Beyond Dirty Paper Coding for Multi-Antenna Broadcast Channel with Partial CSIT: A Rate-Splitting Approach

Yijie Mao, Bruno Clerckx, Senior Member, IEEE

Abstract

Imperfect Channel State Information at the Transmitter (CSIT) is inevitable in modern wireless communication networks, and results in severe multi-user interference in multi-antenna Broadcast Channel (BC). While the capacity of the multi-antenna (Gaussian) BC with perfect CSIT is known and achieved by Dirty Paper Coding (DPC), the capacity and the capacity-achieving strategy of the multi-antenna BC with imperfect CSIT remain unknown. Conventional approaches therefore rely on applying communication strategies designed for perfect CSIT to the imperfect CSIT setting. In this work, we break this conventional routine and make two major contributions. First, we show that linearly precoded Rate-Splitting (RS), relying on the split of messages into common and private parts and linear precoding at the transmitter, and successive interference cancellation at the receivers, can achieve larger rate regions than DPC in multi-antenna BC with partial CSIT. Second, we propose a novel achievable scheme, denoted as Dirty Paper Coded Rate-Splitting (DPCRS), that relies on RS to split the user messages into common and private parts, and DPC to encode the private parts. We show that the rate region achieved by DPCRS in Multiple-Input Single-Output (MISO) BC with partial CSIT is enlarged beyond that of conventional DPC and that of linearly precoded RS. Gaining benefits from the capability of RS to partially decode the interference and partially treat interference as noise, DPCRS is less sensitive to CSIT inaccuracies, networks loads and user deployments compared with DPC and other existing transmission strategies. The proposed DPCRS acts as a new benchmark and the best-known achievable strategy for multi-antenna BC with partial CSIT.

Index Terms

Dirty Paper Coding (DPC), Multiple-Input Single-Output (MISO), Broadcast Channel (BC), Rate-Splitting Multiple Access (RSMA), partial Channel State Information (CSI) at the Transmitter (CSIT)

This work has been partially supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under grant EP/N015312/1, EP/R511547/1.
I. INTRODUCTION

Current wireless communication networks rely increasingly on multi-antenna/Multiple-Input Multiple Output (MIMO) processing to boost rate performance and manage interference. Although appealing in their concept, multi-antenna networks are nevertheless hampered by several practical factors. Among these, the acquisition of accurate Channel State Information (CSI) knowledge at the Transmitter (CSIT) is a major challenge. The availability of accurate CSIT is crucial for downlink multi-user multi-antenna wireless networks. The beamforming and interference management performance heavily depends on the channel estimation accuracy. Unfortunately, pilot reuse tends to impair channel estimation with inter-cell interference in Time Division Duplex (TDD) and a significant feedback overhead is required to guarantee sufficient feedback accuracy in Frequency Division Duplex (FDD) due to the potentially large number of antennas. Delay, mobility, Radio Frequency (RF) impairments (e.g. phase noise) and inaccurate calibrations of RF chains also contribute to making the CSIT inaccurate. Moreover, CSI may be known only for a subset of links in the network, may be estimated only at the subband level (and not for each subcarrier) and may not be known instantaneously but only statistically. This CSIT inaccuracy results in a multi-user interference problem that is the primary bottleneck of MIMO wireless networks. As an illustration of the severity of the problem, in 4G Long-Term Evolution (LTE)-Advanced, the CSIT inaccuracy leads to significant losses of Multi-User MIMO (MU-MIMO) of at least 30% in terms of cell average throughput, and 42% in terms of cell edge throughput [1]. Similarly Coordinated Multi-Point (CoMP) transmission based on coordinated scheduling and beamforming across a full network leads to performance even worse than single-cell processing because of the inaccurate CSIT in the presence of subband-based feedback in 4G LTE-Advanced [2].

Looking backward, the problem has been to strive to apply techniques designed for perfect CSIT to scenarios with partial CSIT [3]. Indeed, multi-antenna in 4G and 5G have been fundamentally motivated by the assumption of perfect CSIT and their performance assessed in the presence of partial CSIT. This is reflected by the conventional approach used in the past 20 years that consists in identifying a communication theoretic channel, e.g. Multiple-Input Single-Output (MISO) Broadcast Channel (BC), characterize its fundamental limits, e.g. capacity region, identify the capacity-achieving strategy, e.g. Dirty Paper Coding (DPC), simplify the strategy, e.g. using linear precoding, and then incorporate partial CSIT and design robust precoders. This leads to the classical linear precoding framework where any residual interference is treated as noise at the receivers. This conventional approach is further illustrated in Fig. 1(a).

While the ability to provide highly accurate and up-to-date CSIT remains questionable, considerable effort has been devoted to improving the performance of those strategies in the presence
of CSIT uncertainties. Unfortunately, such approaches have been shown partially disappointing (e.g. CoMP in 4G as discussed above) and it is conjectured that following the same path will increase the gap between theory and practice as the density of antennas increases. The caveat of this conventional approach is that the underlying communication strategies were motivated by perfect CSIT, and partial CSIT is brought into the picture only at the end of the design. The rationale why such a conventional approach has been extensively used is that, while the capacity of the multi-antenna (Gaussian) BC with perfect CSIT is known and achieved by DPC [4], [5], the capacity and the capacity-achieving strategy of the multi-antenna BC with imperfect CSIT remain unknown.

In this paper, we consider another approach and wonder whether it would be wiser to design MIMO wireless networks from scratch accounting for partial CSIT and its resulting multi-user interference [3]. The fundamental question and first motivation for this paper is can we design and optimize transmission strategies for multi-user multi-antenna communications under the assumption of partial CSIT? Interestingly, new communication and information theoretic understanding of the fundamental role of partial CSIT on the performance of MIMO wireless networks has appeared. It is now known that to benefit from partial CSIT and tackle the multi-user interference, the transmitter should take a Rate-Splitting (RS) approach that splits the messages into a common and a private parts, encode the common parts into a common stream, and private parts into private streams and superpose in a non-orthogonal manner the common stream on top of all private streams [3]. The common stream is drawn from a codebook shared by all receivers and is intended to one but is decodable by all users, while the private streams are to be decoded by their corresponding receivers only. Such approach is optimal from an information theoretic perspective (Degrees-of-Freedom, DoF) in a $K$-user MISO BC with partial CSIT [6]–[9], and brings partial CSIT early on in the picture as illustrated in Fig. 1(b).
This proposed approach contrasts with the conventional approach (as used in 4G and 5G) that is entirely designed based on private streams transmission. Importantly, the proposed RS approach is a more general framework that boils down to conventional precoding whenever no power is allocated to the common stream. That has for consequence that RS-based approaches achieve equal or better performance compared to conventional precoding. Over the past few years, the benefits of RS-based network design have been shown in the literature of MIMO wireless networks. The dawn of RS in multi-antenna BC is from an information theoretic analysis. RS is shown to achieve the optimal sum DoF [7] and further proved to achieve the entire DoF region [8], [9] of the $K$-user underloaded MISO BC with partial CSIT. The DoF benefits of RS are also studied in the underloaded BC with multiple transmitters [10] and multi-antenna receivers [11] in partial CSIT. In the overloaded MISO BC, RS has been shown to achieve the optimal DoF region with heterogeneous CSIT qualities by superimposing degraded symbols on top of linearly precoded RS symbols. The merits of RS discovered from DoF analysis motivate recent studies of precoder design for RS at finite Signal-to-Noise Ratio (SNR) with both perfect and partial CSIT. Specifically, RS linear precoders have been designed in the conventional MISO BC for sum rate maximization with partial CSIT [7] and perfect CSIT [12], max-min fair transmission with partial CSIT [13], energy efficiency maximization with perfect CSIT [14], transmit power control with partial CSIT [15] and minimizing the mean square error with finite feedback [16]. Besides linearly precoded RS, precoder design of RS with non-linear Tomlinson-Harashima Precoding (THP) in MISO BC has been studied in [17]. Though THP technique does not achieve the performance of DPC, it is less complex and considered as a practical implementation of DPC. The precoders of RS have also been designed in other variants of MISO BC, such as multi-group multicast [18], massive MIMO [19], millimeter-wave systems [20], MISO BC with hardware impairments [21], CoMP joint transmission [22], Cloud Radio Access Networks (C-RAN) [23], Simultaneous Wireless Information and Power Transfer (SWIPT) [24], Non-Orthogonal Unicast and Multicast (NOUM) transmission [25], cooperative RS in MISO BC with user relaying [26], [27]. The capability of 1-layer RS discovered in the literature to partially decode the interference and partially treat residual interference as noise makes RS the fundamental building block for a more general and powerful transmission framework for downlink BC, namely, Rate-Splitting Multiple Access (RSMA) [12]. RSMA uses linearly precoded RS at the transmitter to split the user messages into multiple common messages and a private message. The common messages are recombined and encoded into the common streams for the intended users. Successive Interference Cancellation (SIC) is required at each user to sequentially decode the intended common streams. Such linearly precoded generalized RSMA has been demonstrated to be a powerful framework to bridge and generalize Space Division Multiple Access (SDMA) and Non-Orthogonal Multiple
Access (NOMA), and further boost system spectral and energy efficiencies for downlink MISO BC with both perfect and partial CSIT. As a summary, the developed framework based on RS is not only optimum from an information theoretic (DoF) perspective, it also provides significant performance benefits over the conventional precoding strategies.

Building upon the progress in the RS literature for multi-antenna BC, this paper studies RS and DPC in MISO BC with partial CSIT, and makes two major and novel contributions:

First, this paper shows that linearly precoded RS outperforms DPC in MISO BC with partial CSIT. The performance benefits come from the inherent robustness of RS to partial CSIT. This is the first paper to explicitly make this observation. This is in sharp contrast with the perfect CSIT case, where DPC is known to be capacity achieving [4] and outperform linearly precoded RS [12]. This has major implications for practical communication system designs. On one extreme, DPC can be seen as a full transmit-side interference cancellation strategy. On the other extreme, power-domain NOMA based on Superposition Coding (SC) and SIC can be seen as a full receive-side interference cancellation strategy. Power-domain NOMA, however, wastes the multiplexing gain of the MISO BC and results in poor performance, as explained in [12], [28]. In between, stands RS that can be seen as a smart combination of transmit-side and receive-side interference cancellation strategy where the contribution of the common stream is adjusted according to the level of interference that needs to be canceled by the receiver. What this paper shows is that, in practical deployments subject to partial CSIT, an RS strategy enabling a mix of transmit-side and receive-side interference cancellation strategy such as DPC. This further demonstrates the power of the proposed approach in Fig. 1(b) over the conventional approach in Fig. 1(a). Additionally, and importantly, RS also comes at a lower complexity since it only relies on linear precoding. RS is therefore a promising, powerful, and robust non-orthogonal transmission technique for real-world applications.

Second, this paper shows that, in MISO BC with partial CSIT, one can get even better performance than linearly precoded RS (and DPC) by marrying RS and DPC, and using DPC to encode the private parts of the messages. This leads to another transmission strategy, denoted as Dirty Paper Coded Rate-Splitting (DPCRS). We show that the rate region achieved by DPCRS in MISO BC with partial CSIT is enlarged beyond that of conventional DPC and that of linearly precoded RS. Gaining benefits from the capability of RS to partially decode the interference and partially treat interference as noise, DPCRS is less sensitive to the variation of CSIT inaccuracies, network loads, and user deployments compared with DPC and other existing transmission strategies. Since the capacity region of the MISO BC with partial CSIT remains an open problem, the corresponding capacity achieving strategy for the MISO BC with partial CSIT is not known. Nevertheless, the proposed DPCRS strategy becomes a new benchmark as
the best-known achievable strategy for the MISO BC with partial CSIT.

Organizations: The rest of the paper is organized as follows. The system model is described in Section II. The problem formulation for the proposed strategies is specified in Section III. In Section IV, the proposed algorithm is described followed by the numerical results in Section V. Finally, conclusion is made in Section VI.

Notations: Bold lower and upper case letters denote vectors and matrices, respectively. \( \| \cdot \| \) represents Euclidian norm. The notations \( ( \cdot )^H, ( \cdot )^T, \text{tr}(\cdot), \mathbb{E}\{\cdot\} \) respectively denote the Hermitian, transpose, trace and expectation operators. \( \mathbf{I} \) denotes the identity matrix. \( \sim \) denotes “distributed as and \( \mathcal{CN}(0, \sigma^2) \) denotes the Circularly Symmetric Complex Gaussian (CSCG) distribution with zero mean and variance \( \sigma^2 \).

II. SYSTEM MODEL

In this work, we consider a MISO BC with one multi-antenna Base Station (BS) simultaneously serving \( K \) single-antenna users. The BS is equipped with \( N_t \) transmit antennas and the users are indexed by \( \mathcal{K} = \{1, \ldots, K\} \). The signal user-\( k \) received at the \( t \)th time slot for a transmission over a sequence of \( T \) discrete time slots is given by

\[
y_k(t) = h_k^H(t)x(t) + n_k(t), \forall k \in \mathcal{K}, t \in \{1, \ldots, T\},
\]

where \( h_k(t) \in \mathbb{C}^{N_t \times 1} \) is the channel between the BS and user-\( k \) at the \( t \)th time slot, \( x(t) \in \mathbb{C}^{N_t \times 1} \) is the signal vector transmitted in a given channel use subject to the transmit power constraint \( \mathbb{E}\{\|x(t)\|^2\} \leq P_t \). \( n_k(t) \sim \mathcal{CN}(0, \sigma^2_k) \) is the Additive White Gaussian Noise (AWGN). Without loss of generality, we assume that \( \sigma^2_k = 1, \forall k \in \mathcal{K} \). Hence, the transmit SNR is equal to \( P_t \).

A. Channel Model

Due to the uplink channel estimation error caused by quantized feedback [29], feedback delay [30], [31], etc, CSIT is commonly imperfect. In this work, we assume perfect Channel State Information at the Receivers (CSIR) and partial CSIT, which is modeled by

\[
\mathbf{H}(t) = \hat{\mathbf{H}}(t) + \tilde{\mathbf{H}}(t),
\]

where \( \mathbf{H}(t) = [h_1(t), \ldots, h_K(t)] \) is the actual CSI known at all users, \( \hat{\mathbf{H}}(t) = [\hat{h}_1(t), \ldots, \hat{h}_K(t)] \) is the estimated CSI known at the BS, \( \tilde{\mathbf{H}}(t) = [\tilde{h}_1(t), \ldots, \tilde{h}_K(t)] \) is the CSIT estimation error matrix with each element of the \( k \)-th-column for user-\( k \) characterized by an independent and identically distributed (i.i.d.) zero-mean complex Gaussian distribution variable with \( \mathbb{E}\{\tilde{h}_k(t)\tilde{h}_k(t)^H\} = \sigma^2_{e,k} \mathbf{I} \). The variance of the error \( \sigma^2_{e,k} \) is assumed to scale exponentially with SNR as \( \sigma^2_{e,k} \sim O(P_t^{-\alpha}) \), where \( \alpha \in [0, \infty) \) is interpreted as the quality of CSIT in the high SNR regime [6], [7], [29]–[31]. \( \alpha = 0 \) represents partial CSIT with finite precision, e.g. a constant number of feedback bits, while \( \alpha = \infty \) represents perfect CSIT. The joint distribution of \( \{\mathbf{H}, \hat{\mathbf{H}}\} \) is assumed to be stationary and ergodic [7]. \( \mathbf{H}(t) \) over the entire \( T \) time slots are unknown but the conditional density \( f_{\mathbf{H}\hat{\mathbf{H}}}(|\mathbf{H}|\hat{\mathbf{H}}) \) is assumed to be known at the BS.
We consider frequency-flat block-fading channels with in total $B$ blocks indexed by $B = \{1, \ldots, B\}$. The $T$ time slots are divided into $B$ blocks with $T_b$ slots included in each block, i.e., $T = BT_b$. $H(t)$ is assumed to be constant over each block, i.e., $H(t) = H[b], \forall t \in [(b - 1)T_b + 1, bT_b]$. As we focus on an ergodic transmission, we assume that $B \gg 1$ ($T \gg T_b$). At the beginning of block-$b$, the BS receives a channel estimate $\hat{H}[b]$ and it remains static in block-$b$.

### B. Conventional Dirty Paper Coding

In the conventional DPC [32]–[34], with a certain encoding order $\pi$, the BS starts encoding from message $W_{\pi(1)}$ for user-$\pi(1)$ to message $W_{\pi(K)}$ for user-$\pi(K)$. The messages are encoded into a set of data symbols $\{s_{\pi(1)}(t)\}_{t=1}^T, \ldots, \{s_{\pi(K)}(t)\}_{t=1}^T$ to be transmitted in the corresponding time slots. To simplify the notation, we omit $(t)$ and focus on an arbitrary time slot. The data vector $s \triangleq [s_{\pi(1)}, \ldots, s_{\pi(K)}]^T$ to be transmitted in each time slot is precoded by $P \triangleq [p_{\pi(1)}, \ldots, p_{\pi(K)}]$, where $p_{\pi(k)} \in \mathbb{C}^{N_t \times 1}$ is the precoder for user-$\pi(k)$, and the resulting superposed transmit signal is

$$x = Ps = \sum_{k \in \mathcal{K}} p_{\pi(k)}s_{\pi(k)}.$$  

Assuming CSCG inputs with $\mathbb{E}\{ss^H\} = I$, the transmit power constraint is equivalent to $\text{tr}(PP^H) = P_t$. If CSIT is perfect, the encoded data stream $s_{\pi(k)}$ experiences no interference from previously encoded data streams $\{s_{\pi(i)} | i < k\}$ according to the principle of DPC [34]. However, as the BS has no access to the exact channel $H$, $P$ is designed at the BS based on the estimated channel state $\hat{H}$. Only part of the interference from $\tilde{\hat{h}}_{\pi(k)}^H \sum_{i < k} p_{\pi(i)} s_{\pi(i)}$ is removed from the signal received at user-$\pi(k)$. The resulting received signal is given by

$$y_{\pi(k)} = \tilde{\hat{h}}_{\pi(k)}^H \sum_{i < k} p_{\pi(i)} s_{\pi(i)} + \hat{h}_{\pi(k)}^H \sum_{j \geq k} p_{\pi(j)} s_{\pi(j)} + n_{\pi(k)}, \forall k \in \mathcal{K}.$$  

Each user directly decodes the intended message by treating any residual interference as noise. As the precoders are designed at the BS based on the channel estimate $\hat{H}$ and each user decodes the intended stream based on the exact channel $H$, the instantaneous rate of decoding $s_{\pi(k)}$ at user-$\pi(k)$ is determined by one joint fading state $\{H, \hat{H}\}$ given as

$$R_{\pi(k)}^\text{DPC}(H, \hat{H}) = \log_2 \left( 1 + \frac{|h_{\pi(k)}^H|^2}{\sum_{i < k} |h_{\pi(k)}^H p_{\pi(i)}|^2 + \sum_{j > k} |h_{\pi(k)}^H p_{\pi(j)}|^2 + 1} \right), \forall k \in \mathcal{K}.$$  

As the BS only knows the channel estimate $\hat{H}$ without any knowledge of the exact channel $H$, $R_{\pi(k)}^\text{DPC}(H, \hat{H})$ may be overestimated and unachievable at user-$\pi(k)$ [7]. A more robust approach is to design the precoders at the BS based on the Ergodic Rate (ER) under the assumption that the transmission is delay-unlimited. The ER characterizes the long-term performance of user-$\pi(k)$ over all possible joint fading states $\{H, \hat{H}\}$, which is defined as

$$ER_{\pi(k)}^\text{DPC} \triangleq \mathbb{E}_{\{H, \hat{H}\}} \left\{ R_{\pi(k)}^\text{DPC}(H, \hat{H}) \right\}.$$  

Transmission at rate $ER_{\pi(k)}^\text{DPC}$ is reliable and it guarantees that $W_{\pi(k)}$ is decodable at user-$\pi(k)$.
C. Dirty Paper Coded Rate-Splitting

1) Motivation: The sum DoF achieved by RS in an underloaded ($N_t \geq K$) $K$-user MISO BC with partial CSIT is given by $1 + (K - 1)\alpha$ [7]. This sum DoF matches the upper bound obtained from the Aligned Image Sets in [6]. As a consequence, RS achieves the optimal DoF in this setting. This contrasts with the conventional approach of Fig. 1(a) that achieves a sum DoF of $\max\{1, K\alpha\}$ [7]. Interestingly, $1 + (K - 1)\alpha$ can be equivalently written as $(1 - \alpha) + K\alpha$. Leveraging the weighted-sum interpretation in [10], [35] and the notion of signal-space partitioning in [36], [37], one can interpret $(1 - \alpha) + K\alpha$ as the DoF achieved by the superposition of two sub-networks in the power domain: a first sub-network consisting of a $K$-user MISO BC with perfect CSIT using a power level $\alpha$ contributing to a sum DoF of $K\alpha$, and a second sub-network consisting of a $K$-user MISO BC with no CSIT using the remaining power level $1 - \alpha$ contributing to a sum DoF of $1 - \alpha$, as illustrated in Fig. 2. Loading data onto those two sub-networks is achieved by a non-orthogonal transmission in the power domain using RS that splits messages into common and private parts, with the private parts loaded onto the first sub-network and the common parts onto the second sub-network. Since the first sub-network can be viewed as a $K$-user MISO BC with perfect CSIT, and DPC is capacity-achieving for such a scenario, it motivates us to encode the private parts using DPC. This leads to the 1-layer Dirty Paper Coded RS discussed in the sequel.

2) One-Layer Dirty Paper Coded Rate-Splitting (1-DPCRS): Though DPC achieves the capacity region of MISO BC with perfect CSIT [32]–[34], it is sensitive to the CSIT inaccuracy [38]. Motivated by the DoF interpretation in Fig. 2 and the recent benefits of RS in multi-antenna BC, we first marry 1-layer RS with DPC so as to combat performance losses of DPC resulting from partial CSIT and explore a larger achievable rate region in MISO BC with partial CSIT. The proposed “1-layer Dirty Paper Coded RS (1-DPCRS)” strategy is illustrated in Fig. 3(a).

In 1-DPCRS, message $W_k$ intended for user-$k$, $\forall k \in K$ is first split into one common part $W_{c,k}$
and one private part $W_{p,k}$. The common parts $W_{c,1}, \ldots, W_{c,K}$ of all users are combined into the common message $W_c$ and encoded into the set of common streams $\{s_c(t)\}_{t=1}^{T}$ to be decoded by all users in the respective time slots. Different from the linearly precoded 1-layer RS strategy studied in the literature [3], [7], [12], [39], the private parts $W_{p,1}, \ldots, W_{p,K}$ are encoded and precoded by DPC with a certain encoding order $\pi$ into the private streams $\{s_{\pi(1)}(t)\}_{t=1}^{T}, \ldots, \{s_{\pi(K)}(t)\}_{t=1}^{T}$ to be decoded by the corresponding users only. The data vector transmitted in each time slot $s \triangleq [s_c, s_{\pi(1)}, \ldots, s_{\pi(K)}]^T$ (with $(t)$ omitted) is precoded by $P \triangleq [p_c, p_1, \ldots, p_K]$, the resulting transmit signal is

$$x = Ps = p_cs_c + \sum_{k \in \mathcal{K}} p_{\pi(k)}s_{\pi(k)}. \quad (7)$$

The transmit power constraint is $\text{tr}(PP^H) = Pt$ and CSCG inputs with $\mathbb{E}\{ss^H\} = I$ are assumed.

At user sides, user-$\pi(k)$ first decodes the common stream $s_c$ into $\hat{W}_c$ by treating the interference from all private streams as noise. With the assist of SIC, the decoded common message $\hat{W}_c$ then goes through the process of re-encoding, precoding, and subtracting from the received signal. After decoding the common stream, user-$\pi(k)$ then decodes the intended private stream $s_{\pi(k)}$ into $\hat{W}_{p,\pi(k)}$ by treating the interference from the private streams encoded after $s_{\pi(k)}$ as noise. Once $\hat{W}_c$ and $\hat{W}_{p,\pi(k)}$ are decoded, user-$\pi(k)$ reconstructs the original message by extracting $\hat{W}_{c,\pi(k)}$ from $\hat{W}_c$, and then combines $\hat{W}_{c,\pi(k)}$ with $\hat{W}_{p,\pi(k)}$ into $\hat{W}_{\pi(k)}$. The instantaneous rate of decoding the common stream $s_c$ at user-$\pi(k)$ is

$$R_{c,\pi(k)}^{1\text{-DPCRS}}(H, \hat{H}) = \log_2 \left( 1 + \frac{|h_{\pi(k)}^H p_c|^2}{\sum_{j \in \mathcal{K}} |h_{\pi(k)}^H p_{\pi(j)}|^2 + 1} \right). \quad (8)$$

The instantaneous rate $R_{\pi(k)}^{1\text{-DPCRS}}(H, \hat{H})$ of decoding the private stream $s_{\pi(k)}$ at user-$\pi(k)$ is the same as equation (5). The respective ERs of decoding $s_c$ and $s_{\pi(k)}$ at user-$\pi(k)$ using 1-DPCRS are defined as

$$ER_{c,\pi(k)}^{1\text{-DPCRS}} \triangleq \mathbb{E}_{(H, \hat{H})}\{R_{c,\pi(k)}^{1\text{-DPCRS}}(H, \hat{H})\}, \quad ER_{\pi(k)}^{1\text{-DPCRS}} \triangleq \mathbb{E}_{(H, \hat{H})}\{R_{\pi(k)}^{1\text{-DPCRS}}(H, \hat{H})\}. \quad (9)$$

To ensure $s_c$ is successfully decoded by all users, the ER of the common stream $s_c$ should not exceed

$$ER_c^{1\text{-DPCRS}} \triangleq \min\{ER_{c,\pi(k)}^{1\text{-DPCRS}} \mid k \in \mathcal{K}\}. \quad (10)$$

As the common stream is shared by all users, by denoting the ER of the common stream allocated to user-$\pi(k)$ as $C_{\pi(k)}$, we have $\sum_{\pi(k) \in \mathcal{K}} C_{\pi(k)} = ER_c^{1\text{-DPCRS}}$. Therefore, the total ER of each user using 1-DPCRS is $ER_{\pi(k),\text{tot}}^{1\text{-DPCRS}} = C_{\pi(k)} + ER_{\pi(k)}^{1\text{-DPCRS}}$.

3) Multi-Layer Dirty Paper Coded Rate-Splitting (M-DPCRS): To further exploit a larger achievable rate region for MISO BC with partial CSIT, we incorporate the generalized RSMA framework proposed in [12] with DPC and propose a novel transmission scheme, namely, “Multi-layer Dirty Paper Coded RS (M-DPCRS)” . We claim that the proposed M-DPCRS achieves the best-known achievable rate so far in the $K$-user MISO BC with partial CSIT.
Compared with the linearly precoded RSMA framework proposed in [12], the main difference of the proposed strategy comes from the non-linear DPC encoding and precoding for the private streams. How the user messages are split and combined follows the framework in [12]. To simplify the explanation, we introduce a three-user M-DPCRS, where the users are indexed as $K = \{1, 2, 3\}$. It can be easily generalized to the $K$-user case if readers understand the three-user M-DPCRS as well as the two-user 1-DPCRS. As illustrated in Fig. 3(b), user message $W_k$ of user-$k$ is split into four parts at the BS as $\{W_{ik}|i \in I_k\}$, where $I_1 = \{123, 12, 13, 1\}$, $I_2 = \{123, 12, 23, 2\}$, $I_3 = \{123, 13, 23, 3\}$. This is different from 1-DPCRS described in Section II-C where the message of each user is only split into two parts as $\{W_{c,k}, W_{p,k}\}$. The intention of splitting user messages into more different parts is to form more layers of common streams for different users, so as to manage interference and disparity of channel strengths more flexibly. The sub-messages $\{W_{123}^k|k \in K\}$ are jointly encoded into the set of common streams $\{s_{123}(t)\}_{t=1}^T$, which are decoded by all the three users in the corresponding time slots. The sub-messages $\{W_{12}^1, W_{12}^2\}$ and $\{W_{13}^1, W_{13}^3\}$ and $\{W_{23}^2, W_{23}^3\}$ are respectively encoded into the partial-common streams $\{s_{12}(t)\}_{t=1}^T$, $\{s_{13}(t)\}_{t=1}^T$, $\{s_{23}(t)\}_{t=1}^T$ to be decoded by the intended two users. The private messages $W_k^k, \forall k \in K$ are encoded via DPC with encoding order $\pi$ into the private streams $\{s_{\pi(1)}(t)\}_{t=1}^T, \{s_{\pi(2)}(t)\}_{t=1}^T, \{s_{\pi(3)}(t)\}_{t=1}^T$ for the respective user only. We assume the private stream of user-$\pi(k)$ is encoded after user-$\pi(i)$ if $k > i$. Define the common stream index set $K_c \triangleq \{123, 12, 13, 23\}$. With $(t)$ omitted, the encoded stream vector in each time slot $s \triangleq \{s_k|k \in K_c \cup K\}$ is precoded by $P \triangleq \{p_k|k \in K_c \cup K\}$. The resulting transmit signal $x \in \mathbb{C}^{N_t \times 1}$ is

$$x = Ps = \sum_{i \in K_c} p_i s_i + \sum_{k \in K} p_{\pi(k)} s_{\pi(k)}. \quad (11)$$

In the three-user M-DPCRS strategy, each user requires three layers of SIC to sequentially decode and remove the intended common streams before decoding the intended private stream. We follow the rule that the data streams intended for a larger number of users have higher decoding priorities [12], [19]. Therefore, $s_{123}$ is decoded first at all users, followed by the
intended partial-common streams and the private stream is decoded at the end. For a certain decoding order \( \pi' \) of the partial-common streams\(^1\) \( s_{12}, s_{13}, s_{23} \), we obtain the corresponding decoding order \( \pi_k' \) of the partial-common streams to be decoded at user-\( k \). For example, when \( \pi' = 12 \rightarrow 13 \rightarrow 23 \) as illustrated in Fig. 3(b), the decoding order at user-1 is \( \pi_1' = 12 \rightarrow 13 \), where \( s_{\pi_1'(1)} \triangleq s_{12} \) is decoded first followed by \( s_{\pi_1'(2)} \triangleq s_{13} \). Hence, user-\( \pi(k) \) sequentially decodes the common streams \( s_{123}, s_{\pi_k'(1)}, s_{\pi_k'(2)} \) and the private stream \( s_{\pi(k)} \) for a given DPC encoding order \( \pi \) of the private streams and a given decoding order \( \pi' \) of the partial-common streams. The instantaneous rates of decoding \( s_{123}, s_{\pi_k'(1)}, s_{\pi_k'(2)}, s_{\pi(k)} \) at user-\( \pi(k) \) are

\[
R^\text{M-DPCRS}_{i, \pi(k)}(H, \hat{H}) = \log_2 \left( 1 + \frac{|\mathbf{h}^H_{\pi(k)} \mathbf{p}_i|^2}{I^i_{\pi(k)}} \right), 
\]

where

\[
I^i_{\pi(k)} \triangleq \begin{cases} 
\sum_{j \in (K_c \setminus \{123\}) \cup K} |\mathbf{h}^H_{\pi(k)} \mathbf{p}_j|^2 + 1, & i = 123 \\
\sum_{j \in (K_c \setminus \{123, \pi_k'(1)\}) \cup K} |\mathbf{h}^H_{\pi(k)} \mathbf{p}_j|^2 + 1, & i = \pi_k'(1) \\
\sum_{j \in (K_c \setminus \{123, \pi_k'(1), \pi_k'(2)\}) \cup K} |\mathbf{h}^H_{\pi(k)} \mathbf{p}_j|^2 + 1, & i = \pi_k'(2) \\
\sum_{j \in (K_c \setminus \{123, \pi_k'(1), \pi_k'(2)\})} |\mathbf{h}^H_{\pi(k)} \mathbf{p}_j|^2 + \sum_{j > k} |\mathbf{h}^H_{\pi(k)} \mathbf{p}_{\pi(j)}|^2 + 1, & i = \pi(k)
\end{cases}
\]

In the three-user M-DPCRS, the ERs of decoding the intended common and private streams \( \{s_i \mid i \in I_{\pi(k)}\} \) at user-\( \pi(k) \) are defined as

\[
ER^\text{M-DPCRS}_{i, \pi(k)} \triangleq \mathbb{E}_{[H, \hat{H}]} \left[ R^\text{M-DPCRS}_{i, \pi(k)}(H, \hat{H}) \right], 
\]

where \( R^\text{M-DPCRS}_{i, \pi(k)}(H, \hat{H}) = R^\text{M-DPCRS}_{\pi(k)}(H, \hat{H}) \) when \( i = \pi(k) \). To ensure the common streams \( \{s_i \mid i \in K_c\} \) are successfully decoded by the intended users, the ERs of the common streams should not exceed \( ER^\text{M-DPCRS}_i \triangleq \min\left\{ ER^\text{M-DPCRS}_{i, \pi(k)} \mid \pi(k) \in K_i \right\} \), where \( K_i \) denotes the group of users decoding the common stream \( s_i \). For instance, \( K_i = \{1, 2\} \) when \( i = 12 \) and \( K_i = K \) when \( i = 123 \). By introducing scalar variable \( C^i_k \) to denote the ER of common stream \( s_i \) allocated to user-\( k \), we have \( \sum_{k \in K_i} C^i_k = ER^\text{M-DPCRS}_i \). As the message of each user is split into four parts, the ER of user-\( \pi(k) \) is the sum of the rate allocated to user-\( \pi(k) \) in the intended common streams, i.e.,

\[
ER^\text{M-DPCRS}_{\pi(k), \text{tot}} = \sum_{i \in I_{\pi(k)} \setminus \{\pi(k)\}} C^i_{\pi(k)} + ER^\text{M-DPCRS}_{\pi(k)}.
\]

D. Extension to MISO BC with a Multicast Message

The transmission models in Section II-B and II-C for MISO BC can be extended to what is called in the information theory literature as a “MISO BC with a common message” \([40, 41]\).

\(^1\)Notice that \( \pi' \) is different from \( \pi \). \( \pi' \) is the decoding order of the linearly precoded common streams at all users while \( \pi \) is the encoding order of the DPC-coded private streams at the BS.
In order to avoid confusion between this common message and the ones originating from RS, we will refer to this scenario as “MISO BC with a multicast message”. From the perspective of fundamental limits, SC combined with DPC has been first studied in [40] and then proved in [41] to achieve the capacity region of the two-user MISO BC with a multicast message in perfect CSIT. As the capacity region in partial CSIT is unknown, we derive a new communication strategy achieving a larger achievable rate than the existing approach (namely SC combined with DPC), similarly to what we did in Section II-C for MISO BC. We also claim that the proposed 1-DPCRS and M-DPCRS for MISO BC with a multicast message achieve larger achievable rate regions than the conventional SC and DPC-assisted strategy in partial CSIT. Specifically, M-DPCRS-assisted strategy achieves the best-known achievable system throughput so far for the $K$-user MISO BC with a multicast message in partial CSIT.

The schemes we propose adopt DPC encoding and precoding for the unicast streams while the multicast streams are linearly precoded as illustrated in Fig. 4. Comparing with Fig. 3(a), the super-common stream $s_0$ to be decoded by all users in Fig. 4(a) contains not only the entire multicast message $W_0$ but also the split common parts of the unicast messages $W_{c,1}, \ldots, W_{c,K}$ [25]. The resulting transmit signal at each time slot is in the same form as (7) by replacing $p_{c}s_c$ with $p_0s_0$. The instantaneous rates $R_{0,\pi(k)}^{1\text{-DPCRS}}(\mathbf{H}, \hat{\mathbf{H}})$, $R_{\pi(k)}^{1\text{-DPCRS}}(\mathbf{H}, \hat{\mathbf{H}})$ and the ERs $ER_{0,\pi(k)}^{1\text{-DPCRS}}$, $ER_{\pi(k)}^{1\text{-DPCRS}}$ of the super-common and private streams are defined in the same way as 1-DPCRS described in Subsection II-C for MISO BC. Denote the ER allocated to the multicast message $W_0$ as $C_0$, we obtain that $C_0 + \sum_{k \in \mathcal{K}} C_k = ER_{0,\pi(k)}^{1\text{-DPCRS}}$, where $ER_{0,\pi(k)}^{1\text{-DPCRS}} = \min \{ ER_{0,\pi(k)}^{1\text{-DPCRS}} \mid \pi \in \mathcal{K} \}$. The total ER of decoding the unicast stream $W_{\pi(k)}$ at user-$\pi(k)$ using 1-DPCRS is $ER_{\pi(k),\text{tot}}^{1\text{-DPCRS}} = C_{\pi(k)} + ER_{\pi(k)}^{1\text{-DPCRS}}$. The system model of M-DPCRS-assisted MISO BC with a multicast message can be easily traced out from M-DPCRS proposed in Section II-C for MISO BC and 1-DPCRS in Fig. 4(a) for MISO BC with a multicast message. Fig. 4(b) illustrates the three-user M-DPCRS strategy for MISO BC with a multicast message with decoding order $\pi' = 12 \rightarrow 13 \rightarrow 23$. 
III. PROBLEM FORMULATION

An intuitive method of precoder design at the BS is to optimize the instantaneous precoder $\mathbf{P}$ based on the knowledge of the estimated channel state $\hat{\mathbf{H}}$ by maximizing the instantaneous Weighted Sum Rate (WSR) subject to instantaneous power constraint $\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t$. However, the partial CSIT may lead to an undecodable rate [7]. An alternative method is to maximize the Weighted Ergodic Sum Rate (WESR) where the transmission takes place over a long sequence of blocks spanning almost all possible channel states. In this section, the precoder and message-split design problems for WESR maximization in MISO BC and MISO BC with a multicast message transmission are formulated.

Though each BS is not able to estimate the instantaneous rates, the Average Rates (ARs) of users are predictable at each user [7]. The ARs are defined in Definition 1.

**Definition 1.** The AR of decoding the stream $s_i$ at user-$k$, $k \in \mathcal{K}$ for a given channel estimate $\hat{\mathbf{H}}$ and precoder $\mathbf{P}(\hat{\mathbf{H}})$ is given by

$$\bar{R}_{x,i,k}(\hat{\mathbf{H}}) \triangleq \mathbb{E}_{\hat{\mathbf{H}}} \left\{ R_{x,i,k}(\mathbf{H}, \hat{\mathbf{H}}) \mid \hat{\mathbf{H}} \right\},$$

where $R_{x,i,k}(\mathbf{P}) = R_{k}(\mathbf{P})$ when $i = k$. $x \in \{"DPC","1-DPCRS","M-DPCRS"\}$.

Notice that AR is a short-term measure for an instantaneous channel estimate $\hat{\mathbf{H}}[b]$ at the BS. It captures the expected rate over the CSIT uncertainty for a given $\hat{\mathbf{H}}[b]$ and conditional density $f_{\mathbf{H} \mid \hat{\mathbf{H}}}(\mathbf{H} \mid \hat{\mathbf{H}})$. This is different from ER, that captures the long-term performance over all joint fading states $\{\mathbf{H}[b], \hat{\mathbf{H}}[b]\}_{b=1}^B$. Following the law of total expectation, we could then obtain the relation between ER and AR [7] as

$$ER_{x,i,k} = \mathbb{E}_{\hat{\mathbf{H}}} \left\{ \bar{R}_{x,i,k}(\hat{\mathbf{H}}) \right\},$$

since $\mathbb{E}_{\hat{\mathbf{H}}} \left\{ R_{x,i,k}(\mathbf{H}, \hat{\mathbf{H}}) \right\} = \mathbb{E}_{\hat{\mathbf{H}}} \left\{ \mathbb{E}_{\hat{\mathbf{H}}} \{ R_{x,i,k}(\mathbf{H}, \hat{\mathbf{H}}) \mid \hat{\mathbf{H}} \} \right\}$.

Under the assumption of finite SNR and bounded channel state realizations, the ER is then approximated over a sufficiently large number of channel blocks $B$ such that all possible channel states are experienced, which is given as

$$ER_{x,i,k} \approx \frac{1}{B} \sum_{b=1}^{B} \bar{R}_{x,i,k}(\hat{\mathbf{H}}[b]),$$

where the precoder $\mathbf{P}[b]$ of the average rate $\bar{R}_{x,i,k}(\hat{\mathbf{H}}[b])$ is designed at the BS based on $\hat{\mathbf{H}}[b]$.

A. MISO BC

1) **DPC**: Following (15)–(17), the WESR, defined by $WESR_{\pi}^{DPC} \triangleq \sum_{k \in \mathcal{K}} u_{\pi(k)}ER_{\pi(k)}^{DPC}$, can be approximated as $WESR_{\pi}^{DPC} \approx \frac{1}{B} \sum_{b=1}^{B} \bar{R}_{\pi}^{DPC}(\hat{\mathbf{H}}[b])$, where $\bar{R}_{\pi}^{DPC}(\hat{\mathbf{H}}[b]) \triangleq \sum_{k \in \mathcal{K}} u_{\pi(k)}\bar{R}_{\pi(k)}^{DPC}(\hat{\mathbf{H}}[b])$ is the Weighted Average Sum Rate (WASR) for a given channel estimate $\hat{\mathbf{H}}[b]$. 
In this work, we aim at designing the precoder $P^{[b]}$ in each block adaptively so as to maximize the WESR subject to the short-term rate and power constraints. For a given weight vector $u = [u_1, \ldots, u_K]$ and a fixed encoding order $\pi$, the WESR achieved by DPC is given by

$$\max_{\{P^{[b]}\}_{b=1}^B} \frac{1}{B} \sum_{b=1}^B \tilde{R}_{\pi}^{DPC}(\hat{H}[b])$$  \hspace{1cm} (18a)

subject to

$$\tilde{R}_{\pi}^{DPC}(\hat{H}[b]) \geq R_{\pi(k)}^{th}, \ \forall k \in K, \forall b \in B$$  \hspace{1cm} (18b)

$$\text{tr}(P^{[b]}(P^{[b]})^H) \leq P_t, \forall b \in B,$$  \hspace{1cm} (18c)

where $R_{\pi(k)}^{th}$ is the QoS rate constraint for user-$\pi(k)$. As the precoder design in (18) is performed within each block without inter-block dependencies, problem (18) is decomposed into the following WASR subproblem for each block with $[b]$ omitted:

$$\max_{P} \sum_{k \in K} u_{\pi(k)} \tilde{R}_{\pi(k)}^{DPC}(\hat{H})$$  \hspace{1cm} (19a)

subject to

$$\tilde{R}_{\pi(k)}^{DPC}(\hat{H}) \geq R_{\pi(k)}^{th}, \ \forall k \in K$$  \hspace{1cm} (19b)

$$\text{tr}(PP^H) \leq P_t.$$  \hspace{1cm} (19c)

The WESR is maximized if the precoder of each block is optimized by solving (19). Denote the optimized rate vector of problem (19) as $\bar{R}_{\pi, u}^{DPC,*} = \{\tilde{R}_{\pi(1)}^{DPC,*}(\hat{H}), \ldots, \tilde{R}_{\pi(K)}^{DPC,*}(\hat{H})\}$. By calculating $\bar{R}_{\pi, u}^{DPC,*}$ for a set of different weight vectors, the boundary points on the AR region for a certain encoding order $\pi$ and channel estimate $\hat{H}$ are attained. The ER $ER_{\pi, u}^{DPC,*}$ for a given encoding order $\pi$ and user weights $u$ is obtained by averaging $\bar{R}_{\pi, u}^{DPC,*}$ over all blocks. The corresponding ER region is the convex hull over all the boundary points as $R_{\pi}^{DPC} = \text{Conv} (\bigcup_u \text{ER}_{\pi, u}^{DPC,*})$. Therefore, we obtain the ER region achieved by DPC with QoS rate constraints $R_{\pi(k)}^{th}, \forall k \in K$ given as $R_{\pi}^{DPC} = \text{Conv} (\bigcup_{\pi, u} R_{\pi, u}^{DPC})$.

Remark 1. When CSIT is perfect, the capacity region of MISO BC is achieved by solving problem (18) with $R_{\pi(k)}^{th} = 0$ for all possible $\pi$ and a set of $u$. Though it is typical to optimize covariance matrix $Q_\pi = P_\pi^H P_\pi$ in the literature of DPC [34], [42], the optimal covariance matrix of DPC when each user has a single antenna is rank one due to the uplink-downlink duality [32], [33]. The dual MISO MAC channel has a single transmit antenna at each user and thus a rank one covariance matrix. The transformation to the corresponding covariance matrix in MISO BC preserves this rank. Hence, the capacity region can be achieved by optimizing the precoder.

2) 1-DPCRS: Following (15)–(17), the WESR, defined by $\text{WESP}_{\pi}^{1-DPCRS} \triangleq \sum_{k \in K} u_{\pi(k)} \text{ER}_{\pi(k), \text{tot}}^{1-DPCRS}$, can be approximated as $\text{WESP}_{\pi}^{1-DPCRS} \approx \frac{1}{B} \sum_{b=1}^B \bar{R}_{\pi}^{1-DPCRS}(\hat{H}[b])$, where $\bar{R}_{\pi}^{1-DPCRS}(\hat{H}[b])$ is the WASR for a given $\hat{H}[b]$, defined as $\bar{R}_{\pi}^{1-DPCRS}(\hat{H}[b]) \triangleq \sum_{k \in K} u_{\pi(k)} \bar{R}_{\pi(k), \text{tot}}^{1-DPCRS}(\hat{H}[b])$. $\bar{R}_{\pi(k), \text{tot}}^{1-DPCRS}(\hat{H}[b]) \triangleq$
$C_{\pi(k)}^{[b]} + \bar{R}_{\pi(k)}^{1-DPCRS}(\hat{H}[b])$ is the total AR of user-$\pi(k)$ and $\tilde{C}_{\pi(k)}^{[b]}$ is the AR of the common stream allocated to user-$\pi(k)$ in each block such that $\sum_{k \in K} C_{\pi(k)}^{[b]} = \bar{R}_c^{1-DPCRS}(\hat{H}[b])$.

Note that the overall ER of the common stream specified in (10) with minimization outside the ER of each user is intractable since the precoder of all fading states are required to be jointly designed. To remove the inter-block dependency, we consider its lower bound by moving minimization inside of the ER based on the inequality (7):

$$\min_{k \in K} \left\{ \mathbb{E}_{\hat{H}} \left\{ R_{\pi(k)}^{1-DPCRS}(\hat{H}) \right\} \right\} \geq \mathbb{E}_{\hat{H}} \left\{ \min \left\{ R_{\pi(k)}^{1-DPCRS}(\hat{H}) \mid k \in K \right\} \right\} ,$$

and approximate the common rate as $ER_c^{1-DPCRS} \approx \frac{1}{B} \sum_{b=1}^B \bar{R}_c^{1-DPCRS}(\hat{H}[b])$, where $\bar{R}_c^{1-DPCRS}(\hat{H}[b]) = \min\{ R_{\pi(k)}^{1-DPCRS}(\hat{H}[b]) \mid k \in K \}$ is the overall AR of the common stream $s_c$.

Due to the inter-block independence, the WESR achieved by 1-DPCRS can be equivalently decomposed into the following WASR maximization subproblem for each block (by omitting $[b]$ for simplicity) with a given weight vector $u = [u_1, \ldots, u_K]$ and a fixed encoding order $\pi$:

$$\max_{e, \mathbf{P}} \sum_{k \in K} u_{\pi(k)} \bar{R}_{\pi(k),tot}^{1-DPCRS}(\hat{H})$$

s.t. $\sum_{k \in K} \bar{C}_k \leq \bar{R}_{\pi(k),tot}^{1-DPCRS}(\hat{H})$,

$$\bar{R}_{\pi(k),tot}^{1-DPCRS}(\hat{H}) \geq R_{\pi(k)}^{th}, \forall k \in K$$

$$|\text{tr}(\mathbf{P}\mathbf{P}^H)| \leq P_t$$

$$\bar{e} \geq 0,$$

where $\bar{e} = [\bar{C}_1, \ldots, \bar{C}_K]$ is the AR allocation for the common stream $s_c$ in each block. It is required to be jointly optimized with the precoder so as to maximize the WASR. Denote the optimized rate vector of problem (21) as $\bar{R}_{\pi,u}^{1-DPCRS} = \{ \bar{R}_{\pi(1)}^{1-DPCRS}(\hat{H}), \ldots, \bar{R}_{\pi(K)}^{1-DPCRS}(\hat{H}) \}$. The boundary points of the ER region $R_{\pi,tot}^{1-DPCRS}$ for a certain encoding order $\pi$ and channel estimate $\hat{H}$ can be calculated by solving problem (21) for a set of different weight vectors $u$ and averaging the rate vector $\bar{R}_{\pi,u}^{1-DPCRS}$ over all blocks for each $u$. The entire ER region of 1-DPCRS under QoS rate constraints $R_{\pi,k}^{th}, \forall k \in K$ is the convex hull of the rate regions of all encoding orders, i.e., $R_{1-DPCRS} = \text{Conv} \left( \bigcup_{\pi} R_{\pi,tot}^{1-DPCRS} \right)$.

3) M-DPCRS: Following the approximation method adopted in DPC and 1-DPCRS, the system WESR of M-DPCRS is approximated by $\text{WESR}_{\pi,M-DPCRS} = \frac{1}{B} \sum_{b=1}^B \bar{R}_{\pi,tot}^{M-DPCRS}(\hat{H}[b])$, where $\bar{R}_{\pi,tot}^{M-DPCRS}(\hat{H}[b]) \triangleq \sum_{k \in K} u_{\pi(k)} \bar{R}_{\pi(k),tot}^{M-DPCRS}(\hat{H}[b])$ is the WASR in block-$[b]$. The total AR of user-$\pi(k)$ is $\bar{R}_{\pi(k),tot}^{M-DPCRS}(\hat{H}[b]) = \sum_{i \in I_{\pi(k)} \setminus \{ \pi(k) \}} \bar{C}_i^{[b]} + \bar{R}_{\pi(k)}^{M-DPCRS}(\hat{H}[b])$, $\bar{C}_i^{[b]}$ is the AR of $s_i$ allocated to user-$\pi(k)$. The rate allocation in block-$[b]$ is specified as $\sum_{k \in K_i} \bar{C}_k^{[b]} = \bar{R}_i^{M-DPCRS}(\hat{H}[b])$, where $\bar{R}_i^{M-DPCRS}(\hat{H}[b]) = \min\{ \bar{R}_{i,k}^{M-DPCRS}(\hat{H}[b]) \mid k \in K_i \}$ is the AR of the common stream $s_i$. 
By omitting $[b]$, we could also obtain the decomposed WASR maximization problem to be solved in each block for the three-user M-DPCRS with a given weight vector $u$ and a fixed encoding order $\pi$, which is given by

$$\max_{\hat{e}, P, \pi'} \sum_{k \in K} u_{\pi(k)} \overline{R}^{M-DPCRS}_{\pi(k), tot} (\hat{H})$$ \hspace{1cm} (22a)

subject to

$$\sum_{k \in K_i} C^i_k \leq R^{M-DPCRS}_{i} (\hat{H}), \forall i \in K_c$$ \hspace{1cm} (22b)

$$\overline{R}^{M-DPCRS}_{\pi(k), tot} (\hat{H}) \geq R^{th}_{\pi(k)}, \forall k \in K$$ \hspace{1cm} (22c)

$$\text{tr}(PP^H) \leq P_t$$ \hspace{1cm} (22d)

$$\overline{c} \geq 0. \hspace{1cm} (22e)$$

where $\hat{e} = \{C^i_k | k \in K_i, i \in K_c\}$ is the AR allocation for all the common streams in each block. Notice that the decoding order $\pi'$ of the common streams for the partial-common streams $s_{12}, s_{13}, s_{23}$ is required to be jointly optimized with precoders in each block so as to maximize the system WESR for a given weight vector $u$ and a fixed DPC encoding order $\pi$. To further maximize the WESR for a given set of user weights $u$, an extra optimization over the DPC encoding order $\pi$ has to be carried out. This can be done by evaluating the performance for all possible encoding orders and choosing the one with the highest WESR.

### B. MISO BC with a Multicast Message

When considering MISO BC with a multicast message, we jointly design precoders and message splits with the objective of maximizing the WESR of the unicast messages while the QoS rate constraints of multicast and unicast messages as well as the power constraint at the BS should be met. Following the method described for 1-DPCRS in MISO BC, we obtain the decomposed subproblem to be solved in each block (with $[b]$ omitted) by decomposing the WESR maximization problem of 1-DPCRS for MISO BC with a multicast message into $B$ WASR maximization problems. For a given weight vector $u$ and a fixed DPC encoding order $\pi$, the WASR maximization problem in each block is

$$\max_{\hat{e}, P} \sum_{k \in K} u_{\pi(k)} \overline{R}^{1-DPCRS}_{\pi(k), tot} (\hat{H})$$ \hspace{1cm} (23a)

subject to

$$\overline{C}_0 + \sum_{k \in K} \overline{C}_k \leq \overline{R}^{1-DPCRS}_{0} (\hat{H})$$ \hspace{1cm} (23b)

$$\overline{C}_0 \geq R^{th}_0$$ \hspace{1cm} (23c)

$$\overline{C}_{\pi(k)} + \overline{R}^{1-DPCRS}_{\pi(k)} (\hat{H}) \geq R^{th}_{\pi(k)}, \forall k \in K$$ \hspace{1cm} (23d)

$$\text{tr}(PP^H) \leq P_t$$ \hspace{1cm} (23e)

$$\overline{c} \geq 0. \hspace{1cm} (23f)$$
The rate vector \( \bar{c} \) for 1-DPCRS in MISO BC with a multicast message contains the rate allocated to the multicast stream \( s_0 \) as well as the common part of the unicast streams, i.e., \( \bar{c} = [\bar{C}_0, \bar{C}_1, \ldots, \bar{C}_K] \). The AR of the super-common stream and the private streams are defined in the same way as in 1-DPCRS for MISO BC. Compared with Problem (21), the main difference of Problem (23) comes from constraint (23b) and (23c) due to the additional multicast message \( W_0 \) to be transmitted for all users. \( R_{th}^{th} \) is the QoS rate constraint of \( W_0 \). The WASR problem of using M-DPCRS for MISO BC with a multicast message can also be formulated if readers understand (22) and (23), which will not be specified here due to page limitation.

### IV. PROPOSED OPTIMIZATION FRAMEWORK

The formulated problem (19), (21), (22) and (23) are stochastic optimization problems since the ARs specified in Definition 1 are expectations with respect to the random variable \( \tilde{H} \). To tackle with the stochastic nature of the two problems, we employ the Sample Average Approximation (SAA) approach proposed in [7] to respectively transform the original stochastic problems into the deterministic problems, which are then solved using the Weighted Minimum Mean Square Error (WMMSE) method. In this section, the SAA and WMMSE based optimization framework of solving M-DPCRS problem (22) are specified followed by the guidance of solving other problems.

The first step is to use SAA to approximate the stochastic ARs into the corresponding deterministic expressions. As the conditional density \( f_{H|\hat{H}}(H | \hat{H}) \) is known at the BS, for a given channel estimate \( \hat{H} \), BS is able to generate a sample of \( M \) user channels, indexed by \( \mathcal{M} = \{1, \ldots, M\} \) as

\[
\mathcal{H}^{(M)} \triangleq \left\{ H^{(m)} = \hat{H} + \tilde{H}^{(m)} \mid \hat{H}, m \in \mathcal{M} \right\},
\]

Following the strong Law of Large Number (LLN), the ARs \( \bar{R}_{x}^{i,k}(\hat{H}) \) specified in equation (15) for decoding the stream \( s_i \) at user-\( k, k \in \mathcal{K} \) with a given channel estimate \( \hat{H} \) is approximated as

\[
\bar{R}_{x}^{i,k}(\hat{H}) \approx \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} R_{x}^{i,k}(H^{(m)}, \hat{H}),
\]

where \( x \in \{ \text{"DPC","1-DPCRS","M-DPCRS"} \} \). \( \bar{R}_{x}^{i,k}(\hat{H}) = \bar{R}_{x}^{k}(\hat{H}) \) when \( i = k \). Notice that the precoder \( P \) in (25) is unaltered over all the \( M \) channel samples. Considering M-DPCRS strategy for MISO BC, \( P \) is designed by solving problem (22) with the average common and private rates approximated by equation (25) for all channel samples, which is given by
where \( \hat{\mathbf{C}}_{\pi}(k) \approx \mathbb{E}\{\mathbf{x} \mathbf{H}^{-1}(\hat{\mathbf{H}})\} \).

Hence, our target is to design precoder \( \mathbf{P} \) and the common stream allocation vector \( \bar{c} \) by solving (26).

Problem (26) is still non-convex due to the non-convex approximated rate expressions of the common stream and the private streams. To solve the problem, we further extend the WMMSE algorithm proposed in [7], [43]. At user sides, user-\( \pi(k) \) decodes data streams \( \{s_i|i \in \mathcal{I}_{\pi}(k)\} \) sequentially based on the decoding order \( \pi' \) by employing the equalizer \( g_{\pi}(k), i \in \mathcal{I}_{\pi}(k) \). The signal received at user-\( \pi(k) \) is 
\[
y_{\pi}(k) = \mathbf{H}_{\pi}(k) \mathbf{x} + n_{\pi}(k),\]
where \( \mathbf{x} \) is specified in equation (11). After decoded first and the estimated common stream \( \hat{s}_{123} \) is \( \hat{s}_{123} = g^{123}_{\pi}(k)y_{\pi}(k) \). Once \( s_{123} \) is successfully decoded and removed from the received signal, user-\( \pi(k) \) then decodes the partial-common streams \( s_{\pi'(k)(1)}, s_{\pi'(k)(2)} \) by employing the equalizer \( g^{i}_{\pi}(k), i \in \{\pi'(k)(1), \pi'(k)(2)\} \) followed by the private stream \( s_{\pi(k)} \) via equalizer \( g^{\pi(k)}_{\pi(k)} \). The estimated streams \( \hat{s}_{\pi'(k)(1)}, \hat{s}_{\pi'(k)(2)} \) and \( \hat{s}_{\pi(k)} \) at user-\( \pi(k) \) are
\[
\hat{s}_{\pi'(k)(1)} = g^{\pi(k)}_{\pi'(k)(1)}(y_{\pi}(k) - \mathbf{H}_{\pi}(k)\mathbf{p}_{123}) \hat{s}_{123}, \quad \hat{s}_{\pi'(k)(2)} = g^{\pi(k)}_{\pi'(k)(2)}(y_{\pi}(k) - \sum_{i \in \{123, \pi'(k)(1)\}} \mathbf{H}_{\pi}(k)\mathbf{p}_{i} \hat{s}_{i}), \quad \text{and} \quad \hat{s}_{\pi(k)} = g^{\pi(k)}_{\pi(k)}(y_{\pi}(k) - \sum_{i \in \{123, \pi'(k)(1), \pi'(k)(2)\}} \mathbf{H}_{\pi}(k)\mathbf{p}_{i} \hat{s}_{i}),
\]
respectively. The Mean Square Error (MSE) of each stream \( s_i, i \in \{123, \pi'(k)(1), \pi'(k)(2), \pi(k)\} \) at user-\( \pi(k) \) is
\[
\varepsilon^i_{\pi(k)} = \mathbb{E}\{|\hat{s}_i - s_i|^2\} = |g^{i}_{\pi}(k)|^2 T^i_{\pi}(k) - 2\Re \{g^{i}_{\pi}(k)\mathbf{H}^{H}_{\pi}(k)\mathbf{p}_i\} + 1,
\] where \( T^i_{\pi}(k) = I^i_{\pi}(k) + |\mathbf{H}^{H}_{\pi}(k)\mathbf{p}_i|^2 \).

Define the Weighted MSE (WMSE) of decoding \( s_i \) at user-\( \pi(k) \) as
\[
\xi^i_{\pi(k)}(\mathbf{H}, \hat{\mathbf{H}}) = w^i_{\pi(k)} \varepsilon^i_{\pi(k)} - \log_2(w^i_{\pi(k)}),
\] where \( w^i_{\pi(k)} \) is the introduced weight for MSE of user-\( \pi(k) \). The corresponding Weighted Minimum MSE (WMMSE) metrics of the common and private streams are
\[
\xi^{\text{WMSE}}_{\text{c}, \pi(k)}(\mathbf{H}, \hat{\mathbf{H}}) \triangleq \min_{w^i_{\pi(k)} \geq 0} \xi^i_{\pi(k)}(\mathbf{H}, \hat{\mathbf{H}}),
\]
with the introduced WMMSEs, we obtain the following proposition.
Proposition 1. The instantaneous rate and the WMMSE of decoding stream $s_i$ at user-$\pi(k)$, $k \in \mathcal{K}$ has the following relationship:

$$\xi_{i,\pi(k)}^M(\hat{H}, \hat{H}) = 1 - R^M_{i,\pi(k)}(\hat{H}, \hat{H}),$$  \hspace{1cm} (30)

where $R^M_{i,\pi(k)}(\hat{H}, \hat{H}) = R^M_{\pi(k)}(\hat{H}, \hat{H})$ when $i = \pi(k)$.

Proof: Following (29), the optimal WMSE weights $(w_{\pi(k)}^{\pi(k)})$ and equalizers $(g_{\pi(k)}^{\pi(k)})$ of minimizing $\xi_{\pi(k)}(\hat{H}, \hat{H})$ satisfy that

$$\frac{\partial \xi_{\pi(k)}(\hat{H}, \hat{H})}{\partial w_{\pi(k)}^\pi} = 0$$

and

$$\frac{\partial \xi_{\pi(k)}(\hat{H}, \hat{H})}{\partial g_{\pi(k)}^\pi} = 0.$$  \hspace{1cm} (31)

We first solve $\frac{\partial \xi_{\pi(k)}(\hat{H}, \hat{H})}{\partial w_{\pi(k)}^\pi} = 0$ and obtain the optimal WMSE equalizer as

$$g_{\pi(k)}^{\pi(k)} = g_{\pi(k)}^{\text{MMSE}} = p_H^i h_{\pi(k)}(T_{\pi(k)}^i)^{-1}.$$  \hspace{1cm} (32)

By further solving $\frac{\partial \xi_{\pi(k)}(\hat{H}, \hat{H})}{\partial w_{\pi(k)}^\pi} = 0$, we obtain that

$$w_{\pi(k)}^{\pi(k)} = w_{\pi(k)}^{\text{MMSE}} = \frac{1}{\xi_{\pi(k)}(g_{\pi(k)}^{\pi(k)})} = \frac{T_{\pi(k)}^i}{T_{\pi(k)}^i}.$$  \hspace{1cm} (33)

Substituting $(w_{\pi(k)}^{\pi(k)}, g_{\pi(k)}^{\pi(k)})$ back to $\xi_{\pi(k)}(\hat{H}, \hat{H})$, $\xi_{i,\pi(k)}^M(\hat{H}, \hat{H})$ is derived as

$$\xi_{i,\pi(k)}^M(\hat{H}, \hat{H}) = \log_2(w_{\pi(k)}^{\pi(k)}) = 1 - R^M_{i,\pi(k)}(\hat{H}, \hat{H}).$$  \hspace{1cm} (34)

The proof is completed. \hfill \Box

The Rate-WMMSE relationships in (30) is established for instantaneous channel realizations. We can also extend it to the average Rate-WMMSE relationships over a sample of $M$ user channels as

$$\bar{\xi}_{i,\pi(k)}^M(\hat{H}, \hat{H}) \triangleq \frac{1}{M} \sum_{m=1}^M \left( \min_{(u_{\pi(k)})^{(m)}, g_{\pi(k)}^{(m)}} \xi_{\pi(k)}^i(H^{(m)}), \hat{H} \right) = 1 - R^M_{i,\pi(k)}(\hat{H}, \hat{H}),$$  \hspace{1cm} (35a)

where $u_{\pi(k)}^{(m)}, g_{\pi(k)}^{(m)}$ are the weights and equalizers associated with the $m$th channel realization in $\mathbb{E}_i[H^{(M)}], \xi_{i,\pi(k)}^M(\hat{H}, \hat{H}) = \xi_{\pi(k)}^i(\hat{H}, \hat{H})$ when $i = \pi(k)$. With the average Rate-WMMSE relationships in (34), Problem (26) is equivalently transformed into the WMMSE problem

$$\min_{P, x, w, g} \sum_{k \in \mathcal{K}} u_{\pi(k)} \left( \sum_{i \in \mathcal{I}_{\pi(k)} \setminus \{\pi(k)\}} \bar{X}_k^i + \sum_{k \in \mathcal{K}} \bar{X}_{\pi(k)}^i \right)$$

s.t. \hspace{1cm} (35a)

$$\sum_{k \in \mathcal{K}} \bar{X}_k^i + 1 \geq \bar{\xi}_{\pi(k)}^i(\hat{H}), \forall i \in \mathcal{K}$$

$$\sum_{i \in \mathcal{I}_{\pi(k)} \setminus \{\pi(k)\}} \bar{X}_k^i + \bar{\xi}_{\pi(k)}^i(\hat{H}) \leq 1 - R^{ih}_{\pi(k)}, \forall k \in \mathcal{K}$$

$$\text{tr}(P^H) \leq P_t$$

$$\bar{x} \leq 0,$$  \hspace{1cm} (35b)

(35c)

(35d)

(35e)
where \( \tilde{x} = \{ \tilde{X}_k \mid k \in \mathcal{K}_i, i \in \mathcal{K}_c \} \) is the transformation of the common rate \( \bar{c} \) with \( \tilde{x} = -\bar{c} \) holds.

\( w = \{ u_{\pi(k)}^{i(m)} \mid k \in \mathcal{K}_i, i \in \mathcal{K}_c, m \in \mathcal{M} \} \) and \( g = \{ g_{\pi(k)}^{i(m)} \mid k \in \mathcal{K}_i, i \in \mathcal{K}_c, m \in \mathcal{M} \} \) are the MSE weights and equalizers, respectively. 

\( \xi^{(M)}(\tilde{H}) = \max \{ \xi^{(M)}_{\pi(k)}(\tilde{H}) \mid \pi(k) \in \mathcal{K}_i \} \).

Though problem (35) that jointly optimizes \((P, \bar{x}, w, g)\) is still non-convex, it is block-wise convex with respect to each block of \(w, g\) and \((P, \bar{x})\) by fixing other two blocks. This motivates us to use Alternating Optimization (AO) algorithm to solve the problem. At each iteration \([n]\), for given \(w^{[n-1]}\) and \((P^{[n-1]}, \bar{x}^{[n-1]}\)), the optimal solution \(g^{[n]}\) of (35) is \(g^{[n]} \triangleq g^{\text{MMSE}}(P^{[n-1]}) = \{ g_{\pi(k)}^{i(m)} \mid k \in \mathcal{K}_i, i \in \mathcal{K}_c, m \in \mathcal{M} \} \) with each element calculated by equation (31) and precoder \(P^{[n-1]}\) for the \(m\)th channel realization in \(\mathbb{H}([M])\). For given \(g^{[n-1]}\) and \((P^{[n-1]}, \bar{x}^{[n-1]}\)), the optimal solution \(w^{[n]}\) of (35) is \(w^{\text{MMSE}}\), where \(w^{[n]} \triangleq w^{\text{MMSE}}(P^{[n-1]}) = \{ w_{\pi(k)}^{i,\text{MMSE},(m)} \mid k \in \mathcal{K}_i, i \in \mathcal{K}_c, m \in \mathcal{M} \} \) with each element calculated by equation (32) and precoder \(P^{[n-1]}\). The optimal solutions of weights and equalizers can be verified through showing that \((w^{[n]}, g^{[n]}\) satisfy the Karush-Kuhn-Tucker (KKT) conditions of (35).

Substituting \((w^{[n]}, g^{[n]}\) back to (35), the optimization problem is equivalently transformed as:

\[
\min_{P, \bar{x}} \sum_{k \in \mathcal{K}} u_{\pi(k)} \left( \sum_{i \in \mathcal{I}_{\pi(k)}(\pi(k))} \bar{X}^i_{\pi(k)} + \sum_{k \in \mathcal{K}} \xi^{\text{M-DPCS}}_{\pi(k)}(\tilde{H}) \right) \tag{36a}
\]

subject to:

\[
\sum_{k \in \mathcal{K}} \bar{X}^i_k + 1 \geq \xi^{\text{M-DPCS}}_{\pi(k)}(\tilde{H}), \forall i \in \mathcal{K}_c \tag{36b}
\]

\[
\sum_{i \in \mathcal{I}_{\pi(k)}(\pi(k))} \bar{X}^i_k + \xi^{\text{M-DPCS}}_{\pi(k)}(\tilde{H}) \leq 1 - R^{th}_{\pi(k)}, \forall k \in \mathcal{K} \tag{36c}
\]

where \(\xi^{\text{M-DPCS}}_{\pi(k)}(\tilde{H}) \triangleq \Omega^{i}_{\pi(k)} + \bar{v}^{i}_{\pi(k)} - 2 \Re \{ (P^i_{\pi(k)})^H P_i \} + \bar{w}^{i}_{\pi(k)} - \bar{v}^{i}_{\pi(k)}\) and \(\xi^{\text{M-DPCS}}_{\pi(k)}(\tilde{H}) = \xi^{\text{M-DPCS}}_{\pi(k)}(\tilde{H}) \) when \(i = \pi(k)\), \(\Omega^{123}_{\pi(k)} \triangleq \sum_{j \in \mathcal{K}, \cup j \cup k} P^j H \tilde{\Psi}^j_{\pi(k)} P_j, \Omega^{n^2}_{\pi(k)} \triangleq \sum_{j \in \mathcal{K}, \cup j \cup k} P^j H \tilde{\Psi}^j_{\pi(k)} P_j, \Omega^{\pi(k)} \triangleq \sum_{j \in \mathcal{K}_c, \cup j \cup k} P^j H \tilde{\Psi}^j_{\pi(k)} P_j + \sum_{j < k} P^j H \tilde{\Phi}^j_{\pi(k)} P_{\pi(j)} + \sum_{j \geq k} P^j H \tilde{\Psi}^j_{\pi(k)} P_{\pi(j)}\), \(\bar{w}^{i}_{\pi(k)}\), \(\bar{v}^{i}_{\pi(k)}\) are constants (or constant vectors/matrices) averaged over a sample of \(M\) user channels, i.e., \(\bar{w}^{i}_{\pi(k)} = \frac{1}{M} \sum_{m=1}^{M} w_{\pi(k)}^{i(m)}\).

Their corresponding values in each channel realization \((m)\) are updated as:

\[
t^{i(m)}_{\pi(k)} = u^{i(m)}_{\pi(k)} \left[ g^{i(m)}_{\pi(k)} \right]_2^2, \\
\tilde{\Psi}^{i(m)}_{\pi(k)} = t^{i(m)}_{\pi(k)} \left( h^{(m)}_{\pi(k)} \right)^H, \\
\tilde{\Phi}^{i(m)}_{\pi(k)} = t^{i(m)}_{\pi(k)} \left( h^{(m)}_{\pi(k)} \right)^H, \\
f^{i(m)}_{\pi(k)} = u^{i(m)}_{\pi(k)} \left[ g^{i(m)}_{\pi(k)} \right]_2^2, \\
v^{i(m)}_{\pi(k)} = \log_2 \left( u^{i(m)}_{\pi(k)} \right). \tag{37}
\]
Algorithm 1: WMMSE-based AO algorithm

1. Initialize: $n \leftarrow 0$, $P$, $WSR^{[n]}$;
2. repeat
   3. $n \leftarrow n + 1$;
   4. $P^{[n-1]} \leftarrow P$;
   5. $w \leftarrow w^{\text{MMSE}}(P^{[n-1]})$; $g \leftarrow g^{\text{MMSE}}(P^{[n-1]})$;
   6. update $(P, \bar{x})$ by solving (36) using the updated $w, g$;
3. until $|WSR^{[n]} - WSR^{[n-1]}| \leq \epsilon$;

Problem (36) is a convex Quadratically Constrained Quadratic Program (QCQP), which can be solved via interior-point methods. Therefore, $(P^{[n]}, \bar{x}^{[n]})$ can be updated by using the optimal solution of (36). The details to the proposed AO algorithm is specified in Algorithm 1. The weights $w$, equalizers $g$, precoders and common rate vectors $(P, \bar{x})$ are updated iteratively until the WSR of the system $WSR^{[n]}$ calculated by $(P, \bar{x})$ at the end of iteration $[n]$ converges.

**Proposition 2.** Denote any stationary point of Problem (22) as $(P^\circ, \bar{c}^\circ)$. For a feasible initial point, the proposed AO algorithm is guaranteed to converge and the convergent solution $(P', \bar{x}')$ is a stationary point of Problem (22) with $P' = P^\circ$ and $\bar{x}' = -\bar{c}^\circ$ holds.

**Proof:** The proof in [7] for the AO algorithm of linearly precoded 1-layer RS is extended here for that of M-DPCRS. We first show that the proposed algorithm is guaranteed to converge. As the optimal solution $P^{[n]}, \bar{x}^{[n]}, w^{[n]}, g^{[n]}$ of Problem (36) at iteration $[n]$ is also a feasible solution of (36) at iteration $[n + 1]$, the corresponding objective function of Problem (36) is decreasing monotonically. For a given solution $(P^{[n]}, \bar{x}^{[n]}, w^{[n]}, g^{[n]})$, Problem (36) can be equivalently transformed into (22) with solution $P^{[n]}$ and $\bar{c}^{[n]} = -\bar{x}^{[n]}$. And WSR$^{[n]}$ is increasing as the objective function of (36) decreases iteratively. Due to the transmit power constraint (22d), Problem (36) is bounded below, and therefore, WSR$^{[n]}$ is bounded above. Hence, the convergence is guaranteed. Next, we show that the solution $(P^{[n]}, \bar{c}^{[n]} = -\bar{x}^{[n]})$ converges to the stationary points of (22). As the AO algorithm is a special instance of the Successive Convex Approximation (SCA) method and SCA method maintains the KKT conditions of the original problem, the AO algorithm $(P^{[n]}, \bar{c}^{[n]} = -\bar{x}^{[n]})$ therefore meets the KKT conditions of (22). Hence, the proposed AO algorithm is guaranteed to converge to one stationary point of (22).

Similarly, problem (19), (21) and (23) as well as the problem of M-DPCRS for MISO BC with a multicast message are solved respectively by approximating each stochastic optimization problem using the SAA approach. The equivalently transformed problems are then reformulated into the WMMSE problems and solved by the corresponding AO algorithm.
V. Numerical Results

In this section, the WSR performance of the proposed 1-DPCRS and M-DPCRS strategies in both MISO BC and MISO BC with a multicast message transmission networks are evaluated. In the following numerical results, all the optimization problems to be solved by using interior-point methods are solved using the CVX toolbox [44]. User channels are randomly generated as specified in [7], [13], [25]. The actual user channel \( h_k \) experienced at user-\( k \) has i.i.d. complex Gaussian entries drawn from the distribution \( \mathcal{CN}(0, \sigma_k^2) \) and the channel estimation error \( \tilde{h}_k \) has i.i.d. complex Gaussian entries drawn from distribution \( \mathcal{CN}(0, \sigma_{e,k}^2) \). The variance of \( \tilde{h}_k \) is defined as \( \sigma_{e,k}^2 \triangleq \sigma_k^2 P_t - \alpha \). As user channels with heterogeneous variances are considered, the corresponding CSIT qualities also scale with the channel variance \( \sigma_k^2 \). \( \alpha \in [0, 1] \) represents SNR scaling as described in Section II-A. We obtain that \( \hat{h}_k = h_k - \tilde{h}_k \) also follows Gaussian distribution \( \mathcal{CN}(0, 1 - \sigma_{e,k}^2) \). The WESR is obtained by averaging the WASR over 100 channel realizations, i.e., there are \( B = 100 \) channel blocks. Within each block, the AR of each user is approximated using SAA method over \( M = 1000 \) samples of user channels \( \mathbb{H}^{(M)} \). For a given \( \hat{H} \), the \( m \)th channel estimation error \( \hat{H}^{(m)} \) is randomly generated from the error distribution. Hence, the \( m \)th conditional channel is calculated by \( \mathbb{H}^{(m)} = \hat{H} + \hat{H}^{(m)} \). The initialization of the precoders \( P \) for Algorithm 1 is designed by the Maximum Ratio Transmission (MRT) and Singular Value Decomposition (SVD) method proposed in [7]. The precoders for the private streams of RS-assisted or other non-RS-assisted transmission strategies are initialized by MRT, i.e., \( p_k = \sqrt{p_k \| h_k \|} \). The precoders for the common streams of RS-assisted strategies are initialized by SVD, i.e., for 1-DPCRS and 1-layer RS, \( p_c = \sqrt{p_c \| \hat{p}_c \|} \), where \( \hat{p}_c \) is the largest left singular vector of the channel estimate \( \hat{H} \). \( p_c \) and \( p_k \) are the power allocated to each precoder, it follows that \( p_c + \sum_{k \in \mathcal{K}} p_k = P_t \). The precoders of the common streams \( s_i, i \in \mathcal{K}_c \) for M-DPCRS and generalized RS are initialized in the same way as \( p_c \) but \( \hat{p}_i \) is chosen based on \( \hat{H}_i \) formed by the channel estimate of users in \( \mathcal{K}_i \).

A. MISO BC

In MISO BC, the following eight transmission strategies are compared:

- **M-DPCRS**: the M-DPCRS strategy proposed in Section II-C. In the \( K \)-user case, there are \( 2^K - 1 \) linearly precoded common streams and \( K \) DPC-coded private streams to be transmitted from the BS.
- **1-DPCRS**: the 1-DPCRS strategy proposed in Section II-C. One linearly precoded common stream and \( K \) DPC-coded private streams are transmitted in the \( K \)-user case.
- **DPC**: the conventional DPC strategy specified in Section II-B. There are \( K \) DPC-coded data streams to be transmitted in the \( K \)-user case.
• **generalized RS**: the multi-layer RS strategy proposed in [12]. User messages are split in the same way as M-DPCRS discussed in Section II-C. The main difference compared with M-DPCRS is the private streams of the generalized RS are linearly precoded. In the $K$-user case, there are $2^K - 1$ linearly precoded common streams and $K$ linearly precoded private streams to be transmitted from the BS.

• **1-layer RS**: the 1-layer RS strategy specified in [3], [7], [12], [45]. Each user message is split into a common part and a private part. There is one linearly precoded common stream and $K$ linearly precoded private streams to be transmitted jointly from the BS. Each user is required to decode the common stream first and uses one layer of SIC to remove the common stream before decoding the intended private stream.

• **SC–SIC**: the power-domain NOMA widely studied in the literature [46]. In the $K$-user case, the streams are linearly precoded and superimposed at the BS before transmission. Users are ordered based on their effective scalar channel strength after precoding. Each user is required to decode and remove the interference from users with weaker effective channel strength sequentially using SIC.

• **SC–SIC per group**: the method of combining SDMA and NOMA in MIMO transmission networks [47]. The $K$ users are clustered into multiple groups. The inner-group interference is coordinated by SC–SIC while the inter-group interference is coordinated by SDMA. At the BS, the $K$-user messages are linearly precoded. Users within the same group are ordered by the corresponding effective channel strength such that each user is able to sequentially decode and remove the interference from weaker users within the same group. The interference from users in different groups is fully treated as noise at each user.

• **MU–LP**: MU–LP is a practical transmission strategy that has been widely studied in MIMO networks and it is the common implementation of SDMA. User messages are linearly precoded and superimposed at the BS and each user directly decodes its intended data stream by fully treating any residual interference as noise.

 Readers are referred to [12] for more details of “generalized RS”, “1-layer RS”, “SC–SIC”, “SC–SIC per group” and “MU–LP” transmission strategies, where the corresponding WSR maximization problems are studied. In the sequel, we evaluate the WESR performance of all the eight strategies in a wide range of user deployments considering a diverse range of CSIT qualities, QoS rate requirements and channel strength disparities among users. In the following, we first illustrate the results in the two-user MISO BC followed by the three-user case.

---

$^2$As we use random channel realizations, the channel strength disparities are manifested by tuning the channel variance $\sigma_k^2$ of each user. It is termed “channel variance disparities” in the following.
Fig. 5: Sum rate versus SNR comparison of different strategies with different partial CSIT inaccuracies, averaged over 100 random channel realizations, $K = 2$, $N_t = 4$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$.

Fig. 6: Sum rate versus SNR comparison of different strategies with different partial CSIT inaccuracies, averaged over 100 random channel realizations, $K = 2$, $N_t = 4$, $\sigma_1^2 = 1$, $\sigma_2^2 = 0.09$.

1) Two-user case: When $K = 2$, M-DPCRS, generalized RS and SC–SIC per group respectively reduces to 1-DPCRS, 1-layer RS and SC–SIC. We use the term “DPCRS” to represent both M-DPCRS and 1-DPCRS and the term “RS” to represent both generalized RS and 1-layer RS in the two-user case. We first illustrate the system Ergodic Sum Rate (ESR) ($u_1 = 1$, $u_2 = 1$) versus SNR comparison of different strategies considering diverse CSIT inaccuracies and channel strength disparities in Fig. 5 and Fig. 6. The individual QoS rate constraint of each user is set to 0, i.e., $R_{th}^k = 0, \forall k \in \{1, 2\}$. Without QoS rate constraint and unequal user weights, the WESR problem for 1-layer RS reduces to the ESR maximization problem studied in [7]. It has been discovered in [7] the sum DoF achieved by solving the ESR problem of 1-layer RS is

$$d^*_RS = 1 + (K - 1)\alpha, \quad (38)$$

and $d^*_RS$ has been shown to be the sum DoF limit that could be achieved in MISO BC with partial CSIT. In comparison, the DoF achieved by optimally solving the ESR problem of MU–LP is $d^*_MULP = \max\{1, K\alpha\}$. As DoF captures the rate’s asymptotic slope with respect to $\log_2(P_t)$, it is easy to identify the DoF according to the slope of the sum rate at high SNR from both Fig.
Fig. 7: Ergodic rate region comparison of different strategies with partial CSIT, averaged over 100 random channel realizations, SNR = 20 dB, $K = 2$, $\alpha = 0.6$, $\sigma_1^2 = 1$.

The values are calculated by scaling the high-SNR slopes to $\log_2(P_t)$. The DoF achieved by RS and DPCRS from both figures are close to the theoretically anticipated values calculated by [28], i.e., $d_{RS}^* = 1.3, 1.6, 1.9$ when $\alpha = 0.3, 0.6, 0.9$, respectively. By using the same method, we obtain the DoF of MU–LP and DPC, which also coincides with the theoretical DoF results, i.e.: $d_{MU-LP}^* = 1, 1.2, 1.8$ when $\alpha = 0.3, 0.6, 0.9$, respectively. As the DoF of RS is optimal, the DoF of DPCRS is the same as RS. Though DPC is capacity-achieving in perfect CSIT, it is very sensitive to the CSIT inaccuracy. It is DoF is the same as MU–LP. As $\alpha$ decreases, the rate of DPC drops rapidly as MU–LP. In contrast, RS-assisted transmission strategies are more robust to the CSIT inaccuracy. Linearly precoded RS is not only more practical and low complex compared with DPC but also achieves non-negligible rate gain over DPC when CSIT is imperfect. The rate gain of DPCRS and RS over all other strategies grows with SNR. Specifically, both DPCRS and RS achieves 16.12% rate improvement over DPC and MU–LP when $\alpha = 0.3$, SNR is 30 dB and users have equal channel variance as illustrated in Fig. 5. When there is a 10 dB average channel variance disparities as in Fig. 6 the rate gain of RS decreases since there is a higher probability that the transmission reduces to single-user transmission by switching off the weaker user. In both figures, SC–SIC has the worst performance due to the fact that it does not exploit efficiently the spatial domain for interference management. The DoF of SC–SIC is the worst and is limited to 1 (same as OMA) [28].

We further investigate the ER region achieved by the users in the two-user case. The boundary of the rate region is a set of points obtained by solving the WESR problem with different pairs of weights assigned to users. For each pair of weights, the problem is solved over 100 channel realizations and each boundary point is obtained by averaging the WASR over all channel realizations. Following [7], the weight of user-1 is fixed as $u_1 = 1$ for each pair of user weights while the weight of user-2 changes as $u_2 \in 10^{-3}, -1, -0.95, ..., 0.95, 1.3]$. As we study the largest achievable rate region comparison, the QoS rate constraint of each user is set to 0. Fig. 7 illustrates the rate region comparison of different strategies considering different number of
transmit antennas and channel strength disparities. SNR is equal to 20 dB. In all subfigures, DPCRS achieves the largest rate region among all strategies. Comparing subfigure (a) and (c) in Fig. 7, we observe that SC–SIC achieves the worst rate region. As there is no channel variance disparities among users, SC–SIC cannot properly utilize the power domain to manage interference. As the number of transmit antennas decreases (from subfigure (a)/(b) to (c)/(d)), the rate region gap between DPCRS and DPC/MU–LP becomes larger. Each user in DPC and MU–LP directly decodes its intended stream, and the pressure of interference management is at the transmitter. In comparison, both DPCRS and RS utilize the common stream to enable each user the capability of partially decoding the interference and partially treating the interference as noise. Both of them are more robust to various number of transmit antennas and user deployments.

2) Three-user case: When $K = 3$, the ESR of all the eight strategies versus CSIT inaccuracy are compared in Fig. 8 with different QoS rate constraints, network loads and user deployments. SNR is 20 dB. In subfigure (a), the individual QoS rate constraint increases with CSIT accuracy. For $\alpha = [0.2, 0.4, 0.6, 0.8, 1]$, the corresponding rate constraint for user-$k$ ($k \in \{1, 2, 3\}$) changes as $r_k^{th} = [0.1, 0.2, 0.3, 0.4, 0.5]$ bit/s/Hz. In all subfigures, the ESR of DPC, MU–LP, and SC–SIC-assisted strategies decrease dramatically as $\alpha$ decreases from 1 to 0.2 due to the drop-off of CSIT accuracy. The ESR gap between M-DPCRS/1-DPCRS and DPC is more obvious in the region with strong CSIT inaccuracy. Thanks to their ability to partially decode interference and partially treat interference as noise, all RS-assisted transmission strategies are more robust to the CSIT inaccuracy. In Fig. 8(a) and Fig. 8(b) with underloaded network loads, generalized RS and 1-layer RS, using linear precoder for all streams, achieve explicit ESR improvement over DPC when $\alpha$ is less than 0.6. This observation further confirms the powerful interference management capability of RS in the multi-antenna BC. In Fig. 8(c) and Fig. 8(d) where network loads are overloaded and users suffer from stronger inter-user interference, we observe that M-DPCRS (generalized RS) achieves higher rate than 1-DPCRS (1-layer RS). By increasing the number of layers of common streams in RS, inter-user interference is better managed and ESR is further improved even though there is no DoF increase.

(a) $r_k^{th} = [0.1, \ldots, 0.5]$ bit/s/Hz, (b) $r_k^{th} = 0$, $N_t = 4$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$. (c) $r_k^{th} = 0$, $N_t = 2$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$. (d) $r_k^{th} = 0$, $N_t = 2$, $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_3^2 = 0.3$.

Fig. 8: Ergodic sum rate versus CSIT inaccuracy $\alpha$ comparison of different strategies, averaged over 100 random channel realizations, $K = 3$, SNR = 20 dB.
averaged over 100 random channel realizations, $K\alpha R$ user ER region comparison of all strategies. The QoS rate constraint of the multicast stream is

strategies proposed in [25] for MISO BC with a multicast message. Fig. 9 illustrates the two-layer RS”, “SC–SIC”, “SC–SIC per group” and “MU–LP” mentioned in this subsection are the encoded multicast stream on top of the DPC encoded unicast streams. “Generalized RS”, “1-layer RS”, “1-RSDPC” and “M-RSDPC” are the strategies in MISO BC is there is one additional multicast message to be transmitted together with the transmission of the unicast streams since additional power can be allocated to the common stream for dynamic interference management. DPCRS maintains the largest rate region compared with the rate regions of all existing strategies. In the three-user case, we study the ESR comparison versus CSIT accuracy in Fig. 10. The QoS rate constraint of the multicast stream is

B. MISO BC with a multicast message

In the MISO BC with a multicast message transmission networks, we also study the aforementioned eight transmission strategies. The main difference compared with the strategies studied in MISO BC is there is one additional multicast message to be transmitted together with the unicast messages for all users. In this subsection, “M-DPCRS” and “1-DPCRS” are the strategies we proposed in Section II-D “DPC” is the strategy proposed in [40] by superimposing the encoded multicast stream on top of the DPC encoded unicast streams. “Generalized RS”, “1-layer RS”, “SC–SIC”, “SC–SIC per group” and “MU–LP” mentioned in this subsection are the strategies proposed in [25] for MISO BC with a multicast message. Fig. 9 illustrates the two-user ER region comparison of all strategies. The QoS rate constraint of the multicast stream is $R^h_0 = 0.5$ bit/s/Hz. Compared with Fig. 7 for MISO BC ($R^h_0 = 0$), the rate regions of all the strategies slightly decrease due to the introduced rate constraint for the multicast message. The merits of RS-assisted transmission manifest especially when higher power is allocated for the transmission of the unicast streams since additional power can be allocated to the common stream for dynamic interference management. DPCRS maintains the largest rate region compared with the rate regions of all existing strategies. In the three-user case, we study the ESR comparison versus CSIT accuracy in Fig. 10. The QoS rate constraint of the multicast stream is $R^h_0 = 0.5$ bit/s/Hz. The rate regions of all existing strategies. In the three-user case, we study the ESR comparison versus CSIT accuracy in Fig. 10. The QoS rate constraint of the multicast stream is $R^h_0 = 0.5$ bit/s/Hz.
bit/s/Hz. Though the ESR of all strategies decrease marginally due to the QoS rate constraint of the multicast message, the ESR performance comparison coincides with the three-user results in Fig.7 for MISO BC. Overall, M-DPCRS achieves the highest ESR in MISO BC with a multicast message with explicit ESR improvement over DPC, MU–LP and SC–SIC-assisted strategies. Linearly precoded RS strategies (generalized RS and 1-layer RS) outperform non-linear DPC especially in the region with strong CSIT inaccuracy.

VI. CONCLUSION

To conclude, we propose a new benchmark, namely DPCRS in this work that achieves the best-known achievable rate region for MISO BC with partial CSIT by incorporating RS with DPC. By splitting the user messages at the transmitter into common and private parts, and use DPC to encode the private parts, DPCRS not only enables the ability to partially decode the interference and partially treat interference as noise, but also further restrains the multi-user interference among private messages. We extend the proposed DPCRS for MISO BC to MISO BC with a multicast message. The proposed DPCRS for MISO BC with a multicast message also becomes a new benchmark that achieves the best-know achievable rate region for MISO BC with a multicast message in partial CSIT. Numerical results show that existing linearly precoded RS, benefiting from its robustness in partial CSIT, outperforms DPC if CSIT is sufficiently inaccurate in MISO BC. This is sharply different from the observations in perfect CSIT where DPC outperforms all linearly precoded strategies. The proposed DPCRS not only achieves the largest rate region in both MISO BC and MISO BC with a multicast message but also less sensitive to CSIT inaccuracies, network loads and user deployments.

ACKNOWLEDGEMENT

The authors are deeply indebted to Dr. Hamdi Joudeh for his useful insights and suggestions.

REFERENCES

[1] B. Clerckx and C. Oestges, MIMO wireless networks: Channels, techniques and standards for multi-antenna, multi-user and multi-cell systems. Academic Press, 2013.
[2] B. Clerckx, H. Lee, Y. Hong, and G. Kim, “A practical cooperative multicell MIMO-OFDMA network based on rank coordination,” IEEE Trans. Wireless Commun., vol. 12, no. 4, pp. 1481–1491, April 2013.
[3] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, “Rate splitting for MIMO wireless networks: A promising PHY-layer strategy for LTE evolution,” IEEE Commun. Mag., vol. 54, no. 5, pp. 98–105, May 2016.
[4] G. Caire and S. Shamai, “On the achievable throughput of a multiantenna Gaussian broadcast channel,” IEEE Trans. Inf. Theory, vol. 49, no. 7, pp. 1691–1706, 2003.
[5] H. Weingarten, Y. Steinberg, and S. S. Shamai, “The capacity region of the Gaussian multiple-input multiple-output broadcast channel,” IEEE Trans. Inf. Theory, vol. 52, no. 9, pp. 3936–3964, Sept 2006.
[6] A. G. Davoodi and S. A. Jafar, “Aligned image sets under channel uncertainty: Settling conjectures on the collapse of degrees of freedom under finite precision CSIT,” *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5603–5618, Oct 2016.

[7] H. Joudeh and B. Clerckx, “Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach,” *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Nov 2016.

[8] E. Piovano and B. Clerckx, “Optimal DoF region of the K-user MISO BC with partial CSIT,” *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2368–2371, Nov 2017.

[9] H. Joudeh and B. Clerckx, “DoF region of the MISO BC with partial CSIT: Proof by inductive Fourier-Motzkin elimination,” in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, July 2019, pp. 1–5.

[10] C. Hao and B. Clerckx, “MISO networks with imperfect CSIT: A topological rate-splitting approach,” *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2164–2179, May 2017.

[11] C. Hao, B. Rassouli, and B. Clerckx, “Achievable DoF regions of MIMO networks with imperfect CSIT,” *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6587–6606, Oct 2017.

[12] Y. Mao, B. Clerckx, and V. O. K. Li, “Rate-splitting multiple access for downlink communication systems: bridging, generalizing, and outperforming SDMA and NOMA,” *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 133, May 2018.

[13] H. Joudeh and B. Clerckx, “Robust transmission in downlink multiuser MISO systems: A rate-splitting approach,” *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6227–6242, Dec 2016.

[14] Y. Mao, B. Clerckx, and V. O. K. Li, “Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA,” in *Proc. IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug 2018, pp. 1–5.

[15] M. Medra and T. N. Davidson, “Robust downlink transmission: An offset-based single-rate-splitting approach,” in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, June 2018, pp. 1–5.

[16] G. Lu, L. Li, H. Tian, and F. Qian, “MMSE-based precoding for rate splitting systems with finite feedback,” *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 642–645, March 2018.

[17] A. R. Flores, B. Clerckx, and R. C. de Lamare, “Tomlinson-Harashima precoded rate-splitting for multiuser multiple-antenna systems,” in *Proc. IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug 2018, pp. 1–6.

[18] H. Joudeh and B. Clerckx, “Rate-splitting for max-min fair multigroup multicast beamforming in overloaded systems,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7276–7289, Nov 2017.

[19] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, “A rate splitting strategy for massive MIMO with imperfect CSIT,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4611–4624, July 2016.

[20] M. Dai and B. Clerckx, “Multiuser millimeter wave beamforming strategies with quantized and statistical CSIT,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7025–7038, Nov 2017.

[21] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, “Rate-splitting to mitigate residual transceiver hardware impairments in massive MIMO systems,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8196–8211, Sept 2017.

[22] Y. Mao, B. Clerckx, and V. O. K. Li, “Rate-splitting multiple access for coordinated multi-point joint transmission,” *Proc. IEEE Int. Conf. Commun. (ICC) Workshop*, 2019.

[23] A. A. Ahmad, H. Dahrouj, A. Chaaban, A. Sezgin, and M. Alouini, “Interference mitigation via rate-splitting in cloud radio access networks,” in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, June 2018, pp. 1–5.

[24] Y. Mao, B. Clerckx, and V. O. K. Li, “Rate-splitting for multi-user multi-antenna wireless information and power transfer,” in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, July 2019, pp. 1–5.

[25] ———, “Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis,” *IEEE Trans. Commun.*, 2019.

[26] J. Zhang, B. Clerckx, J. Ge, and Y. Mao, “Cooperative rate-splitting for MISO broadcast channel with user relaying, and performance benefits over cooperative NOMA,” *IEEE Signal Process. Lett.*, 2019.

[27] Y. Mao, B. Clerckx, J. Zhang, V. O. K. Li, and M. Arafah, “Max-min fairness of K-user cooperative rate-splitting in MISO broadcast channel with user relaying,” *arXiv preprint arXiv: 1910.07843*, 2019.
[28] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, “Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis,” IEEE Wireless Commun. Lett., 2019.

[29] N. Jindal, “MIMO broadcast channels with finite-rate feedback,” IEEE Trans. Inf. Theory, vol. 52, no. 11, pp. 5045–5060, Nov 2006.

[30] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, “Multiuser MIMO achievable rates with downlink training and channel state feedback,” IEEE Trans. Inf. Theory, vol. 56, no. 6, pp. 2845–2866, June 2010.

[31] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, “Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT,” IEEE Trans. Inf. Theory, vol. 59, no. 1, pp. 315–328, Jan 2013.

[32] N. Jindal, S. Vishwanath, and A. Goldsmith, “Duality, dirty paper coding, and capacity for multiuser wireless channels,” in Information, Coding and Mathematics. Springer, 2002, pp. 239–256.

[33] S. Vishwanath, N. Jindal, and A. Goldsmith, “Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels,” IEEE Trans. Inf. Theory, vol. 49, no. 10, pp. 2658–2668, Oct 2003.

[34] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, “Capacity limits of MIMO channels,” IEEE J. Sel. Areas Commun., vol. 21, no. 5, pp. 684–702, June 2003.

[35] C. Hao, B. Rassouli, and B. Clerckx, “Degrees-of-freedom region of MISO-OFDMA broadcast channel with imperfect CSIT,” arXiv preprint arXiv:1310.6669, 2013.

[36] A. G. Davoodi and S. A. Jafar, “GDoF of the MISO BC: Bridging the gap between finite precision CSIT and perfect CSIT,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), July 2016, pp. 1297–1301.

[37] B. Yuan and S. A. Jafar, “Elevated multiplexing and signal space partitioning in the 2 user MIMO IC with partial CSIT,” in Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), July 2016, pp. 1–6.

[38] R. W. Heath Jr and A. Lozano, Foundations of MIMO communication. Cambridge University Press, 2018.

[39] Y. Mao, B. Clerckx, and V. O. K. Li, “Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission,” in Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), June 2018, pp. 1–5.

[40] H. Weingarten, Y. Steinberg, and S. Shamai, “On the capacity region of the multi-antenna broadcast channel with common messages,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), July 2006, pp. 2195–2199.

[41] Y. Geng and C. Nair, “The capacity region of the two-receiver gaussian vector broadcast channel with private and common messages,” IEEE Trans. Inf. Theory, vol. 60, no. 4, pp. 2087–2104, April 2014.

[42] H. Viswanathan, S. Venkatesan, and H. Huang, “Downlink capacity evaluation of cellular networks with known-interference cancellation,” IEEE J. Sel. Areas Commun., vol. 21, no. 5, pp. 802–811, June 2003.

[43] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” IEEE Trans. Wireless Commun., vol. 7, no. 12, pp. 4792–4799, Dec 2008.

[44] M. Grant, S. Boyd, and Y. Ye, “CVX: Matlab software for disciplined convex programming,” 2008.

[45] C. Hao, Y. Wu, and B. Clerckx, “Rate analysis of two-receiver MISO broadcast channel with finite rate feedback: A rate-splitting approach,” IEEE Trans. Commun., vol. 63, no. 9, pp. 3232–3246, Sept 2015.

[46] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, “A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems,” IEEE Trans. Signal Process., vol. 64, no. 1, pp. 76–88, Jan 2016.

[47] J. Choi, “On generalized downlink beamforming with NOMA,” J. Commun. Networks, vol. 19, no. 4, pp. 319–328, August 2017.