Revisiting MMSE Combining for Massive MIMO Over Heterogeneous Propagation Channels

Harsh Tataria*, Peter J. Smith†, Michail Matthaiou*, Hien Quoc Ngo*, and Pawel A. Dmochowski‡
*Institute of Electronics, Communications and Information Technology (ECIT), Queen's University Belfast, Belfast, U.K.
†School of Mathematics and Statistics, Victoria University of Wellington, Wellington, New Zealand
‡School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand
e–mail:{h.tataria, m.matthaiou, hien.ngo}@qub.ac.uk, peter.smith@vuw.ac.nz, pawel.dmochowski@ecs.vuw.ac.nz

***Abstract*—We consider a massive multiple–input multiple–output system with minimum–mean–squared–error processing on the uplink. A novel analytical framework is proposed to approximate the instantaneous signal–to–interference–plus–noise–ratio (SINR) of an arbitrary user terminal, as well as, the system sum spectral efficiency. Unlike previous studies, our methodology considers spatially correlated Ricean fading, with unequal Ricean K–factors, spatial correlation matrices and link gains across all terminals. Under this fully heterogeneous setting, we demonstrate that the SINR of a terminal can be tightly approximated by a linear combination of non–central chi–squared random variables, where the scaling depends on the individual link gains, K–factors, and eigenvalues of the terminal specific correlation matrices. Our approximations remain tight across the considered spatial correlation models, K–factor models, average uplink signal–to–noise–ratios and number of receive antennas. Leveraging the general form of the SINR and sum spectral efficiency, an analytical method to approximate their statistical moments is presented utilizing the moment generating function. The generality of the aforementioned analytical results is demonstrated via several special cases of practical relevance.**

I. Introduction

It is now widely accepted that massive multiple–input multiple–output (massive MIMO) systems will define the physical layer characteristics of fifth–generation cellular systems [1]. This has led to a flurry of research activities in this area. With its ability to aggressively multiplex tens of user terminals within the same time–frequency resource, studies on massive MIMO have continued to focus on linear signal processing techniques to achieve near–to optimal spectral efficiencies [2, 3]. Though massive MIMO is primarily intended to operate in the centimeter–wave (cmWave) frequency band ($<$ 6 GHz), its adoption in the millimeter–wave (mmWave) bands (30–300 GHz) has been the subject of much recent debate [1, 4]. While massive MIMO is beneficial at cmWave frequencies, it is essential at mmWave, providing the much needed array gain in order to mitigate the large propagation losses.

Understanding the differences between the cmWave and mmWave propagation channel has prompted numerous measurement campaigns, shaping unique features in state–of–the–art directional propagation models (see e.g., [2, 5–8]). These include channel sparsity, prominence of line–of–sight (LoS), as well as channel heterogeneity, where large–scale spatial parameters, e.g. power in the LoS and non LoS (NLoS) components, cluster angular spreads, etc., of a particular terminal are variable not only with the link distance, but also with the terrain on the link between the terminal and the point of service [9]. Due to its extremely short wavelengths, mmWaves are unable to propagate over large distances, and are primarily envisaged to operate in small–cellular scenarios, where there is a considerably higher probability of a terminal experiencing LoS propagation. To this end, several investigations have examined the performance of massive MIMO with Ricean fading [10–13], including studies which consider variability in the Ricean K–factors across multiple terminals [9, 14].

Nevertheless, almost all of these studies neglect the presence of spatially correlated diffuse multipath components (MPCs), which are necessary to capture the generality of the propagation manifestations. On the other hand, without considering LoS, the literature also reports several studies which consider variable MPC statistics across multiple terminals [15–17]. These studies investigate the limiting signal–to–interference–plus–noise–ratio (SINR) and sum spectral efficiency. However, all of the analytical results are left in terms of numerical fixed point algorithms, which require extensive effort to generalize to different types of propagation channels, as well as uplink/downlink multiuser processing techniques. Moreover, it is difficult to obtain insights into the operation of the massive MIMO system from such numerical solutions. Ultimately, it is desirable to have a performance analysis, which is simple and insightful, yet general and able to cater for full heterogeneity (unequal LoS, spatial correlation, and link gains) in the propagation channel. To the best of our knowledge, such analysis only exists for the simple case of maximum–ratio combining on the uplink, shown in our recently published work [9]. It remains a substantial analytical challenge to devise a general methodology towards characterizing the performance of massive MIMO systems in heterogeneous channels, with minimum–mean–squared-error (MMSE) combining, which gives near optimal massive MIMO performance across the entire operating signal–to–noise–ratio (SNR) range. This is precisely the focus of the paper.

Existing works specific to MMSE processing have concentrated on homogeneous scenarios, where a wide range of analytical results are derived (see e.g., [3, 18–23] and references therein). Unlike these studies, we introduce a novel, general, analytical framework for approximating the instantaneous SINR and sum spectral efficiency of a massive MIMO system with uplink MMSE processing under a fully heterogeneous propagation channel. Our analysis methodology caters for unequally correlated Ricean fading channels with variable K-factors and spatial correlation matrices across terminals. Under this general setting, we show that the MMSE SINR of

978-1-5386-3180-5/18/$31.00 ©2018 IEEE

an arbitrary terminal, as well as the sum spectral efficiency can be expressed as a linear combination of non–central chi–squared random variables, where the scaling is a function of the individual link gains, Ricean K–factors, and eigenvalues of the terminal correlation matrix. To the best of the authors' knowledge, this is the only available result analyzing MMSE processing with full heterogeneity in the propagation channel. We further leverage the general form of the MMSE SINR and present an analytical method to approximate its statistical moments via the moment generating function (MGF). The MGF results lead to simple first and second–order statistics of the MMSE SINR, demonstrating the average behavior along with the variability of the MMSE SINR and consequently sum spectral efficiency. It is noteworthy that, even though not shown in the paper, the MGF results are useful for further characterization of metrics such as symbol error rates and sum spectral efficiency variance of fully heterogeneous fading channels. To demonstrate the generality of the derived results, we present several special cases in scenarios such as pure NLoS propagation with fixed and variable correlation matrices, as well as, Ricean fading with equal correlation matrices.

Notation. Boldface upper and lower case symbols are used to denote matrices and vectors, while lightface upper and lower case symbols denote scalar quantities. The $M \times M$ identity matrix is denoted as $\mathbf{I}_M$. The (i,j)–th element of a matrix $\mathbf{H}$ is denoted by $[\mathbf{H}]_{i,j}$, and diag $(h_1, h_2, \ldots, h_M)$ denotes a $M \times M$ diagonal matrix with $h_1, h_2, \ldots, h_M$ on the main diagonal and zeros elsewhere. The transpose, Hermitian transpose, inverse, and trace operators are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^{-1}$, and tr $\{\cdot\}$, respectively. $||\cdot||_{\mathrm{F}}$ and $|\cdot|$ denote the Frobenius and scalar norms. We use $\mathbf{h} \sim \mathcal{CN}(\mathbf{m}, \mathbf{Q})$ to denote a complex Gaussian distribution for $\mathbf{h}$ with mean $\mathbf{m}$ and covariance matrix $\mathbf{Q}$. Similarly, $h \sim \mathcal{U}[a,b]$ is used to denote a uniform random variable for h taking on values from a to b. Finally, $\mathbb{E}\{h\}$ and Var $\{h\}$ denote the statistical expectation and variance of the random variable h, respectively.

II. System and Propagation Models

The uplink of a massive MIMO system operating in an urban microcellular environment (UMi) is considered. The base station (BS) is located at the center of a circular cell with radius $\mathrm{R_c}$, and is equipped with a uniform linear array (ULA) of M elements. The ULA communicates with L single–antenna user terminals in the same time–frequency resource ($M \gg L$). Narrowband transmission is assumed with no uplink power control. Without loss of generality, terminal 1 is considered as desired, and hence terminals $2, \ldots, L$ are considered as interferers. With channel knowledge at the BS, the $M \times 1$ received signal at the BS array can be written as

$$\mathbf{y} = \rho^{\frac{1}{2}} \mathbf{G} \mathbf{D}^{\frac{1}{2}} \mathbf{s} + \mathbf{n}. \tag{1}$$

Remark 1. The assumption of perfect channel knowledge may, at first sight, seem rather naive. However, there are several reasons for this: Firstly, the central focus of the paper is to devise a general analytical method to approximate the performance of uplink MMSE processing under fully heterogeneous channels. Here, each terminal has a specific correlation matrix, Ricean K–factor, and link gain. Under this general scenario, it is extremely difficult, if not intractable, to make analytical progress without perfect channel knowledge. Secondly, in line with [4], this assumption is reasonable in scenarios with low mobility, where a large amount of the coherence interval can be spent for uplink training. Thirdly, it is noteworthy that the results obtained from the subsequent analysis can be regarded as a useful upper bound on what may be achieved in practice with imperfect channel knowledge.

In (1), ρ denotes the average uplink transmit power, $\mathbf{G}$ is the $M \times L$ small–scale fading propagation matrix between the M BS antennas and L terminals (discussed later in the section), $\mathbf{D}$ is an $L \times L$ diagonal matrix of link gains, where the link gain for terminal 1 is given by $[\mathbf{D}]_{1,1} = \beta_1$, and is composed of the large-scale fading effects in geometric attenuation and shadow-fading. We note that $\beta_1 = A\zeta_1 (r_0/r_1)^{\alpha}$. Specifically, A is the unit–less constant for geometric attenuation at the reference distance r_0, r_1 is the link distance between terminal 1 and the BS, α is the attenuation exponent, and ζ_1 models the effects of shadow–fading via the log–normal density, such that $10\log_{10}(\zeta_1) \sim \mathcal{N}\left(0, \sigma_{\mathrm{sh}}^2\right)$, where σ_{sh} is the shadow–fading standard deviation. The $L \times 1$ vector of uplink data symbols is denoted by $\mathbf{s}$, such that the ℓ–th entry of $\mathbf{s}$, s_ℓ, has $\mathbb{E}\{|s_\ell|^2\} = 1, \forall \ell = 1, 2, \ldots, L$. The $M \times 1$ vector of additive Gaussian noise is denoted by $\mathbf{n}$, such that the ℓ–th entry of $\mathbf{n}$, $n_\ell \sim \mathcal{CN}\left(0, \sigma^2\right)$. Note that $\sigma^2 = 1$, and is fixed for any ℓ. As such, the average uplink SNR, defined as $\rho/\sigma^2 = \rho$. Further to the above, in line with [7–9, 14], we employ a probability based approach to determine if a given terminal experiences LoS or NLoS propagation. The LoS and NLoS probabilities are a function of the link distance, from which the LoS and NLoS geometric attenuation and other link characteristics are obtained. The terminal dependent K–factors are assumed to follow a log–normal density with the mean and variance specified in [7] and [8]. Further details of this and other large–scale parameters is provided in Section VI.

The $M \times 1$ small–scale fading vector from terminal 1 to the BS is denoted by $\mathbf{g}_1$, and forms the first column of the $M \times L$ composite channel matrix $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_L]$. Specifically,

$$\mathbf{g}_1 = \sqrt{\frac{K_1}{K_1+1}}\widehat{\mathbf{g}}_1 + \sqrt{\frac{1}{1+K_1}}\mathbf{R}_1^{\frac{1}{2}}\widetilde{\mathbf{g}}_1 = \underbrace{\widehat{\kappa}_1\widehat{\mathbf{g}}_1}_{\mathbf{v}_1} + \underbrace{\widetilde{\kappa}_1\mathbf{R}_1^{\frac{1}{2}}\widetilde{\mathbf{g}}_1}_{\mathbf{w}_1}. \tag{2}$$

Note that the $M \times 1$ vectors of specular and diffuse MPCs are denoted by $\widehat{\mathbf{g}}_1$ and $\widetilde{\mathbf{g}}_1$. Moreover, K_1 denotes the K–factor unique to terminal 1, and is a function of the terrain between the BS and the terminal's physical location. Note that $\widetilde{\mathbf{g}}_1 \sim \mathcal{CN}(0, \mathbf{I}_M)$ and $\widehat{\mathbf{g}}_1 = [1, e^{j2\pi d\cos(\phi_1)}, \ldots, e^{j2\pi d(M-1)\cos(\phi_1)}]$. Here, d is the antenna spacing between successive elements normalized by λ, the wavelength associated with the operating carrier frequency, f_{c}. Note that ϕ_1 is the LoS angle for terminal 1. In addition to the specular components, we consider spatially correlated MPCs. Hence, unlike previous works (see e.g., [10–12]), we define a terminal specific $M \times M$ spatial correlation matrix for terminal 1 as $\mathbf{R}_1$. For consistency, we delay further discussion on the possible models for $\mathbf{R}_1$ to Section VI.

Let $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_L]$, be the $M \times L$ linear combiner, which relies on $\mathbf{G}$. Using $\mathbf{T}$, the received signal is demultiplexed into individual data streams by multiplication with $\mathbf{T}^{\mathrm{H}}$. This results in an $L \times 1$ signal $\mathbf{r} = \mathbf{T}^{\mathrm{H}}\mathbf{y} = \rho^{\frac{1}{2}}\mathbf{T}^{\mathrm{H}}\mathbf{G}\mathbf{D}^{\frac{1}{2}}\mathbf{s} + \mathbf{T}^{\mathrm{H}}\mathbf{n}$. For terminal 1, we take the first component of $\mathbf{r}$, giving

$$r_1 = \rho^{\frac{1}{2}}\beta_1^{\frac{1}{2}}\mathbf{t}_1^{\mathrm{H}}\mathbf{g}_1 s_1 + \rho^{\frac{1}{2}}\sum_{i=2}^{L}\beta_i^{\frac{1}{2}}\mathbf{t}_1^{\mathrm{H}}\mathbf{g}_i s_i + \mathbf{g}_1^{\mathrm{H}}\mathbf{n}. \quad (3)$$

Considering the structure of the MMSE combiner we know that, $\mathbf{T} = \mathbf{G}(\mathbf{G}^{\mathrm{H}}\mathbf{G} + \sigma^2\mathbf{I}_L)^{-1}$. Following the methodology in [19, 20], the corresponding SINR for terminal 1 with uplink MMSE processing can be written as

$$\mathrm{SINR}_1 = \rho\beta_1\mathbf{g}_1^{\mathrm{H}}\left(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}} + \sigma^2\mathbf{I}_M\right)^{-1}\mathbf{g}_1, \quad (4)$$

where

$$\mathbf{X}_1 = \rho^{\frac{1}{2}}\mathbf{G}_1\mathbf{D}_1^{\frac{1}{2}}, \quad (5)$$

where $\mathbf{G}_1 = [\mathbf{g}_2, \mathbf{g}_3, \ldots, \mathbf{g}_L]$ and $\mathbf{D}_1 = \mathrm{diag}\,(\beta_2, \beta_3, \ldots, \beta_L)$ denote the concatenated channel and link gain matrices for the interfering set of user terminals (i.e., excluding terminal 1). Following this, some straightforward algebraic manipulations allows one to write the MMSE SINR as

$$\mathrm{SINR}_1 = \rho\beta_1\mathbf{g}_1^{\mathrm{H}}\left(\mathbf{G}_1\mathbf{D}_1\mathbf{G}_1^{\mathrm{H}} + \frac{\sigma^2}{\rho}\mathbf{I}_M\right)^{-1}\mathbf{g}_1. \quad (6)$$

As $\sigma^2 = 1$, we can write (6) as

$$\mathrm{SINR}_1 = \rho\beta_1\mathbf{g}_1^{\mathrm{H}}\left(\mathbf{G}_1\mathbf{D}_1\mathbf{G}_1^{\mathrm{H}} + \frac{1}{\rho}\mathbf{I}_M\right)^{-1}\mathbf{g}_1. \quad (7)$$

The SINR in (7) can be translated into a uplink spectral efficiency for terminal 1 (in bps/Hz), denoted by $\mathrm{SE}_1 = \log_2(1 + \mathrm{SINR}_1)$. As such, the sum spectral efficiency across all L terminals is given by

$$\mathrm{SE}_{\mathrm{sum}} = \sum_{\ell=1}^{L}\log_2\left(1 + \mathrm{SINR}_\ell\right). \quad (8)$$

In the section which follows, we present a general analytical methodology to approximate (7) and consequently (8) under fully heterogeneous propagation channels.

III. Analysis Methodology and Implications

Via the Matrix Inversion Lemma [19], we can express the inverse in (7) as (9), shown on top of the following page for reasons of space. Now, we know that

$$\frac{1}{M}\mathbf{G}_1^{\mathrm{H}}\mathbf{G}_1 = \frac{1}{M}\left(\mathbf{V}_1^{\mathrm{H}} + \mathbf{W}_1^{\mathrm{H}}\right)\left(\mathbf{V}_1 + \mathbf{W}_1\right) \quad (10)$$
$$= \underbrace{\frac{1}{M}\left(\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1 + \mathbf{W}_1^{\mathrm{H}}\mathbf{W}_1\right)}_{\mathbf{X}} + \underbrace{\frac{1}{M}\left(\mathbf{V}_1^{\mathrm{H}}\mathbf{W}_1 + \mathbf{W}_1^{\mathrm{H}}\mathbf{V}_1\right)}_{\mathbf{Y}},$$

where $\mathbf{V}_1 = [\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_L]$ and $\mathbf{W}_1 = [\mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_L]$. We further note that $\mathbb{E}\{\mathbf{Y}\} = \mathbf{0}$, and $\mathbb{E}\{\mathbf{X}\} = \frac{1}{M}\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_1 + \mathrm{diag}((\widetilde{\kappa}_2)^2, (\widetilde{\kappa}_3)^2, \ldots, (\widetilde{\kappa}_L)^2) = \mathbf{Q}_1$, so that $\frac{1}{M}\mathbf{G}_1^{\mathrm{H}}\mathbf{G}_1 = \mathbf{Q}_1 + \mathbf{\Xi}$, where $\mathbf{\Xi}$ is a zero–mean matrix. Some lengthy, yet straightforward calculations show that $\mathrm{Var}\{[\mathbf{\Xi}]_{i,j}\} \to 0$, as $M \to \infty$, $\forall i, j$, under the following conditions:

$$1)\ \frac{\mathrm{tr}\left\{\mathbf{R}_\ell\mathbf{R}_k\right\}}{M^2} \to 0,\ \text{as } M \to \infty,\ \forall \ell, k$$
$$2)\ \frac{\widehat{\mathbf{g}}_i^{\mathrm{H}}\mathbf{R}_j\widehat{\mathbf{g}}_i}{M^2} \to 0,\ \text{as } M \to \infty,\ \forall i, j. \quad (11)$$

Hence, under the conditions in (11), we have $\frac{1}{M}\mathbf{G}_1^{\mathrm{H}}\mathbf{G}_1 \overset{\text{m.s.}}{\to} \mathbf{Q}_1$, and this motivates us to replace $\mathbf{G}_1^{\mathrm{H}}\mathbf{G}_1$ in (9) by $M\mathbf{Q}_1$. We denote the mean–squared (m.s.) convergence of $\mathbf{A}$ to $\mathbf{B}$ by $\mathbf{A} \overset{\text{m.s.}}{\to} \mathbf{B}$. This holds for the remainder of the paper.

Remark 2. Closely inspecting conditions 1) and 2) in (11) shows that the trace results are valid unless extremely high spatial correlation exists, where one or more eigenvalues of a correlation matrix must remain of $\mathcal{O}(M)$, as $M \to \infty$. For the quadratic forms involving LoS steering vectors, these are convergent, unless, a LoS steering vector has a non–vanishing component in the direction of an eigenvector of a correlation matrix, which has a corresponding eigenvalue of $\mathcal{O}(M)$. Hence, both a very large eigenvalue and strong alignment are required to prevent convergence. Overall, we conclude that the approximation based on mean–squared (m.s.) convergence is reasonable except in extreme scenarios which lead to very peculiar propagation channels. Following this argument, for any finite value of M, we can state that

$$\left(\mathbf{G}_1\mathbf{D}_1\mathbf{G}_1^{\mathrm{H}} + \frac{1}{\rho}\mathbf{I}_M\right)^{-1} \approx \rho\mathbf{I}_M - \rho^2\mathbf{G}_1\mathbf{S}_1\mathbf{G}_1^{\mathrm{H}}, \quad (12)$$

where $\mathbf{S}_1 = (\mathbf{D}_1^{-1} + \rho M\mathbf{Q}_1)^{-1}$. Following this, we can write

$$\mathrm{SINR}_1 \approx \rho\beta_1\mathbf{g}_1^{\mathrm{H}}\left(\rho\mathbf{I}_M - \rho^2\mathbf{G}_1\mathbf{S}_1\mathbf{G}_1^{\mathrm{H}}\right)\mathbf{g}_1. \quad (13)$$

To make further progress, we introduce a second approximation, which relies on replacing $\mathbf{G}_1\mathbf{S}_1\mathbf{G}_1^{\mathrm{H}}$ by $\mathbb{E}\{\mathbf{G}_1\mathbf{S}_1\mathbf{G}_1^{\mathrm{H}}\} = \mathbf{T}_1$. In order to motivate this, we write (14), shown on top of the following page. Now, the structure of (14) is $\mathbf{g}_1^{\mathrm{H}}(\rho\mathbf{I}_M - \rho^2\mathbf{G}_1\mathbf{S}_1\mathbf{G}_1^{\mathrm{H}})\mathbf{g}_1 = \mathbf{g}_1^{\mathrm{H}}\left(\rho\mathbf{I}_M - \rho^2\mathbf{T}_1\right)\mathbf{g}_1\left[1 + (\delta_1/\delta_2)\right]$. In what follows, we record three key properties of δ_1 and δ_2:

1) $\mathbb{E}\{\delta_1\} = 0$ and $\mathbb{E}\{\delta_2\} > 0$, such that the correction term has zero mean.
2) If the interfering terminals are experiencing pure LoS propagation, then $\delta_1 = 0$.
3) If the interfering terminals experience pure NLoS propagation, then $\mathbf{Q}_1 = \mathbf{I}_L$, and the diagonal matrix $\mathbf{S}_1$ contains entries of $\mathcal{O}(1/M)$.

From the above properties, we conclude that purely scattered channels give the largest values of δ_1/δ_2, but here the numerator has an extram $1/M$ scaling due to $\mathbf{S}_1$, which makes $\delta_1 \overset{\text{m.s.}}{\to} 0$, as $M \to \infty$. An in–depth analysis for this convergence is not shown here in the interest of space, and will feature in the upcoming journal version of the paper. This motivates the simple approximation, allowing us to write

$$\mathrm{SINR}_1 \approx \rho\beta_1\mathbf{g}_1^{\mathrm{H}}\left(\rho\mathbf{I}_M - \rho^2\mathbf{T}_1\right)\mathbf{g}_1. \quad (15)$$

We further note that even though the aforementioned approximations are tailored for massive MIMO scenarios, by relying on the averaging offered by large values of M, in Section VI, we demonstrate that they remain tight even for moderate values of M.

Substituting the LoS as well as diffuse MPCs of terminals $2, 3, \ldots, L$ allows us to express $\mathbf{G}_1 = \mathbf{V}_1 + \mathbf{W}_1$, so that

$$\mathbf{T}_1 = \mathbb{E}\left\{\mathbf{G}_1\mathbf{S}_1\mathbf{G}_1^{\mathrm{H}}\right\} = \mathbb{E}\left\{\left(\mathbf{V}_1 + \mathbf{W}_1\right)\mathbf{S}_1\left(\mathbf{V}_1^{\mathrm{H}} + \mathbf{W}_1^{\mathrm{H}}\right)\right\}$$
$$= \mathbf{V}_1\mathbf{S}_1\mathbf{V}_1^{\mathrm{H}} + \mathbb{E}\left\{\mathbf{W}_1\mathbf{S}_1\mathbf{W}_1^{\mathrm{H}}\right\}. \quad (16)$$

Evaluating the expectation in (16), and performing some routine algebraic manipulations results in

$$\left(\mathbf{G}_1\mathbf{D}_1\mathbf{G}_1^{\mathrm{H}}+\frac{1}{\rho}\mathbf{I}_M\right)^{-1}=\rho\,\mathbf{I}_M-\rho^2\mathbf{G}_1\left(\mathbf{D}_1^{-1}+\rho\,\mathbf{G}_1^{\mathrm{H}}\mathbf{G}_1\right)^{-1}\mathbf{G}_1^{\mathrm{H}}. \tag{9}$$

$$\mathbf{g}_1^{\mathrm{H}}\left(\rho\mathbf{I}_M-\rho^2\mathbf{G}_1\mathbf{S}_1\mathbf{G}_1^{\mathrm{H}}\right)\mathbf{g}_1=\mathbf{g}_1^{\mathrm{H}}\left(\rho\mathbf{I}_M-\rho^2\mathbf{T}_1\right)\mathbf{g}_1\left[1+\frac{\rho^2\mathbf{g}_1^{\mathrm{H}}\left(\mathbf{G}_1\mathbf{S}_1\mathbf{G}_1^{\mathrm{H}}-\mathbf{T}_1\right)\mathbf{g}_1/M}{\mathbf{g}_1^{\mathrm{H}}\left(\rho\mathbf{I}_M-\rho^2\mathbf{T}_1\right)\mathbf{g}_1/M}\right] \tag{14}$$

$$\begin{aligned}\mathbb{E}\left\{\mathbf{W}_1\mathbf{S}_1\mathbf{W}_1^{\mathrm{H}}\right\}&=\sum_{i=2}^{L}\mathbb{E}\left\{\widetilde{\kappa}_i\mathbf{R}_i^{\frac{1}{2}}\widetilde{\mathbf{h}}_i\left[\mathbf{S}_1\right]_{i,i}\widetilde{\mathbf{h}}_i^{\mathrm{H}}\mathbf{R}_i^{\frac{1}{2}}\widetilde{\kappa}_i\right\}\\&=\sum_{i=2}^{L}\left(\widetilde{\kappa}_i\right)^2\left[\mathbf{S}_1\right]_{i,i}\mathbf{R}_i.\end{aligned} \tag{17}$$

Working backwards to substitute (17) into (16), (16) into (14), and (14) into (13), allows one to express (15) as

$$\begin{aligned}\mathrm{SINR}_1\approx\rho\beta_1\mathbf{g}_1^{\mathrm{H}}\Bigg\{\rho\mathbf{I}_M-\rho^2\Bigg[&\mathbf{V}_1\mathbf{S}_1\mathbf{V}_1^{\mathrm{H}}\\&+\sum_{i=2}^{L}\left(\widetilde{\kappa}_i\right)^2\left[\mathbf{S}_1\right]_{i,i}\mathbf{R}_i\Bigg]\Bigg\}\mathbf{g}_1.\end{aligned} \tag{18}$$

Denoting the constant matrix $\mathbf{\Theta}_1=\rho\mathbf{I}_M-\rho^2\mathbf{T}_1=\rho\mathbf{I}_M-\rho^2[\mathbf{V}_1\mathbf{S}_1\mathbf{V}_1^{\mathrm{H}}+\sum_{i=2}^{L}(\widetilde{\kappa}_i)^2[\mathbf{S}_1]_{i,i}\mathbf{R}_i]$, we can write (18) as

$$\mathrm{SINR}_1\approx\rho\beta_1\mathbf{g}_1^{\mathrm{H}}\mathbf{\Theta}_1\mathbf{g}_1. \tag{19}$$

Substituting the definition of $\mathbf{g}_1$ from (2), the MMSE SINR for terminal 1 can be written as

$$\begin{aligned}\mathrm{SINR}_1&\approx\rho\beta_1\left(\widehat{\kappa}_1\widehat{\mathbf{g}}_1^{\mathrm{H}}+\widetilde{\kappa}_1\widetilde{\mathbf{g}}_1^{\mathrm{H}}\mathbf{R}_1^{\frac{1}{2}}\right)\mathbf{\Theta}_1\left(\widehat{\kappa}_1\widehat{\mathbf{g}}_1+\widetilde{\kappa}_1\mathbf{R}_1^{\frac{1}{2}}\widetilde{\mathbf{g}}_1\right) \quad (20)\\&=\rho\beta_1\left(\widetilde{\kappa}_1\right)^2\left(\frac{\widehat{\kappa}_1}{\widetilde{\kappa}_1}\widehat{\mathbf{g}}_1^{\mathrm{H}}\mathbf{R}_1^{-\frac{1}{2}}+\widetilde{\mathbf{g}}_1^{\mathrm{H}}\right)\mathbf{R}_1^{\frac{1}{2}}\mathbf{\Theta}_1\mathbf{R}_1^{\frac{1}{2}}\left(\frac{\widehat{\kappa}_1}{\widetilde{\kappa}_1}\mathbf{R}_1^{-\frac{1}{2}}\widehat{\mathbf{g}}_1+\widetilde{\mathbf{g}}_1\right).\end{aligned}$$

Via an eigenvalue decomposition, $\mathbf{R}_1^{\frac{1}{2}}\mathbf{\Theta}_1\mathbf{R}_1^{\frac{1}{2}}=\mathbf{\Phi}_1^{\mathrm{H}}\mathbf{\Lambda}_1\mathbf{\Phi}_1$, one can express (20) as

$$\begin{aligned}\mathrm{SINR}_1&\approx\rho\,\beta_1\left(\widetilde{\kappa}_1\right)^2\left(\sqrt{K_1}\,\widehat{\mathbf{g}}_1^{\mathrm{H}}\mathbf{R}_1^{-\frac{1}{2}}+\widetilde{\mathbf{g}}_1^{\mathrm{H}}\right)\mathbf{\Phi}_1^{\mathrm{H}}\mathbf{\Lambda}_1\mathbf{\Phi}_1\\&\quad\times\left(\sqrt{K_1}\mathbf{R}_1^{-\frac{1}{2}}\widehat{\mathbf{g}}_1+\widetilde{\mathbf{g}}_1\right)\\&=\frac{\rho\beta_1}{K_1+1}\left(\sqrt{K_1}\,\widehat{\mathbf{g}}_1^{\mathrm{H}}\mathbf{R}_1^{-\frac{1}{2}}\mathbf{\Phi}_1^{\mathrm{H}}+\bar{\mathbf{g}}_l^{\mathrm{H}}\right)\mathbf{\Lambda}_1\\&\quad\times\left(\sqrt{K_1}\mathbf{\Phi}_1\mathbf{R}_1^{-\frac{1}{2}}\widehat{\mathbf{g}}_1+\bar{\mathbf{g}}_1\right),\end{aligned} \tag{21}$$

where $\bar{\mathbf{g}}_1\sim\mathcal{CN}(0,\mathbf{I}_M)$. Denoting $\mathbf{z}_1=K_1^{\frac{1}{2}}\mathbf{\Phi}_1\mathbf{R}_1^{-\frac{1}{2}}\widehat{\mathbf{g}}_1$, the result in (21) can be written as

$$\begin{aligned}\mathrm{SINR}_1&\approx\frac{\rho\beta_1}{K_1+1}\left(\mathbf{z}_1^{\mathrm{H}}+\bar{\mathbf{g}}_1^{\mathrm{H}}\right)\mathbf{\Lambda}_1\left(\mathbf{z}_1+\bar{\mathbf{g}}_1\right)\\&=\frac{\rho\beta_1}{K_1+1}\sum_{j=1}^{M}\left[\mathbf{\Lambda}_1\right]_{j,j}\left|\mathbf{z}_{1,j}+\bar{\mathbf{g}}_{1,j}\right|^2,\end{aligned} \tag{22}$$

where $\mathbf{z}_{1,j}$ and $\bar{\mathbf{g}}_{1,j}$ are the j–th elements of $\mathbf{z}_1$ and $\bar{\mathbf{g}}_1$. Separating the real and imaginary components of $\mathbf{z}_1$ and $\bar{\mathbf{g}}_1$, in (22), for all $j=1,2,\ldots,M$, we have

$$\begin{aligned}\mathrm{SINR}_1\approx\frac{\rho\beta_1}{K_1+1}\sum_{j=1}^{M}\left[\mathbf{\Lambda}_1\right]_{j,j}\Big[&\left(\Re\left(\bar{\mathbf{g}}_{1,j}\right)+\Re\left(\mathbf{z}_{1,j}\right)\right)^2\\&+\left(\Im\left(\bar{\mathbf{g}}_{1,j}\right)+\Im\left(\mathbf{z}_{1,j}\right)\right)^2\Big].\end{aligned} \tag{23}$$

Further manipulations allows us to express (23) as (24), shown on top of the following page due to space reasons.

Remark 3. The expression in (24) presents a solution to an extremely complex scenario with the presence of heterogeneous spatially correlated Ricean fading. The fundamental structure of (24) contains real and imaginary components of $\bar{\mathbf{g}}_1$, a zero–mean complex Gaussian random variable, along with $\mathbf{z}_{1,j}$, a non–zero mean component as a result of dominant specular presence on each link. These are scaled by $[\mathbf{\Lambda}_1]_{j,j}$, from the decomposition of $\mathbf{R}_1^{\frac{1}{2}}\mathbf{\Theta}_1\mathbf{R}_1^{\frac{1}{2}}$. The overall form of (24) is a scaled sum of squares of M independent non–zero mean Gaussian random variables. From this, one can recognize that (24) is in the form of a linear combination of non–central chi–squared random variables. Naturally, the scaling depends on the terminal link gain, Ricean K–factor, as well as eigenvalues of a terminal's specific correlation matrix. To the best of the authors' knowledge, this is the first approximation of the instantaneous SINR of a terminal with heterogeneous fading. The result is extremely general, and can be simplified to a variety of simpler statistical channel models, as shown in Section V. Further manipulation of (24) yields

$$\mathrm{SINR}_1\approx\frac{\rho\beta_1}{2\left(K_1+1\right)}\sum_{j=1}^{2M}\widetilde{\lambda}_{1,j}\left(\mathbf{u}_{1,j}+\mathbf{p}_{1,j}\right)^2, \tag{25}$$

where $\widetilde{\lambda}_{1,2k-1}=\widetilde{\lambda}_{1,2k}=[\mathbf{\Lambda}_1]_{k,k}$, $\mathbf{p}_{1,2k-1}=\sqrt{2}\,\Re\left(\mathbf{z}_{1,k}\right)$, and $\mathbf{p}_{1,2k}=\sqrt{2}\,\Im\left(\mathbf{z}_{1,k}\right)$ with $k=1,2,\ldots,M$. Furthermore, $\mathbf{u}_{1,j}$ is a standard Gaussian random variable with zero–mean and unit variance. Via (8), the sum spectral efficiency over all L terminals can also be approximated, using the SINR in (25). In the following section, we utilize the general form of (25) to approximate the statistical moments of the MMSE SINR, and hence, sum spectral efficiency of the massive MIMO system.

IV. Statistical Moments of the MMSE SINR and Sum Spectral Efficiency

Leveraging (25), via the definition of the MGF, we can write

$$\begin{aligned}\mathrm{MGF}&=\mathbb{E}\left\{\exp\left(t\,\mathrm{SINR}_1\right)\right\}\\&=\mathbb{E}\left\{\exp\left(\frac{\rho\beta_1}{2\left(K_1+1\right)}\sum_{j=1}^{2M}\widetilde{\lambda}_{1,j}\,t\left(\mathbf{u}_{1,j}+\mathbf{p}_{1,j}\right)^2\right)\right\}\\&=\mathbb{E}\left\{\prod_{j=1}^{2M}\exp\left(\frac{\rho\beta_1}{2\left(K_1+1\right)}\widetilde{\lambda}_{1,j}\,t\left(\mathbf{u}_{1,j}+\mathbf{p}_{1,j}\right)^2\right)\right\}.\end{aligned} \tag{26}$$

Evaluating the expectation in (26) from first principles,

$$\begin{aligned}\mathrm{MGF}=\prod_{j=1}^{2M}\int_{-\infty}^{\infty}&\exp\left(\frac{\rho\beta_1}{2\left(K_1+1\right)}\widetilde{\lambda}_{1,j}\,t\left(\mathbf{u}_{1,j}+\mathbf{p}_{1,j}\right)^2\right)\\&\times\frac{1}{\sqrt{2\pi}}e^{-(\mathbf{u}_{1,j})^2/2}\,d\,\mathbf{u}_{1,j}.\end{aligned} \tag{27}$$

The integral in (27) is of the form $\int_{-\infty}^{\infty}e^{\omega t(x+y)^2-x^2/2}dx=\int_{-\infty}^{\infty}e^{\omega tx^2+2\omega tyx+\omega ty^2-x^2/2}dx$. Evaluating the integral after some routine algebra, one can express (27) as

$$\text{SINR}_1 \approx \frac{\rho\beta_1}{2\left(K_1+1\right)}\sum_{j=1}^{M}\left[\mathbf{\Lambda}_1\right]_{j,j}\left[\left(\sqrt{2}\,\Re\left(\bar{\mathbf{g}}_{1,j}\right)+\sqrt{2}\,\Re\left(\mathbf{z}_{1,j}\right)\right)^2+\left(\sqrt{2}\,\Im\left(\bar{\mathbf{g}}_{1,j}\right)+\sqrt{2}\,\Im\left(\mathbf{z}_{1,j}\right)\right)^2\right]. \tag{24}$$

$$\begin{aligned}\text{MGF}&=\prod_{j=1}^{2M}\frac{\exp\left(\omega_j t y_i^2\left(1-\frac{2\omega_j t}{1-2\omega_j t}\right)\right)}{\left(1-2\omega_j t\right)^{\frac{1}{2}}}\\&=\frac{\exp\left(\sum_{j=1}^{2M}\omega_j t y_j^2\left(1-\frac{2\omega_j t}{1-2\omega_j t}\right)\right)}{\prod_{j=1}^{2M}\left(1-2\omega_j t\right)^{\frac{1}{2}}}.\end{aligned} \tag{28}$$

Converting back to the case at hand, we have

$$\text{MGF}=\frac{\exp\left(\sum_{j=1}^{2M}\frac{\widetilde{\lambda}_{1,j}\rho\beta_1 t\,\mathbf{p}_{1,j}^2}{2(K_1+1)}\left(1-\frac{2\widetilde{\lambda}_{1,j}\rho\beta_1 t}{2(K_1+1)\left[1-\frac{2\lambda_{1,j}\rho\beta_1 t}{2(K_1+1)}\right]}\right)\right)}{\prod_{j=1}^{2M}\left(1-\frac{2\,\widetilde{\lambda}_{1,j}\rho\beta_1 t}{2(K_1+1)}\right)^{\frac{1}{2}}}. \tag{29}$$

Now, since $\widetilde{\lambda}_{j,1}=\widetilde{\lambda}_{j,2}=\lambda_{j,1}$, $\widetilde{\lambda}_{j,3}=\widetilde{\lambda}_{j,4}=\lambda_{j,2}$,..., $\widetilde{\lambda}_{j,L-1}=\widetilde{\lambda}_{j,L}=\lambda_{j,L-2}$, it follows that

$$\text{MGF}=\frac{\exp\left(\sum_{j=1}^{M}\frac{\lambda_{1,j}\rho\beta_1 t}{2(K_1+1)}\left(1-\frac{2\lambda_{1,j}\rho\beta_1 t}{2(K_1+1)-2\lambda_{1,j}\rho\beta_1 t}\right)2\left|\mathbf{z}_{1,j}\right|^2\right)}{\prod_{i=1}^{M}\left[1-\frac{2\lambda_{1,j}\rho\beta_1 t}{2(K_1+1)}\right]}. \tag{30}$$

Differentiating (30) once and setting $t=0$ yields $\mathbb{E}\{\text{SINR}_1\}$, while differentiating twice and setting $t=0$ results in $\mathbb{E}\{\text{SINR}_1^2\}$. This allows us to state

$$\begin{aligned}\mathbb{E}\left\{\text{SINR}_1\right\}&\approx\frac{\rho\beta_1}{2\left(K_1+1\right)}\sum_{j=1}^{2M}\widetilde{\lambda}_{1,j}\left(1+\mathbf{p}_{1,j}^2\right)\\&=\frac{\rho\beta_1}{\left(K_1+1\right)}\sum_{j=1}^{M}\lambda_{1,j}\left(1+\left|\mathbf{z}_{1,j}\right|^2\right),\end{aligned} \tag{31}$$

and

$$\text{Var}\left\{\text{SINR}_1\right\}\approx\frac{\rho^2\beta_1^2}{\left(K_1+1\right)^2}\sum_{j=1}^{M}\lambda_{1,j}^2\left(1+2\left|\mathbf{z}_{1,j}\right|^2\right). \tag{32}$$

Remark 4. By inspection, one can observe that the results in (31) and (32) are independent of the diffuse MPCs, as these form the ensemble over which the statistical expectations are performed, in order to obtain $\mathbb{E}\{\text{SINR}_1\}$ and $\text{Var}\{\text{SINR}_1\}$. With a fixed M, ρ, β_1, $\lambda_{1,j}$, and $\mathbf{z}_{1,j}$, increasing K_1 yields a loss in the expected SINR proportional to $1/K_1$, and reduces the variability in the SINR of terminal 1 proportional to $1/K_1^2$. This is a result of the propagation channel becoming increasingly deterministic, leading to a loss in spatial selectivity. As will be shown in Section VI, both (31) and (32) remain extremely tight with changes in M, ρ, λ_1, K_1 and $\mathbf{z}_1$, respectively. Though not shown here due to space constraints, (31) and (32) are central to the analysis of other useful performance metrics, such as symbol error rates, as well as capacity variance of spatially correlated Ricean channels. Moreover, we note that the mean and variance results can be easily simplified for simpler scenarios, such as equally correlated Ricean fading, uncorrelated Ricean fading, unequal and equally correlated Rayleigh fading, as well as uncorrelated Rayleigh fading by modifying the constants preceding the summation, $\lambda_{1,j}$ and in $\mathbf{z}_{1,j}$. Finally, we note that the expected SINR result can naturally be translated to approximate the ergodic sum spectral efficiency by application of $\text{SE}_{\text{sum}}^{\text{ergodic}}\approx\sum_{\ell=1}^{L}\log_2(1+\mathbb{E}\{\text{SINR}_\ell\})$, while a similar approach can be taken to compute the variance of the sum spectral efficiency. This is not shown in the interest of space.

V. Special Cases

In the sequel, we present several special cases of the instantaneous MMSE SINR demonstrating its generality.

1) *Equally Correlated Ricean Fading:* The expression in (25) remains as it is, with the exception that $\mathbf{R}_1=\mathbf{R}_i$, for $i=2,3,\ldots,L$ in $\mathbf{\Theta}_1$ defined just after (18). From this, the structure of $\widetilde{\lambda}_{1,j}$ is modified accordingly.
2) *Uncorrelated Ricean Fading:* Similar to above, in this case, (25) has the same form with the exception that $\mathbf{R}_1=\mathbf{R}_i=\mathbf{I}_M$, for $i=2,3,\ldots,L$ in $\mathbf{\Theta}_1$. As a consequence, the structure of $\widetilde{\lambda}_{1,j}$ is modified.
3) *Unequally Correlated Rayleigh Fading:* With the absence of dominant specular components from each terminal, SINR_1 collapses to
$$\text{SINR}_1\approx\frac{\rho\beta_1}{2}\sum_{j=1}^{2M}\widetilde{\lambda}_{1,j}\,\mathbf{u}_{1,j}^2, \tag{33}$$
where $\widetilde{\lambda}_{1,j}$ and $\mathbf{u}_{1,j}$ are as defined after (25). However, to obtain $\widetilde{\lambda}_{1,j}$, $\mathbf{\Theta}_1=\rho\mathbf{I}_M-\rho^2(\mathbf{S}_1+\sum_{i=2}^{L}[\mathbf{S}_1]_{i,i}\mathbf{R}_i)$, where $\mathbf{S}_1=(\mathbf{D}_1^{-1}+\rho\,\mathbf{I}_L)^{-1}$, due to the absence of specular components from the propagation channel.
4) *Equally Correlated Rayleigh Fading:* The expression in (33) still holds for the case where each terminal sees pure NLoS fading, with the exception that $\mathbf{R}_1=\mathbf{R}_i$, for $i=2,3,\ldots,L$. This is used to compute $\widetilde{\lambda}_{1,j}$.
5) *Uncorrelated Rayleigh Fading:* The MMSE SINR of terminal 1 under uncorrelated Rayleigh fading has the same form as (33), where $\mathbf{R}_1=\mathbf{R}_i=\mathbf{I}_M$ for $i=2,3,\ldots,L$, allowing us to recompute $\mathbf{\Theta}_1$ and $\widetilde{\lambda}_{1,j}$.

VI. Numerical Results

Unless otherwise specified, the parameters described below are utilized for all subsequent numerical results, and are obtained from [7]. A cell radius of $R_{\text{c}}=100$ m is chosen with a reference distance $r_0=10$ m, such that the terminals are randomly located outside r_0, and inside R_{c}, following a uniform distribution in area on $[0,2\pi]$. The LoS and NLoS attenuation exponents, α, are given by 2.2 and 3.67, respectively. Moreover, the unit–less constant for geometric attenuation, A, is chosen such that the tenth–percentile of the instantaneous SINR of terminal 1 with MMSE processing is 0 dB, with $\rho=0$ dB with $M=32$, and $L=5$. Note that the exponential correlation model (described further in the section) with a correlation coefficient $\xi=0.9$ was chosen with a fixed $K=5$ dB for all terminals to obtain A. The LoS and NLoS shadow–fading standard deviations, σ_{sh}, are 4 and 6 dB. The terminal specific K–factor has a log–normal distribution with a mean of 9 dB and a variance of 5 dB, or a mean of 12 dB and variance of 3 dB. These reflect the

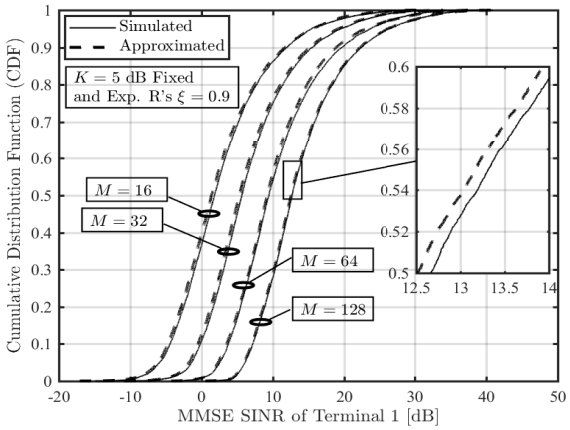

Fig. 1. CDFs of the MMSE SINR for terminal 1 with $M = 16, 32, 64, 128$ and $L = 5$ with $\rho = 0$ dB.

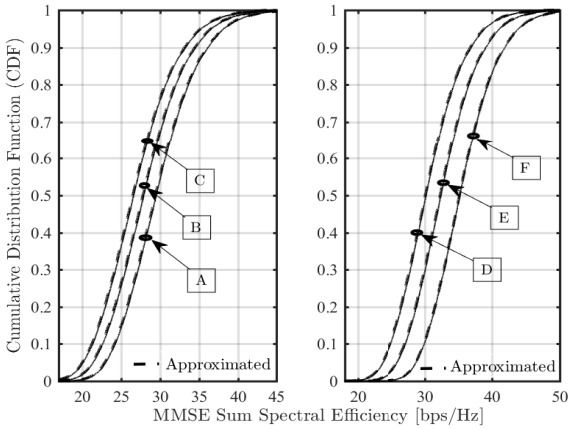

Fig. 2. Sum spectral efficiency CDFs with $M = 128$ and $L = 5$ and $\rho = 0$ dB. Note that [A]: $K \sim \ln(9, 5)$ dB and Exp. R's $\xi = 0.9$, [B]: $K = 9$ dB Fixed and Exp. R's $\xi = 0.9$, [C]: $K \sim \ln(12, 3)$ dB and Exp. R's $\xi = 0.9$, [D]: $K = 9$ dB Fixed and O.R. R's $\mathcal{N}(35°, 15°)$, [E]: $K \sim \ln(9, 5)$ dB and O.R. R's $\mathcal{N}(35°, 15°)$, [F]: Uncorrelated Rayleigh Fading.

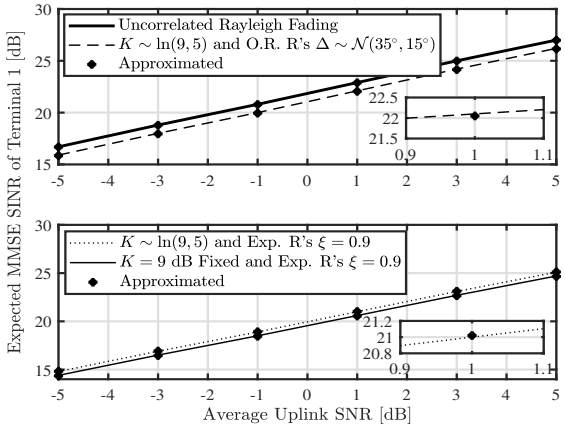

Fig. 3. Expected MMSE SINR vs. average SNR (ρ) when $M = 128$, $L = 5$.

measured K–factor characteristics at 2 GHz [7] and 28 GHz [8], and are denoted by $K \sim \ln(9, 5)$ dB and $K \sim \ln(12, 3)$ dB. The probability of terminal 1 experiencing LoS is given by $\mathbb{P}_{\text{LoS}}(r_1) = (\min(18/r_1, 1)(1-e^{-r_1/36}))+e^{-r_1/36}$. Naturally, $\mathbb{P}_{\text{NLoS}}(r_1) = 1 - \mathbb{P}_{\text{LoS},1}(r_1)$.

For each subsequent result, 10^6 Monte–Carlo realizations were generated with an inter–element spacing, $d = 0.5\lambda$ at the BS. We model fixed correlation matrices at each terminal with the widely used exponential (Exp.) model [12], where the (i, j)–th element of $\mathbf{R}_1$ is modeled as $[\mathbf{R}_1]_{i,j} = \xi^{|i-j|}$, for any $i, j = 1, 2, \ldots, M$, with $0 \leq \xi \leq 1$. Unless otherwise specified, $\rho = 0.9$ is used throughout the evaluation. With variable correlation, we employ the one–ring (O.R.) correlation model [15, 17], where $[\mathbf{R}_1]_{i,j} = \frac{1}{2\Delta_1} \int_{-\Delta_1+\phi_0^1}^{\Delta_1+\phi_0^1} e^{-j2\pi d(i,j)\sin(\phi_1)} d\phi_1$, where Δ_1 denotes the azimuth angular spread, specific to terminal 1, ϕ_0^1 denotes the mean direction–of–arrival (DoA), while ϕ_1 is the actual LoS direction, uniformly distributed within the angular spread around the mean DoA. In order to model variable correlation matrices, we consider $\Delta_1 \sim \mathcal{N}(35°, 15°)$. We further note that $d(i, j)$ captures the normalized antenna spacing between the i–th and j–th elements.

Figure 1 depicts the cumulative distribution functions (CDFs) of the MMSE SINR with $M = 16, 32, 64, 128$ and $L = 5$ at $\rho = 0$ dB. A fixed K–factor of 5 dB is assumed for all terminals with equal correlation matrices from the exponential model. Two trends can be observed: Firstly, each increment of M results in a median (CDF $= 0.5$) SINR increase of 3 dB. This suggests that the MMSE SINR is effectively doubled if one fixes all other system parameters, and increases M exponentially (in powers of 2). This is due to the additional degrees of freedom, which is able to better condition the inverse in the MMSE SINR. Secondly, the proposed approximations remain accurate across the entire range of M values under consideration. We note that this is despite the approximation methodology relying on having a large M. For this reason, it is interesting to observe that as M is increased, the approximated CDFs get increasingly tighter. Overall, this result is able to determine the accuracy of the SINR approximation for both small and large value of M, as well as when the massive MIMO system has homogeneous propagation with fixed K–factors and correlation matrices.

Fig. 2 demonstrates the sum spectral efficiency CDFs with MMSE combining. Both homogeneous and heterogeneous scenarios are considered, with equal and unequal K–factors and spatial correlation structures. Inspecting the left–hand subfigure, one can observe that $K \sim \ln(12, 3)$ dB (case C) results in lower sum spectral efficiencies than a fixed K–factor of 9 dB (case B) and when $K \sim \ln(9, 5)$ dB (case A). This is due to the loss in the spatial selectivity of the propagation channel, reducing the composite channel rank and consequently reducing the SINR of each terminal and system sum spectral efficiency. The right–hand subfigure shows that the sum spectral efficiency increases with unequal K–factors and spatial correlation matrices. Fundamentally, variable correlation induces a random angular spread, controlling the amount of spatial selectivity in the channel, while variable K–factors also bring instances of lower K values, contributing to stronger presence of diffuse MPCs and enhancing the performance. This can be observed when comparing cases D and E, while case F represents that of uncorrelated Rayleigh fading. Again, the developed approximations remain tight and cater to both homogeneous and heterogeneous scenarios.

Figure 3 shows the expected SINR of a terminal as a function of the average uplink SNR. One can observe the same trends as seen for the sum spectral efficiency in Fig.

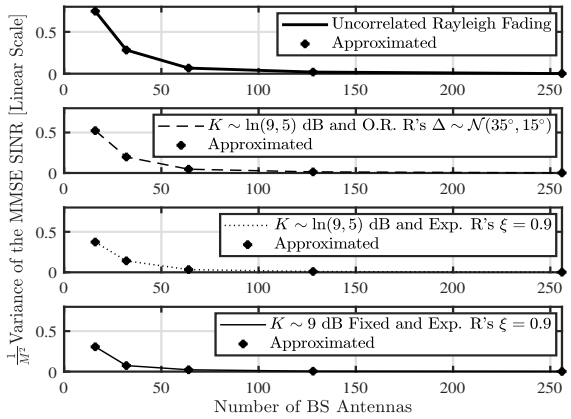

Fig. 4. Variance of the MMSE SINR vs. the number of BS antennas (M) for $L = 5$ and $\rho = 0$ dB.

2, where the expected SINR increases with increasing variability in K–factors and spatial correlation matrices, due to the stronger presence of diffuse MPCs relative to the LoS levels. The derived expected value approximation via the MGF yields extremely tight results against the numerically simulated responses. This result demonstrates that not only can our analysis methodology approximate the instantaneous SINRs, it is able to also approximate the mean SINR. Further to this, Fig. 4 depicts the scaled SINR variance as a function of the number of BS antennas. Two trends can be seen: Firstly, when complete NLoS fading is present, the SINR variability is the largest, as seen in the top–most subfigure. As the propagation channel becomes more deterministic, the SINR variability reduces significantly. The SINR resulting from $K \sim \ln(9, 5)$ dB with O.R. correlation exhibits more variability than their counterparts with fixed correlation matrices, due to the enhanced variability of diffuse MPCs. The case with fixed K–factors and correlation matrices naturally result in the lowest SINR variability. The aforementioned discussion is analytically predicted in Remark 4. Secondly, the simple variance approximation very accurately predicts the simulated performance over a wide range of scenarios.

VII. Conclusions

The paper presents a general analytical framework for approximating the instantaneous SINR of a terminal, as well as, the system sum spectral efficiency of an uplink massive MIMO system, with MMSE combining. We show that the MMSE SINR is well approximated by a linear combination of non–central chi–squared random variables. Unlike many studies, our methodology is able to handle full heterogeneity with terminal specific K–factors, spatial correlation matrices, and link gains. The generality of the SINR and sum spectral efficiency is further exploited to develop an analytical method for computing its statistical moments. Several insightful special cases are presented, which encapsulate simpler homogeneous and heterogeneous scenarios. All proposed approximations are seen to remain tight across considered spatial correlation models, K–factor models, average uplink SNRs and the number of BS antennas. Such a general treatment of the analysis and evaluation of MMSE combining has been missing from the vast massive MIMO literature.

VIII. Acknowledgment

The work of H. Tataria and M. Matthaiou was supported by the EPSRC, UK, under grant EP/P000673/1. The work of P. J. Smith was supported by the Royal Academy of Engineering, UK, via the Visiting Fellowship Program DVF1617/6/29.

References

[1] M. Shafi, *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201-1221, Jun. 2017.

[2] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899-3911, Jul. 2015.

[3] 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, Apr. 2013.

[4] E. G. Larsson, T. L. Marzetta, H. Q. Ngo, and H. Yang, "Antenna count for massive MIMO: 1.9 GHz versus 60 GHz," Available online, https://arxiv.org/abs/1702.06111, Feb. 2017.

[5] J. Ko, *et al.*, "Millimeter-wave channel measurements and analysis for statistical spatial channel model in in-building and urban environments at 28 GHz," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5853-5868, Sep. 2017.

[6] T. S. Rappaport, S. Sun, and M. Shafi, "Investigation and comparison of 3GPP and NYUSIM channel models for 5G wireless communications," in *Proc. IEEE VTC-Fall*, Sep. 2017.

[7] 3GPP TR 36.873 v.12.2.0, *Study on 3D channel models for LTE*, 3GPP, Jun. 2015.

[8] 3GPP TR 38.901, v.14.2.0, *Study on channel model for frequencies from 0.5 to 100 GHz*, 3GPP, Sep. 2017.

[9] H. Tataria, *et al.*, "Impact of line-of-sight and unequal spatial correlation on uplink MU-MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 634-637, Oct. 2017.

[10] Q. Zhang, *et al.*, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966-981, Oct. 2014.

[11] J. Zhang, *et al.*, "On the spectral efficiency of space-constrained massive MIMO with linear receivers," in *Proc. of IEEE ICC*, May 2016.

[12] H. Falconet, L. Sanguinetti, A. Kammoun, and M. Debbah, "Asymptotic analysis of downlink MISO systems over Rician fading channels," in *Proc. IEEE ICASSP*, May 2016, pp. 3926-3930.

[13] S. Jin, *et al.*, "Statistical eigenmode transmission for the MU-MIMO downlink in Rician fading," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6650-6663, Dec. 2015.

[14] H. Tataria, P. J. Smith, L. J. Greenstein, and P. A. Dmochowski, "Zero-forcing precoding performance in multiuser MIMO systems with heterogeneous Ricean fading," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 74-77, Feb. 2017.

[15] J. Nam, G. Caire, and J. Ha, "On the role of transmit correlation diversity in multiuser MIMO systems," *IEEE Trans. Inf. Thoery*, vol. 63, no. 1, pp. 336-354, Jan. 2017.

[16] E. Björnson, J. Hoydis, and L. Sanguinetti, "Pilot contamination is not a fundamental asymptotic limitation in massive MIMO," in *Proc. IEEE ICC*, May 2017.

[17] A. Adhikary, *et al.*, "Joint spatial division and multiplexing for mm-Wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239-1255, Jun. 2014.

[18] X. Li, *et al.*, "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1-15, Jun. 2017.

[19] B. Wang, Y. Chang, and D. Yang, "On the SINR in massive MIMO networks with MMSE receivers," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 1979-1982, Nov. 2014.

[20] M. R. McKay, I. B. Collings, and A. M. Tulino, "Achievable sum rate of MIMO MMSE receivers: A general analytic framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 396-410, Jan. 2010.

[21] A. L. Moustakas, "SINR distribution of MIMO MMSE receiver," *in Proc. IEEE ISIT*, Aug. 2011, pp. 938-942.

[22] A. H. Mehana and A. Nosratinia, "Diversity of MMSE MIMO receivers," *IEEE Trans. Inf. Thoery*, vol. 58, no. 11, pp. 6788-6805, Nov. 2012.

[23] R. H. Y. Louie, M. R. McKay, and I. B. Collings, "New performance results for multiuser optimum combining in the presence of Rician fading," *IEEE Trans. Commun.*, vol. 57, no. 8, pp. 2348-2358, Aug. 2009.