Analytical Framework of Hybrid Beamforming in Multi-Cell Millimeter-Wave Systems

Analytical Framework of Hybrid Beamforming in Multi-Cell Millimeter-Wave Systems

Shu Sun, *Student Member, IEEE*, Theodore S. Rappaport, *Fellow, IEEE*,
Mansoor Shafi, *Life Fellow, IEEE*, and Harsh Tataria, *Member, IEEE*

***Abstract*—Multi-cell wireless systems usually encounter both intra-cell and inter-cell interference, which can be mitigated via coordinated multipoint (CoMP) transmission. Previous works on multi-cell analysis in the microwave band generally consider fully digital beamforming, requiring a complete radio-frequency chain behind each antenna. This is practically infeasible for millimeter-wave (mmWave) systems where large amounts of antennas are necessary to provide sufficient gain and to enable transmission/reception of multiple streams to/from a user. This article provides a general methodology to analytically compute the expected per-cell spectral efficiency of a mmWave multi-cell single-stream system using phase-shifter-based analog beamforming and regularized zero-forcing digital beamforming. Four analog-digital hybrid beamforming techniques for multi-cell multi-stream mmWave communication are proposed, assuming that base stations in different cells share channel state information to cooperatively transmit signals to their home-cell users. Spectral efficiency of the proposed hybrid beamforming approaches are investigated and compared using two channel models suitable for fifth-generation cellular systems, namely the 3rd Generation Partnership Project model and the NYUSIM model. Numerical results show that the benefits of base station coordination (as compared to the no-coordination case) are governed by the underlying propagation model, and the aggregate interference levels proportional to the cell radius and number of users per cell.**

***Index Terms*—5G, CoMP, hybrid beamforming, millimeter wave (mmWave), multi-cell, MIMO.**

I. Introduction

Millimeter-wave (mmWave) technologies are expected to play a key role in fifth-generation (5G) mobile communications due to the tremendous amount of available bandwidth [1], [2]. MmWave cellular systems are expected to be densely deployed to guarantee acceptable coverage, spectral efficiency, as well as energy efficiency [3], [4]. In dense networks, a major challenge that needs to be solved is inter-cell interference. Extensive research has been done on mitigating inter-cell interference. For instance, power control and adaptive beamforming are two classical approaches for controlling multi-user interference [5], yet power control mainly improves the quality of weak links by equalizing the signal-to-interference-plus-noise ratio (SINR) for all users in a cell. On the other hand, adaptive antenna arrays can improve signal quality while mitigating interference via adjustment of spatial beam patterns. To reduce interference using arrays, one promising solution is to let base stations (BSs) or transmission points (TPs) in different geographical cells coordinate in transmission and/or reception. The aim of TP coordination is to prevent the transmitted signals from/to other TPs from incurring serious interference.

A. Prior Work

The 3rd Generation Partnership Project (3GPP) studied coordinated multipoint (CoMP) techniques for both downlink and uplink for fourth-generation (4G) communications in 2013 [6]. The different CoMP strategies in [6] entail various requirements with respect to channel state information (CSI) feedback and CSI sharing, which are detailed below in increasing order of complexity. (1) *Coordinated Scheduling/Beamforming*: Data for a user equipment (UE) is available at and transmitted from one TP in the CoMP cooperating set (downlink data transmission is done from that TP) for a time-frequency resource, but user scheduling/beamforming decisions are made with coordination among TPs [6]. (2) *Dynamic Point Selection/Muting*: Data is available simultaneously at multiple TPs but is transmitted from one TP in a time-frequency resource, and the transmitting/muting TP may change from one subframe to another [6]. (3) *Joint Transmission*: Data for a UE is available at multiple TPs and is simultaneously transmitted from multiple TPs to a single UE or multiple UEs in a time-frequency resource [6].

BS coordination for interference suppression has been extensively explored in the literature [7]–[10], yet those works focused on fully digital beamforming with one radio-frequency (RF) chain behind each antenna, which is not likely to be suitable for mmWave systems with large amounts (e.g., hundreds) of antennas at BSs due to hardware complexity, power consumption, and cost. BS cooperation in mmWave multi-cell networks was investigated in [11]–[13], but the mobile receiver was equipped with merely a single antenna, hence leading to only single-stream communication in those works. In fifth-generation (5G) mmWave systems, however, antenna arrays will also be employed at the mobile receiver to provide array gain and beamforming and/or spatial multiplexing capability for multiple streams.

Furthermore, the majority of the prior work did not provide an analytical framework when analyzing the multi-cell system performance, which could be done via eigenvalue distributions (EVDs) for uncorrelated and correlated Wishart matrices, which are well known [14]–[18]. Eigenvalue densities of the complex non-central Wishart channel correlation matrix were first derived in [17], [18] in matrix tensor form, and were later extended in [19] to a finite summation representation to facilitate analytical approximations of the regularized zero-forcing (RZF) expected SINR and spectral

This work is sponsored by the NYU WIRELESS Industrial Affiliates program, and NSF research grants 1320472, 1302336, and 1555332.

S. Sun is with Intel Corporation, Santa Clara, CA 95054, USA (e-mail: shu.sun@intel.com).

T. S. Rappaport is with the NYU WIRELESS Research Center, Tandon School of Engineering, New York University, Brooklyn, NY 11201, USA (e-mail: tsr@nyu.edu).

M. Shafi is with Spark New Zealand, Wellington 6011, New Zealand (e-mail: Mansoor.Shafi@spark.co.nz).

H. Tataria is with the Institute of Electronics, Communications, and Information Technology (ECIT), Queen's University Belfast, Belfast BT3 9DT, U.K. (e-mail: H.Tataria@qub.ac.uk).

efficiency for the general case of uncorrelated Ricean fading. The authors in [20] demonstrated the equivalent analysis with i.i.d. (independent and identically distributed) and semi-correlated Rayleigh fading channels, by averaging the analytical expressions over the arbitrary eigenvalue densities of the channel correlation matrix. However, the above eigenvalue densities are *all* for channel matrices with known statistics, facilitating analysis with digital processing, rather than the channel matrix multiplied by an RF precoding matrix that is used in analog-digital hybrid beamforming (HBF). Moreover, due to the analytical complications, none of the above studies consider channel models developed for 5G systems (including mmWave frequencies) in the derivation of the relevant channel eigenvalue densities.

B. Contributions of This Work

In this paper, we investigate the performance of multi-cell, multi-user, multi-stream analog-digital HBF for mmWave multiple-input multiple-output (MIMO) systems, where HBF is used at both the TP and UE, which has not been studied before to the best of our knowledge. The main contributions of this paper are as follows.

- Built on the multi-cell framework, EVDs for channels after RF precoding in a multi-cell multi-user system with a single stream per user are investigated for both signal and interference channels, which has not been studied before to our best knowledge. HBF based on RZF is employed at each TP. The channel matrices are generated using both the *3GPP TR 38.901 Release 14* [21] and the *NYUSIM (New York University SIMulator)* channel models [22]. The eigenvalue densities are approximated with a gamma distribution. The approximation of eigenvalue densities is motivated by the fact that exact densities are extremely challenging to derive, so that the best trade-off approach is to approximate.
- Leveraging the approximate channel eigenvalue densities from both channel models, for a single-stream multi-cell system employing hybrid processing, we give a methodology to derive tight analytical approximations of the expected per-user SINR and expected per-cell sum spectral efficiency. Our analyses assume a bank of phase shifters for the analog precoding and RZF processing for digital beamforming. Due to the joint design of both analog and digital processing matrices, we note the tremendous analytical complexity involved in deriving the aforementioned expressions. Hence, to the best of our knowledge, such general analysis of mmWave systems has been missing from the literature to date.
- A novel coordination-based HBF method with **s**ignal-maximizing and **l**eakage-minimizing **a**nalog **b**eamforming (SLAB) is proposed, which is improved upon the leakage-suppressing and signal-maximizing HBF in [23] by adding UE beamforming to enhance signal. Four multi-cell multi-stream downlink HBF approaches, where two use coordinated beamforming and two do not use TP coordination (including a baseline and SLAB), are compared in terms of spectral efficiency under various conditions (e.g., different cell radii, user numbers, and stream numbers per user). Equal power allocations are used for each stream, and numerical results demonstrate that benefits of multi-cell coordination depend on the underlying channel model and the aggregate interference levels, as shown in Fig. 5.

II. Multi-Cell System Layout and Underlying Channel Models

A. Multi-Cell System Layout

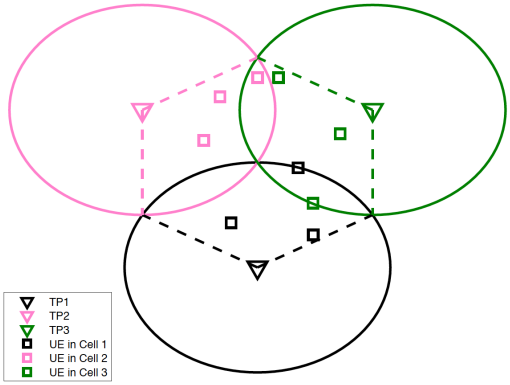
We assume the TPs in different cells have full CSI and can exchange the CSI among each other to mitigate inter-cell interference, which corresponds to coordinated beamforming as defined by 3GPP [6]. First, a multi-cell communication framework based upon today's conventional three-sector BS antenna configuration is formulated, where each 120° sector (i.e., cell [6]) uses a uniform rectangular array (URA) with 256 antenna elements (eight rows by 16 columns by two polarization states) for each TP.

Each antenna is a pair of co-located slant polarized antennas, slanted at $\pm 45^\circ$ [21]. The spacing between adjacent co-polarized elements is $\lambda/2$ in azimuth and λ in elevation, with λ denoting the carrier wavelength, and the radiation pattern of each TP antenna element given in Table I, which provides a half-power beamwidth resolution of about 8° in both azimuth and elevation in the broadside direction of the URA. A number of UEs (3 or 12 in this work), each with an eight-element URA (two rows by two columns by two polarization states) and four RF chains (for up to four streams per user), are randomly dropped in each cell over 2D distances of 10 m to the cell radius (e.g., 50 m or 200 m) [23]. URAs are considered because they are able to form beams in both azimuth and elevation dimensions, since exploiting the zenith characteristics of the propagation channel will be essential for enhanced performance at mmWave frequencies [25].

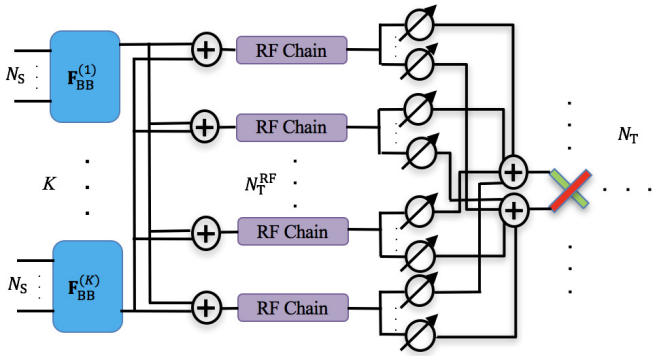
We consider a mmWave system with three cells (where each cell is a 120° sector), each having one TP and multiple UEs, as shown in Fig. 1(a). Only three cells are considered since interference caused by farther cells will be reduced, and analysis is simplified for a homogeneous multi-cell network with both intra- and inter-cell interference. The users in each cell are distributed uniformly and randomly with T-R separation distances ranging from 10 m to the cell radius [21], [23]. By assuming 95% of the area in a cell has an SNR larger than or equal to 5 dB, the upper bound of the T-R separation distance is calculated and rounded to 200 m for both models for fair comparison. The 50 m cell radius is set for comparison purposes. It is assumed that perfect CSI is available at both the home-cell TP and interfering TPs. This assumption, at first sight, may seem naive. However, there are several reasons for this: First, unlike previous studies, the central focus of this paper is to devise a general multi-cell analysis methodology to approximate the downlink expected per-cell spectral efficiency with hybrid processing (joint design of analog and digital beamforming networks) and channel models developed for 5G mmWave frequencies. Under this most general scenario, it is extremely difficult, if not intractable, to make analytical progress without perfect channel knowledge. Second, in line with [26], this assumption is reasonable in scenarios with low terminal mobility, where exploiting time division duplexing, a large fraction of the channel coherence interval can be spent on uplink training. Finally, the results obtained from the subsequent analysis and evaluation can be treated as a useful upper bound on what may be achieved in practice with imperfect channel knowledge. This paper considers a carrier frequency of 28 GHz with a 100 MHz RF bandwidth [21]. However, for the purpose of our study and following [27], we consider a narrowband block fading propagation model since orthogonal frequency-division

Table I
SIMULATION SETTINGS USING THE 3GPP [21] AND NYUSIM [22] MODELS.

Parameter	Setting
Carrier Frequency	28 GHz
Bandwidth	100 MHz [21]
Transmit Power Without Array Gain	35.2 dBm per UE (46 dBm for a cell with 12 UEs)
95% Cell-Edge SNR	5 dB
BS Antennas	three panels for the three sectors, where each panel is a uniform rectangular array with 256 cross-polarized elements in the x-z plane [21]
BS Antenna Spacing	half wavelength in azimuth; one wavelength in elevation
BS Antenna Element Gain	8 dBi [21]
BS Antenna Element Pattern	Model 2, Page 18 in 3GPP TR 36.873 Release 12 [24]
UE Antennas	uniform rectangular array consisting of eight cross-polarized elements in the x-z plane [21]
UE Antenna Spacing	half wavelength in azimuth; one wavelength in elevation
UE Antenna Element Gain	0 dBi [21]
UE Antenna Element Pattern	omnidirectional [21]
Receiver Noise Figure	10 dB



(a)



(b)

Figure 1. (a) An example of the three-cell layout where there is one TP and three UEs per cell, where each cell is a sector with an azimuth span of 120° served by one URA. (b) Multi-cell HBF architecture at the TP in each cell.

multiplexing (OFDM)-like modulations are likely to be used in 5G.

B. Channel Models Used in Analysis and Simulations

Two popular channel models for 5G wireless are the 3GPP [21] and NYUSIM [22], [28] channel models. Both models are stochastic channel models that include basic channel model components such as line-of-sight (LOS) probability model, large-scale path loss model, large-scale parameters, and small-scale parameters. However, the approaches and parameter values used in each channel modeling steps can be significantly different, as shown in [28].

III. MULTI-CELL MULTI-USER SINGLE-STREAM HYBRID BEAMFORMING

In this section, we investigate HBF for a multi-cell MU-MIMO system where each TP communicates with each of its home-cell users via a single data stream. The HBF architecture at each TP is illustrated in Fig. 1(b), where there are K baseband precoding units with one for each user in the same cell, one data stream is transmitted for each home-cell user, and each baseband unit is connected with $N_{\mathrm{T}}^{\mathrm{RF}}$ RF chains with $N_{\mathrm{T}}^{\mathrm{RF}} = K$. Each RF chain is connected to all of the N_{T} TP antennas through a network of N_{T} phase shifters, yielding a total of $N_{\mathrm{T}}^{\mathrm{RF}} N_{\mathrm{T}}$ phase shifters. The large numbers of antenna elements in mmWave systems require RF precoding techniques to provide antenna beamforming for multi-user separation. Digital precoding requires dedicated baseband and RF hardware for each antenna element, which increases cost, complexity, and power consumption. The spectral efficiency achieved via this approach is called the fully digital spectral efficiency. Reduction of implementation complexity is a motivation to look at other forms of precoding that achieve spectrum efficiency similar to the fully digital case. Coverage improvement in the spatially sparse mmWave channel motivates the use of transceiver structures with RF antenna processing, where the mmWave multipath spatial sparsity limits the numbers of simultaneous users. Therefore, best-case system spectral efficiency (close to the fully digital spectrum efficiency) can be achieved with hybrid beamforming (HBF) using much less hardware (especially RF chains) [23], [25]. HBF has two types of precoding, analog precoding and digital precoding. Analog precoding is implemented via phase shifters connecting each antenna element in an array to form the required spatially sparse beam patterns. The analog beamforming stage also plays a significant role in forming beam patterns to either enhance desired signal power or mitigate pattern leakage from a TP to undesired users. Digital precoding is implemented at baseband and is lower in dimension relative to the analog precoders as the numbers of multiple users are few due to sparsity. Both precoders work in tandem to separate the users as shown in Figs. 1(b) and 4. The angle information in the phase shifters is based on feedback from the UE — typically given in the form of a beam ID index — enabling a TP to choose from a number of preselect beams. However, in this work, we assume the ideal case with perfect angular resolution for the phase shifters available at the transmitter.

Each user is assumed to have either a single antenna or an antenna

array with analog beamforming only, for analytical and practical feasibility. For TP i and user k in cell l, the $1 \times N_{\mathrm{T}}$ downlink channel is denoted as $\mathbf{h}_{k,l,i}$, the $N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{RF}}$ RF precoding matrix is $\mathbf{F}_{\mathrm{RF}_i}$ ($N_{\mathrm{T}}^{\mathrm{RF}} \ll N_{\mathrm{T}}$), and the $N_{\mathrm{T}}^{\mathrm{RF}} \times 1$ baseband precoding vector is $\mathbf{f}_{\mathrm{BB}_{k,i}}$. The $1 \times N_{\mathrm{T}}^{\mathrm{RF}}$ effective channel $\check{\mathbf{h}}_{k,l,i}$ after RF precoding is:

$$\check{\mathbf{h}}_{k,l,i} = \mathbf{h}_{k,l,i}\mathbf{F}_{\mathrm{RF}_i} \tag{1}$$

where $\mathbf{F}_{\mathrm{RF}_i}$ is designed based on Algorithm 1 in [29]. The $K \times N_{\mathrm{T}}^{\mathrm{RF}}$ composite effective channel from TP i to all the K users in cell l is expressed as:

$$\check{\mathbf{H}}_{l,i} = [\check{\mathbf{h}}_{1,l,i}^H, \check{\mathbf{h}}_{2,l,i}^H, ..., \check{\mathbf{h}}_{K,l,i}^H,]^H \tag{2}$$

where the superscript H denotes conjugate transpose. The received signal at user k in cell l is:

$$y_{k,l} = \underbrace{\sqrt{\frac{P_{\mathrm{T}}}{\eta_l \mathrm{PL}_{k,l,l}}}\check{\mathbf{h}}_{k,l,l}\mathbf{f}_{\mathrm{BB}_{k,l}}s_{k,l}}_{\text{Desired Signal}} + \underbrace{\sum_{(m,i)\neq(k,l)}\sqrt{\frac{P_{\mathrm{T}}}{\eta_i \mathrm{PL}_{k,l,i}}}\check{\mathbf{h}}_{k,l,i}\mathbf{f}_{\mathrm{BB}_{m,i}}s_{m,i}}_{\text{Interference}} + \underbrace{n_{k,l}}_{\text{Noise}} \tag{3}$$

where P_{T} represents the total transmit power in Watts at each TP, $\mathrm{PL}_{k,l,i}$ denotes the large-scale distance-dependent path loss in Watts[1], including shadow fading, from TP i to user k in cell l. Furthermore, $\eta_l = ||\mathbf{F}_{\mathrm{RF}_l}\mathbf{F}_{\mathrm{BB}_l}||_F^2$ is a scaling factor to satisfy the total transmit power constraint $||\sqrt{P_{\mathrm{T}}}\mathbf{F}_{\mathrm{RF}_l}\mathbf{F}_{\mathrm{BB}_l}/\sqrt{\eta_l}||_F^2 = P_{\mathrm{T}}$, where F denotes the Frobenius norm, and $\mathbf{F}_{\mathrm{BB}_l} = [\mathbf{f}_{\mathrm{BB}_{1,l}}, ..., \mathbf{f}_{\mathrm{BB}_{K,l}}] \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{RF}} \times K}$. Note that $s_{k,l}$ represents the desired transmitted signal for user k in cell l with $\mathbb{E}[|s_{k,l}|^2] = 1$, and $n_{k,l} \sim \mathcal{CN}(0, N_0)$ is complex Gaussian noise with variance N_0. The signal model (3) generally applies to both LOS and non-line-of-sight (NLOS) environments, and the LOS/NLOS state in each channel realization is stochastic and determined by the LOS probability model in the underlying channel model. The SINR of user k in cell l is therefore:

$$\mathrm{SINR}_{k,l} = \frac{\frac{P_{\mathrm{T}}}{\eta_l \mathrm{PL}_{k,l,l}}|\check{\mathbf{h}}_{k,l,l}\mathbf{f}_{\mathrm{BB}_{k,l}}|^2}{\sum_{(m,i)\neq(k,l)}\frac{P_{\mathrm{T}}}{\eta_i \mathrm{PL}_{k,l,i}}|\check{\mathbf{h}}_{k,l,i}\mathbf{f}_{\mathrm{BB}_{m,i}}|^2 + N_0} \tag{4}$$

The expected per-user SINR can be obtained by calculating $\mathbb{E}[\mathrm{SINR}_{k,l}]$. The exact evaluation of $\mathbb{E}[\mathrm{SINR}_{k,l}]$, however, requires knowledge of the "exact" probability density of $\mathrm{SINR}_{k,l}$. This is usually unknown (particularly for ray-based channel models) and is extremely difficult if not intractable to characterize analytically in an a priori fashion. This has led many related works using simple statistical channel models to approximate the SINR expectation via the classical first-order Delta approximation as in [20], [30] and references therein. In line with these, we also leverage this approximation and note that the ultimate accuracy of the approximation relies on the variance of the interference power being small relative to its mean value. This is possible to achieve for scenarios with moderately large values of N_{T}, and can be mathematically seen via the application of the multivariate Taylor series expansion of the SINR around the mean of the signal over the mean of the interference powers [20], [30]. While exact analysis of this approximation is outside the scope of our study since we leverage the simple approximation introduced in [20], [30], we point out that its accuracy is characterized by Lemma 1 and Appendix I in [30]. Hence, the expected per-user SINR can be approximated as:

$$\mathbb{E}[\mathrm{SINR}_{k,l}] \approx \frac{\frac{P_{\mathrm{T}}}{\tilde{\eta}_l \mathrm{PL}_{k,l,l}}\mathbb{E}[|\check{\mathbf{h}}_{k,l,l}\mathbf{f}_{\mathrm{BB}_{k,l}}|^2]}{\sum_{(m,i)\neq(k,l)}\frac{P_{\mathrm{T}}}{\tilde{\eta}_i \mathrm{PL}_{k,l,i}}\mathbb{E}[|\check{\mathbf{h}}_{k,l,i}\mathbf{f}_{\mathrm{BB}_{m,i}}|^2] + N_0} \tag{5}$$

where $\tilde{\eta}_l = \mathbb{E}[\eta_l]$ and $\tilde{\eta}_i = \mathbb{E}[\eta_i]$. In what follows, the expected values in the numerator and denominator of (5) are derived separately using approximated densities for an arbitrary eigenvalue and a joint pair of arbitrary eigenvalues of both signal and interference channels for both models. The approximated density for an arbitrary eigenvalue is obtained via simulations, followed by the derivation of the approximated density for a joint pair of arbitrary eigenvalues detailed below.

A. Channel Eigenvalue Distribution

The EVDs are to determine the expected SINR, which is in turn needed to determine the ergodic spectral efficiency. The EVDs for uncorrelated and correlated Wishart matrices are presented in [14]–[18]. EVDs for channels after RF precoding in HBF, however, have not been investigated to the authors' best knowledge. This is because joint processing of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ twice alters both the magnitude and phase of the preferential channel directions, and therefore the complexity of exact expressions is very high. While the computation of *exact* eigenvalue densities with such complex channel models remains an open problem in multivariate statistics, we resort to accurate approximations to facilitate the subsequent analysis. In this subsection, we study EVDs of $\check{\mathbf{H}}\check{\mathbf{H}}^H$ for the 3GPP channel model and the NYUSIM channel model, where $\check{\mathbf{H}}$ denotes the effective channel matrix after TP RF precoding, i.e., $\check{\mathbf{H}} = \mathbf{H}\mathbf{F}_{\mathrm{RF}}$. Below are existing works on EVDs and the rationale for deriving the approximated EVDs in our work.

- In the simplest case of uncorrelated scattering, the entries of $\mathbf{H}$ are i.i.d. complex Gaussian random variables, widely known as Rayleigh fading, $\mathbf{H}\mathbf{H}^H$ is an uncorrelated central complex Wishart matrix, and the corresponding probability density function (PDF) of an arbitrary eigenvalue of $\mathbf{H}\mathbf{H}^H$ is derived in [14] via the orthogonal basis expansion of $\mathbf{H}\mathbf{H}^H$ as it is non-trivial to compute the density of each eigenvalue, even for this case.
- For the case of semi-correlated Rayleigh fading with spatial correlation at either transmit or receive end of the link, $\mathbf{H}\mathbf{H}^H$ takes the form of a correlated central complex Wishart distribution. The corresponding arbitrary eigenvalue densities are derived in [15], [17], [18], [20], [31] for various types of spatial correlation models.
- For an uncorrelated Ricean channel, $\mathbf{H}\mathbf{H}^H$ follows an uncorrelated non-central complex Wishart structure, whose eigenvalue densities were derived in [16].
- LOS components pointing in certain directions can be regarded as inducing additional spatial correlation [19]. The resultant $\mathbf{H}\mathbf{H}^H$ is a correlated non-central complex Wishart matrix, and the arbitrary eigenvalue densities for such channels were studied in [17]–[19].
- For a channel matrix combined with RF precoding and RF combining, it is conjectured that this is akin to inducing

[1]Note that generally path loss is defined in the dB scale representing a signal scale level, while in (3) it is defined in the linear scale for ease of calculation.

spatial correlation at both ends of the link in the direction of the boresight of the antenna (array). The antenna elements of the array are closely located (e.g., half wavelength) hence inducing spatial correlation as well. Furthermore, with a fixed number of scattering clusters and subpaths within each cluster, the channel models can be statistically treated as an arbitrary link gain pre-multiplied by a correlated random variable dependent on the antenna array configuration and the direction-of-departure/arrival distribution. Thus, the resultant arbitrary eigenvalue density will be similar to the second point mentioned above.

For the first four types of channels above, the mathematical form of the arbitrary eigenvalue density is a product of exponential functions with a finite power of the arbitrary eigenvalue upper bounded by the minimum of the transmit and receive antenna dimension. This is equivalent to the mathematical form of the density of a *gamma*-distributed random variable [32]. Moreover, while the chi-square and beta distributions also exhibit the above mathematical form, they are special cases of the gamma distribution with specific shape and rate parameters. Furthermore, the gamma distribution results in the highest Kolmogorov-Smirnov (KS) test statistic among all other contending distributions[2]. Therefore, it is reasonable to use the gamma distribution to approximate the eigenvalue density distribution. In what follows, to obtain approximated EVDs, we first plot the PDFs of ordered eigenvalues via simulations, and then fit the PDF curves with the gamma distribution by optimizing its shape and scale parameters. Specifically, the fitting is done via Algorithm 1 given below.

To further justify and verify the generality (e.g., applicability to different numbers of antennas at each TP and different numbers of users per cell) of the gamma distribution to approximate PDFs of the eigenvalues, we employ a multivariate statistical technique known as the *moment method*, which matches the moments of the true distribution with the one approximated. Define $X = \sum_{k=1}^{K} \lambda_k$ where λ_k denotes the k-th largest eigenvalue of $\check{\mathbf{H}}_{l,l}\check{\mathbf{H}}_{l,l}^H$, if $\lambda_1, ..., \lambda_K$ can be proved to follow a gamma distribution, then X is also gamma-distributed. Assuming the shape and rate parameters in the gamma distribution for X are α and β, respectively, the mean of X is μ, and the mean of X^2 is $\tilde{\mu}$, then through some fundamental mathematical derivation, we obtain $\mu = \frac{\alpha}{\beta},\ \tilde{\mu} = \frac{\alpha(\alpha+1)}{\beta^2}$. Equivalently, we have

$$\alpha = \frac{\mu^2}{\tilde{\mu} - \mu^2},\ \beta = \frac{\mu}{\tilde{\mu} - \mu^2} \tag{6}$$

Therefore, it is necessary and sufficient to verify (6) for various scenarios, e.g., different numbers of users per cell and different numbers of TP antenna elements, where α and β are to be obtained via mathematical fitting using the gamma distribution on simulated eigenvalues, while μ and $\tilde{\mu}$ are to be obtained through direct simulations. To verify (6), we considered the following cases: (i) 64 TP antenna elements and 6 users per cell, and (ii) 256 TP antenna elements and 3 users per cell. For each case, 1000 random channel realizations were performed to compute the eigenvalues of $\check{\mathbf{H}}_{l,l}\check{\mathbf{H}}_{l,l}^H$ and to obtain their sum X. Then the distribution of X was fitted using the gamma distribution which gives the shape and rate parameters α and β. On the other hand, μ and $\tilde{\mu}$ were calculated from X, which were then used to compute α and β using (6) via the moment method. Finally, these two sets of α and β were compared and the relative error was calculated, which are shown in Table II. As can be observed from Table II, for both cases with different numbers of TP antenna elements and users per cell, the relative differences in both α and β are within $\pm 13\%$, which is small, revealing the rationality and good generality of the gamma distribution when used to fit the eigenvalues.

[2]The KS test is a widely used measure in communications theory to determine the accuracy of an approximate statistical distribution relative to a specific system related metric [33].

Algorithm 1 Algorithm for Fitting Eigenvalues Using the Gamma Distribution

Require: Number of transmit antenna elements N_{T}, number of total receive antennas K in a cell, number of transmit RF chains $N_{\mathrm{T}}^{\mathrm{RF}}$ ($N_{\mathrm{T}}^{\mathrm{RF}} \geq K$), and number of simulation realizations N_{sim}

1: **for** $n_{\mathrm{sim}} = 1 : N_{\mathrm{sim}}$ **do**
2: Generate an $K \times N_{\mathrm{T}}$ propagation channel matrix $\mathbf{H}$ (for both models)
3: Construct the $N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{RF}}$ RF precoding matrix $\mathbf{F}_{\mathrm{RF}}$ given the array structure, using Algorithm 1 in [29]
4: Multiply $\mathbf{H}$ with $\mathbf{F}_{\mathrm{RF}}$ to give $\check{\mathbf{H}} = \mathbf{H}\mathbf{F}_{\mathrm{RF}}$
5: Compute the eigenvalues $\lambda' s$ of $\check{\mathbf{H}}\check{\mathbf{H}}^H$ by taking the eigenvalue decomposition of $\check{\mathbf{H}}\check{\mathbf{H}}^H$: $\check{\mathbf{H}}\check{\mathbf{H}}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$, where $\mathbf{\Lambda}$ is a $K \times K$ diagonal matrix consisting of eigenvalues
6: Extract all of the K eigenvalues, λ_1,..., λ_K (in descending order), of $\check{\mathbf{H}}\check{\mathbf{H}}^H$ and store in an array as a function of n_{sim}
7: **end for**
8: Obtain the approximated PDF of the k-th largest eigenvalue of $\check{\mathbf{H}}\check{\mathbf{H}}^H$: For all the eigenvalues generated from the N_{sim} channel realizations, extract the k-th eigenvalue λ_k from each channel realization, plot their PDF using MATLAB, and fit the density using the gamma distribution by adjusting the shape and rate parameters
9: Find the mathematical trend of the shape (rate) parameter for the K eigenvalues, and derive a common mathematical expression of the shape (rate) parameter for the K eigenvalues as a function of k.

For both models, the approximated PDF of the n-th largest eigenvalue, λ_n, of $\check{\mathbf{H}}_{l,l}\check{\mathbf{H}}_{l,l}^H$ is fitted using the gamma distribution based on the rationale described above, which is expressed as:

$$f_{\lambda_n}(\lambda_n) \approx \frac{b_n^{a_n} \lambda_n^{a_n - 1} e^{-b_n \lambda_n}}{\Gamma(a_n)},\ n = 1, ..., N \tag{7}$$

where a_n and b_n are the shape and rate parameters to be determined via simulations. When $K = 3$, for instance, $a_n = 1 + \frac{20}{30^n}$ and $b_n = \frac{3}{20} \prod_{s=1}^{n} s!$ for the 3GPP channel model, while $a_n = 1 + \frac{1}{3^n}$ and $b_n = \frac{3 \times 5^n}{5000}$ for NYUSIM. $\Gamma(\cdot)$ denotes the complete gamma function, and N is the smaller dimension of $\check{\mathbf{H}}_{l,l}$, which equals K for the single-stream-per-user case considered in (3). Based on the PDFs $f_{\lambda_n}(\lambda_n)$ of the ordered eigenvalues, the approximated PDF of an arbitrary eigenvalue of $\check{\mathbf{H}}_{l,l}\check{\mathbf{H}}_{l,l}^H$ is derived and expressed as [15]:

$$f_{\lambda_{\mathrm{arb}}}(\lambda_{\mathrm{arb}}) \approx \frac{1}{K} \sum_{n=1}^{K} \frac{b_n^{a_n} \lambda_{\mathrm{arb}}^{a_n - 1} e^{-b_n \lambda_{\mathrm{arb}}}}{\Gamma(a_n)}. \tag{8}$$

The approximated joint density of two arbitrary unordered eigenvalues of $\check{\mathbf{H}}_{l,l}\check{\mathbf{H}}_{l,l}^H$ is given by (see Appendix-A for detailed derivation):

Table II
SHAPE AND RATE PARAMETERS IN THE GAMMA DISTRIBUTION OBTAINED VIA MATHEMATICAL FITTING AND CALCULATION USING THE MOMENT METHOD.

	Scenario	Fitted Using Gamma Distribution	Calculated Using Moment Method	Relative Error
Shape Parameter α	64 TP Antennas, 6 UEs Per Cell	0.9786	0.8745	11.9%
	256 TP Antennas, 3 UEs Per Cell	0.5329	0.6058	-12.0%
Rate Parameter β	64 TP Antennas, 6 UEs Per Cell	0.0036	0.0032	12.5%
	256 TP Antennas, 3 UEs Per Cell	0.0020	0.0023	-13.0%

$$f_{\lambda,\text{unord}}(\lambda_1,\lambda_2) \approx \frac{1}{K(K-1)} \sum_{n=1}^{K} \sum_{\substack{q=1 \\ q\neq n}}^{K} (\lambda_1\lambda_2)^{-1} \Big[\big(\phi_n(\lambda_1)\big)^2 \big(\phi_q(\lambda_2)\big)^2 - \phi_n(\lambda_1)\phi_q(\lambda_1)\phi_n(\lambda_2)\phi_q(\lambda_2) \Big] \tag{9}$$

where $\phi_n(\lambda)$ is given by (56). The approximated PDF for the n-th largest eigenvalue of $\check{\mathbf{H}}_{l,i}^H \check{\mathbf{H}}_{l,i}$ (or equivalently $\check{\mathbf{H}}_{l,i} \check{\mathbf{H}}_{l,i}^H$) ($i \neq l$), where $\check{\mathbf{H}}_{l,i}$ represents the effective other-cell interference (OCI) channel, is found to be:

$$f_{\sigma_n}(\sigma_n) \approx \frac{d_n^{c_n} \sigma_n^{c_n-1} e^{-d_n\sigma_n}}{\Gamma(c_n)}, \; n = 1, ..., K \tag{10}$$

where $c_n = 1 + \frac{20}{100^n}$ and $d_n = \frac{10^{n-1}}{4}$ for the 3GPP model, and $c_n = 1 + \frac{1}{30^n}$ and $d_n = 6^{n-3}$ for NYUSIM, when $K = 3$. Note that there is variation with the coefficients in (7) and (10) for both models. One reason for this variation is the way the underlying channel impulse responses are generated from both models that results in very different eigenvalues [28] shown in Fig. 2 below. The approximated PDF for an arbitrary eigenvalue of $\check{\mathbf{H}}_{l,i}^H \check{\mathbf{H}}_{l,i}$ is given by [15]:

$$f_{\sigma_{\text{arb}}}(\sigma_{\text{arb}}) \approx \frac{1}{K} \sum_{n=1}^{K} \frac{d_n^{c_n} \sigma_{\text{arb}}^{c_n-1} e^{-d_n\sigma_{\text{arb}}}}{\Gamma(c_n)} \tag{11}$$

Fig. 2 illustrates the PDFs of an arbitrary (unordered) eigenvalue of $\check{\mathbf{H}}\check{\mathbf{H}}^H$ for both desired signal and interference channels generated for both models, which shows that the analytical expressions given by (8) and (11) match the simulated PDFs very well.

B. Expected Per-User Signal Power

The expected per-user signal power in (5) is:

$$\delta_{k,l} = \frac{P_{\text{T}}}{\tilde{\eta}_l \text{PL}_{k,l,l}} \mathbb{E}[|\check{\mathbf{h}}_{k,l,l} \mathbf{f}_{\text{BB}_{k,l}}|^2] \tag{12}$$

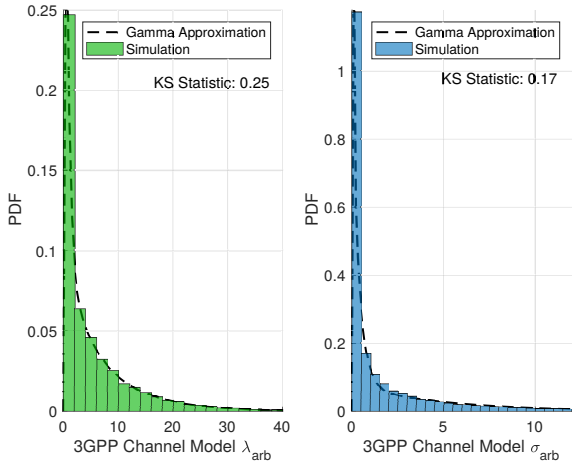
When RZF[3] precoding is employed at baseband, the un-normalized RZF precoding vector for user k in cell l, $\mathbf{f}_{\text{BB}_{k,l}}$, is the k-th column of the $N_{\text{T}}^{\text{RF}} \times K$ matrix $\mathbf{F}_{\text{BB}_l}$, such that

$$\mathbf{F}_{\text{BB}_l} = \check{\mathbf{H}}_{l,l}^H (\check{\mathbf{H}}_{l,l} \check{\mathbf{H}}_{l,l}^H + \xi_l \mathbf{I}_K)^{-1} \tag{13}$$

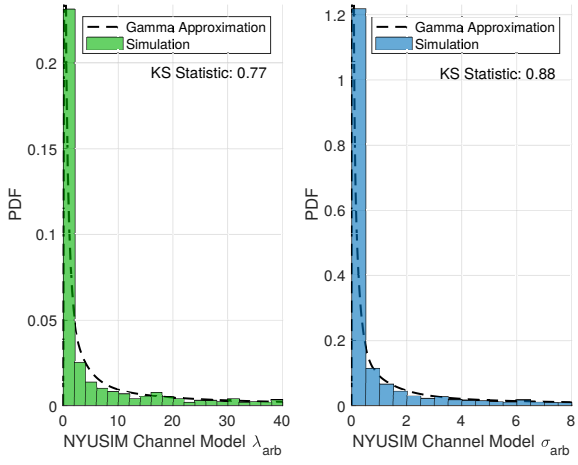
The constant $\xi_l > 0$ represents the regularization parameter specific to TP l. In this work, ξ_l is set to the following value based on [29], [34]:

$$\xi_l = \frac{KN_0}{P_{\text{T}}} \tag{14}$$

[3]Note that the performance of RZF approximates maximum ratio (MR) transmission for low SNRs and zero-forcing (ZF) for high SNRs [34], hence it is sufficient to study RZF instead of MR and ZF.



(a)



(b)

Figure 2. PDF of an arbitrary eigenvalue of $\check{\mathbf{H}}\check{\mathbf{H}}^H$ for signal channels (λ_{arb}) and interference channels (σ_{arb}) for three users per cell, using the (a) 3GPP model and (b) NYUSIM model, where $\check{\mathbf{H}}$ denotes the effective channel matrix after transmit RF precoding, i.e., $\check{\mathbf{H}} = \mathbf{H}\mathbf{F}_{\text{RF}}$.

Through an eigenvalue decomposition, we obtain $\check{\mathbf{H}}_{l,l} \check{\mathbf{H}}_{l,l}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$[4]. The entries in $\mathbf{U}$ have an isotropic distribution for NLOS conditions. For LOS conditions where there is a dominant specular component, $\mathbf{U}$ is not isotropic, but the averaging over the random angles-of-departure (AoDs)/angles-of-arrival (AoAs) in the array

[4]To facilitate the analytical study later on, a singular value decomposition (SVD) or an economy-size SVD is first performed such that $\check{\mathbf{H}}_{l,l} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^H$, which leads to $\check{\mathbf{H}}_{l,l} \check{\mathbf{H}}_{l,l}^H = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^H\mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{U}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$.

steering vectors makes $\mathbf{U}$ retain its isotropicity. Thus, the expected value in (12) over the isotropicity of $\mathbf{U}$ is expressed as [19], [20], [34]:

$$\varrho_{k,l} = \mathbb{E}[|\check{\mathbf{h}}_{k,l,l}\mathbf{f}_{\mathrm{BB}_{k,l}}|^2] = \mathbb{E}\left[\left(\sum_{a=1}^{K}\frac{\lambda_a}{\lambda_a+\xi_l}|u_{l,a}|^2\right)^2\right] \tag{15}$$

The above (15) can be further averaged over the entries of $\mathbf{U}$ and reformulated as [20], [34]:

$$\varrho_{k,l} = \frac{1}{K(K+1)}\left\{\mathbb{E}_\lambda\left[\left(\sum_{a=1}^{K}\frac{\lambda_a}{\lambda_a+\xi_l}\right)^2\right] + \mathbb{E}_\lambda\left[\sum_{a=1}^{K}\left(\frac{\lambda_a}{\lambda_a+\xi_l}\right)^2\right]\right\} \tag{16}$$

where $\mathbb{E}_\lambda[\cdot]$ represents the expectation over the eigenvalues of $\check{\mathbf{H}}_{l,l}\check{\mathbf{H}}_{l,l}^H$. Now we aim to calculate the expected values in (16) for 3GPP and NYUSIM channel models using the approximated PDFs of eigenvalues derived above. For the first expectation term in (16), it is recognized that

$$\begin{aligned}\mathbb{E}_\lambda\left[\left(\sum_{k=1}^{K}\frac{\lambda_k}{\lambda_k+\xi_l}\right)^2\right] =& \mathbb{E}_\lambda\left[\sum_{k=1}^{K}\left(\frac{\lambda_k}{\lambda_k+\xi_l}\right)^2\right] \\ &+ \mathbb{E}_\lambda\left[\sum_{a=1}^{K}\sum_{\substack{b=1\\b\neq a}}^{K}\left(\frac{\lambda_a}{\lambda_a+\xi_l}\right)\left(\frac{\lambda_b}{\lambda_b+\xi_l}\right)\right]\end{aligned} \tag{17}$$

We begin by evaluating the first term on the right-hand side of (17), yielding

$$\begin{aligned}s_l &= \mathbb{E}_\lambda\left[\sum_{k=1}^{K}\left(\frac{\lambda_k}{\lambda_k+\xi_l}\right)^2\right] = \sum_{n=1}^{K}\int_0^\infty \frac{\lambda^2}{(\lambda+\xi_l)^2}f_{\lambda_n}(\lambda)d\lambda \\ &= K\left[\int_0^\infty \frac{\lambda^2}{(\lambda+\xi_l)^2}f_{\lambda_{\mathrm{arb}}}(\lambda)d\lambda\right]\end{aligned} \tag{18}$$

where $f_{\lambda_n}(\cdot)$ denotes the approximated PDF for the n-th largest eigenvalue as expressed in (7). By utilizing the joint density of two arbitrary eigenvalues in (9), the second term on the right-hand side of (17) can be written as (19). The second expectation in (16) equals:

$$p_l = \mathbb{E}_\lambda\left[\sum_{k=1}^{K}\left(\frac{\lambda_k}{\lambda_k+\xi_l}\right)^2\right] = s_l \tag{20}$$

Therefore, the expected signal power in (12) is given by:

$$\delta_{k,l} = \left(\frac{P_{\mathrm{T}}}{\tilde{\eta}_l \mathrm{PL}_{k,l,l}}\right)\left[\frac{2s_l+\epsilon_l}{K(K+1)}\right] \tag{21}$$

in which [29]

$$\begin{aligned}\tilde{\eta}_l =& \mathbb{E}[||\mathbf{F}_{\mathrm{RF}_l}\mathbf{F}_{\mathrm{BB}_l}||_F^2] \approx \mathbb{E}_\lambda\left[\sum_{k=1}^{K}\frac{\lambda_k}{(\lambda_k+\xi_l)^2}\right] \\ =& K\int_0^\infty \frac{\lambda}{(\lambda+\xi_l)^2}f_{\lambda_{\mathrm{arb}}}(\lambda)d\lambda\end{aligned} \tag{22}$$

where the approximation stems from the fact that the array response vectors of $\mathbf{F}_{\mathrm{RF}_l}$ become orthogonal to each other as $N_{\mathrm{T}} \to \infty$, such that $\mathbf{F}_{\mathrm{RF}_l}^H\mathbf{F}_{\mathrm{RF}_l} = \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{RF}}}$ (see (21) in [29]). Through numerous numerical results we find that the eigenvalues of $\check{\mathbf{H}}_{l,l}\check{\mathbf{H}}_{l,l}^H$ are at least three orders of magnitude larger than ξ_l, thus $\frac{\lambda}{(\lambda+\xi_l)^2} \approx \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$. Consequently, (22) can be approximated as:

$$\begin{aligned}\tilde{\eta}_l =& K\int_0^\infty \frac{\lambda}{(\lambda+\xi_l)^2}f_{\lambda_{\mathrm{arb}}}(\lambda)d\lambda \approx K\int_{\lambda_{\min}}^\infty \frac{1}{\lambda}f_{\lambda_{\mathrm{arb}}}(\lambda)d\lambda \\ =& \sum_{n=1}^{K}\int_{\lambda_{\min}}^\infty \frac{b_n^{a_n}\lambda^{a_n-2}e^{-b_n\lambda}}{\Gamma(a_n)}d\lambda \\ =& \sum_{n=1}^{K}\left(\frac{b_n}{\Gamma(a_n)}\Upsilon\Big(a_n-1, b_n\lambda_{\min}\Big)\right)\end{aligned} \tag{23}$$

where $\lambda_{\min} = \min(\lambda_{\mathrm{arb}})$, and $\Upsilon\Big(a_n-1, b_n\lambda_{\min}\Big)$ is defined in (24) in which $\tilde{\Gamma}(\cdot)$ represents the upper incomplete gamma function and $E_1(\cdot)$ denotes the exponential integral [20].

C. Expected Per-User Interference Power

The expected interference power at user k in cell l in (5) is given by:

$$\begin{aligned}\varsigma_{k,l} =& \sum_{(m,i)\neq(k,l)}\frac{P_{\mathrm{T}}}{\tilde{\eta}_i \mathrm{PL}_{k,l,i}}\mathbb{E}[|\check{\mathbf{h}}_{k,l,i}\mathbf{f}_{\mathrm{BB}_{m,i}}|^2] \\ =& \frac{P_{\mathrm{T}}}{\tilde{\eta}_l \mathrm{PL}_{k,l,l}}\sum_{\substack{m=1\\m\neq k}}^{K}\mathbb{E}[|\check{\mathbf{h}}_{k,l,l}\mathbf{f}_{\mathrm{BB}_{m,l}}|^2] \\ &+ \sum_{\substack{i=1\\i\neq l}}^{L}\frac{P_{\mathrm{T}}}{\tilde{\eta}_i \mathrm{PL}_{k,l,i}}\sum_{m=1}^{K}\mathbb{E}[|\check{\mathbf{h}}_{k,l,i}\mathbf{f}_{\mathrm{BB}_{m,i}}|^2]\end{aligned} \tag{25}$$

The first term on the right-hand side of (25) denotes the inter-user interference (IUI) within the same cell, and can be evaluated as the difference between the total (signal plus intra-cell interference) power from TP l to user k in cell l and the desired signal power at user k in cell l [20]. The expected total power from TP l to user k in cell l is given by:

$$\begin{aligned}\gamma_{k,l} =& \mathbb{E}[||\check{\mathbf{h}}_{k,l,l}\mathbf{F}_{\mathrm{BB}_l}||^2] = \mathbb{E}\left[\sum_{a=1}^{K}\frac{|u_{k,a}|^2\lambda_a^2}{(\lambda_a+\xi_l)^2}\right] \\ =& \frac{1}{K}\mathbb{E}_\lambda\left[\sum_{a=1}^{K}\frac{\lambda_a^2}{(\lambda_a+\xi_l)^2}\right] = \frac{s_l}{K}\end{aligned} \tag{26}$$

where s_l is given by (18). Consequently, the intra-cell interference in (25) can be expressed as:

$$\begin{aligned}\varsigma_{k,l_{\mathrm{IUI}}} =& \frac{P_{\mathrm{T}}}{\tilde{\eta}_l \mathrm{PL}_{k,l,l}}\sum_{\substack{m=1\\m\neq k}}^{K}\mathbb{E}[|\check{\mathbf{h}}_{k,l,l}\mathbf{f}_{\mathrm{BB}_{m,l}}|^2] = \frac{P_{\mathrm{T}}}{\tilde{\eta}_l \mathrm{PL}_{k,l,l}}\gamma_{k,l} - \delta_{k,l} \\ =& \frac{P_{\mathrm{T}}}{\tilde{\eta}_l \mathrm{PL}_{k,l,l}}\left[\frac{s_l}{K} - \frac{2s_l+\epsilon_l}{K(K+1)}\right] = \frac{P_{\mathrm{T}}}{\tilde{\eta}_l \mathrm{PL}_{k,l,l}}\left[\frac{(K-1)s_l-\epsilon_l}{K(K+1)}\right]\end{aligned} \tag{27}$$

where (21) is utilized to obtain the third equality. The second term in (25) denotes the inter-cell interference, or OCI, and can be formulated as:

$$\begin{aligned}\mathbb{E}[|\check{\mathbf{h}}_{k,l,i}\mathbf{f}_{\mathrm{BB}_{m,i}}|^2] =& \mathbb{E}[\mathrm{tr}\{\check{\mathbf{h}}_{k,l,i}^H\check{\mathbf{h}}_{k,l,i}\mathbf{f}_{\mathrm{BB}_{m,i}}\mathbf{f}_{\mathrm{BB}_{m,i}}^H\}] \\ =& \mathrm{tr}\{\mathbb{E}[\check{\mathbf{h}}_{k,l,i}^H\check{\mathbf{h}}_{k,l,i}]\mathbb{E}[\mathbf{f}_{\mathrm{BB}_{m,i}}\mathbf{f}_{\mathrm{BB}_{m,i}}^H]\} \\ =& \frac{1}{K}\mathrm{tr}\{\mathbb{E}[\check{\mathbf{h}}_{k,l,i}^H\check{\mathbf{h}}_{k,l,i}]\mathbb{E}[\mathbf{F}_{\mathrm{BB}_i}\mathbf{F}_{\mathrm{BB}_i}^H]\}\end{aligned} \tag{28}$$

The second equality in (28) holds because $\check{\mathbf{h}}_{k,l,i}$ and $\mathbf{f}_{\mathrm{BB}_{m,i}}$

$$\begin{aligned}\epsilon_l =&\mathbb{E}_\lambda\left[\sum_{a=1}^{K}\sum_{\substack{b=1\\b\neq a}}^{K}\left(\frac{\lambda_a}{\lambda_a+\xi_l}\right)\left(\frac{\lambda_b}{\lambda_b+\xi_l}\right)\right]=K(K-1)\int_0^\infty\int_0^\infty\left(\frac{\lambda_a}{\lambda_a+\xi_l}\right)\left(\frac{\lambda_b}{\lambda_b+\xi_l}\right)f_{\lambda,\text{unord}}(\lambda_a,\lambda_b)d\lambda_b d\lambda_a\\ =&\sum_{n=1}^{K}\sum_{\substack{q=1\\q\neq n}}^{K}\int_0^\infty\int_0^\infty\left(\frac{\lambda_a}{\lambda_a+\xi_l}\right)\left(\frac{\lambda_b}{\lambda_b+\xi_l}\right)(\lambda_a\lambda_b)^{-1}\times\Big[\big(\phi_n(\lambda_a)\big)^2\big(\phi_q(\lambda_b)\big)^2-\phi_n(\lambda_a)\phi_q(\lambda_a)\phi_n(\lambda_b)\phi_q(\lambda_b)\Big]d\lambda_b d\lambda_a\\ =&\sum_{n=1}^{K}\sum_{\substack{q=1\\q\neq n}}^{K}\left\{\left[\int_0^\infty\left(\frac{\lambda_a}{\lambda_a+\xi_l}\right)\big(\phi_n(\lambda_a)\big)^2\lambda_a^{-1}d\lambda_a\right]^2-\left[\int_0^\infty\left(\frac{\lambda_a}{\lambda_a+\xi_l}\right)\phi_n(\lambda_a)\phi_q(\lambda_a)\lambda_a^{-1}d\lambda_a\right]^2\right\}\\ =&\sum_{n=1}^{K}\sum_{\substack{q=1\\q\neq n}}^{K}\left\{\left[\int_0^\infty\left(\frac{\lambda_a}{\lambda_a+\xi_l}\right)f_{\lambda_n}(\lambda_a)d\lambda_a\right]^2-\left[\int_0^\infty\left(\frac{\lambda_a}{\lambda_a+\xi_l}\right)\phi_n(\lambda_a)\phi_q(\lambda_a)\lambda_a^{-1}d\lambda_a\right]^2\right\}\end{aligned}\tag{19}$$

$$\Upsilon\Big(a_n-1,b_n\lambda_{\min}\Big)=\begin{cases}\tilde{\Gamma}\big(a_n-1,b_n\lambda_{\min}\big) & \text{if}\quad a_n>1\\ E_1(b_n\lambda_{\min}) & \text{if}\quad a_n=1\\ -\frac{(b_n\lambda_{\min})^{a_n-1}e^{-b_n\lambda_{\min}}}{a_n-1}+\frac{1}{a_n-1}\tilde{\Gamma}\Big(a_n,b_n\lambda_{\min}\Big) & \text{if}\quad 0<a_n<1\end{cases}\tag{24}$$

are independent, since $\mathbf{f}_{\text{BB}_{m,i}}$ is only related to $\check{\mathbf{H}}_{i,i}$ which is independent of $\check{\mathbf{h}}_{k,l,i}$ when $l\neq i$ according to (2). Note that $\mathbf{F}_{\text{BB}_i}=\check{\mathbf{H}}_{i,i}^H(\check{\mathbf{H}}_{i,i}\check{\mathbf{H}}_{i,i}^H+\xi_i\mathbf{I}_K)^{-1}$, the second expectation in (28) can be recast as:

$$\begin{aligned}\mathbb{E}[\mathbf{F}_{\text{BB}_i}\mathbf{F}_{\text{BB}_i}^H]=&\mathbb{E}[\check{\mathbf{H}}_{i,i}^H(\check{\mathbf{H}}_{i,i}\check{\mathbf{H}}_{i,i}^H+\xi_l\mathbf{I}_K)^{-2}\check{\mathbf{H}}_{i,i}]\\ =&\mathbb{E}[\check{\mathbf{H}}_{i,i}^H(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H+\xi_l\mathbf{I}_K)^{-2}\check{\mathbf{H}}_{i,i}]\\ =&\mathbb{E}[\check{\mathbf{H}}_{i,i}^H\mathbf{U}(\mathbf{\Lambda}+\xi_l\mathbf{I}_K)^{-2}\mathbf{U}^H\check{\mathbf{H}}_{i,i}]\end{aligned}\tag{29}$$

where the second equality stems from $\check{\mathbf{H}}_{i,i}\check{\mathbf{H}}_{i,i}^H=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$. For the case $N_{\text{T}}^{\text{RF}}=K$ considered in this work, it follows from singular value decomposition (SVD) that $\check{\mathbf{H}}_{i,i}=\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^H$. Thus, (29) is transformed to:

$$\begin{aligned}\mathbb{E}[\mathbf{F}_{\text{BB}_i}\mathbf{F}_{\text{BB}_i}^H]=&\mathbb{E}[\mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{U}^H\mathbf{U}(\mathbf{\Lambda}+\xi_l\mathbf{I}_K)^{-2}\mathbf{U}^H\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^H]\\ =&\mathbb{E}[\mathbf{V}\mathbf{\Lambda}^{1/2}(\mathbf{\Lambda}+\xi_l\mathbf{I}_K)^{-2}\mathbf{\Lambda}^{1/2}\mathbf{V}^H]\end{aligned}\tag{30}$$

For the first expectation in (28), one can denote $\check{\mathbf{H}}_{l,i}^H\check{\mathbf{H}}_{l,i}=\mathbf{Q}^H\mathbf{\Sigma}\mathbf{Q}$, where $\mathbf{\Sigma}=\text{diag}(\sigma_1,...,\sigma_K)$, then the trace in (28) becomes:

$$\begin{aligned}\varkappa_i=&\text{tr}\{\mathbb{E}[\check{\mathbf{h}}_{k,l,i}^H\check{\mathbf{h}}_{k,l,i}]\mathbb{E}[\mathbf{F}_{\text{BB}_i}\mathbf{F}_{\text{BB}_i}^H]\}\\ =&\frac{1}{K}\text{tr}\{\mathbb{E}[\check{\mathbf{H}}_{l,i}^H\check{\mathbf{H}}_{l,i}]\mathbb{E}[\mathbf{F}_{\text{BB}_i}\mathbf{F}_{\text{BB}_i}^H]\}\\ =&\frac{1}{K}\mathbb{E}[\text{tr}\{\mathbf{Q}^H\mathbf{\Sigma}\mathbf{Q}\mathbf{V}\mathbf{\Lambda}^{1/2}(\mathbf{\Lambda}+\xi_i\mathbf{I}_K)^{-2}\mathbf{\Lambda}^{1/2}\mathbf{V}^H\}]\\ =&\frac{1}{K}\mathbb{E}[\text{tr}\{\mathbf{V}^H\mathbf{Q}^H\mathbf{\Sigma}\mathbf{Q}\mathbf{V}\mathbf{\Lambda}^{1/2}(\mathbf{\Lambda}+\xi_i\mathbf{I}_K)^{-2}\mathbf{\Lambda}^{1/2}\}]\\ =&\frac{1}{K}\mathbb{E}\left[\sum_{k=1}^{K}\sum_{a=1}^{K}\sigma_a|w_{a,k}|^2\frac{\lambda_k}{(\lambda_k+\xi_i)^2}\right]\end{aligned}\tag{31}$$

where $w_{a,k}$ denotes the (a,k)-th entry of the unitary matrix $\mathbf{QV}$. Let $r=|w_{a,k}|^2$, then the PDF of r is given by [34]:

$$f_r(r)=(K-1)(1-r)^{K-2},\quad 0\leq r\leq 1\tag{32}$$

which implies

$$\mathbb{E}[|w_{a,k}|^2]=\int_0^1 r(K-1)(1-r)^{K-2}dr=\frac{1}{K}\tag{33}$$

Therefore,

$$\begin{aligned}\varkappa_i=&\frac{1}{K}\mathbb{E}\left[\sum_{k=1}^{K}\sum_{a=1}^{K}\sigma_a\mathbb{E}[|w_{a,k}|^2]\frac{\lambda_k}{(\lambda_k+\xi_i)^2}\right]\\ =&\frac{1}{K^2}\mathbb{E}\left[\sum_{k=1}^{K}\sum_{a=1}^{K}\sigma_a\frac{\lambda_k}{(\lambda_k+\xi_i)^2}\right]\\ =&\frac{1}{K^2}\mathbb{E}_\sigma\left[\sum_{a=1}^{K}\sigma_a\right]\mathbb{E}_\lambda\left[\sum_{k=1}^{K}\frac{\lambda_k}{(\lambda_k+\xi_i)^2}\right]\approx\frac{1}{K}\tilde{\eta}_i\int_0^\infty\sigma f_{\sigma_{\text{arb}}}(\sigma)d\sigma\end{aligned}\tag{34}$$

where the approximation follows from (22). Based on (11), the integral in (34) can be recast as:

$$\begin{aligned}\varpi=&\int_0^\infty\sigma f_{\sigma_{\text{arb}}}(\sigma)d\sigma=\frac{1}{K}\sum_{n=1}^{K}\int_0^\infty\sigma\frac{d_n^{c_n}\sigma^{c_n-1}e^{-d_n\sigma}}{\Gamma(c_n)}d\sigma\\ =&\frac{1}{K}\sum_{n=1}^{K}\frac{\Gamma(c_n+1)}{d_n\Gamma(c_n)}\end{aligned}\tag{35}$$

Plugging (35) and (23) into (34) results in:

$$\varkappa_i=\frac{1}{K}\tilde{\eta}_i\varpi\tag{36}$$

Combining the results in (27), (28), and (36), the expected per-user interference in (25) is:

$$\begin{aligned}\varsigma_{k,l}=&\frac{P_{\text{T}}}{\tilde{\eta}_l\text{PL}_{k,l,l}}\left[\frac{(K-1)s_l-\epsilon_l}{K(K+1)}\right]+\sum_{\substack{i=1\\i\neq l}}^{L}\frac{P_{\text{T}}\varkappa_i}{\tilde{\eta}_i\text{PL}_{k,l,i}}\\ =&\frac{P_{\text{T}}}{\tilde{\eta}_l\text{PL}_{k,l,l}}\left[\frac{(K-1)s_l-\epsilon_l}{K(K+1)}\right]+\sum_{\substack{i=1\\i\neq l}}^{L}\frac{P_{\text{T}}\varpi}{K\text{PL}_{k,l,i}}\end{aligned}\tag{37}$$

D. Expected Per-User SINR and Expected Per-Cell Spectral Efficiency

The expected per-user SINR in (5) can be now be expressed as a function of $\delta_{k,l}$ and $\varsigma_{k,l}$, i.e.

$$\mathbb{E}[\mathrm{SINR}_{k,l}] \approx \frac{\delta_{k,l}}{\varsigma_{k,l} + N_0} \tag{38}$$

In the derivation of (38) from (4), four approximations are made (excluding the approximations on eigenvalue densities), i.e. in (5), (22), (23), and (34). The approximations in (22) and (34) are tight (usually with a relative error within $\pm 5\%$). The approximation in (23) is larger than the true value if $\lambda_{\min} = 0$ and can equal the true value for some $\lambda_{\min}$ larger than 0. The aggregate tightness of the approximations can be seen from the subsequent numerical results. The expected ergodic spectral efficiency for cell l can be approximated from $\mathbb{E}[\mathrm{SINR}_{k,l}]$ as:

$$\mathbb{E}[R_l] = \mathbb{E}\left[\sum_{k=1}^{K} \log_2(1 + \mathrm{SINR}_{k,l})\right] \approx \sum_{k=1}^{K} \log_2\left(1 + \mathbb{E}[\mathrm{SINR}_{k,l}]\right) \tag{39}$$

Note that (39) applies to the full range of SNR and arouses an approximation instead of an upper bound via Jensen's inequality, as the value of $\mathbb{E}[\mathrm{SINR}_{k,l}]$ is itself an approximation [20], [35]. The generality of the results derived above is worth mentioning. The results derived above are applicable for any link SNR and channel model, including potential special cases such as the presence of a fixed LOS component in the channel (as long as the necessary eigenvalue densities are known). If there is a change in the transmit or the receive dimension, then the analytical approach is still valid, but the approximated gamma distributed eigenvalue densities need to be re-fitted. This is because of the mathematical complexity of finding closed-form expressions when using such advanced channel models as the 3GPP and NYUSIM models, as well as the additional presence of RF beamforming.

E. Numerical Results and Discussion

The accuracy of the derived expected per-user SINR in (38) and expected per-cell spectral efficiency in (39) is evaluated in this subsection through comparison with numerical results for the three-cell homogeneous network introduced in Section II with three users per cell and the HBF architecture in Fig. 1(b). In the simulations, the number of TP antennas was 256, the number of UE antennas was one, the number of RF chains at each TP was three, and the cell radius was 200 m. For each channel model, 500 random channel realizations were carried out for each set of parameter settings. The cumulative distribution functions (CDFs) of simulated and approximated expected per-user SINR and per-cell spectral efficiency are illustrated in Figs. 3(a) and 3(b). The expected SINR and spectral efficiency curves denote (5) (for simulated CDF) or (38) (for approximated CDF) and (39), respectively, where the expectation is taken over the small-scale fading with the distribution representing the randomness in user location (i.e., large-scale path loss and shadow fading). It is observed from Figs. 3(a) and 3(b) that the derived SINR and spectral efficiency approximations closely follow the corresponding simulated values over the entire probability range. Furthermore, the expected per-user SINR as a function of the cell-edge SNR is illustrated in Fig. 3(c), where the average is performed globally over both the link gains and the multipath fading. As shown by Fig. 3(c), for both models, the analytical expressions remain sufficiently accurate over the entire cell-edge SNR range investigated, revealing the tightness and generality of the derived SINR approximations.

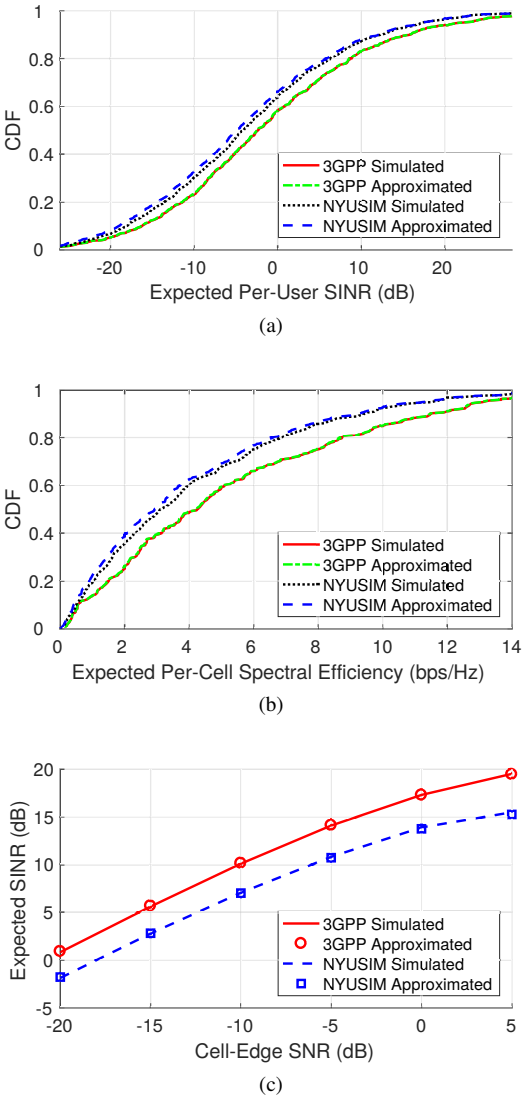


Figure 3. (a) CDFs of expected per-user SINR and (b) CDFs of expected per-cell spectral efficiency, with a cell radius of 200 m, a cell-edge SNR of 5 dB, and three users per cell. (c) Expected per-user SINR versus cell-edge SNR with a cell radius of 200 m and three users per cell.

IV. Multi-Cell Multi-User Multi-Stream Hybrid Beamforming

In this section, we investigate multi-cell multi-user HBF schemes when multiple streams are transmitted from each TP to each of its serving users. As the analytical derivation for the expected per-user SINR is extremely cumbersome for the multi-stream-per-user case, we resort to numerical simulations to evaluate the performance of various multi-cell HBF approaches. Furthermore, it is found through simulations that the spectral efficiency (not shown here due to space limitations) obtained by using the TP HBF architecture in Fig. 1(b) is lower than using the structure shown in Fig. 4, due to the increased IUI in the former. Therefore, we focus on the HBF architecture in Fig. 4 for multi-stream-per-user beamforming, where at each TP the $N_{\mathrm{T}}^{\mathrm{RF}}$ RF chains are divided into K subsets with $M_{\mathrm{T}}^{\mathrm{RF}}$ RF chains in each subset, such that the total number of TP RF chains is $N_{\mathrm{T}}^{\mathrm{RF}} = KM_{\mathrm{T}}^{\mathrm{RF}}$. Additionally, there is a baseband digital precoder which is connected to a subset dedicated to a user

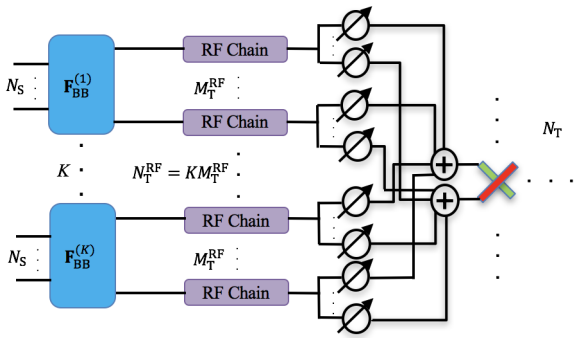


Figure 4. Multi-cell HBF architecture at the TP in each cell. N_{S} denotes the number of data streams per user in each cell, K is the number of users in each cell, $N_{\mathrm{T}}^{\mathrm{RF}}$ represents the total number of RF chains at each TP, $M_{\mathrm{T}}^{\mathrm{RF}}$ is the number of RF chains connected to the baseband precoder for one user, and N_{T} denotes the number of TP antenna elements in each cell.

in the home cell. At each user, there are N_{R} antennas with $N_{\mathrm{R}}^{\mathrm{RF}}$ RF chains and one baseband processing unit. Note that the architecture in Fig. 4 is not suitable for the single-stream-per-user case discussed in Section III, since for the single-stream case, there would be only one RF chain connected with one baseband unit for each user at the serving TP, which becomes analog beamforming, rather than analog-digital HBF. For TP i and user k in cell l, the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ downlink channel is denoted as $\mathbf{H}_{k,l,i}$, the $N_{\mathrm{T}} \times M_{\mathrm{T}}^{\mathrm{RF}}$ RF precoding matrix is $\mathbf{F}_{\mathrm{RF}_{k,l}}$, and the $M_{\mathrm{T}}^{\mathrm{RF}} \times N_{\mathrm{S}}$ baseband precoding matrix is $\mathbf{F}_{\mathrm{BB}_{k,l}}$. The $N_{\mathrm{R}} \times N_{\mathrm{R}}^{\mathrm{RF}}$ RF combining matrix and the $N_{\mathrm{R}}^{\mathrm{RF}} \times N_{\mathrm{S}}$ baseband combining matrix is $\mathbf{W}_{\mathrm{RF}_{k,l}}$ and $\mathbf{W}_{\mathrm{BB}_{k,l}}$, respectively. The received signal at user k in cell l is formulated as:

$$\begin{aligned}\mathbf{y}_{k,l} = &\underbrace{\sqrt{\frac{P_{\mathrm{t}}}{\eta_{k,l}\mathrm{PL}_{k,l,l}}}\mathbf{W}_{\mathrm{BB}_{k,l}}^H\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}\mathbf{F}_{\mathrm{RF}_{k,l}}\mathbf{F}_{\mathrm{BB}_{k,l}}\mathbf{s}_{k,l}}_{\text{Desired Signal}} \\ &+ \underbrace{\sum_{\substack{(m,i)\\ \neq(k,l)}}\sqrt{\frac{P_{\mathrm{t}}}{\eta_{m,i}\mathrm{PL}_{k,l,i}}}\mathbf{W}_{\mathrm{BB}_{k,l}}^H\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,i}\mathbf{F}_{\mathrm{RF}_{m,i}}\mathbf{F}_{\mathrm{BB}_{m,i}}\mathbf{s}_{m,i}}_{\text{Interference}} \\ &+ \underbrace{\mathbf{W}_{\mathrm{BB}_{k,l}}^H\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{n}_{k,l}}_{\text{Noise}}\end{aligned} \tag{40}$$

where P_{t} represents the transmit power for each user in Watts, and is assumed to be constant regardless of the user number per cell and the cell radius. $\mathrm{PL}_{k,l,i}$ denotes the large-scale distance-dependent path loss in Watts, including shadow fading, from TP i to user k in cell l, $\eta_{k,l} = ||\mathbf{F}_{\mathrm{RF}_{k,l}}\mathbf{F}_{\mathrm{BB}_{k,l}}||_F^2$ is a scaling factor to satisfy the per-user transmit power constraint $||\sqrt{P_{\mathrm{t}}}\mathbf{F}_{\mathrm{RF}_{k,l}}\mathbf{F}_{\mathrm{BB}_{k,l}}/\sqrt{\eta_{k,l}}||_F^2 = P_{\mathrm{t}}$. $\mathbf{s}_{k,l}$ represents the desired transmitted signal for user k in cell l with $\mathbb{E}[\mathbf{s}_{k,l}\mathbf{s}_{k,l}^H] = \mathbf{I}_{N_{\mathrm{S}}}$, and $\mathbf{n}_{k,l} \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I}_{N_{\mathrm{R}}})$ is circularly symmetric complex Gaussian noise with variance N_0. The spectral efficiency of user k in cell l is calculated as in (41), where $\mathbf{D}$ in (41) is given by:

$$\mathbf{D} = \sum_{\substack{(m,i)\\ \neq(k,l)}}\frac{P_{\mathrm{t}}}{\eta_{m,i}\mathrm{PL}_{k,l,i}}\mathbf{H}_{k,l,i}\mathbf{F}_{\mathrm{RF}_{m,i}}\mathbf{F}_{\mathrm{BB}_{m,i}}\mathbf{F}_{\mathrm{BB}_{m,i}}^H\mathbf{F}_{\mathrm{RF}_{m,i}}^H\mathbf{H}_{k,l,i}^H \tag{42}$$

1) Baseline Case — No Coordination Among Cells: Let us first consider the interference-ignorant baseline case where there is no TP coordination among cells. Assuming only local CSI is available at each TP, a reasonable precoding scheme is eigenmode transmission. Let us define the effective channel matrix $\check{\mathbf{H}}_{k,l,k,l} \in \mathbb{C}^{N_{\mathrm{R}}^{\mathrm{RF}} \times M_{\mathrm{T}}^{\mathrm{RF}}}$ for user k in cell l as $\frac{1}{\sqrt{\mathrm{PL}_{k,l,l}}}\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}\mathbf{F}_{\mathrm{RF}_{k,l}}$, where the RF precoding and combining matrices $\mathbf{F}_{\mathrm{RF}_{k,l}}$ and $\mathbf{W}_{\mathrm{RF}_{k,l}}$ are designed such that $||\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}\mathbf{F}_{\mathrm{RF}_{k,l}}||_F^2$ is maximized. The RF beamforming approach in Eqs. (12)-(14) proposed in [36] is applied to obtain $\mathbf{F}_{\mathrm{RF}_{k,l}}$ and $\mathbf{W}_{\mathrm{RF}_{k,l}}$, in which the codebooks for $\mathbf{F}_{\mathrm{RF}_{k,l}}$ and $\mathbf{W}_{\mathrm{RF}_{k,l}}$ consist of the TP and UE antenna array response vectors corresponding to the AoDs and AoAs associated with the desired user, respectively [27]. Specifically, maximization of the matrix norm, $||\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}\mathbf{F}_{\mathrm{RF}_{k,l}}||_F^2$, is summarized as follows [36]:

i Initialize $\mathbf{W}_{\mathrm{RF}_{k,l}}$ by maximizing $||\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}||_F^2$ [36].

ii Compute $\mathbf{H}_{k,l,l}^H\mathbf{W}_{\mathrm{RF}_{k,l}}\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}$ and set $\mathbf{F}_{\mathrm{RF}_{k,l}} = \frac{1}{\sqrt{N_{\mathrm{T}}}}e^{j\angle\mathbf{U}(:,1:M_{\mathrm{T}}^{\mathrm{RF}})}$, where $\angle\mathbf{U}(:,1:M_{\mathrm{T}}^{\mathrm{RF}})$ denotes the phases of the eigenvectors associated with the $M_{\mathrm{T}}^{\mathrm{RF}}$ largest eigenvalues of $\mathbf{H}_{k,l,l}^H\mathbf{W}_{\mathrm{RF}_{k,l}}\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}$ [36].

iii Compute $\mathbf{H}_{k,l,l}\mathbf{F}_{\mathrm{RF}_{k,l}}\mathbf{F}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}^H$ and set $\mathbf{W}_{\mathrm{RF}_{k,l}} = \frac{1}{\sqrt{N_{\mathrm{R}}}}e^{j\angle\mathbf{V}(:,1:N_{\mathrm{R}}^{\mathrm{RF}})}$, where $\angle\mathbf{V}(:,1:N_{\mathrm{R}}^{\mathrm{RF}})$ denotes the phases of the eigenvectors associated with the $N_{\mathrm{R}}^{\mathrm{RF}}$ largest eigenvalues of $\mathbf{H}_{k,l,l}\mathbf{F}_{\mathrm{RF}_{k,l}}\mathbf{F}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}^H$ [36].

The baseband precoding matrix $\mathbf{F}_{\mathrm{BB}_{k,l}}$ is composed of the dominant N_{S} right singular vectors obtained from the SVD of $\check{\mathbf{H}}_{k,l,k,l}$, and the baseband combining matrix $\mathbf{W}_{\mathrm{BB}_{k,l}}$ is constituted by the dominant N_{S} left singular vectors obtained from the SVD of $\check{\mathbf{H}}_{k,l,k,l}\mathbf{F}_{\mathrm{BB}_{k,l}}$.

2) SLAB: In SLAB, the RF beamforming is aimed at mitigating the dominant leakage to all the other users while enhancing the strength of the desired signal, improved based on the leakage-suppressing and signal-maximizing HBF in [23] by adding UE beamforming to enhance signal. First, the $(K-1)N_{\mathrm{R}} \times N_{\mathrm{T}}$ cascaded channel matrix consisting of all the channel matrices in cell l except the one for user k in cell l is obtained through CSI exchange among TPs as:

$$\begin{aligned}\overline{\mathbf{H}}_{k,l} = \Bigg[&\frac{1}{\sqrt{\mathrm{PL}_{1,l,l}}}\mathbf{H}_{1,l,l}^T, ..., \frac{1}{\sqrt{\mathrm{PL}_{k-1,l,l}}}\mathbf{H}_{k-1,l,l}^T, \\ &\frac{1}{\sqrt{\mathrm{PL}_{k+1,l,l}}}\mathbf{H}_{k+1,l,l}^T, ..., \frac{1}{\sqrt{\mathrm{PL}_{K,l,l}}}\mathbf{H}_{K,l,l}^T\Bigg]^T\end{aligned} \tag{43}$$

The columns of RF beamforming matrices at each TP and UE are selected from pre-defined beamforming codebooks that consist of antenna array response vectors $\mathbf{a}_{\mathrm{T}}$ and $\mathbf{a}_{\mathrm{R}}$ at the TP and UE, respectively. The matrix $\mathbf{A}_{\mathrm{T}}$ and $\mathbf{A}_{\mathrm{R}}$ are composed of $\mathbf{a}_{\mathrm{T}}$'s and $\mathbf{a}_{\mathrm{R}}$'s corresponding to the AoDs and AoAs associated with the desired user, respectively [27]. The first $M_{\mathrm{T}}^{\mathrm{RF}} - 1$ columns in the RF precoding matrix $\mathbf{F}_{\mathrm{RF}_{k,l}}$ is chosen from $\mathbf{A}_{\mathrm{T}}$ such that $||\overline{\mathbf{H}}_{k,l}\mathbf{F}_{\mathrm{RF}_{k,l}}(:,1:M_{\mathrm{T}}^{\mathrm{RF}}-1)||_F^2$ is minimized, whose physical meaning is using the first $M_{\mathrm{T}}^{\mathrm{RF}} - 1$ RF precoding vectors to suppress the leakage to all the other users in cell l. The remaining column in $\mathbf{F}_{\mathrm{RF}_{k,l}}$ and the columns in $\mathbf{W}_{\mathrm{RF}_{k,l}}$ are used to maximize $||\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{H}_{k,l,l}\mathbf{F}_{\mathrm{RF}_{k,l}}(:,M_{\mathrm{T}}^{\mathrm{RF}})||_F^2$, the physical meaning of which is utilizing the remaining RF precoding vector and all the RF combining vectors to maximize the desired signal power to user k in cell l. The baseband precoding matrix $\mathbf{F}_{\mathrm{BB}_{k,l}}$ and the baseband combining matrix $\mathbf{W}_{\mathrm{BB}_{k,l}}$ are designed in the manner as in the baseline. The key difference between SLAB and the signal-to-leakage-plus-noise-

$$R_{k,l} = \log_2 \Bigg| \mathbf{I}_{N_\mathrm{S}} + \frac{P_\mathrm{t}}{\eta_{k,l}\mathrm{PL}_{k,l,l}} \big(\mathbf{W}^H_{\mathrm{BB}_{k,l}} \mathbf{W}^H_{\mathrm{RF}_{k,l}} (N_0 \mathbf{I}_{N_\mathrm{R}} + \mathbf{D}) \mathbf{W}_{\mathrm{RF}_{k,l}} \mathbf{W}_{\mathrm{BB}_{k,l}}\big)^{-1} \times \mathbf{W}^H_{\mathrm{BB}_{k,l}} \breve{\mathbf{H}}_{k,l,k,l} \mathbf{F}_{\mathrm{BB}_{k,l}} \mathbf{F}^H_{\mathrm{BB}_{k,l}} \breve{\mathbf{H}}^H_{k,l,k,l} \mathbf{W}_{\mathrm{BB}_{k,l}} \Bigg| \tag{41}$$

ratio (SLNR)-based approach to be introduced next is that SLAB utilizes the *RF* beamforming to mitigate leakage and enhance signal, while the SLNR-based approach uses the *baseband* precoder to maximize SLNR. Their performance difference will be shown via numerical results in Section V.

3) SLNR-Based Precoding: Directly maximizing the SINR involves a challenging optimization problem with coupled variables, thus the SLNR is utilized as an alternative optimization criterion. In the SLNR-based TP coordination, the effective channel matrix $\breve{\mathbf{H}}_{m,i,k,l} \in \mathbb{C}^{N_\mathrm{R}^\mathrm{RF} \times M_\mathrm{T}^\mathrm{RF}}$ is defined as $\frac{1}{\sqrt{\mathrm{PL}_{m,i,l}}} \mathbf{W}^H_{\mathrm{RF}_{m,i}} \mathbf{H}_{m,i,l} \mathbf{F}_{\mathrm{RF}_{k,l}}$, and the $(KL-1)N_\mathrm{R}^\mathrm{RF} \times M_\mathrm{T}^\mathrm{RF}$ leakage matrix is:

$$\tilde{\mathbf{H}}_{k,l} = \Big[\breve{\mathbf{H}}^T_{1,1,k,l}, ..., \breve{\mathbf{H}}^T_{k-1,l,k,l}, \breve{\mathbf{H}}^T_{k+1,l,k,l}, ..., \breve{\mathbf{H}}^T_{K,L,k,l}\Big]^T \tag{44}$$

The RF precoding and RF combining matrices $\mathbf{F}_{\mathrm{RF}_{k,l}}$ and $\mathbf{W}_{\mathrm{RF}_{k,l}}$ are designed to maximize $||\mathbf{W}^H_{\mathrm{RF}_{k,l}} \mathbf{H}_{k,l,l} \mathbf{F}_{\mathrm{RF}_{k,l}}||^2_F$, where $\mathbf{F}_{\mathrm{RF}_{k,l}}$ and $\mathbf{W}_{\mathrm{RF}_{k,l}}$ are obtained in the same manner as in the baseline case. The baseband precoding matrix $\mathbf{F}_{\mathrm{BB}_{k,l}}$ is designed to maximize the SLNR as follows. The expected received signal power prior to the baseband combining process is given by $\mathbb{E}\big[\frac{P_\mathrm{t}}{\eta_{k,l}} \mathbf{s}^H_{k,l} \mathbf{F}^H_{\mathrm{BB}_{k,l}} \breve{\mathbf{H}}^H_{k,l,k,l} \breve{\mathbf{H}}_{k,l,k,l} \mathbf{F}_{\mathrm{BB}_{k,l}} \mathbf{s}_{k,l}\big]$, the expected leakage power is expressed as $\mathbb{E}\big[\sum_{(m,i)\neq(k,l)} \frac{P_\mathrm{t}}{\eta_{k,l}} \mathbf{s}^H_{k,l} \mathbf{F}^H_{\mathrm{BB}_{k,l}} \breve{\mathbf{H}}^H_{m,i,k,l} \breve{\mathbf{H}}_{m,i,k,l} \mathbf{F}_{\mathrm{BB}_{k,l}} \mathbf{s}_{k,l}\big]$, and the expected noise power is $\mathbb{E}\big[\mathbf{n}^H_{k,l} \mathbf{W}_{\mathrm{RF}_{k,l}} \mathbf{W}^H_{\mathrm{RF}_{k,l}} \mathbf{n}_{k,l}\big]$. The SLNR is hence formulated as in (45) [9], where $\tilde{\mathbf{H}}_{k,l}$ is given by (44), and the first equality in (45) holds since $\mathbb{E}[\mathbf{s}_{k,l}\mathbf{s}^H_{k,l}] = \mathbf{I}_{N_\mathrm{S}}$ and $\mathbb{E}[\mathbf{n}_{k,l}\mathbf{n}^H_{k,l}] = N_0 \mathbf{I}_{N_\mathrm{R}}$. The optimal $\mathbf{F}_{\mathrm{BB}_{k,l}}$ that maximizes the SLNR in (45) can be derived similarly to the precoding matrix in [9] and is composed of the leading N_S columns of $\mathbf{T}_{k,l}$ which contains the generalized eigenvectors of the pair $\big\{\breve{\mathbf{H}}^H_{k,l,k,l} \breve{\mathbf{H}}_{k,l,k,l},\ \tilde{\mathbf{H}}^H_{k,l} \tilde{\mathbf{H}}_{k,l} + \gamma \mathbf{I}_{M_\mathrm{T}^\mathrm{RF}}\big\}$, where γ satisfies:

$$\mathrm{tr}(\gamma \mathbf{F}^H_{\mathrm{BB}_{k,l}} \mathbf{F}_{\mathrm{BB}_{k,l}}) = \frac{\eta_{k,l}}{P_\mathrm{t}} N_0\ \mathrm{tr}(\mathbf{W}_{\mathrm{RF}_{k,l}} \mathbf{W}^H_{\mathrm{RF}_{k,l}}) \tag{46}$$

$\mathbf{W}_{\mathrm{BB}_{k,l}}$ is designed as a matched filter at the receiver [9]:

$$\mathbf{W}_{\mathrm{BB}_{k,l}} = \frac{\breve{\mathbf{H}}_{k,l,k,l} \mathbf{F}_{\mathrm{BB}_{k,l}}}{||\breve{\mathbf{H}}_{k,l,k,l} \mathbf{F}_{\mathrm{BB}_{k,l}}||_F} \tag{47}$$

4) RZF: The fourth HBF strategy studied in this paper is the HBF based on RZF transmission, which has the same RF precoding and RF combining procedures as the baseline and SLNR-based approaches. In RZF, the effective channel for user k in cell l after RF precoding and combining is denoted as the $N_\mathrm{R}^\mathrm{RF} \times M_\mathrm{T}^\mathrm{RF}$ matrix $\breve{\mathbf{H}}_{m,i,k,l}$ defined as

$$\breve{\mathbf{H}}_{m,i,k,l} = \frac{1}{\sqrt{\mathrm{PL}_{m,i,l}}} \mathbf{W}^H_{\mathrm{RF}_{m,i}} \mathbf{H}_{m,i,l} \mathbf{F}_{\mathrm{RF}_{k,l}} \tag{48}$$

and the $KLN_\mathrm{R}^\mathrm{RF} \times M_\mathrm{T}^\mathrm{RF}$ concatenated effective channel matrix is:

$$\tilde{\mathbf{H}}_{k,l} = [\breve{\mathbf{H}}^T_{1,1,k,l}, ..., \breve{\mathbf{H}}^T_{k,l,k,l}, ..., \breve{\mathbf{H}}^T_{K,L,k,l}]^T \tag{49}$$

If $N_\mathrm{S} = N_\mathrm{R}^\mathrm{RF}$, then the baseband precoding matrix can be set as the $N_\mathrm{S}(K(l-1)+k-1)+1$th to the $N_\mathrm{S}(K(l-1)+k)$th columns of $\overline{\mathbf{F}_{\mathrm{BB}}}$ yielded by the RZF transmission matrix:

$$\overline{\mathbf{F}_{\mathrm{BB}}} = \tilde{\mathbf{H}}^H_{k,l} \left(\tilde{\mathbf{H}}_{k,l} \tilde{\mathbf{H}}^H_{k,l} + \frac{KN_0}{P_\mathrm{t}} \mathbf{I}\right)^{-1} \tag{50}$$

where the regularization factor $\frac{KN_0}{P_\mathrm{t}}$ is set to maximize the SINR of the desired user in a single cell based on Eq. (30) in [34]. The optimal regularization factor for the multi-cell scenario is extremely challenging to derive and is outside the scope of this paper and left for future work. Note that RZF entails the dimension requirement $N_\mathrm{S} = N_\mathrm{R}^\mathrm{RF}$, which can be satisfied by turning off some receive RF chains when necessary [23].

V. Simulation Results and Analysis

Using the multi-cell MU-MIMO HBF procedures proposed above and the system layout and settings demonstrated in Section II, spectral efficiency is studied using both 3GPP [21] and NYUSIM [22] models via MATLAB simulations. It is assumed that there are N_R^RF RF chains at each UE, and each TP communicates with each UE via N_S ($N_\mathrm{S} \leq N_\mathrm{R}^\mathrm{RF}$) data streams. For each channel model, 400 random channel realizations were carried out for the three-user-per-cell case, while 100 random channel realizations were carried out for the 12-user-per-cell case. In each channel realization, UE locations in each cell are randomly and uniformly generated with 2D T-R separation distances ranging from 10 m to the cell radius (i.e., 50 m or 200 m).

The CDFs of per-user spectral efficiency in the three-cell MU-MIMO system are illustrated in Fig. 5 for different cell radii and user numbers with two steams per user. Fig. 5 shows that for both models, the SLNR-based CoMP HBF outperforms all the other HBF schemes in most cases, revealing its effectiveness in suppressing both intra-cell and inter-cell interference and noise. Another distinguishing feature is that non-CoMP SLAB appears more effective in NYUSIM than in 3GPP as the dominant leakage is stronger, and yields even higher spectral efficiency than the SLNR-based CoMP method. This implies that CoMP does not necessarily outperform non-CoMP approaches in sparse spatial channels like NYUSIM, especially for UEs located closer to the TP. NYUSIM predicts higher spectral efficiency as compared to the 3GPP model, likely due to the stronger two dominant eigenmodes per user yielded by NYUSIM channel matrices. Moreover, by comparing Figs. 5(a) and 5(c), or Figs. 5(b) and 5(d), we see that for the majority (about 70%-90%) of the users, the spectral efficiency for the 200 m cell radius is lower than the 50 m cell radius for all the proposed HBF schemes with the same user number per cell and the same transmit power per user, except for the peak spectral efficiency. This indicates that path loss/noise, rather than interference, dictates the spectral efficiency, since the 200 m cell radius corresponds to weaker interference but has lower spectral efficiency in most cases.

Next, we consider the case where each TP communicates with each of its home-cell users via one, two, and four data streams per user. Fig. 6 depicts the 5%, 50%, and 95% CDF points of the per-user spectral efficiency for both models for one to four streams

$$
\begin{aligned}
\mathrm{SLNR} \approx & \frac{\mathbb{E}\left[\frac{P_{\mathrm{t}}}{\eta_{k,l}}\mathbf{s}_{k,l}^H\mathbf{F}_{\mathrm{BB}_{k,l}}^H\breve{\mathbf{H}}_{k,l,k,l}^H\breve{\mathbf{H}}_{k,l,k,l}\mathbf{F}_{\mathrm{BB}_{k,l}}\mathbf{s}_{k,l}\right]}{\mathbb{E}\left[\sum\limits_{(m,i)\neq(k,l)}\frac{P_{\mathrm{t}}}{\eta_{k,l}}\mathbf{s}_{k,l}^H\mathbf{F}_{\mathrm{BB}_{k,l}}^H\breve{\mathbf{H}}_{m,i,k,l}^H\breve{\mathbf{H}}_{m,i,k,l}\mathbf{F}_{\mathrm{BB}_{k,l}}\mathbf{s}_{k,l}\right]+\mathbb{E}\left[\mathbf{n}_{k,l}^H\mathbf{W}_{\mathrm{RF}_{k,l}}\mathbf{W}_{\mathrm{RF}_{k,l}}^H\mathbf{n}_{k,l}\right]} \\
= & \frac{\mathrm{tr}(\frac{P_{\mathrm{t}}}{\eta_{k,l}}\mathbf{F}_{\mathrm{BB}_{k,l}}^H\breve{\mathbf{H}}_{k,l,k,l}^H\breve{\mathbf{H}}_{k,l,k,l}\mathbf{F}_{\mathrm{BB}_{k,l}})}{\mathrm{tr}(\sum\limits_{\substack{(m,i)\\\neq(k,l)}}\frac{P_{\mathrm{t}}}{\eta_{k,l}}\mathbf{F}_{\mathrm{BB}_{k,l}}^H\breve{\mathbf{H}}_{m,i,k,l}^H\breve{\mathbf{H}}_{m,i,k,l}\mathbf{F}_{\mathrm{BB}_{k,l}})+N_0\mathrm{tr}(\mathbf{W}_{\mathrm{RF}_{k,l}}\mathbf{W}_{\mathrm{RF}_{k,l}}^H)} \\
= & \frac{\mathrm{tr}(\mathbf{F}_{\mathrm{BB}_{k,l}}^H\breve{\mathbf{H}}_{k,l,k,l}^H\breve{\mathbf{H}}_{k,l,k,l}\mathbf{F}_{\mathrm{BB}_{k,l}})}{\mathrm{tr}(\mathbf{F}_{\mathrm{BB}_{k,l}}^H\tilde{\mathbf{H}}_{k,l}^H\tilde{\mathbf{H}}_{k,l}\mathbf{F}_{\mathrm{BB}_{k,l}})+\frac{\eta_{k,l}}{P_{\mathrm{t}}}N_0\mathrm{tr}(\mathbf{W}_{\mathrm{RF}_{k,l}}\mathbf{W}_{\mathrm{RF}_{k,l}}^H)}
\end{aligned}
\tag{45}
$$

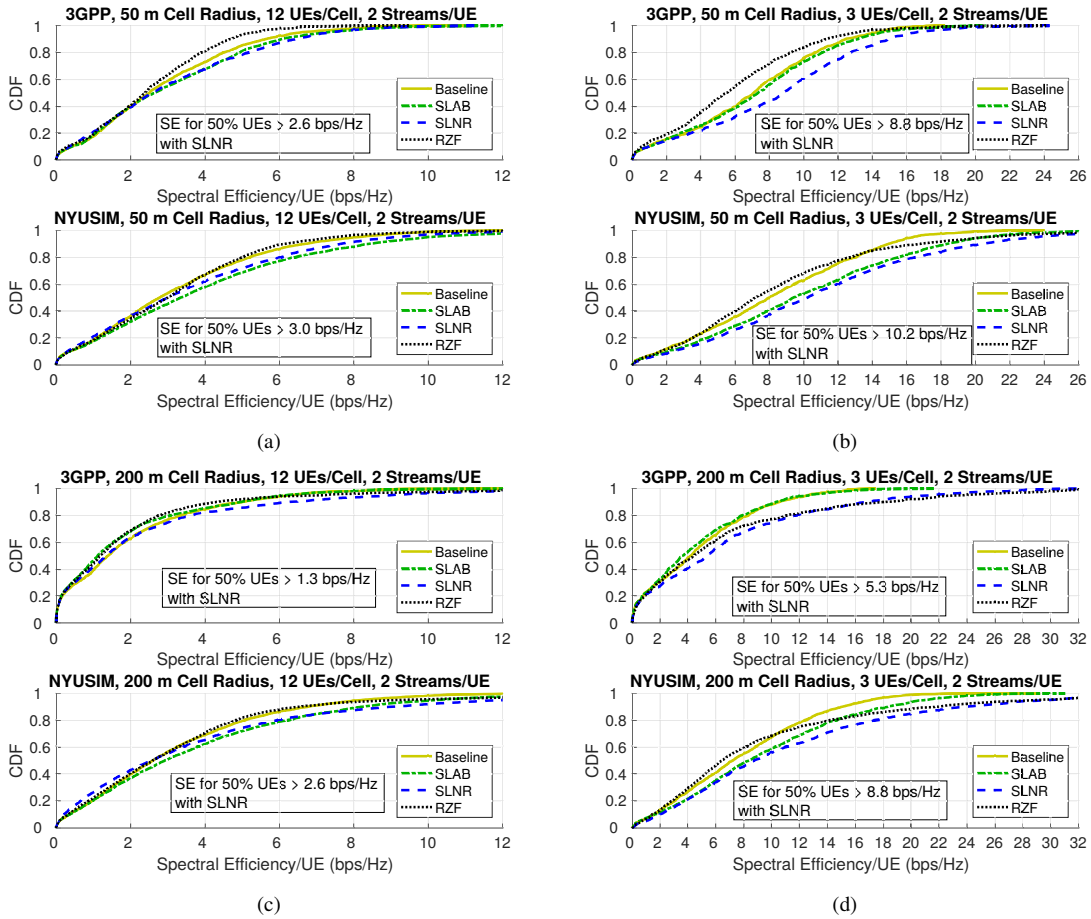


Figure 5. CDFs of the spectral efficiency per user with (a) a 50 m cell radius and 12 users per cell, (b) a 50 m cell radius and three users per cell, (c) a 200 m cell radius and 12 users per cell, and (d) a 200 m cell radius and three users per cell, in the three-cell multi-user MIMO system. Each TP has four RF chains per user, and 48 and 12 TP RF chains for 12 and three users.

with a cell radius of 50 m and 12 users per cell. As revealed by Fig. 6, for the one-stream case, SLNR and RZF yield the highest and comparable spectral efficiency using both channel models. In contrast, for the two-stream and four-stream cases, non-CoMP SLAB provides comparable or even better performance than CoMP SLNR and RZF, especially for non-cell-edge users, indicating that SLAB is more capable of suppressing inter-stream interference, and that coordinated scheduling/beamforming may only be needed for cell-edge users.

Besides spectral efficiency, energy efficiency is also an important performance metric for wireless systems [3], [4]. In fact, the original motivation to consider HBF in [27] was to reduce hardware, complexity, and power consumption — to thereby improve energy efficiency. To investigate energy efficiency of mmWave systems using HBF and the 3GPP and NYUSIM channel models, we compare the energy efficiency using SLNR HBF for corresponding to the spectral efficiency shown in Fig. 5(a), with a 100 MHz RF bandwidth, where power consumptions of the RF components in this table are based on [37]. Table III lists the energy efficiency comparison results, which demonstrates that NYUSIM generally yields higher energy efficiency.

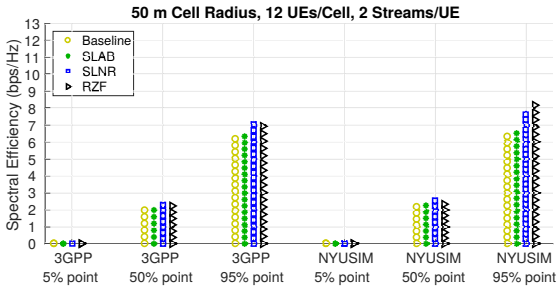
VI. Conclusions

In this paper, we considered multi-cell multi-user communication in mmWave systems, derived analytical expressions for expected SINR and spectral efficiency for the single-stream-per-user case, and

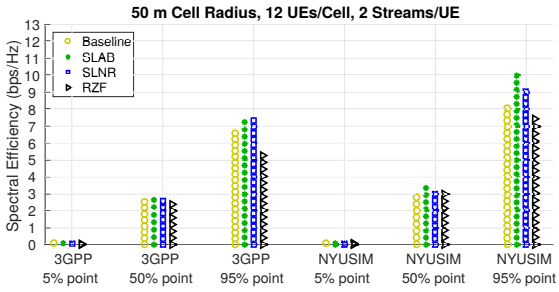
Table III
ENERGY EFFICIENCY COMPARISON BETWEEN 3GPP AND NYUSIM MODELS USING SLNR HBF FOR CORRESPONDING TO THE SPECTRAL EFFICIENCY SHOWN IN FIG. 5(A), WITH A 100 MHZ RF BANDWIDTH. POWER CONSUMPTIONS OF THE RF COMPONENTS IN THIS TABLE ARE BASED ON [37]. P DENOTES POWER CONSUMPTION.

TP	**RF Chains #→P**	**Phase Shifters #→P**	**PAs #→P**	**Splitters #→P**	**Combiners #→P**	**DACs #→P**
	4→163.2 mW	1024→2048 mW	4→6622.6 mW	4→78 mW	256→4992 mW	4→3200 mW
UE	**RF Chains #→P**	**Phase Shifters #→P**	**LNAs #→P**	**Splitters #→P**	**Combiners #→P**	**ADCs #→P**
	4→163.2 mW	32→64 mW	4→156 mW	8→156 mW	4→78 mW	4→3200 mW

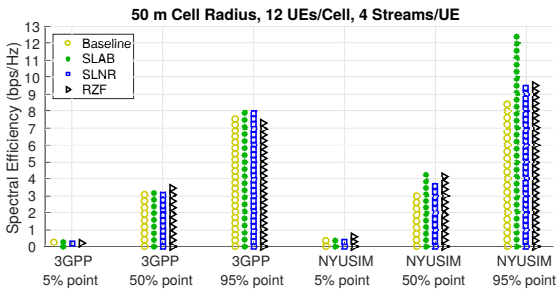
Beamforming Approach	**Channel Model**	**50% CDF Spectral Efficiency (bps/Hz)**	**Energy Efficiency (Mbits/J)**
SLNR (Fig. 5(a))	3GPP	2.6	12.4
	NYUSIM	3.0	14.3



(a)



(b)



(c)

Figure 6. CDFs of the per-user spectral efficiency of the three-cell multi-user MIMO system using the HBF algorithms proposed in this paper for 3GPP [21] and NYUSIM [22] channel models for the cases of (a) two streams, and (b) four streams per user.

proposed and compared four HBF approaches for the multi-stream-per-user case based on the assumption that base stations in different cells have full CSI and can exchange the CSI, but not the user data, among each other, such that they can take into account both intra-cell and inter-cell interference when designing precoding matrices. Numerical results show that the derived analytical expected SINR and spectral efficiency have good accuracy and analytical tractability. Non-CoMP HBF methods (e.g., SLAB) can provide comparable or even higher spectral efficiency than CoMP based on coordinated scheduling/beamforming in most cases, thus CoMP may only be needed for cell-edge users. Moreover, the behaviors of the four proposed multi-stream HBF approaches are affected by the model used, and the interference and SNR level proportional to the cell radius, the number of users per cell, and the number of streams per user.

APPENDIX

Eigenvalue distribution in (9)*:* The joint density of the ordered eigenvalues $\lambda_1 \geq ... \geq \lambda_K$ of $\check{\mathbf{H}}_{l,l}\check{\mathbf{H}}_{l,l}^H$ is given by [14]:

$$f_{\boldsymbol{\lambda},ordered}(\lambda_1,...,\lambda_K) = A^{-1}\prod_{n=1}^{K} f_{\lambda_n}(\lambda_n)\prod_{n<j}^{K}(\lambda_n-\lambda_j)^2, \quad \lambda_1 \geq ... \geq \lambda_K \geq 0 \tag{51}$$

where A is a normalizing factor. The unordered eigenvalues then have the density [14]:

$$f_{\boldsymbol{\lambda}}(\lambda_1,...,\lambda_K) = (K!A)^{-1}\prod_{n=1}^{K} f_{\lambda_n}(\lambda_n)\prod_{n<j}^{K}(\lambda_n-\lambda_j)^2 \tag{52}$$

Note that $\prod_{n<j}^{K}(\lambda_n-\lambda_j)$ is the determinant of a Vandermonde matrix [14]. By applying the Gram-Schmidt orthogonalization procedure to the sequence $1, \lambda, ..., \lambda^{K-1}$ in the space of real-valued functions with the orthogonality relationship:

$$\int_0^{\infty} \phi_n(\lambda)\phi_j(\lambda)\lambda^{-1}d\lambda = \delta_{nj} \tag{53}$$

(52) can be transformed to:

$$f_{\boldsymbol{\lambda}}(\lambda_1,...,\lambda_K) = C\sum_{\alpha,\beta}(-1)^{\mathrm{per}(\alpha)+\mathrm{per}(\beta)}\prod_{n}\phi_{\alpha_n}(\lambda_n)\phi_{\beta_n}(\lambda_n)\lambda_n^{-1} \tag{54}$$

where the sum is over all possible permutations α, β of $\{1,...,K\}$, and $\mathrm{per}(\cdot)$ denotes the sign of the permutation. Integrating the right

hand side of (54) over $\lambda_2, ..., \lambda_K$, we obtain:

$$\begin{aligned} f(\lambda_1) =& C\sum_{\alpha,\beta}(-1)^{\mathrm{per}(\alpha)+\mathrm{per}(\beta)}\phi_{\alpha_1}(\lambda_1)\phi_{\beta_1}(\lambda_1)\lambda_1^{-1}\prod_{n\geq 2}\delta_{\alpha_n\beta_n} \\ =& C(K-1)!\sum_{n=1}^{K}\left(\phi_n(\lambda_1)\right)^2\lambda_1^{-1} \\ =& \frac{(K-1)!}{K!}\sum_{n=1}^{K}\left(\phi_n(\lambda_1)\right)^2\lambda_1^{-1} = \frac{1}{K}\sum_{n=1}^{K}\left(\phi_n(\lambda_1)\right)^2\lambda_1^{-1} \end{aligned} \tag{55}$$

where the third equality follows from the fact that $\left(\phi_n(\lambda_1)\right)^2\lambda_1^{-1}$ integrates to unity and hence C must equal $1/K!$. Comparing (55) with (8), we observe that

$$\phi_n(\lambda) = \sqrt{\frac{b_n^{a_n}\lambda^{a_n}e^{-b_n\lambda}}{\Gamma(a_n)}} \tag{56}$$

Integrating the right hand side of (54) over $\lambda_3, ..., \lambda_K$, we obtain the joint density in (9):

$$\begin{aligned} f(\lambda_1,\lambda_2) =& C\sum_{\alpha,\beta}(-1)^{\mathrm{per}(\alpha)+\mathrm{per}(\beta)}\phi_{\alpha_1}(\lambda_1)\phi_{\beta_1}(\lambda_1)\lambda_1^{-1} \\ & \times \phi_{\alpha_2}(\lambda_2)\phi_{\beta_2}(\lambda_2)\lambda_2^{-1}\prod_{n\geq 3}\delta_{\alpha_n\beta_n} \\ =& \frac{(K-2)!}{K!}\sum_{n=1}^{K}\sum_{\substack{q=1\\q\neq n}}^{K}(\lambda_1\lambda_2)^{-1}\Big[\left(\phi_n(\lambda_1)\right)^2\left(\phi_q(\lambda_2)\right)^2 \\ & - \phi_n(\lambda_1)\phi_q(\lambda_1)\phi_q(\lambda_2)\phi_n(\lambda_2)\Big] \\ =& \frac{1}{K(K-1)}\sum_{n=1}^{K}\sum_{\substack{q=1\\q\neq n}}^{K}(\lambda_1\lambda_2)^{-1}\Big[\left(\phi_n(\lambda_1)\right)^2\left(\phi_q(\lambda_2)\right)^2 \\ & - \phi_n(\lambda_1)\phi_q(\lambda_1)\phi_n(\lambda_2)\phi_q(\lambda_2)\Big] \end{aligned} \tag{57}$$

REFERENCES

[1] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[2] M. Shafi *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[3] A. Mesodiakaki *et al.*, "Energy- and spectrum-efficient user association in millimeter-wave backhaul small-cell networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1810–1821, Feb. 2017.

[4] J. N. Murdock and T. S. Rappaport, "Consumption factor and power-efficiency factor: A theory for evaluating the energy efficiency of cascaded communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 2, pp. 221–236, Feb. 2014.

[5] H. Yang and T. L. Marzetta, "Massive MIMO with max-min power control in line-of-sight propagation environment," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 4685–4693, Nov. 2017.

[6] 3GPP, "Coordinated multi-point operation for LTE physical layer aspects," 3rd Generation Partnership Project (3GPP), TR 36.819 V11.2.0, Sep. 2013.

[7] D. Lee *et al.*, "Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 148–155, Feb. 2012.

[8] S. Schwarz and M. Rupp, "Exploring coordinated multipoint beamforming strategies for 5G cellular," *IEEE Access*, vol. 2, pp. 930–946, 2014.

[9] M. Sadek *et al.*, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1711–1721, May 2007.

[10] H. J. Yang *et al.*, "Opportunistic downlink interference alignment for multi-cell MIMO networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1533–1548, Mar. 2017.

[11] D. Maamari *et al.*, "Coverage in mmWave cellular networks with base station co-operation," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2981–2994, Apr. 2016.

[12] N. A. Muhammad *et al.*, "Multi-cell coordination via disjoint clustering in dense millimeter wave cellular networks," in *Proceedings of the IEEE International Conference on Communications*, 2017, pp. 1–6.

[13] G. Zhu *et al.*, "Hybrid beamforming via the kronecker decomposition for the millimeter-wave massive MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2097–2114, Sep. 2017.

[14] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov.-Dec. 1999.

[15] G. Alfano *et al.*, "Capacity of MIMO channels with one-sided correlation," in *Proceedings of the IEEE Eight Int. Symp. on Spread Spectrum Tech. and App.*, 2004, pp. 515–519.

[16] ——, "Mutual information and eigenvalue distribution of MIMO ricean channels," in *Proceedings of the Int. Symposium on Info. Theory and Applications*, 2004, pp. 1–6.

[17] M. Chiani and A. Zanella, "Joint distribution of an arbitrary subset of the ordered eigenvalues of wishart matrices," in *Proceedings of the IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2008, pp. 1–6.

[18] A. Zanella *et al.*, "On the marginal distribution of the eigenvalues of wishart matrices," *IEEE Transactions on Communications*, vol. 57, no. 4, pp. 1050–1060, Apr. 2009.

[19] H. Tataria *et al.*, "Performance and analysis of downlink multiuser MIMO systems with regularized zero-forcing precoding in Ricean fading channels," in *Proceedings of the IEEE International Conference on Communications*, 2016, pp. 1–7.

[20] ——, "On the general analysis of coordinated regularized zero-forcing precoding: An application to two-tier small-cell networks," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3133–3150, Jul. 2017.

[21] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), TR 38.901 V14.3.0, Dec. 2017.

[22] S. Sun *et al.*, "A novel millimeter-wave channel simulator and applications for 5G wireless communications," in *Proceedings of the IEEE International Conference on Communications*, 2017, pp. 1–7.

[23] ——, "Hybrid beamforming for 5G millimeter-wave multi-cell networks," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Apr. 2018, pp. 589–596.

[24] 3GPP, "Technical specification group radio access network; Study on 3D channel model for LTE (Release 12)," 3rd Generation Partnership Project (3GPP), TR 36.873 V12.2.0, Jun. 2015.

[25] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[26] E. G. Larsson *et al.*, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[27] O. E. Ayach *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[28] T. S. Rappaport *et al.*, "5G channel model with improved accuracy and efficiency in mmWave bands," *IEEE 5G Tech Focus*, vol. 1, no. 1, Mar. 2017.

[29] J. Noh *et al.*, "Zero-forcing based hybrid beamforming for multi-user millimeter wave systems," *IET Communications*, vol. 10, no. 18, pp. 2670–2677, Dec. 2016.

[30] Q. Zhang *et al.*, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 966–981, Oct. 2014.

[31] S. Jin *et al.*, "Ordered eigenvalues of complex noncentral wishart matrices and performance analysis of SVD MIMO systems," in *Proceedings of the IEEE International Symposium on Information Theory*, 2006, pp. 1564–1568.

[32] P. G. Moschopoulos, "The distribution of the sum of independent gamma random variables," *Ann. Inst. Statist. Math. (Part A)*, vol. 37, pp. 541–544, 1985.

[33] G. Castro *et al.*, "Outdoor-to-indoor empirical path loss models: Analysis for pico and femto cells in street canyons," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 542–545, Aug. 2017.

[34] C. B. Peel *et al.*, "A vector-perturbation technique for near-capacity multi-antenna multiuser communication-part I: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley Sons 1991, vol. 3.

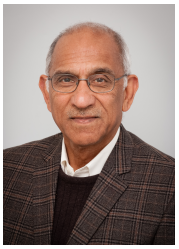
[36] N. Song *et al.*, "Coordinated hybrid beamforming for millimeter wave multi-user massive MIMO systems," in *Proceedings of the IEEE Global Communications Conference*, 2016, pp. 1–6.

[37] W. B. Abbas *et al.*, "Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADCs," *IEEE Trans-*

actions on Wireless Communications, vol. 16, no. 12, pp. 8131–8146, Dec. 2017.

Shu Sun (S'13) received the B.S. degree in applied physics from Shanghai Jiao Tong University, Shanghai, China, in 2012, and the Ph.D. degree in electrical engineering from New York University (NYU) Tandon School of Engineering, Brooklyn, NY, USA, in 2018, with a topic on fifth-generation (5G) millimeter-wave (mmWave) wireless communications under the supervision of Prof. Theodore S. Rappaport. She received the 2017 Paul Baran Young Scholar Award from the Marconi Society and the 2018 Dante Youla Award for graduate research excellence in Electrical and Computer Engineering in NYU. She has authored or coauthored more than 40 technical papers in the field of mmWave wireless communications. She is now a systems engineer in Intel Corporation.

Theodore S. Rappaport (S'83-M'84-SM'91-F'98) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1982, 1984, and 1987, respectively. He is the David Lee/Ernst Weber Professor of electrical and computer engineering with the New York University Tandon School of Engineering, New York University (NYU), Brooklyn, NY, USA, and the Founding Director of the NYU WIRELESS Research Center. He founded major wireless research centers with the Virginia Polytechnic Institute and State University (MPRG), The University of Texas at Austin (WNCG), and NYU (NYU WIRELESS) and founded two wireless technology companies that were sold to publicly traded firms, and has served as an advisor to many other successful companies. He is a highly sought-after technical consultant having testified before the U.S. Congress and having served ITU. He has advised over 100 students, has over 100 patents issued and pending, and has authored or co-authored several books, including the best-seller Wireless Communications: Principles and Practice-Second Edition (Prentice Hall, 2002). His latest book Millimeter Wave Wireless Communications (Pearson/Prentice Hall, 2015) is the first comprehensive text on the subject.



Mansoor Shafi (S69-M82-SM87-F93-LF16) received the B.Sc. (Eng.) and Ph.D. degrees in electrical engineering from the University of Engineering and Technology Lahore and The University of Auckland in 1970 and 1979, respectively. From 1975 to 1979, he was a Junior Lecturer with The University of Auckland, he then joined the New Zealand Post Office, that later evolved to Telecom NZ, and recently to Spark New Zealand. He is currently a Telecom Fellow (Wireless at Spark NZ) and an Adjunct Professor with the School of Engineering, Victoria University. He is a Delegate of NZ to the meetings of ITU-R and APT and has contributed to a large number of wireless communications standards. His research interests include radio propagation, the design and performance analysis for wireless communication systems, especially antenna arrays, MIMO, cognitive radio, and massive MIMO and mmWave systems. He has authored over 100 papers in these areas. He has coshared two IEEE prize winning papers: the IEEE Communications Society, Best Tutorial Paper Award, 2004 (co-shared with D. Gesbert, D.-S. Shiu, A. Naguib, and P. Smith) for the paper, From Theory to Practice: An overview of MIMO Space Time Coded Wireless Systems, IEEE JSAC, April 2003, and the IEEE Donald G Fink Award 2011, (co shared with A. Molisch and L. J. Greenstein), for their paper in IEEE Proceedings April 2009, Propagation Issues for Cognitive Radio. Dr. Shafi has also received the IEEE Communications Society Public Service Award, 1992 For Leadership in the Development of Telecommunications in Pakistan and Other Developing Countries, and was made a member of the New Zealand Order of Merit, Queens Birthday Honors 2013, For Services to Wireless Communications. He has been a Co-Guest Editor for three previous JSAC editions, the IEEE Proceedings, and the IEEE Communications Magazine, and a Co-Chair of ICC 2005 Wireless Communications Symposium, and has held various editorial and TPC roles in the IEEE journals and conferences.

Harsh Tataria (S'13-M'17) received the B.E. (Honors) and Ph.D. degrees in Electronic Engineering from Victoria University of Wellington, New Zealand, in 2013 and 2017, respectively. In May 2017, he acquired a Research Fellowship with the Institute of Electronics, Communications, and Information Technology (ECIT), Queens University Belfast, U.K. From May to July 2018, he held a Visiting Faculty Member appointment in the Wireless Devices and Systems (WiDeS) Group at the Viterbi School of Engineering, University of Southern California, Los Angeles, USA. He has been on several technical program committiees for flagship IEEE ComSoc conferences, such as ICC and GLOBECOM. His research interests include the design of low-complexity microwave and millimeter-wave transceivers, as well as radio propagation modeling and measurement techniques.