}}-j, N_{\text{dl}}\}} (1 - \sigma_j)^+$$
$$= \sum_{j=1}^{M_{\text{ul}}-N_{\text{dl}}} \min\{M_{\text{ul}} - j, N_{\text{dl}}\}(1 - \sigma_j)^+$$
$$= \sum_{j=1}^{M_{\text{ul}}-N_{\text{dl}}} N_{\text{dl}}(1 - \sigma_j)^+.$$

Combining the results above, the objective function in (133) reduces to

$$d_{\text{sum}}^{\text{CSIT}} = \min \sum_{j=1}^{M_{\text{ul}}-N_{\text{dl}}} (M + M_{\text{ul}} + 1 - 2j)\sigma_j$$
$$+ N_{\text{dl}}(N_{\text{dl}} - M_{\text{ul}}) + \sum_{j=1}^{M_{\text{ul}}-N_{\text{dl}}} N_{\text{dl}}(1 - \sigma_j)^+$$
$$= \sum_{j=1}^{M_{\text{ul}}-N_{\text{dl}}} (M + M_{\text{ul}} - N_{\text{dl}} + 1 - 2j)\sigma_j$$

$$\text{Subject to} \quad \sum_{j=1}^{M_{\text{ul}}-N_{\text{dl}}} (1 - \sigma_j)^+ \leq r_{\text{sum}} - N_{\text{dl}}W\alpha_S - N_{\text{dl}},$$
$$0 \leq \sigma_1 \leq \cdots \leq \sigma_{M_{\text{ul}}-N_{\text{dl}}}$$
$$(134)$$

Thus, the optimization problem above has the following solution

$$d_{\mathfrak{B}_{\text{sum}}}^{\text{CSIT}} = d_{M_{\text{ul}} - N_{\text{dl}}, M} \left(r_{\text{sum}} - N_{\text{dl}}(W\alpha_S + 1) \right),$$
$$N_{\text{dl}}(W\alpha_S + 1) \le r_{\text{sum}} \le N_{\text{dl}} W \alpha_S + M_{\text{ul}}.$$

Case 2

In this case, θ_k has the steepest descent, i.e., $\frac{\partial d_{\text{sum}}^{\text{CSIT}}}{\partial \theta_{N_{\text{dl}}}} \ge \frac{\partial d_{\text{sum}}^{\text{CSIT}}}{\partial v_1'}$. Thus when $W \ge \frac{M_{\text{ul}} + N_{\text{dl}} - 1}{2M + M_{\text{ul}} - N_{\text{dl}} + 1}$, for $k - 1 \le r_{\text{sum}} \le k$, the objective function in (127) decays fastest first along the decreasing values of θ_k with $\mu_i = \sigma_j = 1$, $v_l' = W\alpha_S, \forall i, j, l$. The optimization problem becomes

$$d_{\text{sum}}^{\text{CSIT}} = \min \sum_{k=1}^{N_{\text{dl}}} (2M + M_{\text{ul}} + N_{\text{dl}} + 1 - 2k)\theta_k$$
$$+ M_{\text{ul}} N_{\text{dl}} \alpha_S + M(M_{\text{ul}} - N_{\text{dl}}),$$

$$\text{Subject to} \quad \sum_{k=1}^{N_{\text{dl}}} (1 - \theta_k)^+ \le r_{\text{sum}},$$
$$0 \le \theta_1 \le \cdots \le \theta_{N_{\text{dl}}}. \tag{135}$$

Invoking Lemma 6, the solution to the optimization problem above is

$$d_{\mathfrak{B}_{\text{sum}}}^{\text{CSIT}} = d_{N_{\text{dl}}, 2M + M_{\text{ul}}}(r_{\text{sum}}) + M_{\text{ul}} N_{\text{dl}} \alpha_S$$
$$+ M(M_{\text{ul}} - N_{\text{dl}}), \ \forall r_{\text{sum}} \le N_{\text{dl}}.$$

If $r_{\text{sum}} \ge N_{\text{dl}}$, the optimal solution has $\theta_k = 0 \ \forall k$. We rewrite the objective function as

$$d_{\text{sum}}^{\text{CSIT}} = \min \sum_{i=1}^{N_{\text{dl}}} (M + N_{\text{dl}} + 1 - 2i)\mu_i$$
$$+ \sum_{j=1}^{M_{\text{ul}}} (M + M_{\text{ul}} + 1 - 2j)\sigma_j - 2MN_{\text{dl}}$$
$$+ \frac{1}{W} \sum_{l=1}^{N_{\text{dl}}} (M_{\text{ul}} + N_{\text{dl}} + 1 - 2l)v_l'$$
$$+ \sum_{i=1}^{N_{\text{dl}}} \sum_{k=1}^{N_{\text{dl}}-i} (1 - \mu_i)^+ + \sum_{j=1}^{M_{\text{ul}}} \sum_{k=1}^{\min\{M_{\text{ul}}-j, N_{\text{dl}}\}} (1 - \sigma_j)^+,$$

$$\text{Subject to} \quad \sum_{i=1}^{N_{\text{dl}}} (1 - \mu_i)^+ + \sum_{j=1}^{M_{\text{ul}}} (1 - \sigma_j)^+$$
$$+ \sum_{l=1}^{N_{\text{dl}}} (W\alpha_S - v_l')^+ \le r_{\text{sum}} - N_{\text{dl}},$$
$$0 \le \mu_1 \le \cdots \le \mu_{N_{\text{dl}}}; \ 0 \le \sigma_1 \le \cdots \le \sigma_{M_{\text{ul}}};$$
$$0 \le v_1' \le \cdots \le v_{N_{\text{dl}}}',$$
$$\mu_i \ge 1, \ \forall (i + k) \ge N_{\text{dl}} + 1, \forall k,$$
$$\sigma_j \ge 1, \ \forall (j + k) \ge M_{\text{ul}} + 1, \forall k. \tag{136}$$

Again, in order to minimize the objective function above, it is clear that $\mu_i = 1, \forall i$ and $\sigma_j = 1, \forall j \ge M_{\text{ul}} - N_{\text{dl}} + 1$. Hence, the objective function in (136) reduces to

$$d_{\text{sum}}^{\text{CSIT}} = \min \sum_{j=1}^{M_{\text{ul}} - N_{\text{dl}}} (M + M_{\text{ul}} - N_{\text{dl}} + 1 - 2j)\sigma_j$$
$$+ \frac{1}{W} \sum_{l=1}^{N_{\text{dl}}} (M_{\text{ul}} + N_{\text{dl}} + 1 - 2l)v_l'$$

$$\text{Subject to} \quad \sum_{j=1}^{M_{\text{ul}} - N_{\text{dl}}} (1 - \sigma_j)^+ + \sum_{l=1}^{N_{\text{dl}}} (W\alpha_S - v_l')^+ \le r_{\text{sum}} - N_{\text{dl}},$$
$$0 \le \sigma_1 \le \cdots \le \sigma_{M_{\text{ul}} - N_{\text{dl}}}, \ 0 \le v_1' \le \cdots \le v_{N_{\text{dl}}}'. \tag{137}$$

Now, we have two subcases for the optimization problem in (137) when $r_{\text{sum}} \ge N_{\text{dl}}$.

Subcase A: Let v_l' have steeper descent than σ_1, i.e., $\frac{\partial d_{\text{sum}}^{\text{CSIT}}}{\partial v_{N_{\text{dl}}}'} \ge \frac{\partial d_{\text{sum}}^{\text{CSIT}}}{\partial \sigma_1}$. Thus, when $W \le \frac{M_{\text{ul}} - N_{\text{dl}} + 1}{M + M_{\text{ul}} - N_{\text{dl}} - 1}$, the steepest descent of the objective function in (137) is along the decreasing value of v_l' with $\sigma_j = 1, \forall j$. Thus the solution problem above is

$$d_{\mathfrak{B}_{\text{sum}}}^{\text{CSIT}} = \alpha_S d_{M_{\text{ul}}, N_{\text{dl}}} \left(\frac{r_{\text{sum}} - N_{\text{dl}}}{W\alpha_S} \right) + M(M_{\text{ul}} - N_{\text{dl}}),$$
$$N_{\text{dl}} \le r_{\text{sum}} \le N_{\text{dl}}(1 + W\alpha_S).$$

It is obvious that when $r \ge N_{\text{dl}}(1 + W\alpha_S)$, $v_l' = 0, \forall l$. We can further simplify the optimization problem in (137) as

$$d_{\text{sum}}^{\text{CSIT}} = \min \sum_{j=1}^{M_{\text{ul}} - N_{\text{dl}}} (M + M_{\text{ul}} - N_{\text{dl}} + 1 - 2j)\sigma_j$$

$$\text{Subject to} \quad \sum_{j=1}^{M_{\text{ul}} - N_{\text{dl}}} (1 - \sigma_j)^+ \le r_{\text{sum}} - N_{\text{dl}}(1 + W\alpha_S),$$
$$0 \le \sigma_1 \le \cdots \le \sigma_{M_{\text{ul}} - N_{\text{dl}}} \tag{138}$$

Hence, the solution to the optimization problem above is

$$d_{\mathfrak{B}_{\text{sum}}}^{\text{CSIT}} = d_{M_{\text{ul}} - N_{\text{dl}}, M} \left(r_{\text{sum}} - N_{\text{dl}}(W\alpha_S + 1) \right),$$
$$N_{\text{dl}}(W\alpha_S + 1) \le r_{\text{sum}} \le N_{\text{dl}} W \alpha_S + M_{\text{ul}}.$$

Subcase B: Let σ_j have steeper descent than v_1', i.e., $\frac{\partial d_{\text{sum}}^{\text{CSIT}}}{\partial \sigma_{M_{\text{ul}} - N_{\text{dl}}}} \ge \frac{\partial d_{\text{sum}}^{\text{CSIT}}}{\partial v_1'}$. Thus, when $W \ge \frac{M_{\text{ul}} + N_{\text{dl}} - 1}{M - M_{\text{ul}} + N_{\text{dl}} + 1}$, the steepest descent of the objective function in (137) is along the decreasing value of σ_j with $v_l' = W\alpha_S, \forall l$. Now, the solution is given as

$$d_{\mathfrak{B}_{\text{sum}}}^{\text{CSIT}} = d_{M_{\text{ul}} - N_{\text{dl}}, M} \left(r_{\text{sum}} - N_{\text{dl}} \right) + M_{\text{ul}} N_{\text{dl}} \alpha_S,$$
$$N_{\text{dl}} \le r_{\text{sum}} \le M_{\text{ul}}.$$

The result above implies that when $r \geq M_{\mathrm{ul}}$, $\sigma_j = 0, \forall j$; hence, the optimization problem in (137) further reduces to

$$d_{\mathrm{sum}}^{\mathrm{CSIT}} = \min \frac{1}{W} \sum_{l=1}^{N_{\mathrm{dl}}} (M_{\mathrm{ul}} + N_{\mathrm{dl}} + 1 - 2l) v_l'$$

$$\text{Subject to} \quad \sum_{l=1}^{N_{\mathrm{dl}}} (W \alpha_{\mathrm{S}} - v_l')^+ \leq r_{\mathrm{sum}} - M_{\mathrm{ul}},$$

$$0 \leq v_1' \cdots \leq v_{N_{\mathrm{dl}}}'.$$

Consequently, we have

$$d_{\mathfrak{B}_{\mathrm{sum}}}^{\mathrm{CSIT}} = \alpha_{\mathrm{S}} d_{M_{\mathrm{ul}}, N_{\mathrm{dl}}} \left(\frac{r_{\mathrm{sum}} - M_{\mathrm{ul}}}{W \alpha_{\mathrm{S}}} \right), M_{\mathrm{ul}} \leq r_{\mathrm{sum}}$$

$$\leq M_{\mathrm{ul}} + N_{\mathrm{dl}} W \alpha_{\mathrm{S}}.$$

The proof will be complete with the analysis for $N_{\mathrm{dl}} > M_{\mathrm{ul}}$, which can be derived following the same argument and thus is skipped to avoid redundancy. By combining all the cases above, we will obtain the results in Lemma 9.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Mobile Data Offload for 3G Networks. White Paper, IntelliNet Technologies, Inc. (2009)
2. K Doppler, M Rinne, C Wijting, CB Ribeiro, K Hugl, Device-to-device communication as an underlay to LTE-advanced networks. Commun. Mag. IEEE. **47**(12), 42–49 (2009). doi:10.1109/MCOM.2009.5350367
3. A Nosratinia, TE Hunter, A Hedayat, Cooperative communication in wireless networks. Commun. Mag. IEEE. **42**(10), 74–80 (2004). doi:10.1109/MCOM.2004.1341264
4. B Han, P Hui, V Kumar, MV Marathe, G Pei, A Srinivasan, in *Proceedings of the 5th ACM Workshop on Challenged Networks*. Cellular traffic offloading through opportunistic communications: a case study (ACM, 2010), pp. 31–38
5. J Bai, A Sabharwal, in *Proceedings of IEEE Asilomar Conference on Signals, Systems and Computers*. Decode-and-cancel for interference cancellation in a three-node full-duplex network (IEEE, Pacific Grove, 2012)
6. J Bai, A Sabharwal, Distributed full-duplex via wireless side-channels: bounds and protocols. Wireless Commun. IEEE Trans. **12**(8), 4162–4173 (2013). doi:10.1109/TWC.2013.071913.122015
7. J Bai, C Liu, A Sabharwal, in *Proceedings of the 4th ACM Workshop on All Things Cellular: Operations, Applications, & Challenges*. Increasing cellular capacity using ISM band side-channels: a first study (ACM, Chicago, Illinois, 2014), pp. 9–14
8. M Duarte, A Sabharwal, V Aggarwal, R Jana, KK Ramakrishnan, CW Rice, NK Shankaranarayanan, Design and characterization of a full-duplex multiantenna system for wifi networks. IEEE Trans. Veh. Technol. **63**(3), 1160–1177 (2014). doi:10.1109/TVT.2013.2284712
9. E Everett, A Sahai, A Sabharwal, Passive self-interference suppression for full-duplex infrastructure nodes. IEEE Trans. Wirel. Commun. **13**(2), 680–694 (2014). doi:10.1109/TWC.2013.010214.130226
10. A Sabharwal, P Schniter, D Guo, DW Bliss, S Rangarajan, R Wichman, In-band full-duplex wireless: challenges and opportunities. Selected Areas Commun. IEEE J. **32**(9), 1637–1652 (2014)
11. 3GPP TSG RAN (2010). Meeting R1-60, SanFrancisco
12. IEEE 802.11ax-High Efficiency WLAN (HEW) Standardization and Potential Technologies, IEEE 802.11 Task Group AX
13. Small Cell Forum. http://www.smallcellforum.org/
14. C Xing, S Ma, Y Zhou, Matrix-monotonic optimization for MIMO systems. IEEE Trans. Signal Process. **63**(2), 334–348 (2015)
15. EG Larsson, O Edfors, F Tufvesson, TL Marzetta, in *IEEE Communications Magazine*. Massive MIMO for next generation wireless systems, vol. 52, (2014), pp. 186–195
16. T Han, K Kobayashi, A New Achievable Rate Region for the Interference Channel. IEEE Trans. Inf. Theor. **27**(1), 49–60 (2006). IEEE Press, Piscataway
17. S Karmakar, MK Varanasi, The generalized degrees of freedom region of the MIMO interference channel and its achievability. Inf. Theory IEEE Trans. **58**(12), 7188–7203 (2012). doi:10.1109/TIT.2012.2210191
18. L Zheng, DNC Tse, Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels. Inf. Theory IEEE Trans. **49**(5), 1073–1096 (2003). doi:10.1109/TIT.2003.810646
19. S Boyd, L Vandenberghe, *Convex Optimization*. (Cambridge university press, Cambridge, 2004)
20. A Sezgin, SA Jafar, H Jafarkhani, The diversity multiplexing tradeoff for interference networks (2009). arXiv preprint arXiv:0905.2447
21. S Karmakar, MK Varanasi, The diversity-multiplexing tradeoff of the MIMO Z interference channel (2012). http://arxiv.org/abs/1006.0496
22. S Karmakar, MK Varanasi, The capacity region of the MIMO interference channel and its reciprocity to within a constant gap (2011). http://arxiv.org/abs/1102.0267
23. RA Horn, CR Johnson, *Matrix Analysis*. (Cambridge university press, Cambridge, 1990)
24. S Karmakar, MK Varanasi, in *IEEE Transactions on Information Theory*. The diversity-multiplexing tradeoff of the dynamic decode-and-forward protocol on a MIMO half-duplex relay channel, vol. 57, (2011), pp. 6569–6590