Location-Aided User Selection and Sum-Rate Analysis for mmWave NOMA

Igbafe Orikumhi, Chee Yen Leow, and Sunwoo Kim

***Abstract*—In this paper, we propose a user selection scheme based on location-aided interference prediction to reduce the training overhead in a non-orthogonal multiple access (NOMA) system. First, we cluster the users based on their location information, enabling the use of non-orthogonal pilot sequence within a cluster and orthogonal pilot sequence between different clusters to reduce the uplink pilot training length. Secondly, we exploit the location information in the computation of the covariance matrices, enabling the prediction of the interference between users. The predicted interference is employed to select the set of users with minimum interference for uplink channel estimation and downlink NOMA data transmission. Finally, the achievable sum-rate of the massive multiple-input multiple-output millimeter wave NOMA system is analyzed. The analytical and numerical results reveal that the location information can be exploited for user selection to reduce the effect of pilot contamination, enhancing the uplink channel estimation and downlink achievable sum-rate.**

***Index Terms*— Location-aware communication, mmWave, multi-user beamforming, NOMA, power allocation, user clustering.**

I. Introduction

THE increasing demand for mobile broadband has become a major challenge in wireless communication. To address this issue, millimeter wave (mmWave) communication has been proposed for the fifth-generation (5G) networks due to the large bandwidth available in this frequency spectrum [2]. However, the mmWave spectrum is known to suffer from severe path-loss and low penetration, a challenge that has led to the use of massive antenna arrays and directional beamforming to compensate for the losses [3].

Manuscript received May 2, 2022 revised October 4, 2022; approved for publication by Xingwang Li Li Division 2 Editor, October 30, 2022.

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2017-0-01637) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(No. 2020R1I1A1A01073438).

This work of C.Y. Leow was supported in part by the Ministry of Higher Education Malaysia under Fundamental Research Grant Scheme FRGS/1/2020/TK0/UTM/02/68, in part by Universiti Teknologi Malaysia under Grant 05E07, Grant 08G83, and Grant 09G15 and in part by Hanyang University Under Grant 4B675. This work was presented in part [1] at the international conference on ICT convergence JeJu, 2020, South Korea.

I. Orikumhi is with the Department of Electronic Engineering, Hanyang University, Seoul, 04763, South Korea, email: oigbafe2@hanyang.ac.kr.

C. Y. Leow is with Wireless Communication Centre, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Skudai, 81310, Johor, Malaysia, email: bruceleow@fke.utm.my.

S. Kim is with the Department of Electronic Engineering, Hanyang University, Seoul, 04783, South Korea, email: remero@hanyang.ac.kr.

I. Orikumhi is the corresponding author.

Digital Object Identifier: 10.23919/JCN.2022.000049

The use of massive antenna provides numerous advantages [4]–[8] such as (i) channel hardening effects which result from averaging out the small scale fading, (ii) energy efficiency improvement with directional transmission towards the desired user equipment (UE) (iii) reduced interference with the aid of narrow beam transmission, and (iv) increased spectral efficiency by supporting a high number of UEs. Moreover, the use of highly directional beams provides large antenna array gains with small inter-beam interference. This feature can be exploited to serve multiple UEs within each beam or a cluster of beams.

Furthermore, the hardware cost and the number of radio frequency (RF) chains in mmWave devices are usually much smaller than the number of antennas which limits the number of active UEs that can be served simultaneously. To support a higher number of UEs in mmWave 5G networks, more UEs can be multiplexed in the power domain using the principle of non-orthogonal multiple access (NOMA) for a given resource block (e.g., time/frequency) [9]. NOMA has become a promising solution and has been studied for 5G communications. mmWave NOMA allows multiple users to share the same resource block by taking advantage of the different rate requirements of the UEs and the highly correlated mmWave channels [10].

Although more bandwidth resource are available in mmWave band, combining NOMA with mmWave is still important for the following reasons: 1) The highly correlated mmWave channel can potentially degrade systems performance. But such correlation is ideal for application of NOMA. 2) The increasing growth of mobile device and mobile internet services will soon dwarf the spectrum resource provided by the mmWave band. Hence, the need for further improvement of spectral efficient schemes. 3) The 5G system and beyond are designed to support massive number of devices in a small area, hence, combining NOMA and mmWave can improve systems design and resource allocation.

The goal of NOMA is to maximize the system throughput subject to the UEs individual rate constraints [11]. Specifically, the NOMA scheduler estimates a set of system utilities at each resource block by activating any subset of UEs. The base station (BS) can then schedule the set of active UEs in a resource block based on the individual UE's quality of service (QoS) [12]–[14]. In [14], a sort-based user scheduling with partial channel state information (CSI) and power allocation (PA) strategy was proposed. In [15], the authors proposed an optimal user scheduling and PA strategies for multi-beam and multi-user mmWave NOMA system, where the authors focused on the use of random beamforming to reduce the

1229-2370/22/$10.00 © 2022 KICS

CSI overhead. A joint PA and beamforming algorithm is proposed to maximize the sum-rate in a two-user mmWave NOMA scenario in [16]. These methods [12]–[16] suffer from large training overhead in CSI acquisition which in turn decreases the achievable sum-rate or poor CSI due to random beamforming. Unlike, [17], [18], where perfect CSI are assumed and the training overhead is ignored, in this paper, we seek to exploit the location information for user interference prediction and pairing while minimize the training overhead.

The challenges posed by high training overhead and random beamforming are exacerbated using highly directional beamforming in massive multiple-input multiple-output (mMIMO) systems. This is because the BS has to search over a large angular space to estimate the channel of the UEs [19], [20] before pairing the UEs for downlink transmission. In [21], it was shown that the challenge of CSI acquisition at the transmitter could be alleviated by increasing the number of antennas at the BS. The idea is rooted in the law of large numbers, which shows that asymptotic orthogonality between the vector channels of the desired UE and that of a randomly selected interfering UE can be achieved as the number of antennas tends to infinity. Hence, by aligning the beam vectors of a massive antenna BS with the channel of the desired UE, interference can be suppressed. However, this conclusion only holds in the absence of pilot contamination. Pilot contamination arises because the available orthogonal pilot sequence length is limited by the length of the channel coherence time, leading to several UEs employing the same pilot sequence.

The existing literature on pilot decontamination methods for time division duplex (TDD) MIMO systems can be grouped into two broad categories: Pilot-based and subspace-based approaches. In pilot-based approaches, each UE transmits independent pilots to the BS in a non-overlapping fashion [22]–[25]. In such settings, the frame structure is modified such that the UEs transmit with orthogonal pilots or orthogonal times slot [24], [25]. In subspace-based approaches, the second-order statistics of the desired user and the interfering UE's channel is exploited [26], [27]. Such methods utilizes covariance-aided channel estimation [28]–[32]. Based on the spatial correlation of the covariance matrices, UEs that are spatially compatible can be identified with the use of non-diagonal covariance matrices. In [33], a closed-form analysis of the downlink achievable sum-rate exploiting the statistical second-order channel covariance matrices is derived. In recent studies, location information has attracted increasing research interest. Recently, location-aided covariance matrix channel estimation and pilot contamination avoidance in multiple cell scenario [34], [35] have been studied. Studies have shown that location-aided communication is a promising technology to reduce CSI acquisition overhead [36]. In [37], the prior location information of the UE is exploited to speed up the initial access phase in a mmWave vehicular communication utilizing adaptive channel estimation and beamforming. It is shown that location-aided channel estimation and data transmission can be greatly improved with reduced implementation complexity. In [38], a location based MIMO-NOMA is proposed for low complexity user selection. However, the proposed scheme required additional radio resource (frequency, and time) to obtain the channel state information of the user which is then used for power allocation. Hence the location information is not applied efficiently. Furthermore, studies [39], [40] have shown that due to the height of the BS, the channel angle spread observed at BS to the UEs is small. As a consequence, UEs can be clustered based on the line-of-sight (LOS) angle of departure (AOD) differences, where the AOD of the LOS path can be acquired from the location information. Since the difference between the AOD of UEs within the same cluster may be small, we propose a location-aided interference prediction based user selection in NOMA systems to improve the channel estimation and minimize the effect of interference.

Specifically, to reduce the intra-cluster interference caused by pilot contamination from UEs within the same cluster, an efficient location-aided interference prediction is proposed which can effectively distinguish the users with different AODs. To this end, we propose a location-aided interference prediction scheme and user selection to reduce the training overhead, intra-cluster and inter-cluster interference and the channel estimation errors. Moreover, asymptotically, we prove that the intra-cluster interference caused by the users with non-overlapping AODs can be perfectly eliminated when the number of BS antennas tends to infinity. For a finite number of antennas and overlapping angular spread, we show that the interference from UEs can be predicted from the location information and hence, UEs with minimum inter-cluster interference can be scheduled for uplink channel estimation and downlink data transmission.

Our contributions are summarized as follows.

- Since the pilot length is limited by the channel coherence time, we propose a location-aided interference prediction based user selection for NOMA systems to reduce the uplink training overhead, inter-user interference and intra-cluster inference. By utilizing the subspace of the desired UE and that of the predicted interfering UEs, we show that as the number of transmit antenna becomes large, the estimated channel of the desired UE with low training overhead tends toward the estimated channel of the scheme with optimal training length.
- Finally, we analyze the performance of the downlink achievable sum-rate NOMA scheme while considering the effect of channel estimation error, imperfect successive interference cancellation (SIC), interference from other UEs and location information error.

Notations and Organization

Throughout this paper, matrices and vector symbols are represented by uppercase and lowercase boldface respectively. $\mathbf{A}^*$, $\mathbf{A}^T$, and $\mathbf{A}^H$ represent the conjugate, transpose and Hermitian transpose of the matrix $\mathbf{A}$, respectively. The mathematical expectation and variance of matrix $\mathbf{A}$ is denoted as $\mathbb{E}[\mathbf{A}]$ and $\mathbb{V}\mathrm{ar}[\mathbf{A}]$ respectively. $\mathrm{tr}(\mathbf{A})$ represents the trace of matrix $\mathbf{A}$. The Kronecker product between two matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted as $\mathbf{A} \otimes \mathbf{B}$.

The rest of the paper is organized as follows. In Section II, the location information, system description and channel

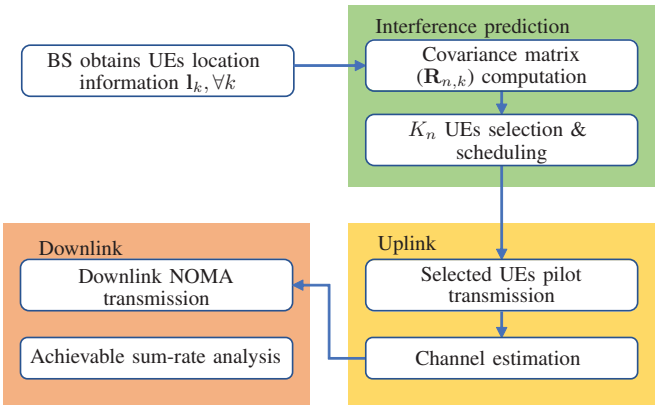

Fig. 1. Design flow for proposed location-aided mmWave-NOMA.

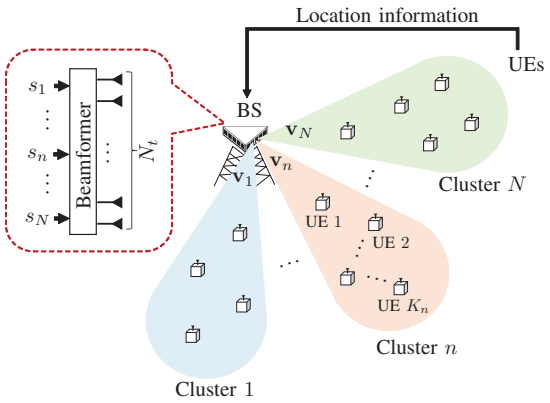

Fig. 2. Proposed mmWave-NOMA downlink MISO scenarios with K_n selected UEs for each cluster.

model are described. In Section III, we present the proposed UE selection based on location-aided interference prediction. The uplink channel estimation and downlink NOMA transmission are discussed in Section IV, followed by the numerical results and discussion in Section V, and finally, the paper is concluded in Section VI. A summary of the proposed system flow is presented in Fig. 1.

II. Location Information, System Description and Channel Model

The proposed scheme relies on exploiting the location information of the UEs. In this section, we present the method to exploit location information and describe the communication scenario.

A. Exploiting Location Information for User Grouping

With the recent advancement in localization and UE position estimation techniques, we assume the BS has access to the location information of the UEs. Such information can be obtained at the initial access phase or by UE position estimation (see [41] and reference therein). However, since the location may not be perfect, we discuss the impact of location errors on the proposed scheme. For analytical tractability, we assume a slow fading environment such that the channel coherence time is much larger than the signal duration. The assumption allows the UEs to update the BS with their location information after the initial access phase.

Let $\mathbf{l}_{\text{BS}} = [x_{\text{BS}}, y_{\text{BS}}]$ denote the location of the BS and $\mathbf{l}_{n,k} = [x_{n,k}, y_{n,k}]$ denotes the location of the kth UE in the nth cluster respectively. Based on the location information, the BS can evaluate the AOD of the kth UE in the nth cluster as follows

$$\theta_{n,k} = \frac{\pi}{2} - \arctan\left(\frac{x_{n,k} - x_{\text{BS}}}{y_{n,k} - y_{\text{BS}}}\right), \quad \forall k. \tag{1}$$

The distance between the BS and the kth UE in the nth cluster can be evaluated as $d_{n,k} = \|\mathbf{l}_{\text{BS}} - \mathbf{l}_{n,k}\|_2$. Define the set of predefined beam vectors as $\mathcal{V} = \{\mathbf{v}_1, \cdots, \mathbf{v}_N\}$ for N total clusters where $\mathbf{v}_n = \mathbf{a}(\phi_n)$ in which the discrete pointing angles ϕ_n is defined as

$$\phi_n = \arccos\left(1 - \frac{2(n-1)}{N-1}\right) \quad \text{for } n = 1, 2, \cdots, N, \tag{2}$$

and $\mathbf{a}(\phi_n)$ is the steering vector. In this paper, a uniform linear array antenna with half wavelength antenna spacing is assumed at the BS and the array response vector is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N_t}}\left[1, e^{-j\pi\cos\theta}, \cdots, e^{-j\pi(N_t-1)\cos\theta}\right]^T, \tag{3}$$

where N_t is the number of transmit antennas at the BS.

With the AOD of the kth UE obtained from the location information, the BS can assign the UE to the best beam cluster as follows

$$\mathbf{v}_{n,k} = \arg\min_{\mathbf{v}\in\mathcal{V}} \|\mathbf{v} - \mathbf{a}(\theta_{n,k})\|_2^2, \tag{4}$$

for $n = 1, \cdots, N$, $k = 1, \cdots, K_n$, where $\mathbf{v}_{n,k}$ is the precoder to the kth UE in the nth cluster.

B. System Description

We consider a mMIMO NOMA communication system operating in the mmWave band. We assume a single BS with N_t antennas serving a total of K UEs, each equipped with a single antenna as shown in Fig. 2. We define P_T as the total power constraint at the BS, then P_T can be expressed as

$$P_T = \sum_{n=1}^{N}\sum_{k=1}^{K_n} P_{n,k}, \tag{5}$$

where N is the total number of beam clusters. We assume that K_T UEs are deployed in each cluster from which K_n UEs subset are selected for uplink channel estimation and downlink data transmission in the nth cluster and $P_{n,k}$ is the power allocated to the kth UE in the nth cluster. We assume that $N \leq N_{\text{RF}}$, where N_{RF} is the number of RF chains at the BS.

C. Channel Model

The uplink channel between the kth UE in the nth cluster and the BS is denoted as $\mathbf{h}_{n,k} \in \mathbb{C}^{N_t}$. We assume the channel consists of L multi-path components and can be modelled as [15]

$$\mathbf{h}_{n,k} = \sqrt{N_t} \sum_{l=1}^{L} \alpha_{n,k}^{l} \mathbf{a}\left(\theta_{n,k}^{l}\right), \tag{6}$$

where $\alpha_{n,k}^{l} \sim \mathcal{CN}\left(0, \rho_{n,k}\right)$ is the complex gain of the lth path to the kth UE in the nth cluster and $\rho_{n,k} = d_{n,k}^{-\delta}$ depends on the large scale path-loss between the BS and kth UE in the nth cluster with path-loss exponent δ. Given that the location information is known, we assume that the channel has an arbitrary AOD distribution but has a support $[\theta_{n,k}^{\max}, \theta_{n,k}^{\min}]$ centered on $\theta_{n,k}$ obtained from the location information with a probability density function (PDF) of $p(\theta_{n,k})$.

Under the assumption of large number of antennas in mMIMO mmWave systems, the number of paths to each UE is small which gives rise to $\mathbf{h}_{n,k}$ having a zero-mean Gaussian distribution with a covariance matrix [35]

$$\begin{aligned}\mathbf{R}_{n,k} &= \mathbb{E}\left[\mathbf{h}_{n,k}\mathbf{h}_{n,k}^{H}\right] \\ &= \rho_{n,k} \int p(\theta_{n,k}) \mathbf{a}\left(\theta_{n,k}\right) \mathbf{a}^{H}\left(\theta_{n,k}\right) \mathrm{d}\theta_{n,k}. \end{aligned} \tag{7}$$

To model the small-scale fading effects, we assume a LOS link between the BS and the UE since the LOS component typically dominates the mmWave wireless propagation. We note from (7) that the covariance matrix depends on the small-scale and large scale fading, where the large scale fading is obtained from the location information and the small scale is inferred from (6). Hence, the evaluation of the covariance matrix is carried out via averaging over this effects. The large scale is approximately constant over the coherence time interval while the small-scale fading varies rapidly and is handled by averaging the entries. In this paper, we exploit the covariance matrix $\mathbf{R}_{n,k}$ for UE selection. Note that the user's channel covariance matrix is computed based on the location information without the need of estimating the covariance matrix directly.

To reduce the overhead in CSI acquisition, we assume a scenario where the UEs within a cluster share the same pilot sequence (non-orthogonal pilots) while employing orthogonal pilots between different clusters. Hence, the goal of the BS is to estimate the uplink channel of the desired UE in the nth cluster in the presence of $K_n - 1$ interfering UEs arising from pilot contamination.

III. Proposed Location-Aided Interference Prediction based UE Selection

It is worth noting that the existing scheme assumes perfect or partial CSI, thereby ignoring the cost of uplink channel estimation and user selection overhead in NOMA system. In this section, we propose a low overhead user selection scheme. Given the location information, we predict the interference between the UEs and exploit the predicted interference for UEs selection.

A. Location-Aided Interference Prediction for UE selection

Given a mMIMO scenario, the performance of the multi-user channel estimation is dependent on the degree to which the subspace of the covariance matrices of the desired UE overlaps that of the other $K_n - 1$ UEs in nth cluster. In this paper, we define the desired UE in the nth cluster as the UE with the strongest channel coefficients to the BS. By exploiting the covariance matrix $\mathbf{R}_{n,k}$ and the AOD obtained from the location information, we can select a set of UEs within a cluster to minimize intra-cluster interference. Hence, the BS can estimate the channel of only the selected UEs in the uplink channel. The covariance matrix, $\mathbf{R}_{n,k}$ is employed in solving the spectral decomposition problem of the high-order uniform linear antenna arrays. Let $\mathbf{R}_{n,k}^{top}$ denotes the Toeplitz assumption of $N_t \times N_t$ matrix [42]. Thus, we can define as follows:

Definition 1: The covariance matrix $\mathbf{R}_{n,k}$ is an $N_t \times N_t$ Toeplitz matrix $\{\mathbf{R}_{n,k}\}_{i,j} = r_{i,j}$ for $i, j = 0, 1, \cdots, N_t - 1$ where $r_{i,j} = r_{i-j}$ and $r_{i,j}$ is the ith row and jth column element of $\{\mathbf{R}_{n,k}\}$.

Proposition 1: The covariance matrix $\mathbf{R}_{n,k}$ is equivalent to a circulant matrix $\mathbf{R}_{n,k}^{circ}$, i.e.,

$$\mathbf{R}_{n,k}^{top} \approx \mathbf{R}_{n,k}^{circ}. \tag{8}$$

Proof: Refer to Appendix A.

Remark 1: From proposition 1 and the assumption of mMIMO at the BS (large N_t), the channel covariance matrices obtained from the location information is a circulant matrix $\mathbf{R}_{n,k}^{circ}$ and can be decomposed as [43]

$$\mathbf{R}_{n,k}^{circ} = \mathbf{U}_{n,k} \mathbf{\Sigma}_{n,k}^{a+b} \mathbf{U}_{n,k}^{H}, \tag{9}$$

where $\mathbf{U}_{n,k} \in \mathbb{C}^{N_t \times b_i}$ is the eigenvector matrix, in which $b_i \leq c_i$, $\mathbf{\Sigma}_{n,k}^{a+b} \in \mathbb{C}^{b_i \times b_i}$ is an eigenvalue matrix and c_i is defined as

$$c_i \triangleq \left(\cos\theta_{n,k}^{\max} - \cos\theta_{n,k}^{\min}\right) \frac{D}{\lambda}, \tag{10}$$

where D is the antenna spacing and λ is the signal wavelength.

Assuming the support of the densities $p(\theta_{n,k})$ of the kth UE do not overlap with the density of other UEs in the same cluster, we have

$$\mathbf{U}_{n,k}^{H} \mathbf{U}_{n,k'} = \begin{cases} \mathbf{I}_{b_i}, & \text{for } k = k' \\ \mathbf{0}, & \text{otherwise} \end{cases} \quad \text{as } N_t \to \infty. \tag{11}$$

Since the covariance matrix $\mathbf{R}_{n,k}^{circ}$ spans $[\theta_{n,k}^{\max}, \theta_{n,k}^{\min}]$, then for any response $\mathbf{a}(\theta_{n,k})$ and $\mathbf{a}(\theta_{n,k'})$ for which $\|\mathbf{U}_{n,k}^{H}\mathbf{a}(\theta_{n,k})\|$ is greater than $\|\mathbf{U}_{n,k}^{H}\mathbf{a}(\theta_{n,k'})\|$ (i.e. $\|\mathbf{U}_{n,k}^{H}\mathbf{a}(\theta_{n,k})\| > \|\mathbf{U}_{n,k}^{H}\mathbf{a}(\theta_{n,k'})\|$) indicates that $\mathbf{a}(\theta_{n,k'})$ causes less interference. Hence, in a finite antenna setting where the span of the spatial angles from the desired UE and other UEs may overlap, $\mathbf{a}(\theta_{n,k'})^{H}\mathbf{R}_{n,k}^{circ}\mathbf{a}(\theta_{n,k'})$ can be used as a measure to predict the interference between user k and user k', where $\mathbf{a}(\theta_{n,k'})^{H}\mathbf{R}_{n,k}^{circ}\mathbf{a}(\theta_{n,k'})$ is given by

$$\mathbf{a}(\theta_{n,k'})^{H}\mathbf{R}_{n,k}^{circ}\mathbf{a}(\theta_{n,k'}) = \mathbf{a}(\theta_{n,k'})^{H}\mathbf{U}_{n,k}\mathbf{\Sigma}_{n,k}^{a+b}\mathbf{U}_{n,k}^{H}\mathbf{a}(\theta_{n,k'}). \tag{12}$$

To reduce the interference between the UEs, the set of interfering UEs for which the predicted interference

$\|\mathbf{a}(\theta_{n,k'})^H \mathbf{R}_{n,k}^{circ} \mathbf{a}(\theta_{n,k'})\|$ is small on the desired channel (i.e., kth UE) is preferred. Hence, we result in selecting the set of UEs for which the interference prediction is small for uplink channel estimation and downlink NOMA transmission. Note that the location-aided interference prediction for all the users can be achieved from the location information without the need for additional radio resources (frequency and time resource).

B. Location-Aided UE Selection

The length of orthogonal pilots is usually limited by the number of UEs and the channel coherence time. To this end, we employ a non-orthogonal pilot sequence within a cluster to reduce the system overhead while employing orthogonal pilots between different clusters to aid channel estimation and reduce inter-cluster interference. Since non-orthogonal pilot sequence is employed within each cluster, the high channel correlation between the UEs may impact the channel estimation. Hence, we focus on exploiting the covariance matrices and the predicted interference in a cluster in selecting the set of UEs with less interference.

To select K_n UEs in a given cluster n, we first determined the desired UE, thereafter, $K_n - 1$ UEs with less interference on the desired UEs are selected. In this paper, we refer to the UE closest to the BS as the desired UE or the primary user (PU). However, this assumption can be relaxed and other UEs may be selected as the PU. The objective here is to exploit the location information in selecting $K_n - 1$ UEs with less interference on the PU, therefore the proposed scheme is still applicable irrespective of the PU. Based on our assumption of the PU, the desired UE can be determined as follows

$$k^* = \underset{k \in \{1,\cdots,K_T\}}{\arg\max}\ \mathbf{a}(\theta_{n,k})^H \mathbf{R}_{n,k}^{circ} \mathbf{a}(\theta_{n,k}), \tag{13}$$

where K_T is the total number of UEs in a cluster from which K_n UEs are selected. Note that the desired UE is determined from the correlation of the steering vectors and the covariance matrices which are obtained from the location information.

Let $\mathcal{K} = \{k' | k' \neq k^*, \text{ and } k' \in \{1, 2, \cdots, K_T\}\}$ be the set of $K_T - 1$ UEs order than the PU which are ordered with ascending order of correlation. The set of $K_n - 1$ interfering UEs from $\mathcal{K}$ that minimize the interference on the PU can then be determined as

$$\mathcal{K}_{n-1} = \left\{ \underset{k' \in \mathcal{K}}{\arg\min} \sum_{k'}^{K_n - 1} \|\mathbf{a}(\theta_{n,k'})^H \mathbf{R}_{n,k}^{circ} \mathbf{a}(\theta_{n,k'})\| \right\}, \tag{14}$$

where $\|\mathbf{a}(\theta_{n,k'})^H \mathbf{R}_{n,k}^{circ} \mathbf{a}(\theta_{n,k'})\|$ corresponds to the predicted interference between user k and user $k^{'}$.

Given the desired UE and the set of $K_n - 1$ UEs with minimum interference, the BS can then schedule the UEs for uplink channel estimation. Note that UE selection complexity is substantially reduced since the BS is not required to estimate the covariance matrix of each UE before UEs selection, instead of the steering vectors $\mathbf{a}(\theta_{n,k'})$ and covariance matrix $\mathbf{R}_{n,k}^{circ}$ can be directly obtained from the location information.

IV. Uplink Channel Estimation and Downlink Achievable Sum-Rate Analysis

The uplink channel estimation of the set of selected UEs and the downlink achievable sum-rate are presented in this section. In addition, we discuss the impact of location information error on the downlink achievable sum-rate.

A. Location-Aided Uplink Channel Estimation

Note that while there may be more UEs within a cluster, only the selected UEs are scheduled for uplink channel estimation and subsequent downlink data transmission.

1) Uplink signal model: By selecting the desired UE and the $K_n - 1$ UEs in each cluster with minimum interference on the desired UE as shown in the previous section, channel estimation can be performed. At the beginning of the coherent time, all the selected UEs from each cluster transmit their pilots sequence to the BS. The received $N_t \times T$ pilot signal at the BS can be expressed as

$$\mathbf{Y}_P = \sqrt{TP} \sum_{n=1}^{N} \sum_{k=1}^{K_n} \mathbf{h}_{n,k} \mathbf{\Phi}_n + \mathbf{N}_P, \tag{15}$$

where T is the pilot sequence length measured in symbol duration, P is the pilot transmit power, $\mathbf{N}_P \sim \mathcal{CN}\left(\mathbf{0}, \sigma_P^2 \mathbf{I}\right)$ is the $N_t \times T$ complex Gaussian noise at the BS, $\mathbf{\Phi}_n \in \mathbb{C}^{1 \times T}$ is the pilot sequence allocated to the UEs in the nth cluster. Based on the received signal $\mathbf{Y}_P$, the BS estimates the channel of the UEs. However, selecting the best set of K_n UEs from the K_T UEs is required to improve the channel estimates, hence, we proposed a location-aided interference prediction scheme to select UEs with minimum interference.

2) Location-aided channel estimation: Given that $\mathbf{\Phi}_n \mathbf{\Phi}_n^H = 1$ and $\mathbf{\Phi}_n \mathbf{\Phi}_{n' \neq n}^H = 0$ then by projecting $\mathbf{Y}_P$ from (15) onto $\mathbf{\Phi}_n$, we obtain $\mathbf{y}_{P,n} = \mathbf{Y}_P \mathbf{\Phi}_n^H$ as

$$\mathbf{y}_{P,n} = \sqrt{TP} \sum_{k=1}^{K_n} \mathbf{h}_{n,k} + \tilde{\mathbf{n}}, \tag{16}$$

where $\tilde{\mathbf{n}} = \mathbf{N}_P \mathbf{\Phi}_n^H$. The MMSE estimate of $\mathbf{h}_{n,k}$ can be derived from (16) as [44]

$$\hat{\mathbf{h}}_{n,k} = \mathbb{E}\left[\mathbf{h}_{n,k} \mathbf{y}_{P,n}^H\right] \mathbb{E}\left[\mathbf{y}_{P,n} \mathbf{y}_{P,n}^H\right]^{-1} \mathbf{y}_{P,n}, \tag{17}$$

where

$$\mathbb{E}\left[\mathbf{h}_{n,k} \mathbf{y}_{P,n}^H\right] = \sqrt{TP} \mathbf{R}_{n,k}^{circ}, \tag{18}$$

$$\mathbb{E}\left[\mathbf{y}_{P,n} \mathbf{y}_{P,n}^H\right] = \left(\mathbf{R}_{n,k}^{circ} + \sum_{k' \neq k}^{K_n} \mathbf{R}_{n,k'}^{circ} + \frac{\sigma_P^2}{TP} \mathbf{I} \right), \tag{19}$$

and $\mathbf{R}_{n,k}^{circ}$ is the covariance matrix of $\mathbf{h}_{n,k}$. Hence, the MMSE estimate of the desired channel $\mathbf{h}_{n,k}$ can be obtained as

$$\hat{\mathbf{h}}_{n,k} = \mathbf{R}_{n,k}^{circ} \left(TP \left(\mathbf{R}_{n,k}^{circ} + \sum_{k' \neq k}^{K_n} \mathbf{R}_{n,k'}^{circ} \right) + \sigma_P^2 \mathbf{I} \right)^{-1} \mathbf{y}_{P,n}. \tag{20}$$

Note that the MMSE channel estimate in (20) is degraded by the interfering channel due to the effect of pilot contamination.

B. Effect of mMIMO Array on the Channel Estimation

We seek to analyze the performance of the location-aided covariance matrices used in the MMSE estimator in the mMIMO regime with large antenna N_t. As previously stated, we assume that the multipath components of the kth UE in (6) are distributed according to the density $p(\theta_{l,n,k})$ with bounded support for some fixed $[\theta_{l,n,k}^{\max},\ \theta_{l,n,k}^{\min}]$. Note that $\theta_{n,k}$ is obtained from the location information, while $\theta_{l,n,k}^{\max}$ and $\theta_{l,n,k}^{\min}$ are assumed to be the maximum and minimum angular spread with mean $\theta_{n,k}$ and can be inferred from the localization techniques used in practice. For instance, the localization error bound of the global positioning system (GPS) could range between 3–20 meters in open areas.

First, we examine the degradation caused by pilot contamination on the channel estimates in the finite antenna regime. From (11), the matrices $\mathbf{R}_{n,k}^{circ}$ and $\sum_{k'\neq k}^{K_n}\mathbf{R}_{n,k'}^{circ}$ span orthogonal subspace in the massive antenna domain. To this end, we restrict ourselves to the asymptotic regime for large N_t, hence, $TP\sum_{k'\neq k}^{K_n}\mathbf{R}_{n,k'}^{circ}$ can be decomposed as

$$TP\sum_{k'\neq k}^{K_n}\mathbf{R}_{n,k'}^{circ}=\mathbf{B}\mathbf{\Sigma}^{a+b}\mathbf{B}^H, \tag{21}$$

where $\mathbf{B}$ is the eigenvector matrix, such that $\mathbf{B}^H\mathbf{B}=\mathbf{I}$. Let $\mathbf{W}$ denote the unitary matrix corresponding to the orthogonal complement of the span of $\mathbf{U}_{n,k}$ and $\mathbf{B}$. Then the $N_t\times N_t$ identity matrix can be decomposed into

$$\mathbf{I}=\mathbf{U}_{n,k}\mathbf{U}_{n,k}^H+\mathbf{B}\mathbf{B}^H+\mathbf{W}\mathbf{W}^H. \tag{22}$$

Hence, as the number of antenna grows large and by substituting (9), (21), and (22) into (20), we have

$$\begin{aligned}\hat{\mathbf{h}}_{n,k}\approx\mathbf{U}_{n,k}\mathbf{\Sigma}_{n,k}^{a+b}\mathbf{U}_{n,k}^H\Big(&TP\mathbf{U}_{n,k}\mathbf{\Sigma}_{n,k}^{a+b}\mathbf{U}_{n,k}^H+\mathbf{B}\mathbf{\Sigma}^{a+b}\mathbf{B}^H\\&+\sigma_P^2\mathbf{U}_{n,k}\mathbf{U}_{n,k}^H+\sigma_P^2\mathbf{B}\mathbf{B}^H+\sigma_P^2\mathbf{W}\mathbf{W}^H\Big)^{-1}\mathbf{y}_{P,n}.\end{aligned} \tag{23}$$

Due to the asymptotic orthogonality between $\mathbf{U}_{n,k}$, $\mathbf{B}$ and $\mathbf{W}$, the channel estimate in (23) is given by

$$\hat{\mathbf{h}}_{n,k}\approx\mathbf{U}_{n,k}\mathbf{\Sigma}_{n,k}^{a+b}\left(TP\mathbf{\Sigma}_{n,k}^{a+b}+\sigma_P^2\mathbf{I}\right)^{-1}\mathbf{U}_{n,k}^H\mathbf{y}_{P,n}. \tag{24}$$

We can simplify (24) as

$$\begin{aligned}\hat{\mathbf{h}}_{n,k}=&\mathbf{U}_{n,k}\mathbf{\Sigma}_{n,k}^{a+b}\left(TP\mathbf{\Sigma}_{n,k}^{a+b}+\sigma_P^2\mathbf{I}\right)^{-1} &(25)\\&\times\left(TP\mathbf{U}_{n,k}^H\mathbf{h}_{n,k}+\tilde{\mathbf{n}}\right). &(26)\end{aligned}$$

The channel estimate in (26) shows the fact that as the number of antennas $N_t\to\infty$, the estimated channel tends to an interference-free channel provided that the span of the covariance matrix of the desired channel does not overlap with the span of the covariance matrices of the interference UEs.

With the assumption of orthogonal pilot sequence between different clusters, the MMSE estimation of the nth PU's channel in the presence of K_n-1 interfering UEs can then be obtained as [43]

$$\hat{\mathbf{h}}_{n,k}=\tilde{\rho}_{n,k}\mathbf{y}_{P,n}, \tag{27}$$

where $\tilde{\rho}_{n,k}=\sqrt{TP}\rho_{n,k}/\left(\sum_{k=1}^{K_n}TP\rho_{n,k}+1\right)$.

Proof: Refer to Appendix B.

By using the MMSE estimation principle of orthogonality [44], the downlink channel can be expressed in terms of the channel estimate and channel estimation error as

$$\mathbf{h}_{n,k}=\hat{\mathbf{h}}_{n,k}+\boldsymbol{\mathcal{E}}_{n,k}, \tag{28}$$

where $\hat{\mathbf{h}}_{n,k}$ is statistically independent from the error $\boldsymbol{\mathcal{E}}_{n,k}$. Finally, the element of $\hat{\mathbf{h}}_{n,k}$ and $\boldsymbol{\mathcal{E}}_{n,k}$ are distributed as $\hat{\mathbf{h}}_{n,k}\sim\mathcal{CN}\left(\mathbf{0},\hat{\rho}_{n,k}\mathbf{I}\right)$ and $\boldsymbol{\mathcal{E}}_{n,k}\sim\mathcal{CN}\left(\mathbf{0},(\rho_{n,k}-\hat{\rho}_{n,k})\mathbf{I}\right)$ respectively, where $\hat{\rho}_{n,k}=TP\rho_{n,k}^2/\left(\sum_{k=1}^{K_n}TP\rho_{n,k}+1\right)$.

C. Downlink Achievable Sum-Rate

After estimating the channel of the selected UE, the BS encodes the data of all the UEs for downlink data transmission where the downlink channel is defined in (28). Note from (28) that channel reciprocity is assumed for the downlink channel. Here, we analyze the downlink achievable sum-rate.

1) Downlink signal model: In the downlink, the BS encodes the signal for each cluster. The received signal at the kth UE in the nth cluster can be expressed as

$$y_{n,k}=\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}s_n+\mathbf{h}_{n,k}^H\sum_{n'=1,n'\neq n}^{N}\mathbf{v}_{n',k}s_{n'}+n_{n,k}, \tag{29}$$

where $s_n=\sum_{k=1}^{K_n}\sqrt{P_{n,k}}x_{n,k}$ is the nth cluster's superimposed signal, $x_{n,k}$ is the message signal of the kth UE in the nth cluster and $n_{n,k}\sim\mathcal{CN}\left(0,\sigma_n^2\right)$ is the complex Gaussian noise at the receiver of kth UE in the nth cluster with zero mean and variance σ_n^2. Note that the first term in (29) consists of the desired signal and intra-cluster interference to the kth UE while the second term denotes the inter-cluster interference. Since the full CSI is difficult to achieve in practice, we have assumed a partial CSI, hence, perfect SIC is not assumed in this paper.

Following the power domain NOMA principle [45], [46], the UEs within a cluster employs the statistical knowledge of the channel to performs SIC which enables the UE to mitigate the effect of intra-cluster interference. Therefore, we assume that the expected channel gains are ordered in descending order as follows

$$\mathbb{E}\left[|\mathbf{h}_{n,1}^H\mathbf{v}_{n,k}|^2\right]\geq\mathbb{E}\left[|\mathbf{h}_{n,2}^H\mathbf{v}_{n,k}|^2\right],\cdots,\mathbb{E}\left[|\mathbf{h}_{n,K_n}^H\mathbf{v}_{n,k}|^2\right]. \tag{30}$$

For a given cluster n, the jth UE is capable of decoding its signal after decoding and successively canceling out the desired signal of the kth UE for $\forall k>j$ where $k,j\in\{1,2,\cdots,K_n\}$, while treating the respective signal for $\forall k<j$ UEs as interference. The decoding order discussed above can be achieved by an optimal scheduling of the K_n UEs in the nth cluster based on their effective channel [15].

By assuming imperfect SIC and imperfect estimated CSI, the received signal can be expressed as

$$y_{n,k} = \underbrace{\sqrt{P_{n,k}}\mathbf{h}_{n,k}^H \mathbf{v}_{n,k} x_{n,k}}_{\text{desired signal}} + \underbrace{\sum_{k'=1}^{k-1} \sqrt{P_{n,k'}}\mathbf{h}_{n,k}^H \mathbf{v}_{n,k} x_{n,k'}}_{\text{residual intra-cluster interference after SIC}} + \underbrace{\sum_{k''=k+1}^{K_n} \sqrt{P_{n,k''}} \left[\mathbf{h}_{n,k}^H \mathbf{v}_{n,k} - \mathbb{E}[\mathbf{h}_{n,k}^H \mathbf{v}_{n,k}]\right] x_{n,k''}}_{\text{residual intra-cluster interference due to imperfect SIC}} + \underbrace{\sum_{n' \neq n}^{N} \sum_{k=1}^{K_n} \sqrt{P_{n',k}}\mathbf{h}_{n,k} \mathbf{v}_{n',k} x_{n',k}}_{\text{inter-cluster interference}} + n_{n,k}\,. \tag{31}$$

Note that users $1, 2, \cdots, K$ with beam $\mathbf{v}_{n,1}, \mathbf{v}_{n,2}, \cdots, \mathbf{v}_{n,Kn}$ share the same beam with index n.

2) Achievable sum-rate analysis: In TDD mMIMO, the uplink channels are estimated at the BS via pilots transmitted by the UEs. To reduce the training overhead and preserve the scalability of the mMIMO antenna array, pilots are not transmitted in the downlink by the BS. Hence, the UEs do not have access to instantaneous downlink channel coefficients and therefore rely on statistical channel knowledge for signal decoding. This in turn can be justified since the mMIMO channel hardens when N_t grows large, and thus, the instantaneous downlink channel coefficients can be approximated by the corresponding statistical counterpart.

From (31), the post-processed received signal can be rewritten when the kth UE in the nth cluster relies upon the statistical CSI knowledge for decoding its desired signal as [47]

$$y_{n,k} = \sqrt{P_{n,k}}\mathbb{E}\left[\mathbf{h}_{n,k}^H \mathbf{v}_{n,k}\right] x_{n,k} + n_{n,k}. \tag{32}$$

The effective signal-to-interference-plus-noise ratio corresponding to kth UE can be defined as

$$\gamma_{n,k} = \frac{P_{n,k}\left|\mathbb{E}\left[\mathbf{h}_{n,k}^H \mathbf{v}_{n,k}\right]\right|^2}{\sum_{m=1}^{4} I_m + \sigma_n^2}, \tag{33}$$

where the values of I_m are given as

$$I_1 = P_{n,k}\,\mathbb{V}\mathrm{ar}\left[\mathbf{h}_{n,k}^H \mathbf{v}_{n,k}\right], \tag{34a}$$

$$I_2 = \sum_{k'=1}^{k-1} P_{n,k'}\mathbb{E}\left[\left|\mathbf{h}_{n,k}^H \mathbf{v}_{n,k}\right|^2\right], \tag{34b}$$

$$I_3 = \sum_{k''=k+1}^{K_n} P_{n,k''}\mathbb{E}\left[\mathbf{h}_{n,k}^H \mathbf{v}_{n,k}\right], \tag{34c}$$

$$I_4 = \sum_{n'=1, n\neq n'}^{N} \sum_{k=1}^{K_{n'}} P_{n',k}\mathbb{E}\left[\left|\hat{\mathbf{h}}_{n,k}^H \mathbf{v}_{n',k}\right|^2\right]. \tag{34d}$$

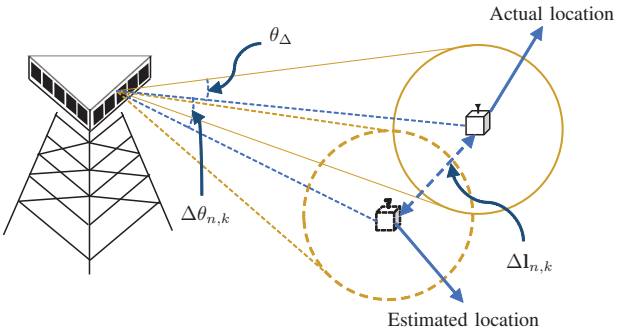

Fig. 3. Scenario showing the actual and estimated location of a single UE, the angular spread θ_Δ, location error $\Delta\mathbf{l}_{n,k}$, and the angular variance due to location error $\Delta\theta_{n,k}$.

The expectation in (33) can be expressed as follows

$$\left|\mathbb{E}\left[\mathbf{h}_{n,k}^H \mathbf{v}_{n,k}\right]\right|^2 = N\hat{\rho}_{n,k} + \rho_{n,k}, \tag{35a}$$

$$I_1 = P_{n,k}\rho_{n,k}, \tag{35b}$$

$$I_2 = \sum_{k'=1}^{k-1} P_{n,k'}\left(N\hat{\rho}_{n,k} + \rho_{n,k}\right), \tag{35c}$$

$$I_3 = \sum_{k''=k+1}^{K_n} P_{n,k''}\sqrt{N\hat{\rho}_{n,k}}, \tag{35d}$$

$$I_4 = \rho_{n,k}. \tag{35e}$$

Proof: The Proof is relegated to Appendix C.

From (33), it can be observed that the NOMA SNR is degraded by the impact of the residual intra-cluster interference due to imperfect SIC, channel estimation error and inter-cluster interference as defined in (31).

The sum-rate of the scheduled UEs, in the nth cluster can be expressed as

$$R_n = \frac{T_c - T}{T_c}\sum_{k=1}^{K_n} \log_2\left(1 + \gamma_{n,k}\right), \tag{36}$$

where T_c is the channel coherence time and the derivation of the interference terms are relegated to Appendix C. The total sum-rate is expressed as

$$R_T = \frac{T_c - T}{T_c}\sum_{n=1}^{N}\sum_{k=1}^{K_n} \log_2\left(1 + \gamma_{n,k}\right). \tag{37}$$

Note that from the location information, the path-loss and angle information are exploited for UE selection to reduce the effect of intra-cluster interference. However, the impact of location uncertainty (location errors) can further degrade the channel estimation process and thereby degrade the downlink achievable sum-rate performance. Hence, we study the impact of the location errors on the system performance.

3) Downlink location-aided power allocation optimization: To maximize the sum-rate in each cluster, we propose a power allocation (PA) strategy that satisfies the minimum quality of service of $K_n - 1$ UEs order than the desired user determined by (14), thereafter the BS allocates the rest of the power to the desired UE determined by (13) in the nth cluster. The problem can be formulated as

$$R_n^* = \max_{P_n, P_{n,k}} R_n \tag{38a}$$

$$\text{s. t. } \sum_{k=1}^{K_n} P_{n,k} = P_n, \tag{38b}$$

$$\sum_{n=1}^{N} P_n = P_T, \tag{38c}$$

$$R_{n,1} \geq \frac{1}{K_n} R_{n,K_n}^c, \tag{38d}$$

$$R_{n,k} = \frac{1}{K_n} R_{n,K_n}^c, \quad k = 2, \cdots, K_n \tag{38e}$$

$$P_{n,k} > 0, \quad \forall k, \tag{38f}$$

where $R_{n,1}$ is the rate of the desired UE (PU), $R_{n,K}^c$ is the rate of the K_nth UE in the nth cluster if it would be supported by conventional beamforming [13] and it is expressed as

$$R_{n,K_n}^c = \log_2 \left(1 + \frac{P_T \left| \mathbf{h}_{n,K_n}^H \mathbf{v}_{n,k} \right|^2}{N\sigma^2} \right). \tag{39}$$

The constraint in (38b) ensures that the total transmit power in the nth cluster is constrained to P_n. Unlike the other $K_n - 1$ UEs, the rate of the PU may be greater than the orthogonal multiple access rate as given by the constraints in (38d) and (38e). The last constraints in (38f) ensure that useful power is allocated to all the scheduled UEs in the nth cluster. Finally, the total sum-rate over all N clusters can be expressed as

$$R_T^* = \sum_{n=1}^{N} R_n^*. \tag{40}$$

The optimal power allocation policy for the kth UE in the nth cluster can be expressed as

$$P_{n,k} = \frac{\lambda b}{(1+\mu) \left| \mathbf{h}_{n,k_n}^H \mathbf{v}_{n,k} \right|^2}, \tag{41}$$

where

$$b = \ln(2) \left(P_n - P_{n,k} \right) \left| \mathbf{h}_{n,k}^H \mathbf{v}_{n,k} \right|^2 + \textstyle\sum_{n' \neq n}^{N} P_{n',k} \left| \mathbf{h}_{n,k}^H \mathbf{v}_{n',k} \right|^2 + \sigma^2. \tag{42}$$

Given the QoS constraints in (38b)–(38f), the allocated power to the PU is given by

$$P_{n,1} = P_n - \sum_{k \neq 1}^{K_n} P_{n,k}. \tag{43}$$

Proof: Refer to Appendix D.

D. Impact of Location Error on the Sum-Rate

In practice, the BS may not have access to the actual location information of the UEs as such information is subject to various localization uncertainty. The accuracy of the location information depends largely on the type of localization technology used and also on the environment (outdoor or indoor). Suppose a location error on the location information of the kth UE and assuming the location error is isotropic with variance $\Delta \mathbf{l}_{n,k}$. The location error will translate into to an error in the spatial domain AOD as shown in Fig. 3 and expressed as [48]

$$\Delta\theta_{n,k} = \frac{\Delta \mathbf{l}_{n,k}}{d_{n,k}\sqrt{2}}, \tag{44}$$

where $\sqrt{2}$ in (44) leads to the correct AOD variance and accounts for the fact that the AOD depends only on $\Delta \mathbf{l}_k$ orthogonal to the LOS direction. The error in AOD will also lead to an error in UEs clustering. Moreover, for $L = 1$ and a small location error, the AOD can be approximated by a first-order Taylor series expansion as

$$\mathbf{a}\left(\theta_{n,k} + \Delta\theta_{n,k}\right) \approx \mathbf{a}\left(\theta_{n,k}\right) + \Delta\theta_{n,k} \mathbf{b}\left(\theta_{n,k}\right), \tag{45}$$

where $\mathbf{b}\left(\theta_{n,k}\right)$ is obtained by taking the derivative of $\mathbf{a}\left(\theta_{n,k}\right)$ with respect to $\theta_{n,k}$. By substituting (45) into (33), we obtain the following

$$\gamma_{n,k} = \frac{P_{n,k} g_{n,k}^2 \left| \mathbb{E}\left[\mathbf{a}\left(\theta_{n,k}\right)^H \mathbf{v}_{n,k} \right] \right|^2}{I_5 + \tilde{g}_{n,k} \sum_{n'=1}^{N} \mathbb{V}\mathrm{ar}\left[\left| \mathbf{b}\left(\theta_{n,k}\right)^H \mathbf{v}_{n',k} \right|^2 \right]}, \tag{46}$$

where $g_{n,k} = \sqrt{\rho_{n,k} N} \alpha_{n,k}$, $\tilde{g}_{n,k} = P_{n,k} g_{n,k}^2 \left(\Delta \mathbf{l}_{n,k} / 2 d_{n,k}^2 \right)$ and

$$I_5 = \sum_{k' \neq k}^{K_n} P_{n,k'} \mathbb{E}\left[\left| \mathbf{h}_{n,k}^H \mathbf{v}_{n,k} \right|^2 \right] + \sum_{n'=1, n \neq n'}^{N} \sum_{k=1}^{K_{n'}} P_{n',k} \mathbb{E}\left[\left| \hat{\mathbf{h}}_{n,k}^H \mathbf{v}_{n',k} \right|^2 \right] + \sigma_n^2, \tag{47}$$

where the expectations in (46) and (47) can be evaluated as

$$\left| \mathbb{E}\left[\mathbf{a}\left(\theta_{n,k}\right)^H \mathbf{v}_{n,k} \right] \right|^2 = N\hat{\rho}_{n,k} + \rho_{n,k} \tag{48a}$$

$$\mathbb{V}\mathrm{ar}\left[\left| \mathbf{b}\left(\theta_{n,k}\right)^H \mathbf{v}_{n',k} \right|^2 \right] = \rho_{n,k} \tag{48b}$$

(48c)

$$I_5 = \sum_{k' \neq k}^{K_n} P_{n,k'} (N\hat{\rho}_{n,k} + \rho_{n,k}) + \sum_{n'=1, n \neq n'}^{N} \sum_{k=1}^{K_{n'}} P_{n',k} \rho_{n,k}. \tag{48d}$$

Proof: Please refer to the Appendix C.

The location error further reduces the SINR by the second term in the denominator of (46). This factor increases with increasing error variance. The performance of the proposed scheme may also be degraded with large errors in the location information. However, we show that the performance of the proposed scheme is robust to small degrees of errors in the location information. Moreover, with increasing advancements in localization techniques, the assumption of small location errors in the proposed scheme can be justified.

Algorithm 1: Proposed location-aided interference prediction and user selection

Input: Location information $\mathbf{l}_{\text{BS}}$, $\mathbf{l}_{n,k}$, AOD distribution $p(\theta_{n,k})$
Output: UE set $\mathcal{K}_n$ cluster beam $\mathbf{v}_{n,k}$
Initialize Evaluate the covariance matrices;
for $Cluster = 1, \cdots, N$ **do**
  Assign UE to cluster from (4);
  Interference prediction and UE selection;
  for *UE* $k := 1, \cdots, K_T$ **do**
    Predict interference by each UE by evaluating (12);
    Evaluate $\mathcal{K}_n = \{k^*, \mathcal{K}_{n-1}\}$ from (13) and (14);
  end
  Evaluated the kth UE uplink channel estimation from (20);
  Evaluated the kth UE power allocation (41);
  Evaluate the downlink achievable sum-rate from (34) and (45);
end

TABLE I
SIMULATION PARAMETERS.

Parameters	Value
Simulation area	200×200 m
BS position	$\mathbf{l}_{\text{BS}} = [0, 100]$
Path-loss exponent	$\delta = 2.75$
Frequency	28 GHz
Pilot length	N
Antenna spacing	$\lambda/2$
Number of paths	$L = 3$
BS total transmit power	P_T
Total deployed UE per cluster	$K_T = 10$
Transmit power to each cluster	$P_i = P_T/N$
Pilot transmit power	$P = P_T/10$
Noise variance	$\sigma_P = 1$, $\sigma_n = 1$
Channel coherence time	$T_c = 196$

Finally, the per cluster and total sum-rate can be evaluated by substituting (46) into (36) and (37), respectively, where the expectation in (46) are evaluated in Appendix C. The proposed location-aided interference prediction and user selection is summarized in Algorithm I.

V. NUMERICAL RESULTS AND DISCUSSION

In this section we discuss the simulation setting and present the simulation results of the proposed scheme. The efficacy of the proposed scheme is evaluated and compared with existing scheme such as the optimal estimation scheme with no inference and also the random UE selection schemes. In addition, the proposed scheme is compared to the UE selection scheme with partial CSI acquisition in [14].

A. Simulation Settings and Performance Metrics

In this section, we present the simulation results to show the efficacy of the proposed scheme and its impact on the NOMA transmission scheme. The basic simulation parameters are presented in Table I. Note that the the simulation parameters in Table I may affect the the simulation result in various ways, however, the trend of the results remains the same since all UEs are subject to similar conditions. The transmit power to each UE ($P_{n,k}$) in a cluster is designed based on the NOMA optimal PA strategy in [15]. The transmit SNR is defined as $SNR = P_T/\sigma_n^2$. In this simulation, $K_T = 10$ UEs are deployed in each cluster from which K_n UEs are selected for NOMA transmission. Throughout the simulations, we set the number of clusters $N = 2$ and UEs per cluster $K_n = 3$, the location error variance $\Delta\theta_{n,k}$ is set to 10 degrees except stated otherwise. We consider two types of AOD distribution namely; (i) uniform and (ii) Gaussian distributions.

1) Uniform distribution: The AODs are uniformly distributed over $[\theta_{n,k} - \theta_\Delta, \theta_{n,k} + \theta_\Delta]$, where $\theta_{n,k}$ is the mean AOD and the θ_Δ is depicted in Fig. 3.

2) Gaussian distribution: In this setting, the AODs of all the paths are i.i.d. Gaussian random variables with mean $\theta_{n,k}$ and standard deviation σ. Such distribution is also unable to provide non-overlapping AOD support. However, the proposed scheme provides some improvement by exploiting the location information.

We consider two performance metrics in this paper. The first metric is the normalized channel estimation error expressed as

$$e_k = 10\log_{10}\left(\frac{\sum_{n=1}^{N}\|\mathbf{h}_{n,k} - \hat{\mathbf{h}}_{n,k}\|_2^2}{\sum_{n=1}^{N}\|\mathbf{h}_{n,k}\|_2^2}\right), \tag{49}$$

where $\mathbf{h}_{n,k}$ and $\hat{\mathbf{h}}_{n,k}$ correspond to the channel of the desired channel and its estimate respectively. Note that we focus only on the error of the desired UE from each cluster. The second performance metric is the achievable sum-rate defined in (37).

In addition, we also compare the achievable sum-rate of the proposed scheme with the orthogonal multiple access (OMA) scheme. The achievable rate of the OMA UE is given by

$$R_{n,k}^{OMA} = \frac{T_c - T}{K_n T_c}\log_2\left(1 + \frac{P_n|\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n,k}|^2}{\mathbb{E}|(\mathbf{h}_{n,k}^H - \hat{\mathbf{h}}_{n,k}^H)\mathbf{v}_{n,k}|^2 + \sigma_n^2}\right), \tag{50}$$

where $1/K_n$ is due to the fact that the OMA scheme results in a multiplexing loss of $1/K_n$ and $\mathbf{h}_{n,k}^H - \hat{\mathbf{h}}_{n,k}^H$ denotes the channel estimation error.

B. Mean Squared Error of Uplink Channel Estimation

1) MSE of channel estimation with uniformly distributed AODs: In Fig. 4, we compare the MSE of the proposed scheme with the random UE selection scheme (i.e., without interference prediction) and the optimal UE selection scheme with optimal training lengths. For the plots with random UE selection, the desired UE is selected as given by (13) while the $K_n - 1$ UEs are randomly selected. From the figure, it is observed that pilot contamination can degrade the performance of the channel estimation if UEs are not properly scheduled within a cluster as shown from the plots with random UE selection. On the other hand, when UEs are selected based on the predicted interference, the proposed scheme achieves close to the performance of the scheme with optimal training length as the number of antennas grows large. Furthermore, with a large angular spread, a non-overlap between the desired and

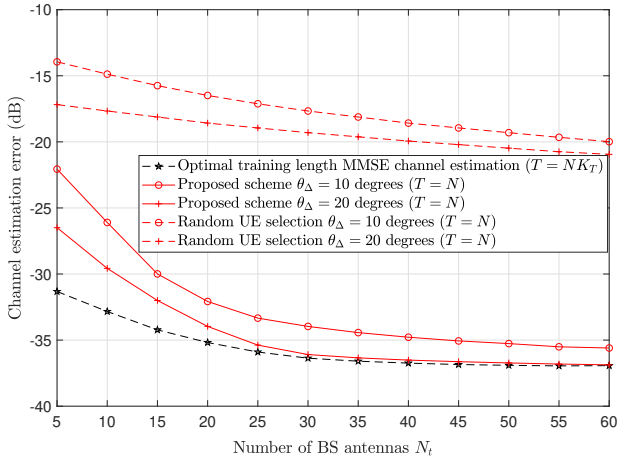

Fig. 4. Comparison of MSE of proposed location-aided interference prediction based UE selection and random UE selection versus number of antennas with uniformly distributed AOD θ_Δ.

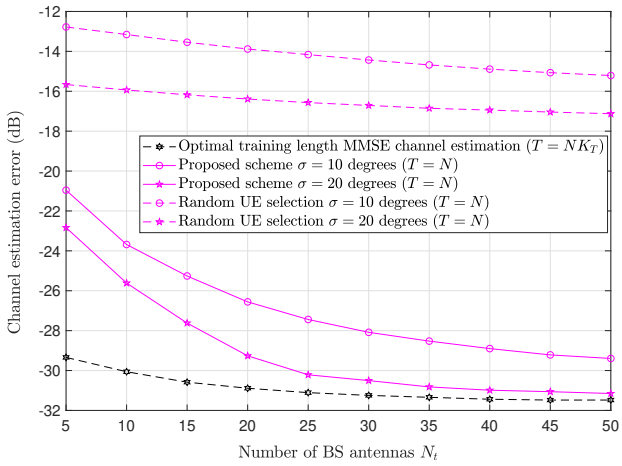

Fig. 5. Comparison of MSE of location-aided interference prediction based UE selection and random UE selection versus number of antennas, Gaussian distributed AOD with variance σ.

interfering UEs multi-path is achieved. In addition, as the number of antenna increases the channel estimation approaches the scheme with optimal training length much faster and the pilot contamination is quickly eliminated.

2) MSE of channel estimation with Gaussian distributed AODs: The MSE of the proposed scheme is compared with the scheme with optimal training length and the random UE selection scheme with Gaussian distributed AODs in Fig. 5. It is observed from the plots that for Gaussian distributed AODs, non-overlapping AODs can not be guaranteed. In addition, by sharing the same pilot sequence, the interference between the UEs increases which leads to a poor channel estimate for the random UE selection schemes. When Fig. 4 is compared with Fig. 5, it is observed from the Gaussian distributed AOD in Fig. 5, that the performance of the estimation error is relatively low due to the non-boundedness of the Gaussian PDF. Nevertheless, when the angular spread of the UEs increases (i.e. $\sigma = 20$ degrees), the proposed scheme achieves close to the scheme with optimal training length channel.

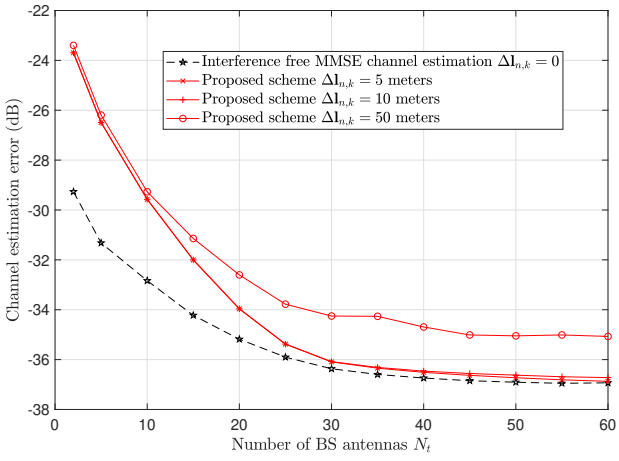

Fig. 6. Comparison of MSE of proposed scheme versus number of antennas with different levels of location errors and uniformly distributed AOD $\theta_\Delta = 10$ degrees.

3) Effect of location error on the channel estimation: In Fig. 6, the impact of location information error on the channel estimation is presented. In this figure, the location-aided covariance matrices is computed based on the location of the UEs. Hence, with an error in the location information, the misalignment of the covariance matrices of the true location and the estimated location affects the channel estimates. It is observed that the proposed scheme is robust to small levels of location uncertainty (up to 10 m) which is within the bounds of a commercial GPS receiver. When the location error is large, a poor channel estimate is achieved due to the fact that the channel used for the estimation may be different from the true channel. Moreover, a much higher number of antennas is required to achieve a performance close to the scheme with optimal training length for large location errors.

C. Downlink Achievable Sum-Rate

In this section, we consider the proposed location-aided schemes with interference prediction and compare its performance with several state-of-the-art schemes.

1) Effect of uniform and Gaussian distributed AODs on the achievable sum-rate: In Fig. 7, the achievable sum-rate is presented with $N = 3$ clusters and $K_n = 3$ UEs per cluster. The performance difference between the proposed scheme, the interference free channel estimation scheme with optimal training period T and the randomly selected UE scheme can be observed from the figure. When K_n UEs are randomly selected, without taking the intra-cluster interference and pilot contamination into account, the system achievable sum-rate degrades especially in the high SNR region which indicates that the degradation is mainly due to interference from other UEs. On the other hand, the proposed scheme does not only improve the channel estimation by reducing the effect of pilot contamination but also improve the achievable sum-rate by selecting UEs with less interference for the NOMA transmission. Finally, the figure reveals that for a distributed Gaussian AOD angular spread, non-overlapping support can

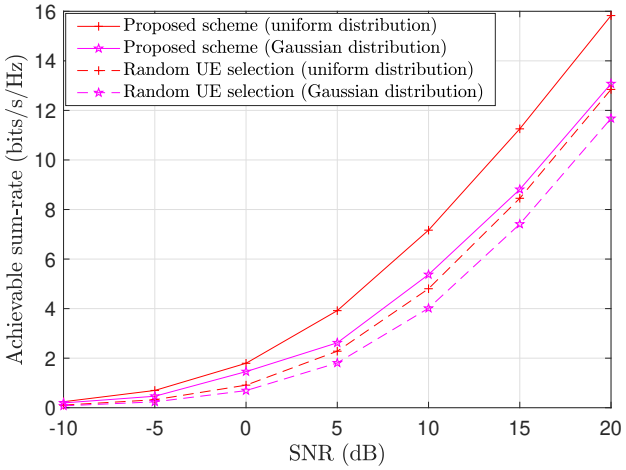

Fig. 7. Comparison of the achievable sum-rate of the proposed scheme with random UE selection versus SNR with $\theta_\Delta = 10$ degrees and $\sigma = 10$ degrees.

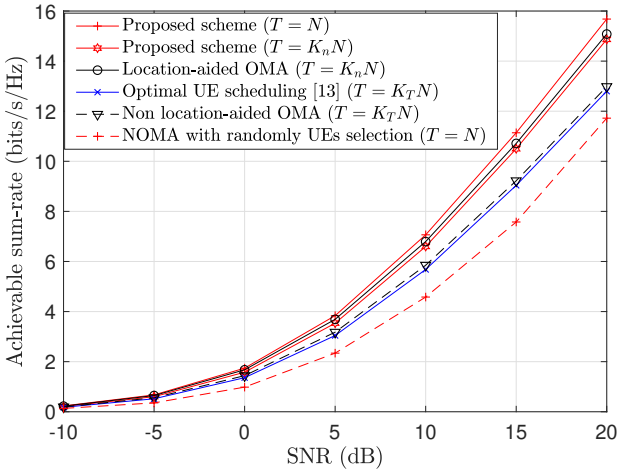

Fig. 8. Achievable sum-rate versus SNR of various schemes with optimal pilot sequence length T with uniformly distributed AODs $\theta_\Delta = 10$ degrees.

not be guaranteed and hence, a performance gap is observed with the uniformly distributed AOD scheme.

2) Comparison of achievable sum-rate of proposed scheme with conventional schemes: Fig. 8 shows the efficacy of the proposed scheme over the exiting NOMA and OMA schemes with respect to the achievable sum-rate. Similar to Fig. 7, the number of cluster $N = 3$ and the number of NOMA users $K_n = 3$. It can be observed that the proposed scheme outperforms the existing NOMA and OMA schemes. We note that the proposed scheme also requires less training sequence length $T = N$. This is due to the fact that the BS is able to select the best set of UEs based on the correlation between the covariance matrix of the UEs obtained from the location information and the steering vectors in the direction of the UE's AOD. The advantages of this selection are in two folds: 1) Firstly, it requires less training overhead since UEs in a cluster can employ the same training sequence while orthogonal pilot sequence can be employed between different clusters, 2) secondly, UEs with less interference can be selected for uplink channel estimation thereby causing less pilot contamination on other UEs.

TABLE II
COMPARISON OF THE COMPLEXITY OF TRAINING OVERHEAD.

Scheme	Overhead
Optimal training	$N \times K_T$
Training for random UE selection	$N \times K_n$
Optimal training with proposed scheme	$N \times K_n$
Proposed training overhead with the proposed scheme	N

TABLE III
COMPLEXITY OF TRAINING OVERHEAD.

Scheme	Complexity
User paring without CSI	$\mathcal{O}(NK_T)^{K_n}$
User paring with partial CSI	$\mathcal{O}(N)^{K_n}$
User paring with perfect CSI	$\mathcal{O}(N)^{K_n}$
Training overhead with the proposed scheme	$\mathcal{O}(N) + K_T$

When the training sequence length is increased from $T = N$ to $T = K_nN$ the channel estimation error improves, however, the detrimental effect of the training overhead can be observed on the proposed scheme's achievable sum-rate. When compared with the optimal UE scheduling scheme in [15], the proposed scheme is observed to achieve better performance due to the reduced training overhead. However, with the training overhead of $T = K_nN$, the OMA scheme achieves better performance compared with the proposed scheme with $T = K_nN$ due to the impact of residual interference. A comparison of the training overhead for the proposed scheme and other schemes are presented in Table II. From the table, it is observed that the proposed scheme and the random UE selection scheme achieves less training overhead, however, the proposed scheme exploits the location information to determine the best set of UEs with less interference to be scheduled for uplink channel estimation. Hence, the performance improvement over the random UE selection and other existing schemes. Note that the improvement of the proposed scheme over the existing schemes is in the fact that the UE paring and training over can be reduced by predicting the interference between UE pairs from the location information. The predicted interference can be used to selected the set of UEs to be paired while existing scheme rely on computationally complex method for channel state information estimation of the UEs before UE pairing.

Furthermore, the location information of the UEs can be exploited to improve the OMA scheme by reducing the training overhead required for the uplink channel estimation. Overall, the non location-aided schemes require more training overhead to achieve better channel estimates which decreases the time for data transmission. Hence, a low achievable sum-rate can be observed. Finally, we show that for a NOMA scheme with $T = N$ training sequence length and randomly selected UEs, the achievable sum-rate is worst due to intra-cluster interference.

3) Impact of location error on the achievable sum-rate: In Fig. 9, the effect of location error on the achievable sum-rate is presented with $N = 3$ and $K_n = 3$. The location error is translated into the spatial domain as given by (44). We note that perfect location information may not always be achieved in practice. From the figure, it is observed that as the deviation

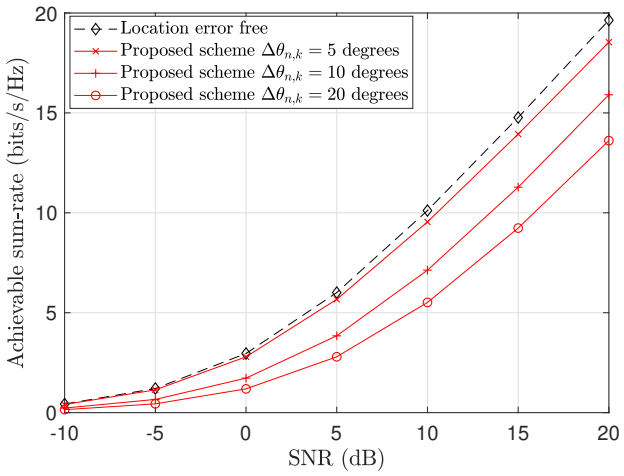

Fig. 9. Achievable sum-rate versus SNR of proposed scheme with varying level of location error with uniformly distributed AODs $\theta_\Delta = 10$ degrees.

of the angle information increases the achievable sum-rates degrades. Note that the channel and covariance matrices used in the estimation of the desired UE's channel are based on the location error. Hence, under high location uncertainty, the channel will be quite different from the true channel which may lead to an overlap on the span of the covariance matrices between UEs. In addition, it may also result in beamforming uncertainty and consequently degrade system performance. However, with a small angular deviation, the channel can still be recovered when the span of the covariance metrics is sufficiently large.

Finally, we present the computational complexity of the proposed scheme in Table III. The MMSE operation, and SIC is similar for each of the schemes after user selection, hence, the complexity analysis for the MMSE and SIC operations are omitted from the analysis. From the table, it is worth noticing that the proposed scheme has a linear computational complexity $(\mathcal{O}(N) + K_T)$ compared to the existing schemes with exponential complexities, the performance is improved due to the interference prediction between the users. This makes the proposed scheme relevant in realistic applications, thanks to the appealing low computational cost.

VI. Conclusion

In this paper, we proposed a user selection scheme based on location-aided interference prediction for mmWave mMIMO NOMA systems. The proposed scheme is exploited for uplink channel estimation and downlink NOMA transmission. In addition, the impact of large mMIMO arrays on the channel estimation and the downlink NOMA achievable sum-rate is also analyzed. We show that although pilot contamination in CSI acquisition can degrade the performance of NOMA system, the challenge can be alleviated by predicting the interference from other UEs on the desired user given that the location information is available at the BS. In addition, it is observed that for a small deviation on the true location, a good channel estimate can be achieved especially in the mMIMO domain. The proposed scheme is found to reduce channel estimation error, intra-cluster interference and consequently improve the achievable sum-rate of NOMA scheme.

Appendix A
Proof of Proposition 1

Proof: From the definition 1,

$$r_{-j} = a_j + b_j \text{ and } j = 0, 1, \cdots, N_t - 1, . \tag{51}$$

and

$$r_{N_t - j} = a_j - b_j, \ j = 1, \cdots, N_t - 1, \tag{52}$$

From (51), we obtain, $a_0 = b_0 = r_0/2$ for $j = 0$. When $j = 1, \cdots, N_t - 1$, a_j and b_j can be obtained from (51) and (52) as

$$a_j = \frac{r_{-j} + r_{N_t - j}}{2}, \quad b_j = \frac{r_{-j} - r_{N_t - j}}{2} \tag{53}$$

Using definition 1 and (53), the channel covariance matrix, $\mathbf{R}_{n,k}^{top}$ is given by

$$\begin{aligned} \mathbf{R}_{n,k}^{top} &= \text{circ}[a_j] + \text{scirc}[b_{i-j}]_{j,i=1} \\ &= \mathbf{A}_{n,k} + \mathbf{B}_{n,k}, \end{aligned} \tag{54}$$

where $\mathbf{A}_{n,k}$ is an $N_t \times N_t$ circulant matrix and $\mathbf{B}_{n,k}$ is an $N_t \times N_t$ skew circulant matrix whose element is $b_{-j} = -b_{N_t - j}$ for $j = 1, 2, \cdots, N_t - 1$ [49]. We observe that in (54), the matrices $\mathbf{A}_{n,k}$ and $\mathbf{B}_{n,k}$ are symmetric or Hermitian, if $\mathbf{R}_{n,k}^{top}$ possesses that property. Hence, we apply eigenvalue decomposition method of $\mathbf{A}_{n,k}$ and $\mathbf{B}_{n,k}$ matrices as

$$\mathbf{A}_{n,k} = \mathbf{U}_{n,k}^{a} \mathbf{\Sigma}_{n,k}^{a} \mathbf{U}_{n,k}^{aH}, \quad \mathbf{B}_{n,k} = \mathbf{U}_{n,k}^{b} \mathbf{\Sigma}_{n,k}^{b} \mathbf{U}_{n,k}^{bH} \tag{55}$$

where $\mathbf{U}_{n,k}^{a}$ and $\mathbf{U}_{n,k}^{b}$ are the $N_t \times N_t$ unitary matrices that satisfy $\mathbf{U}_{n,k}^{a} \mathbf{U}_{n,k}^{aH} = N_t \mathbf{I}$ and $\mathbf{U}_{n,k}^{b} \mathbf{U}_{n,k}^{b} \mathbf{U}_{n,k}^{bH} = N_t \mathbf{I}$, the diagonal matrices are $\mathbf{\Sigma}_{n,k}^{a} = \text{diag}\{\Sigma_{n,k,1}^{a}, \cdots, \Sigma_{n,k,N_t}^{a}\}$ and $\mathbf{\Sigma}_{n,k}^{b} = \text{diag}\{\Sigma_{n,k,1}^{b}, \cdots, \Sigma_{n,k,N_t}^{b}\}$. By substituting (55) in (54) and decomposing, we obtain the following

$$\begin{aligned} \mathbf{R}_{n,k}^{top} &= \mathbf{A}_{n,k} + \mathbf{B}_{n,k} \\ &= \mathbf{U}_{n,k}^{a} \mathbf{\Sigma}_{n,k}^{a} \mathbf{U}_{n,k}^{aH} + \mathbf{U}_{n,k}^{b} \mathbf{\Sigma}_{n,k}^{b} \mathbf{U}_{n,k}^{bH} \\ &= \mathbf{U}_{n,k}^{a} (\mathbf{\Sigma}_{n,k}^{a+b} \odot \mathbf{A}_0) \mathbf{U}_{n,k}^{aH} \\ &\approx \mathbf{R}_{n,k}^{circ}, \end{aligned} \tag{56}$$

where $\mathbf{A} \odot \mathbf{B}$ denotes the Hadamard product of matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{\Sigma}_{n,k}^{a+b} = \mathbf{\Sigma}_{n,k}^{a} + \mathbf{\Sigma}_{n,k}^{b}$ and $\mathbf{A}_0 = \mathbf{U}_{n,k}^{a} \text{diag}\{1, \omega^{N_t - 1}, \cdots, \omega^1\} \mathbf{U}_{n,k}^{aH} = \mathbf{U}_{n,k}^{aH} \mathbf{U}_{n,k}^{b}$, where $\omega = 2j\pi/N_t$. ■

Appendix B
Property of $\mathbf{R}_{n,k}^{circ}$

Another important property of $\mathbf{R}_{n,k}^{circ}$ is the fact that as the number of antennas $N_t \to \infty$

$$\mathbf{R}_{n,k}^{circ} \approx \mathbb{E}\left[\mathbf{h}_{n,k} \mathbf{h}_{n,k}^{H}\right]. \tag{57}$$

From the laws of large number, it follows that when N_t is large, and the channel coefficients are i.i.d., then the channel

vectors between the users and the BS become pairwisely orthogonal [43] such that

$$\frac{1}{N_t}\mathbf{h}_{n,k}^H\mathbf{h}_{n,k} \to \rho_{n,k}, \text{ as } N_t \to \infty. \tag{58}$$

Multiplying $\mathbf{R}_{n,k}^{circ}$ by $\mathbf{y}_{P,n}$, we obtain

$$\mathbf{R}_{n,k}^{circ}\mathbf{y}_{P,n} \approx \mathbf{h}_{n,k}(\sqrt{TP}\rho_{n,k}+1), \tag{59}$$

Therefore, (59) can be considered as a characteristic equation for the covariance matrix $\mathbf{R}_{n,k}^{circ}$. As a consequence $\mathbf{h}_{n,k}$ is the eigenvector corresponding to the $\sqrt{TP}\rho_{n,k}+1$ eigenvalue of $\mathbf{R}_{n,k}^{circ}$. By exploiting this property in (26), we obtain the channel estimates in (27).

APPENDIX C
DERIVATION OF THE INTERFERENCE TERMS IN EVALUATING THE ACHIEVABLE SUM-RATE

Here, we evaluate the expectation and variance in (33) and (34a)–(34d). Starting from (28) and exploiting the following properties of Gaussian random variables [50]

$$\mathbb{E}\left[\mathbf{z}_n^H\mathbf{z}_{n'}\right] = \begin{cases} N, & n = n' \\ 0, & \text{otherwise}, \end{cases} \tag{60a}$$

$$\mathbb{E}\left[|\mathbf{z}_n^H\mathbf{z}_{n'}|^2\right] = N^2+N, \tag{60b}$$

$$\mathbb{E}\left[\boldsymbol{\mathcal{E}}_{n,k}^T\mathbf{v}_{n,k}\right] = 0 \quad \text{for } k \in \{1,\cdots,K_n\}. \tag{60c}$$

The expectation of $\mathbb{E}\left[|\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n,k}|^2\right]$ can then be derived as

$$\mathbb{E}\left[|\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n,k}|^2\right] = \frac{\hat{\rho}_{n,k}}{N}\mathbb{E}\left[|\mathbf{z}_k^H\mathbf{z}_k|^2\right] = (N+1)\hat{\rho}_{n,k}. \tag{61}$$

Next, we derive the expectation of $\mathbb{E}\left[\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}\right]$ by substituting (28) as follows

$$\begin{aligned}\mathbb{E}\left[\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}\right] &= \mathbb{E}\left[\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n,k} + \boldsymbol{\mathcal{E}_{n,k}}^H\mathbf{v}_{n,k}\right] \\ &= \left(\sqrt{\hat{\rho}_{n,k}/N}\mathbb{E}\left[|\mathbf{z}_n^H\mathbf{z}_n|^2\right] + \mathbb{E}\left[\boldsymbol{\mathcal{E}_{n,k}}^H\mathbf{z}_n\right]\right) \\ &= \sqrt{N\hat{\rho}_{n,k}}. \end{aligned} \tag{62}$$

By exploiting the property defined in (60a), we evaluate $\mathbb{E}\left[|\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}|^2\right]$ as follows

$$\begin{aligned}&\mathbb{E}\left[|\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}|^2\right] \\ &= \mathbb{E}\left[|\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n,k} + \boldsymbol{\mathcal{E}_{n,k}}^H\mathbf{v}_{n,k}|^2\right] \\ &= \mathbb{E}\left[\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n,k}\mathbf{v}_{n,k}^H\hat{\mathbf{h}}_{n,k}\right] + \mathbb{E}\left[\boldsymbol{\mathcal{E}_{n,k}}^H\mathbf{v}_{n,k}\mathbf{v}_{n,k}^H\boldsymbol{\mathcal{E}_{n,k}}\right] \\ &= \mathbb{E}\left[|\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n,k}|^2\right] + \mathbb{E}\left[\boldsymbol{\mathcal{E}_{n,k}}^H\boldsymbol{\mathcal{E}_{n,k}}\right]\mathbb{E}\left[\mathbf{z}_n^H\mathbf{z}_n\right]/N^2 \\ &= (N+1)\hat{\rho}_{n,k} + (\rho_{n,k}-\hat{\rho}_{n,k}) \\ &= N\hat{\rho}_{n,k} + \rho_{n,k}. \end{aligned} \tag{63}$$

The variance $\mathbb{V}\text{ar}\left[\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}\right]$ can be evaluated as

$$\begin{aligned}\mathbb{V}\text{ar}\left[\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}\right] &= \mathbb{E}\left[|\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n,k}|^2\right] - \left|\mathbb{E}\left[\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n,k}\right]\right|^2 \\ &= N\rho\tilde{_{n,k}} + \rho_{n,k} - N\rho\tilde{_{n,k}} = \rho_{n,k}. \end{aligned} \tag{64}$$

Finally, $\mathbb{E}\left[|\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n',k}|^2\right]$ for $n' \neq n$ can be derived as

$$\begin{aligned}\mathbb{E}\left[|\hat{\mathbf{h}}_{n,k}^H\mathbf{v}_{n',k}|^2\right] &= \mathbb{E}\left[\hat{\mathbf{h}}_{n,k}^H\mathbf{z}_{n'}\mathbf{z}_{n'}^H\mathbf{h}_{n,k}\right]/N \\ &= \mathbb{E}\left[\hat{\mathbf{h}}_{n,k}^H\mathbf{h}_{n,k}\right]\mathbb{E}\left[\mathbf{z}_{n'}\mathbf{z}_{n'}^H\right]/N^2 = \rho_{n,k}. \end{aligned} \tag{65}$$

By substituting the results from (62), (63), (64), and (65) in (34a) to (34d), the achievable sum-rate can be evaluated.

APPENDIX D
POWER ALLOCATION

To determine the optimal PA to each UE in a cluster, we rewrite (29) in terms of the instantaneous SINR as

$$\gamma_{n,k} = \frac{P_{n,k}\left|\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}\right|^2}{(P_n - P_{n,k})\left|\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}\right|^2 + \sum_{n'\neq n}^{N} P_{n',k}\left|\mathbf{h}_{n,k}^H\mathbf{v}_{n',k}\right|^2 + \sigma^2}, \tag{66}$$

where P_n is the power allocated to the nth cluster. Note that the power allocation is carried out at the BS to enable the UEs perform SIC after the received signal. Our goal therefore is to optimize the power allocated to the kth UE in the nth cluster such that the system sum-rate is maximized.

To this end, we form a Lagrangian from (38a)–(38f) as follows

$$\begin{aligned}\mathcal{L}(P_T,\lambda,\mu) = &\sum_{k=1}^{K_n}\log_2(1+\gamma_{n,k}) + \lambda\left(\sum_{k=1}^{K_n}P_{n,k} - P_n\right) \\ &+ \mu\left(\log_2(1+\gamma_{n,k}) - \log_2(1+\bar{\gamma})\right), \end{aligned} \tag{67}$$

where λ and μ are the Lagrange multipliers and

$$\bar{\gamma} = \frac{P_T\left|\mathbf{h}_{n,K_n}^H\mathbf{v}_{n,k}\right|^2}{N\sigma^2}. \tag{68}$$

The constraints are Karush-Kuhn-Tucker (KKT) conditions for optimizing the power allocation. For a fixed Lagrange multiplier, the problem is a standard optimization problem with the KKT conditions.

$$\frac{\partial}{\partial P_{n,k}}\mathcal{L}(P_T,\lambda,\mu) = 0 \tag{69}$$

The optimal power allocation policy for the kth UE in the nth cluster can be expressed as

$$P_{n,k} = \frac{\lambda b}{(1+\mu)\left|\mathbf{h}_{n,k_n}^H\mathbf{v}_{n,k}\right|^2}, \tag{70}$$

where

$$\begin{aligned} b = &\ln(2)(P_n - P_{n,k})\left|\mathbf{h}_{n,k}^H\mathbf{v}_{n,k}\right|^2 \\ &+ \sum_{n'\neq n}^{N}P_{n',k}\left|\mathbf{h}_{n,k}^H\mathbf{v}_{n',k}\right|^2 + \sigma^2. \end{aligned} \tag{71}$$

while the dual variable can be updated with gradient descent method as

$$\lambda(l+1) = \left[\lambda(l) - \epsilon_1(l)\left(P_T - \sum_{n}^{N}\sum_{k}^{K_n} P_{n,k}\right)\right]^+$$
$$\mu(l+1) = \left[\mu(l) - \epsilon_2(l)\left(R_{n,k} - \frac{1}{K_n}R^c_{n,K_n}\right)\right]^+ \quad (72)$$

where l is the iteration index. $\epsilon_1(l)$ and $\epsilon_2(l)$ are positive step sizes at iteration l. Based on an appropriate step size, the iteration converges to an optimal solution to problem (38a). Hence, with the QoS constraints given in (38b)–(38f), the power required to guarantee the downlink throughput can be iteratively obtained from (72), while the allocated power to the PU is given by

$$P_{n,1} = P_n - \sum_{k\neq 1}^{K_n} P_{n,k}. \quad (73)$$

REFERENCES

[1] I. Orikumhi, H. Jwa, J. Na, and S. Kim, "Location-aided user clustering and power allocation for NOMA in 5G mmWave networks," in *Proc. IEEE ICTC*, 2020.

[2] V. W. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key Technologies for 5G Wireless Systems*. Cambridge Univ. Press, 2017.

[3] S. Deng, M. K. Samimi, and T. S. Rappaport, "28 GHz and 73 GHz millimeter-wave indoor propagation measurements and path loss models," in *Proc. IEEE ICCW*, 2015.

[4] Z. Chen, F. Sohrabi, and W. Yu, "Multi-cell sparse activity detection for massive random access: Massive MIMO versus cooperative MIMO," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4060–4074, 2019.

[5] J. Li *et al.*, "Deep learning-based massive MIMO CSI feedback," in *Proc. IEEE ICOCN*, 2019.

[6] Y. Mehmood, W. Afzal, F. Ahmad, U. Younas, I. Rashid, and I. Mehmood, "Large scaled multi-user MIMO system so called massive MIMO systems for future wireless communication networks," in *Proc. ICAC*, 2013, pp. 1–4.

[7] K. Nishimori, "Novel technologies using massive MIMO transmission toward 5G and its beyond systems," in *Proc. IEEE ISAP*, 2018.

[8] H. Nguyen, "Low complexity max-min uplink power control in massive MIMO system with experiment on testbed," in *Proc. IEEE ICT*, 2019.

[9] D. Zhang *et al.*, "Capacity analysis of NOMA with mmWave massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1606–1618, Jul. 2017.

[10] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[11] A. Misra, M. P. Sarma, K. K. Sarma, and N. Mastorakis, "Temporal deep learning assisted UAV communication channel model for application in EH-MIMO-NOMA set-up," *J. of Commun. Netw.*, vol. 24, no. 2, pp. 166-183, 2022.

[12] P. Xu, K. Cumanan, and Z. Yang, "Optimal power allocation scheme for NOMA with adaptive rates and alpha-fairness," in *Proc. IEEE GLOBECOM*, 2017.

[13] B. Kim *et al.*, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE MILCOM*, 2013.

[14] J. Guo, X. Wang, J. Yang, J. Zheng, and B. Zhao, "User pairing and power allocation for downlink non-orthogonal multiple access," in *Proc. IEEE GLOBECOM*, 2016.

[15] J. Cui, Y. Liu, Z. Ding, P. Fan, and A. Nallanathan, "Optimal user scheduling and power allocation for millimeter wave NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1502–1517, Mar. 2018.

[16] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X. G. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2961–2974, May. 2018.

[17] S. Sobhi-Givi, M. G. Shayesteh, and H. Kalbkhani, "Energy-efficient power allocation and user selection for mmWave-NOMA transmission in M2M communications underlaying cellular heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9866–9881, 2020.

[18] S. A. R. Naqvi and S. A. Hassan, "Combining NOMA and mmWave technology for cellular communication," in *Proc. IEEE VTC-Fall*, 2016.

[19] S. Moon, H. Kim, and I. Hwang, "Deep learning-based channel estimation and tracking for millimeter-wave vehicular communications," *J. Commun. Netw.*, vol. 22, no. 3, pp. 177–184, 2020.

[20] P. J. Okoth, Q. N. Nguyen, D. R. Dhakal, D. Nozaki, Y. Yamada, and T. Sato, "An efficient codebook-based beam training technique for millimeter-wave communication systems," in *Proc. APMC*, 2018.

[21] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[22] Y. Zhao, Z. Lin, K. Liu, C. Zhang, and L. Huang, "Pilot contamination reduction in massive MIMO system," in *Proc. IEEE ICCE-TW*, 2018.

[23] M. Boulouird, A. Riadi, and M. M. Hassani, "Pilot contamination in multi-cell massive-MIMO systems in 5G wireless communications," in *Proc. ICEIT*, 2017.

[24] M. Boudaya, I. Kammoun, and M. Siala, "Efficient pilot design based on the divide and conquer approach for pilot contamination mitigation in massive MIMO," in *Proc. IWCMC*, 2019.

[25] H. Gao, T. Zhang, C. Feng, and Y. Wang, "Clustering based pilot allocation algorithm for mitigating pilot contamination in massive MIMO systems," in *Proc. APMC*, 2018.

[26] Y. You and L. Zhang, "Bayesian matching pursuit-based channel estimation for millimeter wave communication," *IEEE Wireless Commun. Lett.*, vol. 24, no. 2, pp. 344–348, 2020.

[27] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, 2016.

[28] D. Nguyen, A. Nguyen, H. Han, V. Nguyen, and M. Zia, "Pilot decontamination using time-shifted pilot and data-aided channel estimation for massive MIMO," in *Proc. IEEE ICCE*, 2018.

[29] Z. Gong, C. Li, and F. Jiang, "Pilot decontamination for cell-edge users in multi-cell massive MIMO based on spatial filter," in *Proc. IEEE ICC*, 2018.

[30] L. Zhang, P. Zhu, and J. Li, "Pilot decontamination based on pilot allocation for large-scale distributed antenna systems," in *Proc. WCSP*, 2018.

[31] M. H. Mazlan, E. Ali, A. Mohd Ramly, R. Nordin, M. Ismail, and A. Sali, "Pilot decontamination using coordinated wiener predictor in massive-MIMO system," *IEEE Access*, vol. 6, pp. 73 180–73 190, 2018.

[32] J. Sheu, W. Sheen, C. Wu, and H. Chang, "Pilot decontamination techniques based on beam-domain channel characteristics in millimetre-wave sparse channels," *IET Commun.*, vol. 12, no. 19, pp. 2493–2501, 2018.

[33] Mingmei Li, Shi Jin, and Xiqi Gao, "Spatial orthogonality-based pilot reuse for multi-cell massive MIMO transmission," in *Proc. IEEE ICWCSP*, 2013, pp. 1–6.

[34] L. S. Muppirisetty, T. Charalambous, J. Karout, G. Fodor, and H. Wymeersch, "Location-aided pilot contamination avoidance for massive MIMO systems," *IEEE Trans. Commun.*, vol. 17, no. 4, pp.2662-2674, April 2018.

[35] L. S. Muppirisetty, H. Wymeersch, J. Karout, and G. Fodor, "Location-aided pilot contamination elimination for massive MIMO systems," in *Proc. IEEE GLOBECOM*, 2015.

[36] R. D. Taranto *et al.*, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 102–112, Nov. 2014.

[37] N. Garcia, H. Wymeersch, E. G. Ström, and D. Slock, "Location-aided mm-wave channel estimation for vehicular communication," in *Proc. IEEE SPAWC*, 2016, pp. 1–5.

[38] J. Wang, Y. Li, C. Ji, Q. Sun, S. Jin, and T. Q. S. Quek, "Location-based mimo-noma: Multiple access regions and low-complexity user pairing," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2293–2307, 2020.

[39] J. Nam, A. Adhikary, J. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, 2014.

[40] S. Kutty and D. Sen, "Impact of intra-cluster angular spread on the performance of nlos millimeter wave links with imperfect beam alignment," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1813–1827, Feb. 2020.

[41] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1822–1835, Mar. 2018.
[42] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations Trends Commun. and Inf. Theory*, vol. 2, no. 3, pp. 155–239, 2006.
[43] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE ICASSP*, 2012.
[44] S. M. Kay, *Fundamentals of statistical signal processing*. Englewood Cliffs, NJ: Prentice Hall, 1993.
[45] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "Capacity comparison between MIMO-NOMA and MIMO-OMA with multiple users in a cluster," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2413–2424, Oct. 2017.
[46] Z. Zhang *et al.*, "Non-orthogonal multiple access for cooperative multicast millimeter wave wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1794–1808, Aug. 2017.
[47] T. L. Marzetta, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.
[48] D. Slock, "Location aided wireless communications," in *Proc. IEEE ISCCSP*, 2012, pp. 1–6.
[49] Z. Liu, S. Chen, W. Xu, and Y. Zhang, "The eigen-structures of real (skew) circulant matrices with some applications," *Computational and Appl. Math.*, vol. 38, no. 4, p. 178, 2019.
[50] A. M. Tulino *et al.*, "Random matrix theory and wireless communications," *Foundations Trends Commun. and Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.

Igbafe Orikumhi (S'14–M'18) is currently an Assistant Professor with the Department of Electronic Engineering and a Research Fellow with the Wireless Systems Laboratory, Hanyang University, Seoul, South Korea. He received his B.Eng. degree in Electrical and Computer Engineering from the Federal University of Technology Minna, Nigeria, in 2008. He received his M.Sc. degree in Electrical Electronics and Telecommunications Engineering and his Ph.D. degree in Electrical Engineering from Universiti Teknologi Malaysia, Skudai, Malaysia, in 2014 and 2017 respectively.

From 2018 to 2020, he was a Research Fellow at the 5G-Unmanned Aerial Vehicle Laboratory, Hanyang University, Seoul, South Korea. His research interests include relay transmission, cooperative communications, vehicular communication, millimeter wave communication, Beam management, nonorthogonal multiple access, 5G and 6G communications.

Chee Yen (Bruce) Leow (S'08-M'12-SM'21) is currently an Associate Professor with the School of Electrical Engineering, Faculty of Engineering and a Research Fellow with the Wireless Communication Centre, Universiti Teknologi Malaysia (UTM). He obtained a PhD degree in Wireless Communications from Imperial College London in September 2011 and a B.Eng. degree in Computer Engineering from UTM in June 2007.

Dr. Leow's current research interest includes non-orthogonal multiple access, drone communication, intelligent surfaces, advanced MIMO, millimeter wave communication and prototype development using software defined radio, for beyond 5G and Internet of Things applications. His IEEE journal papers won the IEEE Malaysia Comsoc/VTS Joint Chapter's Best Paper Awards 2016, 2017 and 2021, and IEEE Malaysia AP/MTT/EMC Joint Chapter's Best Paper award 2017, 2018, and 2020. He is among the pioneers for 5G initiatives in Malaysia to promote 5G R&D collaboration between industry and academia. He is currently the Secretary for IMT and Future Networks Working Group under the Malaysian Technical Standards Forum Berhad to accelerate the adoption of 5G IMT-2020 in Malaysia. In addition, he regularly conducts short courses on 4G and 5G for the telecommunication industry. Dr. Leow is a registered Chartered Engineer (CEng) of the Engineering Council UK.

Sunwoo Kim (S'99-M'05-SM'17) received his B.S degree from Hanyang University, Seoul, Korea in 1999, and his Ph.D. degree, in 2005, from the Department of Electrical and Computer Engineering, University of California, Santa Barbara. Since 2005, he has been working in the Department of Electronic Engineering at Hanyang University, Seoul, Korea, where he is currently a professor. He is also the director of the 5G/Unmanned Vehicle Research Center, funded by the Ministry of Science and ICT of Korea. He was a visiting scholar to the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology from 2018 to 2019. He is an Associate Editor of IEEE Transactions on Vehicular Technology. He is a senior member of the IEEE. His research interests include wireless communication/positioning/localization, statistical signal processing.