Asymptotic Analysis of MRT Over Double Scattering Channels With MMSE Estimation

Item Type	Article
Authors	Ye, Jia; Nadeem, Qurrat-Ul-Ain; Kammoun, Abla; Alouini, Mohamed-Slim
Citation	Ye, J., Nadeem, Q.-U.-A., Kammoun, A., & Alouini, M.-S. (2020). Asymptotic Analysis of MRT Over Double Scattering Channels With MMSE Estimation. IEEE Transactions on Wireless Communications, 1–1. doi:10.1109/twc.2020.3017053
Eprint version	Post-print
DOI	10.1109/TWC.2020.3017053
Publisher	Institute of Electrical and Electronics Engineers (IEEE)
Journal	IEEE Transactions on Wireless Communications
Rights	Archived with thanks to IEEE Transactions on Wireless Communications
Download date	27/08/2022 20:54:34
Link to Item	http://hdl.handle.net/10754/664599

Asymptotic Analysis of MRT Over Double Scattering Channels With MMSE Estimation

Jia Ye, *Student Member, IEEE*, Qurrat-Ul-Ain Nadeem, *Member, IEEE*,

Abla Kammoun, *Member, IEEE*, and Mohamed-Slim Alouini, *Fellow, IEEE*

Abstract

This paper studies the ergodic rate performance of maximum ratio transmission (MRT) precoding in the downlink of a multi-user multiple-input single-output (MISO) system, where the channel between the base station (BS) and each user is modeled using the double scattering model. We utilize the minimum-mean-square-error (MMSE) channel estimate for this model, which is used in the design of the MRT precoding. Within this setting, we are interested in deriving tight approximations of the ergodic rate under the assumption that the number of BS antennas (N), the number of users (K) and that of scatterers (S) grow large with the same pace. These approximations are expressed in simplified closed-form expressions for the special case of multi-keyhole channels. They reveal that unlike the standard Rayleigh channel in which the SINR grows as $\mathcal{O}\left(\frac{N}{K}\right)$, the SINR associated with a multi-keyhole channel scales as $\mathcal{O}\left(\frac{S}{K}\right)$. This particularly shows that the reaped gains of the large-scale MIMO over double scattering channels do not linearly increase with the number of antennas and are limited by the number of scatterers. We further provide simulation results that confirm the close match provided by the asymptotic analysis for moderate system dimensions and provide some useful insights into the interplay between N, K and S.

Index Terms

Double scattering channel, multiple-input single-output (MISO), multi-user systems, maximum ratio transmission (MRT) precoding, minimum-mean-square-error (MMSE) channel estimation.

I. INTRODUCTION

Large-scale multiple-input multiple-output (LS-MIMO) has emerged as a key technology to significantly increase the spectral efficiency of wireless communication systems and ensure massive connectivity [1], [2]. Specifically, LS-MIMO achieves large multiplexing gains by using hundreds of antennas to serve tens of users simultaneously [3]. The use of more antennas also helps focus energy into smaller regions, thereby mitigating the inter-user interference [4], [5]. Similar to any other technology, many researchers put their efforts to investigate the properties of LS-MIMO under different propagation environment [6], [7]. Several works have shown that under full rank Rayleigh channels, LS-MIMO systems provide large capacity gains that linearly scale with the number of antennas [8]. However, in practice, it has been observed that full-rank Rayleigh channels represent an optimistic model that ignores the presence of spatial correlation and poor scattering conditions. These two factors were shown in [9] to govern the performance of MIMO systems, which is at a much lower level than the performances predicted over full-rank Rayleigh channels [10], [11]. In light of this observation, Gesbert *et al.* propose a new channel model coined "double scattering channel model" which accounts for the spatial correlation, the channel rank deficiency and the limited scattering. Unlike the standard Rayleigh channel, the double scattering channel model is non-Gaussian, making its analysis rather challenging.

A. Prior Work

The analysis of the performance of MIMO systems over double scattering channels is relatively scarce and is mainly represented by the works in [12]–[24]. Specifically, the performance of point-to-point MIMO system under double scattering channels has been studied in [12]–[19], where the authors in [12] showed that the required number of scatterers in the channel to achieve full diversity is the product of the number of transmit (Tx) antennas and receive (Rx) antennas. In [13], the optimal capacity-achieving beamforming directions under double scattering channels were studied in [13] and were proven to correspond to the eigenvectors of the transmit spatial correlation matrix, while a tight upper bound on the ergodic capacity was provided [14]. Interestingly, the results reveal that the use of mutilple antennas in keyhole channels do not allow for spatial multiplexity gains, but only for diversity gains. In this line, closed-form expressions for

the diversity-multiplexing trade-off of double scattering MIMO channels were obtained in [15], while the diversity gain and array gain associated with the symbol error rate (SER) were studied in [16]. It was shown that the array gain of double scattering channel varies with the signal-to-noise ratio (SNR) rather than being a constant as in conventional Rayleigh and Rician channels. Closed-form upper bounds on the sum capacity of a MIMO multiple access system under double scattering channel were provided in [25]. The impact of the residual imperfections on the ergodic channel capacity for optimal receivers, and on the ergodic sum rates for linear minimum mean-squared-error (MMSE) receivers under product Rayleigh channel was investigated in [18]. The authors in [17] derived the asymptotic variance of mutual information under Rayleigh product channels and characterized the tradeoff between multiplexing gain and diversity gain under the uniform power transmission. The authors in [19] investigated the error performance of vehicle-to-vehicle communication systems undergo double Rayleigh fading considering physical-layer network coding and space-time trellis coding techniques.

Only few papers have investigated thus far the double scattering channel model in LS-MIMO settings [20]–[24]. The authors in [20] studied large antenna arrays without considering the correlation between Tx and Rx antennas. They derived implicit expressions for the channel capacity and signal-to-interference-plus-noise ratio (SINR) of the MMSE detector. The authors in [21] derived deterministic approximation of the mutual information in a MIMO multiple-access system with double scattering channels by leveraging tools from random matrix theory while the spectral efficiency of a multi-cell MIMO system under double scattering channels was assessed through Monte Carlo simulations in [22]. Recently, the performance of regularized zero-forcing (RZF) precoding in LS multi-user MISO systems under double scattering channels has been studied in [23] where tight deterministic approximations of SINR and ergodic rate have been derived based on results from random matrix theory.

The analysis in [23] has allowed to shed light on the limitations of double scattering channels. However, the results are extremely involved and could not be interpreted easily. On another note, the RZF precoder is recognized for its prohibitively high computational complexity since it requires the inversion of a large Gram matrix of the joint channels of all users. This motivates our work in which we aim to study the maximum-ratio transmission (MRT) precoding. In contrast to zero-forcing (ZF) precoding, which attempts to eliminate inter-user interference with a significant loss in the received energy of the desired signal [26], the MRT is devised to maximize the signal gain at the user of interest and is optimal in the noise-limited low SNR regime. It has also been

shown to not only present low computational complexity but also to be asymptotically optimal when the number of antennas is much larger than that of users [27], making it particularly suitable for LS-MIMO systems.

B. Main Contributions

Motivated by the above works, we study the downlink SINR and sum rate in a single-cell multi-user MISO system where the channel between the BS and the users is modeled by the double scattering model with S scatterers. For simplicity, we assume that the users are divided into G groups and share common correlation matrices in the same group. We consider a time-division duplex (TDD) protocol where the BS equipped with N Tx antennas estimates the downlink channels using the MMSE estimation technique based on uplink pilot signaling and employs MRT precoding in the downlink link to serve K users. Under this setting, we derive tight deterministic approximations of the SINR and ergodic rate under the assumption that the number of BS antennas N, scatterers S and users K grow large. Interestingly, we observe through numerical results that the obtained results are accurate for moderate system dimensions as well. Moreover, we simplify the obtained expressions in the case of a multi-keyhole channel, which provides interesting insights into the impact of the number of scatterers on the system performance. Particularly, we show that the SINR with MRT precoding over a double scattering channel does not grow unboundedly with the number of antennas and scales as $\frac{S}{K}$ in contrast to the scaling under Rayleigh fading channels for which the SINR grows as $\frac{N}{K}$. This scaling is also compared with the result in [23] showing that the SINR with RZF precoding grows unboundedly with the number of antennas if the number of scatterers is larger than that of users.

C. Outline and Notation

The remainder of the paper is organized as follows. Section II describes the system model. In Section III, the asymptotically tight deterministic equivalents of the SINR and user rates with MRT precoding are derived. Simulation results are presented in Section IV, and conclusions are drawn in Section V.

In this paper, x denotes a scalar; $\mathbf{x}$ represents a vector; $\mathbf{X}$ stands for a matrix. $(\cdot)^H$ is the conjugate transpose. $\mathbb{E}[\cdot]$ represents the expectation operation. $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix whose diagonal entries are from vector $\mathbf{x}$. $\mathrm{tr}(\mathbf{X})$ denotes the trace of $\mathbf{X}$. $\|\mathbf{X}\|$ is the spectral norm of $\mathbf{X}$. $\mathbf{I}_N$ denotes an $N \times N$ identity matrix. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the circularly symmetric

complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Big O notation represented as $a = \mathcal{O}(b)$ serves as a flexible abbreviation for $|a| \leq \beta b$, where β is a generic constant. The notation $\xrightarrow{\text{a.s.}} 0$ denotes almost sure convergence.

II. SYSTEM MODEL

We consider a single-cell multi-user MISO system where a BS equipped with N antennas serves K single-antenna users, divided into G groups of K_g, $g = 1, \cdots, G$, users, such that the users in the same group experience similar propagation conditions. The received complex baseband signal $y_{k,g}$ at user k in group g is given as,

$$y_{k,g} = \mathbf{h}_{k,g}^H \mathbf{x} + n_{k,g}, \ k = 1, \cdots, K_g, \ g = 1, \cdots, G, \tag{1}$$

where $\mathbf{h}_{k,g}^H \in \mathbb{C}^{1 \times N}$ is the double scattering channel vector [23] from the BS to user k in group g, $\mathbf{x} \in \mathbb{C}^{N \times 1}$ is the Tx signal vector and $n_{k,g} \sim \mathcal{CN}(0, \sigma^2)$ is the receiver noise. The Tx signal vector $\mathbf{x}$ is given as

$$\mathbf{x} = \sum_{g=1}^{G} \sum_{k=1}^{K_g} \sqrt{p_{k,g}} \mathbf{g}_{k,g} s_{k,g}, \tag{2}$$

where $\mathbf{g}_{k,g} \in \mathbb{C}^{N \times 1}$ is precoding vector, $p_{k,g} \geq 0$ is the signal power and $s_{k,g} \sim \mathcal{CN}(0, 1)$ is the data symbol for user k in group g respectively. The precoding vectors satisfy the following total power constraint:

$$\mathbb{E}\left[\|\mathbf{x}\|^2\right] = \mathbb{E}\left[\text{tr}\left(\mathbf{P}\mathbf{G}^H\mathbf{G}\right)\right] \leq \bar{P}, \tag{3}$$

where $\bar{P} > 0$ is the average total Tx power, $\mathbf{P} = \text{diag}\left(p_{1,1}, \cdots, p_{K_1,1}, p_{1,2}, \cdots, p_{K_G,G}\right) \in \mathbb{R}^{K \times K}$ and $\mathbf{G} = [\mathbf{G}_1, \cdots, \mathbf{G}_G] \in \mathbb{C}^{N \times K}$ is the precoding matrix, where $\mathbf{G}_g = \left[\mathbf{g}_{1,g}, \cdots, \mathbf{g}_{K_g,g}\right] \in \mathbb{C}^{N \times K_g}$.

We assume that the channel between the BS and user k in group g follows the double scattering channel model proposed in [9]. An illustration of the double scattering channel model is provided in Fig. 1 where $\sigma_{t,g}$ and $\sigma_{s,g}$ denote the angular spread of the radiated signal from the BS array and the Tx scatterers respectively, and $\mu_{t,g}$ and $\mu_{s,g}$ represent the mean angle of departure of the radiated signal from the BS array and Tx scatterers respectively, while d_t and d_s denote the distance between adjacent BS antennas and the distance between adjacent scatterers. The double scattering channel vector $\mathbf{h}_{k,g}$ is modeled as [9]:

$$\mathbf{h}_{k,g} = \sqrt{S_g} \left(\frac{1}{\sqrt{S_g}} \mathbf{R}_{BS_g}^{1/2} \mathbf{W}_g \bar{\mathbf{S}}_g^{1/2}\right) \tilde{\mathbf{w}}_{k,g}, \tag{4}$$
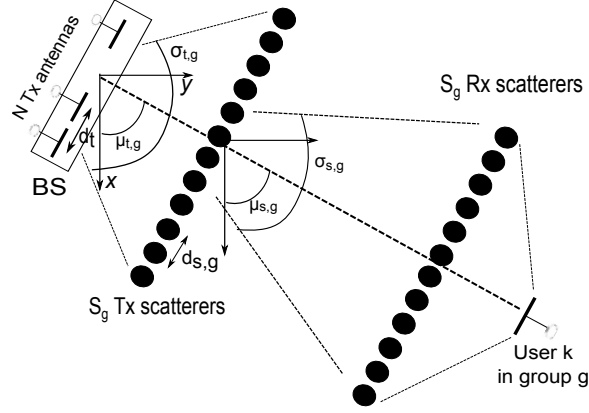
Fig. 1: Geometric model of the double scattering channel between the BS and the user k in group g.

where S_g is the number of scatterers at the Tx and the Rx sides in group g, $\mathbf{R}_{BS_g} \in \mathbb{C}^{N \times N}$ is the correlation matrix between the BS antennas and the S_g Tx scatterers, $\mathbf{W}_g \in \mathbb{C}^{N \times S_g}$ is a standard complex Gaussian matrix that describes the small-scale fading between the BS and the scattering cluster at the Tx side, $\bar{\mathbf{S}}_g \in \mathbb{C}^{S_g \times S_g}$ is the correlation matrix between the S_g Tx and Rx scatterers and $\tilde{\mathbf{w}}_{k,g} \sim \mathcal{CN}\left(\mathbf{0}, \frac{1}{S_g}\mathbf{I}_{S_g}\right) \in \mathbb{C}^{S_g \times 1}$ describes the small-scale fading between the user k in group g and the scattering cluster at the receiver side. In view of (4), the channel lies in a space with a rank that is determined by both the spatial correlation between the antennas at the BS and the structure of scattering in the propagation environment. Finally, it is worth mentioning that $\bar{\mathbf{S}}_g$ can be assumed to be diagonal without any loss of generality for the statistics of the received signal, a fact that can be easily proven by leveraging the bi-unitarily invariant property of standard Gaussian random matrices.

A. Channel Estimation

The acquisition of accurate CSI in a LS-MIMO system in a timely manner using downlink pilot signaling necessitates the number of downlink pilots to scale with the number of BS antennas, causing a large overhead. A common workaround to this issue is to operate in TDD mode in which downlink and uplink transmissions are sent over the same frequency, thus allowing the physical channel to be reciprocal between the downlink and the uplink links. Based on that, the uplink channel is estimated at the BS during a dedicated uplink phase prior to be used in a data transmission phase to implement precoding in the downlink.

More specifically, during the uplink training phase, the users transmit mutually orthogonal pilot sequences that allow the BS to compute the MMSE estimates $\hat{\mathbf{h}}_{k,g}$ of the channel vectors $\mathbf{h}_{k,g}$. After correlating the received training signal with the pilot sequence of user k in group g, the BS estimates the channel vector $\mathbf{h}_{k,g}$ based on the received observation, $\mathbf{y}_{k,g}^{tr} \in \mathbb{C}^{N \times 1}$, given as

$$\mathbf{y}_{k,g}^{tr} = \mathbf{h}_{k,g} + \frac{1}{\sqrt{\rho_{tr}}}\mathbf{n}_{k,g}^{tr}, \ k = 1, \cdots, K_g, \ g = 1, \cdots, G, \tag{5}$$

where $\mathbf{n}_{k,g}^{tr} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $\rho_{tr} > 0$ is the effective training SNR, assumed to be given here. Thus, the MMSE estimate $\hat{\mathbf{h}}_{k,g} \in \mathbb{C}^{N \times 1}$ of the non-Gaussian channel $\mathbf{h}_{k,g}$ is given by [23],

$$\hat{\mathbf{h}}_{k,g} = d_g \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{y}_{k,g}^{tr}, \tag{6}$$

where $d_g = \frac{1}{S_g}\mathrm{tr}\left(\bar{\mathbf{S}}_g\right)$ and $\mathbf{Q}_g = \left(d_g \mathbf{R}_{BS_g} + \frac{1}{\rho_{tr}}\mathbf{I}_N\right)^{-1}$.

Remark 1. *In this work, we do not investigate the effect of pilot contamination, which occurs when pilot signals are re-used due to the number of users being very large or when we have multiple cells (which is not the focus of this work). While we do not account for pilot contamination in the asymptotic analysis in this work, we do provide a simple extension of the MMSE estimate expressions for this case. When the pilot symbols are re-used by the users in different groups, the received observation is expressed as $\mathbf{y}_{k,g}^{tr,c} = \sum_{i=1}^{G} \mathbf{h}_{k,i} + \frac{1}{\sqrt{\rho_{tr}}}\mathbf{n}_{k,g}^{tr}$. Using the independence of channels, the correlation matrix between $\mathbf{h}_{k,g}$ and $\mathbf{y}_{k,g}^{tr,c}$ is found to be $\mathbf{C}_{\mathbf{h}_{k,g}\mathbf{y}_{k,g}^{tr,c}} = d_g \mathbf{R}_{BS_g}$, while the covariance matrix of $\mathbf{y}_{k,g}^{tr,c}$ is calculated as*

$$\mathbf{C}_{\mathbf{y}_{k,g}^{tr,c}\mathbf{y}_{k,g}^{tr,c}} = \mathbb{E}\left[\mathbf{y}_{k,g}^{tr,c}\mathbf{y}_{k,g}^{tr,c^H}\right] = \sum_{i=1}^{G} d_i \mathbf{R}_{BS_i} + \frac{1}{\rho_{tr}}\mathbf{I}_N. \tag{7}$$

Thus, the MMSE estimate $\hat{\mathbf{h}}_{k,g}$ can be expressed for this case as $\hat{\mathbf{h}}_{k,g} = d_g \mathbf{R}_{BS_g}\left(\sum_{i=1}^{G} d_i \mathbf{R}_{BS_i} + \frac{1}{\rho_{tr}}\mathbf{I}_N\right)^{-1} \mathbf{y}_{k,g}^{tr,c}$. In the following analysis, we only use the MMSE estimate shown in (6). The influence of pilot contamination is not the focus of this work, which will be studied in future work on multi-cell scenario.

B. Precoding Scheme and Achievable Rate

As explained earlier, the BS utilizes the MMSE estimate in (6) to implement the MRT precoding in which the symbol of each user is precoded by the Hermitian of its channel vector. Compared to RZF/ZF precoding which involves the inversion of the Gram matrix of joint

users' channel matrix, the MRT precoding presents a much lower computational complexity, thus making it attractive from a practical standpoint. More formally, the MRT precoding vector is given as $\mathbf{g}_{k,g} = \zeta \hat{\mathbf{h}}_{k,g}$ while the precoding matrix writes as ,

$$\mathbf{G}^H = \zeta \hat{\mathbf{H}}, \tag{8}$$

where $\hat{\mathbf{H}} = \left[\hat{\mathbf{H}}_1^H, \hat{\mathbf{H}}_2^H, \cdots, \hat{\mathbf{H}}_G^H \right]^H \in \mathbb{C}^{K \times N}$, with $\hat{\mathbf{H}}_g = \left[\hat{\mathbf{h}}_{1,g}, \hat{\mathbf{h}}_{2,g}, \cdots, \hat{\mathbf{h}}_{K_g,g} \right]^H \in \mathbb{C}^{K_g \times N}$, and ζ set in such a way that it meets the power constraint described in (3), which implies

$$\zeta^2 = \frac{\bar{P}}{\mathbb{E}\left[\mathrm{tr}\left(\mathbf{P}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}} \right) \right]} = \frac{\bar{P}}{\Theta}, \tag{9}$$

where $\Theta = \mathbb{E}\left[\mathrm{tr}\left(\mathbf{P}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}} \right) \right]$.

For the users to decode the transmitted signals in the downlink, the CSI which is estimated in uplink is needed. These can be exchanged by the BS but would cause a large overhead given the high number of transmitting antennas. To circumvent this issue, the channel hardening property of LS-MIMO systems according to which the effective useful channel $\mathbf{h}_{k,g}^H \mathbf{g}_{k,g}$ of a user converges to its average value when N, S grow large is invoked. Hence, it suffices for each user to acquire the knowledge of the statistical CSI for the computation of $\mathbb{E}\left[\mathbf{h}_{k,g}^H \mathbf{g}_{k,g} \right]$. For simplicity, we assume that $\mathbb{E}\left[\mathbf{h}_{k,g}^H \mathbf{g}_{k,g} \right]$ is perfectly known at the user, then $y_{k,g}$ can be decomposed as [1]

$$y_{k,g} = \underbrace{\sqrt{p_{k,g}}\mathbb{E}\left[\mathbf{h}_{k,g}^H \mathbf{g}_{k,g} \right] s_{k,g}}_{\text{effective signal}} + \underbrace{\sum_{(k',g') \neq (k,g)} \sqrt{p_{k',g'}}\mathbf{h}_{k,g}^H \mathbf{g}_{k',g'} s_{k',g'}}_{\text{interference}}$$

$$+ \underbrace{\sqrt{p_{k,g}}\left(\mathbf{h}_{k,g}^H \mathbf{g}_{k,g} - \mathbb{E}\left[\mathbf{h}_{k,g}^H \mathbf{g}_{k,g} \right] \right) s_{k,g} + n_{k,g}}_{\text{noise}}. \tag{10}$$

By treating the interference and channel uncertainty as a noise term, the ergodic achievable rate and the ergodic achievable sum rate with MRT precoding can be written respectively as

$$R_{k,g} = \log\left(1 + \gamma_{k,gMRT} \right) \tag{11}$$

and

$$R_{sum} = \sum_{g=1}^{G} \sum_{k=1}^{K_g} R_{k,g}, \tag{12}$$

---

[1]Superposition coding can help in alleviating the interference term by using self-interference cancellation decoding scheme at the users. However, this is not our focus in this work, we only give the simulation results and leave the exact performance analysis for future work.

where the effective average SINR of user k in group g defined as the ratio of the effective signal power to the sum of interference power and noise power is given by (13),

$$\gamma_{k,gMRT} = \frac{p_{k,g} \overbrace{\left| \mathbb{E} \left[\mathbf{h}_{k,g}^H \hat{\mathbf{h}}_{k,g} \right] \right|^2}^{\text{effective signal power}}}{\underbrace{\mathbb{E} \left[\mathbf{h}_{k,g}^H \hat{\mathbf{H}}_{[k,g]}^H \mathbf{P}_{[k,g]} \hat{\mathbf{H}}_{[k,g]} \mathbf{h}_{k,g} \right]}_{\text{interference power}} + \underbrace{p_{k,g} \text{var} \left[\mathbf{h}_{k,g}^H \hat{\mathbf{h}}_{k,g} \right] + \frac{\Theta}{\rho}}_{\text{noise power}}}, \tag{13}$$

where $\rho = \frac{\bar{P}}{\sigma^2}$, $\hat{\mathbf{H}}_{[k,g]} = \left[\hat{\mathbf{H}}_1^H, \cdots \hat{\mathbf{H}}_{g-1}^H, \hat{\mathbf{h}}_{1,g}, \cdots, \hat{\mathbf{h}}_{k-1,g}, \hat{\mathbf{h}}_{k+1,g}, \cdots, \hat{\mathbf{h}}_{K_g,g}, \cdots, \hat{\mathbf{H}}_G^H \right]^H \in \mathbb{C}^{K-1 \times N}$ and $\mathbf{P}_{[k,g]} = \text{diag} \left(p_{1,1}, \cdots, p_{K_{g-1},g-1}, p_{1,g}, \cdots, p_{k-1,g}, p_{k+1,g}, \cdots, p_{K_g,g}, \cdots, p_{K_G,G} \right) \in \mathbb{C}^{K-1 \times K-1}$.

Remark 2. *In this work, we consider user grouping, which is not necessary under MRT pre-coding. Our analysis applies for any G. However, user grouping with a finite G was needed for the asymptotic analysis of double scattering channel under RZF precoding in [23]. For the sake of comparison, we consider the same system setting as [23] under RZF precoding and extend some special cases to the scenario with no user grouping.*

III. MAIN RESULTS

In this section, we will leverage tools from random matrix theory to derive the deterministic approximations of the SINR and user rates under MRT precoding and assuming the channel to follow a double scattering fading model.

A. Assumptions

Analyzing the expression of the SINR in (13) for fixed dimensions over double scattering channels is not an easy problem. The main idea in this work is to assume that all system dimensions N, K_g, and S_g for each group g tend to infinity and to resort to asymptotic tools based on random matrix theory results to approximate the ergodic user rates. Such an approach has already been pursued in previous works [3], [28], but mainly for Rayleigh channels. The approximation is expected to be tight for large dimensions, an assumption that is not restrictive given that with the advent of 5G, BSs are equipped with a high number of antennas to serve a large number of users. Prior to stating the main results, we shall first formalize the growth rate regime in the following assumption:

Assumption 1. In large (S_g, K_g, N) regime, S_g, K_g and N tend to infinity such that

$$0 < \lim\inf \frac{S_g}{N} \leq \lim\sup \frac{S_g}{N} < \infty, \tag{14}$$

$$0 < \lim\inf \frac{K_g}{N} \leq \lim\sup \frac{K_g}{N} < \infty. \tag{15}$$

In the sequel, this assumption will be represented as $N \to \infty$. Moreover, the deterministic equivalent of a sequence of random variable X_N is represented by the deterministic sequence X_N^o, which approximates X_N such that $X_N - X_N^o \xrightarrow[N\to\infty]{\text{a.s.}} 0$. Accordingly, the deterministic equivalent $\gamma_{k,gMRT}^o$ of the SINR $\gamma_{k,gMRT}$ with MRT precoding satisfies,

$$\gamma_{k,gMRT} - \gamma_{k,gMRT}^o \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{16}$$

Moreover, the fixed diagonal matrix $\mathbf{P}$ whose diagonal elements are power allocation weights for each user should satisfy the following assumption.

Assumption 2. The diagonal values $p_{k,g}$ in $\mathbf{P} = \text{diag}\,(p_{1,1}, \cdots, p_{K_1,1}, p_{1,2}, \cdots, p_{K_G,G})$ are positive and of order $\mathcal{O}\left(\frac{1}{K}\right)$.

Additionally, we require the following two assumptions: **Assumption 3.** For all groups, $\lim\sup_N \left\| \mathbf{R}_{BS_g} \right\| < \infty$ and $\lim\sup_{S_g} \left\| \overline{\mathbf{S}}_g \right\| < \infty$.

Assumption 4. For all groups, $\lim\inf_N \frac{1}{N} \text{tr}\, \mathbf{R}_{BS_g} > 0$ and $\lim\inf_{S_g} \frac{1}{N} \text{tr}\, \overline{\mathbf{S}}_g > 0$.

The analysis starts by expressing the double-scattering channel model in (4) as,

$$\mathbf{h}_{k,g} = \sqrt{S_g}\mathbf{Z}_g\tilde{\mathbf{w}}_{k,g}, \tag{17}$$

where $\mathbf{Z}_g = \frac{1}{\sqrt{S_g}}\mathbf{R}_{BS_g}^{1/2}\mathbf{W}_g\overline{\mathbf{S}}_g^{1/2}$. We will make use of the Fubini theorem [23] to first assume $\mathbf{Z}_g$ to be deterministic. Under this setting, the estimate of the double scattering model in (6) can be interpreted as

$$\hat{\mathbf{h}}_{k,g} = \mathbf{\Phi}_g^{1/2}\bar{\mathbf{q}}_{k,g}, \tag{18}$$

where $\bar{\mathbf{q}}_{k,g} \sim \mathcal{CN}\,(\mathbf{0}, \mathbf{I}_N)$ and $\mathbf{\Phi}_g$ is the covariance matrix of the channel estimate given as,

$$\mathbf{\Phi}_g = d_g^2 \mathbf{R}_{BS_g}\mathbf{Q}_g \left(\mathbf{Z}_g\mathbf{Z}_g^H + \frac{1}{\rho_{tr}}\mathbf{I}_N \right) \mathbf{Q}_g^H\mathbf{R}_{BS_g}^H. \tag{19}$$

Once the deterministic equivalent of the SINR is derived under this setting, we will allow $\mathbf{Z}_g$ to be random and compute the final deterministic equivalent.

B. Useful Results

From (13), we can see that the deterministic approximation of the SINR explicitly depends on the effective signal term, interference term and noise term. Therefore, we first derive the deterministic equivalents of these quantities in the following lemmas.

Lemma 1. Under the the setting of **Assumption 1-4**, the deterministic equivalent of the effective signal power is

$$\frac{p_{k,g}}{K}\left|\mathbb{E}\left[\mathbf{h}_{k,g}^H\hat{\mathbf{h}}_{k,g}\right]\right|^2 - \frac{p_{k,g}}{K}\left[d_g^2\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{R}_{BS_g}\mathbf{Q}_g\right)\right]^2 \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{20}$$

Proof. The proof of **Lemma 1** is provided in Appendix A. □

Lemma 2. The deterministic equivalent of the interference term under **Assumption 1-4** is calculated as

$$\frac{1}{K}\mathbb{E}\left[\mathbf{h}_{k,g}^H\hat{\mathbf{H}}_{[k,g]}^H\mathbf{P}_{[k,g]}\hat{\mathbf{H}}_{[k,g]}\mathbf{h}_{k,g}\right] - \sum_{i=1}^{G}\sum_{l=1}^{K_i}p_{l,i}\left(\Upsilon_{i,g}^o + \Psi_{i,g}^o\right) \xrightarrow[N\to\infty]{\text{a.s.}} 0, \tag{21}$$

where $\Upsilon_{i,g}^o = \frac{d_i^2 d_g}{K\rho_{tr}}\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{R}_{BS_i}\mathbf{Q}_i\mathbf{Q}_i^H\mathbf{R}_{BS_i}^H\right)$ and $\Psi_{i,g}^o$ is shown as

$$\Psi_{i,g}^o = \begin{cases} \frac{d_i^3 d_g}{K}\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{R}_{BS_i}\mathbf{Q}_i\mathbf{R}_{BS_i}\mathbf{Q}_i^H\mathbf{R}_{BS_i}^H\right), & \text{if } i\neq g \\ \frac{d_g^2}{K}\sum\limits_{j=1}^{S_g}\frac{\bar{s}_{g,j}^2}{S_g^2}\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\right)\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{R}_{BS_g}\mathbf{Q}_g\right) \\ +\frac{d_g^2}{K}\sum\limits_{j=1}^{S_g}\frac{\bar{s}_{g,j}}{S_g}\sum\limits_{\substack{n=1\\n\neq j}}^{S_g}\frac{\bar{s}_{g,n}}{S_g}\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{R}_{BS_g}\mathbf{Q}_g\mathbf{R}_{BS_g}\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\right), & \text{if } i=g \end{cases} \tag{22}$$

Since the interference term converges to $\sum_{i=1}^{G}\sum_{l=1}^{K_i}p_{l,i}\left(\Upsilon_{i,g}^o + \Psi_{i,g}^o\right)$, we can conclude that the intra-group interference is $\sum_{l=1}^{K_g}p_{l,g}\left(\Upsilon_{g,g}^o + \Psi_{g,g}^o\right)$, while the inter-group interference should be $\sum_{i=1,i\neq g}^{G}\sum_{l=1}^{K_i}p_{l,i}\left(\Upsilon_{i,g}^o + \Psi_{i,g}^o\right)$.

Proof. The proof of **Lemma 2** is provided in Appendix B. □

Lemma 3. Under the the setting of **Assumption 1-4**, the deterministic equivalent of the noise power term is shown as

$$
\frac{1}{K}\left[p_{k,g}\mathrm{var}\left[\mathbf{h}_{k,g}^{\mathrm{H}}\hat{\mathbf{h}}_{k,g}\right] + \frac{\Theta}{\rho}\right]
$$

$$
- \frac{1}{K\rho}\sum_{g=1}^{G}\sum_{k=1}^{K_g}p_{k,g}\left[\frac{d_g^2}{\rho_{tr}}\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\right) + d_g^3\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\mathbf{R}_{BS_g}\mathbf{Q}_g\right)\right] \xrightarrow[N\to\infty]{\text{a.s.}} 0.
$$

$$(23)$$

Proof. The proof of **Lemma 3** is provided in Appendix C. □

C. Deterministic Approximation of the SINR

Based on the results in **Lemma 1**, **Lemma 2** and **Lemma 3**, the deterministic equivalent of the SINR in (13) can be derived and is stated in the next theorem.

Theorem 1. Under the the setting of **Assumption 1-4**, the downlink SINR of user k in group g defined in (13) converges almost surely as,

$$
\gamma_{k,gMRT} - \gamma_{k,gMRT}^{o} \xrightarrow[N\to\infty]{\text{a.s.}} 0, \tag{24}
$$

where $\gamma_{k,gMRT}^{o}$ is shown as

$$
\gamma_{k,gMRT}^{o} = \frac{\frac{p_{k,g}}{K}\left[d_g^2\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{R}_{BS_g}\mathbf{Q}_g\right)\right]^2}{\sum\limits_{i=1}^{G}\sum\limits_{l=1}^{K_i}p_{l,i}\left(\Upsilon_{i,g}^{o} + \Psi_{i,g}^{o}\right) + \sum\limits_{g=1}^{G}\sum\limits_{k=1}^{K_g}\frac{p_{k,g}}{K\rho}\left[\frac{d_g^2}{\rho_{tr}}\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\right) + d_g^3\mathrm{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\mathbf{R}_{BS_g}\mathbf{Q}_g\right)\right]}.
$$

$$(25)$$

Proof. Substituting the results of **Lemma 1**, **Lemma 2** and **Lemma 3** into (13) completes the proof. □

Corollary 1. The individual downlink rates $R_{k,g}$ of users converge as,

$$
R_{k,g} - R_{k,g}^{o} \xrightarrow[N\to\infty]{\text{a.s.}} 0, \tag{26}
$$

where

$$
R_{k,g}^{o} = \log\left(1 + \gamma_{k,gMRT}^{o}\right), \tag{27}
$$

where $\gamma_{k,gMRT}^{o}$ is given by (25).

Proof. The proof follows from the application of the continuous mapping theorem [29] to the logarithm function along with the almost sure convergence of $\gamma_{k,gMRT}$ in (24). □

An approximation of R_{sum} can be obtained by replacing $R_{k,g}$ in (12) with its asymptotic approximation from (27) as follows

$$R_{sum}^o = \sum_{g=1}^{G} \sum_{k=1}^{K_g} \log\left(1 + \gamma_{k,gMRT}^o\right),$$

(28)

such that $\frac{1}{K}\left(R_{sum} - R_{sum}^o\right) \to 0$ almost surely. The asymptotic approximation of R_{sum} will be shown to be tight using simulations.

We can see that the deterministic equivalents depend only on the 'slowly varying' covariance matrices $\mathbf{R}_{BS_g}$ and $\bar{\mathbf{S}}_g$ instead of the instantaneous channels that vary very fast. These correlation matrices, although huge in size in the large N, S_g regime, can be computed at the BS using knowledge of only the large-scale channel statistics.

Since the expressions derived above do not offer direct insights, we simplify them for the Rayleigh product channel and obtain scaling laws for different operating conditions. We also compare them with the results in [23] for the RZF precoder .

D. Rayleigh Product Channel

The Rayleigh product channel does not exhibit any form of correlation [30], and is popularly known as the multi-keyhole channel. Under this model, **Theorem 1** can be simplified as in the following corollaries.

Corollary 2. Consider a single group, i.e. $G = 1$ in which all channels are assumed to share the same covariance matrix. Let S and K denote respectively the number of scatterers and that of users. Denote by $\mathbf{R}_{BS}$ the transmit correlation matrix and by $\mathbf{Q}$ the inverse of the covariance matrix of the received observation vector during channel estimation in (6). We also assume that the correlation matrix of the channel between the Tx and Rx scatterers $\bar{\mathbf{S}}$ is equal to identity, that is $\bar{\mathbf{S}} = \mathbf{I}_S$. Then $\gamma_{k,gMRT}^o$ in **Theorem 1** can be simplified as

$$\gamma_{kMRT}^o = \frac{\frac{p_k}{K}\left[\text{tr}\left(\mathbf{R}_{BS}\mathbf{R}_{BS}\mathbf{Q}\right)\right]^2}{P\left(\Upsilon^o + \Psi^o\right) + \frac{P}{K\rho}\left[\frac{1}{\rho_{tr}}\text{tr}\left(\mathbf{R}_{BS}\mathbf{Q}\mathbf{Q}^H\mathbf{R}_{BS}^H\right) + \text{tr}\left(\mathbf{R}_{BS}\mathbf{Q}^H\mathbf{R}_{BS}^H\mathbf{R}_{BS}\mathbf{Q}\right)\right]},$$

(29)

where $P = \sum_{k=1}^{K} p_k$ denotes the total Tx power at BS, $\Upsilon^o = \frac{1}{K\rho_{tr}}\text{tr}\left(\mathbf{R}_{BS}\mathbf{R}_{BS}\mathbf{Q}\mathbf{Q}^H\mathbf{R}_{BS}^H\right)$ and $\Psi^o = \frac{S-1}{SK}\text{tr}\left(\mathbf{R}_{BS}\mathbf{R}_{BS}\mathbf{Q}\mathbf{R}_{BS}\mathbf{Q}^H\mathbf{R}_{BS}^H\right) + \frac{1}{SK}\text{tr}\left(\mathbf{R}_{BS}\mathbf{Q}^H\mathbf{R}_{BS}^H\right)\text{tr}\left(\mathbf{R}_{BS}\mathbf{R}_{BS}\mathbf{Q}\right)$ represent the interference term.

Proof. For a single group and $\bar{\mathbf{S}} = \mathbf{I}_S$, $d_1 = \frac{1}{S}\text{tr}\left(\mathbf{S}\right) = 1$. Substituting the value of d into (25) and reducing (25) to a single group completes the proof. $\qquad\square$

Corollary 3. Under the setting of **Corollary 1**, and further assume that the channel is uncorrelated, i.e. $\mathbf{R}_{BS} = \mathbf{I}_N$. Such an assumption generally holds when the angular spread or the inter-antenna separation at the BS is large and describes the multi-keyhole channel [14], [30]. Then $\gamma^o_{k,gMRT}$ in **Theorem 1** can be given in a closed-form as,

$$\gamma^o_{kMRT} = \frac{p_k N}{P\left(\frac{1}{\rho_{tr}} + 1 + \frac{N-1}{S}\right) + \frac{P}{\rho}\left(1 + \frac{1}{\rho_{tr}}\right)}. \tag{30}$$

Proof. The proof of **Corollary 3** is postponed to Appendix D. $\qquad\square$

It should be noted that an upper-bound of the asymptotic received SINR performance for MRT can be derived by tending $\frac{P}{\rho}$ to zero in (30), which leads to $\gamma^o_{kMRT,upper} = \frac{p_k N}{P\left(\frac{1}{\rho_{tr}} + 1 + \frac{N-1}{S}\right)}$.

Remark 3. *In the case without user grouping, that is, $G = K$, the influence of the number of scatterers disappears under Rayleigh Product channel with $\bar{\mathbf{S}}_g = \mathbf{I}_{S_g}$ and $\mathbf{R}_{BS_g} = \mathbf{I}_N$ because the channels between users are orthogonal. Through a similar derivation, the SINR of each user in **Corollary 3** is simplified as $\gamma^o_{k,gMRT} = \frac{p_k N}{P\left(\frac{1}{\rho_{tr}} + 1\right) + \frac{P}{\rho}\left(1 + \frac{1}{\rho_{tr}}\right)}$.*

1) Standard Rayleigh channel: If the number of scatterers S is very large as compared to N and K, the Rayleigh product channel approaches the Rayleigh channel, the behavior of which can be predicted by assuming that $\frac{S}{N} \to \infty$ and $\frac{S}{K} \to \infty$,

Corollary 4. Under the setting of **Corollary 3**, let $\frac{S}{N}, \frac{S}{K} \to \infty$. Then γ^o_{kMRT} defined in **Theorem 1** approaches the limit $\gamma_{kMRT, \frac{S}{N}, \frac{S}{K} \to \infty}$ which is given as

$$\gamma^o_{kMRT, \frac{S}{N}, \frac{S}{K} \to \infty} = \frac{p_k N}{P\left(\frac{1}{\rho_{tr}} + 1\right) + \frac{P}{\rho}\left(1 + \frac{1}{\rho_{tr}}\right)}. \tag{31}$$

As expected, we retrieve the asymptotic SINR expression of an uncorrelated Rayleigh fading channel obtained in [23].

Remark 4. *Based on **Assumption 2**, p_k is of order $\mathcal{O}\left(\frac{1}{K}\right)$, we can conclude that the SINR under a Rayleigh fading channel scales as $\frac{N}{K}$. This proves that the SINR increases linearly with the number of transmit antennas. As the Rayleigh channel model does not account for the number of scatterers, their effect is naturally not reflected in the asymptotic SINR expression.*

In the denominator of (31), the term $\frac{P}{\rho_{tr}} + \frac{P}{\rho\rho_{tr}}$ represents the loss incurred due to an imperfect CSI. The perfect CSI case can be studied by taking the limit of the asymptotic SINR expression as $\rho_{tr} \to \infty$. In doing so, we obtain the following result:

Corollary 5. Under the setting of **Corollary 4**, let $\rho_{tr} \to \infty$. Then the SINR for the perfect CSI case is given by

$$\gamma^o_{kMRT,\frac{S}{N},\frac{S}{K}\to\infty} = \frac{p_k N}{P + \frac{P}{\rho}}. \tag{32}$$

Let us now compare the performance of the MRT precoding with RZF precoding in the perfect CSI case in the case $\frac{S}{N} \to \infty$ and $\frac{S}{K} \to \infty$. In [23], the $\gamma^o_{kRZF,\frac{S}{N},\frac{S}{K}\to\infty}$ under RZF precoding and perfect CSI for Rayleigh product channel is given as,

$$\gamma^o_{kRZF,\frac{S}{N},\frac{S}{K}\to\infty} = \frac{\frac{p_k}{P/K}\left(\frac{1}{\overline{m}} - 1\right)\left(1 + \frac{\alpha}{\overline{m}^2}\right)}{1 + \frac{1}{\rho\overline{m}^2}}, \tag{33}$$

where α is the regularization parameter and $\overline{m} = \frac{1 - \frac{N}{K} - \alpha + \sqrt{\left(\alpha + \frac{N}{K} - 1\right)^2 + 4\alpha}}{2}$. Since it is difficult to compare these two term, we re-express $\overline{m}$ through binomial series $(1 + x)^{1/2} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \mathcal{O}(x^3)$. By only taking the term with first order $\mathcal{O}\left(\frac{K}{N}\right)$, we can approximate $\overline{m}$ as $\overline{m} \approx \alpha\frac{K}{N}$. Substituting the approximated $\overline{m}$ into (33) and only keeping the terms of order $\frac{N}{K}$ in the numerator and the denominator, we can approximate the SINR with RZF precoding under Rayleigh fading channel as,

$$\gamma^o_{kRZF,\frac{S}{N},\frac{S}{K}\to\infty} = \frac{p_k N}{\frac{P}{\rho}}. \tag{34}$$

Remark 5. *Comparing $\gamma^o_{kRZF,\frac{S}{N},\frac{S}{K}\to\infty}$ with $\gamma^o_{kMRT,\frac{S}{N},\frac{S}{K}\to\infty}$, we can clearly see that the performance with MRT precoding is close to the performance with RZF precoding under Rayleigh fading channel in the low SNR regime, i.e. small ρ. For high SNR regime, the SINR under RZF precoding is much better than SINR under MRT precoding because MRT introduces more interference power P than RZF precoding. In the particular case when $p_k = \frac{1}{K}$, $\gamma^o_{kRZF,\frac{S}{N},\frac{S}{K}\to\infty} = C_{RZF}\frac{N}{K}$, where $C_{RZF} = \frac{\rho}{P}$ and $\gamma^o_{kMRT,\frac{S}{N},\frac{S}{K}\to\infty} = C_{MRT}\frac{N}{K}$, where $C_{MRT} = \frac{\rho}{P\rho+P}$. It can be clearly seen that the multiplicative constant of RZF is greater than that of MRT but become very close to it in the low SNR regime.*

2) Multi-keyhole Channel with Limited Scatterers: The number of scatterers is a parameter that depends on the propagation environment and over which the network operator cannot have

any control, while the number of antennas can be increased by equipping the BS with many of them. This motivates us to study whether an increase in the number of antennas can always result in an enhancement of the performance when a multi-keyhole channel with a limited number of scatterers is considered. Such a situation is modeled by the assumption $\frac{N}{S} \to \infty$ and $\frac{N}{K} \to \infty$.

Corollary 6. Under the setting of **Corollary 3**, let $\frac{N}{S}, \frac{N}{K} \to \infty$ and $p_k = \frac{P}{K}$. Then $\gamma^o_{k,gMRT}$ defined in **Theorem 1** approaches the limit $\gamma^o_{kMRT, \frac{N}{S}, \frac{N}{K} \to \infty}$ which is given by,

$$\gamma^o_{kMRT, \frac{N}{S}, \frac{N}{K} \to \infty} = \frac{S}{K}. \tag{35}$$

Remark 6. *Corollary 6 implies that for $\frac{N}{S}, \frac{N}{K} \to \infty$, $\gamma^o_{kMRT, \frac{N}{S}, \frac{N}{K} \to \infty}$ saturates at a fixed value S/K. Different from the result of **Corollary 4** under Rayleigh fading, the multi-keyhole channel provides the SINR scaling of the order $\mathcal{O}\left(\frac{S}{K}\right)$, which is independent of the number of transmit antennas at the BS. This not only proves that the achievable SINR is determined by the number of scatterers in realistic environments but also that improving the number of transmit antennas at the BS is not a feasible scheme to improve the SINR in LS-MIMO system under limited scattering.*

The performance of a LS-MIMO system with MRT precoding is limited by the number of scatterers in the propagation environment. This is also the case of RZF precoding studied in [23]. As a matter of fact, according to the corollary 5 of [23], $\gamma^o_{kRZF, \frac{N}{S}, \frac{N}{K} \to \infty}$ with $K > S$ is given by

$$\gamma^o_{kRZF, \frac{N}{S}, \frac{N}{K} \to \infty} = \frac{S}{K - S}, \tag{36}$$

while it grows unboundedly with N when $S < K$. Comparing the performance of RZF precoding with MRT precoding, we can see that MRT and RZF have similar performances when $K \gg S$, proving that in a very limited rank channel, there is not much gain to reap in going for the RZF precoding. However, if the number of users K is less than that of scatterers ($K < S$), RZF precoding benefit from an increase of the number of antennas, similarly to what happens if a standard Rayleigh channel is considered. On the other hand, MRT precoding could not gain in this situation from an increase in the number of transmit antennas.

Remark 7. *Based on the above derived results, we can conclude that under independent Rayleigh fading channel (which describes a scenario with a large number of scatterers), the sum-rate can be greatly improved by increasing the number of transmit antennas at the BS due to the massive*

MIMO effect [3]. However, when the environment has poor scattering conditions and suffers from rank loss, the BS can not simultaneously schedule more users than the number of degrees of freedom the MIMO channel can offer (i.e. the number of scatterers). In fact, scaling the number of scheduled users with the number of BS antennas as done in a lot of massive MIMO literature will significantly deteriorate the sum-rate in such propagation environments. In practice, the number of users to be scheduled simultaneously will depend on both the number of BS antennas and the scattering conditions in the propagation environment and therefore the latter should not be ignored in the system design.

IV. SIMULATION RESULTS

We consider a double scattering channel model in which the correlation matrices $\mathbf{R}_{BS_g}$ and $\bar{\mathbf{S}}_g$ are given as $\mathbf{R}_{BS_g} = \mathbf{F}\left(\mu_{t,g}, \sigma_{t,g}, d_t, S_g\right)$ and $\bar{\mathbf{S}}_g = \mathbf{F}\left(\mu_{s,g}, \sigma_{s,g}, d_{s,g}, S_g\right)$, where $\mathbf{F}(\mu, \sigma, d, n)$ is given as [21],

$$[\mathbf{F}(\mu, \sigma, d, n)]_{k,l} = \frac{1}{n} \sum_{j=\frac{1-n}{2}}^{\frac{n-1}{2}} \exp\left[-i2\pi d(k-l)\cos\left(\frac{\pi}{2} + \frac{j\sigma}{n-1} + \mu\right)\right]. \tag{37}$$

The main parameters are set as $G = 4$, $K = 128$, $S_g = \{130, 140, 160, 150\}$, $d_t = 0.5$, $d_{s,g} = 2$, $\mu_{t,g} = \mu_{s,g} = \{-\pi/3, -\pi/9, \pi/9, \pi/3\}$, $\sigma_{t,g} = \{\pi/5, \pi/6, \pi/5, \pi/7\}$ and $\sigma_{s,g} = \{\pi/6, \pi/6, \pi/6, \pi/6\}$. We assume an equal number of users in each group, i.e. $K_g = K/G$ with power allocation $\mathbf{P} = \frac{1}{K}\mathbf{I}_K$. Fig. 2 compares the downlink system sum rate using 2000 Monto-Carlo realizations of the SINR in (13) to the asymptotic approximation provided in (28) using (25). As can be seen, the asymptotic result yields a very good approximation for moderate system dimensions. The mismatch between the theoretical and the Monte-Carlo results starts to increase for high SNR values due to the slower convergence of $\gamma_{k,gMRT}$ to its deterministic approximation.

The comparison between RZF precoding, MRT precoding and MRT with superposition coding (MRT-SC) is also shown in Fig. 2. Even though MRT experiences performance loss as compared to RZF since it does not minimize the interference directly, the performance of MRT precoding can be improved by using a higher number of Tx antennas as done in the plotted figure while saving on the computational complexity. MRT precoding with $N = 128$ antennas outperforms RZF precoding with $N = 64$ antennas when pilot power is 10dB for almost all of the SNR range. In fact, in the low SNR regime MRT precoder with only 1 dB pilot training power and
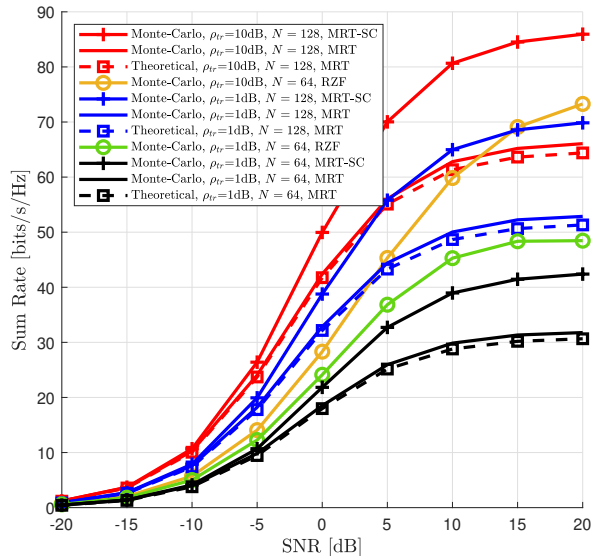
Fig. 2: Sum rate versus SNR.

128 antennas can achieve higher sum rate than RZF precoding with 10 dB pilot training power and 64 antennas. Moreover, MRT precoding is less affected from the decrease in training SNR in high SNR regime as compared to RZF. By adopting the superposition coding which results in much lower interference among the users, the system performance could be increased greatly compared to MRT precoding alone. Moreover, the performance gap between MRT-SC and RZF for the same number of Tx antennas is small as well.

By re-setting $K = S_g = 100$ and the regularization parameter α in RZF precoding, we re-compare the theoretical results of MRT precoding with the one of RZF precoding. In this experiment, we try to find when the MRT precoding outperforms RZF precoding under a fixed regularization parameter. As shown in Fig 3, when $N/S = 1$ and $\alpha = 1$, MRT precoding is worse than RZF precoding when SNR is bigger than -5 dB. As α increases, the gap between the RZF precoding and MRT precoding becomes smaller. Especially when $\alpha = 100$, the sum rate under RZF precoding is almost the same as the one under MRT precoding even in the high SNR regime.

Fig. 4 studies the effect of the number of scatterers on the system sum rate for a single group with multi-keyhole channel, i.e. $G = 1$, $\mathbf{R}_{BS} = \mathbf{I}_N$ and $\bar{\mathbf{S}} = \mathbf{I}_S$ with $N = 32$, $K = 200$ and $\rho_{tr} = 10$dB. We can see that the plotted approximation of the sum rate using (30) is close to the
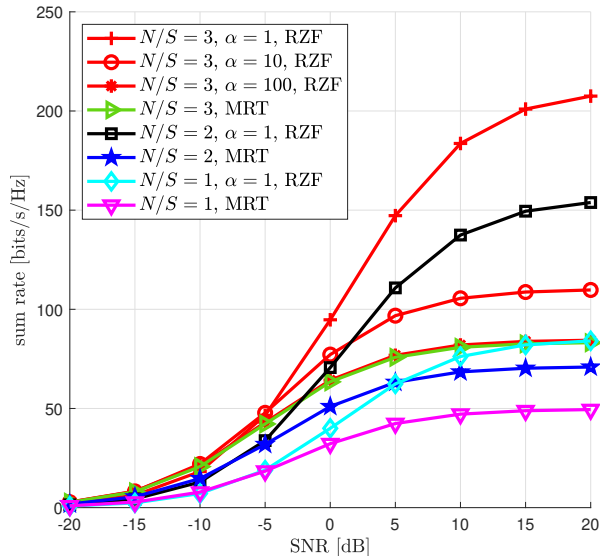
Fig. 3: Sum rate versus SNR.

Monte-Carlo result even for a very low number of scatterers. The spatial multiplexing gains are seen to increase linearly with S when $S < N$ while the gains start to decrease when $S > N$. The reason is that the degrees of freedom are limited by the number of antennas at the BS. The limiting sum rate as $\frac{S}{N}, \frac{S}{K} \to \infty$ is also plotted using the SINR in (31). As the number of scatterers increases, the performance approaches to that of a Rayleigh fading channel. Moreover, the sum-rate with RZF precoding when $\frac{S}{N}, \frac{S}{K} \to \infty$ is shown to be close to the sum-rate with MRT precoding in the low SNR regime, which verifies the correctness of Remark 5. As the SNR increases, the gap between the sum-rate with RZF precoder and that with MRT precoder grows since MRT brings in more interference power as shown in (32) and (34). In conclusion, MRT precoding approaches the performance of RZF in the low SNR regime with much lower computational complexity.

We study the performance of the Rayleigh product channel in Fig. 5 and Fig. 6 as the number of BS antennas increases for a single group. It can be seen that the performance saturates at a limiting sum rate given by the result of **Corollary 6**. This result confirms that it is useless to deploy more antennas when the number of scatterers in the environment is limited. As shown in Fig. 5, the sum-rate with RZF precoder is also bounded by a fixed value given by substituting (36) in the sum-rate, when $K > S$. The sum-rate with MRT precoding is close to the one with
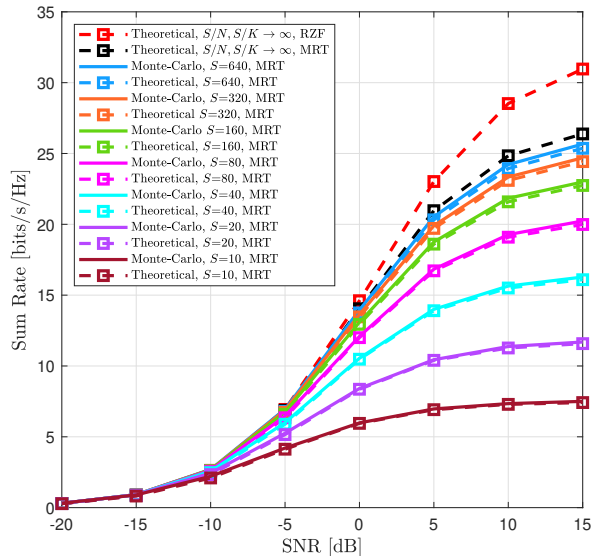
Fig. 4: Sum rate versus SNR for multi-keyhole channel.

RZF precoding when $K \gg S$ while the gap between the sum-rate under RZF and MRT becomes larger and larger as S grows. Such an observation is an agreement with what we deduced earlier from the comparison between (36) and (35).

Now coming to the performance when $S > K$ in Fig. 6, we observe that the performance of MRT precoder is bounded when $S > K$ to the limit S/K, while the performance of RZF precoder grows unboundedly with N. Moreover, a higher number of N is needed to reach the bound under MRT as the value of S increases. In our case, the performance of the Rayleigh product channel grows unboundedly only when $\frac{S}{K} \to \infty$, which is in accordance with the results of **Corollary 6**. In conclusion, the spatial multiplexing gains in realistic implementations will always be limited by the number of scatterers when using MRT precoder. Moreover, MRT precoding could provide near-optimal performance with much lower computational complexity when there is a limited number of scatterers and many active users.

V. CONCLUSION

In this paper, we derived the deterministic approximations of the SINR and ergodic rates with MRT precoding for a LS multi-user MISO system over the double scattering channel model. Unlike the standard Rayleigh fading channel, this model accounts for keyhole channel effect, a practical situation encountered in case of limited number of scatterers. Although being more
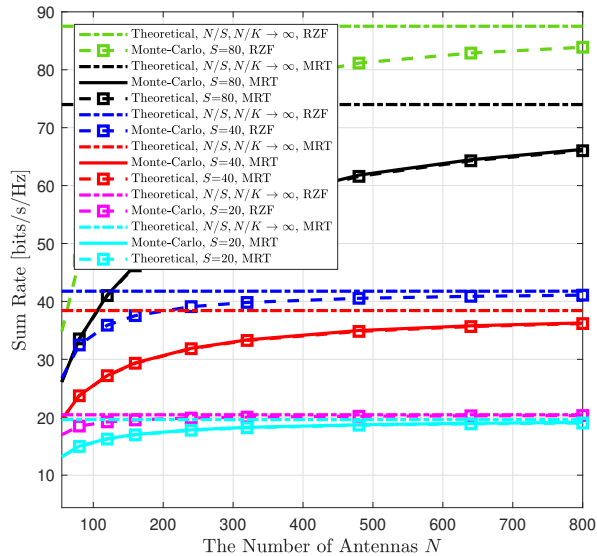
Fig. 5: Sum rate versus the number of antennas N with the number of users $K = 480$.
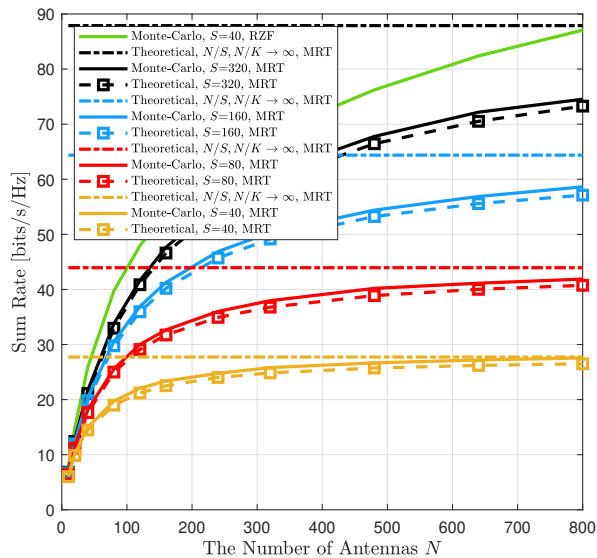
Fig. 6: Sum rate versus the number of antennas N with the number of users $K = 40$.

realistic, the study of this model has remained scarce, thus motivating the present work. Even though derived under the assumption of large system dimensions, these deterministic equivalents where shown to be close to Monte-Carlo results for moderate system dimensions as well. In an effort to draw meaningful insights into the performance of MRT precoding over the double

scattering channels, a simplified multi-keyhole channel is considered. Particularly, it has been shown that contrary to a Rayleigh fading channel whose performance improves with the number of Tx antennas, the performance of MRT precoding over multi-keyhole channels saturates at a limit that depends on the number of scatterers. A comparison with the RZF precoding reveals that MRT results in a comparable performance with RZF at low SNR or when the number of scatterers is limited and a large number of users is served. In such a situation, it becomes an interesting alternative since it allows to avoid the prohibitively high computational complexity of RZF.

APPENDIX A

PROOF OF LEMMA 1

It has been shown in [31] (Appendix E) that $\mathbf{Z}_g$ in (17) satisfies $\limsup_N \left\| \mathbf{Z}_g \mathbf{Z}_g^H \right\| < \infty$ almost surely. As $\mathbf{Z}_g$ is independent of $\tilde{\mathbf{w}}_{k,g}$ for $k = 1, \cdots, K_g$, we proceed in two steps. First, we assume $\mathbf{Z}_g$ to be deterministic and use trace lemma (Lemma 3 from [32]). Then, we leverage the random model of $\mathbf{Z}_g$ to find a deterministic equivalent of the SINR.

Substituting (17) into the MMSE estimate given in (6), we obtain:

$$\frac{1}{K}\mathbf{h}_{k,g}^H \hat{\mathbf{h}}_{k,g} = \frac{d_g}{K}\bar{\mathbf{w}}_{k,g}^H \mathbf{Z}_g^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{Z}_g \bar{\mathbf{w}}_{k,g} + \frac{d_g}{K\sqrt{\rho_{tr}}}\bar{\mathbf{w}}_{k,g}^H \mathbf{Z}_g^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{n}_{k,g}^{tr}, \tag{38}$$

where $\bar{\mathbf{w}}_{k,g} = \sqrt{S_g}\tilde{\mathbf{w}}_{k,g} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_{S_g}\right)$.

The second term in the right-hand side of the above equation converges to zero, while the first term converges to $\frac{d_g}{K}\text{tr}\left(\mathbf{Z}_g^H \mathbf{R}_{BS_g} \mathbf{Q}_g\right)$, thus yielding:

$$\frac{1}{K}\mathbf{h}_{k,g}^H \hat{\mathbf{h}}_{k,g} - \frac{d_g}{K}\text{tr}\left(\mathbf{Z}_g^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{Z}_g\right) \xrightarrow[N \to \infty]{\text{a.s.}} 0. \tag{39}$$

In the double scattering channel model, $\mathbf{Z}_g$s are random and modeled as $\frac{1}{\sqrt{S_g}}\mathbf{R}_{BS_g}^{1/2}\mathbf{W}_g\bar{\mathbf{S}}_g^{1/2}$. We thus obtain:

$$\frac{d_g}{K}\text{tr}\left(\mathbf{Z}_g^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{Z}_g\right) = \frac{d_g}{K}\sum_{j=1}^{S_g}\mathbf{z}_{g,j}^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{z}_{g,j}$$

$$= \frac{d_g}{K}\sum_{j=1}^{S_g}\frac{\bar{s}_{g,j}}{S_g}\mathbf{w}_{g,j}^H \mathbf{R}_{BS_g}^{1/2^H} \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{R}_{BS_g}^{1/2} \mathbf{w}_{g,j}, \tag{40}$$

where $\mathbf{z}_{g,j} = \sqrt{\frac{\bar{s}_{g,j}}{S_g}} \mathbf{R}_{BS_g}^{1/2} \mathbf{w}_{g,j}$ is the jth column of $\mathbf{Z}_g$, $\bar{s}_{g,j}$ is the jth diagonal element of $\bar{\mathbf{S}}_g$ and $\mathbf{w}_{g,j}$ is the jth column of $\mathbf{W}_g$, which is composed of standard complex Gaussian elements. Thus, we can have following convergence by applying trace lemma on $\mathbf{w}_{g,j}$ as,

$$\frac{d_g}{K}\text{tr}\left(\mathbf{Z}_g^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{Z}_g\right) - \frac{d_g^2}{K}\text{tr}\left(\mathbf{R}_{BS_g}^2 \mathbf{Q}_g\right) \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{41}$$

Now combining (41) with (39) will yield,

$$\frac{1}{K}\mathbf{h}_{k,g}^H \hat{\mathbf{h}}_{k,g} - \frac{d_g^2}{K}\text{tr}\left(\mathbf{R}_{BS_g}^2 \mathbf{Q}_g\right) \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{42}$$

The above convergence not only implies almost sure convergence but also convergence in mean as the sequence of random variables $\frac{1}{K}\mathbf{h}_{k,g}^H \hat{\mathbf{h}}_{k,g}$ is uniformly integrable. Therefore, we have

$$\frac{p_{k,g}}{K}\left|\mathbb{E}\left[\mathbf{h}_{k,g}^H \hat{\mathbf{h}}_{k,g}\right]\right|^2 - \frac{p_{k,g}}{K}\left[d_g^2 \text{tr}\left(\mathbf{R}_{BS_g} \mathbf{R}_{BS_g} \mathbf{Q}_g\right)\right]^2 \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{43}$$

This completes the proof of **Lemma 1**.

APPENDIX B

PROOF OF LEMMA 2

Similar to the proof of **Lemma 1**, we also consider $\mathbf{Z}_g$ to be deterministic and use (17) to have

$$\frac{1}{K}\mathbf{h}_{k,g}^H \hat{\mathbf{H}}_{[k,g]}^H \mathbf{P}_{[k,g]} \hat{\mathbf{H}}_{[k,g]} \mathbf{h}_{k,g} = \frac{1}{K}\bar{\mathbf{w}}_{k,g}^H \mathbf{Z}_g^H \hat{\mathbf{H}}_{[k,g]}^H \mathbf{P}_{[k,g]} \hat{\mathbf{H}}_{[k,g]} \mathbf{Z}_g \bar{\mathbf{w}}_{k,g}. \tag{44}$$

Using the trace lemma, we have

$$\frac{1}{K}\mathbf{h}_{k,g}^H \hat{\mathbf{H}}_{[k,g]}^H \mathbf{P}_{[k,g]} \hat{\mathbf{H}}_{[k,g]} \mathbf{h}_{k,g} - \frac{1}{K}\text{tr}\left(\mathbf{Z}_g^H \hat{\mathbf{H}}_{[k,g]}^H \mathbf{P}_{[k,g]} \hat{\mathbf{H}}_{[k,g]} \mathbf{Z}_g\right) \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{45}$$

As the difference between $\text{tr}\left(\mathbf{Z}_g^H \hat{\mathbf{H}}^H \mathbf{P} \hat{\mathbf{H}} \mathbf{Z}_g\right) = \sum\limits_{i=1}^{G}\sum\limits_{l=1}^{K_i} p_{l,i}\hat{\mathbf{h}}_{l,i}^H \mathbf{Z}_g \mathbf{Z}_g^H \hat{\mathbf{h}}_{l,i}$ and $\text{tr}\left(\mathbf{Z}_g^H \hat{\mathbf{H}}_{[k,g]}^H \mathbf{P}_{[k,g]} \hat{\mathbf{H}}_{[k,g]} \mathbf{Z}_g\right)$ $= \sum\limits_{i=1}^{G}\sum\limits_{\substack{l=1 \\ (l,i)\neq(k,g)}}^{K_i} p_{l,i}\hat{\mathbf{h}}_{l,i}^H \mathbf{Z}_g \mathbf{Z}_g^H \hat{\mathbf{h}}_{l,i}$ is of order $O(1/K)$, we may substitute in (45) $\hat{\mathbf{H}}_{[k,g]}$ by $\hat{\mathbf{H}}$ and $\mathbf{P}_{[k,g]}$ by $\mathbf{P}$. Using (18) and applying trace lemma, we have following convergence

$$\frac{1}{K}\mathbf{h}_{k,g}^H \hat{\mathbf{H}}_{[k,g]}^H \mathbf{P}_{[k,g]} \hat{\mathbf{H}}_{[k,g]} \mathbf{h}_{k,g} - \sum_{i=1}^{G}\sum_{l=1}^{K_i} p_{l,i}T_{i,g} \xrightarrow[N\to\infty]{\text{a.s.}} 0, \tag{46}$$

where $T_{i,g} = \frac{1}{K}\text{tr}\left(\mathbf{\Phi}_i \mathbf{Z}_g \mathbf{Z}_g^H\right)$.

Replacing $\mathbf{\Phi}_i$ by (19), $T_{i,g}$ could be further expressed as

$$T_{i,g} = \frac{d_i^2}{K\rho_{tr}} \text{tr} \left(\mathbf{Z}_g^H \mathbf{R}_{BS_i} \mathbf{Q}_i \mathbf{Q}_i^H \mathbf{R}_{BS_i}^H \mathbf{Z}_g \right) + \frac{d_i^2}{K} \text{tr} \left(\mathbf{Z}_g^H \mathbf{R}_{BS_i} \mathbf{Q}_i \mathbf{Z}_i \mathbf{Z}_i^H \mathbf{Q}_i^H \mathbf{R}_{BS_i}^H \mathbf{Z}_g \right)$$

$$= \Upsilon_{i,g} + \Psi_{i,g}. \tag{47}$$

Note that $\Upsilon_{i,g}$ is still a function of $\mathbf{Z}_g$ which is random. We treat it in a same way as shown in (40) and use trace lemma to obtain the following deterministic approximation,

$$\Upsilon_{i,g} - \frac{d_i^2 d_g}{K\rho_{tr}} \text{tr} \left(\mathbf{R}_{BS_g} \mathbf{R}_{BS_i} \mathbf{Q}_i \mathbf{Q}_i^H \mathbf{R}_{BS_i}^H \right) \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{48}$$

To remove the dependence of $\Psi_{i,g}$ on $\mathbf{Z}_g$ and $\mathbf{Z}_i$, we need to divide it into two cases: $\bar{\Psi}_{i,g}$ for $i \neq g$ and $\Psi_{g,g}$ for $i = g$. For the first case, $\mathbf{Z}_g$ and $\mathbf{Z}_i$ are independent. We assume $\mathbf{Z}_i$ to be random and do a similar analysis as (40) and apply trace lemma on $\mathbf{w}_{i,j}$ to obtain,

$$\bar{\Psi}_{i,g} - \frac{d_i^3}{K} \text{tr} \left(\mathbf{R}_{BS_i} \mathbf{Q}_i^H \mathbf{R}_{BS_i}^H \mathbf{Z}_g \mathbf{Z}_g^H \mathbf{R}_{BS_i} \mathbf{Q}_i \right) \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{49}$$

Extending the analysis to random $\mathbf{Z}_g$ in a same way as done in (40), we have

$$\bar{\Psi}_{i,g} - \frac{d_i^3 d_g}{K} \text{tr} \left(\mathbf{R}_{BS_g} \mathbf{R}_{BS_i} \mathbf{Q}_i \mathbf{R}_{BS_i} \mathbf{Q}_i^H \mathbf{R}_{BS_i}^H \right) \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{50}$$

For the second case: $i = g$, $\Psi_{g,g}$ can be re-expressed as

$$\Psi_{g,g} = \frac{d_g^2}{K} \sum_{j=1}^{S_g} \mathbf{z}_{g,j}^H \mathbf{Q}_g^H \mathbf{R}_{BS_g}^H \left(\mathbf{Z}_g \mathbf{Z}_g^H - \mathbf{z}_{g,j} \mathbf{z}_{g,j}^H \right) \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{z}_{g,j}$$

$$+ \frac{d_g^2}{K} \sum_{j=1}^{S_g} \mathbf{z}_{g,j}^H \mathbf{Q}_g^H \mathbf{R}_{BS_g}^H \mathbf{z}_{g,j} \mathbf{z}_{g,j}^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{z}_{g,j}$$

$$= Z_{1,g} + Z_{2,g}. \tag{51}$$

Since we have removed the dependence of $Z_{1,g}$ on the vector $\mathbf{z}_{g,j}$, we can apply trace lemma to have,

$$Z_{1,g} - Z_{3,g} \xrightarrow[N\to\infty]{\text{a.s.}} 0, \tag{52}$$

where $Z_{3,g}$ could be expressed as

$$Z_{3,g} = \frac{d_g^2}{K} \sum_{j=1}^{S_g} \frac{\bar{s}_{g,j}}{S_g} \text{tr} \left(\mathbf{R}_{BS_g} \mathbf{Q}_g^H \mathbf{R}_{BS_g}^H \left(\mathbf{Z}_g \mathbf{Z}_g^H - \mathbf{z}_{g,j} \mathbf{z}_{g,j}^H \right) \mathbf{R}_{BS_g} \mathbf{Q}_g \right). \tag{53}$$

Next we re-express $Z_{3,g}$ as,

$$Z_{3,g} = \frac{d_g^2}{K} \sum_{j=1}^{S_g} \frac{\bar{s}_{g,j}}{S_g} \sum_{\substack{n=1 \\ n \neq j}}^{S_g} \mathbf{z}_{g,n}^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{R}_{BS_g} \mathbf{Q}_g^H \mathbf{R}_{BS_g}^H \mathbf{z}_{g,n}. \tag{54}$$

Using the trace lemma, we obtain the following convergence result for $Z_{3,g}$,

$$Z_{3,g} - \frac{d_g^2}{K} \sum_{j=1}^{S_g} \frac{\bar{s}_{g,j}}{S_g} \sum_{\substack{n=1 \\ n \neq j}}^{S_g} \frac{\bar{s}_{g,n}}{S_g} \text{tr} \left(\mathbf{R}_{BS_g} \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{R}_{BS_g} \mathbf{Q}_g^H \mathbf{R}_{BS_g}^H \right) \xrightarrow[N \to \infty]{\text{a.s.}} 0. \tag{55}$$

Substituting the results of (55) into (52), the final deterministic equivalent of $Z_{1,g}$ is obtained. Moreover, using the trace lemma on $Z_{2,g}$, we obtain,

$$Z_{2,g} - \frac{d_g^2}{K} \sum_{j=1}^{S_g} \frac{\bar{s}_{g,j}^2}{S_g^2} \text{tr} \left(\mathbf{R}_{BS_g} \mathbf{Q}_g^H \mathbf{R}_{BS_g}^H \right) \text{tr} \left(\mathbf{R}_{BS_g} \mathbf{R}_{BS_g} \mathbf{Q}_g \right) \xrightarrow[N \to \infty]{\text{a.s.}} 0. \tag{56}$$

Substituting the result of (56) and the result of (52) into (51), we obtain the deterministic approximation of $\Psi_{g,g}$. Combining this result with (49), we obtain the deterministic equivalent of $\Psi_{i,g}$, which is given by (22). Using (46) and (47), we finally obtain,

$$\frac{1}{K} \mathbf{h}_{k,g}^H \hat{\mathbf{H}}_{[k,g]}^H \mathbf{P}_{[k,g]} \hat{\mathbf{H}}_{[k,g]} \mathbf{h}_{k,g} - \sum_{i=1}^{G} \sum_{l=1}^{K_i} p_{l,i} \left(\Upsilon_{i,g}^o + \Psi_{i,g}^o \right) \xrightarrow[N \to \infty]{\text{a.s.}} 0. \tag{57}$$

In a similar way to the previous analysis, based on the fact that the sequence of random variables $\frac{1}{K} \mathbf{h}_{k,g}^H \hat{\mathbf{H}}_{[k,g]}^H \mathbf{P}_{[k,g]} \hat{\mathbf{H}}_{[k,g]} \mathbf{h}_{k,g}$ is uniformly integrable, the deterministic equivalent of **Lemma 2** is obtained.

APPENDIX C

PROOF OF LEMMA 3

First, we derive the deterministic equivalent of the variance term. Using the expression from (38) and denoting $\mathbf{X} = \frac{d_g}{K} \bar{\mathbf{w}}_{k,g}^H \mathbf{Z}_g^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{Z}_g \bar{\mathbf{w}}_{k,g}$ and $\mathbf{Y} = \frac{d_g}{K\sqrt{\rho_{tr}}} \bar{\mathbf{w}}_{k,g}^H \mathbf{Z}_g^H \mathbf{R}_{BS_g} \mathbf{Q}_g \mathbf{n}_{k,g}^{tr}$, $\text{var}\left(\frac{1}{K} \mathbf{h}_{k,g}^H \hat{\mathbf{h}}_{k,g} \right)$ can be re-expressed as

$$\text{var}\left(\frac{1}{K} \mathbf{h}_{k,g}^H \hat{\mathbf{h}}_{k,g} \right) = \text{var}\left(\mathbf{X} + \mathbf{Y} \right). \tag{58}$$

It can be seen that $\text{var}\left(\mathbf{X} + \mathbf{Y} \right)$ is bounded by $2\text{var}\left(\mathbf{X} \right) + 2\text{var}\left(\mathbf{Y} \right)$. The variance of $\mathbf{Y}$ can be bounded by $\mathbb{E}\left[\mathbf{Y}^2 \right]$ which converges to zero. As for the variance of $\mathbf{X}$, using the trace Lemma and based on the fact that $\sup_N \left\| \mathbf{Z}_g \mathbf{Z}_g^H \right\|_2 < \infty$ almost surely, we can prove that $\text{var}(\mathbf{X})$ converges to zero.

Thus, we can conclude that the noise term can be approximated in the almost sure sense by $\frac{\Theta}{\rho}$, which can be derived as follows.

Using the definition of $\mathbf{h}_{k,g}$ from (18), we have

$$\frac{1}{K} \hat{\mathbf{h}}_{k,g}^H \hat{\mathbf{h}}_{k,g} = \frac{1}{K} \bar{\mathbf{q}}_{k,g}^H \mathbf{\Phi}_g \bar{\mathbf{q}}_{k,g}. \tag{59}$$

Based on the trace lemma, we have

$$\frac{1}{K}\hat{\mathbf{h}}_{k,g}^{H}\hat{\mathbf{h}}_{k,g} - \frac{1}{K}\text{tr}\left(\mathbf{\Phi}_g\right) \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{60}$$

Using the definition of $\mathbf{\Phi}_g$ from (19), we have

$$\text{tr}\left(\mathbf{\Phi}_{\text{g}}\right) = d_g^2\text{tr}\left(\mathbf{Z}_g^H\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\mathbf{R}_{BS_g}\mathbf{Q}_g\mathbf{Z}_g\right) + \frac{d_g^2}{\rho_{tr}}\text{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\right). \tag{61}$$

The deterministic equivalent of $\frac{1}{K}\text{tr}\left(\mathbf{\Phi}_g\right)$ can be shown as

$$\frac{1}{K}\text{tr}\left(\mathbf{\Phi}_g\right) - \frac{1}{K}\left[d_g^3\text{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\mathbf{R}_{BS_g}\mathbf{Q}_g\right) + \frac{d_g^2}{\rho_{tr}}\text{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\right)\right] \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{62}$$

Combining (62) and (60) yields the deterministic equivalent of $\frac{1}{K}\hat{\mathbf{h}}_{k,g}^{H}\hat{\mathbf{h}}_{k,g}$. We can see that $\frac{1}{K}\text{tr}\left(\mathbf{P}\hat{\mathbf{H}}\hat{\mathbf{H}}^{H}\right) = \frac{1}{K}\sum_{g=1}^{G}\sum_{k=1}^{K_g}p_{k,g}\hat{\mathbf{h}}_{k,g}^{H}\hat{\mathbf{h}}_{k,g}$ leading to

$$\frac{\Theta}{K\rho} - \frac{1}{K\rho}\sum_{g=1}^{G}\sum_{k=1}^{K_g}p_{k,g}\left[\frac{d_g^2}{\rho_{tr}}\text{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\right) + d_g^3\text{tr}\left(\mathbf{R}_{BS_g}\mathbf{Q}_g^H\mathbf{R}_{BS_g}^H\mathbf{R}_{BS_g}\mathbf{Q}_g\right)\right] \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{63}$$

Combining the deterministic equivalent of $\frac{\Theta}{K\rho}$ with $\text{var}\left(\frac{1}{K}\mathbf{h}_{k,g}^{H}\hat{\mathbf{h}}_{k,g}\right) \xrightarrow[N\to\infty]{\text{a.s.}} 0$, the proof of the **Lemma 3** is completed.

APPENDIX D

PROOF OF COROLLARY 3

For a single group and $\bar{\mathbf{S}} = \mathbf{I}_S$ and $\mathbf{R}_{BS} = \mathbf{I}_N$, $d = \frac{1}{S}\text{tr}\left(\mathbf{S}\right) = 1$ and $\mathbf{Q} = \frac{\rho_{tr}}{1+\rho_{tr}}\mathbf{I}_N$. One can show by straightforward but tedious calculations that (43) and (63) can be reduced as,

$$\frac{p_k}{K}\left|\mathbb{E}\left[\mathbf{h}_k^H\hat{\mathbf{h}}_k\right]\right|^2 - \frac{p_{k,g}N^2\rho_{tr}^2}{K\left(1+\rho_{tr}\right)^2} \xrightarrow[N\to\infty]{\text{a.s.}} 0 \tag{64}$$

and,

$$\frac{\Theta}{K\rho} - \frac{P}{\rho}\left[\frac{N\rho_{tr}^2}{K\left(1+\rho_{tr}\right)^2} + \frac{N\rho_{tr}^2}{K\rho_{tr}\left(1+\rho_{tr}\right)^2}\right] \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{65}$$

Similarly, Υ and Ψ can also be reduced for the case $i = g$ as,

$$\Upsilon - \frac{N\rho_{tr}^2}{K\rho_{tr}\left(1+\rho_{tr}\right)^2} \xrightarrow[N\to\infty]{\text{a.s.}} 0 \tag{66}$$

and,

$$\Psi - \left[\frac{N\left(S-1\right)\rho_{tr}^2}{KS\left(1+\rho_{tr}\right)^2} + \frac{N^2\rho_{tr}^2}{KS\left(1+\rho_{tr}\right)^2}\right] \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{67}$$

Substituting (66) and (67) into (47) considering single group, we can further obtain the simplified expression for the interference term. Finally, combining all these terms into (13) and dividing the same terms $\frac{N\rho_{tr}^2}{K(1+\rho_{tr})^2}$ both on numerator and denominator, the proof of the **Corollary 3** is completed.

REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[2] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surv. Tutor.*, vol. 21, no. 4, pp. 3109–3132, 2019.

[3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[4] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[5] Q. Nadeem, A. Kammoun, and M.-S. Alouini, "Elevation beamforming with full dimension MIMO architectures in 5G systems: A tutorial," *IEEE Commun. Surv. Tutor.*, pp. 1–1, 2019.

[6] O. Ozdogan, E. Bjornson, and E. G. Larsson, "Massive MIMO with spatially correlated rician fading channels," *IEEE Trans. on Commun.*, vol. 67, no. 5, pp. 3234–3250, 2019.

[7] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in sub-6 Ghz and mmWave: Physical, practical, and use-case differences," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 100–108, 2019.

[8] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2002, pp. 406–410.

[9] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, 2002.

[10] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.

[11] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. commun.*, vol. 48, no. 3, pp. 502–513, 2000.

[12] H. Shin and M. Z. Win, "MIMO diversity in the presence of double scattering," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2976–2996, Jul. 2008.

[13] X. Li, S. Jin, X. Gao, M. R. Mckay, and K. Wong, "Transmitter optimization and beamforming optimality conditions for double-scattering MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3647–3654, Sep. 2008.

[14] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2636–2647, 2003.

[15] S. Yang and J.-C. Belfiore, "Diversity-multiplexing tradeoff of double scattering MIMO channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2027–2034, 2011.

[16] H. Zhang, S. Jin, M. R. McKay, X. Zhang, and D. Yang, "High-SNR performance of MIMO multi-channel beamforming in double-scattering channels," *IEEE Trans. Commun.*, vol. 59, no. 6, pp. 1621–1631, 2011.

[17] Z. Zheng, L. Wei, R. Speicher, R. R. Müller, J. Hämäläinen, and J. Corander, "Asymptotic analysis of rayleigh product channels: A free probability approach," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1731–1745, 2017.

[18] A. Papazafeiropoulos, S. K. Sharma, T. Ratnarajah, and S. Chatzinotas, "Impact of residual additive transceiver hardware impairments on rayleigh-product MIMO channels with linear receivers: Exact and asymptotic analyses," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 105–118, 2018.

[19] S. Ö. Ata and I. Altunbaş, "STTC design for vehicular communication systems employing fixed-gain AF PLNC over cascaded fading channels," *IET Commun.*, vol. 12, no. 11, pp. 1283–1289, 2018.

[20] R. R. Muller and H. Hofstetter, "A random matrix model of communication via antenna arrays," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2495–2506, 2002.

[21] J. Hoydis, R. Couillet, and M. Debbah, "Asymptotic analysis of double-scattering channels," in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2011, pp. 1935–1939.

[22] T. Van Chien, E. Björnson, and E. G. Larsson, "Multi-cell massive MIMO performance with double scattering channels," in *Proc. Int. Workshop Comput. Aided Modelling Design Commun. Links and Netw. (CAMAD)*, Oct. 2016, pp. 231–236.

[23] Q. Nadeem, A. Kammoun, M. Debbah, and M.-S. Alouini, "Asymptotic analysis of RZF over double scattering channels with MMSE estimation," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2509–2526, May 2019.

[24] ——, "Asymptotic analysis of regularized zero-forcing in double scattering channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[25] X. Li, S. Jin, X. Gao, and M. R. McKay, "Capacity bounds and low complexity transceiver design for double-scattering MIMO multiple access channels," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2809–2822, 2010.

[26] K. Ntougias, D. Ntaikos, and C. B. Papadias, "Coordinated MIMO with single-fed load-controlled parasitic antenna arrays," in *Proc. IEEE Signal Process. Adv. Wireless Commun. (SPAWC)*, 2016, pp. 1–5.

[27] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: A survey," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3920–3931, 2017.

[28] A. Kammoun, A. Müller, E. Björnson, and M. Debbah, "Linear precoding based on polynomial expansion: Large-scale multi-cell MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 861–875, 2014.

[29] P. Billingsley, *Probability and Measure*. Hoboken, NJ, USA: Wiley, 1995.

[30] S. Jin, M. R. McKay, K.-K. Wong, and X. Gao, "Transmit beamforming in rayleigh product MIMO channels: Capacity and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5204–5221, 2008.

[31] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the analysis of correlated MIMO multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3493–3514, 2011.

[32] A. Mueller, A. Kammoun, E. Björnson, and M. Debbah, "Linear precoding based on polynomial expansion: Reducing complexity in massive MIMO," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 1, p. 63, 2016.