1

A Comparison of MIMO Techniques in Downlink Millimeter Wave Cellular Networks with Hybrid Beamforming

Mandar N. Kulkarni, Amitava Ghosh and Jeffrey G. Andrews

Abstract—Large antenna arrays will be needed in future millimeter wave (mmWave) cellular networks, enabling a large number of different possible antenna architectures and multiple-input multiple-output (MIMO) techniques. It is still unclear which MIMO technique is most desirable as a function of different network parameters. This paper, therefore, compares the coverage and rate performance of hybrid beamforming enabled multi-user (MU) MIMO and single-user spatial multiplexing (SM) with single-user analog beamforming (SU-BF). A stochastic geometry model for coverage and rate analysis is proposed for MU-MIMO mmWave cellular networks, taking into account important mmWave-specific hardware constraints for hybrid analog/digital precoders and combiners, and a blockagedependent channel model which is sparse in angular domain. The analytical results highlight the coverage, rate and power consumption tradeoffs in multiuser mmWave networks. With perfect channel state information at the transmitter and round robin scheduling, MU-MIMO is usually a better choice than SM or SU-BF in mmWave cellular networks. This observation, however, neglects any overhead due to channel acquisition or computational complexity. Incorporating the impact of such overheads, our results can be re-interpreted so as to quantify the minimum allowable efficiency of MU-MIMO to provide higher rates than SM or SU-BF.

I. INTRODUCTION

A classical question in multi-antenna wireless communications has been to determine which MIMO technique performs better in different scenarios, for example based on the channel and interference characteristics. At mmWave frequencies, several important new factors must be considered, due to different hardware constraints on the precoders/combiners and a significantly different outdoor channel, which is both blockage-dependent and sparse (low rank) [2]–[4]. In order to compensate for the large near-field path loss, SU-BF has been the primary focus of several existing system capacity evaluations for mmWave cellular networks [2], [5]–[7]. However, recently there has been significant work on enabling MU-MIMO and SM under different antenna architectures that respect the necessary hardware constraints at mmWave frequencies. Hybrid analog-digital precoders and combiners, receivers with low resolution analog to digital converters, and continuous aperture phase MIMO with lens-based beamformers (also called CAP-MIMO) are prominent antenna architectures being considered [8]-[10]. Most existing studies on mmWave MIMO, except for SU-BF, rely on single cell analysis for evaluating performance of the MIMO techniques and/or system level simulations for understanding the impact of base station (BS) deployment scenarios or blockages in the environment on the coverage and rate performance. Although analytical models for studying coverage and rate in SU-BF mmWave networks have been studied [6], [11], these cannot be directly used for studying other MIMO techniques like MU-MIMO and SM, as will be explained in Section I-A.

The goals of this paper are two-fold. First, we propose a stochastic geometry-based model to study coverage and per user rate distribution in fully-connected hybrid beamforming-enabled MU-MIMO mmWave cellular networks. Second, we use this analytical model as a tool for comparing coverage, rate and power consumption for MU-MIMO, SM and SU-BF mmWave cellular networks.

A. Background and Related Work

Conventionally, BSs are equipped with fully-digital baseband processing. However, this approach requires a radio frequency (RF) chain per antenna which is impractical for mmWave BSs equipped with large antenna arrays. Fully analog solutions, on the other hand, require only a single RF chain for the entire antenna array but have no capability of digital processing. Hybrid beamforming strikes a balance between these two solutions, wherein the number of RF chains can be designed to be between 1 (analog beamforming) and the number of antennas (digital beamforming). In a fully-connected architecture, each RF chain has phase shifters connected to all antennas in the array. On the other hand, in

This work was supported by Nokia. Part of this work appeared in Proc. *IEEE Asilomar*, Nov. 2015 [1]. M. N. Kulkarni and J. G. Andrews are with the University of Texas at Austin, TX. A. Ghosh is with Nokia Networks, Arlington Heights, IL. Email: mandar.kulkarni@utexas.edu, amitava.ghosh@nokia.com,jandrews@ece.utexas.edu

the array of sub-arrays architecture, the entire array is divided into sub-arrays and all antennas in a subarray are connected via phase shifters to exactly one RF chain. The fully-connected architecture has higher beamforming gain than array of sub-arrays, for a fixed number of antennas. However, the power consumption and hardware complexity of precoder/combiner is lower in the latter. With low-complexity yet near optimal precoding/combining algorithms for MU-MIMO and SM being proposed with the fully-connected architecture [4], [12], this approach looks promising for practical implementation and is the focus of our discussion.

In [12], a joint baseband-RF precoder solution for MU-MIMO was proposed and proven to be asymptotically optimal as the number of antennas become large. Using this scheme, it was observed that MU-MIMO can offer higher sum rates than SU-BF. Another simulation-based work [13] highlighted that per user rates, including the cell edge rates, can be much higher with MU-MIMO with appropriate user pairing. It was observed that exploiting polarization diversity for two stream transmission to each user further enhances the gains in using MU-MIMO. This is one particular way in which SM gains can be obtained in tandem with MU-MIMO. Another way to get SM gains would be to rely on the scatterers in the environment [3], [4]. The simulations in [3], [4] showed that SM and SU-BF could work in tandem to improve capacity. However, all these works implicitly neglected the aspect of power consumption at BSs and UEs when comparing the MIMO techniques. If we were to compare the coverage and rate performance of SU-BF and MU-MIMO or SM with fixed power consumption per unit area and fixed number of antennas per BS, we can deploy a much denser mmWave network with SU-BF than MU-MIMO or SM. This significantly affects the comparisons as will be shown in Section V, since unlike in conventional cellular networks [14], densifying a mmWave network boosts the coverage and capacity notably [6], [7].

The above mentioned studies either rely on systemlevel simulations or on single cell analysis. There is no analytical model for MU-MIMO or SM mmWave networks that incorporates the impact of hybrid precoders and combiners and the channel sparsity. Analysis for MIMO cellular networks has conventionally been done by capturing the impact of linear precoding and combining into the distribution of an effective small scale fading random variable [15]–[18]. In [15], [16], it was shown that Gamma distribution can be used to model the small scale fading gain in MU-MIMO cellular networks employing ZF precoding. Most successive analytical studies on MU-MIMO cellular networks using stochastic geometry have relied on this result, for example [17], [18]. However, justifying this result assumes fully digital processing and full rank Rayleigh fading channels. At mmWave frequencies the channel is expected to be sparse and blockage-dependent [2], [3], [19], [20]. Thus, the full rank assumption is far from reality. A recent work [21] proposes an analytical model for SINR (signal to interference plus noise ratio) in MU-MIMO mmWave cellular networks but assumes fully digital processing. But, as described earlier, fully digital processing is also not realistic at mmWave. Analysis of multiuser mmWave cellular networks, thus, demands a new approach. Also, other existing analytical models for SU-BF enabled mmWave networks assume an equivalent SISO-like system with directional antenna gains by abstracting underlying signal level details [6], [11]. Further, the analysis in these papers is done for single path channels. An analytical framework that can be used as a tool for comparing with different MIMO techniques needs to incorporate multipath in the channel, which is a primary feature enabling SM. The key contributions in this work are as follows.

B. Contributions

1) Tractable Model for Coverage and Rate in MU-MIMO mmWave Cellular Networks. The analytical model captures the following mmWave-specific features: (i) precoding and combining with hybrid beamforming, and (ii) sparse blockage-dependent multipath channel model. For simplicity the channel model is assumed to be non-selective in both time and frequency to focus only on the spatial aspects. Using Monte-Carlo simulations, the model is shown to be reasonably accurate for a large number of antennas at the BSs and user equipments (UEs) in noise-limited scenarios. In interference-limited scenarios, upper and lower bounds to the distribution of the proposed approximate SINR model are derived under some assumptions and validated with Monte-Carlo simulations. The fact that our model incorporates different channel rank for line-of-sight (LOS) and non-LOS (NLOS) makes it possible to fairly compare analytical results with Monte-Carlo simulations for SM, which strongly depend on the rank of the channel. Numerical results reveal the following insights: (i) In interference-limited scenarios, SINR coverage has a nonmonotonic trend with BS density. The optimum BS density for SINR coverage decreases with increasing degree of multiuser transmission. (ii) Although SINR coverage decreases with MU-MIMO, the median and peak per user rate increases due to increasing number of time slots available per user. However, the cell edge rates suffer with round robin scheduling, which motivates that the scheduler must explicitly safeguard the rates of edge users to use MU-MIMO.

2) Comparison of MIMO Techniques Considering Coverage, Rate and Power Consumption Tradeoffs. With perfect channel state information at the transmitter and neglecting channel acquisition and computational complexity overheads, MU-MIMO usually provides higher per user throughput compared to SM and SU-BF in mmWave networks for a fixed density of BSs and fixed number of antennas per BS/UE. Further note that enabling MU-MIMO requires only single RF chain at UEs, whereas enabling SM requires some baseband combining at UEs with multiple RF chains. We provide a stochastic ordering argument which highlights that SNR coverage normalized by the antenna gains is better for MU-MIMO than SM asymptotically with the number of antennas at the BSs and users. SM can outperform MU-MIMO in scenarios when SM can support more streams than the number of users that can be served with MU-MIMO. This boils down to having sufficiently low user density coupled with sufficiently large number of RF chains at UEs/BSs and multipath in the channel. Instead of fixing the density of BSs if power consumption per unit area is fixed, a denser SU-BF network outperforms MU-MIMO and SM in terms of per user cell edge rates. However, the sum rate with MU-MIMO is still usually better than SU-BF and SM. The above results on sum or per user rates neglect the possibly increased overheads with MU-MIMO due to channel acquisition or computational complexity. Incorporating such factors, our results can be re-interpreted so as to quantify the minimum allowable efficiency for MU-MIMO to provide higher data rates than SM or SU-BF. The definition of minimum allowable efficiency is formally given in Section III-B.

C. Organization and Notation

Section II sets up the system model. The analytical model for coverage and rate in MU-MIMO mmWave networks is developed in Section III. Heuristic comparison of coverage and rate with SM is discussed in Section IV. Section V and VI discusses the numerical results and conclusions.¹.

II. SYSTEM MODEL

Consider a downlink mmWave cellular network operating at carrier frequency f_c and with bandwidth B. It is assumed that BSs and UEs are distributed in \mathbb{R}^2 as independent and homogeneous Poisson point processes (PPPs) Φ_{BS} and Φ_{UE} , with intensities λ_{BS} and λ_{UE} , respectively [14]. Each BS and user is assumed to employ a uniform linear array (ULA) of size N_{BS} and N_{UE}, respectively. Full buffer traffic is assumed in this work.

A. Propagation Model

Path loss from BS at $x \in \Phi_{BS}$ to a user at $u \in \Phi_{UE}$ is given in dB by

$$L(x, u) = \beta + 10\alpha \log_{10}(||x - u||) + S_{x, u}, \quad (1)$$

where $\beta = 20 \log_{10} \left(\frac{4\pi}{\lambda_c}\right)$ is the reference distance path loss at 1 meter, λ_c is the wavelength in meters, α is the path loss exponent, $S_{x,u} \sim N(0,\xi^2)$ denotes Gaussian distribution with zero mean and standard deviation ξ . Note that α and ξ are different for LOS and NLOS links. A subscript 'L' and 'N' to α and ξ denote the respective parameters for LOS and NLOS links, respectively. A probabilistic blockage model proposed and validated in [11], [22] is used in this work. According to this model, the probability that a link of length ||x - u|| is LOS is p_{LOS} if $||x - u|| \leq D$, for some value of D. All links longer than D are NLOS.

MmWave channels are expected to be sparse with very few angles of arrival (AOAs) and departure (AODs) capturing most of the energy [2], [3], [19], [20]. In this work, we assume a narrowband geometric channel model [4], [12], where the channel matrix between BS at x and user u is given by

$$\mathbf{H}_{x,u} = \sqrt{\frac{\mathbf{N}_{\mathrm{BS}}\mathbf{N}_{\mathrm{UE}}}{L(x,u)\eta_{x,u}}} \sum_{i=1}^{\eta_{x,u}} \gamma_{i,x,u} \mathbf{a}_{\mathrm{UE}}(\phi_{i,x,u}) \mathbf{a}_{\mathrm{BS}}^*(\theta_{i,x,u}).$$
(2)

Here, $\eta_{x,u}$ is the number of paths between BS at x and user at u, $\gamma_{i,x,u}$ is the small scale fading on i^{th} path (assumed to be complex normal with zero mean and unit variance for both LOS and NLOS to enhance analytical tractability), $\theta_{i,x,u}$ is the virtual AOD and $\phi_{i,x,u}$ is the virtual AOA for the i^{th} path. The number of paths $\eta_{x,u}$ equals $\eta_{\rm L}$ or $\eta_{\rm N}$ depending on whether the link is LOS or NLOS, respectively². It is expected that $\eta_{\rm N} > \eta_{\rm L}$ [3], [19], [20]. The virtual AOA or AOD are related to the corresponding physical angles as $\theta = 2\pi d \sin(\varphi) / \lambda_c$, where d is the inter-antenna spacing (chosen to be $\lambda_c/2$), φ is the physical angle and θ is the virtual angle. The array response vectors for ULAs, \mathbf{a}_{BS} and $\mathbf{a}_{\mathrm{UE}},$ are of the form $\mathbf{a}(\theta) = [1 \ e^{-j\theta} \dots e^{-j(N-1)\theta}]^* / \sqrt{N}$, where $N \in \{N_{BS}, N_{UE}\}$. We assume that for every BS-UE link, scatterers in environment are uniformly distributed in $[0, 2\pi]$ and thus, the physical angles are also uniformly distributed in $[0, 2\pi]$. We call this the "physical model", which will be the basis of our simulation results, whereas

¹Variables in italics are scalar random variables. Small and capital bold letters indicate vectors and matrices, respectively. An exception are random spatial locations in \mathbb{R}^2 , which are italicized small letters x, y, u, v or w. The complex conjugate transpose and pseudo inverse of **A** is **A**^{*} and **A**[†], respectively. Convergence in distribution is denoted by $\stackrel{d}{\rightarrow}$.

 $^{^{2}\}eta_{\rm L} > 1$ indicates more than 1 LOS *like* paths. In this work, we either have LOS or NLOS multipaths. A more general channel model would incorporate scenarios with 1 or more LOS like paths along with NLOS paths. However, an optimal power allocation would nearly allocate all power to LOS-like paths, thus, justifying our model. For simplicity, it is assumed that each scatterer gives rise to a single dominant path [12], [23], [24]. Extension to a clustered model [4], [23] is desirable in future.

for tractable analysis we leverage the virtual channel approximation [23] in Section III-B.

B. Fully Connected Hybrid Beamforming Architecture

A fully-connected two layer hybrid beamforming architecture with $N_{\rm RF}^{\rm BS}$ and $N_{\rm RF}^{\rm UE}$ RF chains at the BS and UE, respectively, is shown in Figure 1. A BS at xsends a total of N_s^{BS} streams of data, which may include data sent to multiple users in the network. The transmit signals first go through a $N_{RF}^{BS} \times N_{s}^{BS}$ baseband precoder matrix $\mathbf{F}_{x}^{\text{BB}} = [\mathbf{f}_{x,1}^{\text{BB}} \dots \mathbf{f}_{x,N_{\text{S}}}^{\text{BB}}]$ followed by a $N_{\text{BS}} \times N_{\text{RF}}^{\text{BS}}$ RF precoder $\mathbf{F}_{x}^{\text{RF}} = [\mathbf{f}_{x,1}^{\text{RF}} \dots \mathbf{f}_{x,N_{\text{RF}}}^{\text{RF}}]$. Note that the RF precoder is generally implemented using phase-shifters [4], [12], although there have been attempts trying to explore alternative methods [25]. Let us denote the RF combiner at user u by $\mathbf{W}_{u}^{\text{RF}}$ and the baseband combiner by $\mathbf{W}_{u}^{\text{BB}} = [\mathbf{w}_{u,1}^{\text{BB}}, \dots, \mathbf{w}_{u,N_{s}^{\text{UE}}}^{\text{BB}}]$. Note that SM, MU-MIMO and SU-BF can all be implemented with this generic architecture. The problem of jointly optimizing over \mathbf{F}_x^{RF} , \mathbf{F}_x^{BB} , \mathbf{W}_u^{RF} and \mathbf{W}_u^{BB} to maximize sum rate or per user rate for SM and MU-MIMO is still an open problem [4], [12]. In the following sections, we use recently proposed near optimal algorithms for designing of precoders and combiners in [4] and [12] to employ SM and MU-MIMO, respectively as baseline for our simulations and analysis.

III. MULTIUSER MIMO IN MMWAVE CELLULAR Networks

For MU-MIMO, it is assumed that each BS serves multiple users with a single stream per user. Thus, analog beamforming with a single RF chain suffices at each UE. Let \mathcal{U}_x be the set of all users in $\Phi_{\rm UE}$ which are scheduled by the BS at x in the same time slot, and the cardinality of \mathcal{U}_x be U_x . We assume $U_x = \min(U_M, N_x)$, where N_x is the total number of users connected to the BS and U_M is the maximum number of users that can be scheduled in a time slot. A more sophisticated algorithm for deciding how many and which users to schedule in a resource block may be implemented as in [13], [26] but we neglect this aspect here for tractability. We further assume $U_M = N_{RF}$, and that unless the load on the BS is less than the number of RF chains, U_M users are served in a time slot. Also, when $U_M > U_x$ only U_x RF chains are used for processing, which means that $\mathbf{F}_x^{\mathrm{BB}} = [\mathbf{f}_{x,1}^{\mathrm{BB}} \dots \mathbf{f}_{x,U_x}^{\mathrm{BB}}]$ is of dimension $U_x \times U_x$ and $\mathbf{F}_x^{\mathrm{RF}} = [\mathbf{f}_{x,1}^{\mathrm{RF}} \dots \mathbf{f}_{x,U_x}^{\mathrm{RF}}]$ is of dimension $N_{\mathrm{BS}} \times U_x$. Under the narrowband assumption, the received signal

Under the narrowband assumption, the received signal at user u from BS at x after passing through w_u , the RF combiner at the user, is given by

$$y_u = \overline{\mathbf{h}}_{x,u}^* \mathbf{f}_{x,u}^{\mathrm{BB}} s_u + \sum_{v \in \mathcal{U}_x, v \neq u} \overline{\mathbf{h}}_{x,u}^* \mathbf{f}_{x,v}^{\mathrm{BB}} s_v + \mathrm{OCI} + \mathrm{noise},$$

where $\overline{\mathbf{h}}_{x,u}^* = \mathbf{w}_u^* \mathbf{H}_{x,u} \mathbf{F}_x^{\mathrm{RF}}$ and OCI is the out-of-cell interference. Here, $s_{(.)}$ are the transmit symbols with energy P/U_x . Thus, the total transmit power of the BS is P. In this work, we will follow the precoding/combining algorithm in [12] and also assume an infinite resolution codebook at BSs and UEs for tractability. The first step is to design the RF precoders and combiners that maximize the received signal power on each of the BS-UE links. Thus, \mathbf{w}_u and $\mathbf{f}_{x,u}^{\mathrm{RF}}$ are designed such that $(\mathbf{w}_u, \mathbf{f}_{x,u}^{\mathrm{RF}}) =$ $\arg \max_{\mathbf{w}, \mathbf{f}} |\mathbf{w}^* \mathbf{H}_{x,u} \mathbf{f}|$.

Lemma 1 (from [24]). The left and right singular vectors corresponding to non-zero eigenvalues of $\mathbf{H}_{x,u}$ with $\eta_{x,u} \ll \min(N_{BS}, N_{UE})$ converge in chordal distance to $\mathbf{a}_{UE}(\phi_{i,x,u})$ and $\mathbf{a}_{BS}(\theta_{i,x,u})$, for $1 \leq i \leq \eta_{x,u}$. The corresponding singular values converge to $\frac{N_{BS}N_{UE}}{L(x,u)\eta_{x,u}}|\gamma_{i,x,u}|^2$.

This lemma indicates that for large number of antennas $\mathbf{w}_u = \mathbf{a}_{\mathrm{UE}}(\phi_{i_m,x,u})$ and $\mathbf{f}_{x,u}^{\mathrm{RF}} = \mathbf{a}_{\mathrm{BS}}(\theta_{i_m,x,u})$, where $i_m = \arg\max_i |\gamma_{i,x,u}|$. This observation will be crucial in developing a tractable model for coverage and rate. Next, the baseband precoder is designed such that the multiuser interference is cancelled. Using a zero forcing (ZF) baseband precoder, $\mathbf{F}_x^{\mathrm{BB}} = \overline{\mathbf{H}}_x^{\dagger} \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix whose entries are chosen such that $||\mathbf{F}_x^{\mathrm{RF}} \mathbf{f}_{x,u}^{\mathrm{BB}}|| = 1$. Here, $\overline{\mathbf{H}}_x = [\overline{\mathbf{h}}_{x,u_1} \dots \overline{\mathbf{h}}_{x,u_{U_x}}]^*$ with $\mathcal{U}_x = \{u_1, \dots, u_{U_x}\}$. Note that $\overline{\mathbf{H}}^{\dagger} = \overline{\mathbf{H}}^* \left(\overline{\mathbf{HH}}^*\right)^{-1}$, if $\overline{\mathbf{H}}$ is full rank.

A. SINR and Rate Model

The SINR of the user at $u \in \Phi_{\text{UE}}$ served by a BS at $x \in \Phi_{\text{BS}}$ connected to U_x total users is given by

$$\operatorname{SINR}_{x,u} = \frac{\frac{||\overline{\mathbf{h}}_{x,u}^* \mathbf{f}_{x,v}^{\operatorname{BB}}||^2}{U_x}}{\frac{\sigma_n^2}{\mathrm{P}} + \sum_{\substack{v \in \mathcal{U}_x \\ v \neq u}} \frac{||\overline{\mathbf{h}}_{x,u}^* \mathbf{f}_{x,v}^{\operatorname{BB}}||^2}{U_x} + \sum_{\substack{y \in \Phi_{\mathrm{BS}} \\ y \neq x}} \sum_{\substack{w \in \mathcal{U}_y \\ y \neq x}} \frac{||\overline{\mathbf{h}}_{y,u}^* \mathbf{f}_{y,w}^{\operatorname{BB}}||^2}{U_y}}{(3)}$$

The second term in the denominator is zero, owing to the ZF precoder and the fact that $\overline{\mathbf{H}}_x$ is almost surely full rank for independently distributed channel gains from BS at x to users in \mathcal{U}_x . The per user rate (in bits per second or bps) of user u served by BS at x is defined as

$$R_{x,u} = \omega_x \frac{\mathrm{B}\,U_x}{N_x} \log_2(1 + \mathrm{SINR}_{x,u}),\tag{4}$$

where $\omega_x < 1$ models the efficiency in implementing MU-MIMO in terms channel acquisition or computational complexity or cyclic prefix while implementing OFDM [21], [27]. The above model implies that each user gets U_x/N_x fraction of resources, which can be achieved using round robin scheduling. The sum rate is



Fig. 1: Fully-connected hybrid architecture at the BSs and UEs.

defined as

$$R_x = \omega_x \mathbf{B} \sum_{u \in \mathcal{U}_x} \log_2(1 + \mathtt{SINR}_{x,u}), \tag{5}$$

which is basically the total number of bits per second (bps) transmitted by the BS, whereas the per user rate is the rate achieved by a typical user in a scheduling cycle.

In general, the efficiency factors vary for different BSs and are dependent on U_M , N_{BS} , N_{UE} , η_N , η_L and OFDM cyclic prefix penalty. For simplicity we assume $\omega_x = \omega_{MU}$, $\forall x \in \Phi_{BS}$. One can interpret $\omega_{MU} = \min_x \omega_x$ to get a lower bound on the rate. Since we expect the overhead to increase with U_x , ω_{MU} corresponds to the efficiency of BSs serving U_M users. Note that ω_{MU} for $U_M = 1$ is the overhead for SU-BF.

B. Coverage and Rate Analysis

Consider a *typical* UE at origin, wherein the notion of typicality for stationary point process is defined through Palm probability [28], and it associates with the BS at x offering minimum path loss L(x, 0). We call this the *tagged* BS. We evaluate the SINR coverage defined as $\mathbb{P}(\text{SINR}_{x,0} > \tau)$, which is the SINR distribution of the *typical* user at origin. Rate coverage is similarly defined. The SINR expression in (3), although exact, is not tractable in terms of finding its distribution. We, thus, provide an accurate yet tractable approximation that captures the dependency of the several parameters in the following analysis.

Definition 1. A random variable Z_1 stochastically dominates another random variable Z_2 , if $\mathbb{P}(Z_1 > z) \ge \mathbb{P}(Z_2 > z)$ for all $z \in \mathbb{R}$. We denote this as $Z_1 \stackrel{\text{st}}{\ge} Z_2$.

1) Rate Distribution in a Noise-limited Network. We first focus on finding the rate distribution in a network with negligible interference effects. Throughout the discussion, the virtual angles of departure/arrival are quantized to take values in $\{\theta : \theta = -\pi + \frac{2\pi i}{N_a}, 1 \le i \le N_a\}$.

Lemma 2. If antenna spacing is half wavelength and the physical AOAs/AODs are uniformly distributed in 0 to 2π , the distribution of the quantized virtual angles is given by

$$q_{a,i} = \frac{\left(\sin^{-1}\left(-1 + \frac{2i+1}{N_a}\right) - \sin^{-1}\left(-1 + \frac{2i-1}{N_a}\right)\right)}{\pi},$$

for $a \in \{\text{UE, BS}\}$ and $i \in \{1, ..., N_a - 1\}$. Further, $q_{a,N_a} = 1 - \sum_{j=1}^{N_a - 1} q_{a,j}$.

Proof: Note that $\theta = \pi \sin(\varphi)$ for half wavelength antenna spacing. Thus, the required probability can be computed by using that φ is uniformly distributed in 0 to 2π .

Proposition 1. Let $U_x = \{u_1, \ldots, u_{U_x}\}$ be the users served by the BS at x. Assuming $\eta_N \ll \min(N_{BS}, N_{UE})$, $U_M \ll \min(N_{BS}, N_{UE})$ and a dense network deployment, SNR at user u_1 can be modelled as

$$\operatorname{SNR}_{x,u_1} \approx \frac{\mathrm{G}}{\eta_{x,u_1} U_x \sigma_n^2} |\gamma_{i_m,x,u_1}|^2 L(x,u_1)^{-1} p_{\mathrm{ZF}},$$
 (6)

where $G = PN_{BS}N_{UE}$, i_m is the index corresponding to arg max $|\gamma_{i,x,u_1}|$, p_{ZF} is a random variable that captures reduction in signal power due to the ZF penalty and has distribution that stochastically dominates p_{MU} , which is a Bernoulli random variable with success probability $\zeta(\eta_{x,u_1}, U_x)$, where

$$\zeta(\eta_{x,u_1}, U_x) = \sum_{j=1}^{N_{BS}} q_{BS,j} B_j(\eta_{x,u_1}, U_x) \left(p_{LOS} A_j(\eta_L) + (1 - p_{LOS}) A_j(\eta_N) \right)^{U_x - 1},$$

$$\begin{split} \mathbf{A}_{j}(\eta) &= \mathbf{C}(\eta) + (1 - q_{\mathrm{BS},j})^{\eta - 1} - (1 - q_{\mathrm{BS},j})^{\eta - 1} \mathbf{C}(\eta), \\ \mathbf{B}_{j}(\eta,, U_{x}) &= \mathbf{C}(\eta) (1 - q_{\mathrm{BS},j})^{U_{x} - 1} + \mathbf{D}_{j}(\eta, U_{x}) - \end{split}$$

$$D_{j}(\eta, U_{x})C(\eta), C(\eta) = \sum_{i=1}^{N_{UE}} q_{UE,i}(1-q_{UE,i})^{\eta-1}, \text{ and}$$
$$D_{j}(\eta, U_{x}) = \sum_{i_{1},...,i_{\eta-1}=1}^{N_{BS}-1} \prod_{n=1}^{\eta-1} l_{i_{n},j} \left(1 - \sum_{unique(i_{(.)})} l_{i_{n},j}\right)^{U_{x}-1}$$

where $l_{n,j} = \frac{q_{BS,n}}{1-q_{BS,j}}$ if n < j and $l_{n,j} = \frac{q_{BS,n+1}}{1-q_{BS,j}}$ if $n \ge j$ and $unique(i_{(.)})$ represents the unique values in the set $\{i_1, \ldots, i_{\eta-1}\}$.

Proof: Without loss of generality, $i_m = 1$. From Lemma 1, $\mathbf{w}_u = \mathbf{a}_{\mathrm{UE}}(\phi_{1,x,u})$ and $\mathbf{f}_{x,u}^{\mathrm{RF}} = \mathbf{a}_{\mathrm{BS}}(\theta_{1,x,u})$. Using the orthogonality of the array response vectors with quantized virtual angles, it can be shown that \mathbf{H}_{x} takes the form

$$\overline{\mathbf{H}}_{x} = \begin{bmatrix} \sqrt{\frac{\mathbf{N}_{\mathrm{BS}}\mathbf{N}_{\mathrm{UE}}}{L(x,u_{1})\eta_{x,u_{1}}}} \gamma_{1,x,u_{1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\tilde{P}}_{x} \end{bmatrix}, \qquad (7)$$

with probability at least $\zeta(\eta_{x,u_1}, U_x)$. See Appendix A for details. Note that here \mathbf{P}_x is a submatrix of $\overline{\mathbf{H}}_x$ of dimension $U_x - 1 \times U_x - 1$. In this case,

$$\overline{\mathbf{H}}_{x}^{\dagger} = \left[\begin{array}{cc} \sqrt{\frac{L(x,u_{1})\eta_{x,u_{1}}}{\mathrm{N}_{\mathrm{BS}}\mathrm{N}_{\mathrm{UE}}}} \gamma_{1,x,u_{1}}^{-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{P}}_{x}^{\dagger} \end{array} \right]$$

We know that $\mathbf{F}^{\mathrm{BB}}_x=\overline{\mathbf{H}}^\dagger_x\mathbf{\Lambda},$ for diagonal matrix $\mathbf{\Lambda}$ that helps satisfy the power constraints. Thus, the first column of the baseband precoder is of the form $\mathbf{f}_{r,u_1}^{\mathrm{BB}} =$ $[c \ 0 \dots 0]$, for some constant c such that $||\mathbf{F}_x^{\text{RF}} \mathbf{f}_{x,u_1}^{\text{BB}}|| = 1$. Thus, $\mathbf{f}_{x,u_1}^{\text{BB}} = [1 \ 0 \dots 0]$ since each term in \mathbf{F}_x^{RF} is unit norm. In this case, the unit norm. In this case, the received signal power of u_1 is equal to $\frac{G}{\eta_{x,0}U_x}|\gamma_{1,x,u_1}|^2 L(x,u_1)^{-1}$, which corresponds to the case when $p_{\rm MU} = 1$ in (6). Since the event that \mathbf{P}_x is not of this form is of low probability and results in intractable expressions the signal power is lower bounded by 0 in this case. Under virtual channel approximation, (6) is a lower bound on SNR.

Remark 1. If the quantized virtual angles are distributed uniformly in their range, instead of the distribution in Lemma 2, $D_j(\eta, U_x)$ takes a much simplified form given by $D(\eta, U_x) = \sum_{d=1}^{\eta-1} {N_{BS}-1 \choose d} (N_{BS} - 1 - d)^{U_x - 1} \times \sum_{i=0}^{d} (-1)^i (d-i)^{\eta-1} {d \choose i}.$

Remark 2. To simplify evaluation of Proposition 1, the

 $\begin{array}{l} (1 - q_{\mathrm{BS},j})^{U_x - 1} \sum_{i=1}^{\mathrm{N}_{\mathrm{UE}}} q_{\mathrm{UE},i} (1 - q_{\mathrm{UE},i})^{\eta - 1} \\ \mathrm{B}_j(\eta, U_x) \leq (1 - q_{\mathrm{BS},j})^{U_x - 1}. \end{array}$ \leq

Remark 3. It can be shown that $\sum_{i=1}^{N_{BS}} q_{BS,i}(1 - \sum_{i=1}^{N_{BS}} q_{BS,i}(1 - \sum$ $(q_{\mathrm{BS},i})^r \to 1$ as $N_{\mathrm{BS}} \to \infty$ for any $r \ge 0$, which is true since $\max_i q_{BS,i} \to 0$ as $N_{BS} \to \infty$. Similar result holds for $q_{\mathrm{UE},j}$ with $N_{\mathrm{UE}} \rightarrow \infty$. All these imply that $\zeta \rightarrow 1$ with $N_{\rm BS} \rightarrow \infty$ and $N_{\rm UE} \rightarrow \infty$.

To find the SNR coverage, we need to find the distribution of several random variables in Proposition 1. First

we focus on the finding the probability mass function (PMF) of the number of multiuser streams of BS at $y \in \Phi_{\rm BS}$ given by $U_y = \min\{U_{\rm M}, N_y\}$. We use an ⁹ approximation proposed in [29] to model the distribution of N_u , which are actually correlated random variables for $y \in \Phi_{BS}$ and particularly known to be intractable since finding the volume of Voronoi association cells is itself an unsolved problem [30]. With notably different propagation channels for LOS and NLOS links, the cell association regions in mmWave networks are not even Voronoi and more irregular [11]. The PMF of N_{μ} is denoted by $\kappa(n)$ is modelled as follows [11]. Let $\rho = \lambda_{\rm UE}/\lambda_{\rm BS}$, then if y = x, that is the BS is serving the typical user, $\kappa(n)$ is given by

$$\frac{3.5^{3.5}}{(n-1)!} \frac{\Gamma(n+3.5)}{\Gamma(3.5)} \rho^{n-1} \left(3.5+\rho\right)^{-n-3.5}, \quad (8)$$

for $n \ge 1$ and $\kappa(0) = 0$. For interfering BSs, $\kappa(n) =$

$$\frac{3.5^{3.5}}{n!} \frac{\Gamma(n+3.5)}{\Gamma(3.5)} \rho^n \left(3.5+\rho\right)^{-n-3.5},\qquad(9)$$

for $n \ge 0$.

Assuming N_y to be independently and identically distributed, we model the PMF of U_y by

$$\mathbb{P}(U_y = n) = \mathbb{1}_{\{0 \le n \le U_M - 1\}} \kappa(n) + \left(1 - \sum_{i=1}^{U_M - 1} \kappa(n)\right) \mathbb{1}_{\{n = U_M\}} \quad (10)$$

We first model the point process Φ_{BS} to be superposition of the point processes Φ_L and Φ_N with intensities $\lambda_{BS} p_{LOS} \mathbb{1}_{\{||x|| \leq D\}}$ and $\lambda_{BS} (1 - p_{LOS}) \mathbb{1}_{\{||x|| \leq D\}} +$ $\lambda_{\rm BS} \mathbb{1}_{\{||x||>D\}}$, respectively. These two point processes correspond to LOS and NLOS BSs. The corresponding propagation processes [31] are given as \mathcal{N}_{L} = $\{||y||^{\alpha_{\rm L}}/S_{y,{\rm L}} : y \in \Phi_{\rm L}\}, \text{ and } \mathcal{N}_{\rm N} = \{||y||^{\alpha_{\rm N}}/S_{y,{\rm N}} :$ $y \in \Phi_{\mathrm{N}}$.

Lemma 3. \mathcal{N}_{L} is a non-homogeneous PPP with intensity $\Lambda_{\rm L}([0,t)) = \lambda_{\rm BS} M_{\rm L}(t)$, where $M_{\rm L}(t)$ is given by (11). Here, $m = -0.1\beta \ln 10$, $\sigma_{\rm L} = 0.1\xi_{\rm L} \ln 10$, $\Upsilon_i(t) =$ $\frac{\ln(\frac{D^{\alpha_j}}{t})-m}{T}$ for $j \in \{L, N\}$ and Q(.) is the Q-function (Standard Gaussian CCDF).

Proof: Special case of Appendix A of [11] and is therefore skipped for brevity.

Lemma 4. \mathcal{N}_{N} is a non-homogeneous PPP with intensity $\Lambda_{\rm N}([0,t)) = \lambda_{\rm BS} M_{\rm N}(t)$, where $M_{\rm N}(t)$ is given by (12).

Proof: Proceeds very similarly to Lemma 3 and thus is omitted.

Note that here $\Lambda_{\rm L}([0,\infty)) = \lambda_{\rm BS} \pi p_{\rm LOS} {\rm D}^2$. The probability that there is no point in the interval $[0,\infty)$ is equal to $\exp(-\lambda_{\rm BS}\pi p_{\rm LOS}D^2)$. This is exactly the

$$M_{\rm L}(t) = \pi p_{\rm LOS} \left[D^2 \mathsf{Q} \left(\Upsilon_{\rm L}(t) \right) + t^{\frac{2}{\alpha_{\rm L}}} \exp\left(\frac{2\sigma_{\rm L}^2}{\alpha_{\rm L}^2} + \frac{2m}{\alpha_{\rm L}}\right) \mathsf{Q} \left(\frac{2\sigma_{\rm L}^2}{\alpha_{\rm L}\sigma_{\rm L}} - \Upsilon_{\rm L}(t)\right) \right]. \tag{11}$$

$$M_{N}(t) = -\pi p_{LOS} D^{2} \mathsf{Q} \left(\Upsilon_{N}(t)\right) + \pi t^{\frac{2}{\alpha_{N}}} \exp\left(\frac{2\sigma_{N}^{2}}{\alpha_{N}^{2}} + \frac{2m}{\alpha_{N}}\right) \left[1 - p_{LOS} \mathsf{Q} \left(\frac{2\sigma_{N}^{2}}{\alpha_{N}\sigma_{N}} - \Upsilon_{N}(t)\right)\right].$$
(12)

probability that there is no point in Φ_L . Let us call the probability that there is at least one point in \mathcal{N}_L to be B_L . The event that number of points in Φ_N is zero is empty and thus, $B_N = 1$.

Corollary 1. Let \mathcal{N} be the point process of propagation losses corresponding to Φ_{BS} . This point process is a PPP with intensity $\Lambda((0,t]) = \lambda_{BS}(M_L(t) + M_N(t)) = \lambda_{BS}M(t)$.

Proof: Follows directly from the Superposition property of PPPs [28, Proposition 1.3.3].

Lemma 5. Given that \mathcal{N}_{L} and \mathcal{N}_{N} are not empty, the probability density function (PDF) of the distance to the point nearest to origin in these point processes is given by $f_{L}(t) = \lambda_{BS} \exp(-\lambda_{BS}M_{L}(t)) M'_{L}(t)/B_{L}$ and $f_{N}(t) = \lambda_{BS} \exp(-\lambda_{BS}M_{N}(t)) M'_{N}(t)/B_{N}$, where $M'_{L}(t)$ and $M'_{N}(t)$ are given in (13) and (14), respectively.

Proof: If l^* is the point nearest to origin in the point process \mathcal{N}_L ,

$$\begin{split} & \mathbb{P}\left(l^* > t \big| \mathcal{N}_L([0,\infty)) > 0\right) \\ &= \mathbb{P}\left(\mathcal{N}_L([0,t)) = 0 \big| \mathcal{N}_L([0,\infty)) > 0\right) \\ &= \frac{\mathbb{P}\left(\mathcal{N}_L([0,t)) = 0 \cap \mathcal{N}_L([0,\infty)) > 0\right)}{\mathbb{P}\left(\mathcal{N}_L([0,\infty)) > 0\right)} \\ &= \mathbb{P}\left(\mathcal{N}_L([0,t)) = 0 \cap \mathcal{N}_L([t,\infty)) > 0\right) / \mathbf{B}_{\mathbf{L}} \end{split}$$

$$= \mathbb{P} \left(\mathcal{N}_{L}([0,t)) = 0 \right) \mathbb{P} \left(\mathcal{N}_{L}([t,\infty)) > 0 \right) / \mathcal{B}_{L}$$

= exp $\left(-\Lambda_{L}([0,t)] \right) \left(1 - \exp\left(-\Lambda_{L}([t,\infty)] \right) \right) / \mathcal{B}_{L}$
= $\left(\exp\left(-\Lambda_{L}([0,t)] \right) - \exp\left(-\Lambda_{L}([0,\infty)] \right) \right) / \mathcal{B}_{L}.$

Thus, taking the negative derivative of the above expression we get the PDF $f_{\rm L}(t)$. Similarly, we can derive the PDF for the NLOS case.

Theorem 1. The SNR coverage of a typical user in the network is given by

$$\overline{\mathcal{S}}(\tau) \triangleq \mathbb{P}(\text{SNR}_{x,0} > \tau) = \mathbb{E}_{U_x} [\mathcal{S}(\tau, U_x)], \text{ where}$$
(15)

$$\mathcal{S}(\tau, \mathbf{U}) \approx \sum_{j \in \{\mathbf{L}, \mathbf{N}\}} \mathbf{B}_{j} \zeta(\eta_{j}, \mathbf{U}) \sum_{n=1}^{\eta_{j}} (-1)^{n+1} {\eta_{j} \choose n} \\ \times \int_{0}^{\infty} \exp\left(-\frac{\eta_{j} \tau n \mathbf{U} l \sigma_{n}^{2}}{\mathbf{G}} - \lambda_{\mathrm{BS}} \mathbf{M}_{\overline{j}}(l)\right) f_{j}(l) \mathrm{d}l,$$

where $G = PN_{BS}N_{UE}$, $\overline{j} = L$ if j = N and vice versa. The terms $\zeta(.)$, $M_j(.)$ and $f_{(.)}$ are derived in Proposition 1, Lemma 3, Lemma 4 and Lemma 5.

Corollary 2. Assuming that user density is much larger than BS density, the SNR coverage can be approximated by $S(\tau, U_M)$.

Theorem 2. In a noise-limited network, the per user rate distribution (or rate coverage) of a typical user at origin served by a BS at x is given by

$$\mathcal{R}(\tau_r) \triangleq \mathbb{P}\left(R_{x,0} > \tau_r\right)$$
$$= \sum_{n>1} \kappa(n) \mathcal{S}\left(2^{\frac{\tau_r n}{\omega_{\mathrm{MU}^{\mathrm{B}\min\left(n,\mathrm{U}_{\mathrm{M}}\right)}}} - 1, \min\left(n,\mathrm{U}_{\mathrm{M}}\right)\right)$$

where S(.) was defined in Theorem 1 and $\kappa(n)$ is given in (8).

Proof: Follows by re-arranging (4) and using SNR = SINR.

Although the above expression is an infinite summation, as verified earlier in [11], [29], it can be accurately represented as a finite summation. For the results in this work, considering the first $\lfloor 12\lambda_{\rm UE}/\lambda_{\rm BS} \rfloor$ terms is sufficient. The following definition will be useful when comparing the rate coverage of MU-MIMO with SM and SU-BF.

Definition 2. The minimum allowable efficiency of scheme A such that it is guaranteed to outperform scheme B in terms of per user rate for p percentile users (that is users with rate coverage p), is given by $\mathcal{O}_{A,B}(p) = \frac{\mathcal{R}_B^{-1}(p)}{\mathcal{R}_A^{-1}(p)}$, where \mathcal{R}_A^{-1} and \mathcal{R}_B^{-1} are inverse of the rate coverage at p (that is rate thresholds τ corresponding to $\mathcal{R}(\tau) = p$) for schemes A and B after setting $\omega_A = \omega_B = 1$, where $\omega_{(.)}$ are the efficiency factors for the respective MIMO techniques as defined in (4). The per user rate of A cannot stochastically dominate that of B, unless the efficiency of A is at least $\min_p \mathcal{O}_{A,B}(p)$.

Note that MU-MIMO implementations with different U_M are considered as separate MIMO schemes in the above definition since they have different efficiency factors. It is clear that $\frac{\mathcal{O}_{A,B}(p)}{\omega_A}$ is an upper bound on the ratio $\mathcal{R}_B^{-1}(p)$ and $\mathcal{R}_A^{-1}(p)$ for non-unity efficiency for scheme

$$M'_{\rm L}(t) = \pi p_{\rm LOS} \left\{ \exp\left(\frac{2\sigma_{\rm L}^2}{\alpha_{\rm L}^2} + \frac{2m}{\alpha_{\rm L}}\right) t^{\frac{2}{\alpha_{\rm L}} - 1} \left[\frac{2}{\alpha_{\rm L}} Q\left(\frac{2\sigma_{\rm L}^2}{\alpha_{\rm L}\sigma_{\rm L}} - \Upsilon_{\rm L}(t)\right) - \frac{1}{\sqrt{2\pi\sigma_{\rm L}^2}} \exp\left(-\left(\frac{\sqrt{2}\sigma_{\rm L}}{\alpha_{\rm L}} - \frac{\Upsilon_{\rm L}(t)}{\sqrt{2}}\right)^2\right)\right] + \frac{D^2}{\sqrt{2\pi}t\sigma_{\rm L}} \exp\left(-\frac{\Upsilon_{\rm L}^2(t)}{2}\right) \right\}.$$
(13)

$$\mathbf{M}_{\mathbf{N}}^{'}(t) = \pi p_{\mathrm{LOS}} \left\{ \exp\left(\frac{2\sigma_{\mathrm{N}}^{2}}{\alpha_{\mathrm{N}}^{2}} + \frac{2m}{\alpha_{\mathrm{N}}}\right) t^{\frac{2}{\alpha_{\mathrm{N}}} - 1} \left[\frac{2}{p_{\mathrm{LOS}}\alpha_{\mathrm{N}}} - \frac{2}{\alpha_{\mathrm{N}}} \mathsf{Q}\left(\frac{2\sigma_{\mathrm{N}}^{2}}{\alpha_{\mathrm{N}}\sigma_{\mathrm{N}}} - \Upsilon_{\mathrm{N}}(t)\right) + \frac{1}{\sqrt{2\pi\sigma_{\mathrm{N}}^{2}}} \exp\left(-\left(\frac{\sqrt{2}\sigma_{\mathrm{N}}}{\alpha_{\mathrm{N}}} - \frac{\Upsilon_{\mathrm{N}}(t)}{\sqrt{2}}\right)^{2}\right)\right] - \frac{\mathrm{D}^{2}}{\sqrt{2\pi}t\sigma_{\mathrm{N}}} \exp\left(-\frac{\Upsilon_{\mathrm{N}}^{2}(t)}{2}\right) \right\}. \quad (14)$$

A and $\omega_{\rm B} = 1$, since $\omega_{\rm A}$ is the minimum efficiency over all BSs in the network. Note that from (4), setting $\omega_{\rm B} = 1$ gives upper bound on rate for scheme B. For scheme A to outperform scheme B for *p* percentile users, we need $\frac{\mathcal{O}_{\rm A,B}(p)}{\omega_{\rm A}} \leq 1$ with equality giving minimum allowable $\omega_{\rm A}$. If the network is such that all BSs see the same overheads $\omega_{\rm A}$ or $\omega_{\rm B}$, then $\mathcal{O}_{\rm A,B}(p)$ is the minimum allowable relative efficiency (that is $\omega_{\rm A}/\omega_{\rm B}$) of scheme A over scheme B. This gives tighter estimates for allowable $\omega_{\rm A}$ especially when comparing MU-MIMO for different U_M or SM for different number of streams.

2) Rate Distribution in an Interference-limited Network. Until now, our analysis focused on noise-limited mmWave cellular networks. In this section, we will discuss how to model interference in these networks.

From (3), the OCI power at user u served by a BS at x is modelled as

$$\begin{split} I_u &= \mathbf{P} \sum_{\substack{y \in \Phi_{\mathrm{BS}} \\ y \neq x}} \sum_{w \in \mathcal{U}_y} \frac{||\overline{\mathbf{h}}_{y,u}^* \mathbf{f}_{y,w}^{\mathrm{BB}}||^2}{U_y} \\ &= \mathbf{P} \sum_{\substack{y \in \Phi_{\mathrm{BS}} \\ y \neq x}} \sum_{w \in \mathcal{U}_y} \frac{||\mathbf{w}_u^* \mathbf{H}_{y,u} \mathbf{F}_y^{\mathrm{RF}} \mathbf{f}_{y,w}^{\mathrm{BB}}||^2}{U_y}. \end{split}$$

Here, $\mathbf{w}_u = \mathbf{a}_{\mathrm{UE}}(\phi_{x,u})$, $\mathbf{H}_{y,u} = \sqrt{\frac{N_{\mathrm{BS}}N_{\mathrm{UE}}}{L(y,u)\eta_{y,u}}} \sum_{i=1}^{\eta_{y,u}} \gamma_{i,y,u} \mathbf{a}_{\mathrm{UE}}(\phi_{i,y,u}) \mathbf{a}_{\mathrm{BS}}^*(\theta_{i,y,u})$, $\mathbf{F}_y^{\mathrm{RF}}$ has columns equal to $\mathbf{a}_{\mathrm{BS}}(\theta_{y,w})$ for all $w \in \mathcal{U}_y$, and $\mathbf{f}_{y,w}^{\mathrm{BB}}$ is designed so as to cancel the multiuser interference of the BS at y. All the AOAs and AODs in the above expression are independent of each other. Leveraging the virtual channel approximation for large number of antennas at the BS and UE, interference due to the link between BS at y and user at w on the user u is non-zero if and only if $\phi_{x,u}$ is equal to at least one of the AOA of $\mathbf{H}_{y,u}$ and $\theta_{y,w}$ equals the corresponding AOD. Note that since multiuser interference was cancelled

by the ZF precoder, the virtual approximation with an ON/OFF model for inner product of two beam steering vectors gave us a tractable and accurate tool to study SNR distribution in the previous section. However, this model may not be accurate when OCI is incorporated.

The virtual channel approximation quantized the angular space into N sectors, where N is the number of antennas. If two angles lie on either sides of a sector boundary, the inner product of beam steering vectors is zero, which can be a main cause of underestimated interference. In order to alleviate this problem, we introduce the notion of sidelobe gain which was also used in [6], [11]. We still assume that the virtual angle space is quantized into N sectors with the angle bisector being a representative of each sector, but the inner product between two beam steering vectors is defined as:

$$\mathbf{a}_{\mathrm{BS}}^{*}(\theta_{1})\mathbf{a}_{\mathrm{BS}}(\theta_{2}) \triangleq \begin{cases} 1 & \text{if } \theta_{1} = \theta_{2} \\ \rho_{\mathrm{BS}} & \text{otherwise,} \end{cases}$$
(16)

where $\rho_{\rm BS} < 1$ introduces a sidelobe gain into the model. Similarly, we model the inner product for beam steering vectors at UEs with parameter $\rho_{\rm UE}$. Note that setting $\rho_{\rm BS} = \rho_{\rm UE} = 0$ reverts back to the virtual channel approximation.

To characterize the interference distribution, we neglect the effect of ZF on interfering links and dependence in $p_{\rm MU}$ and I_u through $\{\mathbf{w}_u\}$ for tractability and show that for a fairly large number of antennas this is a reasonable approximation. First we deal with the $\eta_{\rm L} = \eta_{\rm N} = 1$ case.

Proposition 2. Assuming the inner product of any two beam steering vectors at BS or UE follow the law given by (16), $\eta_{\rm L} = \eta_{\rm N} = 1$ and propagation loss on the service link is l, the OCI power at the typical user can be modelled as $I_0 = \sum_{y \in \Phi_{\rm BS}, y \neq x} G|\gamma_{y,0}|^2 L(y,0)^{-1} \chi_y/U_y$, where $\gamma_{y,0}$ is complex normal random variable with unit variance and zero mean, U_y are i.i.d random variables with distribution given in (10) and χ_y is defined as

$$\chi_{y} = \begin{cases} k + (U_{y} - k)\rho_{\rm BS}^{2} & w.p. \; (\sum_{i=1}^{N_{\rm UE}} q_{{\rm UE},i}^{2}) \times \\ \begin{pmatrix} U_{y} \\ k \end{pmatrix} \sum_{j=1}^{N_{\rm BS}} q_{{\rm BS},j}^{k+1} (1 - q_{{\rm BS},j})^{U_{y}-k} \\ \rho_{\rm UE}^{2} (k + (U_{t} - k)\rho_{{\rm BS}}^{2}) \; w.p. \; (1 - \sum_{i=1}^{N_{\rm UE}} q_{{\rm UE},i}^{2}) \times \\ \begin{pmatrix} U_{y} \\ k \end{pmatrix} \sum_{j=1}^{N_{\rm BS}} q_{{\rm BS},j}^{k+1} (1 - q_{{\rm BS},j})^{U_{y}-k}, \end{cases}$$

for $k = 1, 2, \ldots, U_y$.

Proof: For single path channel, the out-of-cell interference is given by

$$I_{0} = \sum_{\substack{y \in \Phi_{\mathrm{BS}}, y \neq x}} \frac{\mathrm{G}|\gamma_{y}|^{2} L(y, u)^{-1}}{U_{y}} \times \sum_{\substack{w \in \mathcal{U}_{y}}} ||\mathbf{a}_{\mathrm{UE}}^{*}(\phi_{x, u}) \mathbf{a}_{\mathrm{UE}}(\phi_{y, u}) \mathbf{a}_{\mathrm{BS}}^{*}(\theta_{y, u}) \mathbf{a}_{\mathrm{BS}}(\theta_{y, w})||^{2}$$

Now using the inner product rule in (16) and the fact that all the virtual angles in the above equation are independent and distributed according to Lemma 2, the proposition can be proved.

Lemma 6. The Laplace functional of the interference power in Proposition 2 conditioned on path loss to the typical user at origin from serving BS is L(x, u) = l, is given by

$$\begin{split} L_{I_0,l}(s) &\triangleq \mathbb{E} \left[\exp\left(-sI_0\right) | L(x,0) = l \right] & \rho \\ &= \exp\left(-\lambda_{\rm BS} \sum_{n=0}^{\rm U_M} \tilde{p}(n) \sum_{k=0}^n \binom{n}{k} \sum_{i=1}^{\rm N_{\rm BS}} q_{{\rm BS},i}^{k+1} (1-q_{{\rm BS},i})^{n-k} \exp\left(\left(\sum_{i=1}^{\rm N_{\rm UE}} q_{{\rm UE},i}^2\right) \int_{t \ge l} \frac{M'(t) dt}{1 + \frac{tn}{sG(k+(n-k)\rho_{{\rm BS}}^2)}} \right) \\ &+ \left(1 - \sum_{i=1}^{\rm N_{\rm UE}} q_{{\rm UE},i}^2\right) \int_{t \ge l} \frac{M'(t) dt}{1 + \frac{M'(t) dt}{sG\rho_{{\rm UE}}^2(k+(n-k)\rho_{{\rm BS}}^2)}} \right\} \right) . \text{ in } dt$$

where $\tilde{p}(.)$ is the distribution of U_y for interfering BSs given in (10).

Theorem 3. The SINR coverage of the typical user is given by (15) with an extra term $L_{I_0,l}\left(\frac{\eta_j \tau n Ul}{G}\right)$ inside the integral over dl.

Proof: Exactly on same lines as Theorem 1. The Laplace functional $L_{I_0,l}(.)$ has been derived in Lemma 6 for single path channel. Upper and lower bounds on $L_{I_0,l}(.)$ for a general number of paths can be found in Appendix D.

From this expression of SINR coverage, the rate coverage can be found similar to Theorem 2. We will validate these analytical results in Section V. In the next section, we will take a brief look into the coverage and rate for SM enabled mmWave cellular networks.

Before that though, we provide a brief discussion on how to choose $\rho_{\rm UE}$ and $\rho_{\rm BS}$. Recall that $N_{\rm BS}\rho_{\rm BS}^2$ and $N_{\rm UE}\rho_{\rm UE}^2$ are the sidelobe gains for beam pattern at BSs and UEs, respectively. An obvious question is whether these parameters depend on the number of antennas and if yes, how should their dependence be modelled?

If $\rho_{(.)}$ were to be a constant, the sidelobe gain will also scale up with an increasing number of antennas. This will violate Lemma 1. Since virtual channel approximation asymptotically tracks physical channel model, $\rho_{\rm BS}$ and $\rho_{\rm UE}$ should decrease and eventually vanish with increasing $N_{\rm BS}$ and $N_{\rm UE},$ respectively. For a uniform linear array with N antennas, the ratio of the gain of the i^{th} sidelobe to the main lobe is equal to $|\frac{\sin(0.5\pi(2i+1))}{N\sin(0.5\pi(2i+1)/N)}|^2$ [32], for $i = 1, 2, \dots, \lfloor \frac{N}{2} - 1 \rfloor$. For $i \ll N$, this ratio is independent of N using the small angle approximation $\sin \theta \approx \theta$. For *i* on the order of N, this ratio decreases approximately as square of N. The regime in which the ratio is independent of N has about fixed beamwidth, which corresponds to the beamwidth in which the small angle approximation of $\sin \theta \approx \theta$ is accurate with p percent relative error. For p = 1, $\theta \approx 0.244$ radians. Since the majority of the angular space corresponds to the regime in which the above ratio varies inversely with the square of N, we model $\rho_{(.)}$ to linearly decrease with N. We choose $ho_{\rm BS}=1/(\sin(0.244){
m N}_{\rm BS})$ and $U_{UE} = 1/(\sin(0.244)N_{UE})$, however in future it is esirable to re-investigate the scaling factor to get a etter fit.

IV. SINGLE USER SPATIAL MULTIPLEXING IN MMWAVE CELLULAR NETWORKS

For spatial multiplexing (SM), we consider a scenario in which every BS transmits more than one stream of data to a single user per resource block. Thus, $N_s^{UE} = N_s^{BS} = N_s$, where N_s is the multiplexing gain. In this section, we will focus mainly on the multipath diversity approach for SM [3], [4] and not on the polarization approach [3], [7].

A. Spatial Multiplexing: UHF versus mmWave

We begin with a brief recap of the theoretically optimal implementation of closed-loop SM in conventional cellular networks, which motivates the main challenges in precoding/combining for SM in mmWave networks. Under unitary power constraint, given the singular value decomposition of the channel matrix $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$, the transmitter pre-multiplies the input symbols with matrix \mathbf{V} and the receiver combines the received signal on all its antennas with matrix \mathbf{U}^* , to effectively achieve N_s parallel channels, where N_s is the multiplexing gain. Since the channel matrix is either full row rank or full column rank with high probability for sub 6 GHz frequency bands, $N_s = \min\{N_{BS}, N_{UE}\}$.

At mmWave frequencies, however, the first challenge is that it is not practically feasible to implement a fully digital precoder and combiner. Using the popular hybrid beamforming approach for mmWave networks [8], the precoder is of the form $\mathbf{F}_x^{\mathrm{RF}}\mathbf{F}_x^{\mathrm{BB}}$, wherein $\mathbf{F}_x^{\mathrm{RF}}$ is generally implemented using phase shifters and thus has constant magnitude entries. Similarly, we have a constraint for the combiner. Another challenge for implementing SM at mmWave is that the channel is sparse [3], [19] and thus obtaining multiplexing gain on the order of number of antennas is nearly impossible.

We now look at a typical implementation of SM using the hybrid beamforming architecture in Fig. 1. Assuming perfect channel estimation, and using the system model from Section II, the received signal at user u from BS x is given by

$$\mathbf{y}_u = \mathbf{H}_{x,u} \mathbf{F}_x^{\text{RF}} \mathbf{F}_x^{\text{BB}} \mathbf{s}_u + \mathbf{n} + \text{OCI}$$

where \mathbf{s}_u are transmit symbols of dimension $N_s \times 1$ with energy per symbol equal to P/N_s , **n** is the noise power (complex Gaussian with zero mean and variance σ_n^2). We assume equal power allocation to all streams. After RF and baseband combining at the receiver, the processed signal is of the form $\mathbf{W}_u^{\text{BB*}} \mathbf{W}_u^{\text{RF*}} \mathbf{y}_u$. When Gaussian symbols are transmitted over the mmWave channel, the spectral efficiency can be at most [4]

$$r = \log_2 \left| \mathbf{I}_{N_s} + \frac{P}{N_s} \mathbf{R}_n^{-1} \mathbf{H}_{\text{eff}} \mathbf{H}_{\text{eff}}^* \right|,$$

where $\mathbf{H}_{\text{eff}} = \mathbf{W}_{u}^{\text{BB}*} \mathbf{W}_{u}^{\text{RF}*} \mathbf{H}_{x,u} \mathbf{F}_{x}^{\text{RF}} \mathbf{F}_{x}^{\text{BB}}$, $\mathbf{I}_{N_{s}}$ is an identity matrix of rank N_{s} and

$$\mathbf{R_n} = \sigma_n^2 \mathbf{W}_u^{\mathrm{BB}*} \mathbf{W}_u^{\mathrm{RF}*} \mathbf{W}_u^{\mathrm{RF}} \mathbf{W}_u^{\mathrm{BB}},$$

We use the near optimal precoding-combining algorithm proposed in [4] for our simulations. Assuming that an equal fraction of resource is allocated to each UE connected to a BS, the per user rate is defined as $R_{x,u} = \omega_{\rm SM} Br/N_x$, where $\omega_{\rm SM}$ is the efficiency factor for SM and recalling that N_x is the total number of users associated with the BS at x. Note that similar to the MU case, $\omega_{\rm SM}$ is dependent on several network parameters like number of antennas, the channel parameters, number of streams, etc. but we drop this in the notation for convenience. Sum rate is defined as the total bits per second transmitted by a BS in Section III. Based on this definition, we define the sum rate for the SM enabled mmWave network to be $R_x = \omega_{\rm SM}Br$.

B. Heuristic Comparison of Coverage and Rate for MU-MIMO and SM

In this section, we denote the SNR with a superscript SM and MU to identify spatial multiplexing and MU-MIMO. Round robin scheduling and $\omega_{MU} = \omega_{SM} = 1$

will be assumed in this section. Recall from Lemma 1 that for a large number of antennas the eigenvalues of the channel matrix $\mathbf{H}_{x,u}$ converge to $\frac{N_{\mathrm{BS}}N_{\mathrm{UE}}|\gamma_{i,x,u}|^2}{L(x,u)^{-1}}$. Thus, the ratio $\frac{\mathbb{SNR}_{i,x,u}^{\mathrm{SM}}}{G} \xrightarrow{d} \frac{|\gamma_{i,x,u}|^2 L(x,u)^{-1}}{\eta_{x,u} N_s}$. From Remark 3, the ratio $\frac{\mathbb{SNR}_{x,u}^{\mathrm{SM}}}{G} \xrightarrow{d} \frac{|\gamma_{i,x,u}|^2 L(x,u)^{-1}}{\eta_{x,u} N_s}$, where $i_m = \arg\max_i \gamma_{i,x,u}$. Since $\gamma_{i,x,u} \le \max_i \gamma_{i,x,u}$, we can conclude that in the limit as $N_{\mathrm{BS}} \to \infty$ and $N_{\mathrm{UE}} \to \infty$, $\frac{\mathbb{SNR}_{i,x,u}^{\mathrm{SM}}}{G} \le \frac{\mathbb{SNR}_{x,u}^{\mathrm{MU}}}{G}$ for all $i \in \{1, \dots, N_s\}$. The above discussion hints that for many antennas

at BS and UE, the SNR with MU-MIMO stochastically dominates the SNR on each stream of SM. If the network were to be noise-limited, the per user and sum rates with MU-MIMO will be higher than SM for a large number of antennas and the same number of streams. Now, let us consider how this result might be affected by OCI. As the number of antennas become large, the effect of zero forcing on the interfering streams is negligible for both MU-MIMO and SM (since the virtual channel approximation in [23] starts to more closely model the actual channel). Thus, if the number of streams transmitted by the BS with SM and MU-MIMO are the same, the interference statistics with MU-MIMO and SM would be similar and one would expect that MU-MIMO still outperforms SM for a large number of antennas at BSs and UEs.

For a finite number of antennas the ZF penalty may be non-negligible. It is expected that the ZF penalty with SM will be less than MU-MIMO since there are more sidelobes that need to be suppressed with MU-MIMO. Thus, the above SNR dominance result holds given that the number of antenna is large enough such that the effect of the smaller ZF penalty with SM does not reverse the inequalties. For a finite number of antennas, it is neither obvious nor analytically tractable to conjecture as to whether the per user and sum rate of SM would dominate or whether MU-MIMO would. We, thus, rely on Monte Carlo simulations for SM while comparing with our validated analytical model for MU-MIMO and SU-BF.

V. NUMERICAL RESULTS

In this section, we first validate the SNR, SINR and rate coverage analysis from Section III. Next, we compare the per user and sum rate for SU-BF, MU-MIMO and SM with fixed number of BSs per unit area as well as fixed power consumption per unit area. The default parameters used for generating the results are given in Table I. The efficiency factors ω_{MU} and ω_{SM} are implementation specific and estimating these is not the focus of this study. Thus, we set the efficiency parameters to 1 and use Definition 2 for quantifying the allowable relative efficiency.



Fig. 2: Validation of SNR analysis in noise-limited scenario shows a tight match with the physical channel model simulations. Tradeoff between SINR and rate coverage is also shown with MU-MIMO.

Parame-	Value(s)	Parame-	Value(s)
ter		ter	
f_c	73 GHz [7]	В	1 GHz [7]
$p_{\rm LOS}, D$	0.11, 200	σ_n^2	-174 +
	m [11]		$10 \log_{10} B + 10$
			dBm
α (LOS,	2, 3.3 [7]	ξ (LOS,	5.2, 7.6 [7]
NLOS)		NLOS)	
$\lambda_{ m UE}$	500/km ²	$\lambda_{\rm BS}$	60/km ²
N _{UE}	16 [2], [33]	N _{BS}	64 [2], [33]
Р	30 dBm [5]	$\eta_{\rm L}, \eta_{\rm N}$	1,3 [2], [12], [20]

TABLE I: Simulation parameters

A. Coverage and Rate with MU-MIMO: Validation and Trends

1) Cases Where Interference is Negligible. Fig. 2a shows the validation of the SNR coverage formula in Theorem 1. As can be seen from the figure, the analysis is a tight approximation with the simulations using the physical channel model even when the virtual angles are equally likely, in which case we have much simplified analytical expressions as compared to when the distribution is as given in Lemma 2. Henceforth, all analysis plots will be with equally likely virtual angles. As expected, the match loosens as U_M approaches N_{BS} and N_{UE} . With increasing U_M , the SINR coverage decreases since the transmit power is split amongst the multiple users served by the BS. However, as seen from Fig. 2b, the median and peak per user rate increases with MU-MIMO. This is due to the fact that in round robin scheduling, each user connected to BS at x now gets $\min(U_M, N_x)$ times more slots to transmit. A re-interpretation of the above result can be made in terms of minimum allowable efficiencies. For example, $\mathcal{O}_{\{U_M=2\},\{U_M=1\}}(0.5) = 62.67\%$ and $\mathcal{O}_{\{U_M=4\},\{U_M=1\}}(0.5) = 42.73\%$. This means that if the efficiency of implementing MU-MIMO with $U_M = 2$ is at least 62.67% of the efficiency with $U_M = 1$, then it is beneficial to employ MU-MIMO with $U_M = 2$ over SU-BF in terms of the median rates.

Since SINR decreases with U_M , the trend for cell edge rates is exactly opposite to peak and median rates. Note that in [13], it was shown that cell edge rates can improve with MU-MIMO. However, the main difference in their model is the user selection and scheduling. In [13], there is a high priority user scheduled in a time slot and additional users are served using MU-MIMO only if the expected sum proportional fair metric does not increase due to addition of more users. This protects the rates achieved by cell edge users. The result in Fig. 2b, thus, highlights the importance of user selection and scheduling to protect the rates achieved by cell edge users with multiuser transmission.

2) Cases Where Interference is Not Negligible. Fig. 3a shows the validation of SINR coverage formula in Theorem 3 for single path scenario. In order to present a case where interference effects are not negligible we consider a network at 28 GHz band with 200 MHz bandwidth, a less blocked scenario with $p_{\text{LOS}} = 0.5$ for D = 200m and much higher $\lambda_{UE} = 1000/km^2$. As per discussion in Section III, $\rho_{BS} = 1/(N_{BS} \sin(0.244))$ and $\rho_{\rm UE} = 1/(N_{\rm UE}\sin(0.244))$. Fig. 3a shows that increasing BS density does not necessarily improve coverage. This trend is similar to that observed in [6] and shows the presence of an optimal BS density in terms of SINR coverage. Approximate analytical results in Lemma 6 and Appendix D capture the essential non-monotonic trend shown with the simulations. Fig. 3b further validates the analysis in Appendix D for multipath scenario as



(b) Validation of $\rho_{(.)}$ and bounds on Laplace functional for $\eta_{\rm L} = 2$, $\eta_{\rm N} = 3$.

Fig. 3: Validation in interference limited setting for $U_M = 4$ shows that the upper and lower bounds are within ± 5 dB of the actual simulations. SINR coverage has a non-monotonic trend with BS density.



(a) Sparse multipath: MU-MIMO outperforms except for very low(b) Enabling 4 stream SM for $\lambda_{UE} = 100/\text{km}^2$ shows benefit over loads. $\eta_L = 2$, $\eta_N = 3$ MU-MIMO for median/peak rates. $\eta_L = 10$, $\eta_N = 12$,

Fig. 4: Comparison of MIMO techniques with fixed BS density. $U_{\rm M} = 4$

well as shows a decreasing gap with the physical model simulations as the number of antennas grows large. Both these plots build confidence in the analysis and derived insights. Using analysis, it can be found that optimum BS density for $U_M = 1,2$ and 4 decreases as 82,72 and 63 BSs/km². Thus, with increasing U_M the optimum BS density reduces due to increasing interference in the network.

B. Comparing Per User and Sum Rate for SU-BF, MU-MIMO and SM

The gains with SM and MU-MIMO are fundamentally driven by distinct network parameters. For example, having more number of multipaths (or larger η_L and η_N)

increases the rank of the channel and thus enables transmitting more number of streams with single user SM, given that there are enough RF chains at the transmitter and the receiver. However, this does not necessarily help in having more multi-user streams. On the other hand, having low load reduces the possible gain with MU-MIMO even if each BS is equipped with a large number of RF chains due to the fact that there are not many users to schedule simultaneously per BS. This does not however affect SM in terms of the number of streams per user. Thus, sufficiently low load and high multipaths may cause SM to outperform MU-MIMO given that there are enough RF chains at the BSs and UEs. This can be seen in Figures 4a and 4b. The plots for MU-MIMO and SU-



Fig. 5: Comparison of per user and sum rate with fixed power consumption. A denser SU-BF network outperforms MU-MIMO and SM in terms of cell edge per user rates but MU-MIMO performs the best in terms of sum rates.

BF in Figure 4a are with analysis. The plots for SM in Figure 4a and the entire Figure 4b is using Monte-Carlo simulations. Note that our analytical model is valid for $\eta_L, \eta_N \gg N_{UE}$ and not for η_L, η_N close to N_{UE} , which is the case in Figure 4b. Figure 4a shows that for moderate and low user densities (which corresponds to $\lambda_{\rm UE} = 500/{\rm km}^2$ and $\lambda_{\rm UE} = 100/{\rm km}^2$) MU-MIMO outperforms SM and SU-BF. However, for very low load (corresponds to 10 UEs/km²) SM outperforms MU-MIMO. This result is due to the fact that although SM can offer 2 streams per user but MU-MIMO cannot provide gains since per km² there are only 10 users that can associate with 60 possible BSs and the probability that a BS connects to more than 1 user is very low. Since our analytical model slightly loose estimates for low SNR users, we compare the cell edge rates using simulations only. The cell edge rates are quite close for the three schemes, although SM and SU-BF slightly outperform MU-MIMO. For low loads, SM is slightly better than SU-BF in terms of cell edge rates. Considering that overhead with MU-MIMO could be the highest, this trend will be more exaggerated afte considering these factors. A better scheduling will be indeed important for protecting cell edge rates with MU-MIMO.

Figure 4b shows the impact of high multipath on the comparison insights. As was observed in Figure 4a, MU-MIMO outperformed SM for $\lambda_{\rm UE} = 100/\rm{km}^2$ when multipath was low. For the same network parameters, that lead to a noise-limited case, increasing the multipath to $\eta_{\rm L} = 10$ and $\eta_{\rm N} = 12$ gives higher rates with SM for even 30 percentile users. This is again due to the fact that since there are 4 RF chains at UEs and BSs, SM can support 4 streams per user. However, since there are about 1.7 UEs per BS, BSs can only transmit to

about 2 UEs per time slot on an average with MU-MIMO. Further the increased multipath leads to higher ZF penalty for MU-MIMO. Similar trend is observed in the interference-limited scenario ($f_c = 28$ GHz, B = 100 MHz, $p_{\text{LOS}} = 0.5$). Since a low blockage scenario is considered, the 4 streams per UE are LOS links with very high probability. Thus, the gains with SM look slightly exaggerated in the interference-limited case. Also note that having a large multipath as considered here could be a unlikely scenario in outdoor mmWave networks [2] but it is interesting to consider from an analytical perspective.

A re-interpretation of the above plots can be made in terms of minimum allowable efficiency of MU-MIMO to outperform SM or SU-BF. For example, when $\lambda_{\rm UE} = 500/\rm{km^2}$ in Figure 4a, MU-MIMO outperforms SM in terms of median users if its efficiency factor is more than 58%. Similarly, such numbers can be extracted for other plots using Definition 2. As mentioned earlier, a separate study on estimating these efficiency factors is needed to make a strong claim on comparison of these MIMO techniques.

The above comparison results were for fixed BS density and the same number of antennas across different schemes. However, with an increasing number of RF chains, the power consumed per BS also increases. In the hybrid precoding as shown in Fig. 1, each RF chain is connected to all antennas through phase shifters. Thus, with increasing number of RF chains the number of phase shifters grows proportionally with the number of antennas, and effectively the power consumption is also increased. Let $\nu(N_{RF})$ denote the ratio of power consumed at a BS with N_RF RF chains to a BS with 1 RF chain. A ballpark value of ν can be found to be 1.38 for $N_{\rm BS}=64$ and $N_{\rm RF}=U_{\rm M}=2$ based on the power consumption model in [25] (refer [1] for a discussion on this). We now scale up the BS density of SU-BF by exactly a factor of ν . Note that UEs need to use only single RF chain for SU-BF and MU-MIMO with hybrid precoding. However, UEs need multiple RF chains for SM with hybrid precoding architecture. Thus, for fair comparison considering power consumption model in [25] we reduce the $N_{\rm UE}$ to 7 for SM. As can be seen from Fig. 5, the gain in per user data rates with MU-MIMO and SM diminishes or completely vanishes if the SU-BF network has 1.38 times denser deployment on an average. Fig. 5 shows that MU-MIMO still has significantly higher sum rates than for a denser SU-BF network. However, per user cell edge rates with a denser SU-BF network are higher in this case. To quantify the cell edge gains in per user rates $\mathcal{O}_{MU,SU}(0.95) = 315\%$, which is huge and strengthens our conclusion that a denser SU-BF network outperforms MU-MIMO in terms of cell edge rates. Also note that $\mathcal{O}_{MU,SU}(0.5) = 99\%$, which implies that most likely even the median gains with SU-BF will be better after incorporating the channel acquisition overheads. However, in terms of sum rates $\mathcal{O}_{\rm MUSU}(0.5) = 73\%$ which implies that median rates with MU-MIMO can still be higher as long as the efficiency is more than 73% of SU-BF efficiency.

VI. CONCLUSIONS AND FUTURE WORK

The analytical model in this work demonstrates the utility of the virtual channel approximation to incorporate different precoder and combiner constraints in network level analysis of dense MIMO cellular networks with many antennas. It would be beneficial to get tighter bounds on the Laplace functional of the out-of-cell interference. The analytical model can also be extended to incorporate more realistic cross-polarized uniform planar arrays instead of ULA. Another important issue that needs to be addressed is to incorporate the effects of imperfect channel state information in the analytical model. Since MU-MIMO requires more channel state information at the transmitter, imperfect channel knowledge may affect the performance of MU-MIMO more than SM or SU-BF. It is essential to know whether this would overshadow the benefits of MU-MIMO over SM and SU-BF observed in this paper.

APPENDIX

A. Derivation of Zero Forcing Penalty in Proposition 1

For simplicity in notation, let us denote by θ_i^i and ϕ_i^i as the AOD and AOA on the j^{th} path from/to the BS at xunder consideration to/from the i^{th} user, $i \in \{1, \dots, U\}$, served by the BS, respectively. $\overline{\mathbf{H}}_{x,u}$ is equal to (7) when all of the following events are true.

- E_1 : $\mathbf{a}_{\mathrm{UE}}^*(\phi_1^k)\mathbf{a}_{\mathrm{UE}}(\phi_j^k)\mathbf{a}_{\mathrm{BS}}^*(\theta_j^k)\mathbf{a}_{\mathrm{BS}}(\theta_1^1) = 0$ for all $j \in \{1, ..., \eta_k\}$ and $k \in \{2, ..., U\}$.
- E_2 : $\mathbf{a}_{\mathrm{UE}}^*(\phi_1^1)\mathbf{a}_{\mathrm{UE}}(\phi_j^1)\mathbf{a}_{\mathrm{BS}}^*(\theta_j^1)\mathbf{a}_{\mathrm{BS}}(\theta_1^k) = 0$ for all $j \in \{1, \ldots, \eta_1\}$ and $k \in \{2, \ldots, U\}$.
- E_3 : $\mathbf{a}_{\text{UE}}^*(\phi_1^1)\mathbf{a}_{\text{UE}}(\phi_i^1)\mathbf{a}_{\text{BS}}^*(\theta_i^1)\mathbf{a}_{\text{BS}}(\theta_1^1) = 0$ for all $j \in \{2,\ldots,\eta_1\}.$

Note that probability of $p_{\rm ZF} = 1$ is given by $\mathbb{P}(E_1 \cap$ $E_2 \cap E_3$). Using the ON/OFF nature of inner products of beamsteering vectors with virtual channel approximation, we can re-write the above conditions as

- $\mathcal{E}_1 = A_1 \cap A_2$, where $A_1 = \bigcap_{k=2}^{U} \{\theta_1^1 \neq \theta_1^k\}$ and $A_2 = \bigcap_{k=2}^{U} \bigcap_{j=2}^{\eta_k} \{\phi_1^k \neq \phi_j^k\} \cup \{\theta_j^k \neq \theta_1^1\}.$ $\mathcal{E}_2 = A_1 \cap A_3$, where $A_3 = \bigcap_{k=1}^{U} \bigcap_{j=2}^{\eta_1} \{\phi_1^1 \neq \phi_j^1\} \cup \{\theta_j^1 \neq \theta_1^k\}.$

Note that $\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) = \mathbb{P}(A_1 \cap A_2 \cap \mathcal{E}_3)$ A_3). Conditioning on θ_1^1 , A_2 is independent of A_1 and A₃. Using (a) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, (b) all distinct AOA or AOD are independently distributed as per the distribution given in Lemma 2, (c) $\begin{array}{l} \bigcap_{k=1}^{\mathrm{U}}\bigcap_{j=2}^{\eta_1}\{\phi_1^1\neq\phi_j^1\}\cup\bigcap_{k=1}^{\mathrm{U}}\bigcap_{j=2}^{\eta_1}\{\theta_j^1\neq\theta_1^k\}\subset\\ \bigcap_{k=1}^{\mathrm{U}}\bigcap_{j=2}^{\eta_1}\{\phi_1^1\neq\phi_j^1\}\cup\{\theta_j^1\neq\theta_1^k\} \text{ and (d) for a highly} \end{array}$ dense network, the probability that the BS is serving a LOS UE is p_{LOS} since the association region of a BS is almost surely covered by the ball of radius D centered at the BS, the required lower bound on the probability of $p_{\rm ZF} = 1$ is derived, also given by $\zeta(.)$. In order to get the more simplified expression in Remark 1, the term $D_i(.)$ in Proposition 1 needs to be simplified. For equally likely virtual angles, this can be found using the following Lemma, which we propose.

Lemma 7. Pick U numbers that take values in range $\{1, \ldots, N\}$. Repetition of values is allowed and order is important. The probability that the first U_1 numbers are mutually exclusive from the remaining $U_2 = U - U_1$ is given by P, where

$$\mathcal{P} = \sum_{d=1}^{U_1} \binom{N}{d} (N-d)^{U_2} \sum_{i=0}^d \binom{d}{i} (-1)^i (d-i)^{U_1}.$$

The idea is to condition that there are d distinct values in first U_1 numbers, in which case the remaining U_2 numbers can take values in $(N - d)^{U_2}$ ways. Further the number of ways in which first U_1 numbers take d distinct values can be found using inclusion exclusion principle, which is given by the inner summation.

B. Proof of Theorem 1

Let $l_{\rm L}^*$ and $l_{\rm N}^*$ denote the points closest to origin in $\mathcal{N}_{\rm L}$ and \mathcal{N}_{N} , respectively. Using Lemma 5, the probability of associating with a LOS BS is given by

$$\begin{aligned} \mathbf{A}_{\mathrm{L}} &= \mathbf{B}_{\mathrm{L}} \int_{0}^{\infty} \mathbb{P}\left(l_{\mathrm{N}}^{*} > t\right) f_{\mathrm{L}}(t) \mathrm{d}t \\ &= \mathbf{B}_{\mathrm{L}} \int_{0}^{\infty} \exp\left(-\lambda_{\mathrm{BS}} \mathbf{M}_{\mathrm{N}}(t)\right) f_{\mathrm{L}}(t) \mathrm{d}t. \end{aligned}$$

Similarly, the probability of associating with NLOS BS is given by $A_N = B_N \int_0^\infty \exp(-\lambda_{BS} M_L(t)) f_N(t) dt$. Similar to Lemma 3 in [6], the PDF of propagation loss to associated BS given that the association is of type LOS, is given by $f_{\rm L}(t) =$ $\frac{B_L}{A_L} f_L(t) \exp{(-\lambda_{BS} M_N(t))}$. Similarly, the PDF of propagation loss given the associated BS is NLOS is given by $\tilde{f}_{\rm N}(t) = \frac{{\rm B}_{\rm N}}{{\rm A}_{\rm N}} f_{\rm N}(t) \exp\left(-\lambda_{\rm BS} {\rm M}_{\rm L}(t)\right)$. Define $\mathcal{S}(\tau, \mathbf{U}) \triangleq \mathbb{P}(SNR_{x,0} > \tau | U_x = \mathbf{U}).$ Thus, $\overline{\mathcal{S}}(\tau) =$

 $\mathbb{E}_{U_x=U}[\mathcal{S}(\tau, U)]$. By the law of total probability,

$$\begin{split} \mathcal{S}(\tau, \mathbf{U}) &= \mathbf{A}_{\mathrm{L}} \mathbb{P} \left(\mathrm{SNR}_{x,0} > \tau \Big| \mathrm{LOS \ connection} \right) \\ &+ \mathbf{A}_{\mathrm{N}} \mathbb{P} \left(\mathrm{SNR}_{x,0} > \tau \Big| \mathrm{NLOS \ connection} \right) \\ \stackrel{(a)}{\approx} \mathbf{A}_{\mathrm{L}} \int_{0}^{\infty} \mathbb{P} \left(|\gamma_{i_{m},x,0}|^{2} > \frac{\eta_{\mathrm{L}} \tau \mathrm{U} t \sigma_{n}^{2}}{p_{\mathrm{MU}} \mathrm{G}} \right) \tilde{f}_{\mathrm{L}}(t) \mathrm{d}t \\ &+ \mathbf{A}_{\mathrm{N}} \int_{0}^{\infty} \mathbb{P} \left(|\gamma_{i_{m},x,0}|^{2} > \frac{\eta_{\mathrm{L}} \tau \mathrm{U} t \sigma_{n}^{2}}{p_{\mathrm{MU}} \mathrm{G}} \right) \tilde{f}_{\mathrm{N}}(t) \mathrm{d}t \\ &= \mathbf{B}_{\mathrm{L}} \int_{0}^{\infty} \mathbb{P} \left(|\gamma_{i_{m},x,0}|^{2} > \frac{\eta_{\mathrm{L}} \tau \mathrm{U} t \sigma_{n}^{2}}{p_{\mathrm{MU}} \mathrm{G}} \right) e^{-\lambda_{\mathrm{BS}} \mathrm{M}_{\mathrm{N}}(t)} f_{\mathrm{L}}(t) \mathrm{d}t \\ &+ \mathbf{B}_{\mathrm{N}} \int_{0}^{\infty} \mathbb{P} \left(|\gamma_{i_{m},x,0}|^{2} > \frac{\eta_{\mathrm{N}} \tau \mathrm{U} t \sigma_{n}^{2}}{p_{\mathrm{MU}} \mathrm{G}} \right) e^{-\lambda_{\mathrm{BS}} \mathrm{M}_{\mathrm{L}}(t)} f_{\mathrm{N}}(t) \mathrm{d}t \end{split}$$

where (a) is obtained using Proposition 1. Note that the first integral is the probability that SNR exceeds the threshold and there is LOS connection, whereas the second term is for NLOS connection. Let us consider the probabilities in each of these two terms separately.

$$\begin{split} & \mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{L}}\tau \mathrm{U}t\sigma_n^2}{p_{\mathrm{MU}}\mathrm{G}}\right) \\ &= \mathbb{P}\left(p_{\mathrm{MU}} = 1\right) \mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{L}}\tau \mathrm{U}t\sigma_n^2}{p_{\mathrm{MU}}\mathrm{G}}\Big| p_{\mathrm{MU}} = 1\right) \\ & \stackrel{(b)}{=} \zeta(\eta_{\mathrm{L}},\mathrm{U}) \mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\mathrm{L}}\tau \mathrm{U}t\sigma_n^2}{\mathrm{G}}\right), \end{split}$$

where (b) is obtained from distribution of $p_{\rm MU}$ in Proposition 1.

Further, using the distribution of maximum of $\eta_{\rm L}$ exponential random variables for $|\gamma_{i_m,x,0}|^2$,

$$\mathbb{P}\left(|\gamma_{i_m,x,0}|^2 > \frac{\eta_{\rm L}\tau {\rm U}t\sigma_n^2}{{\rm G}}\right)$$
$$= \sum_{n=1}^{\eta_{\rm L}} (-1)^{n+1} \binom{\eta_{\rm L}}{n} \exp\left(-\eta_{\rm L}\tau n {\rm U}t\sigma_n^2/{\rm G}\right)$$

Similarly, we can find the NLOS probability term, which completes the proof.

C. Proof of Lemma 6

The Laplace functional of the out-of-cell interference to a user at origin, given the path loss to the serving BS, is defined as $L_{I_0,l}(s) \triangleq \mathbb{E} \left[\exp \left(-sI_0 \right) | L(x,0) = l \right].$

$$\begin{split} {}_{l}(s) &= \mathbb{E}\left[\exp\left(-s\sum_{y\in\Phi_{\mathrm{BS}},y\neq x}\frac{\mathrm{G}|\gamma_{y,0}|^{2}\chi_{y}}{L(y,0)U_{y}}\right) \left| L(x,0) = l\right] \\ &\stackrel{(a)}{=} \mathbb{E}\left[\exp\left(-s\sum_{t\in\mathcal{N},t\geq l}\frac{\mathrm{G}|\gamma_{t}|^{2}t^{-1}\chi_{t}}{U_{t}}\right)\right] \\ &\stackrel{(b)}{=} \mathbb{E}\left[\prod_{t\in\mathcal{N},t\geq l}\exp\left(-\frac{s\mathrm{G}|\gamma_{t}|^{2}t^{-1}\chi_{t}}{U_{t}}\right)\right] \\ &= \mathbb{E}\left[\prod_{t\in\mathcal{N},t\geq l}\mathbb{E}_{|\gamma_{t}|^{2}}\left[\exp\left(-\frac{s\mathrm{G}|\gamma_{t}|^{2}t^{-1}\chi_{t}}{U_{t}}\right)\right]\right] \\ &= \mathbb{E}\left[\prod_{t\in\mathcal{N},t\geq l}\frac{1}{1+\psi_{t}}\right] \\ &\stackrel{(d)}{=}\exp\left(-\int_{l}^{\infty}\left(1-\mathbb{E}_{\psi_{t}}\left[\frac{1}{1+\psi_{t}}\right]\right)\Lambda(dt)\right) \\ &=\exp\left(-\int_{l}^{\infty}\left(\mathbb{E}_{\psi_{t}}\left[\frac{1}{1+\psi_{t}^{-1}}\right]\right)\Lambda(dt)\right), \end{split}$$

where (a) is obtained by displacing each point $y \in \Phi_{\rm BS}, y \neq x$ to $L(y,0) = t \in \mathcal{N}, t \geq l$. Note that $\gamma_{y,0}, U_y$ and χ_y are independent marks of $y \in \Phi_{\rm BS}$, whose distributions are themselves independent of the location y. After one to one mapping of each point $y \in \mathbb{R}^2$ to $t \in \mathbb{R}^+$ and each mark to itself, we associate each feasible point $t \in \mathcal{N}$ with independent marks γ_t, U_t and χ_t , with same distribution as the corresponding earlier marks. Here, (b) is obtained using independence of the marks of the displaced PPP and (c) since γ_t are exponentially distributed random variables with unit mean and $\psi_t = \frac{s G t^{-1} \chi_t}{U_t}$. Using the PGFL (probability generating functional) [28] we obtain (d). Using the distribution of U_y and χ_t , we get the required result.

D. Laplace Functional of Out-of-cell Interference for General Number of Paths

The out-of-cell interference from a BS at y to user at origin, served by BS at x is given by

$$I_{y,0} = \frac{\mathrm{G}L(y,0)^{-1}}{\eta_{y,u}U_y} \sum_{w \in \mathcal{U}_y} || \sum_{j=1}^{\eta_{y,u}} \gamma_j \mathbf{a}_{\mathrm{UE}}^*(\phi_{x,u}) \mathbf{a}_{\mathrm{UE}}(\phi_{j,y,u}) \times \mathbf{a}_{\mathrm{BS}}^*(\theta_{j,y,u}) \mathbf{a}_{\mathrm{BS}}(\theta_{y,w}) ||^2,$$

Thus,

 L_{I_0}

$$I_{y,0} = \frac{\mathrm{G}L(y,0)^{-1}}{\eta_{y,u}U_y} \sum_{w \in \mathcal{U}_y} ||\sum_{j=1}^{\eta_{y,u}} \gamma_j \chi_{j,w}||^2$$

where $\chi_{j,w}$ is given by,

$$\chi_{j,w} = \begin{cases} 1 & \text{if } \phi_{x,u} = \phi_{j,y,u} \text{ and } \theta_{y,w} = \theta_{j,y,u} \\ \rho_{\text{BS}} & \text{if } \phi_{x,u} = \phi_{j,y,u} \text{ and } \theta_{y,w} \neq \theta_{j,y,u} \\ \rho_{\text{UE}} & \text{if } \phi_{x,u} \neq \phi_{j,y,u} \text{ and } \theta_{y,w} = \theta_{j,y,u} \\ \rho_{\text{BS}}\rho_{\text{UE}} & \text{otherwise.} \end{cases}$$

Now let us look at the Laplace functional of this interference power.

$$L_{I_0}(s) = \mathbb{E} \left[\exp \left(-s \sum_{y \in \Phi_{\mathrm{BS}}, y \neq x} I_{y,0} \right) \right]$$

$$= \mathbb{E} \left[\exp \left(-s \sum_{y \in \Phi_{\mathrm{BS}}, y \neq x} \frac{\mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_y} \times \sum_{w \in \mathcal{U}_y} || \sum_{j=1}^{\eta_{y,0}} \gamma_j \chi_{j,w} ||^2 \right) \right]$$

$$\stackrel{(a)}{=} \mathbb{E} \left[\prod_{y \in \Phi_{\mathrm{BS}}, y \neq x} \mathbb{E}_{\chi_{(.,.)}, \gamma_{(.)}} \left[\exp \left(\frac{-s \mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_y} \times \sum_{w \in \mathcal{U}_y} || \sum_{j=1}^{\eta_{y,0}} \gamma_j \chi_{j,w} ||^2 \right) \right] \right]$$

where (a) follows since χ and γ have distributions independent of location y. Finding the exact distribution from this expression is intractable. The main bottleneck is that the small scale fading random variables γ_j , are together clubbed in a single norm expression and thus, although these random variables are assumed to be independent, the distribution of the norm squared for different users in \mathcal{U}_y are correlated exponential random variables. We, thus, find upper and lower bounds in this work.

1) Upper Bound on the Laplace Functional. In order to find an upper bound, we use the fact that $\chi_{j,w} \ge \rho_{\rm BS}\rho_{\rm UE}$. Thus,

$$L_{I_0}(s) \leq \mathbb{E} \left[\prod_{y \in \Phi_{\mathrm{BS}}, y \neq x} \mathbb{E}_{\chi_{(.,.)},\gamma_{(.)}} \left[\exp\left(-\frac{s \mathrm{G}\rho_{\mathrm{BS}}^2 \rho_{\mathrm{UE}}^2}{L(y,0)\eta_{y,0}U_y} \right) \right] \right] \\ \times \sum_{w \in \mathcal{U}_y} \left| \sum_{j=1}^{\eta_{y,0}} \gamma_j \right| \right]$$
$$\stackrel{(a)}{=} \mathbb{E} \left[\prod_{y \in \Phi_{\mathrm{BS}}, y \neq x} \mathbb{E}_{\Xi} \left[\exp\left(\frac{-s \mathrm{G}\rho_{\mathrm{BS}}^2 \rho_{\mathrm{UE}}^2 \Xi}{L(y,0)\eta_{y,0}}\right) \right] \right] \\ = \mathbb{E} \left[\prod_{y \in \Phi_{\mathrm{BS}}, y \neq x} \frac{1}{1 + s \mathrm{G}L(y,0)^{-1} \rho_{\mathrm{BS}}^2 \rho_{\mathrm{UE}}^2} \right],$$

where Ξ is an exponential random variable with mean $\eta_{y,0}$ in (a). In order to find the SINR distribution, we are interested in Laplace functional conditioned on path loss

to serving BS. Thus, conditioning on L(x, 0) = l and displacing the points in Φ to \mathcal{N} , similar to Appendix C we get,

$$L_{I_0,l}(s) \leq \mathbb{E}\left[\prod_{t\in\mathcal{N},t\geq l}\frac{1}{1+s\mathrm{G}t^{-1}\rho_{\mathrm{BS}}^2\rho_{\mathrm{UE}}^2}\right]$$
$$=\exp\left(-\int_l^\infty\frac{\Lambda(\mathrm{d}t)}{1+\frac{1}{s\mathrm{G}t^{-1}\rho_{\mathrm{BS}}^2\rho_{\mathrm{UE}}^2}}\right).$$

2) Lower Bound on the Laplace Functional. One obvious lower bound can be obtained using $\chi_{j,w} = 1$. The Laplace functional in this case is the same as for the upper bound with $\rho_{BS}^2 \rho_{UE}^2$ replaced by 1. However, with the narrow beamwidth for a large number of antennas, this approximation is clearly very pessimistic. We can get a tighter lower bound using the Cauchy-Schwarz inequality as follows.

$$L_{I_0}(s) \geq \mathbb{E} \left[\prod_{y \in \Phi, y \neq x} \mathbb{E}_{\chi_{(.,.)},\gamma_{(.)}} \left[\exp\left(\frac{-s \mathrm{G}L(y,0)^{-1}}{\eta_{y,0} U_y} \right) \times \left(\sum_{j=1}^{\eta_{y,0}} ||\gamma_j||^2\right) \sum_{w \in \mathcal{U}_y} \sum_{j=1}^{\eta_{y,0}} \chi_{j,w}^2 \right) \right] \right]$$
$$= \mathbb{E} \left[\prod_{y \in \Phi, y \neq x} \mathbb{E}_{\chi_{(.,.)}} \left[\left(1 + \frac{s \mathrm{G}L(y,0)^{-1}}{\eta_{y,0} U_y} \right) \times \sum_{w \in \mathcal{U}_y} \sum_{j=1}^{\eta_{y,0}} \chi_{j,w}^2 \right)^{-\eta_{y,0}} \right] \right]$$

Simplifying the term

$$\Psi_{y} = \mathbb{E}_{\eta_{y,0}} \left[\left(1 + \frac{s \mathrm{G} L(y,0)^{-1}}{\eta_{y,0} U_{y}} \sum_{w \in \mathcal{U}_{y}} \sum_{j=1}^{\eta_{y,0}} \chi_{j,w}^{2} \right)^{-\eta_{y,0}} \right],$$

we get

$$\begin{split} \Psi_{y} &= \sum_{i=1}^{N_{\rm BS}} \sum_{m=0}^{\eta_{y,0}} \binom{\eta_{y,0}}{m} q_{{\rm UE},i}^{m+1} (1-q_{{\rm UE},i})^{\eta_{y,0}-m} \\ &\times \sum_{k_{1},\ldots,k_{U_{y}}=1}^{N_{\rm BS}} \sum_{j_{1},\ldots,j_{U_{y}}=0}^{\eta_{y,0}} \prod_{n=1}^{U_{y}} \binom{\eta_{y,0}}{j_{n}} q_{{\rm BS},k_{n}}^{j_{n}+1} (1-q_{{\rm BS},k_{n}})^{\eta_{y,0}-j_{n}} \\ &\quad \times \left(1 + \frac{s {\rm G}L(y,0)^{-1}}{\eta_{y,0}U_{y}} \sum_{j=1}^{\eta_{y,0}} \left(\sum_{n=1}^{U_{y}} \left(\rho_{{\rm BS}}^{2} \right) \right) \\ &+ \mathbb{1}(j \leq j_{n}))(1-\rho_{{\rm BS}}^{2}) \left(\rho_{{\rm UE}}^{2} + \mathbb{1}(j \leq m)(1-\rho_{{\rm UE}}^{2}) \right) \right)^{-\eta_{y,0}} \end{split}$$

The above expression boils down to Lemma 6, for a single path channel. This expression can be further simplified assuming equiprobable virtual angles,

$$\begin{split} \Psi_{y} &= \sum_{m=0}^{\eta_{y,0}} \binom{\eta_{y,0}}{m} \left(\frac{1}{N_{\mathrm{UE}}}\right)^{m} \left(1 - \frac{1}{N_{\mathrm{UE}}}\right)^{\eta_{y,0}-m} \\ &\times \sum_{j_{1},\dots,j_{U_{y}}=0}^{\eta_{y,0}} \prod_{n=1}^{U_{y}} \binom{\eta_{y,0}}{j_{n}} \left(\frac{1}{N_{\mathrm{BS}}}\right)^{j_{n}} \left(1 - \frac{1}{N_{\mathrm{BS}}}\right)^{\eta_{y,0}-j_{n}} \\ &\times \left(1 + \frac{s\mathrm{G}L(y,0)^{-1}}{\eta_{y,0}U_{y}} \sum_{j=1}^{\eta_{y,0}} \left(\sum_{n=1}^{U_{y}} \left(\rho_{\mathrm{BS}}^{2} + \mathbbm{1}(j \leq j_{n})\right)\right) \\ &\times (1 - \rho_{\mathrm{BS}}^{2})\right) \left(\rho_{\mathrm{UE}}^{2} + \mathbbm{1}(j \leq m)(1 - \rho_{\mathrm{UE}}^{2})\right)\right)^{-\eta_{y,0}}. \end{split}$$

Now seperating the LOS and NLOS terms and using the Displacement theorem as for the upper bound, the Laplace functional can be given as

$$L_{I_0,l}(s) \ge \exp\left(-\int_l^\infty (1 - \mathbb{E}\left[\Psi_{t,\mathrm{L}}\right])\Lambda_{\mathrm{L}}(\mathrm{d}t)\right)$$
$$\times \exp\left(-\int_l^\infty (1 - \mathbb{E}\left[\Psi_{t,\mathrm{N}}\right])\Lambda_{\mathrm{N}}(\mathrm{d}t)\right).$$

where $\Psi_{t,j}$ is same as Ψ_y with y replaced by t and $\eta_{y,0}$ replaced by η_j , for $j \in \{L, N\}$.

ACKNOWLEDGMENT

The authors thank A. Alkhateeb for his comments on early drafts of this work. The authors also thank anonymous reviewers for their constructive comments.

REFERENCES

- M. N. Kulkarni, A. Alkhateeb, and J. G. Andrews, "A tractable model for per user rate in multiuser millimeter wave cellular networks," *in Proc. IEEE Asilomar Conference on Signals, Systems and Computers*, Nov. 2015.
- [2] M. R. Akdeniz, Y. Liu, M. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, June 2014.
- [3] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter wave wireless communications: beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.
- [4] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Jan. 2014.
- [5] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
- [6] T. Bai and R. W. Heath Jr., "Coverage and rate analysis for millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Oct. 2014.
- [7] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. MacCartney, S. Sun, and S. Nie, "Millimeter wave enhanced local area systems: A high data rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, June 2014.
- [8] A. Alkhateeb, J. Mo, and R. W. Heath, "MIMO precoding and combining solutions for millimeter wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

- [9] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.
- [10] A. Sayeed and N. Behdad, "Continuous aperture phased MIMO: Basic theory and applications," in Proc. Allerton Conference on Communications, Control and Computers, pp. 1196–1203, Sep. 2010.
- [11] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2196–2211, Oct. 2015.
- [12] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [13] F. W. Vook, T. Thomas, and E. Visotsky, "Massive MIMO for mmWave systems," in Proc. Asilomar Conference on Signals, Systems and Computers, Nov. 2014.
- [14] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.
- [15] R. W. Heath, T. Wu, Y. H. Kwon, and A. C. K. Soong, "Multiuser MIMO in distributed antenna systems with out-ofcell interference," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4885–4899, Oct. 2011.
- [16] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO HetNets: Modeling, ordering results and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5208–5222, Sept. 2013.
- [17] H. Zhuang and T. Ohtsuki, "A model based on Poisson point process for analyzing MIMO cellular networks utilizing fractional frequency reuse," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6839–6850, Dec. 2014.
- [18] A. K. Gupta, H. S. Dhillon, S. Vishwanath, and J. G. Andrews, "Downlink multi-antenna heterogeneous cellular network with load balancing," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 1830–1844, Nov. 2014.
- [19] M. K. Samimi and T. S. Rappaport, "Ultra-wideband statistical channel model for non line of sight millimeter wave urban channels," *IEEE Globecom*, pp. 3483–3489, Dec. 2014.
- [20] —, "Local multipath model parameters for generating 5G millimeter-wave 3GPP-like channel impulse response," in Proc. of 10th European Conference on Antennas and Propagation, April 2016, Available on ArXiv:1511.06941.
- [21] T. Bai and R. W. Heath, "Asymptotic SINR for millimeter wave massive MIMO cellular networks," *IEEE International Workshop* on Signal Processing Advances in Wireless Communications (SPAWC), pp. 620–624, June 2015.
- [22] M. N. Kulkarni, S. Singh, and J. G. Andrews, "Coverage and rate trends in dense urban millimeter wave cellular networks," *IEEE Globecom*, Dec. 2014.
- [23] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, 2002.
- [24] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPACWC)*, pp. 100–104, 2012.
- [25] R. Mendez-Rial, C. Rusu, A. Alkhateeb, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" *Proc. of Information Theory and Applications Workshop*, Feb. 2015.
- [26] A. Adhikary, E. A. Safadi, M. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mmWave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, May 2014.
- [27] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590 – 3600, Nov. 2010.

- [28] F. Baccelli and B. Blaszczyszyn, Stochastic Geometry and Wireless Networks, Volume I – Theory. NOW: Foundations and Trends in Networking, 2009.
- [29] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.
- [30] J.-S. Ferenc and Z. Néda, "On the size distribution of Poisson Voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518 – 526, Nov. 2007.
- [31] B. Blaszczyszyn, M. K. Karray, and H.-P. Keeler, "Using Poisson processes to model lattice cellular networks," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 773–781.
- [32] C. Balanis, Antenna Theory. Wiley, 1997.
- [33] T. Bai, A. Alkhateeb, and R. W. Heath, "Coverage and capacity of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70–77, Sept. 2014.