

Sum-Rate Maximization for Linearly Precoded Downlink Multiuser MISO Systems with Partial CSIT: A Rate-Splitting Approach

Hamdi Joudeh and Bruno Clerckx

Abstract—This paper considers the Sum-Rate (SR) maximization problem in downlink MU-MISO systems under imperfect Channel State Information at the Transmitter (CSIT). Contrary to existing works, we consider a rather unorthodox transmission scheme. In particular, the message intended to one of the users is split into two parts: a common part which can be recovered by all users, and a private part recovered by the corresponding user. On the other hand, the rest of users receive their information through private messages. This Rate-Splitting (RS) approach was shown to boost the achievable Degrees of Freedom (DoF) when CSIT errors decay with increased SNR. In this work, the RS strategy is married with linear precoder design and optimization techniques to achieve a maximized Ergodic SR (ESR) performance over the entire range of SNRs. Precoders are designed based on partial CSIT knowledge by solving a stochastic rate optimization problem using means of Sample Average Approximation (SAA) coupled with the Weighted Minimum Mean Square Error (WMMSE) approach. Numerical results show that in addition to the ESR gains, the benefits of RS also include relaxed CSIT quality requirements and enhanced achievable rate regions compared to conventional transmission with NoRS.

Index Terms—MISO-BC, ergodic sum-rate, degrees of freedom, sample average approximation, WMMSE approach.

I. INTRODUCTION

THE utilization of multiple antennas at the Base Station (BS) combined with multiple single-antenna mobile devices tremendously increases the spectral efficiencies of wireless networks. High multiplexing gains are realized with far less restrictions on the scattering environment compared to point-to-point Multiple Input Multiple Output (MIMO) systems [1]. However, this comes with a price of higher restrictions imposed on the quality of the Channel State Information (CSI) required at the BS, specifically in the Downlink (DL) mode. This stems from the necessity to deal with the interference through preprocessing at the BS, as receivers cannot coordinate. While precise CSI at the Receivers (CSIR) can be obtained through DL training, the ability to provide highly accurate and up-to-date CSI at the Transmitter (CSIT) remains

questionable. Therefore, considerable effort has been devoted to the characterization and improvement of the performance in the presence of CSIT uncertainties.

It is well established that linear transmission strategies, e.g. Zero Forcing Beamforming (ZF-BF), achieve the optimum Degrees of Freedom (DoF) of the MISO Broadcast Channel (BC) under perfect CSIT [2], [3]. It is also possible to maintain the full DoF in the presence of imperfect CSIT given that the errors decay with increased Signal to Noise Ratio (SNR) as $O(\text{SNR}^{-1})$ [4]–[6]. For example, it has been shown that employing ZF-BF in limited channel feedback systems achieves the maximum DoF as long as the feedback rate increases linearly with the SNR in dB, satisfying an error decay rate of $O(\text{SNR}^{-1})$ [5]. However, maintaining such quality can be exhausting in terms of resources required for CSIT acquisition. Furthermore, the CSIT accuracy is expected to drop due to the Doppler effect, feedback delays and/or mismatches across multiple uses of the channel. While understanding the MISO-BC capacity region under CSIT imperfections is still far from complete, considerable progress has been made towards characterizing the achievable and optimum DoF – as an asymptotic tractable alternative – for a variety of imperfect CSIT scenarios [7]–[10]. In this work, we focus on the case where imperfect instantaneous CSIT is available, with errors decaying as $O(\text{SNR}^{-\alpha})$ for some constant $\alpha \in [0, 1]$. Such errors are detrimental to the DoF achieved using conventional transmission schemes initially designed assuming perfect CSIT. For example, ZF-BF achieves a fraction α of the full DoF [5], [6]. Therefore we consider a rather unconventional Rate-Splitting (RS) transmission strategy.

The message intended to one of the users (labeled as the RS-user) is split into two parts: a common part decoded by all users, and a private part decoded by the corresponding user. On the other hand, the information intended to the remaining users is sent through private messages. At the receivers, the common message is decoded first by treating all private signals as noise. Then, each receiver decodes its private message after removing the common message from the received signal. This RS strategy, inspired by similar techniques used for the Interference Channel (IC) [11], was recently shown to achieve a DoF gain of $1 - \alpha$ over conventional transmission (relying solely on the transmission of private messages and termed NoRS here) in the MISO-BC [7], [8]. An interpretation of this result in the context of limited feedback and Random Vector Quantization (RVQ) of the channel vectors is given in [12], where the RS rate performance is analysed and compared to

This work is partially supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under grant EP/N015312/1. A preliminary version of this paper was presented at the IEEE International Conference on Communications (ICC), London, U.K., June 2015.

H. Joudeh is with the Communications and Signal Processing group, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (email: hamdi.joudeh10@imperial.ac.uk)

B. Clerckx is with the Communications and Signal Processing group, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K., and also with the School of Electrical Engineering, Korea University, Seoul 136-701, Korea (e-mail: b.clerckx@imperial.ac.uk).

its NoRS counterpart.

While RS in the MISO-BC is fairly understood from an information theoretic point of view (at least in the DoF sense), it is considerably less treated and analysed in the transceiver/precoder design and optimization literature. For example, none of the aforementioned works considers an optimized design of precoders. One reason is that theoretical analysis is made possible by simpler designs, e.g. random precoding for the common message and ZF-BF for private messages [12], as it is easier to derive closed-form expressions of performance metrics. Such closed-forms are usually sufficient for DoF analysis, where the asymptotically high SNR regime is considered. However, this is not the case for finite SNRs where the design of precoders is highly influential. Precoder optimization for Multi-User (MU) MISO/MIMO systems has been considered in a number of works under a variety of metrics, design objectives and constraints, and CSI assumptions [13]–[22]. In general, problems that involve Sum-Rate (SR) expressions in such setups are known to be very challenging to solve in their raw forms due to the non-convexity arising from decentralized receivers with coupled rate expressions¹. Nevertheless, the novel relationship proposed in [14] enables the reformulation of SR problems into equivalent special forms of Weighted Mean Square Error (WMSE) problems. This unveils a block-wise convexity property which can be exploited using Alternating Optimization (AO).

Contributions: With the aim of maximizing the SR performance in MU-MISO systems with partial CSIT, this paper marries the RS strategy with precoder design and optimization techniques. In doing so, we generalize previous works on SR maximization under imperfect CSIT in several directions:

- First, we consider a general channel fading model where the channel state, and its corresponding estimate at the BS, are allowed to change throughout the transmission according to some stationary process. This model contains as special cases the Rician-fading scenario in [19]–[21] where a fixed channel estimate (e.g. line-of-sight) throughout the transmission was considered, and the fast-fading scenario in [24] where the instantaneous channel estimate is completely suppressed. A feasible solution is defined as a sequence of precoders, with one for each incoming channel estimate.
- Second, the adoption of the RS strategy ensures that the proposed design performs at least as well as the corresponding NoRS-based design across the entire range of SNRs. This follows from the fact that NoRS can be seen as a special (or restricted) case of RS. We demonstrate that RS designs outperform conventional NoRS designs at high SNRs through DoF analysis.

In order to average out the effects of CSIT errors, we consider the Ergodic SR (ESR) achieved over a long sequence of fading states as the main performance metric [25]. Given partial CSIT, the ESR is maximized by updating the precoding matrix according to the incoming channel state estimate such

¹This can be resolved by resorting to asymptotic regimes, e.g. high SNR or large antenna arrays [23]. From a design and optimization point of view, this work is mainly concerned with finite SNRs and small antenna arrays.

that a conditional Average SR (ASR) metric is maximized. This metric corresponds to the average performance w.r.t CSIT errors for a given channel estimate, and can be computed by the BS using imperfect instantaneous knowledge. To solve the stochastic ASR problem, it is first converted into a deterministic counterpart using the Sampling Average Approximation (SAA) method [26]. The deterministic problem is then solved using the Weighted Minimum MSE (WMMSE) approach [14]. A simplified method to approximate the stochastic problem based on the conservative method in [19], [20] is also proposed, and its limitations are discussed. The benefits of employing the RS strategy are demonstrated through simulations. These include enhanced rate performances and relaxed CSIT quality requirements. Moreover, the two-user RS rate region is numerically obtained by solving a sequence of Weighted ASR (WASR) problems, where the splitting procedure is extended to both users. Examining such achievable rate region brings us one step closer to understanding the behaviour of the MISO-BC's capacity region under partial CSIT.

It should be noted that part of the results presented here were reported in a preliminary version of this paper [27]. Since then, a number of RS design problems have been addressed, namely: achieving max-min fairness [28], robust transmission under bounded CSIT errors [29], and the application to massive MIMO [30]. In a recent work, a topological RS strategy is proposed for MISO networked systems [31].

Organization: The rest of the paper is organized as follows. The system model and CSIT assumptions are described in Section II. In Section III, the problem is formulated and DoF analysis is carried out. The SAA based WMMSE algorithm is proposed in Section IV, then the conservative WMMSE algorithm is described in Section V. Simulation results are presented in Section VI, and Section VII concludes the paper.

Notation: Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors and standard letters denote scalars. The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate-transpose (Hermitian) operators, respectively. $\text{tr}(\cdot)$ and $\text{diag}(\cdot)$ are the trace and diagonal entries respectively. $\|\cdot\|$ is the Euclidean norm. $\mathbb{E}_X\{\cdot\}$ denotes the expectation w.r.t the random variable X .

II. SYSTEM MODEL

Consider a MU-MISO system operating in DL, where a BS equipped with N_t antennas serves a set of single-antenna users $\mathcal{K} \triangleq \{1, \dots, K\}$, where $K \leq N_t$. The signal received by the k th user in a given channel use (time or frequency) is described as

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k \quad (1)$$

where $\mathbf{h}_k \in \mathbb{C}^{N_t}$ is the channel vector between the BS and the k th user, $\mathbf{x} \in \mathbb{C}^{N_t}$ is the transmit signal, and $n_k \sim \mathcal{CN}(0, \sigma_{n,k}^2)$ is the Additive White Gaussian Noise (AWGN). The input signal is subject to the power constraint $\mathbb{E}\{\mathbf{x}^H \mathbf{x}\} \leq P_t$. Without loss of generality, we assume equal noise variances across users, i.e. $\sigma_{n,k}^2 = \sigma_n^2, \forall k \in \mathcal{K}$. The transmit SNR writes as $\text{SNR} \triangleq P_t / \sigma_n^2$. Moreover, σ_n^2 is non-zero and fixed. Hence, $\text{SNR} \rightarrow \infty$ is equivalent to $P_t \rightarrow \infty$.

A. Channel State Information

We assume a fading model where the channel state, given by $\mathbf{H} \triangleq [\mathbf{h}_1, \dots, \mathbf{h}_K]$, varies during the transmission according to an ergodic stationary process with probability density $f_{\mathbf{H}}(\mathbf{H})$. Receivers are assumed to estimate and track their channel vectors with high accuracy, i.e. perfect CSIR. The BS on the other hand has an imperfect instantaneous channel estimate given by $\hat{\mathbf{H}} \triangleq [\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_K]$. This is typically obtained through Uplink (UL) training in Time Division Duplex (TDD) systems [32] or quantized feedback in Frequency Division Duplex (FDD) systems [33]. Hence, the joint fading process is characterized by the joint distribution of $\{\mathbf{H}, \hat{\mathbf{H}}\}$, assumed to be stationary and ergodic [25]. For a given estimate, the estimation error matrix is denoted by $\tilde{\mathbf{H}} \triangleq [\tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_K]$, from which we write the relationship $\mathbf{H} = \hat{\mathbf{H}} + \tilde{\mathbf{H}}$. The CSIT error is characterized by the conditional density $f_{\mathbf{H}|\hat{\mathbf{H}}}(\mathbf{H} | \hat{\mathbf{H}})$.

Taking each user separately, the marginal density of the k th channel conditioned on its estimate writes as $f_{\mathbf{h}_k|\hat{\mathbf{h}}_k}(\mathbf{h}_k | \hat{\mathbf{h}}_k)$. The mean of the distribution is assumed to be given by the estimate, i.e. $\mathbb{E}_{\mathbf{h}_k|\hat{\mathbf{h}}_k}\{\mathbf{h}_k | \hat{\mathbf{h}}_k\} = \hat{\mathbf{h}}_k$, and $\mathbb{E}_{\mathbf{h}_k|\hat{\mathbf{h}}_k}\{\mathbf{h}_k \mathbf{h}_k^H | \hat{\mathbf{h}}_k\} = \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{R}_{e,k}$, where $\mathbf{R}_{e,k}$ is the k th user's CSIT error covariance matrix assumed to be independent of $\hat{\mathbf{h}}_k$. The CSIT quality is allowed to scale with SNR [5], [7]. In particular, the maximum entry of $\text{diag}(\mathbf{R}_{e,k})$ scales as $O(P_t^{-\alpha})$, where $\alpha \in [0, \infty)$ is some constant. Equivalently, we write $\sigma_{e,k}^2 = O(P_t^{-\alpha})$, where $\sigma_{e,k}^2 \triangleq \mathbb{E}_{\tilde{\mathbf{h}}_k}\{\|\tilde{\mathbf{h}}_k\|^2\} = \text{tr}(\mathbf{R}_{e,k})$ is the corresponding average CSIT error power. The constant $\alpha \triangleq \lim_{P_t \rightarrow \infty} -\frac{\log(\sigma_{e,k}^2)}{\log(P_t)}$ is known as the quality scaling factor (or exponent), which quantifies the CSIT quality as SNR grows large. For example, $\alpha \rightarrow \infty$ corresponds to perfect CSIT as $\sigma_{e,1}^2, \dots, \sigma_{e,K}^2 \rightarrow 0$. The opposite extreme of $\alpha = 0$ represents a fixed quality w.r.t SNR, e.g. a constant number of quantization (and hence feedback) bits in FDD systems. A finite non-zero α corresponds to CSIT quality that improves with increased SNR, e.g. by increasing the number of feedback bits. The exponent is truncated such that $\alpha \in [0, 1]$, which is customary in DoF analysis as $\alpha = 1$ corresponds to perfect CSIT in the DoF sense [7]. It is worth highlighting that α assumes various practical interpretations, e.g. in addition to the aforementioned channel quantization and feedback interpretation [5], [12], it can also correspond to the Doppler process in delayed/outdated CSIT [6], [7].

B. Rate-Splitting and Transmit Signal Model

The BS wishes to communicate the independent messages $W_{t,1}, \dots, W_{t,K}$, drawn from the message sets $\mathcal{W}_{t,1}, \dots, \mathcal{W}_{t,K}$, to users $1, \dots, K$ respectively. In conventional linearly precoded systems, $W_{t,1}, \dots, W_{t,K}$ are independently encoded into the private symbol streams $s_1, \dots, s_K$ respectively. Symbols are mapped to the transmit antennas through a precoding matrix denoted by $\mathbf{P}_p \triangleq [\mathbf{p}_1, \dots, \mathbf{p}_K]$, where $\mathbf{p}_k \in \mathbb{C}^{N_t}$ is the k th precoding vector. This yields the linear transmit signal model described as $\mathbf{x} = \sum_{k=1}^K \mathbf{p}_k s_k$. At each receiver, the desired stream is decoded while treating interference from all other streams as noise. It is, therefore, the BS's duty to minimize MU interference through a proper

design of $\mathbf{P}_p$. Unfortunately, this is not possible when the BS experiences high CSIT uncertainty. Hence, we resort to the unconventional RS transmission model, where part of the interference is broadcasted such that it is decoded and cancelled by all users before decoding their own streams [12].

The RS scheme is described as follows. The message intended to the k_{RS} th user (the RS-user) is split into a common part $W_c \in \mathcal{W}_c$ and a private part $W_{k_{\text{RS}}} \in \mathcal{W}_{k_{\text{RS}}}$, where $\mathcal{W}_c \times \mathcal{W}_{k_{\text{RS}}} = \mathcal{W}_{t,k_{\text{RS}}}$. W_c is encoded into the stream s_c using a common (or public) codebook, and hence is decoded by all users, while $W_{t,1}, \dots, W_{k_{\text{RS}}}, \dots, W_{t,K}$ are encoded into the private stream $s_1, \dots, s_K$ in the conventional manner. The $K+1$ streams are linearly precoded using $\mathbf{P} \triangleq [\mathbf{p}_c, \mathbf{p}_1, \dots, \mathbf{p}_K]$, where $\mathbf{p}_c \in \mathbb{C}^{N_t}$ is the common precoder. The resulting transmit signal writes as

$$\mathbf{x} = \mathbf{P}\mathbf{s} = \mathbf{p}_c s_c + \sum_{i=1}^K \mathbf{p}_i s_i \quad (2)$$

where $\mathbf{s} \triangleq [s_c, s_1, \dots, s_K]^T \in \mathbb{C}^{K+1}$ groups the symbols in a given channel use. Assuming that $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$, the transmit power constraint reduces to $\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t$. When $|\mathcal{W}_c| = 0$, i.e. no splitting is carried out, it is natural to allocate zero power to the common precoder, and the signal in (2) simply reduces to the conventional transmit signal. While the model in (2) can be seen as a super-position of MU beamforming [34] and multicast beamforming [35]², the transmission of a common message has a fundamentally different purpose here. In particular, multicast transmission sends information requested by, and intended to, all users in the system. On the other hand, the common message in RS encapsulates part of the RS-user's private message, which is decoded by all users in the system for interference mitigation and performance enhancement purposes as we see later in this paper³.

C. SINRs and Rates

At the k th receiver, the average receive power for a given channel state is written as

$$T_{c,k} \triangleq \mathbb{E}\{|y_k|^2\} = \underbrace{|\mathbf{h}_k^H \mathbf{p}_c|^2}_{S_{c,k}} + \underbrace{|\mathbf{h}_k^H \mathbf{p}_k|^2}_{S_k} + \underbrace{\sum_{i \neq k} |\mathbf{h}_k^H \mathbf{p}_i|^2}_{I_k} + \sigma_n^2. \quad (3)$$

$I_{c,k} = T_k$

Each receiver decodes two streams, the common stream and its corresponding private stream. The common stream is first decoded by treating interference from all private signals as noise. Successive Interference Cancellation (SIC) is then used to remove the common signal from y_k , in order to improve the detectability of the private stream. This is followed by decoding the private stream in the presence of the remaining interference. The instantaneous (for a given channel state) Signal to Interference plus Noise Ratios (SINRs) of the

²A BC with both private messages and a common message from an information theoretic point of view [11].

³Note that decoding other user(s) data at the physical layer does not necessarily breach secrecy as encryption may be implemented at higher layers.

common stream and the private stream at the output of the k th receiver are given by

$$\gamma_{c,k} \triangleq S_{c,k} I_{c,k}^{-1} \quad \text{and} \quad \gamma_k \triangleq S_k I_k^{-1}. \quad (4)$$

Assuming Gaussian codebooks, the instantaneous achievable information rates for the common stream and the k th private stream, from the point of view of the k th user, write as

$$R_{c,k} = \log_2(1 + \gamma_{c,k}) \quad \text{and} \quad R_k = \log_2(1 + \gamma_k). \quad (5)$$

We assume that the transmission is delay-unlimited, and hence channel coding (from messages to symbols) can be performed over a long sequence of channel states [36]. This is considered to achieve an average CSIT error performance as we see in the next section. Precoders on the other hand are adapted throughout the transmission depending on the available estimate of the instantaneous channel state [6], [25]. As far as the codewords (symbol streams) are concerned, adaptive precoders are observed as part of the fading channel. It follows that sending the common message and the k th private message at the Ergodic Rates (ERs) given by $E_H \{R_{c,k}\}$ and $E_H \{R_k\}$ respectively guarantees successful decoding by the k th user. To guarantee that W_c is successfully decoded (and hence cancelled) by all users, it should be transmitted at an ER not exceeding $\min_j \{E_H \{R_{c,j}\}\}_{j=1}^K$. Finally, the total ER achieved by the k th user writes as $E_H \{R_k\}$ for $k \neq k_{RS}$, and $E_H \{R_k\} + \min_j \{E_H \{R_{c,j}\}\}_{j=1}^K$ for $k = k_{RS}$.

III. MOTIVATION, PROBLEM FORMULATION AND DOF ANALYSIS

In this section, we formulate the precoder design problem for ESR maximization in the RS transmission scheme described in the previous section. In doing so, we propose the ASR optimization framework which exploits partial instantaneous CSIT to achieve a robust maximized ergodic performance. Moreover, we derive the DoF performances of the ASR-optimized NoRS and RS designs. To gain some insight, we start by looking at the conventional (NoRS) ESR maximization problem in the presence of perfect instantaneous CSIT. This is formulated as

$$\max_{\{P_p\}} E_H \left\{ \sum_{k=1}^K R_k \right\}. \quad (6)$$

The availability of perfect CSIT enables the BS to adapt the precoding matrix P_p according to H such that the instantaneous SR (inside the expectation) is maximized. The ergodic nature of the transmission allows the long-term power constraint in (6), yielding a form of inter-state (temporal or frequency) power allocation. A feasible solution for such problem consists of a set of precoding matrices, each for a channel state. Moreover, we define a precoding scheme as a family of feasible solutions for all possible power levels, i.e. $\{P_p(P_t, H)\}_{P_t, H}$. It is common to replace the long-term power constraint in (6) with a short-term one given by $\text{tr}(P_p P_p^H) \leq P_t$ [6], [33], which corresponds to a precoder peak power constraint in practice. While this may result in a smaller ESR, it has a significant impact on the tractability of the problem. In particular, the inter-state power allocation

disappears as precoders for different channel states are no longer coupled. As a result, the maximization in (6) moves inside the expectation, and the optimum solution is obtained by optimizing P_p separately for each H , i.e.

$$\max_{\text{tr}(P_p P_p^H) \leq P_t} \sum_{k=1}^K R_k \quad (7)$$

which corresponds to the problem in [14]. Due to its tractability, we adopt the short-term power constraint.

A. Average Sum Rate Maximization

Now what if the BS has partial instantaneous CSIT as described in Section II-A? The channel state is characterized by $\hat{H}$ from the BS's perspective, and it is natural to define a precoding scheme as $\{P_p(P_t, \hat{H})\}_{P_t, \hat{H}}$. A naive approach would be to design each precoding matrix by solving problem (7) assuming that $\hat{H}$ is perfect. The resulting design is not only unable to cope with the MU interference, it is also unaware of it. In addition, this may lead to an overestimation of the instantaneous and ergodic rates by the BS, yielding transmission at undecodable rates. A robust approach on the other hand employs the available CSIT knowledge to: 1) design an informed precoding scheme that enhances the instantaneous channel condition experienced by the receivers in each channel state, 2) perform transmission at reliable rates and hence guarantee decodability. In the following, we describe how this can be achieved.

While the BS is unable to predict the instantaneous rates, it has access to the the Average Rates (ARs) defined as: $\bar{R}_k \triangleq E_{H|\hat{H}} \{R_k | \hat{H}\}$ for all $k \in \mathcal{K}$, determined by the imperfect channel state. The AR and the ER should not be confused: while the latter describes the long-term performance over all channel states, the former is a short-term (instantaneous) measure that captures the expected performance over the CSIT error distribution for a given channel state estimate. It turns out that the ERs can be characterized by averaging the ARs over the variation in $\hat{H}$. In particular, for a given fading process and adaptive precoding strategy, it is evident that $P_p(\hat{H})$ and $\bar{R}_k(\hat{H})$ depend on $\hat{H}$, while $R_k(H, \hat{H})$ is determined by $\{H, \hat{H}\}$, as $P_p(\hat{H})$ is selected by the BS based on the estimate. From the law of total expectation, the ER experienced by the user is expressed as

$$\begin{aligned} E_{\{H, \hat{H}\}} \{R_k(H, \hat{H})\} &= E_{\hat{H}} \{E_{H|\hat{H}} \{R_k(H, \hat{H}) | \hat{H}\}\} \\ &= E_{\hat{H}} \{\bar{R}_k(\hat{H})\} \end{aligned} \quad (8)$$

from which the ESR for a given precoding strategy writes as $E_{\hat{H}} \left\{ \sum_{k=1}^K \bar{R}_k \right\}$. Following the same logic in (6) and (7), maximizing the ESR under partial CSIT and a short-term power constraint is achieved by optimizing $P_p(\hat{H})$ such that the ASR, i.e. $\sum_{k=1}^K \bar{R}_k$, is maximized for each $\hat{H}$. For a class of information theoretic channels under imperfect CSIT with a Markov property, it has been shown that the ergodic capacity is found in a similar manner, while replacing the ASR with the average mutual-information, and maximizing over the input

distribution rather than the precoding matrix [25]. This gives rise to the ASR optimization problem formulated as

$$\mathcal{R}(P_t) : \begin{cases} \max_{\mathbf{P}_p} & \sum_{k=1}^K \bar{R}_k \\ \text{s.t.} & \text{tr}(\mathbf{P}_p \mathbf{P}_p^H) \leq P_t. \end{cases} \quad (9)$$

It follows that the maximized ESR given by $E_{\hat{\mathbf{H}}} \{\mathcal{R}(P_t)\}$ is achievable. The fixed-coding/adaptive-precoding strategy described in Section II-C can be employed, where the knowledge of the long-term properties of CSIT is leveraged to predict the ERs and adjust the code rates, while precoders are adapted each time a new channel estimate is revealed. It is evident that for perfect CSIT, (9) reduces to the SR problem in (7). A discussion on the influence of CSIT uncertainty on the ESR performance follows in the next subsection. It is worth noting that the considered CSIT model reduces to the one in [19], [20] when $\hat{\mathbf{H}}$ is fixed over all channel states, and the fast-fading model in [24] when $\hat{\mathbf{H}}$ is completely suppressed. Both lead to special cases of the problem formulation, where the ESR problem and the ASR problem coincide, and a precoding scheme is given by a single precoding matrix for each P_t .

Now we turn to designing a precoding scheme $\{\mathbf{P}(P_t, \hat{\mathbf{H}})\}_{P_t, \hat{\mathbf{H}}}$ that maximizes the ESR performance of the RS strategy by formulating the corresponding ASR problem. The k th common AR is defined as $\bar{R}_{c,k} \triangleq E_{\mathbf{H}|\hat{\mathbf{H}}} \{R_{c,k} \mid \hat{\mathbf{H}}\}$, from which the common AR is given by $\bar{R}_c \triangleq \min_j \{\bar{R}_{c,j}\}_{j=1}^K$. The ASR is expressed as $\bar{R}_c + \sum_{k=1}^K \bar{R}_k$, and the ASR maximization problem is formulated as

$$\mathcal{R}_{\text{RS}}(P_t) : \begin{cases} \max_{\bar{R}_c, \mathbf{P}} & \bar{R}_c + \sum_{k=1}^K \bar{R}_k \\ \text{s.t.} & \bar{R}_{c,k} \geq \bar{R}_c, \forall k \in \mathcal{K} \\ & \text{tr}(\mathbf{P} \mathbf{P}^H) \leq P_t \end{cases} \quad (10)$$

where the inequality constraints involving $\bar{R}_c$ are equivalent to the pointwise minimization in the definition of $\bar{R}_c$. Next, we show that the ESR given by $E_{\hat{\mathbf{H}}} \{\mathcal{R}_{\text{RS}}(P_t)\}$ is achievable. The reliability of the private ERs follows from (8). As for the common ER, we write

$$\min_{j \in \mathcal{K}} \{E_{\{\mathbf{H}, \hat{\mathbf{H}}\}} \{R_{c,j}\}\} = \min_{j \in \mathcal{K}} \{E_{\hat{\mathbf{H}}} \{\bar{R}_{c,j}\}\} \geq E_{\hat{\mathbf{H}}} \left\{ \min_{j \in \mathcal{K}} \bar{R}_{c,j} \right\} \quad (11)$$

where the equality follows from the law of total expectation as in (8), and the inequality follows from the fact that moving the minimization inside the expectation does not increase the value. Leaving the minimization outside the expectation couples the common ARs, and hence $\mathbf{P}(\hat{\mathbf{H}})$, across the different channel states, leading to an intractable formulation. After formulating the NoRS and RS instantaneous ASR problems, the next two questions that come to mind are: 1) how does the instantaneous CSIT quality influence the long-term performances given by $E_{\hat{\mathbf{H}}} \{\mathcal{R}(P_t)\}$ and $E_{\hat{\mathbf{H}}} \{\mathcal{R}_{\text{RS}}(P_t)\}$? and 2) How do the NoRS and RS performances compare? To answer these questions, we resort to DoF analysis.

B. DoF Performance

The DoFs for the NoRS and RS strategies are defined as

$$\lim_{P_t \rightarrow \infty} \frac{E_{\hat{\mathbf{H}}} \{\mathcal{R}(P_t)\}}{\log_2(P_t)} \quad \text{and} \quad \lim_{P_t \rightarrow \infty} \frac{E_{\hat{\mathbf{H}}} \{\mathcal{R}_{\text{RS}}(P_t)\}}{\log_2(P_t)}. \quad (12)$$

The DoF can be roughly interpreted as the total number of interference-free streams that can be simultaneously supported in a single channel use. This follows by noting that the rate of a single interference-free data stream scales as $\log_2(P_t) + O(1)$. The significance of the DoF in such setups comes from the detrimental effects MU interference may have, and the role of CSIT in dealing with such effect. For the proof of our next result, we assume isotropically distributed CSIT error vectors, i.e. $\mathbf{R}_{e,k} = \frac{\sigma_{e,k}^2}{N_t} \mathbf{I}$ for all $k \in \mathcal{K}$. Moreover, we assume that $\hat{\mathbf{H}}$ is of full column rank with probability one. Note that these assumptions are not necessary for the optimization in the following sections.

Theorem 1. The DoF of an optimum NoRS scheme is

$$\lim_{P_t \rightarrow \infty} \frac{E_{\hat{\mathbf{H}}} \{\mathcal{R}(P_t)\}}{\log_2(P_t)} = \max \{1, K\alpha\} \quad (13)$$

while the DoF of an optimum RS scheme is given by

$$\lim_{P_t \rightarrow \infty} \frac{E_{\hat{\mathbf{H}}} \{\mathcal{R}_{\text{RS}}(P_t)\}}{\log_2(P_t)} = 1 + (K-1)\alpha. \quad (14)$$

The proof is relegated to Appendix A. It is evident that when $\alpha = 1$, the maximum attainable DoF is achieved, i.e. K . Although CSIT may still be erroneous, it is good enough to design private precoders that reduce the MU interference to the level of additive noise. However, this is not possible when α drops below 1, as private streams start leaking substantial interference. The NoRS scheme exhibits loss in DoF until $\alpha = 1/K$. As α drops below $1/K$, the CSIT quality is not good enough to support multi-stream transmission, and it is preferred from a DoF perspective to switch to SU transmission by allocating all power to one stream, hence achieving a DoF of 1. In naive designs with fixed and uniform power allocation across users [5], [6], the DoF reduces to zero for $\alpha = 0$, and the gain from switching to SU transmission is not realized. Now looking at RS, as soon as α drops below 1, the scheme carefully allocates powers to the private streams such that MU interference is always received at the level of additive noise. This is achieved by scaling down private powers to $O(P_t^\alpha)$, hence maintaining a DoF of $K\alpha$. The rest of the power, which scales as $O(P_t)$, is allocated to the common stream which is broadcasted to all users. This achieves a DoF gain of $1 - \alpha$, as interference from private messages is treated as noise. It should be noted that the RS DoF is strictly greater than the NoRS DoF for all $\alpha \in (0, 1)$.

The RS DoF in (14) is inline with the results in [7], [8]. Although optimizing the precoders does not improve the achievable DoF in these works, simulation results in Section VI show that optimized RS schemes are superior from a rate performance perspective. It is worth highlighting that the converse result recently reported in [37] proves the optimality of the DoF in (14) for the Gaussian MISO-BC when one user has perfect CSIT. Since improving the CSIT quality of one user

cannot decrease the DoF, the optimality of the result in (14) naturally follows. We conclude this section by highlighting that further to the high SNR analysis, $\mathcal{R}_{\text{RS}}(P_t) \geq \mathcal{R}(P_t)$ is guaranteed over the entire range of SNRs. This is seen by noting that solving (9) is equivalent to solving (10) over a subset of its domain characterized by restricting $\|\mathbf{p}_c\|^2$ to zero. In the following, we focus on solving the RS design problem, achieved by solving (10).

IV. SAMPLE AVERAGE APPROXIMATED WMMSE ALGORITHM

The RS ASR problem in (10) is of a stochastic nature and appears to be challenging to solve in its current form. In fact, deterministic subproblems of (10) are known to be non-convex and non-trivial [14], [35]. We propose a two-step approach to obtain a tractable form of (10). In the first step, a deterministic approximation is obtained using the SAA method [26]. In the second step, the deterministic ASR problem is transformed into an equivalent, and solvable, augmented Average Weighted Sum MSE (AWSMSE) problem. Before we proceed, it should be highlighted that in the following analysis, the dependencies of some functions on certain relevant variables is occasionally emphasized in the notation.

A. Sample Average Approximation

For a given estimate $\hat{\mathbf{H}}$ and index set $\mathcal{M} \triangleq \{1, \dots, M\}$, let

$$\mathbb{H}^{(M)} \triangleq \{\mathbf{H}^{(m)} = \hat{\mathbf{H}} + \tilde{\mathbf{H}}^{(m)} \mid \hat{\mathbf{H}}, m \in \mathcal{M}\}$$

be a sample of M i.i.d realizations drawn from the conditional distribution with density $f_{\mathbf{H}|\hat{\mathbf{H}}}(\mathbf{H} \mid \hat{\mathbf{H}})$. This sample is used to approximate the ARs through their Sample Average Functions (SAFs) defined as: $\bar{R}_{c,k}^{(M)} \triangleq \frac{1}{M} \sum_{m=1}^M R_{c,k}^{(m)}$ and $\bar{R}_k^{(M)} \triangleq \frac{1}{M} \sum_{m=1}^M R_k^{(m)}$, where $R_{c,k}^{(m)} \triangleq R_{c,k}(\mathbf{H}^{(m)})$ and $R_k^{(m)} \triangleq R_k(\mathbf{H}^{(m)})$ are rates associated with the m th realization. This leads to the SAA of problem (10) formulated as

$$\mathcal{R}_{\text{RS}}^{(M)}(P_t) : \begin{cases} \max_{\bar{R}_c, \mathbf{P}} & \bar{R}_c + \sum_{k=1}^K \bar{R}_k^{(M)} \\ \text{s.t.} & \bar{R}_{c,k}^{(M)} \geq \bar{R}_c, \forall k \in \mathcal{K} \\ & \text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t. \end{cases} \quad (15)$$

It should be noted that $\mathbf{P}$ is fixed over the M realizations, which follows from the definition of the ARs. Before we proceed, we make the following finite SNR/SINR assumption.

Assumption 1. In the following we assume that $\text{SNR} < \infty$, i.e. $\sigma_n^2 > 0$ and $P_t < \infty$. Moreover, we assume that channel realizations are bounded. Hence, $\gamma_{c,k}, \gamma_k < \infty$ for all $k \in \mathcal{K}$.

It follows that all rates are bounded for any P_t and all channel realizations. Therefore, from the strong Law of Large Numbers (LLN), we have

$$\lim_{M \rightarrow \infty} \bar{R}_{c,k}^{(M)}(\mathbf{P}) = \bar{R}_{c,k}(\mathbf{P}), \text{ a.s. } \forall \mathbf{P} \in \mathbb{P} \quad (16a)$$

$$\lim_{M \rightarrow \infty} \bar{R}_k^{(M)}(\mathbf{P}) = \bar{R}_k(\mathbf{P}), \text{ a.s. } \forall \mathbf{P} \in \mathbb{P} \quad (16b)$$

where the dependency on $\mathbf{P}$ is highlighted, a.s. denotes almost surely, and $\mathbb{P} \triangleq \{\mathbf{P} \mid \text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t\}$ is the feasible set of

precoders. (16) suggests that the optimum solution of problem (15) converges to its counterpart of problem (10) as $M \rightarrow \infty$. In the following, we show that this is indeed the case.

First, we note that $\mathbb{P}$ is compact and the rate functions are bounded and continuously differentiable in $\mathbf{P}$. It follows that the convergence in (16) is uniform in $\mathbf{P} \in \mathbb{P}$. The ARs are also continuously differentiable with gradients given by $\nabla_{\mathbf{P}} \bar{R}_{c,k}(\mathbf{P}) = \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \{\nabla_{\mathbf{P}} R_{c,k}(\mathbf{P}) \mid \hat{\mathbf{H}}\}$ and $\nabla_{\mathbf{P}} \bar{R}_k(\mathbf{P}) = \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \{\nabla_{\mathbf{P}} R_k(\mathbf{P}) \mid \hat{\mathbf{H}}\}$, which follows from the bounded convergence theorem [38]. On the other hand, the objective functions of (10) and (15) can be equivalently reformulated by incorporating the common rates as

$$\bar{R}_s(\mathbf{P}) \triangleq \min_j \{\bar{R}_{c,j}(\mathbf{P})\}_{j=1}^K + \sum_{k=1}^K \bar{R}_k(\mathbf{P}) \quad (17a)$$

$$\bar{R}_s^{(M)}(\mathbf{P}) \triangleq \min_j \{\bar{R}_{c,j}^{(M)}(\mathbf{P})\}_{j=1}^K + \sum_{k=1}^K \bar{R}_k^{(M)}(\mathbf{P}) \quad (17b)$$

where $\bar{R}_s(\mathbf{P})$ and $\bar{R}_s^{(M)}(\mathbf{P})$ are the equivalent ASRs. Hence, from (16), (17) and the continuity of the pointwise minimization, we have

$$\lim_{M \rightarrow \infty} \bar{R}_s^{(M)}(\mathbf{P}) = \bar{R}_s(\mathbf{P}), \text{ a.s. } \forall \mathbf{P} \in \mathbb{P}. \quad (18)$$

From the continuity of the ARs, we observe that $\bar{R}_s(\mathbf{P})$ is continuous in $\mathbf{P}$ although not necessarily differentiable at all points due to the embedded pointwise minimization. Hence, it can be shown that the convergence in (18) is also uniform. Combining these observations with [26, Theorem 5.3], it is concluded that the set of global optimum solutions of the SAA problem in (15) converges to that of the stochastic problem in (10) a.s. as $M \rightarrow \infty$. Extending this result to the set of points that satisfy the first-order optimality conditions (i.e. KKT points) can be found in [39, Section 4.3]. Now, we turn to solving the sampled problem in (15).

B. Augmented AWSMSE Minimization

Although (15) is deterministic, it is non-convex and very challenging to solve. Therefore, the approach proposed in [14] is employed to reformulate (15) into an equivalent augmented WMSE form. Augmented WMSE problems differ from their conventional WMSE counterparts (e.g. [16]) in two ways: 1) the weights are considered as optimization variables, 2) the cost function is extended to incorporate the logarithms of the weights. A Rate-WMMSE relationship is built upon the two aforementioned features.

Let $\hat{s}_{c,k} = g_{c,k} y_k$ be k th user's estimate of s_c , where $g_{c,k}$ is a scalar equalizer. After successfully removing the common stream, the estimate of s_k is obtained as $\hat{s}_k = g_k(y_k - \mathbf{h}_k^H \mathbf{p}_c s_{c,k})$, where g_k is the corresponding equalizer. At the output of the k th receiver, the common and private MSEs are defined as $\varepsilon_{c,k} \triangleq \mathbb{E}\{|\hat{s}_{c,k} - s_c|^2\}$ and $\varepsilon_k \triangleq \mathbb{E}\{|\hat{s}_k - s_k|^2\}$ respectively, which write as:

$$\varepsilon_{c,k} = |g_{c,k}|^2 T_{c,k} - 2\Re\{g_{c,k} \mathbf{h}_k^H \mathbf{p}_c\} + 1 \quad (19a)$$

$$\varepsilon_k = |g_k|^2 T_k - 2\Re\{g_k \mathbf{h}_k^H \mathbf{p}_k\} + 1 \quad (19b)$$

where $T_{c,k}$ and T_k are defined in (3). Optimum Minimum MSE (MMSE) equalizers are given as

$$g_{c,k}^{\text{MMSE}} = \mathbf{p}_c^H \mathbf{h}_k T_{c,k}^{-1} \quad \text{and} \quad g_k^{\text{MMSE}} = \mathbf{p}_k^H \mathbf{h}_k T_k^{-1} \quad (20)$$

obtained by solving $\frac{\partial \varepsilon_{c,k}}{\partial g_{c,k}} = 0$ and $\frac{\partial \varepsilon_k}{\partial g_k} = 0$. Substituting (20) into (19), the MMSEs write as

$$\varepsilon_{c,k}^{\text{MMSE}} \triangleq \min_{g_{c,k}} \varepsilon_{c,k} = T_{c,k}^{-1} I_{c,k} \quad (21a)$$

$$\varepsilon_k^{\text{MMSE}} \triangleq \min_{g_k} \varepsilon_k = T_k^{-1} I_k \quad (21b)$$

from which the SINRs are rewritten as $\gamma_{c,k} = (1/\varepsilon_{c,k}^{\text{MMSE}}) - 1$ and $\gamma_k = (1/\varepsilon_k^{\text{MMSE}}) - 1$, and the rates write as $R_{c,k} = -\log_2(\varepsilon_{c,k}^{\text{MMSE}})$ and $R_k = -\log_2(\varepsilon_k^{\text{MMSE}})$. Now, the augmented WMSEs are given by

$$\xi_{c,k} = u_{c,k} \varepsilon_{c,k} - \log_2(u_{c,k}) \quad \text{and} \quad \xi_k = u_k \varepsilon_k - \log_2(u_k) \quad (22)$$

where $u_{c,k}, u_k > 0$ are weights associated with the k th user's MSEs. In the following, $\xi_{c,k}$ and ξ_k are referred to as the WMSEs, where "augmented" is dropped for brevity. By taking the equalizers and weights as optimization variables, the Rate-WMMSE relationship is established as

$$\xi_{c,k}^{\text{MMSE}} \triangleq \min_{u_{c,k}, g_{c,k}} \xi_{c,k} = 1 - R_{c,k} \quad (23a)$$

$$\xi_k^{\text{MMSE}} \triangleq \min_{u_k, g_k} \xi_k = 1 - R_k \quad (23b)$$

which is obtained as follows. From $\frac{\partial \xi_{c,k}}{\partial g_{c,k}} = 0$ and $\frac{\partial \xi_k}{\partial g_k} = 0$, the optimum equalizers are given as $g_{c,k}^* = g_{c,k}^{\text{MMSE}}$ and $g_k^* = g_k^{\text{MMSE}}$. Substituting this back into (22) yields

$$\xi_{c,k}(g_{c,k}^{\text{MMSE}}) = u_{c,k} \varepsilon_{c,k}^{\text{MMSE}} - \log_2(u_{c,k}) \quad (24a)$$

$$\xi_k(g_k^{\text{MMSE}}) = u_k \varepsilon_k^{\text{MMSE}} - \log_2(u_k). \quad (24b)$$

Furthermore, from $\frac{\partial \xi_{c,k}(g_{c,k}^{\text{MMSE}})}{\partial u_{c,k}} = 0$ and $\frac{\partial \xi_k(g_k^{\text{MMSE}})}{\partial u_k} = 0$, we obtain the optimum MMSE weights given by $u_{c,k}^* = u_{c,k}^{\text{MMSE}} \triangleq (\varepsilon_{c,k}^{\text{MMSE}})^{-1}$ and $u_k^* = u_k^{\text{MMSE}} \triangleq (\varepsilon_k^{\text{MMSE}})^{-1}$, where a scaling factor of $(\ln(2))^{-1}$ has been omitted as it has no effect on the solution. Substituting this back into (24) yields the relationship in (23).

By taking the expectation over the conditional distribution of $\mathbf{H}$ given $\hat{\mathbf{H}}$, an average version of the Rate-WMMSE relationship, denoted by AR-AWMMSE, is expressed as

$$\bar{\xi}_{c,k}^{\text{MMSE}} \triangleq \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \left\{ \min_{u_{c,k}, g_{c,k}} \xi_{c,k} \mid \hat{\mathbf{H}} \right\} = 1 - \bar{R}_{c,k} \quad (25a)$$

$$\bar{\xi}_k^{\text{MMSE}} \triangleq \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \left\{ \min_{u_k, g_k} \xi_k \mid \hat{\mathbf{H}} \right\} = 1 - \bar{R}_k \quad (25b)$$

where the expectations are taken outside the minimizations to account for the dependencies of the optimum equalizers and weights on $\mathbf{H}$. The rationale behind employing the Rate-WMMSE relationship is explained as follows. The incorporation of the equalizers and weights unveils a block-wise convexity property, which can be checked by noting that the WMSEs are convex in each of their corresponding variables when all other variables are fixed. This will prove crucial for solving the problem.

To formulate a deterministic equivalent of (25), the Average WMSEs (AWMSEs) are approximated by their SAFs. The sampling extends to the equalizers and weights such that

$$\bar{\xi}_{c,k}^{(M)} \triangleq \frac{1}{M} \sum_{m=1}^M \xi_{c,k}^{(m)} \quad \text{and} \quad \bar{\xi}_k^{(M)} \triangleq \frac{1}{M} \sum_{m=1}^M \xi_k^{(m)} \quad (26)$$

where $\xi_{c,k}^{(m)} \triangleq \xi_{c,k}(\mathbf{h}_k^{(m)}, g_{c,k}^{(m)}, u_{c,k}^{(m)})$ and $\xi_k^{(m)} \triangleq \xi_k(\mathbf{h}_k^{(m)}, g_k^{(m)}, u_k^{(m)})$, $g_{c,k}^{(m)} \triangleq g_{c,k}(\mathbf{h}_k^{(m)})$ and $g_k^{(m)} \triangleq g_k(\mathbf{h}_k^{(m)})$, and $u_{c,k}^{(m)} \triangleq u_{c,k}(\mathbf{h}_k^{(m)})$ and $u_k^{(m)} \triangleq u_k(\mathbf{h}_k^{(m)})$, are all associated with the m th realization in $\mathbb{H}^{(M)}$. For compactness, we define the set of sampled equalizers as: $\mathbf{G} \triangleq \{\mathbf{g}_{c,k}, \mathbf{g}_k \mid k \in \mathcal{K}\}$, where $\mathbf{g}_{c,k} \triangleq \{g_{c,k}^{(m)} \mid m \in \mathcal{M}\}$ and $\mathbf{g}_k \triangleq \{g_k^{(m)} \mid m \in \mathcal{M}\}$. In a similar manner, we define: $\mathbf{U} \triangleq \{\mathbf{u}_{c,k}, \mathbf{u}_k \mid k \in \mathcal{K}\}$, where $\mathbf{u}_{c,k} \triangleq \{u_{c,k}^{(m)} \mid m \in \mathcal{M}\}$ and $\mathbf{u}_k \triangleq \{u_k^{(m)} \mid m \in \mathcal{M}\}$. The same approach used to prove (23) is employed to demonstrate the following relationship

$$\bar{\xi}_{c,k}^{\text{MMSE}(M)} \triangleq \min_{\mathbf{u}_{c,k}, \mathbf{g}_{c,k}} \bar{\xi}_{c,k}^{(M)} = 1 - \bar{R}_{c,k} \quad (27a)$$

$$\bar{\xi}_k^{\text{MMSE}(M)} \triangleq \min_{\mathbf{u}_k, \mathbf{g}_k} \bar{\xi}_k^{(M)} = 1 - \bar{R}_k \quad (27b)$$

where optimality conditions are checked separately for each conditional realization. The sets of optimum MMSE equalizers associated with (27) are defined as $\mathbf{g}_{c,k}^{\text{MMSE}} \triangleq \{g_{c,k}^{\text{MMSE}(m)} \mid m \in \mathcal{M}\}$ and $\mathbf{g}_k^{\text{MMSE}} \triangleq \{g_k^{\text{MMSE}(m)} \mid m \in \mathcal{M}\}$. In the same manner, the sets of optimum MMSE weights are defined as $\mathbf{u}_{c,k}^{\text{MMSE}} \triangleq \{u_{c,k}^{\text{MMSE}(m)} \mid m \in \mathcal{M}\}$ and $\mathbf{u}_k^{\text{MMSE}} \triangleq \{u_k^{\text{MMSE}(m)} \mid m \in \mathcal{M}\}$. For the K users, the MMSE solution is composed as $\mathbf{G}^{\text{MMSE}} \triangleq \{\mathbf{g}_{c,k}^{\text{MMSE}}, \mathbf{g}_k^{\text{MMSE}} \mid k \in \mathcal{K}\}$ and $\mathbf{U}^{\text{MMSE}} \triangleq \{\mathbf{u}_{c,k}^{\text{MMSE}}, \mathbf{u}_k^{\text{MMSE}} \mid k \in \mathcal{K}\}$. Motivated by the relationship in (27), the deterministic augmented AWSMSE minimization problem is formulated as

$$\mathcal{A}_{\text{RS}}^{(M)}(P_t) : \begin{cases} \min_{\bar{\xi}_c, \mathbf{P}, \mathbf{U}, \mathbf{G}} & \bar{\xi}_c + \sum_{k=1}^K \bar{\xi}_k^{(M)} \\ \text{s.t.} & \bar{\xi}_{c,k}^{(M)} \leq \bar{\xi}_c, \forall k \in \mathcal{K} \\ & \text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \end{cases} \quad (28)$$

where $\bar{\xi}_c$ represents the common AWMSE.

C. Equivalence

The fact that the AWMSEs are decoupled in their corresponding equalizers and weights suggests that optimizing (28) w.r.t $(\mathbf{U}, \mathbf{G})$ is achieved by minimizing each of the AWMSEs individually as shown in (27). This can be confirmed by showing that the MMSE solution $(\mathbf{U}^{\text{MMSE}}, \mathbf{G}^{\text{MMSE}})$ satisfies the KKT optimality conditions of (28) for a given $\mathbf{P}$. As a result, it is easily shown that under the MMSE solution, (28) boils down to (15) with affine transformations applied to the objective function and the common rate, i.e. $\mathcal{A}_{\text{RS}}^{(M)}(P_t) = K + 1 - \mathcal{R}_{\text{RS}}^{(M)}(P_t)$ and $\bar{\xi}_c = 1 - \bar{R}_c$. This Rate-WMMSE equivalence is not restricted to the global optimum solutions, and in fact can be extended to the whole set of stationary points. In particular, for any point $(\bar{\xi}_c^*, \mathbf{P}^*, \mathbf{U}^*, \mathbf{G}^*)$ satisfying the KKT optimality conditions of (28), the solution given by $(\bar{R}_c^* = 1 - \bar{\xi}_c^*, \mathbf{P}^*)$ satisfies the KKT optimality conditions

of (15). This can be demonstrated by following the steps in the proof of [40, Proposition 1], where the max-min fairness problem in a MIMO Interfering BC is considered. Despite the different setup, the absence of RS and the assumption of perfect CSIT in [40], the proofs are directly extendable to the problem considered here due to a key similarity: both objective functions are non-smooth and can be expressed as a max-min of sum-rate expressions. At this point, it is concluded that solving (28) yields a solution for (15), which in turn, converges to a solution for the ASR problem in (10) a.s. as $M \rightarrow \infty$.

D. Alternating Optimization Algorithm

Although problem (28) is non-convex in the joint set of optimization variables, it is convex in each of the blocks $\mathbf{P}$, $\mathbf{U}$ and $\mathbf{G}$, while fixing the other two. Moreover, $(\mathbf{G}, \mathbf{U})$ assume the closed-form MMSE solution for a given $\mathbf{P}$. These properties are exploited using an AO algorithm. Each iteration of the proposed algorithm consists of two steps: 1) updating $(\mathbf{G}, \mathbf{U})$ for a given $\mathbf{P}$, 2) updating $\mathbf{P}$ (and $\bar{\xi}_c$) for given $(\mathbf{G}, \mathbf{U})$. Next, each of the steps is described in detail.

1) *Updating the Equalizers and Weights:* In n th iteration of the AO algorithm, the equalizers and weights are updated such that $(\mathbf{G}, \mathbf{U}) = (\mathbf{G}^{\text{MMSE}}(\mathbf{P}^{[n-1]}), \mathbf{U}^{\text{MMSE}}(\mathbf{P}^{[n-1]}))$, where the dependency of the MMSE solution on a given precoder is highlighted in the notation, and $\mathbf{P}^{[n-1]}$ is the precoding matrix obtained in the $n-1$ th iteration. To facilitate the formulation of the precoder optimization problem in the following step, we introduce the SAFs listed as: $\bar{t}_{c,k}, \bar{t}_k, \bar{\Psi}_{c,k}, \bar{\Psi}_k, \bar{\mathbf{f}}_{c,k}, \bar{\mathbf{f}}_k, \bar{u}_{c,k}, \bar{u}_k, \bar{v}_{c,k}$ and $\bar{v}_k$, which are obtained using the updated $(\mathbf{G}, \mathbf{U})$. In particular, $\bar{u}_{c,k}$ and $\bar{u}_k$ are calculated by taking the ensemble averages over the M realizations of $u_{c,k}^{(m)}$ and $u_k^{(m)}$, respectively. The rest are calculated in a similar manner by averaging over their corresponding realizations given by:

$$\begin{aligned} t_{c,k}^{(m)} &= u_{c,k}^{(m)} \left| g_{c,k}^{(m)} \right|^2 & \text{and} & \quad t_k^{(m)} = u_k^{(m)} \left| g_k^{(m)} \right|^2 \\ \Psi_{c,k}^{(m)} &= t_{c,k}^{(m)} \mathbf{h}_k^{(m)} \mathbf{h}_k^{(m)H} & \text{and} & \quad \Psi_k^{(m)} = t_k^{(m)} \mathbf{h}_k^{(m)} \mathbf{h}_k^{(m)H} \\ \mathbf{f}_{c,k}^{(m)} &= u_{c,k}^{(m)} \mathbf{h}_k^{(m)} g_{c,k}^{(m)H} & \text{and} & \quad \mathbf{f}_k^{(m)} = u_k^{(m)} \mathbf{h}_k^{(m)} g_k^{(m)H} \\ v_{c,k}^{(m)} &= \log_2 \left(u_{c,k}^{(m)} \right) & \text{and} & \quad v_k^{(m)} = \log_2 \left(u_k^{(m)} \right). \end{aligned}$$

2) *Updating the Precoders:* Following the previous step, the problem of updating $\mathbf{P}$ is formulated by plugging $\mathbf{G}^{\text{MMSE}}(\mathbf{P}^{[n-1]})$ and $\mathbf{U}^{\text{MMSE}}(\mathbf{P}^{[n-1]})$ into (28). This yields

$$\begin{aligned} \mathcal{A}_{\text{RS}}^{[n]}(P_t) : \\ \left\{ \begin{array}{ll} \min_{\bar{\xi}_c, \mathbf{P}} & \bar{\xi}_c + \sum_{k=1}^K \left(\sum_{i=1}^K \mathbf{p}_i^H \bar{\Psi}_k \mathbf{p}_i + \sigma_n^2 \bar{t}_k \right. \\ & \left. - 2\Re\{\bar{\mathbf{f}}_k^H \mathbf{p}_k\} + \bar{u}_k - \bar{v}_k \right) \\ \text{s.t.} & \mathbf{p}_c^H \bar{\Psi}_{c,k} \mathbf{p}_c + \sum_{i=1}^K \mathbf{p}_i^H \bar{\Psi}_{c,k} \mathbf{p}_i + \sigma_n^2 \bar{t}_{c,k} \\ & - 2\Re\{\bar{\mathbf{f}}_{c,k}^H \mathbf{p}_c\} + \bar{u}_{c,k} - \bar{v}_{c,k} \leq \bar{\xi}_c, \quad \forall k \in \mathcal{K} \\ & \|\mathbf{p}_c\|^2 + \sum_{k=1}^K \|\mathbf{p}_k\|^2 \leq P_t. \end{array} \right. \quad (29) \end{aligned}$$

The expressions of the AWMSEs in (29) are obtained as follows. The MSEs in (19) are substituted into (22), followed by substituting the result into the M realizations in (26). This is carried out while considering the updated equalizers and

Algorithm 1 Alternating Optimization

```

1: Initialize:  $n \leftarrow 0, \mathcal{A}_{\text{RS}}^{[n]} \leftarrow 0, \mathbf{P}$ 
2: repeat
3:    $n \leftarrow n + 1, \mathbf{P}^{[n-1]} \leftarrow \mathbf{P}$ 
4:    $\mathbf{G} \leftarrow \mathbf{G}^{\text{MMSE}}(\mathbf{P}^{[n-1]}), \mathbf{U} \leftarrow \mathbf{U}^{\text{MMSE}}(\mathbf{P}^{[n-1]})$ 
5:   update  $\bar{\Psi}_{c,k}, \bar{\Psi}_k, \bar{\mathbf{F}}_{c,k}, \bar{\mathbf{F}}_k, \bar{t}_{c,k}, \bar{t}_k, \bar{u}_{c,k}, \bar{u}_k, \bar{v}_{c,k}, \bar{v}_k$ 
     for all  $k \in \mathcal{K}$ 
6:    $\mathbf{P} \leftarrow \arg \mathcal{A}_{\text{RS}}^{[n]}$ 
7: until  $|\mathcal{A}_{\text{RS}}^{[n]} - \mathcal{A}_{\text{RS}}^{[n-1]}| < \epsilon_R$ 

```

weights from the previous step, which are embedded in the SAFs listed above. Problem (29) is a convex Quadratically Constrained Quadratic Program (QCQP) which can be solved using interior-point methods [41].

3) *Algorithm and Convergence:* The two steps are repeated in an alternating manner until convergence as summarized in Algorithm 1, where ϵ_R is some arbitrarily small constant⁴.

Proposition 1. For a given $\mathbb{H}^{(M)}$, the iterates generated by Algorithm 1 converge to the set of KKT solutions of the corresponding sampled ASR problem in (15). As $M \rightarrow \infty$, the iterates converge, a.s., to the set of KKT solutions of the ASR problem in (10).

Proof: The sequence $\{\mathcal{A}_{\text{RS}}^{[n]}\}_{n=1}^{\infty}$ is monotonically decreasing, as each optimization step decreases the objective function. Since $\mathcal{R}_{\text{RS}}^{(M)}(P_t)$ is bounded above for a given power constraint, it follows that $\mathcal{A}_{\text{RS}}^{(M)}(P_t)$ is bounded below, and hence $\mathcal{A}_{\text{RS}}^{[n]}$ is guaranteed to converge. Next, we show that the corresponding sequence $\{\mathbf{P}^{[n]}\}_{n=1}^{\infty}$ converges to the set of KKT points of problem (15).

It can be seen from the relationship in (27) that for the n th iteration, problem $\mathcal{A}_{\text{RS}}^{[n]}(P_t)$ in (29) is in fact a convex approximation of the sampled ASR problem $\mathcal{R}_{\text{RS}}^{(M)}(P_t)$ in (15) around the point $\mathbf{P}^{[n-1]}$, obtained from the previous iteration. Hence, the AO procedure in Algorithm 1 in an instance of the Successive Convex Approximation (SCA) method in [42, Section 2.1]. Combining this with the conditions in [42, Assumption 1] and the fact that the iterates $\{\mathbf{P}^{[n]}\}_{n=1}^{\infty}$ lie in the compact feasible set $\mathbb{P}$, the convergence to the set of KKT points of problem (15) follows from [42, Theorem 1] and [43, Corollary 1]. The a.s. convergence to the set of KKT points of problem (10) as $M \rightarrow \infty$ follows from Section IV-A. ■

It should be highlighted that due to the non-convexity of the problem, the global optimality of the solution cannot be guaranteed in general, and different initializations of $\mathbf{P}$ may lead to different limit points. This is examined through simulations in Section VI-A where it is shown that appropriate initialization yields good convergence and rate performances. Moreover, the influence of varying the finite sample size M on the performance is investigated in Section VI-B3.

⁴Contrary to the algorithm in [24], Algorithm 1 applies the entire Monte-Carlo sample in each iteration. This comes at the cost of updating more equalizers and weights, yet liberates the convergence (and hence the number iterates in which $\mathbf{P}$ is updated) from the Monte-Carlo sample size. This is more suitable here as equalizers and weights are scalars obtained in closed-form, while updating $\mathbf{P}$ is the most expensive step in each iteration.

V. CONSERVATIVELY APPROXIMATED WMMSE ALGORITHM

In this section, we consider an alternative deterministic approximation of the AR-AWMMSE relationship in (25). This approximation is based on relaxing the dependencies of the equalizers and weights in (25) on the channel state $\mathbf{H}$. As a result, the minimizations are taken outside the expectations, and the new AWMSEs assume closed-form expressions [19], [20]. This eliminates the SAA and the need for a possibly large Monte-Carlo sample of conditional channel realizations. Moreover, the sets of equalizers and weights to be updated in the AO procedure reduce to two pairs for each user. However, this simplification comes at the expense of the achievable performance, as the approximation is in fact conservative.

A. Conservative Approximation

The the AWMSEs are approximated by $\hat{\xi}_{c,k} \triangleq \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \{\xi_{c,k} | \hat{\mathbf{H}}\}$ and $\hat{\xi}_k \triangleq \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \{\xi_k | \hat{\mathbf{H}}\}$, which are expressed as

$$\hat{\xi}_{c,k} = \hat{u}_{c,k} \left(|\hat{g}_{c,k}|^2 \bar{T}_{c,k} - 2\Re\{\hat{g}_{c,k} \hat{\mathbf{h}}_k^H \mathbf{p}_c\} + 1 \right) - \log_2(\hat{u}_{c,k}) \quad (30a)$$

$$\hat{\xi}_k = \hat{u}_k \left(|\hat{g}_k|^2 \bar{T}_k - 2\Re\{\hat{g}_k \hat{\mathbf{h}}_k^H \mathbf{p}_k\} + 1 \right) - \log_2(\hat{u}_k) \quad (30b)$$

where $(\hat{g}_{c,k}, \hat{g}_k)$ and $(\hat{u}_{c,k}, \hat{u}_k)$ are the relaxed equalizers and weights respectively, and $\bar{T}_{c,k}$ and $\bar{T}_k$ are given by

$$\bar{T}_{c,k} = \mathbf{p}_c^H (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{R}_{e,k}) \mathbf{p}_c + \bar{T}_k \quad (31a)$$

$$\bar{T}_k = \sum_{i=1}^K \mathbf{p}_i^H (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{R}_{e,k}) \mathbf{p}_i + \sigma_n^2. \quad (31b)$$

The optimum relaxed equalizers are derived from (30) as: $\hat{g}_{c,k}^{\text{MMSE}} = \mathbf{p}_c^H \hat{\mathbf{h}}_k \bar{T}_{c,k}^{-1}$ and $\hat{g}_k^{\text{MMSE}} = \mathbf{p}_k^H \hat{\mathbf{h}}_k \bar{T}_k^{-1}$. Plugging them back into (30), the optimum relaxed weights are obtained as $\hat{u}_{c,k}^{\text{MMSE}} = (\hat{\xi}_{c,k}^{\text{MMSE}})^{-1}$ and $\hat{u}_k^{\text{MMSE}} = (\hat{\xi}_k^{\text{MMSE}})^{-1}$, where $\hat{\xi}_{c,k}^{\text{MMSE}} \triangleq 1 - \bar{T}_{c,k}^{-1} |\hat{\mathbf{h}}_k^H \mathbf{p}_c|^2$ and $\hat{\xi}_k^{\text{MMSE}} \triangleq 1 - \bar{T}_k^{-1} |\hat{\mathbf{h}}_k^H \mathbf{p}_k|^2$. It is clear that the relaxed equalizers and weights are functions of $\hat{\mathbf{H}}$. Incorporating this approximation into the AR-AWMMSE relationship in (25), the expectations are taken inside the minimizations such that

$$\hat{\xi}_{c,k}^{\text{MMSE}} \triangleq \min_{\hat{u}_{c,k}, \hat{g}_{c,k}} \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \{\xi_{c,k} | \hat{\mathbf{H}}\} = 1 - \hat{R}_{c,k} \quad (32a)$$

$$\hat{\xi}_k^{\text{MMSE}} \triangleq \min_{\hat{u}_k, \hat{g}_k} \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \{\xi_k | \hat{\mathbf{H}}\} = 1 - \hat{R}_k \quad (32b)$$

where $\hat{R}_{c,k} \triangleq -\log_2(\hat{\xi}_{c,k}^{\text{MMSE}})$ and $\hat{R}_k \triangleq -\log_2(\hat{\xi}_k^{\text{MMSE}})$. The approximations in (32) are upper-bounds for the expressions in (25), i.e. $\hat{\xi}_{c,k}^{\text{MMSE}} \geq \xi_{c,k}^{\text{MMSE}}$ and $\hat{\xi}_k^{\text{MMSE}} \geq \xi_k^{\text{MMSE}}$. This follows from the fact that moving the expectation inside the minimization does not decrease the value. Hence, we have

$$\hat{R}_{c,k} \leq \bar{R}_{c,k} \quad \text{and} \quad \hat{R}_k \leq \bar{R}_k. \quad (33)$$

By employing (30), a conservative version of the AWSMSE problem is formulated as

$$\hat{\mathcal{A}}_{\text{RS}}(P_t) : \begin{cases} \min_{\hat{\xi}_c, \mathbf{P}, \hat{\mathbf{u}}, \hat{\mathbf{g}}} & \hat{\xi}_c + \sum_{k=1}^K \hat{\xi}_k \\ \text{s.t.} & \hat{\xi}_{c,k} \leq \hat{\xi}_c, \forall k \in \mathcal{K} \\ & \text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \end{cases} \quad (34)$$

where $\hat{\mathbf{u}} \triangleq \{\hat{u}_{c,k}, \hat{u}_k \mid k \in \mathcal{K}\}$ and $\hat{\mathbf{g}} \triangleq \{\hat{g}_{c,k}, \hat{g}_k \mid k \in \mathcal{K}\}$. Using the same argument in Section IV-B, it can be shown that the aforementioned MMSE solution is optimal for (34), and the cost function at any optimum point is an affine transformation of the conservative ASR given by $\min_j \{\hat{R}_{c,j}\}_{j=1}^K + \sum_{k=1}^K \hat{R}_k$. Problem (34) is solved by slightly modifying Algorithm 1. Specifically, $(\mathbf{G}, \mathbf{U})$ are reduced to $(\hat{\mathbf{g}}, \hat{\mathbf{u}})$, and step 5 is eliminated as (30) is directly employed to formulate the QCQP.

B. Conservative Performance Limitations

It can be seen that (34) is a restricted version of the AWSMSE problem in (28), obtained by restricting the domain of (28) such that the sets of equalizers and weights remain unchanged across the M realizations, and driving $M \rightarrow \infty$. By definition, solving (34) is equivalent to maximizing a lower-bound on the ASR. Hence, the conservative ESR given by $K + 1 - \mathbb{E}_{\hat{\mathbf{H}}} \{\hat{\mathcal{A}}_{\text{RS}}(P_t)\}$ is achievable. As the equalizers are updated using partial CSIT, the availability of highly accurate CSIR is completely ignored by the BS in the design (also called an ignorant approach [44]), which may lead to a loss in performance compared to the SAA method. This leaves us wondering about the effectiveness of the conservative approximation, i.e. the looseness of the lower-bound. A similar ignorant approach was used for worst-case robust designs under bounded errors, where it was observed that the rate lower-bound fails to achieve the anticipated DoF [45]. This was analysed in [29], where it was shown that the loss in performance is due to self-interference terms induced by the approximation. The same argument applies here, where it can be seen that the equivalent conservative SINRs, defined as $\hat{\gamma}_{c,k} = (1 - \hat{\xi}_{c,k}^{\text{MMSE}}) / \hat{\xi}_{c,k}^{\text{MMSE}}$ and $\hat{\gamma}_k = (1 - \hat{\xi}_k^{\text{MMSE}}) / \hat{\xi}_k^{\text{MMSE}}$, have the self-interference terms, given by $\mathbf{p}_c^H \mathbf{R}_{e,k} \mathbf{p}_c$ and $\mathbf{p}_k^H \mathbf{R}_{e,k} \mathbf{p}_k$ respectively, within their denominators. Such terms may scale significantly with increased power, depending on the CSIT quality, hence limiting the rate performance as we observe in Section VI-B4.

It is worth noting that while relaxing the dependencies of equalizers on $\mathbf{H}$ is carried out by the BS to simplify the design algorithm, is not necessarily followed by receivers which are still expected to apply perfect CSIR to update their equalizers. However, predicting the possibly higher ERs (achievable by the users) at the BS, and hence adjusting the transmission rates, requires the evaluation of the expressions in (8) and (11), which in turn requires numerical calculations that bring us back to Monte-Carlo samplings. Hence for this approach, we assume that the BS transmit at the conservative ERs.

VI. NUMERICAL RESULTS AND ANALYSIS

In this section, the proposed algorithms are evaluated through simulations. The channel $\mathbf{H}$ has i.i.d. complex Gaussian entries with unit variance, i.e. $\mathcal{CN}(0, 1)$. The noise variance is fixed as $\sigma_n^2 = 1$, from which the long-term SNR is given as P_t . Entries of $\hat{\mathbf{H}}$ are also i.i.d complex Gaussian drawn from $\mathcal{CN}(0, \sigma_e^2)$, where $\sigma_e^2 = N_t^{-1} \sigma_{e,k}^2, \forall k \in \mathcal{K}$. The error variance is given as $\sigma_e^2 = \beta P_t^{-\alpha}$, where $\beta \geq 0$ and $\alpha \in [0, 1]$ are varied to represent different CSIT accuracies and

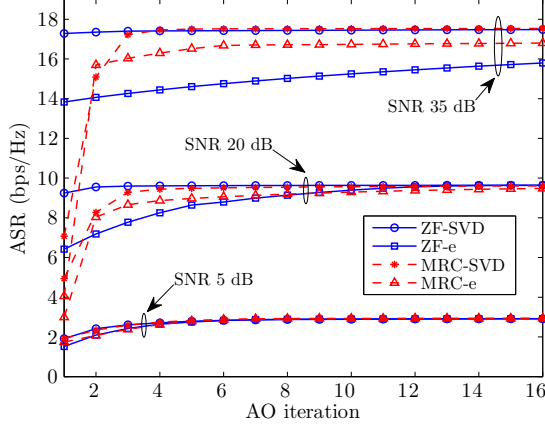


Fig. 1. ASR convergence of Algorithm 1 using four different initializations of $\mathbf{P}$ for 1 randomly generated channel estimate, $K, N_t = 2$, $\sigma_e^2 = P_t^{-0.6}$, and SNR = 5, 20 and 35 dB.

SNR scalings. It follows that the channel estimate $\hat{\mathbf{H}} = \mathbf{H} - \tilde{\mathbf{H}}$ is also Gaussian. The channel realization $\mathbf{H}$ should not be confused with a conditional realization $\mathbf{H}^{(m)}$. While the former represents the actual channel experienced by the users and unknown to the BS, the latter is part of a sample $\mathbb{H}^{(M)}$ available at the BS and used to calculate the SAFs. The size of the sample is set to $M = 1000$ throughout the simulations, unless otherwise stated. For a given estimate $\hat{\mathbf{H}}$, the m th conditional realization is obtained as $\mathbf{H}^{(m)} = \hat{\mathbf{H}} + \tilde{\mathbf{H}}^{(m)}$, where $\tilde{\mathbf{H}}^{(m)}$ is drawn from the error distribution. More on generating the channel realizations in the simulations is given in Appendix B. Convex optimization problems are solved using the CVX toolbox [46].

A. Convergence

First, we examine the convergence of Algorithm 1 under four different initializations of $\mathbf{P}$. The first initialization, denoted by ZF-e, is taken as the DoF-motivated design in [7], [8]. To recall, we have $\mathbf{p}_c = \sqrt{q_c} \hat{\mathbf{p}}_c$ and $\mathbf{p}_k = \sqrt{q_k} \hat{\mathbf{p}}_k$, where $q_c = P_t - P_t^\alpha$ and $q_k = P_t^\alpha / K$, $\hat{\mathbf{p}}_c = \mathbf{e}_1$ and $\{\hat{\mathbf{p}}_1, \dots, \hat{\mathbf{p}}_K\}$ are normalized ZF-BF vectors constructed using $\hat{\mathbf{H}}$. The second initialization, labeled as ZF-SVD, is a modification of ZF-e, where $\hat{\mathbf{p}}_c$ is slightly optimized by choosing it as the dominant left singular vector of $\hat{\mathbf{H}}$. The third and fourth initializations, labeled as MRC-e and MRC-SVD respectively, maintain the same common precoders as ZF-e and ZF-SVD respectively. However, Maximum Ratio Combining (MRC) is employed instead of ZF-BF, i.e. $\hat{\mathbf{p}}_k = \hat{\mathbf{h}}_k / \|\hat{\mathbf{h}}_k\|$. The ASR convergence of Algorithm 1 for $\sigma_e^2 = P_t^{-0.6}$ and SNRs 5, 20 and 35 dB is shown in Fig. 1. It is evident that the algorithm converges to a limit point regardless of the initialization. However, the speed of convergence is influenced by the initial state, which may also influence the limit point as the problem is non-convex. The initialization effect becomes more visible as SNR grows large. For example, initializing the common precoder using SVD enhances the convergence at high SNRs. In the following results, MRC-SVD is adopted as it provides good overall performance over various channel realizations and a

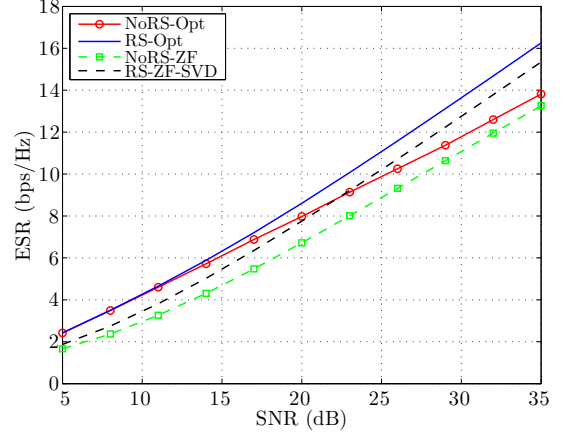


Fig. 2. Comparison between the ESR performances of NoRS and RS transmission schemes for $K, N_t = 2$ and $\sigma_e^2 = P_t^{-0.6}$.

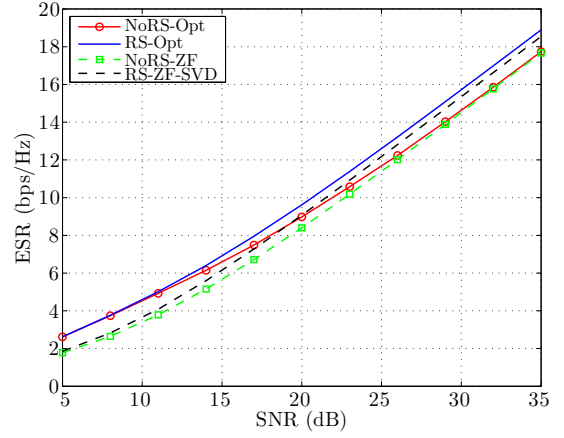


Fig. 3. Comparison between the ESR performances of NoRS and RS transmission schemes for $K, N_t = 2$ and $\sigma_e^2 = P_t^{-0.9}$.

wide range of SNRs.

B. Ergodic Sum Rate Performance

Here we evaluate the achievable ESRs obtained by averaging the ASRs over 100 channel realizations. For each channel realization, Algorithm 1 is used to obtain the RS ASR, and a modification of the algorithm where the common precoder is switched off is used for NoRS.

1) *RS vs. NoRS*: The performance of the optimized RS scheme (RS-Opt) is compared to the optimized NoRS scheme (NoRS-Opt). ZF-BF with Water-Filling (WF), where power allocation is carried out assuming that the estimate $\hat{\mathbf{H}}$ is perfect, is considered as a baseline for NoRS transmission. This scheme is termed as NoRS-ZF. On the other hand, we consider a modified version of the ZF-SVD initialization in the previous subsection as a baseline for RS transmission. In particular, the power splitting between the common message and the private messages is maintained, while WF is used to allocate the power among the private messages. This scheme is termed RS-ZF-SVD.

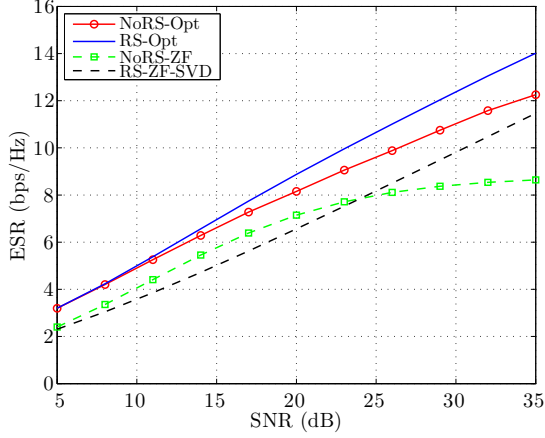


Fig. 4. Comparison between the ESR performances of NoRS and RS transmission schemes for $K, N_t = 2$ and $\sigma_e^2 = 0.063$.

The 2-user ESRs obtained using the different optimization schemes for $\sigma_e^2 = P_t^{-0.6}$ and $P_t^{-0.9}$ are shown in Fig. 2 and Fig. 3, respectively. Comparing RS-Opt and NoRS-Opt, the two schemes perform similarly at low SNRs. In particular, both reduce to SU transmission by switching off the weaker user. As SNR grows, MU transmission starts to take over. However, reducing MU interference to the level of noise (or eliminating it) is not possible due to imperfect CSIT. Therefore, the overall performance does not necessarily benefit from additional power inaccurately directed towards a given user, as it may cause more damage by interfering with other users. At this stage, RS-Opt starts to part from NoRS-Opt by switching on the common message. The contribution of the common message primarily manifests at high SNRs, where the gap between RS-Opt and NoRS-Opt exceeds 4 dB for $\alpha = 0.6$. For $\alpha = 0.9$, which is almost perfect from a DoF perspective, RS is not as instrumental as it is for $\alpha = 0.6$. However, rate gains can still be observed at high SNRs. Regarding the baseline schemes, their inferiority compared to optimized schemes is evident over the entire SNR range. In the low SNR regime, baselines fall behind due to the strict application of ZF-BF, which is not ideal in this case. However, the RS-ZF-SVD scheme slightly compensates for this through the common message. However, it fails to match the optimized transmission in RS-Opt and NoRS-Opt. In the high SNR regime, each baseline scheme achieves the same DoF (slope of the curve) as its optimized counterpart with a rate gap, which is particularly noticeable for $\alpha = 0.6$.

The performance under CSIT errors that do not scale with SNR ($\alpha = 0$) is shown in Fig. 4, where $\sigma_e^2 = (10^{\frac{20}{10}})^{-0.6} = 0.063$, which corresponds to the CSIT quality obtained at 20 dB SNR w.r.t Fig. 2. First, it can be seen that the rate of NoRS-ZF saturates. This is due to the naive employment of $\hat{\mathbf{H}}$ to design the ZF-BF vectors and allocate power as if it was a perfect estimate. At high SNRs, residual interference dominates and caps the performance. RS-ZF-SVD employs the DoF-motivated power splitting (i.e. $q_c = P_t$ and $q_k = 0$) and hence reduces to multicast transmission (sending only the common message) maintaining a DoF of 1. On the other hand,

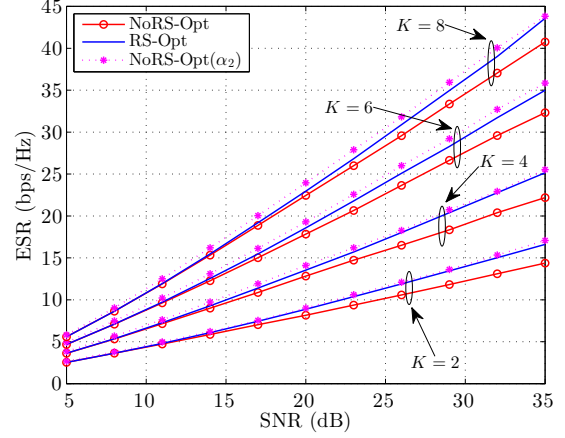


Fig. 5. ESR performances of NoRS-Opt and RS-Opt with $\sigma_e^2 = P_t^{-\alpha}$, and NoRS-Opt with $\sigma_e^2 = P_t^{-\alpha_2}$, where $\alpha_2 = \frac{1+(K-1)\alpha}{K}$, for $\alpha = 0.6$, and $K, N_t = 2, 4, 6$ and 8 .

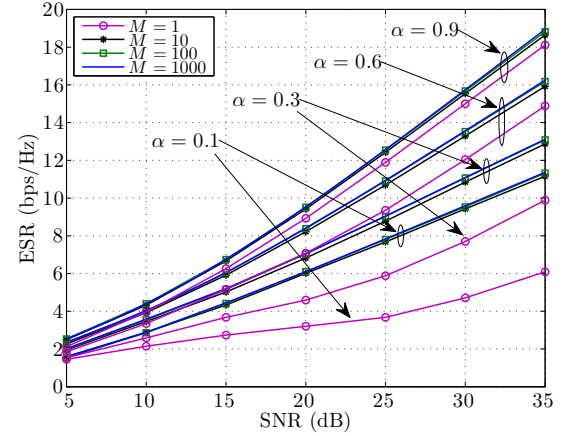


Fig. 6. The effect of changing M on the ESR performance of the RS SAA algorithm for $K, N_t = 2$, $\sigma_e^2 = P_t^{-\alpha}$, and $\alpha = 0.1, 0.3, 0.6$ and 0.9 .

NoRS-Opt and RS-Opt schemes achieve significantly better performances compared to their corresponding baselines, over the entire range of SNRs. Moreover, although RS-Opt is not expected to achieve DoF gains over NoRS-Opt as $P_t \rightarrow \infty$, the former still manages to deliver superior rate performance at medium and high SNRs.

2) *Increased Number of Users*: Theorem 1 implies that the relative DoF gain achieved by RS over NoRS decreases as K increases. Fig. 5 shows the ESRs achieved by NoRS-Opt and RS-Opt for $K = 2, 4, 6$ and 8 users, assuming $\alpha = 0.6$. Fig. 5 also shows the NoRS-Opt performance for a CSIT error exponent given as: $\alpha_2 = \frac{1+(K-1)\alpha}{K}$. This corresponds to a conventional system that achieves the same sum DoF as the RS system, but at the cost of higher CSIT quality requirements since $\alpha_2 > \alpha$, $\forall \alpha \in (0, 1)$. It is evident that the two perform closely, which highlights another benefit of RS, i.e. relaxed CSIT requirements compared to NoRS.

3) *SAA Sample Size*: Next, we look at the influence of changing the sample size M on the SAA algorithm. The 2-user ESRs achieved by solving Algorithm 1 for $M = 1, 10, 100$ and

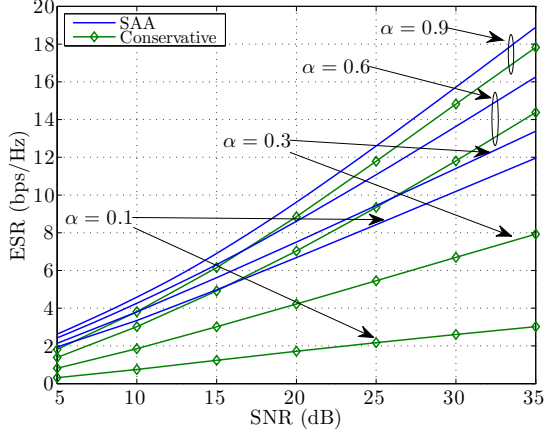


Fig. 7. ESR performances of the SAA and conservative approaches for $K, N_t = 2$, $\sigma_e^2 = P_t^{-\alpha}$, and $\alpha = 0.1, 0.3, 0.6$ and 0.9 .

1000 are given in Fig. 6. Different scaling CSIT qualities are considered where $\sigma_e^2 = P_t^{-\alpha}$, and $\alpha = 0.1, 0.3, 0.6$ and 0.9 . Using a sample of one realization significantly degrades the performance, as the resulting design lacks statistical knowledge of the CSIT uncertainty. On the other hand, the SAA algorithm performs very well with a sample size as small as $M = 10$ under the specified settings. Higher sample sizes of $M = 100$ and 1000 perform almost identically. Note that the achievable ESRs in this part are obtained by numerically evaluating the right-most expressions in (8) and (11), as taking the expectations of the sampled ASR objective functions for low M s is not reflective of the achievable performance.

4) *Conservative Algorithm:* In this part, we look at the performance of the conservative design in Section V. For the Gaussian CSIT error described earlier in this section, we have $\mathbf{R}_{e,k} = \sigma_e^2 \mathbf{I}$, $\forall k \in \mathcal{K}$. The conservative ESR is obtained by averaging the conservative ASRs obtained by solving (34). The ESRs predicted by the BS using the SAA approach and the conservative approach are given in Fig. 7 for $K = 2$ and $\sigma_e^2 = P_t^{-\alpha}$, with $\alpha = 0.1, 0.3, 0.6$ and 0.9 . The SAA scheme achieves significant gains over the conservative scheme, which increase with decreased CSIT qualities. This is explained by the discussion in Section V-B. In particular, self-interference terms become dominant and the ASR lower-bound becomes looser. This effect is extremely detrimental for low CSIT qualities, as the achievable ESRs are highly undermined by the BS.

C. Ergodic Rate Region

In the final part, we investigate the NoRS and RS achievable ER regions in a two-user scenario. ER regions are obtained by solving Weighted ESR (WESR) problems, where different boundary points are realized by varying the weights. For NoRS, this is achieved by incorporating the weights into the ASR problem in (9). For a fixed pair of weights, the problem is solved for several channel realizations. A boundary ER region point is obtained by averaging the resulting AR realizations for a given pair of weights. For the RS problem, applying the weights directly to (10) results in a region with the common

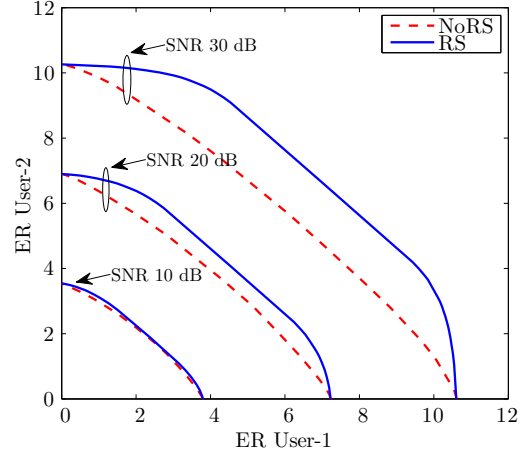


Fig. 8. 2-user ER regions for $\sigma_e^2 = P_t^{-0.6}$, and SNR = 10, 20 and 30 dB.

message allocated to one user the entire time, e.g. $k_{RS} = 1$. A second region is obtained by allocating the common message to the other user, and the full region is given by the convex-hull enclosing the two regions by time-sharing of the extremity points. Alternatively, all boundary points for the RS region can be obtained by assuming that $\bar{R}_c$ is shared between users such that $\bar{C}_k$ is the k th user's portion of the common AR with $\sum_{k=1}^K \bar{C}_k = \bar{R}_c$. The WASR problem is formulated as

$$\mathcal{R}_{WRS}(P_t) : \begin{cases} \max_{\mathbf{P}, \{\bar{C}_k\}_{k=1}^K} & \sum_{k=1}^K w_k (\bar{R}_k + \bar{C}_k) \\ \text{s.t.} & \bar{R}_{c,k} \geq \sum_{j=1}^K \bar{C}_j, \forall k \in \mathcal{K} \\ & \bar{C}_k \geq 0, \forall k \in \mathcal{K} \\ & \text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \end{cases} \quad (35)$$

where the weights $\{w_k \mid k \in \mathcal{K}\}$ are fixed parameters. (35) is solved by formulating its equivalent AWSMSE problem, and modifying Algorithm 1 accordingly. To obtain the two-user regions shown in Fig. 8, the corresponding problems are solved for 43 different pairs of weights. The first weight is fixed as $w_1 = 1$, and the second weight changes as $w_2 \in \{10^{-3}, 10^{-1}, 10^{-0.95}, \dots, 10^{0.95}, 10, 10^3\}$. What is meaningful for each pair is the ratio between the weights rather than their absolute levels [14]. Each point on the ER region is characterized by the tuple $E_{\hat{\mathbf{H}}} \{(\bar{R}_1, \bar{R}_2)\}$ for NoRS, and $E_{\hat{\mathbf{H}}} \{(\bar{R}_1 + \bar{C}_1, \bar{R}_2 + \bar{C}_2)\}$ for RS, where each pair of ARs is obtained from solving the corresponding WASR problem⁵.

The ER regions shown in Fig. 8 are obtained for SNRs: 10, 20 and 30 dB, and $\sigma_e^2 = P_t^{-0.6}$. As expected, the gap between RS and NoRS grows with SNR. The performance for 30 dB SNR is particularly interesting, as it is evident that RS enlarges the whole ER region significantly. For example, while guaranteeing an ER of 10 bps/Hz for user-1, an ER of almost 4 bps/Hz can be achieved by user-2. On the other hand, guaranteeing the same user-1 ER using NoRS restricts the ER of user-2 to just over 1.5 bps/Hz. This observation is of special interest for designs exploiting different points of the rate regions, e.g. max-min fairness and Quality of Service

⁵More accurately, we take the convex hulls enclosing the resulting points.

(QoS) based designs as seen in [28], [29].

VII. CONCLUSION

The problem of ESR maximization in MU-MISO systems with linear precoding and partial CSIT was addressed by employing the RS transmission strategy, where part of the MU interference is broadcasted to (and decoded by) all users. The precoding matrix is optimized based on the available channel estimate to maximized a conditional ASR metric, computed using partial CSIT knowledge. The stochastic ASR problem was transformed into a deterministic counterpart using the SAA method. This was followed by applying the Rate-WMMSE transformation and an AO algorithm, which converges to a stationary solution. An alternative deterministic approximation of the ASR problem was developed based on the conservative method in [19], [20]. While this approach avoids the sampling process, its guaranteed achievable performance is shown to be inferior. The effectiveness of the proposed algorithms and the benefits of adopting the RS strategy were demonstrated through simulations. In particular, it was shown that the gains anticipated by the DoF analysis can be realized at finitely high SNRs where RS designs outperform their NoRS counterparts. Moreover, additional benefits of RS including relaxed CSIT quality requirements were highlighted. Finally, the two-user ER region was numerically obtained by solving a sequence of WASR problems. RS proved to enlarge the entire achievable ER region, drawing attention to the potentials of employing RS in QoS constrained designs.

APPENDIX A PROOF OF THEOREM 1

Since $\mathbf{P}$ is separately designed for each $\hat{\mathbf{H}}$, we focus on a precoding scheme defined for a given $\hat{\mathbf{H}}$ as $\{\mathbf{P}\}_{P_t}$. The associated powers allocated to the common precoder and the k th private precoder are given as: $q_c \triangleq \|\mathbf{p}_c\|^2$ and $q_k \triangleq \|\mathbf{p}_k\|^2$ respectively, which scale with increased P_t as $O(P_t^{a_c})$ and $O(P_t^{a_k})$ respectively, where $a_c, a_k \in [0, 1]^6$. The streams interfering with the k th user are dominated by a power scaling factor of $\bar{a}_k \triangleq \max_{j \neq k} \{a_j\}$. We define the corresponding k th user's conditional average DoFs as

$$\bar{d}_{c,k} \triangleq \lim_{P_t \rightarrow \infty} \frac{\bar{R}_{c,k}(P_t)}{\log_2(P_t)} \quad \text{and} \quad \bar{d}_k \triangleq \lim_{P_t \rightarrow \infty} \frac{\bar{R}_k(P_t)}{\log_2(P_t)}. \quad (36)$$

For a given precoding scheme defined over all $\hat{\mathbf{H}}$, the k th user's long-term achievable DoFs are given by $d_{c,k} \triangleq \lim_{P_t \rightarrow \infty} \frac{E_{\hat{\mathbf{H}}} \{\bar{R}_{c,k}(P_t)\}}{\log_2(P_t)} = E_{\hat{\mathbf{H}}} \{\bar{d}_{c,k}\}$ and $d_k \triangleq \lim_{P_t \rightarrow \infty} \frac{E_{\hat{\mathbf{H}}} \{\bar{R}_k(P_t)\}}{\log_2(P_t)} = E_{\hat{\mathbf{H}}} \{\bar{d}_k\}$, which follows from the bounded convergence theorem. The same definitions extend to NoRS precoding schemes while discarding the common part. Without loss of generality, $\sigma_n^2 = 1$ is assumed throughout the proof. Each of the results in (13) and (14) is obtained through two steps. First, we show that the DoF is upper-bounded by its corresponding result. Then, we show that the upper-bounds are achievable via feasible precoding schemes. Before proceeding, we define the function $(x)^+ \triangleq \max\{x, 0\}$.

⁶Note that $a_c, a_k < 0$ can be replaced with $a_c, a_k = 0$ without influencing the DoF results derived next.

1) *Proof of (13)*: For an arbitrary precoding scheme, let us first find an upper-bound for

$$\bar{R}_k = E_{\mathbf{H}|\hat{\mathbf{H}}} \{ \log_2(T_k) - \log_2(I_k) \mid \hat{\mathbf{H}} \} \quad (37)$$

by upper-bounding and lower-bounding the first and second right-hand-side terms respectively. From Jensen's inequality, we write $E_{\mathbf{H}|\hat{\mathbf{H}}} \{ \log_2(T_k) \mid \hat{\mathbf{H}} \} \leq \log_2(\bar{T}_k)$, where $\bar{T}_k$ is defined in (31b). From the Cauchy-Schwarz inequality and the isotropic property of the CSIT errors, we have $\bar{T}_k \leq (\|\hat{\mathbf{h}}_k\|^2 + \sigma_e^2) \sum_{i=1}^K q_i + 1$, where $\sigma_e^2 = N_t^{-1} \sigma_{e,k}^2$. Since the actual channel state $\mathbf{H}$ does not depend on the SNR, we have $\|\mathbf{h}_k\|^2, \|\hat{\mathbf{h}}_k\|^2 = O(1)$. It follows that $\bar{T}_k \leq O(P_t^{\max\{a_k, \bar{a}_k\}})$, from which we write

$$E_{\mathbf{H}|\hat{\mathbf{H}}} \{ \log_2(T_k) \mid \hat{\mathbf{H}} \} \leq \max\{a_k, \bar{a}_k\} \log_2(P_t) + O(1). \quad (38)$$

From the isotropic property and [7, Lemma 1], we write $E_{\mathbf{H}|\hat{\mathbf{H}}} \{ \log_2(I_k) \mid \hat{\mathbf{H}} \} \geq \log_2(2^\kappa \sigma_e^2 \lambda_1 + 1) + O(1)$, where $\kappa \triangleq E_{\mathbf{H}|\hat{\mathbf{H}}} \left\{ \log_2 \left(\frac{|\mathbf{e}_i^T \tilde{\mathbf{h}}_k|^2}{\sigma_e^2} \right) \mid \hat{\mathbf{H}} \right\}$ is bounded, and λ_1 is the dominant eigenvalue of $\sum_{i \neq k} \mathbf{p}_i \mathbf{p}_i^H$. Since the maximum is lower-bounded by the average, we have $\lambda_1 \geq N_t^{-1} \sum_{i \neq k} q_i = O(P_t^{\bar{a}_k})$ from which we obtain

$$E_{\mathbf{H}|\hat{\mathbf{H}}} \{ \log_2(I_k) \mid \hat{\mathbf{H}} \} \geq (\bar{a}_k - \alpha)^+ \log_2(P_t) + O(1). \quad (39)$$

Form (36), (37), (38) and (39), we have $\bar{d}_k \leq \max\{a_k, \bar{a}_k\} - \max\{\bar{a}_k - \alpha, 0\}$. When $a_k \geq \bar{a}_k$, the upper-bound is given by $\min\{\alpha + a_k - \bar{a}_k, a_k\}$, otherwise we have 0. Combining the two cases, we write $\bar{d}_k \leq \min\{(\alpha + a_k - \bar{a}_k)^+, a_k\}$, from which the sum DoF is upper-bounded by

$$\sum_{k=1}^K \bar{d}_k \leq \sum_{k=1}^K \min\{(\alpha + a_k - \bar{a}_k)^+, a_k\}. \quad (40)$$

To obtain the maximum upper-bound for (40), we define $\mathcal{J} \subseteq \mathcal{K}$ as the subset composed of all users with non-zero DoF. Due to the symmetry in the CSIT qualities, it is sufficient to assume, without loss of generality, that $\mathcal{J} \triangleq \{1, \dots, J\}$ with $J \triangleq |\mathcal{J}|$. We define $\bar{d}(J) \triangleq \sum_{j=1}^J \bar{d}_j$. It is evident that $\bar{d}(1) \leq 1$. On the other hand, we have $\bar{d}(J) \leq J\alpha$ for $J > 1$. This is shown from

$$\sum_{k=1}^J \bar{d}_k \leq J\alpha + \sum_{k=1}^J (a_k - \bar{a}_k) \quad (41a)$$

$$\leq J\alpha + (a_J - a_1) + \sum_{k=1}^{J-1} (a_k - a_{k+1}) \quad (41b)$$

where (41a) follows from $\min\{x, y\} \leq x, y$, and the strict positivity of the DoF in $\mathcal{J}$. (41b) follows from $\bar{a}_k \geq a_j, \forall j \neq k$, and is equal to $J\alpha$. We conclude that the maximum upper-bound is obtained when $J = K$ for $\alpha \geq K^{-1}$, and $J = 1$ otherwise. Hence, $\sum_{k=1}^K \bar{d}_k \leq \max\{1, K\alpha\}$. This is also an upper-bound for $\lim_{P_t \rightarrow \infty} \frac{E_{\hat{\mathbf{H}}} \{\bar{R}(P_t)\}}{\log_2(P_t)}$, which follows from the bounded convergence theorem.

This part of the proof is completed by showing that the upper-bound can be achieved by a feasible precoding scheme. In particular, a DoF of 1 is achieved by Single User (SU) transmission in a Time Division Multiple Access (TDMA)

fashion. On the other hand, $K\alpha$ is achieved using ZF-BF vectors designed using $\hat{\mathbf{H}}$, and allocating the powers such that $q_k = P_t^\alpha/K$, for all $k \in \mathcal{K}$ as shown in [7].

2) *Proof of (14)*: From the definitions of the ARs, we write

$$\bar{R}_c + \bar{R}_1 \leq \bar{R}_{c,1} + \bar{R}_1 = E_{\mathbf{H}|\hat{\mathbf{H}}} \{ \log_2(T_{c,1}) - \log_2(I_1) \mid \hat{\mathbf{H}} \}.$$

Following the same approach used in the previous part, we obtain $E_{\mathbf{H}|\hat{\mathbf{H}}} \{ \log_2(T_{c,1}) \mid \hat{\mathbf{H}} \} \leq \log_2(P_t) + O(1)$, and $E_{\mathbf{H}|\hat{\mathbf{H}}} \{ \log_2(I_1) \mid \hat{\mathbf{H}} \} \geq (\bar{a}_1 - \alpha)^+ \log_2(P_t) + O(1)$, from which we obtain the following upper-bound

$$\bar{d}_c + \bar{d}_1 \leq \bar{d}_{c,1} + \bar{d}_1 \leq \min\{1 + \alpha - \bar{a}_1, 1\}. \quad (42)$$

Next, we define $\bar{d}_{\text{RS}}(J) \triangleq \bar{d}_c + \sum_{k=1}^J \bar{d}_k$ for J users with strictly positive DoFs. From (42), we have $\bar{d}_{\text{RS}}(1) \leq 1$. For $J = 2$, we have $\bar{d}_{\text{RS}}(2) \leq 1 + \alpha - \bar{a}_1 + a_2 \leq 1 + \alpha$, obtained from (42), $\bar{d}_2 \leq a_2$, and $\bar{a}_1 \geq a_2$. For $J > 2$, we have $\bar{d}_{\text{RS}}(J) \leq 1 + (J-1)\alpha$ obtained by combining (42) and (41). Hence, the maximum upper-bound is $1 + (K-1)\alpha$ obtained when $J = K$ regardless of α . It follows that this is also an upper-bound for $\lim_{P_t \rightarrow \infty} \frac{E_{\hat{\mathbf{H}}} \{ \mathcal{R}_{\text{RS}}(P_t) \}}{\log_2(P_t)}$.

Finally, this is achieved using the DoF-motivated design in [7]. The private precoders are given as shown in the previous part, while $q_c = P_t - P_t^\alpha$ and $\mathbf{p}_c = \sqrt{q_c} \mathbf{e}_1$.

APPENDIX B GENERATING CHANNEL STATES

For a given channel realization, a normalized channel estimate denoted by $\hat{\mathbf{H}}_n$ is generating with entries drawn from $\mathcal{CN}(0, 1)$. Moreover, a set of M normalized estimation errors defined as $\tilde{\mathbf{H}}_n^{(M)} \triangleq \{ \tilde{\mathbf{H}}_n^{(m)} \mid m \in \mathcal{M} \}$ is generated in the same manner. This is followed by constructing $\mathbb{H}^{(M)}$, where the m th conditional realization is calculated as $\mathbf{H}^{(m)} = \sqrt{1 - \sigma_e^2} \hat{\mathbf{H}}_n + \sqrt{\sigma_e^2} \tilde{\mathbf{H}}_n^{(m)}$. It is evident that the fading assumptions made at the beginning of Section VI are preserved. For $\alpha > 0$, i.e. σ_e^2 relies on SNR, the set $\mathbb{H}^{(M)}$ varies with SNR. To produce the ESR curves, averaging is carried out over 100 realizations of $\hat{\mathbf{H}}_n$, where $\tilde{\mathbf{H}}_n^{(M)}$ is reused across realizations. Reusing the same normalized channels across SNRs in this manner produces the smooth ESR curves shown in the figures.

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] 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, Jul. 2003.
- [3] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.
- [4] G. Caire, N. Jindal, and S. Shamai, "On the required accuracy of transmitter channel state information in multiple antenna broadcast channels," in *Proc. Asilomar*, Nov. 2007, pp. 287–291.
- [5] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [6] 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, Jun. 2010.
- [7] 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, 2013.
- [8] C. Hao and B. Clerckx, "MISO BC with imperfect and (un)matched CSIT in the frequency domain: DoF region and transmission strategies," in *Proc. IEEE PIMRC*, Sep. 2013, pp. 1–6.
- [9] R. Tandon, S. Jafar, S. Shamai Shitz, and H. Poor, "On the synergistic benefits of alternating CSIT for the MISO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4106–4128, Jul. 2013.
- [10] J. Chen and P. Elia, "Optimal DoF region of the two-user MISO-BC with general alternating CSIT," in *Proc. Asilomar*, Nov. 2013, pp. 1860–1864.
- [11] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.
- [12] 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, Sep. 2015.
- [13] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate maximization in multi-antenna broadcast channels with linear preprocessing," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2338–2342, Sep. 2006.
- [14] S. Christensen, R. Agarwal, E. Carvalho, and J. 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.
- [15] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.
- [16] M. Shenouda and T. Davidson, "On the design of linear transceivers for multiuser systems with channel uncertainty," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 1015–1024, Aug. 2008.
- [17] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, Sep. 2009.
- [18] T. Bogale, B. Chalise, and L. Vandendorpe, "Robust transceiver optimization for downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 446–453, Jan. 2011.
- [19] M. Bashar, Y. Lejosne, D. Slock, and Y. Yuan-Wu, "MIMO broadcast channels with Gaussian CSIT and application to location based CSIT," in *Proc. ITA*, Feb. 2014, pp. 1–7.
- [20] F. Negro, I. Ghauri, and D. T. M. Slock, "Sum Rate maximization in the noisy MIMO interfering broadcast channel with partial CSIT via the expected weighted MSE," in *Proc. ISWCS*, Aug. 2012, pp. 576–580.
- [21] R. Fritzsche and G. P. Fettweis, "Robust sum rate maximization in the multi-cell MU-MIMO downlink," in *Proc. IEEE WCNC*, Apr. 2013, pp. 3180–3184.
- [22] Y. Wu, S. Jin, X. Gao, M. R. McKay, and C. Xiao, "Transmit designs for the MIMO broadcast channel with statistical CSI," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4451–4466, Sep. 2014.
- [23] E. Björnson, N. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Process. Magazine*, vol. 31, no. 4, pp. 142–148, Jul. 2014.
- [24] M. Razaviyayn, M. Boroujeni, and Z.-Q. Luo, "A stochastic weighted MMSE approach to sum rate maximization for a MIMO interference channel," in *Proc. IEEE SPAWC*, Jun. 2013, pp. 325–329.
- [25] G. Caire and K. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, pp. 2274–2298, Dec. 2007.
- [26] A. Shapiro, D. Dentcheva *et al.*, *Lectures on stochastic programming: modeling and theory*. SIAM, 2009, vol. 9.
- [27] H. Joudeh and B. Clerckx, "Sum rate maximization for MU-MISO with partial CSIT using joint multicasting and broadcasting," in *Proc. IEEE ICC*, Jun. 2015, pp. 6349–6354.
- [28] —, "Achieving max-min fairness for MU-MISO with partial CSIT: A multicast assisted transmission," in *Proc. IEEE ICC*, Jun. 2015, pp. 6355–6360.
- [29] —, "Robust transmission in downlink multiuser MISO systems: A rate-splitting approach," *IEEE Trans. Signal Process.*, Accepted 2016.
- [30] 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, Jul. 2016.
- [31] C. Hao and B. Clerckx, "MISO networks with imperfect CSIT: A topological rate-splitting approach," *arXiv preprint arXiv:1602.03768*, 2016.
- [32] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

- [33] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.
- [34] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.
- [35] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.
- [36] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [37] 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*, Accepted 2016.
- [38] B. E. Fristedt and L. F. Gray, *A modern approach to probability theory*. Springer, 1997.
- [39] S. Kim, R. Pasupathy, and S. G. Henderson, "A guide to sample average approximation," in *Handbook of Simulation Optimization*. Springer, 2015, pp. 207–243.
- [40] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "Linear transceiver design for a MIMO interfering broadcast channel achieving max–min fairness," *Signal Processing*, vol. 93, no. 12, pp. 3327 – 3340, 2013.
- [41] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.
- [42] M. Razaviyayn, "Successive convex approximation: Analysis and applications," Ph.D. dissertation, UNIVERSITY OF MINNESOTA, 2014.
- [43] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.
- [44] H. Joudeh and B. Clerckx, "AMMSE optimization for multiuser MISO systems with imperfect CSIT and perfect CSIR," in *Proc. IEEE GLOBE-COM*, Dec. 2014, pp. 3308–3313.
- [45] A. Tajer, N. Prasad, and X. Wang, "Robust linear precoder design for multi-cell downlink transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 235–251, Jan. 2011.
- [46] M. Grant, S. Boyd, and Y. Ye, "CVX: MATLAB software for disciplined convex programming [Online]." Available: <http://www.stanford.edu/boyd/cvx>, 2008.



Hamdi Joudeh (S'13) received his PhD in Electrical Engineering and MSc in Communications and Signal Processing from Imperial College London, U.K., in 2016 and 2011 respectively, and his BEng in Electrical Engineering from the Islamic University of Gaza, Palestine, in 2010. In 2011–2012, he was an intern in the Mobile Communication Division at Samsung Electronics, Suwon, South Korea. He is currently a Research Associate in the Department of Electrical and Electronic Engineering at Imperial College London. His research interests include signal processing for wireless networks and communication theory.



Bruno Clerckx (M'08) is a Senior Lecturer (Associate Professor) in the Department of Electrical and Electronic Engineering at Imperial College London (London, United Kingdom). He received his M.S. and Ph.D. degrees in applied science from the Université catholique de Louvain (Louvain-la-Neuve, Belgium) in 2000 and 2005, respectively. From 2006 to 2011, he was with Samsung Electronics (Suwon, South Korea) where he actively contributed to 3GPP LTE/LTE-A and IEEE 802.16m and acted as the rapporteur for the 3GPP Coordinated Multi-Point (CoMP) Study Item. Since 2011, he has been with Imperial College London, first a Lecturer and now as a Senior Lecturer. Since March 2014, he also occupies an Associate Professor position at Korea University, Seoul, Korea. He also held visiting research positions at Stanford University, EURECOM, National University of Singapore and The University of Hong Kong. He is the author or co-author of 2 books, numerous peer-reviewed research papers, 150 standard contributions and the inventor of 75 issued or pending patents among which 15 have been adopted in the specifications of 4G (3GPP LTE/LTE-A and IEEE 802.16m) standards. Dr. Clerckx served as an editor for IEEE TRANSACTIONS ON COMMUNICATIONS from 2011–2015 and is currently an editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. His research area is communication theory and signal processing for wireless networks.