Downlink User Capacity of Massive MIMO Under Pilot Contamination

Juei-Chin Shen, *Member, IEEE*, Jun Zhang, *Member, IEEE*, and Khaled B. Letaief, *Fellow, IEEE*

***Abstract*—Pilot contamination has been regarded as a main limiting factor of time division duplexing (TDD) massive multiple-input-multiple-output (Massive MIMO) systems, as it will make the signal-to-interference-plus-noise ratio (SINR) saturated. However, how pilot contamination will limit the user capacity of downlink Massive MIMO, i.e., the maximum number of users whose SINR targets can be achieved, has not been addressed. This paper provides an explicit expression of the Massive MIMO user capacity in the pilot-contaminated regime where the number of users is larger than the pilot sequence length. This capacity expression characterizes a region within which a set of SINR requirements can be jointly satisfied. The size of this region is fundamentally limited by the pilot sequence length. Furthermore, the scheme for achieving the user capacity, i.e., the uplink pilot training sequences and downlink power allocation, has been identified. Specifically, the generalized Welch bound equality sequences are exploited and it is shown that the power allocated to each user should be proportional to its SINR target. With this capacity-achieving scheme, the SINR requirement of each user can be satisfied and energy-efficient transmission is achieved in the large-antenna-size (LAS) regime. The comparison with two non-capacity-achieving schemes highlights the superiority of our proposed scheme in terms of achieving higher user capacity. Furthermore, for the practical scenario with a finite number of antennas, the actual antenna size required to achieve a significant percentage of the asymptotic performance has been analytically quantified.**

***Index Terms*—Massive multiple-input-multiple-output (Massive MIMO), user capacity, pilot contamination, pilot-aided channel estimation, power allocation.**

I. Introduction

MASSIVE MIMO is regarded as an efficient and scalable approach for multicell multiuser MIMO implementation, and it can significantly improve both the spectrum efficiency and energy efficiency [1]–[3]. By equipping each base station (BS) with an antenna array whose size is greatly larger than the number of active user equipments (UEs), the asymptotic orthogonality among MIMO channels to different UEs is achieved and in turn it makes intra- and inter-cell interference more manageable. Hence, using simple linear precoder and detector can approach the optimal dirty-paper coding capacity [4]. Channel side information (CSI) at BSs plays an important role in the exploitation of channel orthogonality. In practice, TDD operation is assumed for CSI acquisition through uplink training. The advantage of uplink training is that the required length of pilot training sequences is proportional to the number of active UEs rather than that of BS antennas. The training length is fundamentally limited by the channel coherence time, which can be short due to UEs of high mobility. It has been shown, nevertheless, in [5] that the effect of using short training sequences diminishes in the LAS regime. Specifically, among the poorly estimated channels, the asymptotic orthogonality can still hold.

However, a major problem with TDD Massive MIMO is that inevitably the same pilot sequences will be reused in multiple cells. The channels to UEs in different cells who share the same pilot sequence will be collectively learned by BSs. In other words, the desired channel learned by a BS is contaminated by undesired channels. Once this contaminated CSI is utilized for transmitting or receiving signals, intercell interference occurs immediately which limits the achievable SINR. This phenomenon, known as *pilot contamination*, can not be circumvented simply by increasing the BS antenna size [6], [7].

Several attempts have been made to tackle the problem of pilot contamination. In [8], a sophisticated precoding method is proposed to minimize intercell interference due to pilot contamination. A more direct approach for pilot decontamination, harnessing second-order channel statistics, can be found in [9]. A recent study [10] claims that pilot contamination is due to inappropriate linear channel estimation. Hence, by using the blind subspace-based estimation, unpolluted CSI is obtainable. In the context of Massive-MIMO-aided OFDM, pilot contamination can be eliminated by a novel combination of downlink training and scheduled uplink training [11]. However, this method, suffering from similar drawbacks as the blind estimation, requires a substantial training period.

The effect of pilot contamination is usually quantified as SINR saturation due to intercell interference. A number of studies have examined this saturation phenomenon and the corresponding uplink or downlink throughput [6], [12], [13]. The former two provide analysis in the LAS regime with a fixed number of active UEs, while in [13], it analyzes asymptotic SINRs with a fixed ratio of the BS antenna size to the active-UE number. All these studies lead to a similar conclusion that the SINR will saturate with an increasing antenna size, making system throughput limited. In [14], several possible definitions of SINR are provided and discussed.

So far, however, there has been little discussion about the user capacity of TDD Massive MIMO, which is the maximum

Manuscript received June 27, 2014; revised November 12, 2014; accepted January 13, 2015. Date of publication February 13, 2015; date of current version June 6, 2015. This work is supported by the Hong Kong Research Grant Council under Grant No. 610212. The associate editor coordinating the review of this paper and approving it for publication was E. A. Jorswieck.

The authors are with the Department of Electronic and Computer Engineering, the Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: eejcshen@ust.hk; eejzhang@ust.hk; eekhaled@ust.hk).

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/TWC.2015.2403317

number of UEs whose SINR requirements can be met for a given pilot sequence length. This terminology "user capacity," originally coined for analyzing CDMA systems [15], is employed here for investigating Massive MIMO from a new angle. In this paper, we confine our discussion to the user capacity of single-cell downlink TDD Massive MIMO, where pilot contamination will occur once the number of UEs is greater than the pilot sequence length. Meanwhile, we consider a more general set of pilot sequences whose cross-correlations can range from -1 to 1. In contrast, in most existing studies of Massive MIMO, the cross-correlations are restricted to be 1 or 0. Our discussion will show that the user capacity can be characterized by a specific region within which there exists a capacity-achieving pilot sequence and power allocation scheme such that the SINR requirements are satisfied and the user data can be energy-efficiently transmitted. This result is significant in the sense that within the specified region we have no worries about pilot contamination by using the proposed allocation scheme. Meanwhile, the derivation of this result only involves a simple least-squares (LS) channel estimator and maximum ratio transmission (MRT) precoding. It means that this region exists without relying on advanced estimation or precoding methods. Though this region is identified under the single-cell assumption, it sheds light on the possible existence of a similar region in the general multicell scenario.

A. Prior Works

The idea of using user capacity as a metric to characterize the system performance has been explored in [15], [16]. In [15], the discussion of uplink/downlink user capacity is confined to CDMA systems, in which the role of the signature sequences is analogous to that of pilot sequences in TDD Massive MIMO. When the signature/pilot sequence length is less than the number of served UEs, it is described as overloading in CDMA while as pilot contamination in Massive MIMO [17]–[19]. However, more emphasis is placed on the single-cell scenario for over-loaded CDMA, while the multicell scenario is stressed for pilot-contaminated Massive MIMO. The recent study of user capacity in multicell CDMA has shifted its focus to the effects of hard-handoffs instead of over-loaded channels [20]. On the other hand, the previous investigation of downlink user capacity of multiuser MIMO in [16] is conducted without considering the effects of pilot contamination. In addition, the capacity depicted in the same work assumes an infinite number of UEs. With perfect CSI at the BS, uplink/downlink user capacity of multiuser MIMO with identical SINR constraints is presented in [21]. Basically, the maximum number of admissible UEs increases with the BS antenna size in the downlink, while it is governed by the pilot sequence length in the uplink. Moreover, the optimal number of UEs for maximizing the ergodic sum throughput can be identified. Briefly, the foregoing studies of the user capacity of multiuser MIMO have not examined the effects of pilot contamination with general SINR constraints. The main purpose of this study is to develop an understanding of the mentioned effects on downlink user capacity of Massive MIMO.

B. Contributions

The major contributions of this paper are summarized as follows:

1) Previous research on Massive MIMO has examined the effects of pilot contamination on the received SINRs of individual users. In contrast, this paper makes the first attempt to study the impact on a group of UEs and to articulate conditions under which their SINR targets can be met simultaneously in the LAS regime. We show that if the SINR requirements are located in a prescribed admissible region, using our proposed joint pilot sequence and power allocation can achieve these targets concurrently as the BS antenna size grows without limit, i.e., in the LAS regime. We find that the size of the prescribed region, characterizing the user capacity, is fundamentally limited by the length of uplink pilot sequences. If cross-correlations of pilot sequences are not well designed, the size of this region can be further reduced.
2) We advance this research by evaluating the performance in the non-LAS regime, which provides a useful guide to practical BS antenna deployment. Specifically, we characterize the average SINR that is actually achieved for a given number of BS antennas. Based on it, the study of how many BS antennas are needed to approach a significant percentage of SINR performance limits is presented. Generally, the number of the required antennas depends on the cross-correlations of pilot sequences and the SINR requirements. When there exists a UE with a relatively high SINR requirement, we provide a fresh insight into how this UE will cause the demand for a larger antenna size. Simulation results will verify the accuracy of this analytical investigation, i.e., the prediction of the required antenna size.
3) Not only is the user capacity characterized as an admissible region of SINR requirements, but we also propose a joint pilot sequence design and transmit power allocation scheme to achieve this capacity. The detailed comparisons between our proposed capacity-achieving scheme and other non-capacity-achieving schemes have been made. Particularly, the conventional pilot sequence design is taken into account in one of the non-capacity-achieving schemes. The results highlight the advantage of using the proposed scheme, i.e., the potential of admitting more UEs into the specified Massive MIMO system. This advantage becomes increasingly apparent when UEs have more diverse SINR requirements.

C. Organization and Notations

The rest of the paper is as follows. In Section II, we specify the system model including the uplink training method and the downlink transmit beamforming scheme. Section III addresses the issue of admissible users in the pilot-contaminated system. This upper bound is further translated into an admissible region in which the SINR requirements can be satisfied by using the proposed joint scheme of pilot sequence design and transmit power allocation. The analysis in the non-LAS regime is

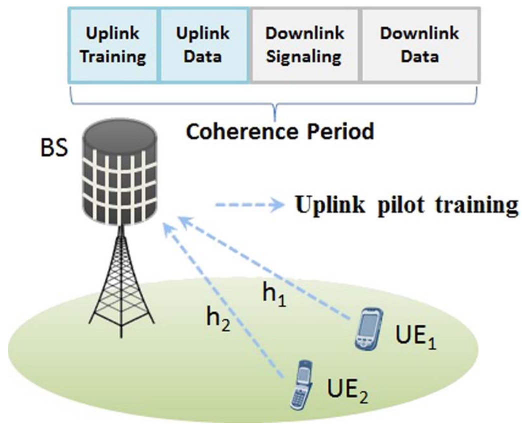

Fig. 1. A single-cell TDD Massive MIMO network, in which the BS acquires CSI via uplink training.

provided in the same section. Numerical results are presented in Section IV, which is followed by the concluding remarks in Section V.

Notations: $\mathbb{R}$: real number, $\mathbb{Z}$: integers, $\|\cdot\|_p$: p-norm, $(\cdot)^T$: transpose, $(\cdot)^H$: Hermitian transpose, $\otimes$: Kronecker product, $\circ$: Hadamard product, $\mathbf{I}_N$: $N \times N$ identity matrix, $\mathcal{CN}(\cdot,\cdot)$: complex normal distribution, $\mathbb{E}[\cdot]$: expectation, $\mathrm{tr}(\cdot)$: trace, $\mathrm{diag}(\cdots)$: diagonal matrix, $\succ, \succeq$: vector inequalities, $\mathbf{0}$: zero vector, Null$(\cdot)$: null space, card$(\cdot)$: cardinality.

II. System Model

Consider a single-cell Massive MIMO system, as shown in Fig. 1, where a BS equipped with M antennas serves K single-antenna UEs. With TDD operation, the BS acquires downlink CSI through uplink pilot training. The acquired CSI will then be utilized to form MRT precoding vectors for downlink spatial multiplexing.

A. Uplink Training

During the uplink training phase, the ith UE transmits its own pilot sequence $\mathbf{s}_i \in \mathbb{R}^{\tau \times 1}$ with $\|\mathbf{s}_i\|_2 = 1$. The correlation $\mathbf{s}_i^T \mathbf{s}_j = \rho_{ij}$ between different training sequences ranges from -1 to 1. The pilot data over the block-fading channel, synchronously received at the BS, can be expressed as

$$\mathbf{y}_{(\tau M \times 1)} = \sum_{i=1}^{K} \left(p_{s,i}^{1/2} \mathbf{S}_i\right) \left(\beta_i^{1/2} \mathbf{h}_i\right) + \mathbf{z}, \tag{1}$$

where $p_{s,i}$ and β_i denote respectively the pilot power and the channel power gain, $\mathbf{S}_i$ is the $\tau M \times M$ matrix given by $\mathbf{S}_i = \mathbf{s}_i \otimes \mathbf{I}_M$, $\mathbf{z}$ is the additive white Gaussian noise distributed as $\mathcal{CN}(\mathbf{0}, \sigma_z^2 \mathbf{I}_{\tau M})$, and $\mathbf{h}_i (1 \le i \le K)$ are independent and identically distributed (i.i.d.) small-scale fading channels from UEs to the BS with distribution $\mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. This channel model is commonly assumed in Massive MIMO systems [6], [8], [12]. We further assume that certain uplink power control is enabled to make the received signals from UEs be of equal power, i.e., $p_{s,i}\beta_i = 1$, and thus a fair estimation performance can be achieved. Without exploiting statistical knowledge of $\mathbf{h}_i$, the LS method is adopted to provide the channel estimate

$$\begin{aligned} \hat{\mathbf{h}}_i &= \mathbf{P}_{\mathrm{LS},i} \mathbf{y}, \\ &= \mathbf{h}_i + \sum_{j \ne i}^{K} \rho_{ij} \mathbf{h}_j + \mathbf{S}_i^T \mathbf{z}, \end{aligned} \tag{2}$$

where $\mathbf{P}_{\mathrm{LS},i} = \mathbf{S}_i^T$. This estimate indicates how the desire channel information is polluted by undesired channels in the pilot-contaminated regime $(K > \tau)$, where ρ_{ij} may not be 0 for some $j \ne i$.

B. Downlink Transmission

Utilizing estimated CSI at the BS, linearly precoded signals are formed and simultaneously transmitted to UEs. The received signal at the ith UE is

$$\mathbf{r}_i = \beta_i^{1/2} \mathbf{h}_i^H \left(\sum_{j=1}^{K} \mathbf{t}_j x_j\right) + w_i, \tag{3}$$

where $\mathbf{t}_i$ is a linear precoding vector, x_i denotes uncorrelated zero-mean data symbols with power $\mathbb{E}[x_i^H x_i] = P_i$, and w_i is the zero-mean noise with variance σ_w^2. Let us devise the ith precoding vector $\mathbf{t}_i = \hat{\mathbf{h}}_i / \sqrt{M(\sum_{j=1}^K \rho_{ij}^2 + \sigma_z^2)}$, where the denominator ensures that $\mathbf{t}_i$ has unit 2-norm as M grows without limit. The fact, $\lim_{M\to\infty} \frac{1}{M} \hat{\mathbf{h}}_i^H \hat{\mathbf{h}}_i = \sum_{j=1}^K \rho_{ij}^2 + \sigma_z^2$, can be easily proved by making use of the following results [12]

$$\lim_{M\to\infty} \frac{1}{M} \mathbf{h}_i^H \mathbf{h}_j = \begin{cases} 0, & \text{if } i \ne j, \\ 1, & \text{if } i = j. \end{cases} \tag{4}$$

Such asymptotic orthogonality among MIMO channels has been experimentally verified in realistic propagation environments [4].

Consider that channel estimates are available at the BS and only channel statistic $\mathbb{E}[\mathbf{h}_i^H \mathbf{t}_i]$ is known by the ith UE. Following this, the received signal (3) is recast as

$$\begin{aligned} \mathbf{r}_i = {} & \underbrace{\beta_i^{1/2} \mathbb{E}\left[\mathbf{h}_i^H \mathbf{t}_i\right] x_i}_{\text{part1}} \\ & \underbrace{+ \beta_i^{1/2} \left(\mathbf{h}_i^H \mathbf{t}_i - \mathbb{E}\left[\mathbf{h}_i^H \mathbf{t}_i\right]\right) x_i + \beta_i^{1/2} \mathbf{h}_i^H \left(\sum_{j\ne i}^{K} \mathbf{t}_j x_j\right) + w_i,}_{\text{part 2}} \end{aligned} \tag{5}$$

where the second part can be regarded as the effective noise which is uncorrelated with the first part. As shown in [8, Theorem 1], the above expansion facilitates defining an ergodic achievable rate $\log_2(1 + \mathrm{SINR}_{i,M})$, where

$$\mathrm{SINR}_{i,M} = \frac{\left(\mathbb{E}\left[\mathbf{h}_i^H \mathbf{t}_i\right]\right)^2 \beta_i P_i}{\mathrm{Var}\left[\mathbf{h}_i^H \mathbf{t}_i\right] \beta_i P_i + \sum_{j\ne i}^{K} \mathbb{E}\left|\mathbf{h}_i^H \mathbf{t}_j\right|^2 \beta_j P_j + \sigma_w^2}, \tag{6}$$

in which $\mathrm{Var}[\mathbf{h}_i^H \mathbf{t}_i] = \mathbb{E}|(\mathbf{h}_i^H \mathbf{t}_i - \mathbb{E}[\mathbf{h}_i^H \mathbf{t}_i])|^2$. The following lemma is a consequence of this definition, describing how the SINR varies with the BS antenna size.

Lemma 1: The $\mathrm{SINR}_{i,M}$ defined in (6) can be expressed as

$$SINR_{i,M} = \frac{\beta_i P_i}{\beta_i \left(\sum_{j \neq i}^K \rho_{ji}^2 P_j + \frac{1}{M} \sum_{j=1}^K \alpha_j P_j \right) + \frac{1}{M} \alpha_i \sigma_w^2}, \tag{7}$$

Proof: Please refer to Appendix A. □

In the LAS regime ($M \to \infty$), we obtain the limit form of the effective SINR

$$\begin{aligned} \mathrm{SINR}_{i,\infty} &= \frac{P_i}{\sum_{j \neq i}^K \rho_{ji}^2 P_j}, \\ &= \frac{P_i}{\mathrm{tr}\left(s_i^T \mathbf{S}\mathbf{D}\mathbf{S}^T s_i \right) - P_i}, \end{aligned} \tag{8}$$

where $\mathbf{D} = \mathrm{diag}(P_1, \cdots, P_K)$, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K]$, the symbol ∞ denotes that the antenna size goes to infinity so that the effect of the noise w_i can be ignored. This SINR expression suggests that the downlink transmission operates in the interference-limited regime because of using a large number of antennas. Moreover, in the pilot-contaminated regime, the interference part, $\sum_{j \neq i}^K \rho_{ji}^2 P_j$, can not be simultaneously eliminated for every UE as non-orthogonal pilot sequences have to be used.

III. Downlink User Capacity Characterization

The asymptotic SINR expression, identified in the previous section, will be further employed to examine the user capacity in the LAS regime. In this examination, $K > \tau$ will be implicitly assumed as the pilot-contaminated regime is of interest. Therefore, the following results will show the effects of pilot contamination on the user capacity, and can be viewed as the performance limits due to the antenna size growing limitlessly.

A group of UEs is said to be *admissible* in the specified TDD Massive MIMO system if there exists a feasible pilot sequence matrix $\mathbf{S}$ and a power allocation vector $\mathbf{p} = [P_1, \cdots, P_K]^T \succ \mathbf{0}$ such that the SINR requirements, $\mathrm{SINR}_{i,\infty} \geq \gamma_i$ for $1 \leq i \leq K$, can be jointly satisfied. A pilot sequence matrix $\mathbf{S}$ is claimed to be feasible if $\mathbf{S} \in \mathcal{S} = \{[\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K], \mathbf{s}_i \in \mathbb{R}^{\tau \times 1} | \|\mathbf{s}_i\|_2 = 1\}$. How the number of admissible UEs is related to the pilot length and SINR requirements will be presented in the proposition below.

Proposition 2: If K UEs are admissible in the TDD Massive MIMO system, then

$$K \leq \left\{ \tau \left[\sum_{i=1}^K \left(1 + \frac{1}{\gamma_i} \right) \right] \right\}^{1/2}. \tag{9}$$

Proof: Please refer to Appendix B. □

It is clear from this proposition that the number of admissible UEs is fundamentally limited once the length of pilot sequences and the SINR requirements are given. For instance, if a group of UEs has high SINR requirements, this group is less likely to be admissible as (9) may not hold with a given pilot length. On the other hand, (9) is always true for $K \leq \tau$, which is not the case under consideration. As the normalized mean-square error $\mathrm{MSE}_{i,\infty}$ seen by the ith UE is equivalent to $1/\mathrm{SINR}_{i,\infty}$, another explanation can be offered. A lower bound on the sum of $\mathrm{MSE}_{i,\infty}$ is given by

$$\begin{aligned} \sum_{i=1}^K \mathrm{MSE}_{i,\infty} &= \mathrm{tr}\left(\mathbf{D}^{-1} \mathbf{S}^T \mathbf{S}\mathbf{D}\mathbf{S}^T \mathbf{S} \right) - K, \\ &\geq \frac{K^2}{\tau} - K, \end{aligned} \tag{10}$$

where (31) and (32) are applied in the second inequality. This bound can be immediately translated into a bound on K, i.e.,

$$K \leq \tau(\overline{\mathrm{MSE}} + 1), \tag{11}$$

where $\overline{\mathrm{MSE}} = \sum_{i=1}^K \mathrm{MSE}_{i,\infty}/K$ represents the average of the normalized mean-square errors. Hence, we reach the conclusion that the intrinsic bound on K is the result of inevitable mean-square errors caused by pilot contamination.

Proposition 2 provides an upper bound for the user capacity. The next question to answer is whether K UEs are admissible if the inequality (9) is satisfied, i.e., the achievability issue. In other words, once the UE number is less than or equal to the upper bound, we wonder if there exists a set of $\mathbf{S} \in \mathcal{S}$ and $\mathbf{p} \succ \mathbf{0}$, fulfilling the SINR requirements. The following section will show that the answer to this question is positive.

A. Capacity-Achieving Pilot Sequence Design and Transmit Power Allocation

Validating the converse of Proposition 2 requires to identify the pair of $\mathbf{S}$ and $\mathbf{p}$ with which we have $\mathrm{SINR}_{i,\infty} \geq \gamma_i$ for $1 \leq i \leq K$. The following proposition will specify a condition under which this pair, referred to as a valid combination of pilot sequence design and transmit power allocation, can exist.

Proposition 3: If

$$\sum_{i=1}^K \left(\frac{\gamma_i}{1 + \gamma_i} \right) \leq \tau, \tag{12}$$

then $K \leq \{\tau[\sum_{i=1}^K (1 + 1/\gamma_i)]\}^{1/2}$ and there exists a valid combination of pilot sequence design and transmit power allocation.

Proof: The Cauchy-Schwarz inequality gives

$$\begin{aligned} \sum_{i=1}^K \left(1 + \frac{1}{\gamma_i} \right) &\geq \frac{K^2}{\sum_{i=1}^K \left(\frac{\gamma_i}{1+\gamma_i} \right)}, \\ &\geq \frac{K^2}{\tau}, \end{aligned}$$

which is equivalent to (9) and proves the first part of the statement. Before going on to the second part, a definition and a lemma to be utilized later are provided.

Definition 4: Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, $\mathbf{x}$ majorizes $\mathbf{y}$ if

$$\sum_{k=1}^n x_{[k]} \geq \sum_{k=1}^n y_{[k]}, \text{ for } n = 1, \cdots, N,$$

where $x_{[k]}$ and $y_{[k]}$ are respectively the elements of $\mathbf{x}$ and $\mathbf{y}$ in decreasing order.

Lemma 5 [22, Theorem 9.B.2]: Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, if $\mathbf{x}$ majorizes $\mathbf{y}$ and $\sum_{k=1}^N x_{[k]} = \sum_{k=1}^N y_{[k]}$, then there exists a real symmetric matrix H with diagonal elements $y_{[k]}$ and eigenvalues $x_{[k]}$.

First consider the case of $\sum_{i=1}^K [\gamma_i/(1+\gamma_i)] = \tau$.

Given that the $1 \times K$ vector of eigenvalues $\mathbf{e} = [\lambda_1, \cdots, \lambda_\tau, 0 \cdots, 0]^T$ majorizes $\mathbf{p}$ and $\sum_{i=1}^\tau \lambda_i = \sum_{i=1}^K P_i$, following Lemma 5, there exists a real symmetric matrix $\mathbf{H} = \mathbf{Q}\Lambda\mathbf{Q}^T$, where the vector of diagonal entries of $\mathbf{H}$ is equal to $\mathbf{p}$, $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_\tau, 0 \cdots, 0)$, and the orthogonal matrix $\mathbf{Q}$ can be presented as

$$\begin{bmatrix} \mathbf{V}_{K\times\tau} & \tilde{\mathbf{V}}_{K\times(K-\tau)} \end{bmatrix}.$$

The approach to constructing $\mathbf{Q}$ as well as $\mathbf{H}$ is provided in [23, Sec. IV-A]. Define

$$\mathbf{S} \triangleq \Sigma^{1/2}\mathbf{V}^T\mathbf{D}^{-1/2}, \tag{13}$$

where $\Sigma = \mathrm{diag}(\lambda_1, \cdots, \lambda_\tau)$. Then, $\mathbf{S} \in \mathcal{S}$ is true as the diagonal entries of $\mathbf{S}^T\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{H}\mathbf{D}^{-1/2}$ are equal to 1. Moreover, we have

$$\mathbf{S}\mathbf{D}\mathbf{S}^T = \Sigma. \tag{14}$$

Let us specify that

$$\lambda_1 = \cdots = \lambda_\tau = \frac{\sum_{i=1}^K P_i}{\tau}, \tag{15}$$

and

$$P_i = c\frac{\gamma_i}{1+\gamma_i}, \text{for some } c > 0. \tag{16}$$

It can be verified that $\mathbf{e}$ majorizes $\mathbf{p}$ since for $1 \le i \le \tau$,

$$\begin{aligned} \lambda_i &= \frac{c\sum_{i=1}^K \left(\frac{\gamma_i}{1+\gamma_i}\right)}{\tau}, \\ &= c, \\ &> \max\left\{P_k, \text{ for } 1 \le k \le K\right\}, \end{aligned}$$

where the second equality is due to the case under consideration.

Next we will check if the SINR requirements are satisfied by using such pilot sequence design and transmit power allocation. Making use of (14), we have

$$\begin{aligned} \mathrm{SINR}_{i,\infty} &= \frac{P_i}{\mathrm{tr}\left(s_i^T\Sigma s_i\right) - P_i}, \\ &= \frac{c\frac{\gamma_i}{1+\gamma_i}}{c - c\frac{\gamma_i}{1+\gamma_i}}, \\ &= \gamma_i, \ \forall i = 1, \cdots, K, \end{aligned}$$

which means that all SINR requirements are met.

Now we turn to the case of $\sum_{i=1}^K [\gamma_i/(1+\gamma_i)] < \tau$. As $f(x) = x/(1+x)$ is monotonically increasing for $x > 0$, there exists a set $\{\hat{\gamma}_i \ge \gamma_i \text{ for } 1 \le i \le K\}$ such that $\sum_{i=1}^K [\hat{\gamma}_i/(1+\hat{\gamma}_i)] = \tau$. At the same time, $K \le \{\tau[\sum_{i=1}^K (1 + 1/\hat{\gamma}_i)]\}^{1/2}$ holds. By exploiting the previous result, we can find a corresponding valid set of $\mathbf{S} \in \mathcal{S}$ and $\mathbf{p} \succ \mathbf{0}$. $\square$

The valid set of $\mathbf{S}$ and $\mathbf{p}$ used in Proposition 3 is summarized here. The matrix $\mathbf{S}$ results from the generalized Welch bound equality (GWBE) sequences, satisfying $\mathbf{S}\mathbf{D}\mathbf{S}^T = c\mathbf{I}_\tau$ for some $c \in \mathbb{R}^+$[24]. The transmit power allocated to UE_i is given by $P_i = c\hat{\gamma}_i/(1+\hat{\gamma}_i)$ where $\hat{\gamma}_i \ge \gamma_i$ and $\sum_{i=1}^K [\hat{\gamma}_i/(1+\hat{\gamma}_i)] = \tau$. In the case of identical SINR requirements, the pilot sequences in use are called Welch bound equality (WBE) sequences. Both GWBE and WBE sequences have been extensively applied for constructing optimal CDMA signature sequences. These sequences also find applications in wavelet expansions and Grassmannian frames [25], [26]. In [27], a comparative evaluation of several sequence-construction algorithms is presented.

An explanation of the constraint, $\sum_{i=1}^K [\gamma_i/(1+\gamma_i)] \le \tau$, is presented as follows. When a UE has a high SINR requirement, the pilot sequence assigned to this UE should be orthogonal to others so that it will receive less interference. Overall, only τ such assignments are allowed in the system. So the system can roughly admit τ UEs with high SINR requirements. From another viewpoint, this constraint indicates that the desired pilot length is lower bounded for the given K and γ_i. In addition, the proposed pilot sequence design and transmit power allocation can be applied to meet the target SINRs using the minimum pilot length. A corollary which follows from Propositions 2 and 3 is provided below.

Corollary 6: Given the identical SINR requirement γ, K UEs are admissible in the TDD Massive MIMO system if and only if

$$K \le \left(1 + \frac{1}{\gamma}\right)\tau. \tag{17}$$

Unlike (9), the right-hand side of (17) does not depend on K, and thus it provides an explicit upper bound of admissible UEs. However, this is true only for identical SINR requirements. In the general case, (9) and (12) do not provide upper bounds of this kind. To have a consistent interpretation of the user capacity in the general case, we intend to characterize the user capacity as the *admissible region* $R_{\mathrm{UC}} = \{\gamma_{1\sim K} \in \mathbb{R}^+ | \sum_{i=1}^K [\gamma_i/(1+\gamma_i)] \le \tau\}$. It means that once the SINR requirements are located within R_{UC}, the corresponding K UEs are admissible. When it comes to identical SINR requirements, this region maintains the same structure, i.e., $R_{\mathrm{UC}} = \{\gamma_{1\sim K} = \gamma \in \mathbb{R}^+ | K\gamma/(1+\gamma) \le \tau\}$. Later on, this characterization will be utilized to evaluate different joint pilot sequence design and power allocation schemes, i.e., different combinations of $\mathbf{S}$ and $\mathbf{p}$.

According to the present analytical results, the following remarks can be made.

1) The valid combination of pilot sequence design and power allocation is also referred to as *the capacity-achieving scheme* or alternatively *the GWBE scheme*. It means that any K UEs having the SINR requirements within R_{UC} can be admitted by using this scheme. In the next section, it will be shown that other non-capacity-achieving schemes can not guarantee this.

2) When using the capacity-achieving pilot sequence and power allocation, the converse of Proposition 3 can be shown to be true.
3) Generally, the region specified by (9) and $\sum_{i=1}^{K}[\gamma_i/(1+\gamma_i)] > \tau$ has not been characterized, in which the existence of a valid set of $\mathbf{S}$ and $\mathbf{p}$ is unknown. However, when all the SINR requirements are the same, (9) turns into $\sum_{i=1}^{K}[\gamma/(1+\gamma)] \leq \tau$, making this uncharacterized region empty.
4) Following the previous remark, we can not remove the possibility that there exists a better scheme which leads to an admissible region larger than R_{UC}. For example, when GWBE sequences are employed, there might exist some superior power allocation scheme other than the proposed one.

B. Comparison With Non-Capacity-Achieving Schemes

The superiority of the proposed capacity-achieving scheme over other existing schemes will be presented in this section. Let us first define two non-capacity-achieving schemes as follows.

1) The WBE Scheme: Pilot sequences in use are the WBE sequences, having properties: $\mathbf{S}\mathbf{S}^T = \frac{K}{\tau}\mathbf{I}_\tau$, and $\rho_{ij}^2 = (K-\tau)/[(K-1)\tau]$ for $i \neq j$[28]. The transmit power P_i allocated to the ith UE is given by $c\gamma_i/(1+\gamma_i)$ for some $c > 0$.
2) The Finite Orthogonal Sequence (FOS) Scheme: Given a pilot sequence length τ, only τ orthogonal pilot sequences will be repeatedly used in the pilot-contaminated regime. Assume that $K = q\tau + r$ where $q, r \in \mathbb{Z}$ and $0 \leq r < \tau$. Each pilot sequence s_i is used by a collection E_i of UEs. Let $\mathrm{card}(E_i) = q+1$ for $1 \leq i \leq r$, $\mathrm{card}(E_i) = q$ for $r+1 \leq i \leq \tau$, and $E_i \cap E_j = \emptyset$ for $i \neq j$. This pilot design is commonly adopted in TDD Massive MIMO. The power allocation is the same as the one in the WBE scheme.

The main difference of these two schemes from the GWBE scheme lies in the pilot sequence design. Therefore, the comparisons made below largely point out the benefit of using the GWBE sequences. Note that the power allocation in the WBE scheme is adopted for ease of comparison. It is possible to have an alternative scheme such as $P_i = c\gamma_i$. However, how to identify the optimal power allocation corresponding to the WBE sequences remains an open question.

The following lemmas will show the potential reduction of the user capacity when the WBE and FOS schemes are applied to the case of general SINR constraints.

Lemma 7: The general SINR requirements γ_i are satisfied by using the WBE scheme if and only if

$$\sum_{i=1}^{K}\left(\frac{\gamma_i}{1+\gamma_i}\right) \leq \min\left\{\tau,\ \kappa - (\kappa-1)\left(\frac{\gamma_{max}}{1+\gamma_{max}}\right)\right\}, \tag{18}$$

where $\kappa = \frac{(K-1)\tau}{(K-\tau)}$ and $\gamma_{\max} = \max\{\gamma_i, 1 \leq i \leq K\}$.

Proof: Please refer to Appendix C. □

Lemma 8: The general SINR requirements γ_i are satisfied by using the FOS scheme if and only if

$$\sum_{k\in E_i}\left(\frac{\gamma_k}{1+\gamma_k}\right) \leq 1,\ for\ 1 \leq i \leq \tau, \tag{19}$$

and

$$\sum_{i=1}^{K}\left(\frac{\gamma_i}{1+\gamma_i}\right) \leq \tau. \tag{20}$$

Proof: Similar to the proof of Lemma 7. □

In Lemma 7, the second term in the curly brackets becomes less than or equal to τ as $\gamma_{\max} \geq \tau/(K-\tau)$. When $\gamma_{\max} \gg 1$, (18) turns into $\sum_{i=1}^{K}[\gamma_i/(1+\gamma_i)] \leq 1+\epsilon\ \forall\,\epsilon > 0$, which implies potentially fewer admissible UEs as the allowable combinations of $\gamma_1 \sim \gamma_K$ become much less. From Lemma 8, it follows that the FOS scheme imposes more constraints, i.e., (19), on the admissibility of UEs. Hence, compared with the GWBE scheme, the FOS scheme potentially can admit fewer UEs. More comparisons, taking account of the admissible region, among these three schemes will be provided in Section IV.

C. How Many BS Antennas Are Required?

So far this paper has focused on the LAS regime, and the previous discussion applies provided the BS antenna size goes to infinity. This section moves on to examine the performance in the non-LAS regime. With the aid of Lemma 1, we are able to answer the question "How many BS antennas are required to achieve a significant fraction of the ultimate $\mathrm{SINR}_{i,\infty}$?" The lemma below presents the required antenna sizes for three joint pilot sequence design and transmit power allocation schemes.

Lemma 9: To have $\mathrm{SINR}_{i,M}/\mathrm{SINR}_{i,\infty} \geq \nu_i$, where $\nu_i \in (0,1)$, for the GWBE scheme, the WBE scheme, and the FOS scheme, we need antenna sizes

$$M_{GWBE,i} \geq \frac{\beta_i\sum_{j=1}^{K}(\alpha_j P_j) + \alpha_i\sigma_w^2}{\left[\beta_i\sum_{j\neq i}^{K}\left(\rho_{ji}^2 P_j\right)\right]\left(\frac{1-\nu}{\nu}\right)}, \tag{21}$$

$$M_{WBE,i} \geq \frac{(K-1)\left(K+\tau\sigma_z^2\right)\left(\beta_i P_{total}+\sigma_w^2\right)}{(K-\tau)\beta_i(P_{total}-P_i)\left(\frac{1-\nu}{\nu}\right)}, \tag{22}$$

$$M_{FOS,i} \geq \frac{\left[card(E_k)+\sigma_z^2\right]\left(\beta_i\sum_{j\in E_k}P_j+\sigma_w^2\right)}{\beta_i\left(\sum_{j\in E_k\setminus\{i\}}P_j\right)\left(\frac{1-\nu}{\nu}\right)}, \tag{23}$$

respectively, where $P_{total} = \sum_{i=1}^{K}P_i$, and $i \in E_k$ is assumed for the FOS scheme.

Proof: (21) results from Lemma 1. (22) and (23) are basically the same as (21) except that the cross-correlations ρ_{ji} are replaced by the actual values. □

There are important implications of this result. With a higher ν, the number of the required BS antennas becomes larger whichever scheme is applied. In addition, the BS antenna size has to be significantly large in order to achieve $100\nu\%$ of the limit $\mathrm{SINR}_{i,\infty}$ when the transmit power allocated to a UE is relatively higher. For example, if the ratio P_i/P_{total} approaches 1 in the WBE scheme, then the corresponding M_{WBE} increases without limit. This observation suggests that if the BS antenna size is limited, UEs with similar SINR requirements, to whom similar powers are assigned, should be clustered together and be served cluster by cluster. Some supplementary remarks are added below.

1) If there is a shortfall in the transmission power budget P_{total}, leading to $\beta_i P_{total} \ll \sigma_w^2$, then the required antenna size to achieve a significant fraction of $\text{SINR}_{i,\infty}$ will become very large. This conclusion can be drawn from Lemma 9.
2) If $\text{SINR}_{i,M_{\text{GWBE}},i} \geq \nu_i \text{SINR}_{i,\infty} \geq \gamma_i$ holds for any index i, then using $M \geq \max\{M_{\text{GWBE},i}, 1 \leq i \leq K\}$ should guarantee that all downlink SINR targets are satisfied when the GWBE scheme is applied. A similar argument also holds for the other two schemes.
3) Knowing the achievable rates that are guaranteed is intrinsically important. Hence, a sensible alternative to the computation in Lemma 9 is to evaluate the required number of antennas to achieve a fraction of the ergodic rate $\log_2(1 + \text{SINR}_{i,\infty})$.

IV. Numerical Results

In this section, comparisons between the capacity-achieving and non-capacity-achieving schemes are made. The first set of analyses examine the impact of different schemes on the size of the admissible region of SINR requirements. Next, we investigate how target SINRs are asymptotically achieved and how many BS antennas are used to attain the desired performance.

A. Admissible Region Characterization in the LAS Regime

To verify the results in Proposition 3 and in Lemmas 7 and 8, we consider a pilot-contaminated Massive MIMO system with $K = 6$ and $\tau = 3$. By fixing certain SINR requirements $\{\gamma_4 = \gamma_5 = \gamma_6 = 1\}$, we look into admissible regions of the remaining SINR requirements given by

$$R_{\text{GWBE}} = \left\{ \gamma_{1\sim3} \in \mathbb{R}^+ \middle| \sum_{i=1}^{3} [\gamma_i/(1+\gamma_i)] \leq 3/2 \right\}, \quad (24)$$

$$R_{\text{WBE}} = R_{\text{GWBE}} \cap \left\{ \sum_{j=1}^{3} \gamma_{1\sim3} \in \mathbb{R}^+ \middle| \sum_{j=1}^{3} [\gamma_j/(1+\gamma_j)] \leq (7/2 - 4\gamma_i/(1+\gamma_i)), \text{ for } 1 \leq i \leq 3 \right\}, \quad (25)$$

and

$$R_{\text{FOS}} = R_{\text{GWBE}} \cap \left\{ \gamma_{1\sim3} \in \mathbb{R}^+ \middle| \gamma_i \leq 1, \text{ for } 1 \leq i \leq 3 \right\}, \quad (26)$$

for the GWBE, WBE, and FOS schemes. Note that it is implicitly assumed that $E_1 = \{\text{UE}_1, \text{UE}_4\}$, $E_2 = \{\text{UE}_2, \text{UE}_5\}$, and $E_3 = \{\text{UE}_3, \text{UE}_6\}$ for the FOS scheme. The upper boundaries of these regions in the positive orthant are plotted in Fig. 2. For the GWBE scheme, an extra restriction $\gamma_3 = \min\{\gamma_3, 5\}$ is placed as the admissible γ_3 can go to infinity. It can be observed that the boundary surface of R_{GWBE} lies well above those of R_{WBE} and R_{FOS}. This implies that R_{GWBE} contains more admissible points than R_{WBE} and R_{FOS}, so more general SINR constraints $\gamma_{1\sim3}$ can be met in the GWBE scheme.

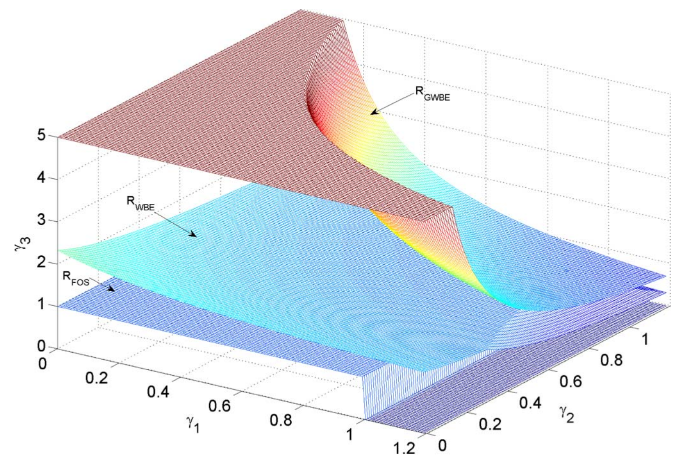

Fig. 2. Upper boundaries of admissible regions for the GWBE, WBE, and FOS schemes.

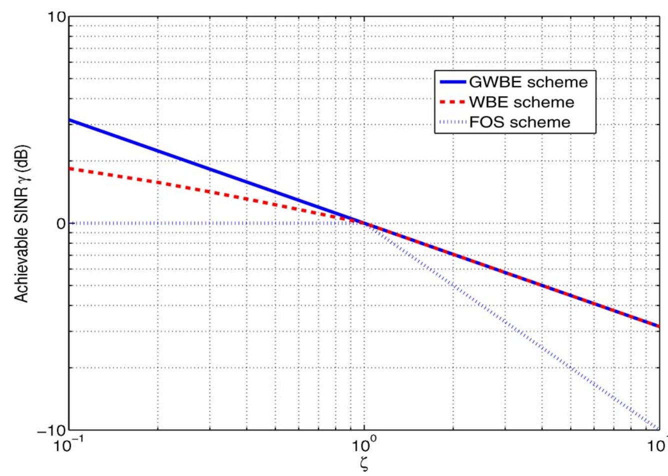

Fig. 3. Achievable SINR γ versus the factor ζ for the GWBE, WBE, and FOS schemes given a fixed SINR-requirement pattern, that is $\{\gamma_1 = \gamma_2 = \gamma_3 = \gamma, \gamma_4 = \gamma_5 = \gamma_6 = \zeta\gamma\}$.

Fig. 3 presents the achievable SINR for different schemes when the pattern of the SINR requirements is specified. The UE grouping, assumed for the FOS scheme, is given by $E_1 = \{\text{UE}_1, \text{UE}_2\}$, $E_2 = \{\text{UE}_3, \text{UE}_4\}$, and $E_3 = \{\text{UE}_5, \text{UE}_6\}$. The factor ζ adjusts a subset of the SINR requirements, making overall requirements of 6 UEs more or less diverse. As shown in this figure, there is a clear trend of decreasing the achievable SINR with increasing ζ. Additionally, it is obvious that the GWBE scheme outperforms the other two in terms of achieving higher SINRs for ζ ranging from 0.1 to 1. For $\zeta < 1$, the advantage of the GWBE scheme remains significant over the FOS scheme while it becomes negligible over the WBE scheme. Compared to the FOS scheme, the GWBE scheme is more

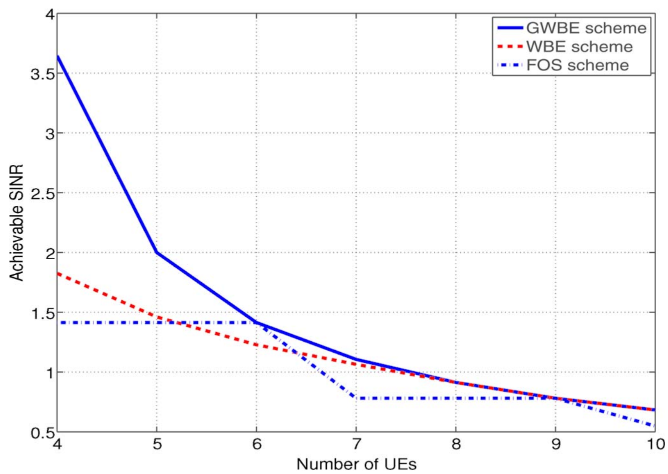

Fig. 4. Achievable SINR versus the number K of UEs for the GWBE, WBE, and FOS schemes given a fixed SINR-requirement pattern, that is $\{\gamma_1 = \gamma_2 = \gamma_3 = \gamma, \gamma_4 = \cdots = \gamma_K = \gamma/2\}$.

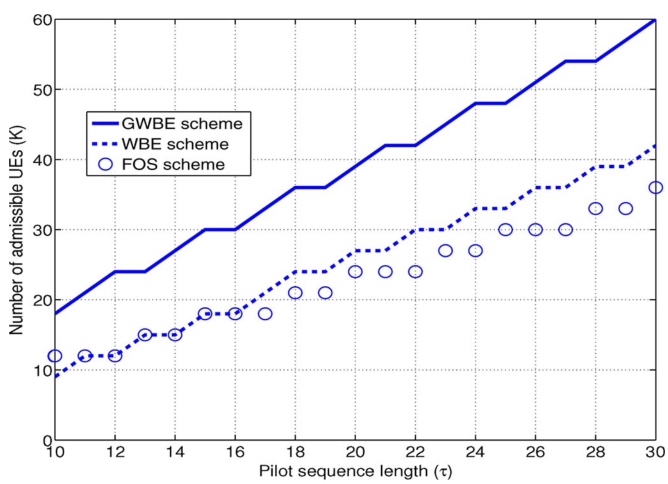

Fig. 5. Number of admissible UEs versus pilot sequence length for the GWBE, WBE, and FOS schemes, given a fixed SINR-requirement pattern, that is $\{\gamma_{1\sim l} = 1/3, \gamma_{(l+1)\sim 2l} = 1, \gamma_{(2l+1)\sim 3l} = 3\}$.

favorable when we have diverse SINR requirements, i.e, when ζ approaches 0.1 or 10.

To explore the effects of having different numbers of UEs, Fig. 4 plots the achievable SINR versus the number of UEs when fixing $\tau = 3$. For the FOS scheme, the grouping among UEs for any given K is assumed to be optimal in the sense of maximizing the achievable SINR. It can be observed that increasing K, making pilot contamination more serious, will decrease achievable SINRs for all three schemes. Our proposed GWBE scheme, however, attains relatively higher SINRs compared with the WBE and FOS schemes. Interestingly, the WBE scheme does not always outperform the FOS scheme for $K < 7$, but does so for $K \geq 7$. This highlights that the GWBE scheme exhibits a consistent superiority over the FOS scheme compared with the WBE scheme.

By specifying the SINR-requirement pattern of $K = 3l$ UEs, how many UEs are admissible for a given pilot length is depicted in Fig. 5. It can be observed that the number of admissible UEs scales almost linearly with the pilot length whatever scheme is adopted. This linear relationship directly

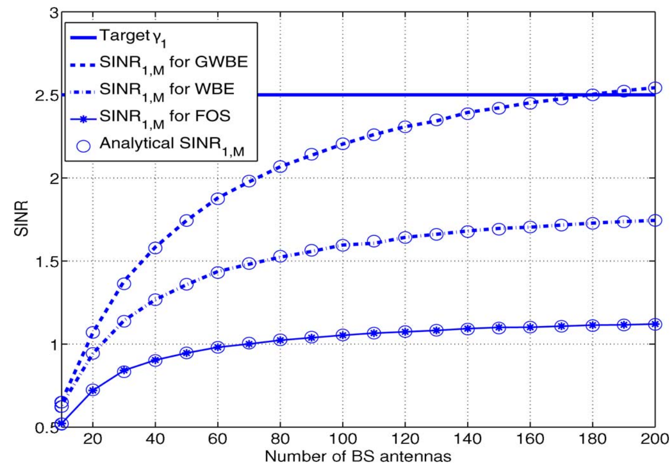

Fig. 6. Achieved $\mathrm{SINR}_{1,M}$ for a given number of BS antennas when using different joint pilot sequence and power allocation schemes.

demonstrates how the user capacity is limited by the pilot length. Also shown in the same figure, the GWBE scheme, without doubt, substantially outperforms the other two schemes in terms of admitting more UEs. In addition, two non-capacity-achieving schemes exhibit comparable user capacities especially with short pilot lengths.

B. Performance Comparison in the Non-LAS Regime

The results discussed in Section III-C will be numerically evaluated in this section. The system of interest still operates in the pilot-contaminated regime with $K = 6 > \tau = 3$. The noise powers, σ_z^2 and σ_w^2, at the BS and UEs, and all channel power gains β_i are set to 1. Three UE sets for the FOS scheme are the same as those specified in the previous section.

We first consider a set of SINR requirements $\gamma_1 \sim \gamma_6$, given by {5/2, 3/2, 2/3, 2/3, 3/8, 2/9}. This SINR set is located in the admissible region for the GWBE scheme, while outside those for the WBE scheme and FOS scheme. The transmit power allocated to the ith UE is $P_i = c\gamma_i/(1+\gamma_i)$ for the WBE scheme and FOS scheme and $P_i = c\hat{\gamma}_i/(1+\hat{\gamma}_i)$ for the GWBE scheme, where $c = 10$ and $\hat{\gamma}_{1\sim 6} = \{3, 2, 1, 1, 1/2, 1/3\}$. Fig. 6 presents the $\mathrm{SINR}_{i,M}$ of UE_1 achieved by different schemes for a given antenna size M. It can be seen that the simulated (dashed, dotted, dash-dot) curves due to the average over 100 000 channel realizations fit well with the analytical (circle-marker) values obtained from (7). Hence, the result in Lemma 1 is verified. Another observation is that the $\mathrm{SINR}_{1,M}$ curves of the WBE scheme and FOS scheme can not approach the target SINR by increasing the antenna size. It is because that both $\mathrm{SINR}_{1,\infty} = 1.92$ for the WBE scheme and $\mathrm{SINR}_{1,\infty} = 1.19$ for the FOS scheme are less than γ_1. In contrast, the GWBE scheme, having $\mathrm{SINR}_{1,\infty} = 3$ greater than γ_1, can achieve the target SINR with a finite number of BS antennas.

Fig. 7 shows the $\mathrm{SINR}_{i,M}$ curves of different UEs for the GWBE scheme and WBE scheme. We can see that the curves of $\mathrm{SINR}_{3,M}$ and $\mathrm{SINR}_{6,M}$ reach their own target SINRs respectively for the WBE scheme. This result means that as the set $\gamma_1 \sim \gamma_6$ falls outside the admissible region for the WBE scheme, only partial SINR requirements can be satisfied by this

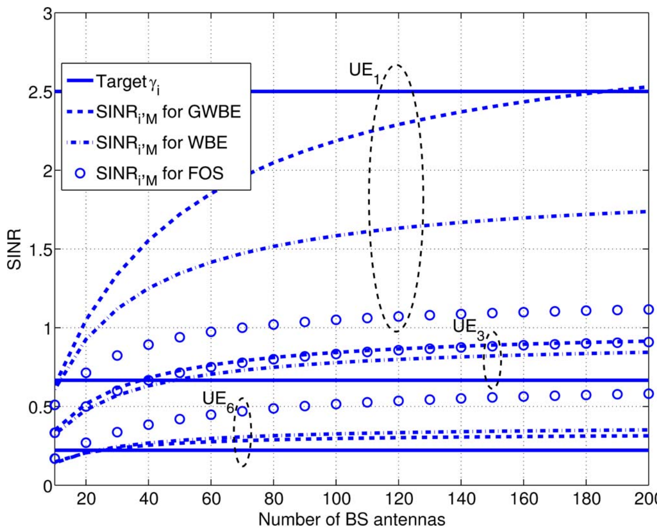

Fig. 7. Average received SINRs at UEs versus the number of BS antennas for the GWBE, WBE, and FOS schemes.

scheme. For UEs with lower SINR constraints, such as UE_3 and UE_6, it can be observed that the difference between the results due to two schemes becomes less significant. Next, the GWBE scheme is taken as an example for demonstrating the accuracy of results in Lemma 9. The SINR limits for this scheme are given by $\mathrm{SINR}_{i,\infty} = \hat{\gamma}_i > \gamma_i$ for $1 \le i \le 6$. If 83.3% of the limit $\mathrm{SINR}_{1,\infty}$ is achieved, the corresponding SINR requirements γ_1 will be met. Based on (22), the required antennas size is expected to be $M_{\mathrm{GWBE}} \ge 186$ for UE_1. Similarly, we need to achieve 66.6% of the limit $\mathrm{SINR}_{3,\infty}$, resulting in $M_{\mathrm{GWBE}} \ge 38$ for UE_3. This analytical inference is verified in Fig. 7, where the simulated $\mathrm{SINR}_{i,M}$ and target γ_i curves intersect at the points around $M = 196$ for UE_1 and $M = 38$ for UE_3. From then onwards, the simulated $\mathrm{SINR}_{i,M}$ is greater than the target SINR.

V. Conclusion

This paper has investigated the user capacity of downlink TDD Massive MIMO systems in the pilot-contaminated regime. The necessary condition for admitting a group of UEs with general SINR requirements has been provided. It has shown an intrinsic capacity upper bound due to the limited length of pilot sequences. Meanwhile, the capacity-achieving GWBE scheme, which can achieve the identified user capacity and satisfy the SINR requirements, has been proposed and compared with the non-capacity-achieving WBE and FOS schemes. The results of this study indicate that the capacity-achieving GWBE scheme is superior in terms of enhancing higher user capacity.

The current investigation has only examined the single-cell scenario. More study is needed to better understand the user capacity in the multicell scenario. Future research might explore the capacity region under intra- and inter-cell pilot contamination, the corresponding capacity-achieving pilot sequence design and transmit power allocation, and an efficient method for performing intercell resource allocation.

Appendix A
Proof of Lemma 1

First, we have

$$\begin{aligned}
&\left|\mathbf{h}_i^H \hat{\mathbf{h}}_j\right|^2 \\
&= \left(\sum_{m=1}^{K} \rho_{jm} \mathbf{h}_i^H \mathbf{h}_m + \mathbf{h}_i^H \mathbf{S}_j^H \mathbf{z}\right)^H \\
&\quad \times \left(\sum_{n=1}^{K} \rho_{jn} \mathbf{h}_i^H \mathbf{h}_n + \mathbf{h}_i^H \mathbf{S}_j^H \mathbf{z}\right), \\
&= \sum_{m=1}^{K}\sum_{n=1}^{K} \rho_{jm}\rho_{jn} \mathbf{h}_m^H \mathbf{h}_i \mathbf{h}_i^H \mathbf{h}_n + \sum_{m=1}^{K} \rho_{jm} \mathbf{h}_m^H \mathbf{h}_i \mathbf{h}_i^H \mathbf{S}_j^H \mathbf{z} \\
&\quad + \sum_{n=1}^{K} \rho_{jn} \mathbf{z}^H \mathbf{S}_j \mathbf{h}_i \mathbf{h}_i^H \mathbf{h}_n + \mathbf{z}^H \mathbf{S}_j \mathbf{h}_i \mathbf{h}_i^H \mathbf{S}_j^H \mathbf{z}.
\end{aligned}$$

Then,

$$\begin{aligned}
&\mathbb{E}\left[\left|\mathbf{h}_i^H \hat{\mathbf{h}}_j\right|^2\right] \\
&= \sum_{m=1}^{K}\sum_{n=1}^{K} \rho_{jm}\rho_{jn} \mathbb{E}\left[\mathbf{h}_m^H \mathbf{h}_i \mathbf{h}_i^H \mathbf{h}_n\right] + \mathbb{E}\left[\mathbf{z}^H \mathbf{S}_j \mathbf{h}_i \mathbf{h}_i^H \mathbf{S}_j^H \mathbf{z}\right], \\
&= \rho_{ji}^2 \mathbb{E}\left[\mathbf{h}_i^H \mathbf{h}_i \mathbf{h}_i^H \mathbf{h}_i\right] + \sum_{m=n\neq i}^{K} \rho_{jm}^2 \mathbb{E}\left[\mathbf{h}_m^H \mathbf{h}_i \mathbf{h}_i^H \mathbf{h}_m\right] \\
&\quad + \sum_{m\neq n}^{K} \rho_{jm}\rho_{jn} \mathbb{E}\left[\mathbf{h}_m^H \mathbf{h}_i \mathbf{h}_i^H \mathbf{h}_n\right] + \mathbb{E}\left[\mathrm{tr}\left(\mathbf{S}_j \mathbf{h}_i \mathbf{h}_i^H \mathbf{S}_j^H \mathbf{z}\mathbf{z}^H\right)\right], \\
&= \rho_{ji}^2 (M^2 + M) + \sum_{m\neq i}^{K} \rho_{jm}^2 \left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{h}_i} \mathbf{R}_{\mathbf{h}_m}\right)\right] \\
&\quad + \mathrm{tr}\left(\mathbf{R}_{\mathbf{h}_i} \mathbf{S}_j^H \mathbf{R}_{\mathbf{z}} \mathbf{S}_j\right), \\
&= \rho_{ji}^2 (M^2 + M) + M \sum_{m\neq i}^{K} \rho_{jm}^2 + M\sigma_z^2, \\
&= M^2 \rho_{ji}^2 + M\alpha_j,
\end{aligned} \tag{27}$$

where $\alpha_j = \sum_{m=1}^{K} \rho_{jm}^2 + \sigma_z^2$, $\mathbf{R}_{\mathbf{h}_i}$ and $\mathbf{R}_{\mathbf{z}}$ respectively denote the covariance matrices of $\mathbf{h}_i$ and $\mathbf{z}$, and $\mathbb{E}[\mathbf{h}_i^H \mathbf{h}_i \mathbf{h}_i^H \mathbf{h}_i] = (M^2 + M)$ because of $\mathbf{h}_i^H \mathbf{h}_i$ being gamma-distributed $\Gamma(M, 1)$.

By definition,

$$\begin{aligned}
&\mathrm{SINR}_{i,M} \\
&= \frac{\frac{\left(\mathbb{E}\left[\mathbf{h}_i^H \hat{\mathbf{h}}_i\right]\right)^2 \beta_i P_i}{M\alpha_i}}{\beta_i \left(\frac{\mathbb{E}\left|\left(\mathbf{h}_i^H \hat{\mathbf{h}}_i - \mathbb{E}\left[\mathbf{h}_i^H \hat{\mathbf{h}}_i\right]\right)\right|^2 P_i}{M\alpha_i} + \sum_{j\neq i}^{K} \frac{\mathbb{E}\left|\mathbf{h}_i^H \hat{\mathbf{h}}_j\right|^2 P_j}{M\alpha_j}\right) + \sigma_w^2}, \\
&= \frac{\frac{M^2}{M\alpha_i} \beta_i P_i}{\beta_i \left[\frac{M\alpha_i}{M\alpha_i} P_i + \sum_{j\neq i}^{K} \frac{M}{M\alpha_i}\left(M\rho_{ji}^2 + \alpha_j\right) P_j\right] + \sigma_w^2}, \\
&= \frac{\beta_i P_i}{\beta_i \left[\frac{\alpha_i}{M} P_i + \sum_{j\neq i}^{K} \frac{1}{M}\left(M\rho_{ji}^2 + \alpha_j\right) P_j\right] + \frac{\alpha_i}{M}\sigma_w^2}, \\
&= \frac{\beta_i P_i}{\beta_i \left(\sum_{j\neq i}^{K} \rho_{ji}^2 P_j + \sum_{j=1}^{K} \frac{\alpha_j}{M} P_j\right) + \frac{\alpha_i}{M}\sigma_w^2},
\end{aligned} \tag{28}$$

where $\mathbb{E}[\mathbf{h}_i^H \hat{\mathbf{h}}_i] = M$, and $\mathrm{Var}[\mathbf{h}_i^H \hat{\mathbf{h}}_i] = M\alpha_i$.

APPENDIX B
PROOF OF PROPOSITION 2

Making use of (8), we have

$$\sum_{i=1}^{K} \frac{1+\mathrm{SINR}_{i,\infty}}{\mathrm{SINR}_{i,\infty}} = \sum_{i=1}^{K} \frac{1}{P_i} \mathrm{tr}\left(s_i^T \mathbf{S}\mathbf{D}\mathbf{S}^T s_i\right), \\ = \mathrm{tr}(\mathbf{D}^{-1}\mathbf{S}^T\mathbf{S}\mathbf{D}\mathbf{S}^T\mathbf{S}), \\ = \mathrm{tr}\left(\mathbf{D}^{-1/2}\mathbf{G_s}\mathbf{D}\mathbf{G_s}\mathbf{D}^{-1/2}\right), \quad (29)$$

where

$$\mathbf{G_s} \triangleq \mathbf{S}^T\mathbf{S}, \\ = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1K} \\ \rho_{12} & 1 & \rho_{23} & \cdots & \rho_{2K} \\ \rho_{13} & \rho_{23} & 1 & \cdots & \rho_{3K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{1K} & \rho_{2K} & \rho_{3K} & \cdots & 1 \end{bmatrix}. \quad (30)$$

Also, we can expand the trace in (29) and obtain its lower bound as below

$$\mathrm{tr}\left(\mathbf{D}^{-1/2}\mathbf{G_s}\mathbf{D}\mathbf{G_s}\mathbf{D}^{-1/2}\right) \\ = K + \sum_{i=1}^{K}\sum_{j>i=1}^{K}\left(\frac{P_i}{P_j}+\frac{P_j}{P_i}\right)\rho_{ij}^2, \\ \geq K + \sum_{i=1}^{K}\sum_{j>i=1}^{K} 2\rho_{ij}^2, \\ = \mathrm{tr}\left(\mathbf{G_s}\mathbf{G_s}\right), \quad (31)$$

where the inequality is due to $(P_i/P_j + P_j/P_i) \geq 2$. The Gram matrix $\mathbf{G_s}$ has an eigendecomposition $\mathbf{U}\mathbf{D_G}\mathbf{U}^T$, where $\mathbf{U}$ is a unitary matrix and $\mathbf{D_G} = \mathrm{diag}(d_1, \cdots, d_K)$ with $d_1 \sim d_\tau > 0$, $d_{\tau+1} \sim d_K = 0$, and $\sum_{i=1}^{\tau} d_i = K$. Then, we have

$$\mathrm{tr}(\mathbf{G_s}\mathbf{G_s}) = \mathrm{tr}\left(\mathbf{U}\mathbf{D}_\mathbf{G}^2\mathbf{U}^T\right), \\ = \sum_{i=1}^{\tau} d_i^2, \\ \geq \frac{1}{\tau}\left(\sum_{i=1}^{\tau} d_i\right)^2 = \frac{K^2}{\tau}. \quad (32)$$

As

$$\sum_{i=1}^{K} \frac{1+\mathrm{SINR}_{i,\infty}}{\mathrm{SINR}_{i,\infty}} \leq \sum_{i=1}^{K} \frac{1+\gamma_i}{\gamma_i}, \quad (33)$$

the desired inequality follows.

APPENDIX C
PROOF OF LEMMA 7

The SINR requirement

$$\frac{P_i}{\sum_{j\neq i}^{K} \rho_{ji}^2 P_j} = \frac{P_i}{\sum_{j=1}^{K} \frac{1}{\kappa}P_j - \frac{1}{\kappa}P_i}, \\ = \frac{c\frac{\gamma_i}{1+\gamma_i}}{\frac{c}{\kappa}\left(\sum_{j=1}^{K}\frac{\gamma_j}{1+\gamma_j}\right) - \frac{c}{\kappa}\frac{\gamma_i}{1+\gamma_i}} \geq \gamma_i,$$

can be recast as

$$\sum_{j=1}^{K} \frac{\gamma_j}{1+\gamma_j} \leq \kappa - (\kappa-1)\left(\frac{\gamma_i}{1+\gamma_i}\right). \quad (34)$$

Let us sum up (34) over $i = 1, \cdots, K$, i.e.,

$$\sum_{i=1}^{K}\sum_{j=1}^{K} \frac{\gamma_j}{1+\gamma_j} = K\sum_{j=1}^{K}\frac{\gamma_j}{1+\gamma_j}, \\ \leq \sum_{i=1}^{K}\left[\kappa - (\kappa-1)\left(\frac{\gamma_i}{1+\gamma_i}\right)\right], \\ = K\kappa - (\kappa-1)\left(\sum_{i=1}^{K}\frac{\gamma_i}{1+\gamma_i}\right),$$

which gives

$$\sum_{j=1}^{K} \frac{\gamma_j}{1+\gamma_j} \leq \frac{K\kappa}{K+\kappa-1} = \tau. \quad (35)$$

The desired result follows from (34) and (35).

REFERENCES

[1] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
[2] H. Yang and T. Marzetta, "Total energy efficiency of cellular large scale antenna system multiple access mobile networks," in *Proc. IEEE Online Conf. GreenCom*, Piscataway, NJ, USA, Oct. 2013, pp. 27–32.
[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
[4] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE VTC*, San Francisco, CA, USA, Sep. 2011, pp. 1–5.
[5] T. Marzetta, "How much training is required for multiuser MIMO," in *Proc. 40th ACSSC*, Pacific Grove, CA, USA, Oct. 2006, pp. 359–363.
[6] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
[7] Y. Shi, J. Zhang, and K. B. Letaief, "Optimal stochastic coordinated beamforming for wireless cooperative networks with CSI uncertainty," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 960–973, Feb. 2015.
[8] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
[9] 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. 264–273, Feb. 2013.
[10] R. Mueller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.
[11] J. Zhang *et al.*, "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 759–772, Oct. 2014.
[12] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
[13] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular Networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
[14] X. Wu and W. Xu, "Downlink performance analysis with enhanced multiplexing gain in multi-cell large-scale MIMO under pilot contamination," in *Proc. IEEE WCNC*, Istanbul, Turkey, Apr. 2014, pp. 242–247.
[15] P. Viswanath, V. Anantharam, and D. Tse, "Optimal sequences, power control, user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1968–1983, Sep. 1999.

[16] C. Zhong, G. Zheng, and K.-K. Wong, “Feasibility conditions of linear multiuser MIMO systems in the asymptotic regime,” *IEEE Commun. Lett.*, vol. 11, no. 12, pp. 979–981, Dec. 2007.
[17] A. Kapur and M. Varanasi, “Multiuser detection for overloaded CDMA systems,” *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1728–1742, Jul. 2003.
[18] S. Ponnaluri and T. Guess, “Signature sequence and training design for overloaded CDMA systems,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1337–1345, Apr. 2007.
[19] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, “Pilot contamination problem in multi-cell TDD systems,” in *Proc. IEEE ISIT*, Seoul, South Korea, Jun./Jul. 2009, pp. 2184–2188.
[20] S. Kishore, L. Greenstein, H. Poor, and S. Schwartz, “Uplink user capacity in a multicell CDMA system with hotspot microcells,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1333–1342, Jun. 2006.
[21] M. Jung, Y. Kim, J. Lee, and S. Choi, “Optimal number of users in zero-forcing based multiuser MIMO systems with large number of antennas,” *J. Commun. Netw.*, vol. 15, no. 4, pp. 362–369, Aug. 2013.
[22] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. New York, NY, USA: Springer-Verlag, 2010.
[23] P. Viswanath and V. Anantharam, “Optimal sequences and sum capacity of synchronous CDMA systems,” *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1984–1991, Sep. 1999.
[24] S. Waldron, “Generalized Welch bound equality sequences are tight frames,” *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2307–2309, Sep. 2003.
[25] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61. Philadelphia, PA, USA: SIAM, 1992.
[26] T. Strohmer and R. W. Heath, Jr., “Grassmannian frames with applications to coding and communication,” *Appl. Comput. Harmon. Anal.*, vol. 14, no. 3, pp. 257–275, May 2003.
[27] J. Tropp, I. Dhillon, and R. Heath, “Finite-step algorithms for constructing optimal CDMA signature sequences,” *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2916–2921, Nov. 2004.
[28] S. Ulukus and R. Yates, “Iterative construction of optimum signature sequence sets in synchronous CDMA systems,” *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1989–1998, Jul. 2001.

Juei-Chin Shen (S’10–M’14) received the Ph.D. degree in communication engineering from University of Manchester, U.K., in 2013. In the same year, he joined the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), as a Research Associate. He is a co-recipient of IEEE PIMRC 2014 Best Paper Award. His research interests include statistical signal processing, cognitive radio, spectrum sensing, compressed sensing, massive MIMO systems, millimeter-wave communications, and random matrix theory.

Jun Zhang (S’06–M’10) received the B.Eng. degree in electronic engineering from the University of Science and Technology of China, in 2004, the M.Phil. degree in information engineering from the Chinese University of Hong Kong, in 2006, and the Ph.D. degree in Electrical and Computer Engineering from the University of Texas at Austin, in 2009. He is currently a Research Assistant Professor in the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST). He co-authored the book *Fundamentals of LTE* (Prentice-Hall, 2010). He received the 2014 Best Paper Award for the EURASIP Journal on Advances in Signal Processing, and the PIMRC 2014 Best Paper Award. He served as a MAC track co-chair for IEEE WCNC 2011. His research interests include wireless communications and networking, green communications, and signal processing.

Khaled B. Letaief (S’85–M’86–SM’97–F’03) received the B.S. degree (with distinction), and the M.S. and Ph.D. degrees, all in electrical engineering, from Purdue University in 1984, 1986 and 1990, respectively.

From January 1985 and as a Graduate Instructor at Purdue, he taught courses in communications and electronics. From 1990 to 1993, he was a faculty member at the University of Melbourne, Australia. Since 1993, he has been with the Hong Kong University of Science and Technology (HKUST) where he is currently Chair Professor and the Dean of Engineering, with expertise in wireless communications and networks. In these areas, he has over 500 journal and conference papers and has given invited keynote talks as well as courses all over the world. He has 13 patents including 11 US patents.

Dr. Letaief serves as a consultant for different organizations and is the founding Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He has served on the editorial board of other prestigious journals including the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS—WIRELESS SERIES (as Editor-in-Chief). He has been involved in organizing a number of major international conferences. These include WCNC’07 in Hong Kong; ICC’08 in Beijing; ICC’10 in Cape Town; TTM’11 in Hong Kong; and ICCC’12 in Beijing.

Professor Letaief has been a dedicated teacher committed to excellence in teaching and scholarship. He received the Mangoon Teaching Award from Purdue University in 1990; the HKUST Engineering Teaching Excellence Award; and the Michael Gale Medal for Distinguished Teaching (Highest University-wide Teaching Award at HKUST). He is also the recipient of many other distinguished awards including the 2007 IEEE Communications Society Publications Exemplary Award; the 2009 IEEE Marconi Prize Award in Wireless Communications; the 2010 Purdue University Outstanding Electrical and Computer Engineer Award; the 2011 IEEE Communications Society Harold Sobol Award; the 2011 IEEE Wireless Communications Technical Committee Recognition Award; and 11 IEEE Best Paper Awards.

Dr. Letaief is a Fellow of HKIE. He has served as an elected member of the IEEE Communications Society (ComSoc) Board of Governors, as an IEEE Distinguished lecturer, IEEE ComSoc Treasurer, and IEEE ComSoc Vice-President for Conferences.

He is currently serving as the IEEE ComSoc Vice-President for Technical Activities as a member of the IEEE Product Services and Publications Board, and is a member of the IEEE Fellow Committee. He is also recognized by Thomson Reuters as an ISI Highly Cited Researcher.