

3D Massive MIMO Systems: Modeling and Performance Analysis

Journal:	<i>IEEE Transactions on Wireless Communications</i>
Manuscript ID:	Paper-TW-Jan-15-0011
Manuscript Type:	Original Transactions Paper
Date Submitted by the Author:	04-Jan-2015
Complete List of Authors:	Nadeem, Qurrat-UI-Ain; King Abdullah University of Science and Technology (KAUST), CEMSE Kammoun, Abla; KAUST, Debbah, Merouane; Supelec, Telecommunications department Alouini, Mohamed; KAUST, EE Program
Keyword:	

3D Massive MIMO Systems: Modeling and Performance Analysis

Qurrat-Ul-Ain Nadeem, Abla Kammoun, *Member, IEEE*,

Mérouane Debbah, *Fellow, IEEE*, and Mohamed-Slim Alouini, *Fellow, IEEE*

Abstract

Multiple-input-multiple-output (MIMO) systems of current LTE releases are capable of adaptation in the azimuth only. Recently, the trend is to enhance system performance by exploiting the channel's degrees of freedom in the elevation, which necessitates the characterization of 3D channels. We present an information-theoretic channel model for MIMO systems that supports the elevation dimension. The model is based on the principle of maximum entropy, which enables us to determine the distribution of the channel matrix consistent with the prior information on the angles. Based on this model, we provide analytical expression for the cumulative density function (CDF) of the mutual information (MI) for systems with a single receive and finite number of transmit antennas in the general signal-to-interference-plus-noise-ratio (SINR) regime. The result is extended to systems with finite receive antennas in the low SINR regime. A Gaussian approximation to the asymptotic behavior of MI distribution is derived for the large number of transmit antennas and paths regime. We corroborate our analysis with simulations that study the performance gains realizable through meticulous selection of the transmit antenna downtilt angles, confirming the potential of elevation beamforming to enhance system performance. The results are directly applicable to the analysis of 5G 3D-Massive MIMO-systems.

Index Terms

Manuscript received January 04, 2015.

Q.-U.-A. Nadeem, A. Kammoun and M.-S. Alouini are with the Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, Building 1, Level 3, King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia 23955-6900 (e-mail: {qurratulain.nadeem,abla.kammoun,slim.alouini}@kaust.edu.sa)

M. Debbah is with Mathematical and Algorithmic Sciences Lab, Huawei France R&D, Paris, France (e-mail: merouane.debbah@huawei.com).

1
2
3
4 Massive Multiple-input multiple-output (MIMO) systems, channel modeling, maximum entropy,
5 elevation beamforming, mutual information, antenna downtilt.
6
7

8 9 I. INTRODUCTION

10
11 Multiple-input multiple-output (MIMO) systems have remained a subject of interest in wireless
12 communications over the past decade due to the significant gains they offer in terms of the
13 spectral efficiency by exploiting the multipath richness of the channel. The increased spatial
14 degrees of freedom not only provide diversity and interference cancellation gains but also
15 help achieve a significant multiplexing gain by opening several parallel sub-channels. Pioneer
16 work in this area has shown that under certain channel conditions, MIMO system capacity
17 can potentially scale linearly with the minimum number of transmit (Tx) and receive (Rx)
18 antennas [1], [2]. This has led to the emergence of Massive MIMO systems, that scale up the
19 conventional MIMO systems by possibly orders of magnitude compared to the current state-of-art
20 [3]. These systems were initially designed to support antenna configurations at the base station
21 (BS) that are capable of adaptation in the azimuth only. However, the future generation of mobile
22 communication standards such as 3GPP LTE-Advanced are targeted to support even higher data
23 rate transmissions, which has stirred a growing interest among researchers to enhance system
24 performance through the use of adaptive electronic beam control over both the elevation and the
25 azimuth dimensions. Several measurement campaigns have already demonstrated that elevation
26 has a significant impact on the system performance [4]–[7]. The reason can be attributed to its
27 potential to allow for a variety of strategies like user specific elevation beamforming and three-
28 dimensional (3D) cell-splitting. This novel approach towards enhancing the system performance
29 uncovers entirely new problems that need urgent attention: deriving accurate 3D channel models
30 that support the elevation dimension, designing beamforming strategies that exploit the extra
31 degrees of freedom, performance prediction and analysis to account for the impact of the channel
32 component in the elevation and finding appropriate deployment scenarios. In fact, 3GPP has
33 started working on defining future mobile communication standards that would help evaluate the
34 potential of 3D beamforming in the future [8].
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 A good starting point for the effective performance analysis of MIMO systems is the evalua-
54 tion and characterization of their information-theoretic mutual information (MI). The statistical
55 distribution of MI is not only useful in obtaining performance measures that are decisive on
56
57
58
59
60

1
2
3
4 the use of several emerging technologies, but also provides a comprehensive picture of the
5 overall MIMO channel. This has motivated many studies on the statistical distribution of the MI
6 in the past decade. The analysis for 2D MIMO channels in different fading environments has
7 appeared frequently in literature [9]–[16]. Most of the work in this area assumes the channel
8 gain matrix to have independent and identically distributed (i.i.d) complex Gaussian entries and
9 resorts to the results developed for Wishart matrices. However, because of the complexity of the
10 distribution of a Wishart matrix, the study on the statistical distribution of MI becomes quite
11 involved. The characteristic function (CF)/moment generating function (MGF) of the MI has
12 been derived in [16]–[18] but the MGF approaches to obtain the statistical distribution of MI
13 involve the inverse Laplace transform which leads to numerical integration methods. However
14 in the large number of antennas regime, Random Matrix Theory (RMT) provides some simple
15 deterministic approximations to the MI distribution. These deterministic equivalents were shown
16 to be quite accurate even for a moderate number of antennas. The authors in [14] showed
17 that the MI can be well approximated by a Gaussian distribution and derived its mean and
18 variance for spatially correlated Rician MIMO channels. More recently, Hachem *et al* derived
19 a deterministic equivalent for the MI of Kronecker channel model and rigorously proved that
20 the MI converges to a standard Gaussian random variable in the asymptotic limit in [19]. The
21 Central Limit Theorem (CLT) was established with appropriate scaling and centering and the
22 convergence result is almost sure. More asymptotic results on the statistical distribution of MI
23 have been provided in [14]–[16], [18] that show the accuracy of the Gaussian approximation to
24 the distribution of the MI of 2D channels. However, the extension of these results to the 3D
25 case has not appeared yet.

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43 In order to allow for the efficient design and better understanding of the limits of wireless
44 systems through the meticulous analysis of the MI requires accurate modeling and characteri-
45 zation of the 3D channels. Given the available knowledge on certain channel parameters like
46 angle of departure (AoD), angle of arrival (AoA), delay, amplitude, number of Tx and Rx
47 antennas, finding the best way to model the channel consistent with the information available
48 and to attribute a joint probability distribution to the entries of the channel matrix is of vital
49 importance. It has been shown in literature that the choice of distribution with the greatest entropy
50 creates a model out of the available information without making any arbitrary assumption on
51 the information that was not available [20], [21]. The authors in [22] showed the principle of
52
53
54
55
56
57
58
59
60

1
2
3
4 maximum entropy to be an appropriate method of scientific inference for translating information
5 into probability assignment, as maximizing any other function might not satisfy the specified
6 consistency arguments and could lead to implicit assumptions that do not reflect upon our true
7 state of knowledge of the problem. In the context of wireless communications, the authors in [23]
8 addressed the question of MIMO channel modeling from an information-theoretic point of view
9 and provided guidelines to determine the distribution of the 2D MIMO channel matrix through
10 an extensive use of the principle of maximum entropy together with the consistency argument.
11 Several channel models were derived consistent with the apriori knowledge on parameters like
12 AoA, AoD, number of scatterers. These results can be extended to the 3D MIMO channel model,
13 which will be one of the aims of this work. As mentioned earlier, almost all existing system
14 level based standards are 2D. However, owing to the growing interest in 3D beamforming,
15 extensions of these channel model standards to the 3D case have started to emerge recently that
16 capture the 3D propagation effects of the environment as well as the spatial structure of the radio
17 channel [24], [8]. An important feature of these models is the vertical downtilt angle, that can
18 be optimized to change the vertical beam pattern dynamically and yield the promised MI gains
19 [25]–[28]. Although studies on 3D channel modeling and impact of antenna downtilt angle on
20 the achievable rates have started to appear in literature, but to the best of authors' knowledge,
21 there has not been any comprehensive study to develop analytical expressions for the MI of these
22 channels, that could help evaluate the performance of the future 3D MIMO systems. Deriving
23 these expressions for the 3D entropy maximizing channel model is another aim of this work.

24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
The aims of this paper are twofold. First is to provide an information-theoretic channel model
for 3D Massive MIMO systems and second is to predict and analyze the performance of these
systems by characterizing the distribution of the MI. We follow the guidelines provided in [23]
to present the entropy maximizing 3D channel model that is consistent with the state of available
knowledge of the channel parameters. The 3D channel used for the MI analysis is inspired from
the models presented in standards like 3GPP SCM [29], WINNER+ [24] and ITU [30]. These
geometry-based standardized MIMO channel models assume single-bounce scattering between
the transmitter and receiver, which allows us to assume the AoDs and AoAs to be fixed and
known apriori. The motivation behind this approach lies in the observation that when a single
bounce on a scatterer occurs, the directions of arrival and departure are deterministically related
by Descartes's laws, and the channel statistics pertaining to AoD and AoA can be assumed

not to change during the modeling phase. The maximum entropy 3D channel model consistent with the state of knowledge of channel statistics pertaining to AoDs and AoAs, turns out to have a systematic structure with a reduced degree of randomness in the channel matrix. This calls for the use of an approach that does not depend on the more generally employed results on the distribution of Wishart matrices for the characterization of MI. However, the systematic structure of the model can be exploited to our aid in characterizing its MI. With this channel model at hand, we derive analytical expressions for the cumulative density function (CDF) of the MI in different signal-to-interference plus noise ratio (SINR) regimes. An exact closed-form expression is obtained for systems with a single Rx and multiple Tx antennas in the general SINR regime. The result is extended to systems with a finite number of Rx antennas in the low SINR regime. We also present an asymptotic analysis of the statistical distribution of the MI and show that it is well approximated by a Gaussian distribution for any number of Rx antennas, as the number of paths and Tx antennas grow large. Finally numerical results are provided that illustrate an excellent fit between the Monte-Carlo simulated CDF and the derived closed-form expressions. The simulation results provide a flavor of the performance gains realizable through the meticulous selection of the downtilt angles. We think our work contributes to cover new aspects that have not yet been investigated by a research publication and addresses some of the current challenges faced by researchers and industrials to fairly evaluate 3D beamforming techniques on the basis of achievable transmission rates. The results have immediate applications in the design of 3D 5G Massive MIMO systems.

The rest of the paper is organized as follows. In section II, we introduce the 3D channel model based on the 3GPP standards and information-theoretic considerations. In Section III, we provide analytical expressions for the CDF of the MI in the general and low SINR regimes for a finite number of Tx antennas. In section IV, we provide an asymptotic analysis that holds for any number of Rx antennas. The derived results are corroborated using simulations in section V and finally in section VI, some concluding remarks are drawn.

II. SYSTEM AND CHANNEL MODEL

Encouraged by the potential of elevation beamforming to enhance system performance, many standardized channel models have emerged that define the next generation 3D channels. We base the evaluation of our work on these models while making some realistic assumptions on

the channel parameters.

A. System Model

We consider a downlink 3D MIMO system, where the BS is equipped with N_{BS} directional Tx antenna ports and the mobile station (MS) has N_{MS} Rx antenna ports. Denoting Δh as the height difference between the MS and the BS, and defining x and y as the relative distance between the MS and the BS in the x and y coordinates respectively, the elevation line of sight (LoS) angle with respect to the horizontal plane at the BS can be written as $\theta_{LoS} = \tan^{-1} \frac{\Delta h}{\sqrt{x^2 + y^2}}$. In addition, θ_{tilt} represents the elevation angle of the antenna boresight. The antenna configuration we consider is shown in Fig. 1. Each antenna port comprises of M vertically stacked antenna elements that determine the effective antenna port pattern. The antenna ports are placed at fixed positions along \hat{e}_y , with the elements in each port aligned along \hat{e}_z . The same Tx signal is fed to all elements with corresponding weights w_k 's, $k = 1, \dots, M$, in order to achieve the desired directivity. The MS sees each antenna port as a single antenna because all the elements carry the same signal.

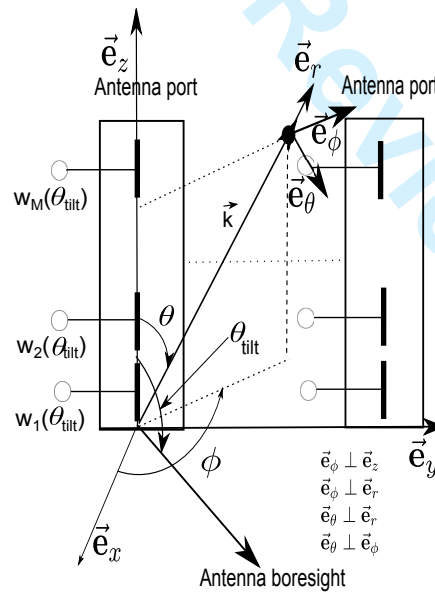


Fig. 1. Antenna configuration.

B. 3D MIMO Channel Model

The mobile communication standards like 3GPP SCM [29], ITU [30] and WINNER [31] follow a system level, stochastic channel modeling approach wherein, the propagation paths are described through statistical parameters, like delay, amplitude, AoA and AoD. The extensions of these 2D models to the 3D case have started to emerge recently [24], [32]. These geometry-based stochastic models generate channel realizations by summing the contributions of N multipaths (plane waves) with specific small-scale parameters like delay, amplitude, AoA and AoD in both azimuth and elevation directions.

Based on the aforementioned standards and for the antenna configuration in Fig. 1, the effective double directional radio channel between BS antenna port s and MS antenna port u is given by,

$$[\mathbf{H}]_{su} = \frac{1}{N} \sum_{n=1}^N \alpha_n \sqrt{g_t(\phi_n, \theta_n, \theta_{tilt})} \exp(ik(s-1)d_t \sin \phi_n \sin \theta_n) \sqrt{g_r(\varphi_n, \vartheta_n)} \exp(ik(u-1)d_r \sin \varphi_n \sin \vartheta_n), \quad (1)$$

where ϕ_n and θ_n are the azimuth and elevation AoD of the n^{th} path respectively, φ_n and ϑ_n are the azimuth and elevation AoA of the n^{th} path respectively and θ_{tilt} is the elevation angle of the antenna boresight, with $\theta_{tilt} = 90^\circ$ corresponding to zero electrical downtilt. α_n is the complex random amplitude of the n^{th} path. Also $\sqrt{g_t(\phi_n, \theta_n, \theta_{tilt})}$ and $\sqrt{g_r(\varphi_n, \vartheta_n)}$ are the global patterns of Tx and Rx antennas respectively. d_t and d_r are the separations between Tx antenna ports and Rx antenna ports respectively and k is the wave number that equals $\frac{2\pi}{\lambda}$, where λ is the wavelength of the carrier frequency. The entries [33],

$$[\mathbf{a}_t(\phi, \theta)]_s = \exp(ik(s-1)d_t \sin \phi \sin \theta), \quad (2)$$

$$[\mathbf{a}_r(\varphi, \vartheta)]_u = \exp(ik(u-1)d_r \sin \varphi \sin \vartheta) \quad (3)$$

are the array responses of s^{th} Tx and u^{th} Rx antenna respectively. Fig. 2 shows the 3D channel model being considered.

This system level approach allows for paths to be described through statistical parameters. For our work, we use the distributions given in the standards to generate the random correlated variables that describe the paths. In the standards, AoDs and AoAs in the elevation (θ_n, ϑ_n) are drawn from the Laplacian elevation density spectrum [24], [34], while those in azimuth (ϕ_n, φ_n)

are drawn from the Wrapped Gaussian azimuth density spectrum [30], [35]. Once generated, the angles are assumed to be fixed over channel realizations as justified in the next subsection.

To enable an abstraction of the role played by antenna elements, standards like 3GPP, ITU approximate the global pattern of each port by a narrow beam as [30],

$$\sqrt{g_t(\phi_n, \theta_n, \theta_{tilt})} = \sqrt{[17\text{dBi} - \min\{-(A_H(\phi) + A_V(\theta, \theta_{tilt})), 20\text{dB}\}]_{\text{lin}}}, \quad (4)$$

where,

$$A_H(\phi) = -\min \left[12 \left(\frac{\phi}{\phi_{3dB}} \right)^2, 20 \right] \text{ dB}, \quad (5)$$

$$A_V(\theta, \theta_{tilt}) = -\min \left[12 \left(\frac{\theta - \theta_{tilt}}{\theta_{3dB}} \right)^2, 20 \right] \text{ dB}. \quad (6)$$

where ϕ_{3dB} is the horizontal 3 dB beamwidth and θ_{3dB} is the vertical 3 dB beamwidth. The individual antenna radiation pattern at the MS $g_r(\varphi_n, \vartheta_n)$ is taken to be 0 dB.

C. Maximum Entropy Channel

It was rigorously proved in [22] that the principle of maximum entropy yields models that express the constraints of our knowledge of model parameters and avoid making any arbitrary assumptions on the information that is not available. Maximizing any quantity other than entropy would result in inconsistencies, and the derived model will not truly reflect our state of knowledge. Motivated by the success of the principle of maximum entropy in parameter estimation and Bayesian spectrum analysis modeling problems, the authors in [23] utilized this framework

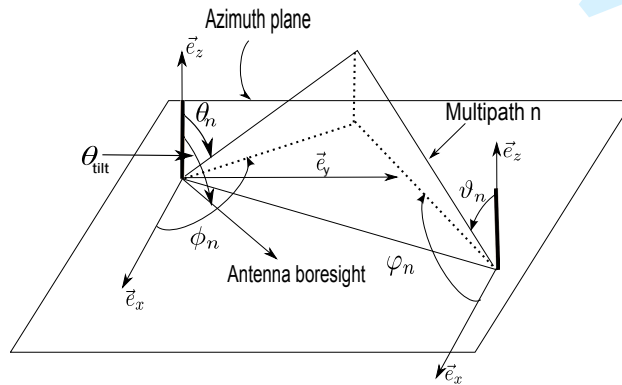


Fig. 2. 3D channel model.

to devise theoretical grounds for constructing channel models for 2D MIMO systems, consistent with the state of available knowledge of channel parameters.

The geometry-based MIMO channel models presented in standards assume single-bounce scattering between the transmitter and receiver, i.e., there is a one-to-one mapping between the AoDs and AoAs, allowing several bounces to occur as long as each AoA is linked to only one AoD [36]. Since the model presented in the last section assumes single-bounce scattering, so $\phi_n, \theta_n, \varphi_n, \psi_n$ are assumed to be known apriori and fixed over channel realizations. The motivation behind this lies in the observation that for single-bounce scattering channel models, the directions of arrival and departure are deterministically related by Descartes's laws, and the channel statistics pertaining to AoD and AoA can be assumed not to change during the modeling phase. Moreover the 3D channel model described by (1) is non-linear in the azimuth and elevation AoD and AoA so the apriori knowledge of these parameters implies that we can condition on the random variables (RV) that exhibit this non-linear dependence with the model. The only random component in the 3D channel model is now α_n , so a suitable distribution needs to be assigned to it such that the obtained model is consistent with the information already available.

It was proved through an extensive analysis in [23] that in the presence of the prior information on the number of propagation paths, AoDs, AoAs and transmitted and received power of the propagation paths, the maximum entropy channel model is given by,

$$\mathbf{H} = \mathbf{\Psi} \mathbf{P}_{Rx}^{\frac{1}{2}} (\mathbf{\Omega} \circ \mathbf{G}) \mathbf{P}_{Tx}^{\frac{1}{2}} \mathbf{\Phi}^H, \quad (7)$$

where \mathbf{P}_{Rx} and \mathbf{P}_{Tx} contain the corresponding transmitted and received powers of the multipaths, $\mathbf{\Omega}$ is the mask matrix that captures the path gains, $\mathbf{\Psi}$ and $\mathbf{\Phi}$ capture the antenna array responses and \circ is the Hadamard product. The solution of the consistency argument that maximizes the entropy is to take \mathbf{G} to be i.i.d zero mean Gaussian with unit variance [23], [36].

To represent the channel model in (1) with a structure similar to that of (7), we define \mathbf{A} and \mathbf{B} as $N_{BS} \times N$ and $N_{MS} \times N$ deterministic matrices given by,

$$\mathbf{A} = \frac{1}{\sqrt{N}} [\mathbf{a}_t(\phi_1, \theta_1), \mathbf{a}_t(\phi_2, \theta_2), \dots, \mathbf{a}_t(\phi_N, \theta_N)] \circ \begin{bmatrix} \sqrt{g(\phi_1, \theta_1, \theta_{tilt_1})} & \dots & \sqrt{g(\phi_N, \theta_N, \theta_{tilt_1})} \\ \vdots & \ddots & \vdots \\ \sqrt{g(\phi_1, \theta_1, \theta_{tilt_{N_{BS}}})} & \dots & \sqrt{g(\phi_N, \theta_N, \theta_{tilt_{N_{BS}}})} \end{bmatrix}, \quad (8)$$

where $\mathbf{a}_t(\phi_n, \theta_n) = [[\mathbf{a}_t(\phi_n, \theta_n)]_{s=1}, \dots, [\mathbf{a}_t(\phi_n, \theta_n)]_{s=N_{BS}}]^H$. Similarly,

$$\mathbf{B} = [\mathbf{a}_r(\varphi_1, \vartheta_1), \dots, \mathbf{a}_r(\varphi_N, \vartheta_N)], \quad (9)$$

where $\mathbf{a}_r(\varphi_n, \vartheta_n) = [[\mathbf{a}_r(\varphi_n, \vartheta_n)]_{u=1}, \dots, [\mathbf{a}_r(\varphi_n, \vartheta_n)]_{u=N_{MS}}]^T$.

Note that the array responses and antenna patterns are captured in \mathbf{A} and \mathbf{B} . Since the transmitted and received powers are incorporated in the antenna patterns so \mathbf{P}_{Rx} and \mathbf{P}_{Tx} in (7) are identity matrices. For the single-bounce scattering model, the mask matrix Ω will be diagonal. Taking Ψ and Φ to be \mathbf{B} and \mathbf{A} respectively, the solution of the maximum entropy problem for the $N_{MS} \times N_{BS}$ single-bounce scattering 3D MIMO channel model in (1) will have a systematic structure as follows,

$$\mathbf{H} = \frac{1}{\sqrt{N}} \mathbf{B} \text{diag}(\boldsymbol{\alpha}) \mathbf{A}^H, \quad (10)$$

where $\boldsymbol{\alpha}$ is an N dimensional vector with entries that are i.i.d zero mean, unit variance Gaussian RVs.

III. MUTUAL INFORMATION ANALYSIS IN FINITE NUMBER OF ANTENNAS REGIME

In this section, we explicitly define the MI for subsequent performance prediction and analysis. We derive an analytical expression for the statistical distribution of the MI in the general SINR regime for a single Rx antenna and any arbitrary finite number of Tx antennas. The result is then extended to an arbitrary finite number of Rx antennas in the low SINR regime.

We consider a point-to-point communication link where \mathbf{H} is a $N_{MS} \times N_{BS}$ MIMO channel matrix from (10). The channel is linear and time-invariant. A time-division duplex (TDD) protocol is considered where BS acquires instantaneous channel state information (CSI) in the uplink and uses it for downlink transmission by exploiting the channel reciprocity.

The received complex baseband signal $\mathbf{y} \in \mathbb{C}^{N_{MS} \times 1}$ at the MS is given by,

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (11)$$

where $\mathbf{s} \in \mathbb{C}^{N_{BS} \times 1}$ is the Tx signal from the BS and $\mathbf{n} \in \mathbb{C}^{N_{MS} \times 1}$ is the complex additive noise such that $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \mathbf{R} + \sigma^2\mathbf{I}$, where \mathbf{R} is the covariance matrix of interference experienced by the MS and σ^2 is the noise variance at the MS.

In this context, the MI of the 3D MIMO system is given by,

$$I(\sigma^2) = \log \det(\mathbf{I}_{N_{MS}} + (\mathbf{R} + \sigma^2\mathbf{I}_{N_{MS}})^{-1}\mathbf{H}\mathbf{H}^H). \quad (12)$$

\mathbf{H} is fixed during the communication interval, so we do not need to time average the MI.

A. General SINR Regime

The exact MI given in (12) has been mostly characterized in literature for the cases when $\mathbf{H}\mathbf{H}^H$ is a Wishart matrix [9]–[11], [13]. Wishart matrices are random matrices with independent entries and have been frequently studied in the context of wireless communications. However the special structure of our channel model reduces the degree of randomness as now \mathbf{H} is a diagonal matrix of random Gaussian RVs as opposed to the whole matrix having random entries, which implies that we can not employ the results developed for Wishart matrices. However fortunately for the single-bounce scattering model in (10), we can obtain results by exploiting the quadratic nature of the entries of $\mathbf{H}\mathbf{H}^H$. For the case, when MS is equipped with a single antenna, an exact closed-form solution can be obtained and is provided in the following theorem.

Theorem 1: Let the MS be equipped with a single Rx antenna and the BS with any arbitrary number of Tx antennas, then the statistical distribution of MI in the general SINR regime can be expressed in a closed-form as,

$$\mathbb{P}[I(\sigma^2) \leq y] = 1 - \sum_{i=1}^N \frac{\lambda_i^N}{\prod_{l \neq i} (\lambda_i - \lambda_l)} \frac{1}{\lambda_i} \exp\left(-\frac{(e^y - 1)}{\lambda_i}\right), \quad (13)$$

where λ_i 's are the eigenvalues of \mathbf{C} given that,

$$\mathbf{C} = \frac{1}{N} (\mathbf{B}^H \mathbf{\Omega} \mathbf{B}) \circ (\mathbf{A}^H \mathbf{A})^T, \quad (14)$$

where $\mathbf{\Omega} = (\mathbf{R} + \sigma^2 \mathbf{I}_{MS})^{-1}$.

Proof: The proof follows from realizing that when the number of antennas at the MS is one, $(\mathbf{R} + \sigma^2 \mathbf{I}_{MS})^{-1} \mathbf{H}\mathbf{H}^H$ is a scalar and can be expressed as a quadratic form in α . The MI is given as a function of this quadratic form as,

$$I(\sigma^2) = \log(1 + \alpha^H \mathbf{C} \alpha). \quad (15)$$

where \mathbf{C} is given by (14).

This quadratic form in Normal RVs will play a crucial role in the subsequent analysis in this work. The CF of a quadratic form in Normal RVs is given by [37], [38],

$$\mathbb{E}[e^{j(\alpha^H \mathbf{C} \alpha)}] = \frac{1}{\det[\mathbf{I}_N - j \frac{1}{N} ((\mathbf{B}^H \mathbf{\Omega} \mathbf{B}) \circ (\mathbf{A}^H \mathbf{A})^T)]}. \quad (16)$$

It is easy to see that the denominator of (16) can be equivalently expressed as $\prod_{i=1}^N (1 - j\lambda_i)$, where λ_i 's are the eigenvalues of \mathbf{C} . (16) has a form similar to the CF of the sum of i.i.d

unit parameter exponential RVs scaled by λ_i 's, i.e. $\boldsymbol{\alpha}^H \mathbf{C} \boldsymbol{\alpha} \sim \sum_{i=1}^N \lambda_i \text{Exp}(1)$. Consequently, the closed-form expression of the statistical distribution of $(\mathbf{R} + \sigma^2 \mathbf{I}_{N_{MS}})^{-1} \mathbf{H} \mathbf{H}^H$ is given by [39],

$$\mathbb{P}[\boldsymbol{\alpha}^H \mathbf{C} \boldsymbol{\alpha} \leq x] = 1 - \sum_{i=1}^N \frac{\lambda_i^N}{\prod_{l \neq i} (\lambda_i - \lambda_l)} \frac{1}{\lambda_i} \exp\left(-\frac{x}{\lambda_i}\right),$$

The rest of the analysis can be completed through the transformation $\log(1 + x)$.

$$\mathbb{P}[I(\sigma^2) \leq y] = \mathbb{P}[\log(1 + x) \leq y] = F_x(e^y - 1). \quad (17)$$

The theoretical result coincides very well with the simulated CDF for the general SINR regime, as will be illustrated later in the simulations section.

B. Low SINR Regime

It is well known that the performance of the channel can severely deteriorate in the low SINR regime, unless intelligent precoding and signaling schemes are employed [40]. It is important to develop analytical expressions for the MI of MIMO channels in this regime to allow for the analysis of various signaling and precoding schemes. We develop such an analytical expression for the 3D channel model in (10) by exploiting its structure. The closed-form expression for the CDF of MI in the low SINR regime is now presented in the following theorem.

Theorem 2: In the low SINR regime, the approximate statistical distribution of the MI can be expressed in a closed-form as,

$$\mathbb{P}[I(\sigma^2) \leq x] = 1 - \sum_{i=1}^N \frac{\lambda_i^N}{\prod_{l \neq i} (\lambda_i - \lambda_l)} \frac{1}{\lambda_i} \exp\left(-\frac{x}{\lambda_i}\right), \quad (18)$$

where λ_i 's are the eigenvalues of \mathbf{C} and,

$$\mathbf{C} = \frac{1}{N} (\mathbf{B}^H \boldsymbol{\Omega} \mathbf{B}) \circ (\mathbf{A}^H \mathbf{A})^T, \quad (19)$$

where $\boldsymbol{\Omega} = (\mathbf{R} + \sigma^2 \mathbf{I}_{N_{MS}})^{-1}$.

Proof: The proof of Theorem 2 follows from making use of the relation that for $\boldsymbol{\delta}$ with $\|\boldsymbol{\delta}\|_2 \approx 0$,

$$\log \det(\mathbf{X} + \boldsymbol{\delta} \mathbf{X}) = \log \det(\mathbf{X}) + \text{Tr}(\mathbf{X}^{-1} \boldsymbol{\delta} \mathbf{X}). \quad (20)$$

In the low SINR regime, $\|(\mathbf{R} + \sigma^2 \mathbf{I}_{N_{MS}})^{-1}\|_2 \approx 0$ so taking δ as $(\mathbf{R} + \sigma^2 \mathbf{I}_{N_{MS}})^{-1} \mathbf{H} \mathbf{H}^H$ and \mathbf{X} as $\mathbf{I}_{N_{MS}}$ yields,

$$I(\sigma^2) = \text{Tr}((\mathbf{R} + \sigma^2 \mathbf{I}_{N_{MS}})^{-1} \mathbf{H} \mathbf{H}^H). \quad (21)$$

Observe from (10) that the trace of $\mathbf{H} \mathbf{H}^H$ can be expressed as a quadratic form given by,

$$\text{Tr}(\mathbf{\Omega} \mathbf{H} \mathbf{H}^H) = \boldsymbol{\alpha}^H \mathbf{C} \boldsymbol{\alpha}, \quad (22)$$

where $\mathbf{\Omega}$ is an arbitrary deterministic matrix and

$$[\mathbf{C}]_{ij} = \frac{1}{N} [\mathbf{B}^H \mathbf{\Omega} \mathbf{B}]_{ij} \circ [\mathbf{A}^H \mathbf{A}]_{ji}. \quad (23)$$

Using (22), $I(\sigma^2)$ in (21) is a quadratic form in zero mean, unit variance Normal i.i.d RVs α 's,

$$I(\sigma^2) = \boldsymbol{\alpha}^H \mathbf{C} \boldsymbol{\alpha}, \quad (24)$$

where \mathbf{C} is given by (19). The rest of the proof follows from that for the general SINR regime, where we already characterized the distribution of $\boldsymbol{\alpha}^H \mathbf{C} \boldsymbol{\alpha}$. The theoretical result coincides very well with the simulated CDF for the low SINR regime, as will be illustrated later in the numerical results.

IV. ASYMPTOTIC ANALYSIS OF THE MUTUAL INFORMATION

As explained earlier, the reduced degree of randomness in our model makes it difficult to characterize the behaviour of the MI. The case for a single Rx antenna was dealt with in the last section and a closed-form was obtained as a consequence of the quadratic nature of the entries of $\mathbf{H} \mathbf{H}^H$. Even when $\mathbf{H} \mathbf{H}^H$ is a Wishart matrix, the study of the exact statistical distribution of the MI becomes rather involved because it often succumbs to taking the inverse Fourier transform of the CF or applying Jacobian transformations on the eigenvalue distribution of Wishart matrices [12], [13]. As a result, a lot of work resorts to the asymptotic analysis and provides Gaussian approximations to the distribution of the MI. The asymptotic analysis provided here differs from the previous analysis because of the non-Wishart nature of $\mathbf{H} \mathbf{H}^H$ and the incorporation of the elevation dimension and antenna tilt angles in the channel model.

In order to study the asymptotic approximation to the MI distribution, the following assumption is required over the number of BS antennas and multipaths.

Assumption A-1. In the large (N_{BS}, N) regime, N_{BS} and N tend to infinity such that

$$0 < \liminf \frac{N_{BS}}{N} \leq \limsup \frac{N_{BS}}{N} < +\infty, \quad (25)$$

a condition we shall refer to by writing $N_{BS}, N \rightarrow \infty$. The analysis stays valid for a moderate number of N_{BS} and N as well, as will be confirmed through the simulation results later.

A. Distribution of $\mathbf{H}\mathbf{H}^H$

The analysis starts by characterizing the distribution of $\mathbf{H}\mathbf{H}^H$, which would later be followed by transformations to complete the characterization of MI. Defining $\bar{\mathbf{b}}_i$ as the i^{th} row vector of \mathbf{B} in (9), the $(k, l)^{th}$ entry of $\mathbf{H}\mathbf{H}^H$ can be expressed as,

$$[\mathbf{H}\mathbf{H}^H]_{kl} = \frac{1}{N} \alpha^H ([\bar{\mathbf{b}}_k^H \bar{\mathbf{b}}_l] \circ (\mathbf{A}^H \mathbf{A})^T) \alpha. \quad (26)$$

To characterize the distribution, an additional assumption is required over the matrices that form the quadratic forms in $\mathbf{H}\mathbf{H}^H$. Denoting N_{MS} by M , the assumption is stated as follows.

Assumption A-2. Under the setting of Assumption A-1,

$$\frac{\|[\bar{\mathbf{b}}_k^H \bar{\mathbf{b}}_l] \circ (\mathbf{A}^H \mathbf{A})^T\|_{sp}}{\|[\bar{\mathbf{b}}_k^H \bar{\mathbf{b}}_l] \circ (\mathbf{A}^H \mathbf{A})^T\|_F} \rightarrow 0, \quad \forall k, l = 1, \dots, M, \quad (27)$$

where $\|\cdot\|_{sp}$ denotes the spectral norm and $\|\cdot\|_F$ denotes the Forbenius norm of a matrix.

Therefore $\mathbf{H}\mathbf{H}^H$ is a $M \times M$ matrix of the following general form,

$$\frac{1}{N} \begin{bmatrix} \alpha^H \mathbf{C}^{1,1} \alpha & \alpha^H \mathbf{C}^{1,2} \alpha & \dots & \alpha^H \mathbf{C}^{1,M} \alpha \\ \alpha^H \mathbf{C}^{2,1} \alpha & \alpha^H \mathbf{C}^{2,2} \alpha & \dots & \alpha^H \mathbf{C}^{2,M} \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^H \mathbf{C}^{M,1} \alpha & \alpha^H \mathbf{C}^{M,2} \alpha & \dots & \alpha^H \mathbf{C}^{M,M} \alpha \end{bmatrix}, \quad (28)$$

where $\mathbf{C}^{k,l} = [\bar{\mathbf{b}}_k^H \bar{\mathbf{b}}_l] \circ (\mathbf{A}^H \mathbf{A})^T$, $\forall k, l$ satisfies assumption A-2. This is a technical assumption required to observe the convergence behavior that will be discussed hereon.

We investigate the asymptotic behaviour of the joint distribution of the M^2 entries of $\mathbf{H}\mathbf{H}^H$ under the assumptions A-1 and A-2. Note that N_{BS} needs to be of at least the same order as N to prevent \mathbf{C} 's from becoming ill-conditioned. This is definitely not a problem in Massive MIMO systems where N_{BS} is generally much higher than N . N_{MS} can be any finite number as

per the practical application. Denoting the $(k, l)^{th}$ quadratic term by $T^{k,l} = \frac{1}{N} \alpha^H \mathbf{C}^{k,l} \alpha$, every quadratic term can be decomposed into its real and imaginary parts increasing the total number of terms that constitute $\mathbf{H}\mathbf{H}^H$ to $2M^2$. These parts are also quadratic forms in α as shown below,

$$T^{k,l} = \frac{1}{N} \Re(\alpha^H \mathbf{C}^{k,l} \alpha) + j \frac{1}{N} \Im(\alpha^H \mathbf{C}^{k,l} \alpha), \quad \forall k, l = 1, \dots, M, \quad (29)$$

where,

$$\begin{aligned} \Re(\alpha^H \mathbf{C}^{k,l} \alpha) &= [\Re(\alpha)^T \quad \Im(\alpha)^T] \begin{bmatrix} \Re(\mathbf{C}^{k,l}) & -\Im(\mathbf{C}^{k,l}) \\ \Im(\mathbf{C}^{k,l}) & \Re(\mathbf{C}^{k,l}) \end{bmatrix} \begin{bmatrix} \Re(\alpha) \\ \Im(\alpha) \end{bmatrix}, \\ &= [\Re(\alpha)^T \quad \Im(\alpha)^T] (\mathbf{C}_{\Re}^{k,l}) \begin{bmatrix} \Re(\alpha) \\ \Im(\alpha) \end{bmatrix}. \end{aligned} \quad (30)$$

$$\begin{aligned} \Im(\alpha^H \mathbf{C}^{k,l} \alpha) &= [\Re(\alpha)^T \quad \Im(\alpha)^T] \begin{bmatrix} \Im(\mathbf{C}^{k,l}) & \Re(\mathbf{C}^{k,l}) \\ -\Re(\mathbf{C}^{k,l}) & \Im(\mathbf{C}^{k,l}) \end{bmatrix} \begin{bmatrix} \Re(\alpha) \\ \Im(\alpha) \end{bmatrix}, \\ &= [\Re(\alpha)^T \quad \Im(\alpha)^T] (\mathbf{C}_{\Im}^{k,l}) \begin{bmatrix} \Re(\alpha) \\ \Im(\alpha) \end{bmatrix}. \end{aligned} \quad (31)$$

Equivalently $\mathbf{H}\mathbf{H}^H$ can be written as a sum of two matrices containing the real and imaginary parts of the quadratic terms.

$$\mathbf{H}\mathbf{H}^H = \frac{1}{N} \begin{bmatrix} \Re(\alpha^H \mathbf{C}^{1,1} \alpha) & \dots & \Re(\alpha^H \mathbf{C}^{1,M} \alpha) \\ \Re(\alpha^H \mathbf{C}^{2,1} \alpha) & \dots & \Re(\alpha^H \mathbf{C}^{2,M} \alpha) \\ \vdots & \ddots & \vdots \\ \Re(\alpha^H \mathbf{C}^{M,1} \alpha) & \dots & \Re(\alpha^H \mathbf{C}^{M,M} \alpha) \end{bmatrix} + \frac{j}{N} \begin{bmatrix} \Im(\alpha^H \mathbf{C}^{1,1} \alpha) & \dots & \Im(\alpha^H \mathbf{C}^{1,M} \alpha) \\ \Im(\alpha^H \mathbf{C}^{2,1} \alpha) & \dots & \Im(\alpha^H \mathbf{C}^{2,M} \alpha) \\ \vdots & \ddots & \vdots \\ \Im(\alpha^H \mathbf{C}^{M,1} \alpha) & \dots & \Im(\alpha^H \mathbf{C}^{M,M} \alpha) \end{bmatrix}. \quad (32)$$

Note that $\mathbb{E}[\Re(\alpha)^2] = \mathbb{E}[\Im(\alpha)^2] = \frac{1}{2}$. With this decomposition at hand, we now state in the following theorem the asymptotic behavior of the joint distribution of the entries of $\Re(\mathbf{H}\mathbf{H}^H)$ and $\Im(\mathbf{H}\mathbf{H}^H)$ stacked in vector \mathbf{x} given by,

$$\mathbf{x} = [\text{vec}(\Re(\mathbf{H}\mathbf{H}^H))^T \quad \text{vec}(\Im(\mathbf{H}\mathbf{H}^H))^T]^T, \quad (33)$$

where the operator $\text{vec}(\cdot)$ maps an $M \times M$ matrix to an $M^2 \times 1$ vector by stacking the rows of the matrix.

Theorem 3: Let \mathbf{H} be a $N_{MS} \times N_{BS}$ MIMO channel matrix from (10), then under the setting of assumptions **A-1** and **A-2**, \mathbf{x} behaves as multivariate Gaussian such that the MGF of \mathbf{x} has the following convergence behavior,

$$\mathbb{E} \left[\exp \left(\frac{\sqrt{N} \mathbf{s}^T (\mathbf{x} - \mathbf{m})}{\sqrt{\mathbf{s}^T \Theta \mathbf{s}}} \right) \right] - \exp \left(\frac{1}{2} \right) \rightarrow 0, \quad (34)$$

where,

$$\mathbf{m} = \frac{1}{2N} [Tr(\mathbf{C}_{\mathcal{R}}^{1,1}) \quad \dots \quad Tr(\mathbf{C}_{\mathcal{R}}^{M,M}), Tr(\mathbf{C}_{\mathcal{S}}^{1,1}) \quad \dots \quad Tr(\mathbf{C}_{\mathcal{S}}^{M,M})]^T, \quad (35)$$

and

$$\Theta = \begin{bmatrix} \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{R}}^{1,1}(\mathbf{C}_{\mathcal{R}}^{1,1})) + Tr(\mathbf{C}_{\mathcal{R}}^{1,1}(\mathbf{C}_{\mathcal{R}}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{R}}^{1,1}(\mathbf{C}_{\mathcal{S}}^{M,M})) + Tr(\mathbf{C}_{\mathcal{R}}^{1,1}(\mathbf{C}_{\mathcal{S}}^{M,M})^H)] \\ \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{R}}^{1,2}(\mathbf{C}_{\mathcal{R}}^{1,1})) + Tr(\mathbf{C}_{\mathcal{R}}^{1,2}(\mathbf{C}_{\mathcal{R}}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{R}}^{1,2}(\mathbf{C}_{\mathcal{S}}^{M,M})) + Tr(\mathbf{C}_{\mathcal{R}}^{1,2}(\mathbf{C}_{\mathcal{S}}^{M,M})^H)] \\ \vdots & \ddots & \vdots \\ \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{R}}^{M,M}(\mathbf{C}_{\mathcal{R}}^{1,1})) + Tr(\mathbf{C}_{\mathcal{R}}^{M,M}(\mathbf{C}_{\mathcal{R}}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{R}}^{M,M}(\mathbf{C}_{\mathcal{S}}^{M,M})) + Tr(\mathbf{C}_{\mathcal{R}}^{M,M}(\mathbf{C}_{\mathcal{S}}^{M,M})^H)] \\ \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{S}}^{1,1}(\mathbf{C}_{\mathcal{R}}^{1,1})) + Tr(\mathbf{C}_{\mathcal{S}}^{1,1}(\mathbf{C}_{\mathcal{R}}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{S}}^{1,1}(\mathbf{C}_{\mathcal{S}}^{M,M})) + Tr(\mathbf{C}_{\mathcal{S}}^{1,1}(\mathbf{C}_{\mathcal{S}}^{M,M})^H)] \\ \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{S}}^{1,2}(\mathbf{C}_{\mathcal{R}}^{1,1})) + Tr(\mathbf{C}_{\mathcal{S}}^{1,2}(\mathbf{C}_{\mathcal{R}}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{S}}^{1,2}(\mathbf{C}_{\mathcal{S}}^{M,M})) + Tr(\mathbf{C}_{\mathcal{S}}^{1,2}(\mathbf{C}_{\mathcal{S}}^{M,M})^H)] \\ \vdots & \ddots & \vdots \\ \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{S}}^{M,M}(\mathbf{C}_{\mathcal{R}}^{1,1})) + Tr(\mathbf{C}_{\mathcal{S}}^{M,M}(\mathbf{C}_{\mathcal{R}}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\mathcal{S}}^{M,M}(\mathbf{C}_{\mathcal{S}}^{M,M})) + Tr(\mathbf{C}_{\mathcal{S}}^{M,M}(\mathbf{C}_{\mathcal{S}}^{M,M})^H)] \end{bmatrix}. \quad (36)$$

The proof of Theorem 3 is postponed to Appendix A. Note that this theorem can be used in applications involving the statistical distribution of $\mathbf{H}\mathbf{H}^H$ other than the MI analysis.

B. Distribution of Mutual Information

After studying the asymptotic behaviour of $\Re(\mathbf{H}\mathbf{H}^H)$, $\Im(\mathbf{H}\mathbf{H}^H)$ for the proposed entropy maximizing channel, the CDF of the MI is worked out. To this end, note that (12) can be equivalently expressed as,

$$I(\sigma^2) = \log \det((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H) - \log \det(\mathbf{R} + \sigma^2 \mathbf{I}_M), \quad (37)$$

Decomposing $((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H)$ into real and imaginary parts as,

$$\begin{aligned} (\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H &= \Re((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H) + j \Im((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H), \\ &= \mathbf{Y} + j \mathbf{Z}, \end{aligned} \quad (38)$$

and denoting the entries of $\mathbf{R} + \sigma^2 \mathbf{I}_M$ by ζ_{ij} , we can extend the result of Theorem 3 to the entries of $\Re((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H)$ and $\Im((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H)$, which will also show the convergence behavior in (34) under assumptions **A-1** and **A-2** with mean vector \mathbf{m} given by,

$$\mathbf{m} = [\Re(\zeta_{11}) + \frac{1}{2N} \text{Tr}(\mathbf{C}_{\Re}^{1,1}), \Re(\zeta_{12}) + \frac{1}{2N} \text{Tr}(\mathbf{C}_{\Re}^{1,2}), \dots, \Re(\zeta_{MM}) + \frac{1}{2N} \text{Tr}(\mathbf{C}_{\Re}^{M,M}), \\ \Im(\zeta_{11}) + \frac{1}{2N} \text{Tr}(\mathbf{C}_{\Im}^{1,1}), \Im(\zeta_{12}) + \frac{1}{2N} \text{Tr}(\mathbf{C}_{\Im}^{1,2}), \dots, \Im(\zeta_{MM}) + \frac{1}{2N} \text{Tr}(\mathbf{C}_{\Im}^{M,M})]^T \quad (39)$$

and the same covariance.

The mean matrix \mathbf{M} for $((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H)$ is expressed as,

$$\mathbf{M} = \mathbf{M}_1 + j \mathbf{M}_2, \quad (40)$$

where,

$$\mathbf{M}_1 = \begin{bmatrix} \Re(\zeta_{11}) + \frac{1}{2N} \text{Tr}(C_{\Re}^{1,1}) & \dots & \Re(\zeta_{1M}) + \frac{1}{2N} \text{Tr}(C_{\Re}^{1,M}) \\ \Re(\zeta_{21}) + \frac{1}{2N} \text{Tr}(C_{\Re}^{2,1}) & \dots & \Re(\zeta_{2M}) + \frac{1}{2N} \text{Tr}(C_{\Re}^{2,M}) \\ \vdots & \ddots & \vdots \\ \Re(\zeta_{M1}) + \frac{1}{2N} \text{Tr}(C_{\Re}^{M,1}) & \dots & \Re(\zeta_{MM}) + \frac{1}{2N} \text{Tr}(C_{\Re}^{M,M}) \end{bmatrix}, \quad (41)$$

$$\mathbf{M}_2 = \begin{bmatrix} \Im(\zeta_{11}) + \frac{1}{2N} \text{Tr}(C_{\Im}^{1,1}) & \dots & \Im(\zeta_{1M}) + \frac{1}{2N} \text{Tr}(C_{\Im}^{1,M}) \\ \Im(\zeta_{21}) + \frac{1}{2N} \text{Tr}(C_{\Im}^{2,1}) & \dots & \Im(\zeta_{2M}) + \frac{1}{2N} \text{Tr}(C_{\Im}^{2,M}) \\ \vdots & \ddots & \vdots \\ \Im(\zeta_{M1}) + \frac{1}{2N} \text{Tr}(C_{\Im}^{M,1}) & \dots & \Im(\zeta_{MM}) + \frac{1}{2N} \text{Tr}(C_{\Im}^{M,M}) \end{bmatrix}. \quad (42)$$

Before presenting the Gaussian approximation to the statistical distribution of the MI in the asymptotic limit, we define a matrix $\tilde{\mathbf{M}}$ and a $2M^2 \times 1$ vector $\mathbf{f}(\tilde{\mathbf{M}})$ as,

$$\tilde{\mathbf{M}} = \begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix}, \quad (43)$$

$$\mathbf{f}(\tilde{\mathbf{M}})|_{l+(k-1)M} = [\det(\mathbf{D}_1) + \dots + \det(\mathbf{D}_{2M})]|_{l+(k-1)M}, \quad k = 1, \dots, 2M, l = 1, \dots, M, \quad (44)$$

where \mathbf{D}_i is identical to $\tilde{\mathbf{M}}$ except that the entries in the i^{th} row are replaced by their derivatives with respect to $(k, l)^{\text{th}}$ entry of $\begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{bmatrix}$. Every entry of the $2M^2 \times 1$ vector, $\mathbf{f}(\tilde{\mathbf{M}})$ will involve the sum of only two non-zero determinants as every $(k, l)^{\text{th}}$ entry occurs in only two rows.

Additionally we make the following assumption,

Assumption A-3. In the large N_{BS}, N regime presented in **A-1**,

$$\liminf \underline{\mathbf{f}}(\tilde{\mathbf{M}})^T \Theta \underline{\mathbf{f}}(\tilde{\mathbf{M}}) > 0, \quad (45)$$

where Θ is the covariance matrix from (36).

With all these definitions at hand, we now present in the following theorem the Gaussian approximation to the statistical distribution of the MI in the asymptotic limit.

Theorem 4: Let \mathbf{H} be a $N_{MS} \times N_{BS}$ MIMO channel matrix from (10), such that the assumptions **A-1**, **A-2** and **A-3** hold, then the statistical distribution of $I(\sigma^2)$ can be approximated as,

$$\mathbb{P}[\sqrt{N}I(\sigma^2) \leq x] - \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{x - \sqrt{N}\mu}{\sqrt{2\sigma_a^2}} \right) \right) \rightarrow 0, \quad (46)$$

where erf is the error function, $\mu = 0.5 \log \det \tilde{\mathbf{M}} - \log \det(\mathbf{R} + \sigma^2 \mathbf{I}_M)$ and σ_a^2 is given by,

$$\left(\frac{.5}{\det \tilde{\mathbf{M}}} \right)^2 \times \underline{\mathbf{f}}(\tilde{\mathbf{M}})^T \Theta \underline{\mathbf{f}}(\tilde{\mathbf{M}}), \quad (47)$$

The proof of Theorem 4 is postponed to Appendix B. The approximation becomes more accurate as N_{BS} and N grow large but yields a good fit for even moderate values of N_{BS} and N . The Gaussian approximation to the MI is an effective tool for the performance evaluation of future 3D MIMO channels in Massive MIMO systems that will scale up the current MIMO systems by possible order of magnitudes compared to the current state-of-the-art. However accommodating more antennas at the MS introduces several constraints for practical implementations which would not affect the presented results since the approximation is valid for any number of Rx antennas

V. NUMERICAL RESULTS

We corroborate the results derived in this work with simulations. A simple but realistic scenario shown in Fig. 3 is studied to help gain insight into the accuracy of the derived CDFs and the impact of the careful selection of antenna downtilt angles on the system MI in the presence of interfering BSs. This multi-cell scenario consists of a single MS and two BSs separated by an inter-site distance of 500m. The BS height is 25m and MS height is 1.5m. The MS is associated to BS 1 and can be located inside a cell of outer radius 250m and inner radius 35m. We will study the worst case performance when the MS is at the cell edge. The MI given in (12) requires

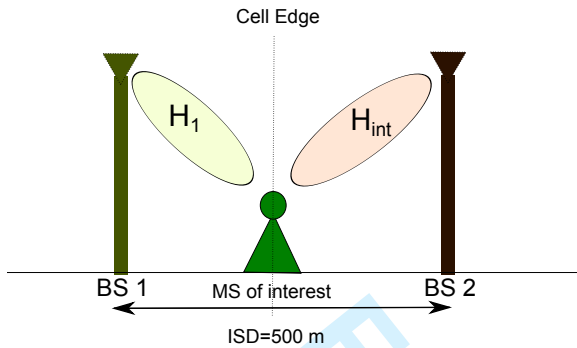


Fig. 3. Example scenario.

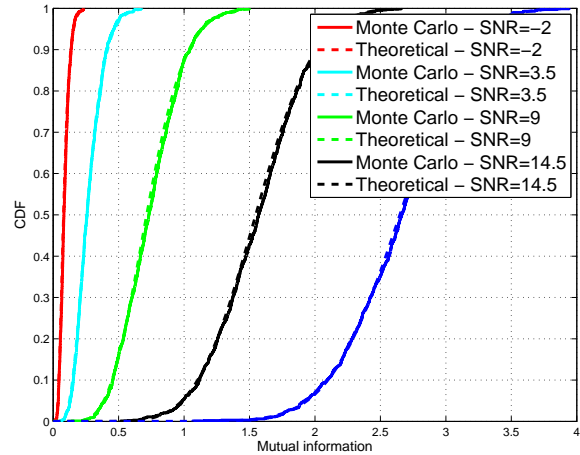


Fig. 4. Comparison of Monte Carlo simulated CDF and theoretical CDF derived in (13) for the general SINR regime.

the computation of the interference matrix. The complex received baseband signal at the MS in the multi-cell scenario will be given as,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \sum_{i=1}^{N_{int}} \mathbf{H}_i \mathbf{x}_i + \mathbf{n}, \quad (48)$$

where N_{int} is the number of interfering BSs ($N_{int}=1$ for the scenario in Fig. 3) and \mathbf{n} is the additive white Gaussian noise with variance σ^2 . The channel between the serving BS and MS is denoted by \mathbf{H} from (10) and the channels \mathbf{H}_i between N_{int} interfering BSs and MS follow a similar matrix representation. The resultant interference matrix is then given by,

$$\mathbf{R} = \sum_{i=1}^{N_{int}} \mathbb{E}[\mathbf{H}_i \mathbf{H}_i^H], \quad (49)$$

where,

$$\mathbb{E}[\mathbf{H}_i \mathbf{H}_i^H] = \frac{1}{N} \mathbf{B}_i \text{diag}(\text{diag}(\mathbf{A}_i^H \mathbf{A}_i)) \mathbf{B}_i^H, \quad i = 1, \dots, N_{int}. \quad (50)$$

Since this paper largely focuses on the guidelines provided in the mobile communication standards used globally, so it is important that we use the angular distributions and antenna patterns specified in the standards for our analysis. In the standards, elevation AoD and AoA are drawn from Laplacian elevation density spectrum given by,

$$f_\theta(\theta) \propto \exp\left(-\frac{\sqrt{2}|\theta - \theta_0|}{\sigma}\right), \quad (51)$$

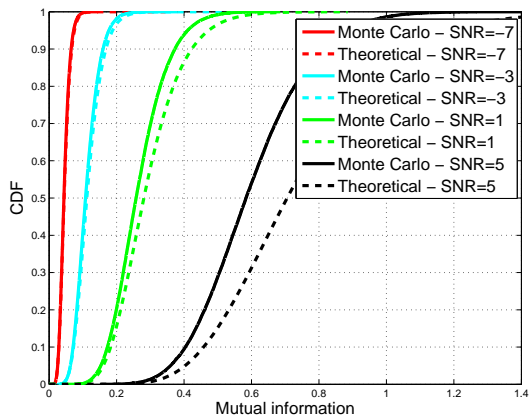


Fig. 5. Comparison of Monte Carlo simulated CDF and theoretical CDF derived in (18) for the low SINR regime.

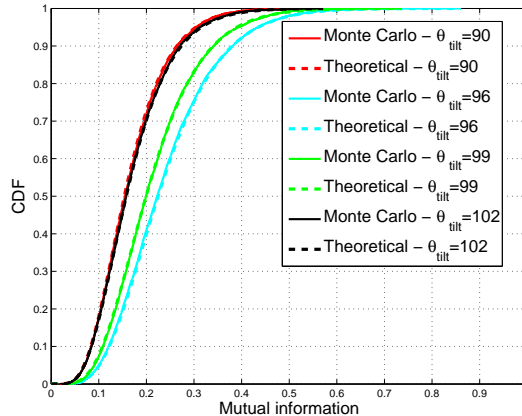


Fig. 6. Comparison in a multi-cell environment for different values of θ_{tilt} .

where θ_0 is the mean AoD/AoA and σ is the angular spread in the elevation [34]. The characteristics of the azimuth angles are well captured by Wrapped Gaussian (WG) density spectrum [30], [35] which can be approximated quite accurately by the Von Mises (VM) distribution [41], [42] given by,

$$f_{\phi}(\phi) = \frac{\exp(\kappa \cos(x - \mu))}{2\pi I_0(\kappa)}, \quad (52)$$

where $I_n(\kappa)$ is the modified Bessel function of order n , μ is the mean AoD/AoA and $\frac{1}{\kappa}$ is a measure of azimuth dispersion. Since the channel assumes single-bounce scattering, so the angles are generated once and are assumed to be constant over the channel realizations.

The radio and channel parameters are set as $\theta_{3dB} = 15^\circ$, $\phi_{3dB} = 70^\circ$, $\sigma_{BS} = 7^\circ$, $\sigma_{MS} = 10^\circ$, $\kappa_{BS}, \kappa_{MS} = 5$ and $\mu = 0$. Moreover θ_0 is set equal to the elevation LoS angle between the BS and MS. Two thousand Monte Carlo realizations of the 3D channel in (1) are generated to obtain the simulated MI for comparison. In the simulations, SNR is used to denote $\frac{1}{\sigma^2}$ and is given in dB. The closed-form expression of the CDF of MI provided in Theorem 1 (13) for the general SINR regime is validated through simulation in Fig. 4 for $N_{BS} = 20$ and $N_{MS} = 1$. The number of propagation paths N is assumed to be 40 and θ_{tilt} is set equal to the elevation LoS angle of the BS with the MS. For this result, BS 2 is assumed to be operating in a different frequency band as BS 1 so the MS does not face any interference and $(\mathbf{R} + \sigma^2 \mathbf{I}_{N_{MS}})^{-1}$ in (12) reduces to $\frac{1}{\sigma^2} \mathbf{I}_{N_{MS}}$. The closed-form expression for the CDF provides a perfect fit to the MI obtained

using Monte Carlo realizations of the channel in (1). In the next numerical result, we validate the expression in (18) for the low SINR regime. The result is plotted for $N_{MS} = 2$ in Fig. 5. The approximation is seen to coincide quite well with the simulated results for the proposed 3D channel model at low values of SNR . However for SNR level of 5 dB or above, the fit is not very good as the approximation in (20) starts to become inaccurate.

In the next simulation, we investigate the impact of the downtilt angle of the transmitting cell on the performance of the users at the cell edge in a multi-cell scenario, when both BSs are transmitting in the same frequency band. The interference matrix \mathbf{R} is computed from (50). Note from Fig. 3 that the MS is in the direct boresight of the interfering cell and has θ_{LoS} of 95.37° with both the serving and interfering BS. It can be clearly seen from Fig. 6 that the performance of the user at the cell edge is very sensitive to the value of the downtilt angle. The simulation result illustrates that the MI of the system is maximized when the serving BS also sets the elevation boresight angles of its antennas equal to the elevation LoS angle with the MS, i.e. $\theta_{tilt} \approx 96^\circ$. We plot the CDF for the exact expression of MI for a single Rx antenna provided in (13) for $SNR = 5$ dB. The perfect fit of the theoretical result to the simulated result for the CDF of the MI is again evident.

We now validate the result obtained for the CDF of MI in the asymptotic limit, as N_{BS} and N grow large, for the cases with finite values of N_{BS} and N . The first result deals with scenario when the BS 2 is operating in a different frequency band as BS 1. Fig. 7 shows that there is quite a good fit between the asymptotic theoretical result and the simulated result for finite sized simulated system with $N_{BS} = 60$, a reasonable number for large MIMO systems and $N_{MS} = 4$. The number of paths is taken to be of the same order as N_{BS} . Even at the highest simulated value of SNR , the asymptotic mean and asymptotic variance show only a 1.2 and 3.2 percent relative error respectively. This close match illustrates the usefulness of the asymptotic approach in characterizing the MI for 3D maximum entropy channels for any number of Rx antennas. Hence, as far as performance analysis based on the ergodic MI is considered, the asymptotic statistical distribution obtained can be used for the evaluation of future 3D beamforming strategies. Finally the multi-cell scenario is considered in Fig. 8, where we plot the CDF of the MI achieved by the proposed entropy maximizing channel in the asymptotic limit at different values of antenna boresight angles of the serving BS. We again deal with the worst case scenario of the MS being in the direct boresight of the interfering cell. Comparison with the simulated CDF of MI obtained

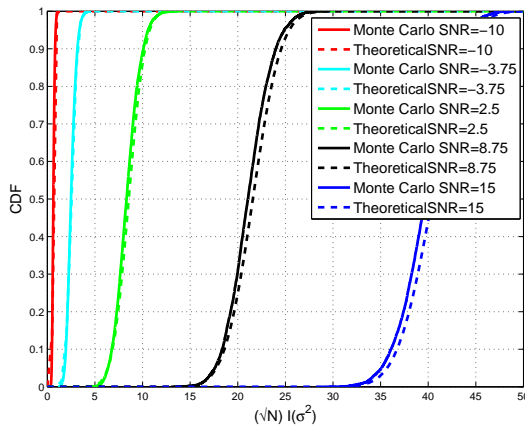


Fig. 7. Comparison of Monte Carlo simulated CDF and asymptotic theoretical CDF in (46).

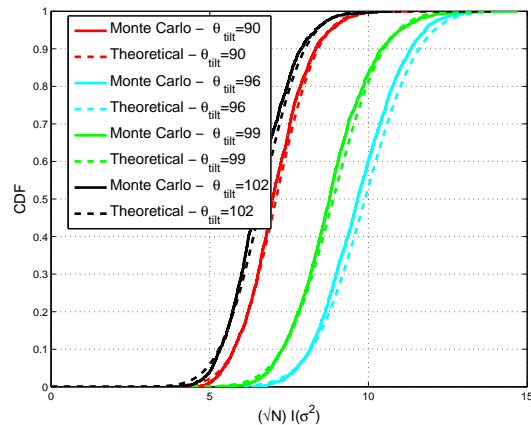


Fig. 8. Comparison of Monte Carlo simulated CDF and asymptotic theoretical CDF in a multi-cell environment.

at $SNR = 5\text{dB}$ validates the accuracy of the derived theoretical asymptotic distribution and illustrates the impact 3D elevation beamforming can have on the system performance through the meticulous selection of downtilt angles. For the case on hand, the MI of the system is again maximized when the serving BS also sets its antenna boresight angles equal to the elevation LoS angle with the MS.

VI. CONCLUSION

The prospect of enhancing system performance through elevation beamforming has stirred a growing interest among researchers in wireless communications. A large effort is currently being devoted to produce accurate 3D channel models that will be of fundamental practical use in analyzing these elevation beamforming techniques. In this work, we proposed a 3D channel model that is consistent with the state of available knowledge on channel parameters, and is inspired from system level stochastic channel models that have appeared in standards like 3GPP, ITU and WINNER. The principle of maximum entropy was used to determine the distribution of the MIMO channel matrix based on the prior information on angles of departure and arrival. The resulting 3D channel model is shown to have a systematic structure that can be exploited to effectively derive the closed-form expressions for the CDF of the MI. Specifically, we characterized the CDF in the general and low SINR regimes. An exact closed-form expression for the general SINR regime was derived for systems with a single receive antenna. An asymptotic

analysis, in the number of paths and BS antennas, of the achievable MI is also provided and validated for finite-sized systems via numerical results. The derived analytical expressions were shown to closely coincide with the simulated results for the proposed 3D channel model. An important observation made was that the meticulous selection of the downtilt angles at the transmitting BS can have a significant impact on the performance gains, confirming the potential of elevation beamforming. We believe that the results presented will enable a fair evaluation of the 3D channels being outlined in the future generation of mobile communication standards.

APPENDIX A

PROOF OF THEOREM 3

The mean and variance of every quadratic term $T^{k,l}$ can be computed as follows [43],

$$\mathbb{E}[T^{k,l}] = \frac{1}{N} \mathbb{E}[|\alpha|^2] \text{Tr}(\mathbf{C}^{k,l}). \quad (53)$$

$$\text{Var}[T^{k,l}] = \frac{1}{N^2} \mathbb{E}[|\alpha|^2]^2 [\text{Tr}(\mathbf{C}^{k,l} \mathbf{C}^{k,l}) + \text{Tr}(\mathbf{C}^{k,l} (\mathbf{C}^{k,l})^H)]. \quad (54)$$

Similarly, the covariance between two entries of $\mathbf{H}\mathbf{H}^H$ is,

$$\text{Cov}[T^{k,l} T^{k',l'}] = \frac{1}{N^2} \mathbb{E}[|\alpha|^2]^2 [\text{Tr}(\mathbf{C}^{k,l} \mathbf{C}^{k',l'}) + \text{Tr}(\mathbf{C}^{k,l} (\mathbf{C}^{k',l'})^H)]. \quad (55)$$

Given the expressions for mean and covariances of quadratic forms in Gaussian RVs, the Central Limit Theorem for the $(k, l)^{th}$ entry of $\mathbf{H}\mathbf{H}^H$ can be established from the result in [44] under assumptions **A-1** and **A-2**. Noting that $E[\alpha] = 0$ and $E[|\alpha|^2] = 1$, this convergence implies that the MGF of $T^{k,l}$ has the following behavior,

$$\mathbb{E} \left[\exp \left(s \frac{\sqrt{N} [T^{k,l} - \frac{1}{N} \text{Tr}(\mathbf{C}^{k,l})]}{\sqrt{\frac{\text{Tr}(\mathbf{C}^{k,l} \mathbf{C}^{k,l}) + \text{Tr}(\mathbf{C}^{k,l} (\mathbf{C}^{k,l})^H)}{N}}} \right) \right] - \exp \left(\frac{1}{2} s^2 \right) \rightarrow 0. \quad (56)$$

We stack the entries of $\Re(\mathbf{H}\mathbf{H}^H)$, $\Im(\mathbf{H}\mathbf{H}^H)$ and their means in \mathbf{x} and \mathbf{m} respectively as,

$$\begin{aligned} \mathbf{x} &= \frac{1}{N} [\Re(\alpha^H \mathbf{C}^{1,1} \alpha) \dots \Re(\alpha^H \mathbf{C}^{M,M} \alpha), \Im(\alpha^H \mathbf{C}^{1,1} \alpha) \dots \Im(\alpha^H \mathbf{C}^{M,M} \alpha)]^T, \\ \mathbf{m} &= \frac{1}{2N} [\text{Tr}(\mathbf{C}_{\Re}^{1,1}) \dots \text{Tr}(\mathbf{C}_{\Re}^{M,M}), \text{Tr}(\mathbf{C}_{\Im}^{1,1}) \dots \text{Tr}(\mathbf{C}_{\Im}^{M,M})]^T. \end{aligned} \quad (57)$$

From [44], each term in \mathbf{x} is asymptotically Normal. To characterize the joint distribution of the entries of $\Re(\mathbf{H}\mathbf{H}^H)$, $\Im(\mathbf{H}\mathbf{H}^H)$, the behaviour of their linear combination, i.e. $\mathbf{g}^T [\mathbf{x} - \mathbf{m}]$, where

\mathbf{g} is any arbitrary vector, is studied. The MGF of this linear combination is given by,

$$\mathbb{E} [\exp (\mathbf{g}^T [\mathbf{x} - \mathbf{m}])] = \mathbb{E} \left[\exp \left(\frac{1}{N} \left[\Re(\boldsymbol{\alpha})^T \quad \Im(\boldsymbol{\alpha})^T \right] \boldsymbol{\Xi} \begin{bmatrix} \Re(\boldsymbol{\alpha}) \\ \Im(\boldsymbol{\alpha}) \end{bmatrix} - \mathbf{g}^T \mathbf{m} \right) \right], \quad (58)$$

where $\boldsymbol{\Xi} = [g_1 \mathbf{C}_{\Re}^{1,1} + \dots + g_{M^2} \mathbf{C}_{\Re}^{M,M} + g_{M^2+1} \mathbf{C}_{\Im}^{1,1} + \dots + g_{2M^2} \mathbf{C}_{\Im}^{M,M}]$. Note that $\begin{bmatrix} \Re(\boldsymbol{\alpha})^T & \Im(\boldsymbol{\alpha})^T \end{bmatrix} \boldsymbol{\Xi} \begin{bmatrix} \Re(\boldsymbol{\alpha}) \\ \Im(\boldsymbol{\alpha}) \end{bmatrix}$ is also a quadratic form in Normal RVs ($\Re(\boldsymbol{\alpha})$, $\Im(\boldsymbol{\alpha})$). Extending the convergence result in (56) for a quadratic form in Normal RVs with $s = 1$, $\mathbb{E}[|\Re(\alpha)|^2] = \mathbb{E}[|\Im(\alpha)|^2] = \frac{1}{2}$, we have

$$\mathbb{E} \left[\exp \left(\frac{\sqrt{N} \mathbf{g}^T [\mathbf{x} - \mathbf{m}]}{\sqrt{\frac{1}{4N} [Tr(\boldsymbol{\Xi} \boldsymbol{\Xi}) + Tr(\boldsymbol{\Xi}(\boldsymbol{\Xi})^H)]}} \right) \right] - \exp \left(\frac{1}{2} \right) \rightarrow 0, \quad (59)$$

where,

$$\frac{1}{4N} [Tr(\boldsymbol{\Xi} \boldsymbol{\Xi}) + Tr(\boldsymbol{\Xi}(\boldsymbol{\Xi})^H)] = \mathbf{g}^T \boldsymbol{\Theta} \mathbf{g}, \quad (60)$$

and $\boldsymbol{\Theta}$ is the $2M^2 \times 2M^2$ covariance matrix of the entries of vector $\sqrt{N} \mathbf{x}$ given by,

$$\boldsymbol{\Theta} = \begin{bmatrix} \frac{1}{4N} [Tr(\mathbf{C}_{\Re}^{1,1}(\mathbf{C}_{\Re}^{1,1})) + Tr(\mathbf{C}_{\Re}^{1,1}(\mathbf{C}_{\Re}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\Re}^{1,1}(\mathbf{C}_{\Im}^{M,M})) + Tr(\mathbf{C}_{\Re}^{1,1}(\mathbf{C}_{\Im}^{M,M})^H)] \\ \frac{1}{4N} [Tr(\mathbf{C}_{\Re}^{1,2}(\mathbf{C}_{\Re}^{1,1})) + Tr(\mathbf{C}_{\Re}^{1,2}(\mathbf{C}_{\Re}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\Re}^{1,2}(\mathbf{C}_{\Im}^{M,M})) + Tr(\mathbf{C}_{\Re}^{1,2}(\mathbf{C}_{\Im}^{M,M})^H)] \\ \vdots & \ddots & \vdots \\ \frac{1}{4N} [Tr(\mathbf{C}_{\Re}^{M,M}(\mathbf{C}_{\Re}^{1,1})) + Tr(\mathbf{C}_{\Re}^{M,M}(\mathbf{C}_{\Re}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\Re}^{M,M}(\mathbf{C}_{\Im}^{M,M})) + Tr(\mathbf{C}_{\Re}^{M,M}(\mathbf{C}_{\Im}^{M,M})^H)] \\ \frac{1}{4N} [Tr(\mathbf{C}_{\Im}^{1,1}(\mathbf{C}_{\Re}^{1,1})) + Tr(\mathbf{C}_{\Im}^{1,1}(\mathbf{C}_{\Re}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\Im}^{1,1}(\mathbf{C}_{\Im}^{M,M})) + Tr(\mathbf{C}_{\Im}^{1,1}(\mathbf{C}_{\Im}^{M,M})^H)] \\ \frac{1}{4N} [Tr(\mathbf{C}_{\Im}^{1,2}(\mathbf{C}_{\Re}^{1,1})) + Tr(\mathbf{C}_{\Im}^{1,2}(\mathbf{C}_{\Re}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\Im}^{1,2}(\mathbf{C}_{\Im}^{M,M})) + Tr(\mathbf{C}_{\Im}^{1,2}(\mathbf{C}_{\Im}^{M,M})^H)] \\ \vdots & \ddots & \vdots \\ \frac{1}{4N} [Tr(\mathbf{C}_{\Im}^{M,M}(\mathbf{C}_{\Re}^{1,1})) + Tr(\mathbf{C}_{\Im}^{M,M}(\mathbf{C}_{\Re}^{1,1})^H)] & \dots & \frac{1}{4N} [Tr(\mathbf{C}_{\Im}^{M,M}(\mathbf{C}_{\Im}^{M,M})) + Tr(\mathbf{C}_{\Im}^{M,M}(\mathbf{C}_{\Im}^{M,M})^H)] \end{bmatrix} \quad (61)$$

The entries of $\boldsymbol{\Theta}$ are obtained using (55). Therefore,

$$\mathbb{E} \left[\exp \left(\frac{\sqrt{N} (\mathbf{g}^T \mathbf{x} - \mathbf{g}^T \mathbf{m})}{\sqrt{\mathbf{g}^T \boldsymbol{\Theta} \mathbf{g}}} \right) \right] - \exp \left(\frac{1}{2} \right) \rightarrow 0. \quad (62)$$

Since $\mathbf{g}^T \mathbf{x}$ behaves as a Gaussian RV with mean $\mathbf{g}^T \mathbf{m}$ and variance given by $\frac{1}{N} \mathbf{g}^T \boldsymbol{\Theta} \mathbf{g}$, so it can be deduced that \mathbf{x} behaves as multivariate Gaussian distribution with mean \mathbf{m} and covariance $\frac{1}{N} \boldsymbol{\Theta}$. Therefore under assumptions **A-1** and **A-2**, the joint distribution of the entries of $\sqrt{N} ([\text{vec}(\Re(\mathbf{H}\mathbf{H}^H))^T, \text{vec}(\Im(\mathbf{H}\mathbf{H}^H))^T]^T - \mathbf{m})$ behaves approximately as multivariate Gaussian with mean 0 and covariance matrix given by $\boldsymbol{\Theta}$. This completes the proof of Theorem 3.

APPENDIX B

PROOF OF THEOREM 4

To complete the characterization of MI, the log det transformation needs to be applied on the entries of \mathbf{Y} and \mathbf{Z} , i.e. $\Re((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H)$ and $\Im((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H)$ respectively. These entries can be stacked in vector \mathbf{x}_n as,

$$\begin{aligned} \mathbf{x}_n = & [\Re(\zeta_{11}) + \frac{1}{N} \Re(\boldsymbol{\alpha}^H \mathbf{C}^{1,1} \boldsymbol{\alpha}) \quad \dots \quad \Re(\zeta_{MM}) + \frac{1}{N} \Re(\boldsymbol{\alpha}^H \mathbf{C}^{M,M} \boldsymbol{\alpha}), \\ & \Im(\zeta_{11}) + \frac{1}{N} \Im(\boldsymbol{\alpha}^H \mathbf{C}^{1,1} \boldsymbol{\alpha}) \quad \dots \quad \Im(\zeta_{MM}) + \frac{1}{N} \Im(\boldsymbol{\alpha}^H \mathbf{C}^{M,M} \boldsymbol{\alpha})]^T, \end{aligned} \quad (63)$$

To obtain the statistical distribution of MI from the joint distribution of the entries of \mathbf{x}_n , we succumb to the Taylor series expansion of a real-valued function $f(\mathbf{x}_n)$,

$$f(\mathbf{x}_n) = f(\mathbf{m}) + \nabla^T f(\mathbf{m})(\mathbf{x}_n - \mathbf{m}) + R_n, \quad (64)$$

where \mathbf{m} is the mean vector in (39) and $R_n = o(\|\mathbf{x}_n - \mathbf{m}\|^2)$. We observe that $o(\|\mathbf{x}_n - \mathbf{m}\|^2)$ is $o(1)$ because $\mathbf{x}_n - \mathbf{m} \xrightarrow{p} 0$. Any mapping of $\mathbf{x}_n - \mathbf{m}$ will therefore be $o(1)$ which implies that $R_n = o(1)$.

$$\mathbb{E}[\exp(f(\mathbf{x}_n) - f(\mathbf{m}))] = \mathbb{E}[\exp(\nabla^T f(\mathbf{m})(\mathbf{x}_n - \mathbf{m}))] + o(1). \quad (65)$$

Note that $\mathbb{E}[\exp(\nabla^T f(\mathbf{m})(\mathbf{x}_n - \mathbf{m}))]$ is analogous to $\mathbb{E}[\exp(\mathbf{g}^T[\mathbf{x}_n - \mathbf{m}])]$ with $\mathbf{g} = \nabla f(\mathbf{m})$. Now using the convergence result in (59) with (60) and invoking Slutsky's Theorem, given that $\liminf \nabla^T f(\mathbf{m}) \Theta \nabla f(\mathbf{m}) > 0$ results in,

$$\mathbb{E} \left[\exp \left(\frac{\sqrt{N}[f(\mathbf{x}_n) - f(\mathbf{m})]}{\sqrt{\nabla^T f(\mathbf{m}) \Theta \nabla f(\mathbf{m})}} \right) \right] - \exp \left(\frac{1}{2} \right) \rightarrow 0. \quad (66)$$

This is analogous to applying Delta Theorem on the entries of $\mathbf{Y} + j\mathbf{Z}$. Therefore for $f(\mathbf{Y} + j\mathbf{Z}) = \log \det(\mathbf{Y} + j\mathbf{Z})$, we can say that $\sqrt{N}(\log \det(\mathbf{Y} + j\mathbf{Z}) - \log \det(\mathbf{M}_1 + j\mathbf{M}_2))$ behaves as a Gaussian RV under assumptions **A-1** and **A-2**, with mean 0, and variance given by $(\nabla^T f(\mathbf{M}) \Theta \nabla f(\mathbf{M}))$.

Define a function $h(\mathbf{x}_n) = f_2 \circ f_1(\mathbf{x}_n)$, where $f_1(\mathbf{x}_n) = (\det(\mathbf{Y} + j\mathbf{Z}))^2$ and $f_2(x) = 0.5 \log x$. For this function the gradient is,

$$\nabla h|_{\mathbf{x}_n = \mathbf{m}} = f_2'(\det(\mathbf{M}_1 + j\mathbf{M}_2)^2) \times \nabla f_1|_{\mathbf{Y} = \mathbf{M}_1, \mathbf{Z} = \mathbf{M}_2}. \quad (67)$$

We use the following to determine the expression for ∇f_1 ,

$$(\det(\mathbf{Y} + j\mathbf{Z}))^2 = \det \left(\begin{bmatrix} \mathbf{Y} & -\mathbf{Z} \\ \mathbf{Z} & \mathbf{Y} \end{bmatrix} \right). \quad (68)$$

Therefore the $(l + (k - 1)M)^{th}$, $k = 1, \dots, 2M, l = 1, \dots, M$ entry of the vector ∇f_1 is given by [Ch 6, [45]],

$$\nabla f_1 \left(\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix} \right) \Big|_{l+(k-1)M} = [\det(\mathbf{D}_1) + \dots + \det(\mathbf{D}_{2M})]_{l+(k-1)M}, \quad (69)$$

where \mathbf{D}_i is identical to $\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix}$ except that the entries in the i^{th} row are replaced by their

derivatives with respect to $(k, l)^{th}$ entry of $\begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{bmatrix}$, $k = 1, \dots, 2M, l = 1, \dots, M$. Every entry of

$\nabla f_1 \left(\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix} \right)$ will involve the sum of only two non-zero determinants as every $(k, l)^{th}$ entry occurs in only two rows and,

$$f'_2(\det(\mathbf{M}_1 + j\mathbf{M}_2)^2) = \frac{.5}{\det \left(\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix} \right)}. \quad (70)$$

Therefore from (66), in the asymptotic limit, the distribution of $\sqrt{N} \left(\log \det((\mathbf{R} + \sigma^2 \mathbf{I}_M) + \mathbf{H}\mathbf{H}^H) - \frac{1}{2} \log \det \left(\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix} \right) \right)$ can be approximated by a Gaussian distribution with mean 0 and variance given by,

$$\left(\frac{.5}{\det \left(\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix} \right)} \right)^2 \times \nabla^T f_1 \left(\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix} \right) \Theta \nabla f_1 \left(\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix} \right), \quad (71)$$

under the assumption that $\liminf \nabla^T f_1 \left(\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix} \right) \Theta \nabla f_1 \left(\begin{bmatrix} \mathbf{M}_1 & -\mathbf{M}_2 \\ \mathbf{M}_2 & \mathbf{M}_1 \end{bmatrix} \right) > 0$.

Using this result together with (37), we can establish the statistical distribution of MI in the asymptotic limit for 3D MIMO channels. This completes the proof of Theorem 4.

REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.
- [3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [4] A. Kuchar, J.-P. Rossi, and E. Bonek, "Directional macro-cell channel characterization from urban measurements," *IEEE Transactions on Antennas and Propagation*, vol. 48, no. 2, pp. 137–146, Feb. 2000.
- [5] J. Fuhl, J.-P. Rossi, and E. Bonek, "High-resolution 3-d direction-of-arrival determination for urban mobile radio," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 4, pp. 672–682, Apr. 1997.
- [6] J. Koppenborg, H. Halbauer, S. Saur, and C. Hoek, "3D beamforming trials with an active antenna array," in *Proc. ITG Workshop Smart Antennas*, pp. 110–114, 2012.
- [7] Q. Nadeem, A. Kammoun, M. Debbah, and M.-S. Alouini, "A generalized spatial correlation model for 3d mimo channels based on the fourier coefficients of power spectrums," *Submitted to IEEE transactions on Signal Processing*, 2014, [Online.] Available: <http://www.laneas.com/publication/generalized-spatial-correlation-model-3d-mimo-channels-based-fourier-coefficients-power>.
- [8] 3GPP TR 36.873 V12.0.0, "Study on 3D channel model for LTE," Sep. 2014.
- [9] Y. Zhu, P. Y. Kam, and Y. Xin, "On the mutual information distribution of MIMO Rician fading channels," *IEEE Transactions on Communications*, vol. 57, no. 5, pp. 1453–1462, May 2009.
- [10] G. Alfano, A. Lozano, A. M. Tulino, and S. Verdú, "Mutual information and eigenvalue distribution of MIMO Ricean channels," in *International Symposium on Information Theory and its Applications (ISITA) 2004*, Oct 2004.
- [11] J. Salo, F. Mikas, and P. Vainikainen, "An upper bound on the ergodic mutual information in Rician fading MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, pp. 1415–1421, June 2006.
- [12] P. Smith and L. Garth, "Exact capacity distribution for dual MIMO systems in Ricean fading," *IEEE Communications Letters*, vol. 8, no. 1, pp. 18–20, Jan 2004.
- [13] P. Smith, L. Garth, and S. Loyka, "Exact capacity distributions for MIMO systems with small numbers of antennas," *IEEE Communications Letters*, vol. 7, no. 10, pp. 481–483, Oct 2003.
- [14] M. McKay and I. Collings, "General capacity bounds for spatially correlated Rician MIMO channels," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3121–3145, Sept 2005.
- [15] P. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *IEEE International Conference on Communications, ICC 2002.*, vol. 1, 2002, pp. 406–410.
- [16] Z. Wang and G. Giannakis, "Outage mutual information of space-time MIMO channels," *IEEE Transactions on Information Theory*, vol. 50, no. 4, pp. 657–662, April 2004.
- [17] M. Chiani, M. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2363–2371, Oct 2003.
- [18] M. Kang and M.-S. Alouini, "Capacity of MIMO Rician channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 112–122, Jan 2006.
- [19] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 3987–4004, 2008.

- 1
2
3
4 [20] E. T. Jaynes, "Information theory and statistical mechanics, part 1," *Phys. Rev.*, vol. 106, pp. 620–630, 1957.
- 5 [21] —, "Information theory and statistical mechanics, part 2," *Phys. Rev.*, vol. 108, pp. 171–190, 1957.
- 6 [22] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum
7 cross-entropy," *IEEE Transactions on Information Theory*, vol. 26, no. 1, pp. 26–37, Jan 1980.
- 8 [23] M. Debbah and R. R. Müller, "MIMO channel modeling and the principle of maximum entropy," *IEEE Transactions on
9 Information Theory*, vol. 51, no. 5, pp. 1667–1690, May 2005.
- 10 [24] J. Meinilä, P. Kyösti *et al.*, "D5.3: WINNER+ Final Channel Models V1.0," [Online.] Available: [http://projects.
11 celtic-initiative.org/winner+/WINNER+%20Deliverables/D5.3_v1.0.pdf](http://projects.celtic-initiative.org/winner+/WINNER+%20Deliverables/D5.3_v1.0.pdf), June. 2010.
- 12 [25] F. Athley and M. Johansson, "Impact of electrical and mechanical antenna tilt on LTE downlink system performance," in
13 *IEEE 71st Vehicular Technology Conference (VTC 2010-Spring)*, May 2010, pp. 1–5.
- 14 [26] B. Partov, D. Leith, and R. Razavi, "Utility fair optimization of antenna tilt angles in LTE networks," *IEEE/ACM
15 Transactions on Networking*, vol. PP, no. 99, pp. 1–1, Jan 2014.
- 16 [27] O. N. C. Yilmaz, S. Hämäläinen, and J. Hämäläinen, "Analysis of antenna parameter optimization space for 3GPP LTE,"
17 in *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*, Sept 2009, pp. 1–5.
- 18 [28] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing - the large-scale array regime,"
19 *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, Oct 2013.
- 20 [29] "Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations," [Online]. Available: [http://www.3gpp.
21 org/ftp/Specs/html-info/25996.htm](http://www.3gpp.org/ftp/Specs/html-info/25996.htm), Sep. 2003.
- 22 [30] Report ITU-R M.2135, "Guidelines for evaluation of radio interface technologies for IMT-advanced," [Online]. Available:
23 <http://www.itu.int/pub/R-REP-M.2135-2008/en>, 2008.
- 24 [31] IST-4-027756 WINNER II, "D1.1.2, WINNER II Channel Models," [Online]. Available: [https://www.ist-winner.org/
25 WINNER2-Deliverables/D1.1.2v1.1.pdf](https://www.ist-winner.org/WINNER2-Deliverables/D1.1.2v1.1.pdf), Sept. 2007.
- 26 [32] R1-122034, "Study on 3D channel Model for elevation beamforming and FD-MIMO studies for LTE," 3GPP TSG RAN
27 no. 58, Dec. 2012.
- 28 [33] A. Kammoun, H. Khanfir, Z. Altman, M. Debbah, and M. Kamoun, "Preliminary results on 3D channel modeling: From
29 theory to standardization," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1219–1229, June. 2014.
- 30 [34] J. Zhang, C. Pan, F. Pei, G. Liu, and X. Cheng, "Three-dimensional fading channel models: A survey of elevation angle
31 research," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 218–226, June. 2014.
- 32 [35] K. Kalliola, K. Sulonen, H. Laitinen, O. Kivekas, J. Krogerus, and P. Vainikainen, "Angular power distribution and mean
33 effective gain of mobile antenna in different propagation environments," *IEEE Transactions on Vehicular Technology*,
34 vol. 51, no. 5, pp. 823–838, Sep. 2002.
- 35 [36] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli Esposti, H. Hofstetter, P. Kyösti, D. Lauren-
36 son, G. Matz, A. F. Molisch, C. Oestges, and H. Özcelik, "Survey of channel and radio propagation models for
37 wireless MIMO systems," *EURASIP Journal on Wireless Communications and Networking Volume 2007*, *url* =
38 <http://www.eurecom.fr/publication/2162>, 02 2007.
- 39 [37] U. Fernández-Plazaola, E. Martos-Naya, J. F. Paris, and J. T. Entrambasaguas, "Comments on Proakis analysis of the
40 characteristic function of complex Gaussian quadratic forms," *CoRR*, vol. abs/1212.0382, 2012. [Online]. Available:
41 <http://arxiv.org/abs/1212.0382>
- 42 [38] G. L. Turin, "The characteristic function of hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47,
43 no. 1/2, pp. pp. 199–201, 1960. [Online]. Available: <http://www.jstor.org/stable/2332977>
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 [39] M. Bibinger, "Notes on the sum and maximum of independent exponentially distributed random variables with
5 different scale parameters," Humboldt-Universität zu Berlin, Tech. Rep., 2013. [Online]. Available: [url{http:
6 //arxiv.org/abs/1307.3945}](http://arxiv.org/abs/1307.3945)
7
- 8 [40] X. Wu and R. Srikant, "MIMO channels in the low-SNR Regime: communication rate, error exponent, and signal peakiness,"
9 *IEEE Transactions on Information Theory*, vol. 53, no. 4, pp. 1290–1309, April 2007.
- 10 [41] W. J. L. Queiroz, F. Madeiro, W. T. A. Lopes, and M. S. Alencar, "Spatial correlation for DoA characterization using Von
11 Mises, Cosine, and Gaussian distributions," *International Journal of Antennas and Propagation*, vol. 2011, July. 2011.
- 12 [42] J. Ren and R. Vaughan, "Spaced antenna design in directional scenarios using the von mises distribution," in *Vehicular
13 Technology Conference Fall (VTC 2009-Fall)*, Sept. 2009, pp. 1–5.
- 14 [43] W. Hachem, M. Kharouf, J. Najim, and J. Silverstein, "A CLT for information-theoretic statistics of non-centered gram
15 random matrices," *Random Matrices: Theory and Applications*, 01 (02), April 2012.
- 16 [44] R. Bhansali, L. Giraitis, and P. Kokoszka, "Convergence of quadratic forms with nonvanishing diagonal," *Statistics
17 & Probability Letters*, vol. 77, no. 7, pp. 726–734, April 2007. [Online]. Available: [http://ideas.repec.org/a/eee/stapro/
18 v77y2007i7p726-734.html](http://ideas.repec.org/a/eee/stapro/v77y2007i7p726-734.html)
19
- 20 [45] C. D. Meyer, *Matrix analysis and applied linear algebra*. Philadelphia, PA, USA: Society for Industrial and Applied
21 Mathematics, 2000.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60