On the Performance of Random Beamforming in Sparse Millimeter Wave Channels

Gilwon Lee, *Student Member, IEEE*, Youngchul Sung, *Senior Member, IEEE*, and Marios Kountouris, *Senior Member, IEEE*

***Abstract*—The performance of random beamforming (RBF) with partial channel state information (CSI) feedback is investigated here for millimeter wave (mmWave) multiuser multiple-input single-output (MU-MISO) downlink systems using the uniform random multipath (UR-MP) channel model. In particular, the required number of users in the cell for RBF to yield linear sum rate scaling with respect to the number of transmit antennas is identified under the UR-MP channel model for different levels of channel sparsity. Then, the problem of user scheduling in MU-MISO downlink is considered when the number of users in the cell is not sufficiently large (sparse user regime) for the system to operate in the identified sufficient user regime. By exploiting the sparsity of mmWave radio channels, several user scheduling algorithms based on beam aggregation with reasonable amount of feedback are proposed for the sparse user regime. Our numerical results show that the proposed algorithms yield very good sum-rate performance in sparse mmWave channels.**

***Index Terms*—mmWave, multiuser MIMO, random beamforming, uniform random multipath channel model, user scheduling.**

I. Introduction

THE unprecedented demand for higher data rates, mainly due to the rapid growth of mobile traffic and the use of smartphones and tablets, are creating challenges for wireless service providers to operate in larger bandwidth. Millimeter wave (mmWave) multiple-input multiple-output (MIMO) operating in the band of 30-300GHz has recently been under research to reveal its potential as a promising technology for obtaining high data rates for 5G wireless communication. Radio propagation in the mmWave band has different characteristics from that in the lower radio bands used in current 3G/4G wireless communication. The propagation in the mmWave band is highly directional with large path loss and very few multi-paths, which are sparse in the departure and/or arrival angle domain [2]–[7]. To perform highly directional downlink beamforming for compensating for the large path loss in the mmWave band, accurate channel estimation and channel state information (CSI) feedback to the base station (BS) are required [8]. However, channel estimation and CSI feedback induce heavy system overhead in MU-MIMO downlink systems.

One way to overcome the burden of channel estimation and CSI feedback is the multi-beam random beamforming (RBF) method proposed in [9]. Therein, the BS generates a set of random orthogonal beams, selects the user that experiences the largest signal-to-interference-plus-noise ratio (SINR) for each beam, and then transmits data streams to the selected users associated with the set of orthogonal beams. Hence, RBF does not require full CSI but yields reasonable (asymptotically optimal) performance with partial CSI feedback (scalar SINR) by exploiting the multiuser diversity (MUD) gain in the network. Due to the aforementioned advantages, RBF and the associated MUD gain have been investigated extensively over the past decade, especially under the rich scattering channel model that assumes independent and identically distributed (i.i.d.) Rayleigh fading for each element of the channel vector with small-scale MIMO in the low frequency band in mind [9]–[13]. It has been shown that under the i.i.d. Rayleigh fading channel model, the RBF sum rate $\mathcal{R}_{RBF}$ in a K-user MISO downlink system with M transmit antennas scales as [9]

$$\mathcal{R}_{RBF} \simeq \begin{cases} M \log\log K, & \text{for fixed } M, \\ cM, & \text{for } M = O(\log K), \end{cases} \tag{1}$$

where $x \simeq y$ indicates that $\lim_{K\to\infty} x/y = 1$, and c is a positive constant. Furthermore, Sharif and Hassibi showed that [9]

$$\lim_{K\to\infty} \frac{\mathcal{R}_{RBF}}{M} = 0, \tag{2}$$

when $\lim_{K\to\infty} \frac{\log K}{M} = 0$, i.e., $K = o(e^{c'M})$ for some constant c'. In other words, in order to achieve linear sum rate scaling by RBF with respect to (w.r.t.) the number of transmit antennas, the number of users in the cell should increase exponentially w.r.t. the number of transmit antennas. This result puts a pessimistic view on the RBF method to be used in large-scale antenna arrays as in massive MIMO [13], [14]. However, this pessimistic result is based on the assumption of i.i.d. Rayleigh fading channel, which captures rich scattering in small-scale MIMO systems at low frequency bands and is not suitable for large-scale MIMO systems in the mmWave band with very few multi-paths and quasi-optical propagation [2]–[5]. Recently, Lee *et al.* [15] considered the RBF method in mmWave MU-MISO downlink and analyzed its performance under the *uniform random single-path (UR-SP) channel*

Manuscript received June 01, 2015; revised January 25, 2016; accepted January 28, 2016. Date of publication February 03, 2016; date of current version April 14, 2016. This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2A10060852) and in part by "The Cross-Ministry Giga KOREA Project" of The Ministry of Science, ICT, and Future Planning, Korea [GK14N0100, 5G mobile communication system development based on mmWave]. A preliminary version of this work was presented in SPAWC 2015 [1]. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Nuria Gonzlez-Prelcic.

G. Lee and Y. Sung are with the School of Electrical Engineering, KAIST, Daejeon 305-701, Korea (e-mail: gwlee@kaist.ac.kr; ysung@ee.kaist.ac.kr).

M. Kountouris is with the Mathematical and Algorithmic Sciences Laboratory, France Research Center, Huawei Technologies Co. Ltd., Paris, France (e-mail: marios.kountouris@huawei.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2016.2524999

TABLE I
SUFFICIENT NUMBER OF USERS FOR LINEAR SUM RATE SCALING W.R.T. M

Scattering richness	Sufficient number of users
L fixed	$K \simeq M$ (linear)
$L = \log(M)$	$K \simeq M^{1+c_u}$ (polynomial)
$L = M^{\beta}$, $0 \leq \beta < 1$	$K \simeq e^{c_u' M^{\beta}}$ (sub-exponential)
$L = M$ or faster	$K \simeq e^{c_u' M}$ (exponential)

which models the quasi-optical propagation characteristics of the mmWave band [4], [5], [15]. They showed very different behaviors of the RBF method under the UR-SP channel model from those under the conventional i.i.d. Rayleigh fading MIMO channel model. That is, linear sum rate scaling by RBF w.r.t. the number of transmit antennas may be achieved with increasing only linearly the number of users in the cell w.r.t. the number of transmit antennas under the UR-SP channel model [15]. This result puts an optimistic prospect for the RBF scheme to be used in mmWave massive MIMO.

In this paper, we analyze the performance of RBF for large antenna arrays in general channels beyond the UR-SP channel model. For that, we first propose a new channel model, coined as *the uniform random multi-path (UR-MP) channel model*, which captures the multiple propagation paths from the BS to each user and is still analytically tractable. Under the UR-MP channel model, the number L of multi-paths determines the sparsity (or richness) of channel scattering. Based on the proposed UR-MP channel model, we answer the following fundamental question regarding the performance of RBF: "how many users in the cell are sufficient for RBF in order to achieve linear sum rate scaling w.r.t. the number M of transmit antennas?" for different levels of channel sparsity, namely fixed and finite L (this case includes the UR-SP channel model), logarithmically increasing $L = \log M$, fractional power function $L = M^{\beta}$ with $0 \leq \beta < 1$, and linear function or faster $L = M^{\beta}$ with $\beta \geq 1$ [1] as M tends to infinity. The results are summarized in Table I.

When the number of users in the cell is sufficient high (dense user regime) to achieve linear sum rate scaling, RBF can be used to operate the system exhibiting satisfactory performance. However, in the sparse user regime, the number of users in the cell may not be sufficiently high for the system to operate under a sufficient number of users, although the sufficient number of users for linear sum rate scaling in sparse mmWave channels is much smaller than that required for rich scattering channels. This scenario may be relevant in small cell networks, in which case the conventional RBF may not be efficient. Thus, in this paper, we next consider the problem of user scheduling for MU-MISO downlink with sparse mmWave channels in the sparse user regime. By exploiting the sparsity of mmWave radio channels and insights from our previous work [15], we propose several user scheduling methods suitable for MU-MISO downlink with sparse mmWave channels. The proposed scheduling methods implement maximum ratio transmission (MRT) beamforming or equal gain combining (EGC) transmit beamforming with a reasonable amount of feedback for the sparse user regime. Furthermore, a two-phase feedback-based scheduling algorithm is proposed based on these methods to enhance the performance. The proposed methods yield good performance in sparse mmWave channels.

[1]This case corresponds to the i.i.d. Rayleigh fading channel model. See Section II-A.

Notation: Vectors and matrices are written in boldface with matrices in capitals. All vectors are column vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^H$ and $\mathbf{A}^T$ indicate the conjugate transpose and transpose of $\mathbf{A}$, respectively. $\mathbf{I}_M$ is the $M \times M$ identity matrix, and $\mathbf{1}$ denotes a column vector with all one elements. $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that random vector $\mathbf{x}$ is complex Gaussian distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\theta \sim \text{Unif}(a, b)$ means that θ is uniformly distributed for $\theta \in [a, b)$. $\mathbb{E}[\cdot]$ denotes statistical expectation. $\mathbb{C}$ is the set of complex numbers. $\iota = \sqrt{-1}$.

II. SYSTEM MODEL

We consider a MU-MISO downlink network composed of a BS equipped with a uniform linear array (ULA) of M transmit antennas and K single-antenna users. We consider the RBF scheme [9] that chooses S users among the K users in the cell and broadcasts independent data streams to the S selected users with beam vectors $\mathbf{w}_1, \cdots, \mathbf{w}_S$. (The RBF scheme will be explained in detail in the next section.) During the data transmission period the received signal y_k at user k is given by

$$y_k = \sqrt{\rho}\mathbf{h}_k^H \mathbf{w}_k x_k + \sqrt{\rho}\sum_{j \neq k} \mathbf{h}_k^H \mathbf{w}_j x_j + n_k, \tag{3}$$

where

$$\mathbf{h}_k = [h_{k,1}, h_{k,2}, \cdots, h_{k,M}]^T \tag{4}$$

is the channel vector from the BS to user k, $\mathbf{w}_k$ is the unit-norm beamforming vector for user k, x_k is the data symbol for user k, and $n_k \sim \mathcal{CN}(0, 1)$ is the additive circularly-symmetric Gaussian noise at user k. Here, we assume $x_k \sim \mathcal{CN}(0, 1)$ and further assume that the total transmit power P_t is equally distributed to each data stream. Hence, ρ is given by $\rho = \frac{P_t}{S}$ and $\rho\mathbb{E}[||\mathbf{h}_k||^2]$ is the average received signal-to-noise ratio (SNR) for user k since the noise variance is one.

A. The Proposed Channel Model

The most widely considered channel model for (4) is the i.i.d. Rayleigh fading channel model. Under this channel model, each element $h_{k,m}$ in $\mathbf{h}_k$ is given by

$$h_{k,m} \overset{i.i.d.}{\sim} \mathcal{CN}(0, 1), m = 1, \cdots, M. \tag{5}$$

The i.i.d. Rayleigh channel model has been used to capture rich scattering environments in conventional lower band MIMO communication and most MUD gain analysis was conducted under this channel model and its variants such as correlated fading [9], [11]–[14], [16]. However, this channel model is

not relevant to the mmWave band in which radio propagation is highly directional and there exist very few multi-paths [6], [17]–[19]. Recently, *the UR-SP channel model* was proposed in [4], [5] to model highly directional wireless channels in the mmWave band. Under the UR-SP model, the channel vector of user k is given by [4], [5], [15]

$$\mathbf{h}_k = \alpha_k \sqrt{M}\mathbf{a}(\theta_k), \text{ for } k = 1, \cdots, K, \tag{6}$$

where the single-path link gain α_k is i.i.d. Gaussian-distributed, i.e. $\alpha_k \overset{i.i.d.}{\sim} \mathcal{CN}(0,1)$, the normalized direction θ_k for user k is i.i.d. with $\theta_k \overset{i.i.d.}{\sim} \text{Unif}[-1,1]$, and $\mathbf{a}(\theta)$ is the ULA steering vector given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}}[1, e^{-\iota\pi\theta}, \cdots, e^{-\iota\pi(M-1)\theta}]^T. \tag{7}$$

Although the UR-SP channel model is analytically tractable and captures well the highly directional radio propagation in the mmWave band, it is an extreme channel model since it captures the limiting case of only a single path. Since our goal in this paper is to analyze the performance of RBF under more realistic, non-extreme, channel model, i.e. models capturing settings in-between the UR-SP and i.i.d. Rayleigh fading MIMO channels, we propose a new channel model that can capture both channel models and represents different levels of channel scattering sparsity by extending the UR-SP channel model. In the proposed channel model, the channel vector $\mathbf{h}_k$ of user k is given by

$$\mathbf{h}_k = \sqrt{\frac{M}{L}} \sum_{i=1}^{L} \alpha_{k,i}\mathbf{a}(\theta_{k,i}), \tag{8}$$

where L is the number of multi-paths, and $\alpha_{k,i} \overset{i.i.d.}{\sim} \mathcal{CN}(0,1)$ and $\theta_{k,i} \overset{i.i.d.}{\sim} \text{Unif}[-1,1]$ are the path gain and normalized angle-of-departure (AoD)[2] of the i-th path of the channel vector $\mathbf{h}_k$, respectively. Here, the normalized AoD $\theta \in [-1,1]$ is related to the physical AoD $\phi \in [-\pi/2, \pi/2]$ as

$$\theta = \frac{2d\sin(\phi)}{\lambda}, \tag{9}$$

where d and λ are the distance between two adjacent antenna elements and the carrier wavelength, respectively. Note that the normalized AoD is the sine function of the actual AoD. We assume here the critical spatial sampling of $\frac{d}{\lambda} = 0.5$. Note that the channel vector $\mathbf{h}_k$ in the proposed channel model is the sum of L uniform random multi-paths with complex Gaussian gains. Thus, we refer to this channel model as the *UR-MP channel model*.

By varying the number L of multi-paths, the UR-MP channel model can capture various levels of channel scattering sparsity. In the simple case of $L = 1$, the proposed channel model is expressed as $\mathbf{h}_k = \sqrt{M}\alpha_{k,1}\mathbf{a}(\theta_{k,1})$, which is the UR-SP channel model (6) considered in [4], [5], [15]. In the asymptotic scenario in which the number M of transmit antennas tends to infinity to model large-scale antenna arrays, we can consider several functions for L as a function of M. We consider

- fixed and finite $L(\geq 1)$ irrespective of M,
- a logarithmic function $L = \log M$,
- a fractional power function $L = M^\beta$ with $0 \leq \beta < 1$, and
- a linear or faster function $L = M^\beta$ with $\beta \geq 1$.

Note that the first three cases correspond to sparse channel models since $L/M \to 0$ as $M \to \infty$. In particular, the UR-MP channel model with the fractional power function $L = M^\beta$ with $\beta > 0$ is useful. It captures the channel scattering sparsity with the single parameter β. When $\beta = 0$, it reduces to the UR-SP channel model. When $\beta > 1$, on the other hand, it converges to the i.i.d. Rayleigh fading channel model in the sense specified in Theorem 1.

Theorem 1: Under the UR-MP channel model with the fractional power function $L = M^\beta$, we have

$$\mathcal{L}\left(\mathbf{h}_k | \theta_{k,1}, \cdots, \theta_{k,L}\right) = \mathcal{CN}(\mathbf{0}, \mathbf{R}(\theta_{k,1}, \cdots, \theta_{k,L})), \tag{10}$$

where $\mathcal{L}\left(\mathbf{h}_k | \theta_{k,1}, \cdots, \theta_{k,L}\right)$ is the distribution of $\mathbf{h}_k$ conditioned on $(\theta_{k,1}, \cdots, \theta_{k,L})$. Furthermore, we have for any $\epsilon > 0$,

$$\Pr(|\mathbf{R}(\theta_{k,1}, \cdots, \theta_{k,L}) - \mathbf{I}|_E \geq \epsilon \mathbf{1}\mathbf{1}^T) \to 0, \tag{11}$$

as $M \to \infty$, if $\beta > 1$, where the probability is computed according to $\theta_{k,i} \overset{i.i.d.}{\sim} \text{Unif}[-1,1]$, and $|\cdot|_E$ represents the element-wise absolute value. That is, the conditional channel covariance matrix $\mathbf{R}(\theta_{k,1}, \cdots, \theta_{k,L})$ converges to $\mathbf{I}$ uniformly in elements in probability. On the other hand, when $0 \leq \beta < 1$, the conditional channel covariance matrix has rank at most $L = M^\beta < M$ and is rank-deficient.

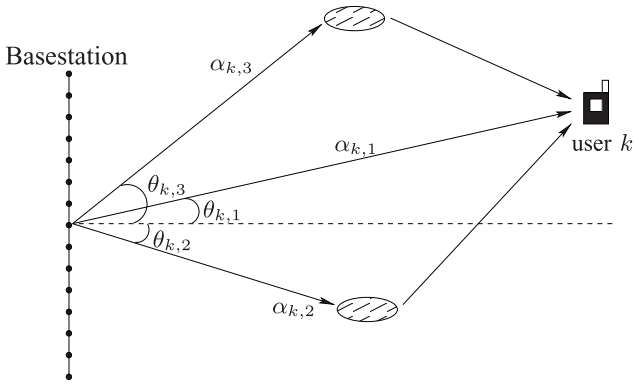

Fig. 1. The proposed uniform random multipath (UR-MP) channel model ($L = 3$): The uniform randomness is in the arrival-of-departure angle domain.

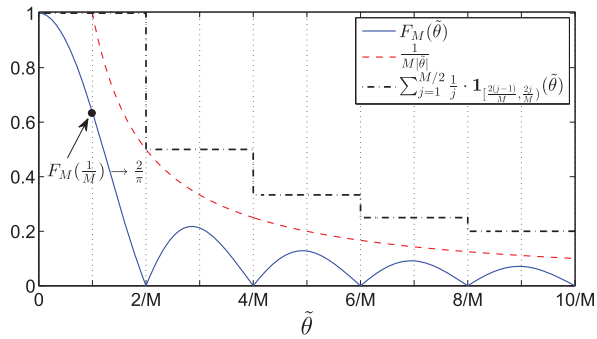

Fig. 2. $F_M(\tilde{\theta})$ in (15) when $M = 100$.

[2]Since we consider the MISO case, AoD matters. We assume that there exists a scatter or reflector at each AoD included in the model (8) to generate a propagation path from the BS and user k at that AoD, as shown in Fig. 1.

Proof: See Appendix A. ■

When $\beta \in (0, 1]$, the channel model lies somewhere in-between the UR-SP and i.i.d. Rayleigh fading channel models. Thus, under this model, the parameter β of the model monotonically represents the richness in scattering in the propagation between the BS and the considered user.

III. Performance Analysis in the Sufficient User Regime

In this section, we first explain the considered RBF scheme, which is referred to as randomly directional beamforming (RDB). Then, we identify how many users K are sufficient for RDB to achieve linear sum rate scaling w.r.t. the number M of antennas under the UR-MP channel model for different levels of channel scattering richness. Since in this section we consider the case that RDB achieves linear sum rate scaling w.r.t. the number M of antennas, we set the number S of selected users for RDB as $S = M$.

The conventional RBF scheme operates as follows [9].

S.1 *Random beam selection/generation:* The BS generates a set of random orthonormal beam vectors $\{\mathbf{w}_b\}_{b=1}^M$ that form an orthonormal basis of $\mathbb{C}^M$.

S.2 *Training:* The BS transmits the selected beam vectors $\{\mathbf{w}_b\}_{b=1}^M$ sequentially in time. Then, each user k in the cell measures the inner product $\mathbf{h}_k^H \mathbf{w}_b$ between its channel vector $\mathbf{h}_k$ and the transmitted beam $\mathbf{w}_b$ for $b = 1, 2, \cdots, M$.

S.3 *SINR computation and feedback:* At the end of the training period, each user k computes the SINR for each beam b and feeds back to the BS the maximum SINR value and the corresponding beam index. Here, SINR for user k at beam b is given by

$$\mathrm{SINR}_{k,b} = \frac{\rho |\mathbf{h}_k^H \mathbf{w}_b|^2}{1 + \rho \sum_{b' \neq b} |\mathbf{h}_k^H \mathbf{w}_{b'}|^2}. \tag{12}$$

Thus, each user feeds back one integer and one real number to the BS. Note that this SINR definition assumes that the b-th beam is assigned to user k and all other $M - 1$ beams are assigned to other users [9]. Thus, all M beams are used for data transmission in this setting.

S.4 *Scheduling and data transmission:* After feedback, the BS groups users according to their maximum SINR beam index and selects one user with maximum SINR for each beam $b = 1, \cdots, M$. After scheduling, the BS transmits data to the selected M users with the beam vectors $\{\mathbf{w}_b\}_{b=1}^M$.

In the considered RDB scheme, assuming highly-directional propagation environments as in the mmWave band, we construct the M transmit beams based on the beam direction. That is, the orthonormal beam basis $\{\mathbf{w}_b\}_{b=1}^M$ is designed as

$$\mathbf{w}_b = \mathbf{a}(\vartheta_b) = \mathbf{a}\left(\vartheta + \frac{2(b-1)}{M}\right), \text{ for } b = 1, \cdots, M, \tag{13}$$

where $\mathbf{a}(\cdot)$ is defined in (7), and $\vartheta \sim \mathrm{Unif}[-1, 1]$ or equivalently $\vartheta \sim \mathrm{Unif}[-1, -1 + \frac{2}{M}]$ is a random offset value. Thus, the RDB scheme uses M beams equispaced in the normalized angle domain with a uniform random offset. Note that in the RDB scheme, the inter-beam interference results when the user is not located at the exact boresight angle $\vartheta + \frac{2(b-1)}{M}$ but located somewhere in-between equispaced boresight beam angles.

The expected sum rate of the RDB scheme is given by [9]

$$\mathcal{R}_{sum} = \sum_{b=1}^{M} \mathcal{R}_{\kappa_b} = \sum_{b=1}^{M} \mathbb{E}\left[\log\left(1 + \mathrm{SINR}_{\kappa_b, b}\right)\right], \tag{14}$$

where $\kappa_b = \arg\max_{1 \leq k \leq K} \mathrm{SINR}_{k,b}$ and $\mathrm{SINR}_{k,b}$ is defined in (12). As seen in (3), the magnitude of the inner product of two normalized array steering vectors plays an important role in computing SINR in the RDB scheme, and is given by

$$\begin{aligned} |\mathbf{a}(\theta)^H \mathbf{a}(\vartheta)| &= \frac{1}{M} \left| \sum_{n=0}^{M-1} e^{-\iota \pi n (\vartheta - \theta)} \right| = \frac{1}{M} \left| \frac{1 - e^{-\iota \pi (\vartheta - \theta) M}}{1 - e^{-\iota \pi (\vartheta - \theta)}} \right| \\ &= \frac{1}{M} \left| \frac{\sin \frac{\pi (\vartheta - \theta) M}{2}}{\sin \frac{\pi (\vartheta - \theta)}{2}} \right| =: F_M(\vartheta - \theta), \end{aligned} \tag{15}$$

which is the square root of the Fejér kernel of order M [20], defining the beam pattern of ULAs. We introduce useful lemmas regarding the square root of the Fejer kernel $F_M(\cdot)$ required to prove several theorems in this section regarding the RDB rate performance under the UR-MP channel model.

Lemma 1: [15] For $|\tilde{\theta}| \in (0, 1]$, we have an upper bound for $F_M(\tilde{\theta})$ as

$$F_M(|\tilde{\theta}|) \leq \frac{1}{M|\tilde{\theta}|}. \tag{16}$$

Lemma 2: [21] $F_M(\tilde{\theta})$ is upper bounded by a piece-wise step function given by

$$F_M(|\tilde{\theta}|) \leq \sum_{j=1}^{M/2} \frac{1}{j} \cdot \mathbf{1}_{[\frac{2(j-1)}{M}, \frac{2j}{M})}(|\tilde{\theta}|), \tag{17}$$

where $\mathbf{1}_A(\cdot)$ is the indicator function.

Basically Lemmas 1 and 2 state that $F_M(x)$ is upper bounded by $\frac{c}{Mx}$ with some c.

We first consider the asymptotic case that M tends to infinity with finite and fixed L regardless of M. The following theorem provides the result in this case.

Theorem 2: [1] Under the UR-MP channel model with finite and fixed L, an asymptotic lower bound on the per-user rate $\mathcal{R}_{\kappa_b}$ for fixed total transmit power $P_t = 1$ is given by

$$\begin{aligned} \mathcal{R}_{\kappa_b} &= \mathbb{E}\left[\log\left(1 + \frac{M^{-1} |\mathbf{h}_{\kappa_b}^H \mathbf{a}(\vartheta_b)|^2}{1 + M^{-1} \sum_{b' \neq b} |\mathbf{h}_{\kappa_b}^H \mathbf{a}(\vartheta_{b'})|^2}\right)\right] \\ &\gtrsim r_{LB,1} > 0, \end{aligned} \tag{18}$$

when $K = \Theta(M)$, where $x \gtrsim y$ indicates $\lim_{M \to \infty} \frac{x}{y} \geq 1$, and $r_{LB,1}$ is a positive constant value.

Proof: See Appendix B. ■

Theorem 2 states that under the UR-MP channel model with a finite number L of multi-paths, linear sum rate scaling w.r.t. the number of transmit antennas is achievable by the RDB scheme when the number of users in the cell increases linearly

w.r.t. the number of transmit antennas. This result is a generalization of the result under the UR-SP channel model provided in [15]. Thus, linear sum rate scaling is achievable by the RDB scheme with a linearly increasing number of users w.r.t. the number of transmit antennas not only for the UR-SP channel model but also for the UR-MP channel model if the number of multi-paths is finite irrespective of the number of transmit antennas.

Next we consider the case in which the number L of multi-paths increases as M increases. In this case, we have the following result.

Theorem 3: Under the UR-MP channel model, if the following conditions are satisfied:

(C.1) L tends to infinity as $M \to \infty$,

(C.2) $L/M \leq 1$,

(C.3) $K = Me^{c_u L}$ for some $c_u > 3$,

then an asymptotic lower bound on the per-user rate $\mathcal{R}_{\kappa_b}$ for fixed total transmit power $P_t = 1$ is given by

$$\mathcal{R}_{\kappa_b} \gtrsim r_{LB,2} > 0, \tag{19}$$

where $r_{LB,2}$ is a positive constant value.

Proof: See Appendix D. ■

Theorem 3 states that linear sum rate scaling by the RDB scheme is achievable under the UR-MP channel model with $L \to \infty$ and $L/M \leq 1$ as $M \to \infty$ when $K = Me^{c_u L}$ for $c_u > 3$. Consider the condition of

$$K = Me^{c_u L}, \tag{20}$$

with $c_u > 3$. Taking logarithm on both sides of (20), we have a sufficient condition for the number of users in the cell for RDB to achieve linear sum rate scaling w.r.t. the number M of transmit antennas, given by

$$\log K = \log M + c_u L. \tag{21}$$

Consider now several functions for L satisfying the condition that $L \to \infty$ and $L/M \leq 1$ as $M \to \infty$. One of such functions is $L = \log M$. In this case, (21) becomes $\log K = \log M + c_u \log M = (1 + c_u) \log M$ and we have a sufficient number of users for linear sum rate scaling expressed as

$$K = \Theta(M^{1+c_u}). \tag{22}$$

Thus, in this case the sufficient number of users in the cell for RDB to achieve linear sum rate scaling w.r.t. M is *polynomial* in M. Next we consider the fractional power function $L = M^\beta$ with $0 \leq \beta \leq 1$, which satisfies the condition that $L \to \infty$ and $L/M \leq 1$ as $M \to \infty$. As seen in the previous section, this model connects the UR-SP channel model and the i.i.d. Rayleigh fading channel model by varying β. In this case, we have $\log K = \log M + c_u M^\beta$, i.e., $\log K = \Theta(c_u M^\beta)$, and thus we have a sufficient number of users for linear sum rate scaling expressed as

$$K = \Theta(e^{c'_u M^\beta}), 0 \leq \beta \leq 1. \tag{23}$$

Thus, in this case the sufficient number of users in the cell for RDB to achieve linear sum rate scaling w.r.t. M is *sub-exponential* in M for $\beta < 1$. Note that at $\beta = 1$, K in (23) is an exponential function of M. Finally, from Theorem 1, we know that the UR-MP model with $L = M^\beta$ with $\beta > 1$ corresponds to the i.i.d. Rayleigh fading channel model under which an *exponentially* increasing number of users in the cell w.r.t. the number of transmit antennas is required for RBF to achieve linear sum rate scaling w.r.t. the number of transmit antennas due to Sharif and Hassibi [9]. Indeed, the result of Sharif and Hassibi coincides with our result (23) by setting $\beta = 1$. Table I summarizes the above results.[3]

Remark 1: In the two extreme cases of the UR-SP channel model and the i.i.d. Rayleigh fading channel model, the conditions are also necessary conditions since $K \geq M$ to schedule M users for the UR-SP model and the necessity of $\Theta(\log K) = M$ for the i.i.d. Rayleigh fading channel was shown in [9].

IV. User Selection in the Sparse User Regime

A. Motivation

In the previous section, we have identified the sufficient number of users in the cell required for RDB to achieve linear sum rate scaling w.r.t. the number of transmit antennas under the UR-MP channel model. Contrary to the i.i.d. Rayleigh fading channel case, linear sum rate scaling w.r.t. M by RDB can be achieved with a much smaller number of users K in the cell under sparse UR-MP channels such as fixed L, $L = \log M$ and $L = M^\beta$ with $\beta < 1$ than that required for the i.i.d. Rayleigh fading channel case. Thus, when the number of active users K in the cell is sufficiently high (dense user regime) as specified in the previous section, conventional RDB with $S = M$ described in Section III can be simply used. However, in the sparse user regime, the number of active users K may not be sufficiently large in order to achieve linear sum-rate scaling. Consider the UR-SP channel case. In this case, a linearly increasing number of users in the cell is required for RDB to achieve linear sum rate scaling w.r.t. M, as shown in the previous section. When M is in order of hundreds as in mmWave massive MIMO, the number of active users that is linear in M may not be a small number, especially in small cell networks. Thus, in this sparse user regime, simple RDB with $S = M$ may not be applied due to the lack of users. Thus, in this section, we consider the sparse user regime and propose several scheduling algorithms with a limited amount of feedback.

B. Insights from the UR-SP Channel Case

In order to gain insights to devise new scheduling algorithms in the sparse user regime, we first review the previous result regarding the RDB performance under the UR-SP channel model provided in [15]. In [15], the *fractional rate order*

[3]In current cellular systems, cells are sectorized and thus the AoDs of propagation paths are confined. In this case, we can model the AoDs of propagation paths as uniformly distributed on $[\theta_{min}, \theta_{max}]$, where $-1 < \theta_{min} < \theta_{max} < 1$. In that case, we can also show that the same asymptotic lower bound on the per-user rate $\mathcal{R}_{\kappa_b}$ of Theorem 3 can be obtained by modifying the proof of Theorem 3 slightly. Therefore, Table I is still valid in the case of $\theta_{k,i} \overset{i.i.d.}{\sim} \mathrm{Unif}[\theta_{min}, \theta_{max}]$, even though the number of training beams can be reduced from M to $\lfloor \frac{M(\theta_{max} - \theta_{min})}{2} \rfloor$ [22]. However, the maximum number of users supported at the same time-frequency resource is also reduced from M to $\lfloor \frac{M(\theta_{max} - \theta_{min})}{2} \rfloor$.

(FRO) γ was defined to compare the asymptotic performance of several scheduling methods in the sparse user regime as a function of number of users in the cell as

$$\gamma := \lim_{M\to\infty} \frac{\log \mathcal{R}}{\log M}. \tag{24}$$

The FRO indicates how the sum rate $\mathcal{R}$ of a method behaves as a fractional power function, i.e. $\mathcal{R} = \Theta(M^\gamma)$ for $\gamma \neq 0$. For $\gamma > 0$, $\mathcal{R}$ increases to infinity as $M \to \infty$, whereas for $\gamma < 0$, $\mathcal{R}$ decreases to zero as $M \to \infty$.

In [15], a general number of users $K = M^q$ with $0 \leq q \leq 1$ is considered to include the linear user increase ($q = 1$) w.r.t. M (the sufficient user regime) and the case of sparse users as a fractional power function ($q < 1$). Among several scheduling methods considered in [15], we consider here the following two methods. The first method is a modified RDB scheme [15]. In the modified RDB scheme, when $K = M^q$ with $0 < q \leq 1$, $S = M^\ell$ (with $\ell < q$) asymptotically orthogonal beams equispaced in the normalized angle domain $[-1, 1]$ are used for both training and data transmission. Thus, in this scheme, the number of orthogonal beams for RDB is reduced according to K although the number of transmit antennas is M. Since all reduced S beams are used for data transmission in this scheme, each user can compute the SINR for each beam based on its channel during the training period under the fact that all other $S - 1$ are also used for data transmission (this is an advantage of this method), and feeds back the maximum SINR value and the corresponding beam index as in the conventional RBF. The FRO in this case is given as a function of q (determining the number of users) by [15].

$$\lim_{M\to\infty} \frac{\log \mathcal{R}_{sum}(S)}{\log M} = 2q - 1, \text{ for } q \in (0, 1). \tag{25}$$

In the second method, all M orthogonal beams in (13) available with M transmit antennas at the BS are used for training. During the training period, each user computes the received power for each training beam based on its channel and feeds back the maximum *received power* value and the corresponding beam index. Then, the BS chooses the user and beam index pair that has the maximum received power out of all beams and all users, and transmits data only to the selected user with the selected beam. The FRO in that case is given by [15]

$$\lim_{M\to\infty} \frac{\log \mathcal{R}_{SU}}{\log M} = 0, \text{ for } q \in (0, 1) \tag{26}$$

since $\mathcal{R}_{SU} = \log(1 + \max_{k,b} |\mathbf{h}_k^H \mathbf{w}_b|)$ in this case.

Fig. 3 compares the FROs (25) and (26) versus q of the two schemes. In the figure, the FRO with the perfect CSI at the transmitter is also shown.[4] Fig. 3 shows which strategy is better for different q determining the number of users in the cell relative to the number of antenna elements. Note that the first method (the reduced beam-size conventional RDB) has larger FRO than the second method for $q \in (\frac{1}{2}, 1)$, whereas the second high-resolution training single-user selection method has

[4]In the perfect CSI case, the capacity of the network composed of a BS with M transmit antennas and $K = M^q$ single-antenna users scales as $\min(M, M^q) \log \text{SNR}$. Hence, the FRO of the capacity is given by q for $0 \leq q \leq 1$.

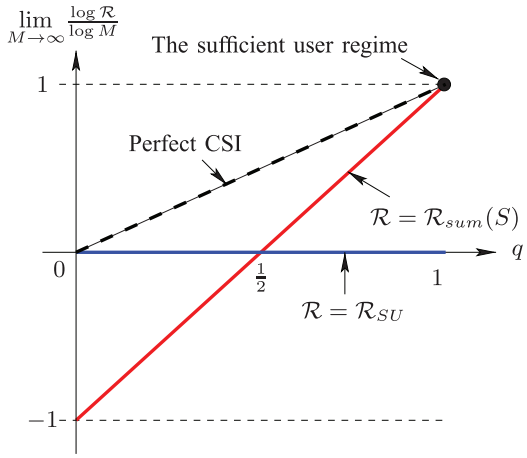

Fig. 3. Fractional rate order versus the number of users in the cell [15].

larger FRO than the first method for $q \in (0, \frac{1}{2})$. The reason why the first method yields worse performance than the second method when q is small is that it does not have enough training due to the requirement that all training beams become transmission beams in the data transmission phase and SINR can be computed at each receiver during the training phase. Note that $\ell < q$ in this case. On the other hand, the reason why the second method yields worse performance when q is large is that it only serves a single user even though the number of users in the cell increases as $q \uparrow 1$. Based on these two facts, one can recognize that if the training overhead can be ignored, to enhance the system performance, all M orthogonal beams available with the M transmit antennas at the BS should be used for training and the number of served users should increase properly as $q \uparrow 1$ although the number of served users is less than M. In those cases, each user cannot compute SINR at the receiver side because all M beams are not used for data transmission and each receiver does not know which of the M beams will be used for data transmission beforehand. Thus, in the sparse user regime with full training, we should solve the joint problem of selecting beams for data transmission among M full training beams and selecting users to be served among K active users in the cell with limited amount of feedback. Thus, in the next subsection we shall propose several scheduling methods incorporating both beam and user selection for mmWave massive MIMO assuming sparse MIMO channels incorporating the above hints, based on *feedback of the received signal power values* from users.

C. Proposed Scheduling Methods With Beam Selection

In the methods below, targeting the sparse user regime, we use all M orthogonal beams available with the M transmit antennas at the BS for training and select beams for data transmission and users to be served. In the proposed methods, the numbers of served users and selected beams are adaptively determined according to each channel realization.

1) Scheduling Based on a Single Beam for a Selected User: The first method handles the case that a single beam is used for each selected user. As aforementioned, since all M training beams are not used for data transmission, each user cannot

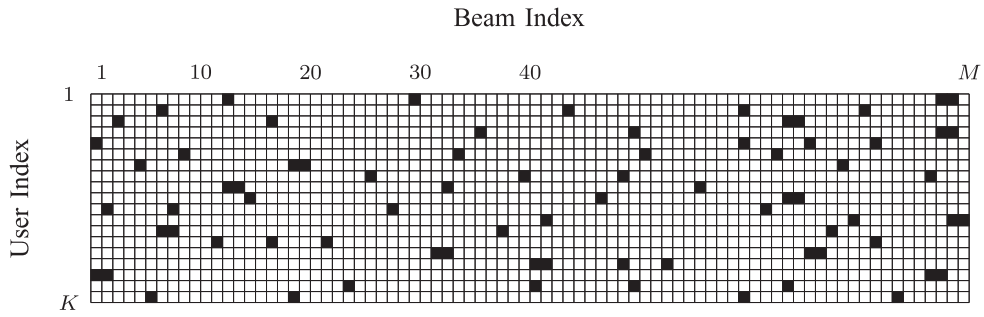

Fig. 4. The received signal power table.

compute the exact SINR value for each beam during the training period. However, each user can still compute the received signal power for each of the M beams during the training period. Suppose that each user feeds back the M received signal power values $\{|\mathbf{h}_k\mathbf{w}_b|^2\}_{b=1}^M$ for all M beams to the BS after the training period. Then, the BS can construct a table that contains the received power value for each beam for each user. Once such a table is constructed, we can apply a scheduling algorithm. Suppose now that we serve only $K_s(< M)$ users from the K users in the cell with one beam $\mathbf{w}_b$ for one selected user. The optimal beam and user selection in this case can be obtained by an exhaustive search with computing SINRs over all the possible subsets of beams and users. However, such an optimal scheme requires a huge amount of feedback and heavy computational complexity. Thus, we here consider a practical selection method with a reduced amount of feedback and reduced complexity.

First, in the proposed feedback method, each user computes the received signal power for all M beams but feeds back only the largest N_{FB} received signal power values and the corresponding indices to the BS. When the feedback from all users is finished, the BS can construct the received power table as shown in Fig. 4 with the feedback information by *simply padding zeros in the unreported boxes.* Such a feedback scheme is effective in sparse mmWave channels since there exists signal only in a few directions [2], [3]. Under the UR-MP channel model, the channel vector $\mathbf{h}_k$ is the sum of L uniform random multi-paths with complex gains. Thus, signal power $|\mathbf{h}_k^H\mathbf{w}_b|^2$ is large when the angle of the training beam $\mathbf{w}_b$ and one of the L paths are roughly aligned.

Second, when the received signal power table like one in Fig. 4 is constructed, one can apply a suboptimal greedy sequential scheduling algorithm to select K_s users, as in [23]. The basic concept of such sequential greedy selection algorithms [23] is that the BS first selects a user-and-beam pair (k_1, b_1) that has the maximum signal power $|\mathbf{h}_{k_1}^H\mathbf{w}_{b_1}|^2$ over all beams and users in the table, and then under the condition that the beam b_1 is allocated for the user k_1, selects another user-and-beam pair (k_2, b_2) that has the maximum sum rate over all beams and users except the selected user and beam at the first step. This sequential procedure is iterated until the number of allocated beams reaches K_s. (For detail, see the algorithm in Table II in [23].)

2) Scheduling Based On Beam Aggregation: In the previous subsection, we considered the case that only one beam is used to serve a selected user. However, when there exist multiple propagation paths at different angles from the BS to a user as in the UR-MP channel model, such a scheme is not optimal, and we should use those multiple paths for data transmission. Thus, in this subsection, we propose several feedback and beam and user selection schemes that use multiple beams for one selected user for data transmission based on *beam aggregation*. In the proposed scheduling methods, we find multiple beam directions that are matched to the angles of the L multiple paths for each user by using the M training beams, schedule users that have roughly orthogonal channel vectors and a large sum rate, and then transmit data streams to the scheduled users. In the data transmission stage, we use an aggregated beam combining the matched multiple beams to each scheduled user to transmit data to the user. In the case of sparse channels as in the mmWave band, the feedback amount required for the proposed methods is not heavy and the proposed methods can effectively implement MRT beamforming or EGC transmit beamforming.

2-1) MRT-Based Scheduling: The detail of the proposed scheduling method aiming at MRT is described in Algorithm 1. In Algorithm 1, at each iteration, the BS chooses a user that has the maximum combined power value in the MRT sense among the users in the set $\mathcal{K}$ that is updated in each iteration to choose a user whose channel is roughly orthogonal to those of the previously selected users, and constructs an MRT beam for the chosen user based on the feedback information of the N_{FB} dominant training beams' received signal magnitudes and phases. Rough orthogonality (and thus interference) among the selected users is controlled by updating the set $\mathcal{K}$ of candidate users. As (32), the set $\mathcal{K}$ is updated as the set of users whose dominant beams do not overlap with the already scheduled user's dominant beams by more than N_{OL} beams. Thus, by controlling N_{OL} we can manage interference among the selected users. Note that in Step 5) in Algorithm 1, by setting $K_{s,max} = \min(K, M)$ we can run the algorithm until we find maximally possible users for scheduling. In this case, the number of scheduled users is not predetermined but determined adaptively according to channel realization. The amount of feedback required for the MRT-based scheduling algorithm is $2N_{FB}$ real numbers and N_{FB} integers.

2-2) EGC-Based Scheduling: To reduce the amount of feedback, we also consider a scheduling method based on EGC. We obtain an EGC-based scheduling algorithm by slightly modifying Algorithm 1. In the EGC-based scheme, each user

k feeds back the sum of the magnitude of the N_{FB} dominant training beams (i.e., $\sum_{b\in\mathcal{B}_k}|\mathbf{h}_k^H\mathbf{w}_b|$), the N_{FB} phase values $\{\angle(\mathbf{h}_k^H\mathbf{w}_b), b\in\mathcal{B}_k\}$, and the corresponding N_{FB} beam indices to the BS. Then, the procedure of EGC-based scheduling is the same as that of MRT-based scheduling except the fact that (30) and (31) are replaced by

$$\kappa_j = \arg\max_{k\in\mathcal{K}} \sum_{b\in\mathcal{B}_k} |\mathbf{h}_k^H\mathbf{w}_b| \tag{27}$$

$$\bar{\mathbf{w}}_j = \sum_{b\in\mathcal{B}_{\kappa_j}} e^{-\iota\angle(\mathbf{h}_{\kappa_j}^H\mathbf{w}_b)}\mathbf{w}_b. \tag{28}$$

Note that the amount of feedback required by this strategy is $N_{FB}+1$ real numbers and N_{FB} integers.

3) Beam Design Beyond MRT and EGC: One way to avoid causing inter-user interference among the scheduled users in the beam and user selection methods in Subsection IV-C2 is to set $N_{OL}=0$, if all beam indices with non zero inner product with each user's channel are reported from each user. However, setting $N_{OL}=0$ reduces the number of users to be served simultaneously by the BS. To increase the number of simultaneously served users, we can set $N_{OL}\geq 1$ to allow a certain amount of channel overlap. In the case of $N_{OL}\geq 1$, we can eliminate this inter-user interference by designing zero-forcing (ZF) or minimum-mean-square-error (MMSE) beams instead of the MRT beams. In this case, after user scheduling is done based on Algorithm 1, the BS designs ZF or MMSE beams for the selected users instead of MRT beams. Here, an approximate ZF or MMSE beam can be designed using the CSI of the dominant beam indices with padding zeros in the unreported user-beam boxes (blocks). Hence, the MMSE beam design based on zero padding requires the same amount of feedback as the MRT beam design.

Furthermore, we can improve the performance of the ZF or MMSE beam design by applying two-phase feedback (TPF) [10], [12], [24]–[26]. In a two-phase feedback scheme, the BS selects users with partial CSI in the first phase and designs a refined beamformer based on the full CSI feedback from the selected users in the second phase. Here, we propose a beam design and scheduling method based on this TPF method as follows. In the first phase, each user k feeds back the N_{FB} dominant beam indices $\mathcal{B}_k$ and the one real value $\sum_{b\in\mathcal{B}_k}|\mathbf{h}_k^H\mathbf{w}_b|^2$ to the BS. Then, the BS chooses users $\{\kappa_i\}_{i=1}^J$ to be served and the corresponding beam indices $\{\mathcal{B}_{\kappa_i}\}_{i=1}^J$ by Algorithm 1 without designing the MRT beam in (31). The value of J will be determined when Algorithm 1 is finished. After user selection, the BS additionally receives the information of magnitudes $\{|\mathbf{h}_k^H\mathbf{w}_b|\}$ and phases $\{\angle(\mathbf{h}_k^H\mathbf{w}_b)\}$ corresponding to the beam indices in the set

$$\mathcal{B}^\star := \cup_{i=1}^J \mathcal{B}_{\kappa_i} \tag{34}$$

from each selected user in the second phase. Then, the BS knows the full CSI $\mathbf{H}_{eff}=\mathbf{W}_{sel}^H\mathbf{H}_{sel}$ within the subspace spanned by the beams $\{\mathbf{w}_b\}_{b\in\mathcal{B}^\star}$ that will be used for data transmission to the selected users, where the $M\times|\mathcal{B}^\star|$ matrix $\mathbf{W}_{sel}=[\mathbf{w}_b]_{b\in\mathcal{B}^\star}$, the $M\times J$ matrix $\mathbf{H}_{sel}=[\mathbf{h}_{\kappa_1},\cdots,\mathbf{h}_{\kappa_J}]$, and the $M\times 1$ vector $\mathbf{h}_{\kappa_j}$ is the actual channel from the BS antenna array to user κ_j. Thus, it is possible for the BS to implement exact MMSE beamforming (or ZF beamforming) for the selected users, given by

$$\mathbf{V} = \xi\cdot\mathbf{H}_{eff}\left(\mathbf{H}_{eff}^H\mathbf{H}_{eff}+\frac{J}{P_t}\mathbf{I}\right)^{-1} \tag{35}$$

where ξ is the power scaling factor to satisfy the power constraint. The BS finally transmits data streams to the scheduled users $\{\kappa_i\}$ with the MMSE beams $\{\bar{\mathbf{w}}_{\kappa_i}\}$ determined as

$$[\bar{\mathbf{w}}_{\kappa_1},\cdots,\bar{\mathbf{w}}_{\kappa_J}] = \mathbf{W}_{sel}\mathbf{V}. \tag{36}$$

The amount of feedback required for this method is $2|\mathcal{B}^\star|+1$ real numbers and N_{FB} integers for each scheduled user. Note that $N_{FB}<|\mathcal{B}^\star|\approx JN_{FB}\ll M$ for small J in the sparse user regime so this method can be applied with a moderate amount of feedback for each user. Note that this TPF-based ZF or MMSE scheme requires more feedback than the ZF or MMSE beam design based on zero padding for the unreported elements in $\mathbf{H}_{eff}$ but yields more accurate beams to eliminate the inter-user interference. If this TPF scheme is used, the number N_{FB}

Algorithm 1. Proposed Beam Aggregation Method: Maximum Ratio Transmission

1) The BS initializes

$$\mathcal{K}_1=\{1,2,\cdots,K\}, \mathcal{B}=\{1,2,\cdots,M\}, \tag{29}$$

and $j=1$.

2) Each user k computes N_{FB} dominant training beam signal powers $\{|\mathbf{h}_k^H\mathbf{w}_b|^2\}$ among the M training beams, and feeds back the magnitudes $\{|\mathbf{h}_k^H\mathbf{w}_b|, b\in\mathcal{B}_k\}$ and phases $\{\angle(\mathbf{h}_k^H\mathbf{w}_b), b\in\mathcal{B}_k\}$ of the N_{FB} dominant training beams and the corresponding beam indices to the BS. $\mathcal{B}_k$ denotes the set of the N_{FB} dominant beam indices.

3) The BS computes

$$\kappa_j = \arg\max_{k\in\mathcal{K}_j}\sum_{b\in\mathcal{B}_k}|\mathbf{h}_k^H\mathbf{w}_b|^2 \tag{30}$$

$$\bar{\mathbf{w}}_{\kappa_j} = \frac{\hat{\mathbf{w}}_{\kappa_j}}{\|\hat{\mathbf{w}}_{\kappa_j}\|}, \tag{31}$$

where $\hat{\mathbf{w}}_{\kappa_j}=\sum_{b\in\mathcal{B}_{\kappa_j}}(\mathbf{h}_{\kappa_j}^H\mathbf{w}_b)^*\mathbf{w}_b$, and updates $\mathcal{K}_j=\mathcal{K}_j\backslash\{\kappa_j\}$ and $\mathcal{B}=\mathcal{B}\backslash\mathcal{B}_{\kappa_j}$.

4) To impose rough orthogonality among the selected users, the BS further updates the set $\mathcal{K}$ of candidate users as

$$\mathcal{K}_{j+1}=\left\{k\in\mathcal{K}_j : |\mathcal{B}_{\kappa_j}\cap\mathcal{B}_k|\leq N_{OL}\right\} \tag{32}$$

where N_{OL} is the number of allowed beam overlaps between the sets of dominant beams.

5) If $j<K_{s,max}$, $\mathcal{K}_{j+1}\neq\emptyset$, and $\mathcal{B}\neq\emptyset$, update $j\leftarrow j+1$ and go to step 3). Otherwise, the algorithm is finished and the set of selected users and corresponding rough MRT beams is given by

$$(\kappa_1,\bar{\mathbf{w}}_{\kappa_1}),\cdots,(\kappa_j,\bar{\mathbf{w}}_{\kappa_j}). \tag{33}$$

of dominant beams for feedback can be reduced while the inter-user interference during the data transmission period is still completely removed.

V. Extension

A. Fairness-Aware Scheduling

If the channel statistics are the same across all users and the channel realizations are i.i.d. across scheduling intervals, providing fairness among users is not an issue [11]. However, in networks where users experience different large-scale fading due to different user locations from the base station, fairness among users has to be considered. One method to implement fairness with obtaining MU diversity gain is to apply the proportional-fair (PF) scheduling policy [27], which was originally proposed for single-user selection with a single beam at the same time-frequency resource. In the PF algorithm, the BS keeps tracking of the average previously service rate $\mu_k(t)$ for every user k at each scheduling interval t and selects the user that has the maximum ratio of the currently supportable rate to the average served rate, i.e.,

$$k^\star = \arg\max_k \frac{R_k(t)}{\mu_k(t)}, \tag{37}$$

where $R_k(t)$ is the currently supportable rate of user k at scheduling interval t. Here, the average supported rate is updated by a simple first-order autoregressive (AR) filter as

$$\mu_k(t+1) = (1-\delta)\mu_k(t) + \delta R_k(t) I_{\{k\in\mathcal{S}(t)\}}, \tag{38}$$

where I_A is the indicator function of event A, $\delta \in (0,1)$ is the parameter of the first-order AR filter, and $\mathcal{S}(t)$ is the set of scheduled users at scheduling interval t. However, in the case of multi-user selection with multiple beams as in the proposed scheduling methods or RBF, the PF scheduling policy cannot be applied directly since the rate of a scheduled user is dependent on the channel vectors of the other scheduled users at the same time, and the instantaneous supportable rate for each scheduled user can be computed after all scheduled users and the corresponding beams are determined. Therefore, all combinations with respect to beams and users should be considered for maximizing the rate in the sense of the PF scheduling policy, and the best pairs of beams and users can be chosen by exhaustive search. However, this difficulty can be overcome by exploiting the property of our scheduling methods that the channel vectors of the scheduled users are almost orthogonal to each other. This property is imposed by the step (32) in Algorithm 1. Then, the instantaneous supportable rates for the selected users required for PF metric computation can be approximated based on this semi-orthogonality by neglecting the interference from other scheduled users as in [10], [11]. That is, the instantaneous supportable rate of user k at scheduling interval t is approximated based only on its channel by

$$\hat{\mathcal{R}}_k(t) = \log\left(1 + \frac{P_t}{|\mathcal{S}(t-1)|}|\mathbf{h}_k(t)^H \bar{\mathbf{w}}_k(t)|^2\right), \tag{39}$$

$$= \log\left(1 + \frac{P_t}{|\mathcal{S}(t-1)|}\sqrt{\sum_{b\in\mathcal{B}_k}|\mathbf{h}_k^H \mathbf{w}_b|^2}\right), \tag{40}$$

where $\mathbf{h}_k(t)$ is the channel vector of user k, $\bar{\mathbf{w}}_k(t) = \frac{\hat{\mathbf{w}}_k}{\|\hat{\mathbf{w}}_k\|}$ is the corresponding beamforming vector at time t, and $\hat{\mathbf{w}}_k = \sum_{b\in\mathcal{B}_k}(\mathbf{h}_k(t)^H\mathbf{w}_b)^*\mathbf{w}_b$. Here, the key point in (40) is that $\hat{\mathcal{R}}_k(t)$ is computed based only on the feedback information $|\mathbf{h}_k^H\mathbf{w}_b|^2, b\in\mathcal{B}_k$ without considering other scheduled users' channel vectors by exploiting the semi-orthogonality among the scheduled users' channel vectors guaranteed by the step (32). Finally, the proposed scheduling method based on beam aggregation incorporating the PF principle is given by modifying Algorithm 1 as follows:

1) At scheduling interval t, perform Algorithm 1 with replacing (30) by

$$\kappa_j = \arg\max_{k\in\mathcal{K}_j} \frac{\hat{\mathcal{R}}_k(t)}{\mu_k(t)}, \tag{41}$$

where $\hat{\mathcal{R}}_k(t)$ is given by (40).

2) Update $\mu_k(t+1)$ by (38).

B. Application of the Proposed Scheduling Methods to Hybrid Beamforming

Real-world implementation of massive mmWave MIMO systems requires complex and costly RF/Analog circuit design. A practical solution to overcome the burden of heavy RF/Analog circuitry for building massive mmWave MIMO systems is hybrid analog/digital beamforming, in which the number of analog-to-digital (AD) converters is far less than that of antennas, and analog beamforming is performed between the AD converters and the antennas by using RF amplifiers and phase shifters [6], [8], [19], [28]–[30]. One of the main issues in the hybrid beamforming scheme is to estimate the CSI of large-array antennas between the BS and each user with a small number of training beams. One candidate for this is to predetermine a set of analog beams with a reduced size (i.e., less than or equal to M) at the BS and training symbols are transmitted by using the set of predetermined analog beams (e.g., one training symbol for each beam in the analog beam set); the training signal itself is beamformed by each of the predetermined analog beams [30]. In this case, each user can only feed back certain information about the effective beamformed channel, but not about the real channel between the user and the BS antenna array. Then, the BS should select analog beams to be used as well as users to be served based on the feedback information. Note that this situation is exactly the same as that considered in Section IV-C and the proposed methods in Section IV-C can be applied to the joint beam and user selection problem in the hybrid beamforming scheme (the analog and digital beamformers become $\mathbf{W}_{sel}$ and $\mathbf{V}$ in (36), respectively.). For example, by using Algorithm 1 we can obtain the set $\{\mathbf{w}_b\}_{b\in\mathcal{B}^\star}$ (i.e., $\mathbf{W}_{sel}$) of analog beams to be used and the set $\{\kappa_i\}_{i=1}^J$ of users to be served, and design the MMSE digital beamformer $\mathbf{V}$ given in (35) based on zero padding on the unreported elements in $\mathbf{H}_{eff}$ or the additional feedback in the TPF case.

VI. Numerical Results

In this section, we provide some numerical results to validate our asymptotic analysis in Section III and to evaluate the

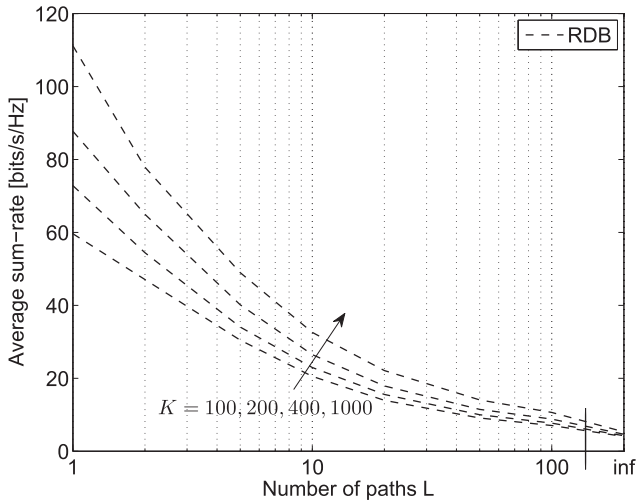

Fig. 5. Average sum rate of the RDB scheme versus L for different K ($M = 100$ and $P_t = 1$).

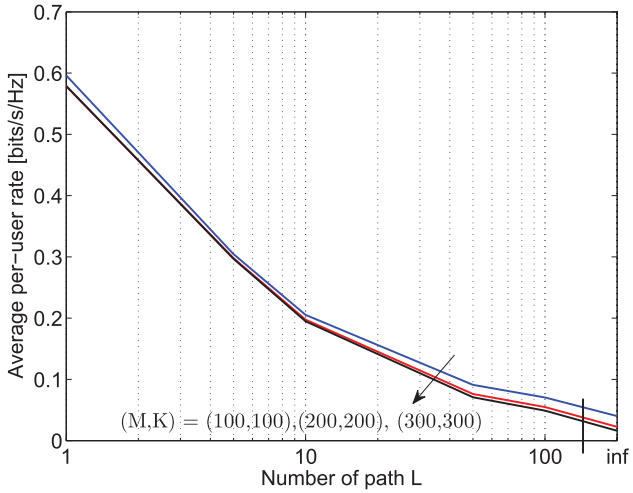

Fig. 6. Average per-user rate of the RDB scheme versus L for different M ($K = M$ for each M and $P_t = 1$).

performance of the proposed scheduling methods in Section IV. All the values in the below are values averaged over 200 channel realizations.

First, we investigated the impact of channel sparsity on the RDB performance. We considered a mmWave MU-MISO downlink system with the UR-MP channel model with different L. Fig. 5 shows the average sum rate of the RDB scheme versus L for $K = 100, 200, 400, 1000$ when $M = 100$ and $P_t = 1$. Note that the last value in the x-axis of the figure is $L = \infty$, i.e., it corresponds to the i.i.d. Rayleigh fading channel. It is seen that the average sum rate of the RDB scheme drastically decreases as the number L of multi-paths increases. This results from the difference in the required sufficient number of users depending on L. As L increases, the sufficient user regime requires more users in the cell, predicted by the analysis in Section III. Hence, if we fix K, the RDB performance degrades as L increases. Note that the performance gap between the two extreme cases $L = 1$ and $L = \infty$ is large.

We then evaluated the average per-user rate of the RDB scheme versus L for different $(M, K) = (100, 100), (200, 200)$, and $(300, 300)$. Fig. 6 shows how the per-user rate changes as $M = K$ increases for different L. It is seen in Fig. 6 that at $L = 1$ (i.e., under the UR-SP channel model) the per-user rates for $M = K = 100$, $M = K = 200$, and $M = K = 300$ are almost the same. This means that the sum rate scales linearly with respect to M with $K = M$. This result coincides with the result in [15]. and Theorem 2. On the other hand, at $L \to \infty$ (under the i.i.d. Rayleigh fading channel model) the per-user rates for $M = K = 100$, $M = K = 200$, and $M = K = 300$ are not the same. The per-user rate for $M = K = 300$ is almost half of that for $M = K = 100$. This means that the sum rate did not scale linearly w.r.t. M with $K = M$, as expected from (1). We need an exponential increase in K as a function of M for linear sum rate scaling w.r.t. M. It is seen that the per-user curves for $M = K = 100$, $M = K = 200$, and $M = K = 300$ deviate from $L = 10$ in this setting.

Next, we considered the sparse user regime $K \leq M$ and evaluated the performance of the proposed beam selection and scheduling methods (MRT, EGC, MMSE with zero padding and TPF-MMSE), and the greedy single beam selection method in [23]. Since the MRT-based scheduling method receives the N_{FB} magnitude values and N_{FB} phase values of the N_{FB} dominant paths, the greedy single beam selection method used $2N_{FB}$ dominant received signal power values for fair comparison with the MRT-based method in terms of feedback. As a reference, we included the performance of conventional RDB in this sparse user regime. In the case of RDB, M orthogonal training beams were used, and each receiver computed the SINR for each beam under the false assumption that all M training beams would be used for data transmission and fed back its maximum SINR and corresponding beam index. Then, the BS chose a user with maximum SINR for each reported beam index. With consideration of beam index overlap, this method used at most K beams and supported at most K users. Note that $K \leq M$ in the considered sparse user regime. As a benchmark for optimal beamforming, we included the performance of optimal MMSE beamforming based on perfect CSI knowledge of all users at the BS, which requires feedback of $2MK (\gg 2N_{FB}K)$ real numbers.[5] For the proposed methods based on Algorithm 1 and the greedy single beam selection method, we set $K_{s,max} = \min(M, K)$ so as to find maximally possible users for scheduling. Fig. 7(a) shows the average sum rate of the considered scheduling algorithms versus K when $M = 100, L = 5$, $P_t = 1$, $N_{FB} = 4$ and $N_{OL} = 1$. It is seen that the proposed methods based on beam aggregation are superior to the other two previous methods (RDB and the greedy single-beam selection in [23]). Since $L = 5$, $K \leq 100$ is not sufficient for the system to be operated in the sufficient user regime (see Table I and thus the performance of RDB is not good in this sparse user regime. It is also seen that MMSE beamforming based on zero padding

[5]Since $K \leq M$, we can construct an MMSE beamformer from the BS antenna array to all K users in the MISO case when all $M \times 1$ channel vectors from the BS to all K users are known.

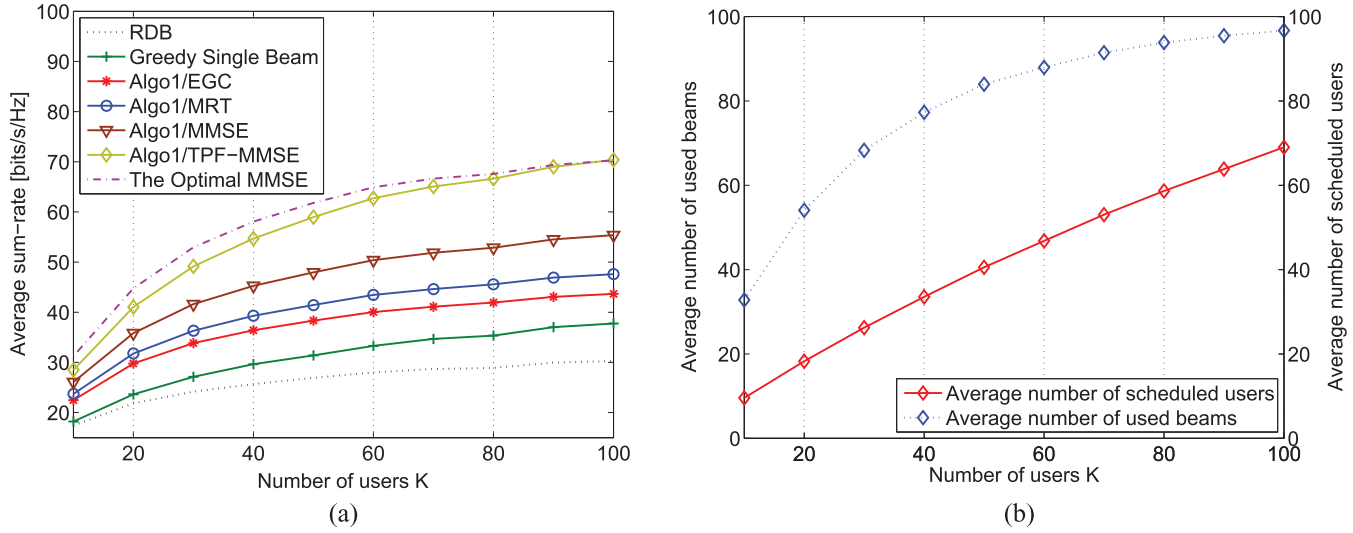

Fig. 7. (a) Average sum rate versus K w.r.t. the different scheduling algorithms and (b) the average number $|\mathcal{B}^*|$ of total selected beams and the average number of scheduled users in the TPF-based scheduling method: $M = 100$, $L = 5$, $P_t = 1$, $N_{FB} = 4$, and $N_{OL} = 1$.

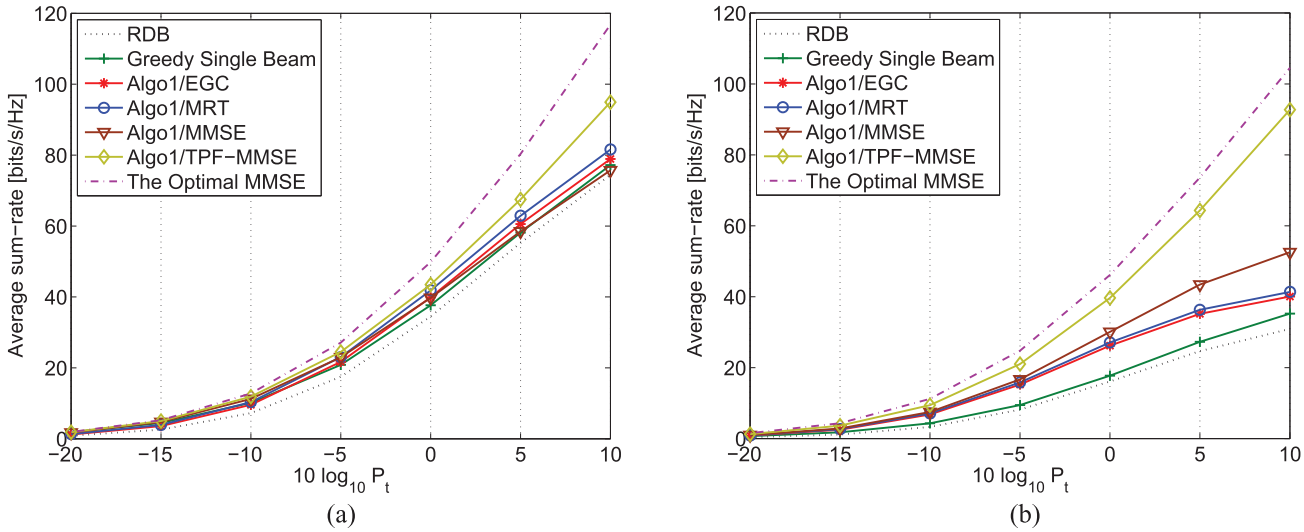

Fig. 8. Average sum rate versus P_t: (a) $M = 100$, $K = 30$, $L = 2$, $N_{FB} = 2$, $N_{OL} = 0$, and $|\mathcal{B}^*| = 38$ and (b) $M = 100$, $K = 20$, $L = 10$, $N_{FB} = 4$, $N_{OL} = 1$, and $|\mathcal{B}^*| = 54$.

with user-and-beam selection by Algorithm 1 outperforms the MRT-based method since it mitigates the inter-user interference. Furthermore, the TPF-based MMSE with user and beam selection by Algorithm 1 almost achieves the performance achieved by the optimal MMSE beamforming based on full CSI of all users at the BS. It is seen in Fig. 7(b) that a moderate amount of feedback is required for the TPF-based scheduling method, i.e., $2N_{FB}(= 8) < 2|\mathcal{B}^\star| + 1 \ll 2M(= 200)$ for small K compared to M.

Figs. 8(a) and (b) show the average sum rate of the proposed scheduling algorithms versus P_t in the cases of $(M, K, L, N_{FB}, N_{OL}) = (100, 30, 2, 2, 0)$ and $(M, K, L, N_{FB}, N_{OL}) = (100, 20, 10, 4, 1)$, respectively. The average numbers $|\mathcal{B}^\star|$ of total selected beams for the two setups in the TPF-based method were 38 and 54, respectively. It is worth comparing the performance of the three schemes based on Algorithm 1: MRT, MMSE with zero padding and TPF-MMSE, shown in Fig. 8(a). When the number of multipaths is very small ($L = 2$ here), major path gains are reported ($N_{FB} = L$)[6], and N_{OL} is set to zero, inter-user interference is already controlled by Algorithm 1 and the MRT-based beamformer yields better performance than the MMSE beamformer with zero padding since Algorithm 1 intends to select users in an MRT optimal sense. On the other hand, when unreported inter-user channel elements are fed back and an MMSE beamformer based on full effective CSI is used, this TPF-MMSE performs better than the MRT-based method. As L becomes large from 2 to 10, the situation changes as shown in Fig. 8(b). The greedy single-beam selection algorithm yields poor performance compared to the scheduling methods based on beam aggregation since it does not exploit the

[6]Since each user's channel may not be on the boresight of a training beam, the number of beams with non-zero inner product with the user's channel may be larger than L.

multi-path combining gain. Also, under Algorithm 1, even the MMSE beamforming with zero padding outperforms the MRT beamforming especially when the transmit power is high. The MMSE beamforming has an advantage in terms of inter-user interference mitigation even though it is designed based on imperfect CSI with padding zeros in unreported boxes. In Fig. 8(b), it can be seen that the rate performance of the three methods, namely MRT, EGC, and MMSE with zero padding, is interference-limited as the total transmit power increases. However, the TPF-MMSE method expectedly overcomes the interference-limited performance. Note that the proposed MRT-based and TPF-based scheduling methods achieve 85% and 88% in Fig. 8 (a) and 66% and 86% in Figs. 8 (b) of the optimal MMSE performance at $P_t = 1$, respectively, whereas the RDB method achieves barely 69% and 35% of the optimal MMSE performance, respectively. The amount of feedback per user of 4 and 8 real numbers for the proposed MRT method for Fig. 8 (a) and (b), respectively, is significantly smaller than that (200 real numbers) of the optimal MMSE scheme and slightly larger than that (1 real number) of RDB. In the case of the TPF-based method, the amount of feedback is 77 and 109 real numbers for Fig. 8 (a) and (b), respectively. Thus, the proposed scheduling methods based on beam aggregation perform very well in sparse mmWave channels with a reasonable amount of feedback.

VII. Conclusion

In this paper, the performance of RBF in mmWave MU-MISO downlink systems under the newly proposed UR-MP channel model was studied. We have derived the number of active users in the cell that is required for RBF to yield linear sum rate scaling w.r.t. the number of transmit antennas under the UR-MP channel model for different levels of channel sparsity. We have shown that in sparse mmWave channels the necessary number of users in the cell for RBF to yield linear sum rate scaling w.r.t. the number of transmit antennas is significantly less than that required in rich scattering channels. We have then considered the problem of user scheduling for MU-MISO downlink when the number of users in the cell is not sufficiently large (sparse user regime) for the system to operate in the linear sum-rate scaling regime. By exploiting the sparsity of mmWave radio channels, we have proposed several user scheduling algorithms with a reasonable amount of feedback for the considered sparse user regime. Furthermore, we have proposed a fairness-aware scheduling algorithm based on the PF principle requiring a reasonable amount of feedback. Numerical results showed that the proposed user selection algorithms provide very good sum-rate performance in sparse mmWave channels.

Appendix A

We first explain why we consider $\mathbf{h}_k|(\theta_{k,1}, \cdots, \theta_{k,L})$ instead of $\mathbf{h}_k$ for the rank property of the channel. Under the UR-MP model, the channel vector for user k is given by $\mathbf{h}_k = \sqrt{\frac{M}{L}} \sum_{i=1}^{L} \alpha_{k,i} \mathbf{a}(\theta_{k,i})$, where the path gain and the path (or ray) angle are given by $\alpha_{k,i} \overset{i.i.d.}{\sim} \mathcal{CN}(0,1)$ and $\theta_{k,i} \overset{i.i.d.}{\sim} \mathrm{Unif}[-1,1]$, respectively. Note that the unconditional channel covariance matrix is given by

$$\begin{aligned}\mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H] &= \mathbb{E}\left[\frac{M}{L} \sum_{i=1}^{L} \alpha_{k,i} \mathbf{a}(\theta_{k,i}) \sum_{j=1}^{L} \alpha_{k,j}^* \mathbf{a}^H(\theta_{k,j})\right] \\ &\overset{(a)}{=} \frac{M}{L} \sum_{i=1}^{L} \mathbb{E}\left[\mathbf{a}(\theta_{k,i}) \mathbf{a}^H(\theta_{k,i})\right] \\ &= \frac{1}{L} \sum_{i=1}^{L} \mathbb{E}\left[\left[e^{-\iota\theta_{k,i}(m-n)}\right]_{m,n}\right] \overset{(b)}{=} \frac{1}{L} \sum_{i=1}^{L} \mathbf{I} = \mathbf{I},\end{aligned}$$

where $\left[e^{-\iota\theta_{k,i}(m-n)}\right]_{m,n}$ denotes the matrix composed of $e^{-\iota\theta_{k,i}(m-n)}$ as its (m,n)-th element. Here, step (a) is valid since $\alpha_{k,i}$ and $\theta_{k,i}$ are independent and $\mathbb{E}[\alpha_{k,i}\alpha_{k,j}^*] = 0$ for $i \neq j$, and (b) holds since $\mathbb{E}[e^{-\iota\pi(m-n)\theta_{k,i}}] = \frac{1}{2}\int_{-1}^{1} e^{-\iota\pi(m-n)\theta_{k,i}} d\theta_{k,i} = \frac{\sin\pi(m-n)}{\pi(m-n)} = 0$ for any integers $m \neq n$ [5]. The reason why the unconditional channel covariance matrix has full rank is as follows. The probability that each ray comes from a certain direction is uniform over the angle domain. So, if we take the expectation over the ray angle, then the unconditional channel covariance matrix becomes of full rank since there is an equal probability for each ray direction. However, the actually realized channel for user k only has L paths (or rays) and the channel covariance matrix should have rank L. Hence, we consider $\mathbf{h}_k|(\theta_{k,1}, \cdots, \theta_{k,L})$ for the actual rank property of the UR-MP channel model.

Proof of Theorem 1: From $\mathbf{h}_k = \sqrt{\frac{M}{L}} \sum_{i=1}^{L} \alpha_{k,i} \mathbf{a}(\theta_{k,i})$ and (7), the m-th component of $\mathbf{h}_k$ is given by

$$h_{k,m} = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} \alpha_{k,i} e^{-\iota\pi\theta_{k,i}(m-1)}. \tag{42}$$

Since $\alpha_{k,i} \sim \mathcal{CN}(0,1)$, for any realized value of $\theta_{k,i}$, the product of $\alpha_{k,i}$ and $e^{-\iota\pi\theta_{k,i}(m-1)}$ follows $\mathcal{CN}(0,1)$ due to the phase invariance property of the circularly-symmetric complex Gaussian distribution. Thus, $h_{k,m}|(\theta_{k,1}, \cdots, \theta_{k,L})$ has Gaussian distribution since it is the normalized sum of L independent Gaussian random variables $\{\alpha_{k,i} e^{\iota\pi\theta_{k,i}(m-1)}\}_{i=1}^{L}$. We have $\mathbb{E}[\mathbf{h}_{k,m}|(\theta_{k,1}, \cdots, \theta_{k,L})] = \mathbf{0}$ and the conditional covariance matrix $\mathbf{R}_{\mathbf{h}}(\theta_{k,1}, \cdots, \theta_{k,L})$ as

$$\begin{aligned}\mathbf{R}_{\mathbf{h}}(\theta_{k,1}, \cdots, \theta_{k,L}) &= \mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H | (\theta_{k,1}, \cdots, \theta_{k,L})] \\ &= \frac{M}{L} \sum_{i=1}^{L} \mathbf{a}(\theta_{k,i}) \mathbf{a}^H(\theta_{k,i}),\end{aligned}$$

and the (m,n)-th element of $\mathbf{R}_{\mathbf{h}}(\theta_{k,1}, \cdots, \theta_{k,L})$ is given by

$$\mathbb{E}[h_{k,m} h_{k,n}^* | \{\theta_{k,i}\}] = \underbrace{\frac{1}{L} \sum_{i=1}^{L} e^{-\iota\pi\theta_{k,i}(m-n)}}_{=: X_{m-n}}, \text{ for } m \neq n. \tag{43}$$

(Note that the rank of $\mathbf{R}_{\mathbf{h}}(\theta_{k,1}, \cdots, \theta_{k,L})$ is L at most.) Hence, we have

$$\mathbf{h}_k|(\theta_{k,1}, \cdots, \theta_{k,L}) \sim \mathcal{CN}(0, \mathbf{R}_{\mathbf{h}}(\theta_{k,1}, \cdots, \theta_{k,L}))$$

Next, we show that $\mathbf{R_h}(\theta_{k,1},\cdots,\theta_{k,L}) \to \mathbf{I}$ in probability as $M\to\infty$, if $\beta>1$. For this, we need to show for any $\epsilon>0$

$$\begin{aligned}&\Pr\{|X_{-M+1}|\le\epsilon,\cdots,|X_{-1}|\le\epsilon,|X_1|\le\epsilon,\cdots,|X_{M-1}|\le\epsilon\}\\&=\Pr\{|X_1|\le\epsilon,|X_2|\le\epsilon,\cdots,|X_{M-1}|\le\epsilon\}\to 1\end{aligned}\tag{44}$$

as $M\to\infty$, where $X_m(\{\theta_{k,i}\})=\frac{1}{L}\sum_{i=1}^{L}e^{-\iota\pi\theta_{k,i}m}$. Note that the random variables $\{X_m\}$ are identically distributed with mean 0 and variance $\frac{1}{L}$ but dependent on one another. Using the union bound and Chebyshev's inequality, we have

$$\begin{aligned}&\Pr\{|X_1|\le\epsilon,|X_2|\le\epsilon,\cdots,|X_{M-1}|\le\epsilon\}\\&=1-\Pr\left\{\bigcup_{i=1}^{M-1}\{|X_i|>\epsilon\}\right\}\\&\ge 1-\sum_{i=1}^{M-1}\Pr\{|X_i|>\epsilon\}\\&\ge 1-\frac{(M-1)}{L\epsilon^2}.\end{aligned}\tag{45}$$

Since $L=M^\beta$ in the considered case, for $\beta>1$ the probability (45) goes to one as $M\to\infty$. Hence, $\mathbf{R_h}(\theta_{k,1},\cdots,\theta_{k,L})$ converges to $\mathbf{I}$ element-wise uniformly in probability as $M\to\infty$, if $\beta>1$. On the other hand, when $\beta<1$, $L=M^\beta<M$ and hence $\mathbf{R_h}(\beta_{k,1},\cdots,\beta_{k,L})$ is rank-deficient. This concludes the proof. ■

APPENDIX B

Proof of Theorem 2: Consider finite $L\ge 1$. Let A be the event that there exists a user k' such that $\theta_{k',1}\in[\vartheta_b-\frac{1}{M},\vartheta_b+\frac{1}{M})$, $|\alpha_{k',1}|\ge L$, and $|\alpha_{k',i}|\le 1$, $\theta_{k',i}\notin[\vartheta_b-\frac{1}{M},\vartheta_b+\frac{1}{M})$ for $i=2,\cdots,L$. Note that we have under the UR-MP model for any k,i

$$\Pr\left\{\theta_{k,i}\in\left[\vartheta_b-\frac{1}{M},\vartheta_b+\frac{1}{M}\right)\right\}=\frac{1}{M}\tag{46}$$

$$\Pr\{|\alpha_{k,i}|\ge L\}=e^{-L^2}.\tag{47}$$

Based on (46) and (47), the asymptotic probability of the event A is given by

$$\begin{aligned}\Pr\{A\}&=1-\Pr\{A^c\}\\&=1-\left(1-\frac{1}{M}e^{-L^2}\left((1-e^{-1})(1-\frac{1}{M})\right)^{L-1}\right)^{K}\\&=1-\left(1-\frac{1}{M}\frac{(M-1)^{L-1}}{M^{L-1}}e^{-L^2}(1-e^{-1})^{L-1}\right)^{c'M}\\&\to e^{-c'e^{-L^2}(1-e^{-1})^{L-1}}>0,\end{aligned}\tag{48}$$

as $M\to\infty$ with $K=\Theta(M)$, where K is replaced with $c'M$ with some $c'\ge 1$ for $K=\Theta(M)$. From the fact that $\mathbb{E}[f(X)]\ge p(A')\mathbb{E}[f(X|A')]$ for a non-negative function $f(X)$, $\mathcal{R}_{\kappa_b}$ is lower bounded by

$$\mathcal{R}_{\kappa_b}\ge\Pr\{A\}\mathbb{E}\left[\log\left(1+\frac{\frac{1}{M}|\mathbf{h}_{\kappa_b}^H\mathbf{a}(\vartheta_b)|^2}{1+\frac{1}{M}\sum_{b'\ne b}|\mathbf{h}_{\kappa_b}^H\mathbf{a}(\vartheta_{b'})|^2}\right)\middle|A\right].\tag{49}$$

Furthermore, the second term in the right-hand side (RHS) of (49) is bounded as

$$\begin{aligned}&\mathbb{E}\left[\log\left(1+\frac{\frac{1}{M}|\mathbf{h}_{\kappa_b}^H\mathbf{a}(\vartheta_b)|^2}{1+\frac{1}{M}\sum_{b'\ne b}|\mathbf{h}_{\kappa_b}^H\mathbf{a}(\vartheta_{b'})|^2}\right)\middle|A\right]\\&\overset{(a)}{\ge}\mathbb{E}\left[\log\left(1+\frac{\frac{1}{L}|\sum_{i=1}^{L}\alpha_{k',i}^*\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_b)|^2}{1+\frac{1}{L}\sum_{b'\ne b}\left|\sum_{i=1}^{L}\alpha_{k',i}^*\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_{b'})\right|^2}\right)\middle|A\right]\\&\overset{(b)}{\ge}\mathbb{E}\left[\log\left(1+\frac{\frac{1}{L}(|\alpha_{k',1}|-(L-1))^2\frac{4}{\pi^2}}{1+\frac{1}{L}(|\alpha_{k',1}|^2+(L-1))\frac{L\pi^2}{3}}\right)\middle|A\right]\\&\overset{(c)}{\ge}\log\left(1+\frac{\frac{4}{L\pi^2}}{1+\frac{(L^2+L-1)\pi^2}{3}}\right)>0,\end{aligned}\tag{50}$$

where (a) holds because the rate of user k' for beam ϑ_b is not larger than that of the optimal user κ_b for beam b; (b) holds by the two facts

$$\left|\sum_{i=1}^{L}\alpha_{k',i}\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_b)\right|^2\ge(|\alpha_{k',1}|-(L-1))^2\frac{4}{\pi^2}\tag{51}$$

and

$$\sum_{b'\ne b}\left|\sum_{i=1}^{L}\alpha_{k',i}\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_{b'})\right|^2\le(|\alpha_{k',1}|^2+(L-1))\frac{L\pi^2}{3}\tag{52}$$

(See Appendix C for (51) and (52).); and (c) follows from the fact that the SINR term is minimized when $|\alpha_{k',1}|=L$ since $|\alpha_{k',1}|\ge L$ under the event A. By (48), (49) and (50), $\mathcal{R}_{\kappa_b}$ is asymptotically lower bounded as

$$\mathcal{R}_{\kappa_b}\ge r_{LB,1}=e^{-c'e^{-L^2}(1-e^{-1})^{L-1}}\log\left(1+\frac{\frac{4}{L\pi^2}}{1+\frac{(L^2+L-1)\pi^2}{3}}\right).$$

This concludes the proof.

APPENDIX C

Proof of eq. (51)*:* We have

$$\begin{aligned}&\left|\sum_{i=1}^{L}\alpha_{k',i}\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_b)\right|\\&=\left|\alpha_{k',1}\mathbf{a}(\theta_{k',1})^H\mathbf{a}(\vartheta_b)+\sum_{i\ne1}\alpha_{k',i}\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_b)\right|\\&\ge\left|\alpha_{k',1}\mathbf{a}(\theta_{k',1})^H\mathbf{a}(\vartheta_b)\right|-\left|\sum_{i\ne1}\alpha_{k',i}\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_b)\right|\\&\ge|\alpha_{k',1}|\left|\mathbf{a}(\theta_{k',1})^H\mathbf{a}(\vartheta_b)\right|-\sum_{i\ne1}|\alpha_{k',i}|\left|\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_b)\right|\\&\overset{(a)}{\ge}(|\alpha_{k',1}|-(L-1))\left|\mathbf{a}(\theta_{k',1})^H\mathbf{a}(\vartheta_b)\right|\overset{(b)}{\ge}(|\alpha_{k',1}|-(L-1))\frac{2}{\pi},\end{aligned}$$

where step (a) is valid because $|\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_b)| \le |\mathbf{a}(\theta_{k',1})^H\mathbf{a}(\vartheta_b)|$ due to $|\theta_{k',1}-\vartheta_b| \le \frac{1}{M} < |\theta_{k',i}-\vartheta_b|$, $i \neq 1$ under the event A (See Fig. 2.), and step (b) follows from the fact that $|\mathbf{a}(\theta_{k',1})^H\mathbf{a}(\vartheta_b)| \ge F_M(\frac{1}{M}) \to \frac{2}{\pi}$ for $|\theta_{k',1}-\vartheta_b| \le \frac{1}{M}$ (See Fig. 2.). Note that $|\alpha_{k',1}| \ge L$ under the event A. This concludes the proof.

Proof of eq. (52)*:* We have

$$\begin{aligned}
&\sum_{b'\neq b}\left|\sum_{i=1}^{L}\alpha_{k',i}\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_{b'})\right|^2 \\
&\le \sum_{b'=1}^{M}\left|\sum_{i=1}^{L}\alpha_{k',i}\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_{b'})\right|^2 \\
&\le \sum_{b'=1}^{M}\left(\sum_{i=1}^{L}|\alpha_{k',i}|\cdot\left|\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_{b'})\right|\right)^2 \\
&\overset{(a)}{\le} \sum_{b'=1}^{M} L\sum_{i=1}^{L}\left(|\alpha_{k',i}|\cdot\left|\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_{b'})\right|\right)^2 \\
&\overset{(b)}{\le} L\sum_{i=1}^{L}|\alpha_{k',i}|^2\sum_{b'=1}^{M}\left|\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_{b'})\right|^2 \\
&\overset{(c)}{\le} L\left(\sum_{i=1}^{L}|\alpha_{k',i}|^2\right)\frac{\pi^2}{3} \overset{(d)}{\le} \left(|\alpha_{k',1}|^2+(L-1)\right)\frac{L\pi^2}{3}
\end{aligned}$$

where (a) follows from Jensen's inequality for the convex function $f(x)=x^2$; (b) holds by interchanging the order of summation; (c) follows from the inequality (53); and (d) holds because $|\alpha_{k',i}| \le 1$ for $i \neq 1$ under the event A. For step (c), we have

$$\sum_{b'=1}^{M}|\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_{b'})|^2 \le 2\sum_{j=1}^{M/2}\frac{1}{j^2} \le 2\sum_{j=1}^{\infty}\frac{1}{j^2} \le \frac{\pi^2}{3}, \quad (53)$$

where the first inequality holds due to Lemma 2 (see the stepwise bounding function for $|\mathbf{a}(\theta_{k',i})^H\mathbf{a}(\vartheta_{b'})|$ in Fig. 2).

APPENDIX D

Proof of Theorem 3: Let $\bar{A}$ be the event that there exists a user $\bar{k}$ such that $|\theta_{\bar{k},1}-\vartheta_b| \in [0,\frac{1}{M}]$ and $|\alpha_{\bar{k},1}| \in [\sqrt{L},\sqrt{2L}]$, and $|\theta_{\bar{k},i}-\vartheta_b| \in [\frac{1}{2},1]$ and $|\alpha_{\bar{k},i}| \le 1$ for $i=2,3,\cdots,L$. Then, based on $\theta_{k,i} \overset{i.i.d.}{\sim} \mathrm{Unif}[-1,1]$ and $\alpha_{k,i} \overset{i.i.d.}{\sim} \mathcal{CN}(0,1)$, the asymptotic probability of the event $\bar{A}$ is given by

$$\begin{aligned}
\Pr\{\bar{A}\} &= 1-\Pr\{\bar{A}^c\} \\
&= 1-\left(1-\left(\frac{e^{-L}-e^{-2L}}{M}\right)\left(\frac{1-e^{-1}}{2}\right)^{L-1}\right)^K \\
&\overset{(a)}{>} 1-\left(1-\frac{1}{Me^{3L}}\right)^K \\
&= 1-e^{K\log(1-\frac{1}{Me^{3L}})} \\
&\overset{(b)}{=} 1-e^{-\frac{Me^{c_uL}}{Me^{3L}}+O(\frac{Me^{c_uL}}{M^2e^{6L}})} \\
&= 1-e^{-\frac{Me^{c_uL}}{Me^{3L}}(1+o(1))} \overset{(c)}{\to} 1 \text{ as } M\to\infty, \quad (54)
\end{aligned}$$

where (a) is by direction comparison and (b) follows from $\log(1-x)=-x+O(x^2)$ and replacing K with Me^{c_uL} by Condition *(C.3)*, and (c) holds by Condition *(C.1)* and $c_u > 3$ in Condition *(C.3)*.

By applying the same technique used in (49), it is left only to show

$$\mathbb{E}\left[\log\left(1+\frac{X}{1+Y}\right)\middle|\,\bar{A}\right] \not\to 0 \text{ as } M\to\infty, \quad (55)$$

where $X=\frac{1}{L}|\sum_{i=1}^{L}\alpha_{\bar{k},i}^*\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_b)|^2$ and $Y=\frac{1}{L}\sum_{b'\neq b}|\sum_{i=1}^{L}\alpha_{\bar{k},i}^*\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_{b'})|^2$. To obtain a non-trivial lower bound on the rate in (55), we first handle the term X and the term Y next. Under the event $\bar{A}$, we have $|\theta_{\bar{k},1}-\vartheta_b| \in [0,\frac{1}{M}]$, $|\alpha_{\bar{k},1}| \in [\sqrt{L},\sqrt{2L}]$, and thus

$$|\alpha_{\bar{k},1}^*\mathbf{a}(\theta_{\bar{k},1})^H\mathbf{a}(\vartheta_b)| \gtrsim \frac{2}{\pi}|\alpha_{\bar{k},1}| \ge \frac{2}{\pi}\sqrt{L} \quad (56)$$

by $|\mathbf{a}(\theta_{\bar{k},1})^H\mathbf{a}(\vartheta_b)| \ge F_M(|\frac{1}{M}|) \to \frac{2}{\pi}$, since $F_M(\tilde{\theta})$ is a monotone decreasing function from $\tilde{\theta}\in[0,2/M]$ [5], [15]. (See Fig. 2.) Also, we have under the event $\bar{A}$

$$\begin{aligned}
\left|\sum_{i=2}^{L}\alpha_{\bar{k},i}^*\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_b)\right| &\le \sum_{i=2}^{L}|\alpha_{\bar{k},i}^*\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_b)| \\
&\overset{(a)}{\le} \sum_{i=2}^{L}|\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_b)| \\
&\overset{(b)}{\le} \sum_{i=2}^{L}\frac{1}{M|\theta_{\bar{k},i}-\vartheta_b|} \overset{(c)}{\le} 2 \quad (57)
\end{aligned}$$

where (a) follows from $|\alpha_{\bar{k},i}| \le 1$ for $i \neq 1$ under the event $\bar{A}$; (b) holds by Lemma 1; (c) is by $|\theta_{\bar{k},i}-\vartheta_b| \ge \frac{1}{2}$ for $i \neq 1$ under the event $\bar{A}$ and by $L/M \le 1$ from Condition *(C.2)*. Using (56) and (57), we have a lower bound on the numerator term X in the SINR expression in (55) as

$$\begin{aligned}
X &= \frac{1}{L}\left|\alpha_{\bar{k},1}^*\mathbf{a}(\theta_{\bar{k},1})^H\mathbf{a}(\vartheta_b)+\sum_{i=2}^{L}\alpha_{\bar{k},i}^*\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_b)\right|^2 \\
&\gtrsim \frac{1}{L}\left|\frac{2}{\pi}\sqrt{L}-2\right|^2 \to \frac{4}{\pi^2}, \quad (58)
\end{aligned}$$

as $M\to\infty$, since $L\to\infty$ as $M\to\infty$ by Condition *(C.1)*. By replacing the numerator term X in the SINR expression in (55) with its lower bound (58), the rate in (55) is lower bounded by $\mathbb{E}\left[\log\left(1+\frac{4/\pi^2}{1+Y}\right)\middle|\,\bar{A}\right]$. By the convexity of the function $f(x)=\log(1+\frac{1}{1+x})$ and Jensen's inequality, we have

$$\mathbb{E}\left[\log\left(1+\frac{4/\pi^2}{1+Y}\right)\middle|\,\bar{A}\right] \ge \log\left(1+\frac{4/\pi^2}{1+\mathbb{E}[Y|\bar{A}]}\right). \quad (59)$$

We now find an upper bound on $\mathbb{E}[Y|\bar{A}]$. To do so, we first derive lower bounds on the terms in

$Y = \frac{1}{L}\sum_{b'\neq b} |\sum_{i=1}^{L} \alpha_{\bar{k},i}^{*}\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_{b'})|^2$. By Lemma 1, we have

$$\sum_{b'\neq b} |\alpha_{\bar{k},1}^{*}\mathbf{a}(\theta_{\bar{k},1})^H\mathbf{a}(\vartheta_{b'})|^2 \leq |\alpha_{\bar{k},1}|^2 \sum_{b'\neq b} |\mathbf{a}(\theta_{\bar{k},1})^H\mathbf{a}(\vartheta_{b'})|^2 \leq 2|\alpha_{\bar{k},1}|^2 \sum_{j=1}^{M/2} \frac{1}{j^2} \leq |\alpha_{\bar{k},1}|^2 \frac{\pi^2}{3}, \tag{60}$$

where the second inequality is obtained by re-arranging the indices of $\{\vartheta_{b'}\}_{b'\neq b}$ in the order of closeness to $\theta_{\bar{k},1}$ and using Lemma 2 and the fact that $\vartheta_{b'+1} = \vartheta_{b'} + \frac{2}{M}$ described in (13). We also have under the event $\bar{A}$, for $i \neq 1$

$$\begin{aligned}
&\mathbb{E}\left[\sum_{b'\neq b} |\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_{b'})|^2 \middle| \bar{A}\right] \\
&\overset{(a)}{\leq} M\mathbb{E}\left[|\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_b+1)|^2 \middle| \bar{A}\right] \\
&\overset{(b)}{\leq} M\mathbb{E}\left[\left|\sum_{j=1}^{M/2} \frac{1}{j}\mathbf{1}_{[\frac{2(j-1)}{M},\frac{2j}{M})}(|\tilde{\theta}|)\right|^2 \middle| \bar{A}\right] \\
&\overset{(c)}{\leq} M\left(\sum_{j=1}^{M/4} \frac{2}{jM}\right)^2 \\
&\overset{(d)}{\leq} \frac{(2(\log\frac{M}{4}+1))^2}{M} \to 0, \text{ as } M\to\infty,
\end{aligned} \tag{61}$$

where (a) follows from the fact that since $|\theta_{\bar{k},i} - \vartheta_b| \in [\frac{1}{2},1] \Leftrightarrow \theta_{\bar{k},i} \in \mathcal{D} := [\vartheta_b + 1 - \frac{1}{2}, \vartheta_b + 1 + \frac{1}{2}]$ under the event $\bar{A}$ due to the periodicity of 2 in the angle domain, the conditional expectation is maximized when $\vartheta_{b'}$ is located in the center of the domain $\mathcal{D}$, i.e., $\vartheta_{b'} = \vartheta_b + 1$ under the assumed uniform distribution of $\theta_{\bar{k},i}$; (b) holds due to Lemma 2 with $\tilde{\theta} := \theta_{\bar{k},i} - (\vartheta_b + 1)$; (c) holds because $\Pr\left\{\mathbf{1}_{[\frac{2(j-1)}{M},\frac{2j}{M})}(|\theta_{\bar{k},i} - (\vartheta_b+1)|)|\bar{A}\right\} = \frac{2}{M}$ for $j = 1, 2, \cdots, \frac{M}{4}$; and (d) is valid because $\sum_{j=1}^{M/4}\frac{2}{j} \leq 2(\log\frac{M}{4}+1)$.

Based on (60) and (61), $\mathbb{E}[Y|\bar{A}]$ is lower bounded as

$$\begin{aligned}
\mathbb{E}[Y|\bar{A}] &\overset{(a)}{\leq} \mathbb{E}\left[\frac{1}{L}\sum_{b'\neq b} 2|\alpha_{\bar{k},1}^{*}|^2|\mathbf{a}(\theta_{\bar{k},1})^H\mathbf{a}(\vartheta_{b'})|^2 \middle| \bar{A}\right] \\
&\quad + \mathbb{E}\left[\frac{1}{L}\sum_{b'\neq b} 2\sum_{i\neq 1}^{L} |\alpha_{\bar{k},i}^{*}|^2|\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_{b'})|^2 \middle| \bar{A}\right] \\
&\overset{(b)}{\leq} \frac{4L}{L}\frac{\pi^2}{3} + 2\mathbb{E}\left[\sum_{b'\neq b} |\mathbf{a}(\theta_{\bar{k},i})^H\mathbf{a}(\vartheta_{b'})|^2 \middle| \bar{A}\right], \\
&\overset{(c)}{\lesssim} \frac{4L}{L}\frac{\pi^2}{3},
\end{aligned} \tag{62}$$

where (a) holds because $(\sum_{i=1}^{n} x_i)^2 \leq \sum_{i=1}^{n} n x_i^2$, (b) follows from (60) with $|\alpha_{\bar{k},1}| \leq \sqrt{2L}$ and $|\alpha_{\bar{k},i}| \leq 1$ for $i \neq 1$ under the event $\bar{A}$, and (c) holds due to (61).

Finally, by (62), the rate (59) is asymptotically lower bounded by

$$\mathcal{R}_{\kappa_b} \gtrsim r_{LB,2} = \log\left(1 + \frac{4/\pi^2}{1 + 4\pi^2/3}\right). \tag{63}$$

This concludes the proof.

REFERENCES

[1] G. Lee, Y. Sung, and M. Kountouris, "On the performance of randomly directional beamforming between line-of-sight and rich scattering channels," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC'15)*, Jun. 2015, pp. 141–145.

[2] S. Sun, T. S. Rappaport, R. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[3] S. Sun and T. S. Rappaport, "Wideband mmWave channels: Implications for design and implementation of adaptive beam antennas," in *Proc. IEEE Int. Microw. Symp. (IMS)*, Jun. 2014, pp. 1–4.

[4] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE Global Telecommun. Conf. (Globecom)*, Dec. 2013, pp. 3679–3684.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *Proc. IEEE 22nd Eur. Signal Process. Conf. (EUSIPCO'14)*, Sep. 2014, pp. 76–80.

[6] A. Alkhateeb, O. E. Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[7] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[8] A. Alkhateeb, G. Leus, and R. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Mar. 2015.

[9] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[10] G. Lee and Y. Sung, "A new approach to user scheduling in massive multi-user MIMO broadcast channels," arXiv:1403.6931, Mar. 2014.

[11] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[12] M. Kountouris, D. Gesbert, and T. Salzer, "Enhanced multiuser random beamforming: Dealing with the not so large number of users case," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1536–1545, Oct. 2008.

[13] A. Tomasoni, G. Caire, M. Ferrari, and S. Bellini, "On the selection of semi-orthogonal users for zero-forcing beamforming," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2009, pp. 1100–1104.

[14] H. Hur, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[15] G. Lee, Y. Sung, and J. Seo, "Randomly-directional beamforming in millimeter-wave multi-user MISO downlink," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1086–1100, Feb. 2016.

[16] T. Al-Naffouri, M. Sharif, and B. Hassibi, "How much does transmit correlation affect the sum-rate scaling of MIMO Gaussian broadcast channels?" *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 562–572, Feb. 2009.

[17] T. S. Rappaport, E. Ben-Dor, J. N. Murdock, and Y. Qiao, "38 GHz and 60 GHz angle-dependent propagation for cellular & peer-to-peer wireless communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 4568–4573.

[18] J. Seo, Y. Sung, G. Lee, and D. Kim, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1228–1242, Mar. 2016.

[19] A. Alkhateeb, O. E. Ayach, G. Leus, and R. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. Inf. Theory Appl. Workshop*, San Diego, CA, USA, 2013, pp. 1–5.

[20] R. S. Strichartz, *The Way of Analysis*. Boston, MA, USA: Jones and Bartlett, 2000.

[21] G. H. Tucci and P. A. Whiting, "Eigenvalue results for large scale random Vandermonde matrices with unit complex entries," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3938–3954, Jun. 2011.
[22] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 246–273, Feb. 2013.
[23] J. L. Vicario, B. Bosisio, C. Anton-Haro, and U. Spagnolini, "Beam selection strategies for orthogonal random beamforming in sparse networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3385–3396, Sep. 2008.
[24] W. Xu and C. Zhao, "Two-phase multiuser scheduling for multiantenna downlinks exploiting reduced finite-rate feedback," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1367–1380, Mar. 2010.
[25] R. Zakhour and D. Gesbert, "A two-stage approach to feedback design in multi-user MIMO channels with limited channel state information," in *Proc. IEEE 18th Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, 2007, pp. 1–5.
[26] C. Swannack, G. Wornell, and E. Uysal-Biyikoglu, "MIMO broadcast scheduling with quantized channel state information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2006, pp. 1788–1792.
[27] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.
[28] O. E. Ayach, S. Rajagopal, S. Abu-Surra, P. Zhouyue, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[29] W. Roh et al., "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
[30] H. Ji, Y. Kim, J. Lee, E. Onggosanusi, Y. Nam, J. Zhang, B. Lee, and B. Shim, "Overview of full-dimension MIMO in LTE-advanced pro," arXiv preprint arXiv:1601.00019v1, Dec. 2015.

Gilwon Lee (S'10) received the B.S. and M.S. degrees in electrical engineering from KAIST, Daejeon, Korea, in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree at the Wireless Information Systems Research Group, KAIST. His research interests include design and analysis of large-scale MIMO systems and signal processing for next wireless communications.

Youngchul Sung (S'92–M'93–SM'09) received the B.S. and M.S. degrees from Seoul National University, Seoul, Korea, in electronics engineering, in 1993 and 1995, respectively, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2005. After working at LG Electronics, Ltd., Seoul, Korea, from 1995 to 2000. From 2005 to 2007, he was a Senior Engineer with the Corporate R&D Center of Qualcomm, Inc., San Diego, CA, USA, and participated in design of WCDMA base station modem. Since 2007, he has been on the Faculty of the Department of Electrical Engineering in KAIST, Daejeon, Korea. His research interests include signal processing for communications, statistical signal processing, asymptotic statistics, and information geometry. He is currently a member of UNESCO/Netexplo University Advisory Board, Signal, and Information Processing Theory and Methods (SIPTM) Technical Committee of Asia-Pacific Signal and Information Processing Association (APSIPA), Vice-Chair of the IEEE ComSoc Asia-Pacific Board ISC, and was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2012 to 2014.

Marios Kountouris (S'04–M'08–SM'15) received the Diploma degree in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from the Ecole Nationale Supérieure des Télécommunications (Télécom ParisTech), Paris, France, in 2004 and 2008, respectively. His doctoral research was carried out at Eurecom Institute, France, and it was funded by Orange Labs, France. From February 2008 to May 2009, he has been with the Department of ECE, University of Texas at Austin, Austin, TX, USA, as a Research Associate, working on wireless ad hoc networks under DARPAs IT-MANET program. From June 2009 to December 2013, he has been an Assistant Professor with the Department of Telecommunications at Supélec, France, where he is currently an Associate Professor. From March 2014 to February 2015, he has been an Adjunct Professor with the School of the EEE at Yonsei University, Seoul, Korea. Since January 2015, he has been a Principal Researcher with the Mathematical and Algorithmic Sciences Laboratory, Huawei Technologies, France. He has authored several papers and patents all in the area of communications, wireless networks, and signal processing. He has served as technical program committee member for several top international conferences and has served as a Workshop Chair for the IEEE Globecom 2010 Workshop on Femtocell Networks, the IEEE ICC 2011 Workshop on Heterogeneous Networks, and the IEEE Globecom 2012 Workshop on Heterogeneous and Small Cell Networks. He is currently an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE WIRELESS COMMUNICATION LETTERS, *EURASIP Journal on Wireless Communications and Networking*, and *Journal of Communications and Networks* (JCN). He is a Professional Engineer of the Technical Chamber of Greece. He was the recipient of the 2013 IEEE ComSoc Outstanding Young Researcher Award for the EMEA Region, the 2014 EURASIP Best Paper Award for EURASIP Journal on Advances in Signal Processing (JASP), the 2012 IEEE SPS Signal Processing Magazine Award, the IEEE SPAWC 2013 Best Student Paper Award, and the Best Paper Award in Communication Theory Symposium at IEEE Globecom 2009.