Achieving Massive Connectivity with Non-orthogonal Pilots and Dirty-RFs

Kyung Jun Choi, *Member, IEEE*, and Kwang Soon Kim†, *Senior Member, IEEE*e

***Abstract*— Emerging 5G wireless networks need to support sufficient connectivity for a large number of machine-type communication devices, which may have poor radio-frequency (RF) circuit quality due to cost- and energy-efficient design. In this paper, the connectivity is analyzed for a training-based large-scale antenna system employing both non-orthogonal pilots and dirty-RFs on both the transmitter and receiver sides. By considering the effects of imperfect hardware and interference caused by non-orthogonal pilots, the performance of the linear minimum mean-square error (MMSE) channel estimator is derived, and the corresponding average rate of the maximum ratio combiner (MRC) is obtained. From these results, the connectivity maximization problem is formulated and the closed-form solutions for the optimal training length and the optimal number of simultaneously served users are provided. Asymptotic analysis further reveals that allowing only orthogonal pilots limits both the connectivity even when a large number of antenna is employed. However, by allowing non-orthogonal pilots, both the connectivity and the energy efficiency can be improved significantly, even when dirty-RFs are taken into account.**

***Index Terms*—Dirty-RFs, non-orthogonal pilots, large-scale antenna system, training-based transmission.**

I. Introduction

New varieties of 5G scenario, such as machine-type communication (MTC) or Internet-of-Things (IoT), require massive connectivity in cellular systems [1]. To achieve this requirement, one promising approach is to increase the spatial degree-of-freedom (DoF) by employing a large-scale antenna system concept (a.k.a. massive multiple input multiple output (MIMO)), in which a base station (BS) equipped with a very large number of antennas serves a number of users simultaneously with the same time-frequency resource [2]. Since the use of multiple antennas has, thus far in practice, extended the number of simultaneous uplink access streams to match the number of antennas, it has been widely believed that employing the large-scale antenna system concept in a cellular system will provide the same, straightforward connectivity gain [3]. However, most MTC or IoT devices are power-limited and cost-limited so that the use of expensive and power-inefficient high-quality radio-frequency (RF) is inappropriate [4], and the use of low-quality RF with a low transmit power is inevitable. In addition, the use of high-quality analog-to-digital converters (ADCs) requires non-negligible power consumption and complexity, and their use may be limited, even for a BS due to the large number of RF chains. Thus, for an MTC or IoT scenario, it would be more reasonable to assume an large-scale antenna system with dirty-RFs on both the transmitter and receiver sides.

With dirty-RFs at the transmitter, various transmitter hardware impairment effects need to be considered, including in-phase/quadrature-phase (I/Q) imbalance, RF non-linearity, and oscillator phase noise. Although introducing exact models of such transmitter hardware impairment effects is nearly impossible, it has been reported that such transmitter hardware impairment effects can be well-modeled as transmit additive Gaussian noise [5]. This model has been used to derive the achievable rate of a point-to-point communication system while assuming either perfect channel state information (CSI) [6], [7] or imperfect CSI [8]. On the other hand, the use of a low-precision ADC is the main receiver hardware impairment, and its quantization effect can also be well-modeled as either additive quantization noise, independent of the transmitted signal, or thermal noise [11]. The achievable rate under such a receiver hardware impairment has been investigated subject to either Rayleigh fading [11] or Rician fading [12]. Recently, the achievable rate of a point-to-point communication system with transceiver hardware impairment has been considered by assuming perfect CSI [13]. However, its quantization effect on the CSI acquisition step is not negligible, and it is necessary to jointly consider the effects of the transceiver hardware impairment on both the CSI acquisition step and the data transmission step to implement an large-scale antenna system and achieve massive connectivity.

On top of the potentially large number of spatial DoFs in an large-scale antenna system, another promising approach is to share each spatial DoF by adopting non-orthogonal multiple access (NOMA), in which more than one user access simultaneously over a single orthogonal resource, and each user's signal is distinguished in either the power domain [14] or the code domain [15]. Information theory indicates that NOMA can provide little gain over its orthogonal counterpart in an large-scale antenna system if perfect CSI is available at the BS, because there are already plenty of spatial DoFs. However, it is impossible to acquire perfect CSI due to the limited transmit power available and the CSI even degrades due to the dirty-RFs. In addition, the pilot overhead needs to be taken into account, which may increase along with the number of users served simultaneously in case of an orthogonal multiple access. Thus, it is unclear ***how the connectivity can***

This work was supported in part by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea Government (MSIT) (2015-0-00300, Multiple Access Technique with Ultra-Low Latency and High Efficiency for Tactile Internet Services in IoT Environments), and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2014R1A2A2A01007254).

The authors are with the Department of Electrical and Electronic Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Korea.

†: Corresponding author (ks.kim@yonsei.ac.kr)

be maximized and what the maximum connectivity is in a training-based large-scale antenna system employing non-orthogonal pilots and dirty-RFs. This paper aims to address these ground questions and present valuable insights.

In order to maximize the connectivity in an large-scale antenna system with dirty-RFs in practice, it is generally preferred to employ non-orthogonal pilots with a consideration for erroneous CSI acquisition. Some optimal designs for non-orthogonal pilots are proposed in [17][18]. In [17], the optimal non-orthogonal pilots to minimize the channel estimation error are derived, and it turns out that finding the optimal non-orthogonal pilots is equivalent to solving the Grassmannian subspace packing problem. In [18], an iterative algorithm is proposed to find the optimal non-orthogonal pilotsto maximize the number of users under a minimum rate constraint in a downlink large-scale antenna system. In addition, a non-orthogonal pilot is employed to reduce network latency in a downlink large-scale antenna system [19]. However, the optimal use of non-orthogonal pilots for massive connectivity in a training-based large-scale antenna system with dirty-RFs has not yet been investigated.

In this paper, the connectivity of an uplink training-based large-scale antenna system employing both non-orthogonal pilots and dirty-RFs on both transmitter and receiver sides is derived and then maximized by optimizing the training length and the number of users served simultaneously. The major concern of this paper is to address the following question and to find the answer:

- How the connectivity can be maximized in a training-based uplink large-scale antenna system employing non-orthogonal pilots and dirty-RFs in both users and base station?

In order to find the answer of the above question, the detailed apporach used in this paper is summarized as follows.

- We first model the distortion of dirty-RFs by adopting the transmit noise model for the transmitter hardware impairment and the additive quantization noise model for the receiver hardware impairment. By using the worst-case uncorrelated Gaussian noise argument, we derive the mean-square error (MSE) performance of the linear minimum mean-square-error (MMSE) channel estimator as well as an accurate and simple approximation for the average user rate of the maximum ratio combiner (MRC) receiver.
- We formulate the connectivity maximization problem with orthogonal pilots and non-orthogonal pilots and derive the closed-form solutions to both the optimal training length and the optimal number of users served simultaneously. Then, we discuss whether the non-orthogonal pilots are beneficial or not according to the different levels of transceiver hardware impairment, frame size, and transmit power. In order to obtain more insight, an asymptotic analysis is performed for the case of using a power-saving scheme.
- Similar to the connectivity maximization problem, the energy efficiency maximization problem is also formulated and solved. It turns out that single-user transmission is the method to achieve the maximum energy efficiency, and both the near-optimal training length and transmit power are obtained by solving two sub-problems iteratively. Then, the achievable region of the connectivity and the energy efficiency is obtained and is shown to be improved significantly by allowing non-orthogonal pilots even with dirty-RFs, as long as they are taken into account appropriately.

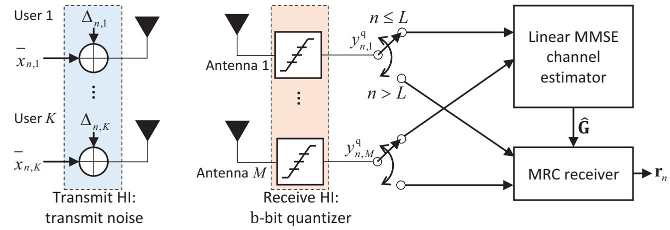


Fig. 1. System model with the transceiver hardware impairments.

Notations: Boldface lower case is used for denoting column vectors, such as $\mathbf{a}$ and boldface upper case is used for matrices, such as $\mathbf{A}$, $[\mathbf{a}]_i$, $\mathbf{A}^H$, and $\|\mathbf{A}\|_F$ denote the ith element of a vector $\mathbf{a}$, the conjugate transpose of $\mathbf{A}$, and the Frobenius norm of $\mathbf{A}$, respectively. A circularly symmetric complex Gaussian random vector is denoted by $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{C})$, where $\mathbf{m}$ is the mean and $\mathbf{C}$ is the covariance matrix. The expectation operator is denoted by $\mathbb{E}[\cdot]$. $\lceil x \rfloor$ denotes the function rounding x towards the nearest integer, $(x)^+ = \max\{x, 0\}$, $[a, b] = \{x | a \le x \le b\}$, and $(a, b) = \{x | a < x < b\}$. Finally, $\exists x \in \mathcal{S}$ indicate that there exists x in a set $\mathcal{S}$. The notations of symbols are summarized in Table I.

II. Modeling Transceiver Hardware Impairments

We consider an uplink training-based large-scale antenna system consisting of a BS with M antennas and K single-antenna users, as shown in Fig. 1.[1] One frame consists of N symbols, where the first L symbols are dedicated to pilot sequences and the remaining $N - L$ symbols are dedicated to data symbols. The (un-quantized) received signal at the nth symbol interval can be written as

$$\mathbf{y}_n = \mathbf{G}\mathbf{x}_n + \mathbf{v}_{\mathsf{an},n}, \tag{1}$$

where $\mathbf{x}_n \in \mathbb{C}^{K \times 1}$ and $\mathbf{v}_{\mathsf{an},n} \sim \mathcal{CN}(0, \mathbf{I}_M)$ denote the transmitted signal vector of K users and the noise vector at the nth symbol interval, respectively. The composite channel fading matrix, $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_K] \in \mathbb{C}^{M \times K}$, can be decomposed into $\mathbf{G} = \mathbf{H}\mathbf{D}^{1/2}$, where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]$ denotes the small-scale fading matrix whose elements are independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ and $\mathbf{D}$ denotes the large-scale fading diagonal matrix with $[\mathbf{D}]_{jj} = \beta_j$. In this paper, we assume a block fading model, where $\mathbf{H}$ remains constant within a frame but is independent across different frames, while $\mathbf{D}$ is invariant over much longer intervals and is perfectly known at the BS. Note that $\mathbf{x}_n$ in (1) may include

[1] Without loss of generality, our system model can be straightforwardly extended to a multi-antenna user case, i.e., it can be considered as the K' users with X antennas such that $K = K'X'$.

TABLE I
SYMBOL NOTATIONS

Symbols	Notations
M, K, N, L	the number of BS antennas, the number of signal-antenna users, frame length, and training interval length, respectively
δ, ρ	the parameters to model transmitter hardware impairment and receiver hardware impairment
$\overline{\mathbf{x}}_n, \mathbf{x}_n, \mathbf{\Delta}_n$	the desired transmit signal, the distorted transmit signal, and the transmitter hardware impairment at the nth symbol interval, respectively
$\overline{\mathbf{y}}_n, \mathbf{y}_n$	the received signal before quantization and the received signal after quantization at the nth symbol interval, respectively
$p_j^{\mathrm{tr}}, p_j^{\mathrm{dt}}, \overline{p^{\mathrm{tr}}}$	the desired signal power of user j during training and data transmission, respectively, and the received target signal power during training
$\mathbf{v}_{\mathsf{an},n}, \mathbf{v}_{\mathsf{tx},n}, \mathbf{v}_{\mathsf{rx},n}$	the additive Gaussian white noise, the noise caused from the transmitter hardware impairment, and the noise caused from the receiver hardware impairment at the nth symbol interval, respectively
$\mathbf{g}_j, \mathbf{h}_j, \beta_j$	the composite channel fading, the small-scale fading and the large-scale fading between user j and BS, respectively
$\mathbf{\Psi}, \boldsymbol{\psi}_j$	the pilot matrix and the pilot sequence of length L allocated to user j
$A, B, C, \phi, \tau, \theta$	Auxiliary variables defined in Theorem 2
μ, ξ, ϖ	Auxiliary variables defined in Theorem 3
D, E	Auxiliary variables defined in Section IV

not only the desired signal but also the distortion caused from dirty-RFs and $\mathbf{y}_n$ in (1) will be further quantized with low-precision ADCs at the BS.

The transmitter hardware impairment is well-modeled as additive Gaussian noise on the transmitter side so that $\mathbf{x}_n$ in (1) can be written as [5], [9], [10]

$$\mathbf{x}_n = \overline{\mathbf{x}}_n + \mathbf{\Delta}_n, \tag{2}$$

where $\overline{\mathbf{x}}_n \in \mathbb{C}^{K\times 1}$ denotes the desired signal to be transmitted at the nth symbol interval with $\mathbb{E}[\overline{\mathbf{x}}_n\overline{\mathbf{x}}_n^H] = \mathbf{P}^{\Upsilon} = \mathrm{diag}(p_1^{\Upsilon}, \ldots, p_K^{\Upsilon})$, $\Upsilon = \mathrm{tr}$ (for training phase) if $n \in [1, L]$ and $\Upsilon = \mathrm{dt}$ (for data transmission phase) if $n \in [L+1, N]$, and $\mathbf{\Delta}_n \in \mathbb{C}^{K\times 1}$ denotes the transmit noise vector with $\mathbb{E}[\mathbf{\Delta}_n\mathbf{\Delta}_n^H] = \delta^2\mathbf{P}^{\Upsilon}$,[2] which is independent to $\overline{\mathbf{x}}_n$. It is worth noting that the transmit-side additive noise, $\mathbf{\Delta}_n$, is modeled to be dependent on the statistics of $\overline{\mathbf{x}}_n$, but not to be dependent to the actual realizations of $\overline{\mathbf{x}}_n$. Such a model has been widely used in literature due to its mathemathical tractability at an acceptably low error. See [5] for the detailed information. Inserting (2) into (1), we have $\mathbf{y}_n = \mathbf{G}\overline{\mathbf{x}}_n + \mathbf{v}_{\mathsf{an},n} + \mathbf{v}_{\mathsf{tx},n}$, where $\mathbf{v}_{\mathsf{tx},n} = \mathbf{G}\mathbf{\Delta}_n$ is the additive noise vector caused by the transmitter hardware impairment. Note that the actually transmitted power after the RF circuit is $(1+\delta^2)p_k^{\Upsilon}$.

Let $\mathcal{Q}_{\mathsf{b}}(\cdot)$ be the quantization function using b-bit ADCs. Then, the quantized version $\mathbf{y}_n^{\mathsf{q}} = \mathcal{Q}_{\mathsf{b}}(\mathbf{y}_n)$ of (1) is given as

$$\begin{aligned}\mathbf{y}_n^{\mathsf{q}} &= (1-\rho)\mathbf{y}_n + \mathbf{v}_{\mathsf{rx},n} \\ &= (1-\rho)(\mathbf{G}\overline{\mathbf{x}}_n + \mathbf{v}_{\mathsf{an},n} + \mathbf{v}_{\mathsf{tx},n}) + \mathbf{v}_{\mathsf{rx},n},\end{aligned} \tag{3}$$

[2]Note that δ appears in practical applications as the error vector magnitude (EVM), which is commonly used to measure the RF quality. For example, the 3rd Generation Partnership Project (3GPP) Long-Term Evolution (LTE) has EVM requirements in the range [0.08 0.175]. Note that the relationship between δ and EVM is defined as $\mathrm{EVM} = \sqrt{\mathbb{E}\|\mathbf{\Delta}_n\|^2/\mathbb{E}\|\overline{\mathbf{x}}_n\|^2} = \delta,\ \forall n$.

where $\mathbf{v}_{\mathsf{rx},n}$ denotes the additive quantization noise vector caused by the receiver hardware impairment, according to the additive quantization noise model (AQNM) [11], [12], and ρ denotes the inverse of the signal-to-quantization noise ratio (SQNR) given as[3]

$$\rho = \frac{\mathbb{E}[|[\mathbf{y}_n]_m - [\mathbf{y}_n^{\mathsf{q}}]_m|^2]}{\mathbb{E}[|[\mathbf{y}_n]_m|^2]}, \tag{4}$$

where values are given in [11] for $\mathsf{b} < 5$ and $\rho \approx \frac{\pi\sqrt{3}}{2}2^{-2\mathsf{b}}$ for $\mathsf{b} \geq 5$. Define $\mathbf{v}_{\mathsf{agg},n} = (1-\rho)\mathbf{v}_{\mathsf{an},n} + (1-\rho)\mathbf{v}_{\mathsf{tx},n} + \mathbf{v}_{\mathsf{rx},n}$ as the aggregated noise. Its covariance matrix, denoted as $\mathbf{C}_{\mathsf{agg},n} = \mathbb{E}[\mathbf{v}_{\mathsf{agg},n}\mathbf{v}_{\mathsf{agg},n}^H]$, can be written as

$$\mathbf{C}_{\mathsf{agg},n} = (1-\rho)^2\mathbf{C}_{\mathsf{an},n} + (1-\rho)^2\mathbf{C}_{\mathsf{tx},n} + \mathbf{C}_{\mathsf{rx},n}, \tag{5}$$

where

$$\begin{aligned}\mathbf{C}_{\mathsf{an},n} &= \mathbb{E}\left[\mathbf{v}_{\mathsf{an},n}\mathbf{v}_{\mathsf{an},n}^H\right] = \mathbf{I}_M, \\ \mathbf{C}_{\mathsf{tx},n} &= \mathbb{E}\left[\mathbf{v}_{\mathsf{tx},n}\mathbf{v}_{\mathsf{tx},n}^H\right] = \delta^2\mathbb{E}\left[\mathbf{G}\mathbf{P}_n\mathbf{G}^H\right], \\ \mathbf{C}_{\mathsf{rx},n} &= \mathbb{E}\left[\mathbf{v}_{\mathsf{rx},n}\mathbf{v}_{\mathsf{rx},n}^H\right] = (1-\rho)\,\rho\left((1+\delta^2)\mathbb{E}\left[\mathbf{G}\mathbf{P}_n\mathbf{G}^H\right] + \mathbf{I}_M\right),\end{aligned}$$

$\mathbf{P}_n = \mathbf{P}^{\mathrm{tr}}$ if $n \in [1, L]$ and $\mathbf{P}_n = \mathbf{P}^{\mathrm{dt}}$ if $n \in [L+1, N]$. Note that the detailed derivation of $\mathcal{C}_{\mathsf{agg},n}$ is shown in Appendix A. Although the covariance matrix of the aggregated noise can be easily estimated, its exact distribution is not easy to obtain in general. However, we may take a worst-case uncorrelated Gaussian noise argument, which can provide quite simple and well-approximated results. In the sequel, we assume $\mathbf{v}_{\mathsf{agg},n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathsf{agg},n})$ and is uncorrelated with the data symbol, i.e., $\mathbb{E}[\overline{\mathbf{x}}_n\mathbf{v}_{\mathsf{agg},n}^H] = \mathbf{0}$.

III. UPLINK PERFORMANCE ANALYSIS

A. Linear MMSE Channel Estimation

Let $\mathbf{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K]$ be the $L \times K$ pilot matrix with $\|\boldsymbol{\psi}_k\|^2 = 1$ for $\forall k$, in which orthogonal pilots can be used if $K \leq L$. Otherwise, if $K > L$, non-orthogonal pilots must be used. User j transmits the jth row of $\mathbf{\Psi}$ with a transmit power p_j^{tr}. In general, the received training signal powers from various users are quite different. In that case, a (cell-edge) user with a low received training signal power may obtain bad channel estimation accuracy and a (cell-center) user with a high received training signal power may obtain good channel estimation accuracy, which may result in too small achievable rate for the cell-edge user. To avoid such a problem, it is natural to equalize all users' received signal power, i.e., an uplink open-loop power control, which is also adopted in 3GGP LTE(-A). To equalize all users' channel estimation quality, the average received power of each user is set to $\overline{p^{\mathrm{tr}}}$ so that the transmit power of user j is set to $p_j^{\mathrm{tr}} = \overline{p^{\mathrm{tr}}}/\beta_j$ at the training phase and the received signal matrix during this phase, $\mathbf{Y}^{\mathsf{q},\mathrm{tr}} = [\mathbf{y}_1^{\mathsf{q}}, \ldots, \mathbf{y}_L^{\mathsf{q}}]$, can be written as

$$\mathbf{Y}^{\mathsf{q},\mathrm{tr}} = (1-\rho)\sqrt{L\overline{p^{\mathrm{tr}}}}\mathbf{H}\mathbf{\Psi}^H + \mathbf{V}_{\mathsf{agg}}^{\mathrm{tr}}, \tag{6}$$

[3]Note that the AQNM model results in some approximation error when the number of quantization levels becomes small. However, it is meaningful in a moderate range of quantization bits, i.e., $4 \sim 10$ bits.

where $\mathbf{V}^{\text{tr}}_{\mathsf{agg}} = [\mathbf{v}_{\mathsf{agg},1}, \dots, \mathbf{v}_{\mathsf{agg},L}]$ and the covariance matrix of $\mathbf{v}_{\mathsf{agg},n}$ for $n \in [1, L]$ is given as

$$\mathbf{C}^{\text{tr}}_{\mathsf{agg},n} = (1-\rho)\left(\left(\rho+\delta^2\right)K\overline{p^{\text{tr}}}+1\right)\mathbf{I}_M. \quad (7)$$

Let $\widetilde{\mathbf{h}}_k = \widehat{\mathbf{h}}_k - \mathbf{h}_k$ denote the CSI error vector of user k, where $\widehat{\mathbf{h}}_k$ is the output of the linear MMSE channel estimator for user k and $\sigma^2_{\text{ch},k} = \frac{1}{M}\mathbb{E}\|\widetilde{\mathbf{h}}\|_F^2$ denote the corresponding mean square error (MSE) of user k. Due to power equalization, every user has the same MSE, i.e., $\sigma^2_{\text{ch}} = \sigma^2_{\text{ch},k}$, for all k.[4] The following lemma informs the properties of the MMSE channel estimation with dirty-RFs and non-orthogonal pilots.

Lemma 1. *With transmitter hardware impairment and receiver hardware impairment, $\widehat{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{0}, (1-\sigma^2_{\text{ch}})\mathbf{I}_M)$ and $\widetilde{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2_{\text{ch}}\mathbf{I}_M)$ are mutually independent, where the MSE is expressed as*

$$\sigma^2_{\text{ch}} = \left(1-\frac{L}{K}\right)^+ + \frac{1}{K}\sum_{i=1}^{\min\{L,K\}}\left(1+\frac{(1-\rho)\,L\overline{p^{\text{tr}}}}{\left(\rho+\delta^2\right)K\overline{p^{\text{tr}}}+1}\lambda_i\right)^{-1} \quad (8)$$

where λ_i represents the ith eigenvalue of $\mathbf{\Psi}^H\mathbf{\Psi}$.

Proof: See the proof of Theorem 1 in [17]. ■

To minimize the MSE, we need to design pilot sequences as follows.

Lemma 2. *Optimal pilot sequences $\mathbf{\Psi}$ for minimizing σ^2_{ch} satisfy*

$$\mathbf{\Psi}\mathbf{\Psi}^H = \max\left\{1, \frac{K}{L}\right\}\mathbf{I}_L. \quad (9)$$

Proof: For $L \ge K$, it is straightforward to prove that the use of orthogonal pilots can minimize the MSE. For $L < K$, the MSE in (8) can be simplified as

$$\begin{aligned}\sigma^2_{\text{ch}} &= 1-\frac{L}{K}+\frac{1}{K}\sum_{i=1}^{L}\frac{\left(\rho+\delta^2\right)K\overline{p^{\text{tr}}}+1}{\left(\rho+\delta^2\right)K\overline{p^{\text{tr}}}+1+L\left(1-\rho\right)\overline{p^{\text{tr}}}\lambda_i}\\ &= 1+\frac{1}{K}\sum_{i=1}^{L}\left(\frac{\left(\rho+\delta^2\right)K\overline{p^{\text{tr}}}+1}{\left(\rho+\delta^2\right)K\overline{p^{\text{tr}}}+1+L\left(1-\rho\right)\overline{p^{\text{tr}}}\lambda_i}-1\right)\\ &= 1-\frac{1}{K}\sum_{i=1}^{L}\frac{L\left(1-\rho\right)\overline{p^{\text{tr}}}\lambda_i}{\left(\rho+\delta^2\right)K\overline{p^{\text{tr}}}+1+L\left(1-\rho\right)\overline{p^{\text{tr}}}\lambda_i}.\end{aligned}$$

So, the minimization of σ_{ch} is equivalent to the maximization of $\sum_{i=1}^{L}\frac{L(1-\rho)\overline{p^{\text{tr}}}\lambda_i}{(\rho+\delta^2)K\overline{p^{\text{tr}}}+1+L(1-\rho)\overline{p^{\text{tr}}}\lambda_i}$ under $\sum_{i=1}^{L}\lambda_i = K$, which is obtained when the eigenvalues are the same, i.e., $\lambda_i = K/L$. ■

Note that the above condition is identical to the condition minimizing the channel estimation error in [20]. In fact, this condition is known as the Welch bound equality (WBE) [21]. From Lemmas 1 and 2, we can calculate that the MSE is minimized at

$$\sigma^2_{\text{ch}} = 1-\frac{(1-\rho)\,L\overline{p^{\text{tr}}}}{(1-\rho)\,(L-K)^+\overline{p^{\text{tr}}}+(1+\delta^2)\,K\overline{p^{\text{tr}}}+1}. \quad (10)$$

[4]The appiled MMSE channel estimator is a conventional MMSE channel estimator by regarding the aggregated noise as the AWGN noise but its covariance matrix $\mathbf{C}^{\text{tr}}_{\mathsf{agg},n}$ in (7).

When the pilot sequences are under-utilized (orthogonal pilots are used), i.e., $K \le L$, σ^2_{ch} is simplified into

$$\begin{aligned}\sigma^2_{\text{ch}} &= \frac{1}{\frac{(1-\rho)L\overline{p^{\text{tr}}}}{(\rho+\delta^2)K\overline{p^{\text{tr}}}+1}+1}\\ &\to \frac{1}{\frac{1-\rho}{\delta^2+\rho}\frac{L}{K}+1}, \text{ as } \overline{p^{\text{tr}}} \to \infty,\end{aligned}$$

which implies that even if an extremely high training power is used, accurate CSI is not achievable at the BS due to both transmitter hardware impairment and receiver hardware impairment. In the sequel, we assume that optimal pilot sequences are employed.

Remark 1 (Optimal pilot sequences)**.** *Obviously, if $K \le L$, optimal pilot sequences can be obtained from K arbitrarily chosen columns of an $L \times L$ unitary matrix. In the case of $K > L$, one option is to over-sample an $L \times L$ unitary matrix. Although there are infinitely many sequences that hold the WBE, one simple example is the discrete Fourier transform (DFT) based sequences, obtained as*

$$\mathbf{\Psi} = \frac{1}{\sqrt{L}}\begin{bmatrix} 1 & e^{-\mathrm{j}\frac{2\pi f_1}{K}} & \dots & e^{-\mathrm{j}\frac{2\pi f_1(K-1)}{K}} \\ 1 & e^{-\mathrm{j}\frac{2\pi f_2}{K}} & \dots & e^{-\mathrm{j}\frac{2\pi f_2(K-1)}{K}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-\mathrm{j}\frac{2\pi f_L}{K}} & \dots & e^{-\mathrm{j}\frac{2\pi f_L(K-1)}{K}} \end{bmatrix},$$

where $\mathrm{j} = \sqrt{-1}$ and $0 < f_1 < f_2 < \dots < f_L < K$ are arbitrarily chosen integers. Note that the DFT-based sequences are widely used in the designs of unitary-space time modulation [22] and the feedback codebook [23].

B. Uplink Average Rate

During the data transmission phase, the received signal vector at the nth symbol interval, $n \in [L+1, N]$, is given by

$$\mathbf{y}^{\mathsf{q}}_n = (1-\rho)\sum_{j=1}^{K}\sqrt{p^{\text{dt}}_j\beta_j}\mathbf{h}_j s_{jn} + \mathbf{v}_{\mathsf{agg},n}, \quad (11)$$

where $s_{jn} \sim \mathcal{CN}(0,1)$ is the information-bearing data symbol for user j at the nth symbol interval. For the given estimated channel matrix $\widehat{\mathbf{G}} = \widehat{\mathbf{H}}\mathbf{D}^{1/2}$, the covariance matrix of $\mathbf{v}_{\mathsf{agg},n}$ is given as

$$\begin{aligned}\mathbf{C}_{\mathsf{agg},n} = &(1-\rho)\left(1+\left(\rho+\delta^2\right)\sigma^2_{\text{ch}}\sum_{j=1}^{K}p^{\text{dt}}_j\beta_j\right)\mathbf{I}_M \\ &+(1-\rho)^2\delta^2\mathbf{\Phi}+\rho\left(1-\rho\right)\left(1+\delta^2\right)\overline{\mathbf{\Phi}},\end{aligned} \quad (12)$$

where $\mathbf{\Phi} = \widehat{\mathbf{G}}\mathbf{P}^{\text{dt}}\widehat{\mathbf{G}}^H$ and $\overline{\mathbf{\Phi}}$ is the diagonal matrix of $\mathbf{\Phi}$. Using the maximum ratio combining (MRC) filter $\mathbf{F} = \widehat{\mathbf{G}}$,[5] the average user rate of user k is given by $R_k = (1-$

[5]Here, we deal with a very simple MRC receiver first, and we show that it becomes optimal in meaningful situations. See Remark 3.

$L/N)\mathbb{E}[\log_2(1+\widetilde{\gamma}_k)]$, where $\widetilde{\gamma}_k$ denotes the received signal-to-interference-plus-noise ratio (SINR), given by

$$\widetilde{\gamma}_k = \frac{p_k^{\mathrm{dt}}\beta_k\|\widehat{\mathbf{h}}_k\|^4}{\sum\limits_{i\neq k} p_i^{\mathrm{dt}}\beta_i|\widehat{\mathbf{h}}_k^H\widehat{\mathbf{h}}_i|^2 + \sigma_{\mathrm{ch}}^2\|\widehat{\mathbf{h}}_k\|^2\sum\limits_{j=1}^{K} p_j^{\mathrm{dt}}\beta_j + \frac{\widehat{\mathbf{h}}_k^H\mathbf{C}_{\mathrm{agg},n}^{\mathrm{dt}}\widehat{\mathbf{h}}_k}{(1-\rho)^2}}, \tag{13}$$

and the average sum rate is denoted by $R_{\mathrm{sum}} = \sum_{k=1}^{K} R_k$. The following theorem presents an approximated version of the average user rate when dirty-RFs are used.

Theorem 1. *With transmitter hardware impairment and receiver hardware impairment, the average user rate of user k is well-approximated by $R_k \approx (1-L/N)\log_2(1+\gamma_k)$, where*

$$\gamma_k = \frac{(1-\rho)(M+1)\,p_k^{\mathrm{dt}}\beta_k}{\Sigma + ((M\delta^2-1)(1-\rho)+\rho(1+\delta^2))\,p_k^{\mathrm{dt}}\beta_k} \tag{14}$$

and $\Sigma = \frac{1}{1-\sigma_{\mathrm{ch}}^2}\left(\left(1+\delta^2\right)\sum_{j=1}^{K} p_j^{\mathrm{dt}}\beta_j + 1\right)$.

Proof: See Appendix B. ■

Let $S_k = \left(1-\sigma_{\mathrm{ch}}^2\right)p_k^{\mathrm{dt}}\beta_k$ and $I_k = \sum_{j=1,j\neq k}^{K} p_j^{\mathrm{dt}}\beta_j + \sigma_{\mathrm{ch}}^2 p_k^{\mathrm{dt}}\beta_k$ denote the desired signal power (without array gain, M) and the interference power, respectively. Then, the effective SINR, γ_k, is found as follows.

- Let $\rho = \delta = 0$ (without both transmitter hardware impairment and receiver hardware impairment). Then, γ_k is simplified as
$$\gamma_k = \frac{(M+1)\,S_k}{1+I_k},$$
where the desired signal power increases according to the channel estimation quality, $1-\sigma_{\mathrm{ch}}^2$, the number of BS antennas, M, the transmit power, p_k^{dt}, and the pathloss β_k, and the interference is composed of other-user interference, $\sum_{j=1,j\neq k}^{K} p_j^{\mathrm{dt}}\beta_j$, and the self-interference is caused by the channel mismatch, $\sigma_{\mathrm{ch}}^2 p_k^{\mathrm{dt}}\beta_k$.
- Let $\rho = 0$ and $\delta > 0$ (transmitter hardware impairment only). Then, γ_k is simplified as
$$\gamma_k = \frac{(M+1)\,S_k}{1+(1+\delta^2)\,I_k + (M+1)\,S_k\delta^2}.$$
Since the total transmit power increases according to the transmitter hardware impairment, the interference power also increases by a factor of $1+\delta^2$. However, the desired signal power does not change and the power increase due to the transmitter hardware impairment, $(M+1)S_k\delta^2$, is additional interference.
- Let $\delta = 0$ and $\rho > 0$ (receiver hardware impairment only). Then, γ_k is simplified as
$$\gamma_k = \frac{(1-\rho)(M+1)\,S_k}{1+I_k+2\rho S_k}.$$
Obviously, the receiver hardware impairment reduces the desired signal power by a factor of $1-\rho$, and interference caused by the receiver hardware impairment, $2\rho S_k$, is additionally generated.

Remark 2. *Theorem 1 can be considered as a generalization of the conventional results in literature. When $\sigma_{\mathrm{ch}}^2 = 0$ and $\delta = 0$ (perfect CSI and no transmitter hardware impairment), (14) reduces to Theorem 1 in [11] and (16) in [12]. When $\rho = 0$ (no receiver hardware impairment), (14) is a simplified version of (24) in [8]. Note that (24) in [8] is also an approximated equation that includes a hypergeometric function and (14) is more insightful. Also note that the average user rate derived in [12] does not match Theorem 1, even for the case of $\delta = 0$, because [12] does not consider the quantization noise in the training phase.*

IV. Connectivity Maximization

For the optimal use of a training-based large-scale antenna system with dirty-RFs for massive connectivity, non-orthogonal pilots need to be employed so that the connectivity does not suffer from the limited number of orthogonal pilots. In addition, the system parameters, such as the number of dedicated training symbols L and the number of simultaneously served users K, need to be selected carefully. Because the connectivity is proportional to the average sum rate, we attempt to maximize the average sum rate.[6] The optimization problem for average sum rate maximization can be formulated as follows.

$$\text{(P1)}\ \max_{(L,K)}\ \left(1-\frac{L}{N}\right)\sum_{k=1}^{K}\log_2\left(1+\gamma_k\right),\ \text{subject to} \tag{15}$$

$$\begin{cases} 1 \le K = L < N, & \text{if orthogonal pilots,} \\ L < K,\ 1 \le L < N, & \text{if non-orthogonal pilots.} \end{cases} \tag{16}$$

In the sequel, the channel-inversion power control is assumed, i.e., $p_j^{\mathrm{tr}}\beta_j = p_j^{\mathrm{dt}}\beta_j = \overline{p}$ for $\forall j$, where $\overline{p}$ is the target received signal power so that all of users have the same quality of received SINR, which is given at the bottom of the next page.

A. Closed-form Optimal Solution

Theorem 2. *For given values of the number of BS antennas M and the target received signal power $\overline{p}$, the optimal system parameters to maximize the connectivity are*

$$(L^\star, K^\star) = \begin{cases} \left(\lceil\varphi\rfloor, \left\lceil\frac{N-\varphi}{(3\varphi-N)\,C}\right\rfloor\right), & \text{if } \exists\varphi\in\left(\frac{N}{3},\frac{\theta N}{3}\right), \\ \left(\lceil\tau\rfloor, \lceil\tau\rfloor\right), & \text{otherwise}. \end{cases} \tag{18}$$

Here, τ is the unique positive root of $f_{A,B,C}(x) = 0$ and φ is the unique positive root of $g_{A,B}(x) = 0$, where

$$f_{A,B,C}(x) = (N-2x)\log\left(1+\frac{Ax}{Bx+(Cx+1)^2}\right) - \frac{Ax}{Bx+(Cx+1)^2}\frac{(N-x)(Cx-1)(Cx+1)}{(A+B)\,x+(Cx+1)^2}, \tag{19}$$

[6] Typically, MTC or IoT devices require a low minimum rate with no latency constraint. Thus, the best way to serve these devices is to maximize the average sum rate and schedule them in a best-effort manner. Connectivity maximization with a latency constraint is outside the scope of this paper and will be considered in future study.

and

$$g_{A,B}(x) = \log\left(1 + \frac{A(3x-N)^2}{4x + B(3x-N)^2}\right) - \frac{A(3x-N)^2}{4x + B(3x-N)^2}\frac{4(N-x)}{4x + (A+B)(3x-N)^2}, \tag{20}$$

with

$$A = (1-\rho)^2 (M+1)\overline{p}^2,$$
$$B = (1-\rho)\left(\left(M\delta^2 - 1\right)(1-\rho) + \rho\left(1+\delta^2\right)\right)\overline{p}^2,$$
$$C = \left(1+\delta^2\right)\overline{p},$$

and

$$\theta = \frac{1}{2} + \frac{\sqrt{(1+\delta^2)^2\overline{p}^2N^2 + 10(1+\delta^2)\overline{p}N + 1} - 1}{2(1+\delta^2)\overline{p}N}. \tag{21}$$

Proof: See Appendix C. ■

Theorem 2 provides a simple close-form solution for optimal system parameters. For example, let $M = 128, N = 100, \delta = \rho = 0.1$. From the definitions of A, B, C, and θ, we have $A = 10449, B = 10.1, C = 31.77$ and $\theta = 1.002$. After some computations, we obtain $\varphi \approx 33.4066$ which is not included in $(33.33, 33.40)$, which means that the use of orthogonal pilots is optimal. And we obtain $\tau = 32.17$ so that the optimal training overhead and users are 32. As shown in the example, Theorem 2 provides a simple criterion to determine whether it is beneficial to use non-orthogonal pilots or not according to the hardware impairment and other parameters. However, it is hard to gain an intuitive insignt due to the complicated equations $g_{A,B(x)}$ and $f_{A,B,C}(x)$. So, we present three interesting asymoptotic results of this theorem as follows:

- *Hardware impairment*: As δ increases, then θ approaches 1, which implies that to use orthogonal pilots only is sufficient when users have extremely low-quality RFs. On the other hand, as ρ increases up to 1 (i.e., the SQNR decreases), φ approaches $N/2$ so that if $\left(1+\delta^2\right)\overline{p}N \leq 1$, using non-orthogonal pilots is beneficial and if not, orthogonal pilots are sufficient. Thus, employing non-orthogonal pilots is better in general but can be useless under extremely high transmitter hardware impairment.
- *Frame Size*: As N increases, then θ approaches 1, which implies that using only orthogonal pilots is sufficient. However, since N is limited due to the finite channel coherence time, employing non-orthogonal pilots is preferred in practice.
- *Target Received Signal Power*: As $\overline{p}$ increases, then θ approaches 1, which implies that to use orthogonal pilots only is sufficient. However, since $\overline{p}$ is limited due to the limited transmit power, employing non-orthogonal pilots is beneficial in practice.

TABLE II
RESULTS OF THEOREM 3 FOR ORTHOGONAL PILOTS

$0 \leq \alpha < 1/2$	$(L^\star, K^\star)$	$\left(\left\lceil \frac{N}{2} \right\rfloor, \left\lceil \frac{N}{2} \right\rfloor\right),$
	User Rate	$\frac{1}{2}\log_2\left(1 + \frac{1}{\delta^2}\right)$
$\alpha = 1/2$	$(L^\star, K^\star)$	$(\lceil \mu \rfloor, \lceil \mu \rfloor)$
	User Rate	$\left(1 - \frac{\mu}{N}\right)\log_2\left(1 + \frac{\mu(1-\rho)^2P^2}{1+\mu\delta^2(1-\rho)^2P^2}\right)$
$\alpha > 1/2$	$(L^\star, K^\star)$	$\left(\left\lceil \frac{2N}{3} \right\rfloor, \left\lceil \frac{2N}{3} \right\rfloor\right)$
	User Rate	$\frac{2N}{9\log 2}(1-\rho)^2P^2M^{1-2\alpha}$

It is worth noting that the condition $\exists\varphi \in (N/3, \theta N/3)$ is equivalent to $g_{A,B}(\theta N/3) > 0$ because

$$\begin{aligned} g_{A,B}\left(\frac{N+\epsilon}{3}\right) &= \log\left(1 + \frac{3A\epsilon^2}{4(N+\epsilon) + 3B\epsilon^2}\right) \\ &\quad - \frac{3A\epsilon^2}{4(N+\epsilon)+3B\epsilon^2}\frac{4(2N-\epsilon)}{4(N+\epsilon) + 3(A+B)\epsilon^2} \\ &= -\frac{3A\varepsilon^2}{4N} + o(\epsilon^2) < 0 \end{aligned}$$

for an arbitrarily small positive value of ϵ. So, we can easily determine whether to use non-orthogonal pilots by checking the sign of $g_{A,B}(\theta N/3)$ without having to compute the real roots of $g_{A,B}(x) = 0$.

B. Asymptotic Analysis

Note that the average user rate is saturated, even for the case of using infinitely high transmit power due to the hardware impairment and the required target received signal power for a given available rate can be reduced in general as the number of BS antennas increases. Thus, the natural inclination is to investigate how the transmit power can be saved as the number of BS antennas increases. In order to show the power saving effect, we construct the relationship between the target received signal power and the number of BS antennas as follows: $\overline{p} = P/M^\alpha$, where $\alpha \geq 0$ and P is a given positive value. Under this power saving model, the taret received signal power is constant when $\alpha = 0$ or is inversely-proportional to the number of BS antennas. The following theorem shows the results of this power saving model.

Theorem 3. *Suppose that the target received signal power is set to $\overline{p} = P/M^\alpha$ for a given value of P. As $M \to \infty$, the optimal system parameters to maximize connectivity are as given in Tables I and II, where*

- *μ is the unique positive root of the function $f_{(1-\rho)^2P^2,\delta^2(1-\rho)^2P^2,0}(x) = 0$,*
- *$\chi \in (N/3, \infty)$ is the unique root of the function $g_{(1-\rho)^2P^2,\delta^2(1-\rho)^2P^2}(x) = 0$, and*
- *ϖ is the unique positive root of the function $h(x) = (1+x)\log(1+x) - 2x\left(1-\delta^2x\right) = 0$.*

Proof: See Appendix D. ■

$$\gamma_k = \frac{(1-\rho)^2 L(M+1)\overline{p}^2}{\left(\left((1-\rho)(L-K)^+ + (1+\delta^2)K\right)\overline{p} + 1\right)\left((1+\delta^2)K\overline{p} + 1\right) + (1-\rho)L\left((M\delta^2-1)(1-\rho) + \rho(1+\delta^2)\right)\overline{p}^2} \tag{17}$$

Remark 3 (Optimal Frame Configuration and Connectivity)**.** *When only orthogonal pilots are allowed:*

- *If the power saving is loose ($0 \leq \alpha < 1/2$), the optimal frame configuration is to allocate $N/2$ symbols to the training phase and to serve $N/2$ users simultaneously. In this case, the connectivity is saturated at $\Theta(1)$ and becomes independent of the receive hardware impairment because loose power saving can compensate for the effects of receiver hardware impairment with the cost of more power consumption.*
- *If the power saving is strict ($\alpha > 1/2$), the optimal frame configuration is to allocate $2N/3$ symbols to the training phase and to serve $2N/3$ users simultaneously. In this case, the connectivity diminishes along with $\Theta(M^{1-2\alpha})$ (i.e., such strict power saving cannot be actually achieved).*
- *If the power saving is moderate ($\alpha = 1/2$), the optimal frame configuration is to allocate μ symbols to the training phase and to serve μ users simultaneously for $N/2 \leq \mu \leq 2N/3$. In this case, the connectivity is saturated and depends on both the transmitter hardware impairment and the receiver hardware impairment.*

Clearly, when only orthogonal pilots are allowed, the connectivity is limited by the number of orthogonal pilots. However, the use of non-orthogonal pilots can break this constraint, as described in the following.

- *If the power saving is loose ($0 \leq \alpha < 1/2$), the optimal training length is $N/3$ but the optimal number of users served simultaneously is $\Theta(\sqrt{M})$. In this case, the connectivity becomes $\Theta(\sqrt{M})$ and depends on both the transmitter hardware impairment and the receiver hardware impairment.*
- *If the power saving is moderate or strict ($\alpha \geq 1/2$), the optimal training length is $N/2$ but the optimal number of users served simultaneously is $\Theta(M^{\alpha})$. In this case, the connectivity becomes $\Theta(M^{1-\alpha})$, which indicates that stricter power saving (or higher α) can be realized with a meaningful tradeoff. Note that the total energy consumption of the whole network becomes independent of M.*

Thus, we can conclude that ***the use of non-orthogonal pilots can increase the connectivity from*** $\Theta(1)$ ***to*** $\Theta(M^{1/2-(\alpha-1/2)^+})$ ***while keeping the total energy consumption constant****. Note that when allowing non-orthogoanl pilots at least $\Theta(\sqrt{M})$ users can be simultantenously served to achieve maximum connectivity.*

Remark 4 (Asymptotic Optimality of MRC receiver)**.** *The simple MRC receiver scheme is widely considered as typical reception scheme for an large-scale antenna system due to its low-complexity. Since the MRC receiver can be asymptotically optimal when the number of BS antennas is far larger than the number of users served simultaneously (i.e., $M \gg K$), our analysis reveals that the MRC receiver is asymptotically optimal for meaningful situations (i.e., $0 \leq \alpha < 1$, where the sum rate is not diminished).*

V. Energy Efficiency Maximization

TABLE III
RESULTS OF THEOREM 3 FOR NON-ORTHOGONAL PILOTS

$0 \leq \alpha < 1/2$	$(L^\star, K^\star)$	$\left(\left\lceil \frac{N}{3} \right\rfloor, \left\lceil \frac{1-\rho}{1+\delta^2} \sqrt{\frac{\varpi^{-1}-\delta^2}{3}} NM \right\rfloor\right)$
	User Rate	$\frac{2}{3}\log_2(1+\varpi)$
$\alpha = 1/2$	$(L^\star, K^\star)$	$\left(\lceil \chi \rfloor, \left\lceil \frac{1}{(1+\delta^2)P} \frac{N-\chi}{(3\chi-N)} \sqrt{M} \right\rfloor\right)$,
	User Rate	$(1-\frac{\chi}{N})\log_2\left(1+\frac{(3\chi-N)^2}{4\chi+\delta^2(1-\rho)^2P^2(3\chi-N)^2}\right)$
$\alpha > 1/2$	$(L^\star, K^\star)$	$\left(\left\lceil \frac{N}{2} \right\rfloor, \left\lceil \frac{\sqrt{2(1+\delta^2)(1-\rho)P^2+1}-1}{2(1+\delta^2)P} M^{\alpha} \right\rfloor\right)$,
	User Rate	$\frac{N}{2\log 2} \frac{(1-\rho)^2P^2}{(1+\delta^2)(1-\rho)P^2+1+\sqrt{2(1+\delta^2)(1-\rho)P^2+1}} M^{1-2\alpha}$

The energy efficiency is another important metric for both MTC and IoT devices. The energy efficiency (bits/Joule) is defiend as $\sum_{k=1}^{K} R_k / \sum_{k=1}^{K} E_k$, where $R_k = (1 - L/N)\log_2(1+\gamma_k)$ denotes the user rate shown in Theorem 1 and (17) and

$$
\begin{aligned}
E_k &= \frac{L}{N}(1+\delta^2)p_k^{\mathrm{tr}} + (1-\frac{L}{N})(1+\delta^2)p_k^{\mathrm{dt}} \\
&= (1+\delta^2)\overline{p}/\beta_k
\end{aligned}
$$

denotes the energy consumed by user k in transmitting one symbol. Recall that $\overline{p} = p_k^{\mathrm{tr}}\beta_k = p_k^{\mathrm{dt}}\beta_k$ is the target received signal power. In this paper, we ignore a circuit energy consumption because the consumed circuit energy is independent to the optimization parameters L, K and $\overline{p}$. However, when users are equipped with multiple antennas, an appropriate circuit energy model should be further taken into account, which is out of scope of this paper and a future work.

Similarly as in the connectivity maximization problem (P1), the energy efficiency optimization problem can be formulated as

$$
\boxed{
\begin{aligned}
&(\mathrm{P2}) \max_{(L,K),\overline{p}} \frac{\left(1-\frac{L}{N}\right)\sum\limits_{k=1}^{K}\log_2\left(1+\gamma_k\right)}{\sum\limits_{k=1}^{K}\left(1+\delta^2\right)\overline{p}\beta_k^{-1}}, \text{ subject to} \quad (22) \\
&\begin{cases} 1 \leq K \leq L < N, & \text{if orthogonal pilots,} \\ L < K,\ 1 \leq L < N, & \text{if non-orthogonal pilots.} \end{cases} \quad (23)
\end{aligned}
}
$$

In contrast to the connectivity maximization problem (P1), the energy efficiency optimization problem requires additional joint optimization of the target received signal power $\overline{p}$ and it allows $K \leq L$ when orthogonal pilots are used. To solve P2, it is decomposed into the two sub-problems: one is solving P2 at a given value of $\overline{p}$ (P2$'$) and the other is solving P2 at a given values of (L, K) (P2$''$). Then, *near-optimal* values of both (L, K) and $\overline{p}$ can be obtained by solving the two sub-problems iteratively as follows.

First, suppose that value of $\overline{p}$ is given. Because energy efficiency decreases monotonically as K increases, we can easily prove that $K^\star = 1$ to maximize the energy efficiency while minimizing the connectivity, which implies that the use of a single-user transmission (i.e., only orthogonal pilots) is sufficient from the perspective of the energy efficiency.[7] Then, the energy efficiency optimization problem for a given value of $\overline{p}$ can be simplified as

$$(\text{P2}') \max_{1\leq L\leq N}\left(1-\frac{L}{N}\right)\log\left(1+\frac{AL}{E+FL}\right),$$

where $E = (C+1)\left(\left(\delta^2-\rho\right)\overline{p}+1\right)$ and $F = B + (1-\rho)\overline{p}(C+1)$. To find $L^\star$, we use the fact $\frac{d}{dL}\left(1-\frac{L}{N}\right)\log\left(1+\frac{AL}{E+FL}\right)|_{L=L^\star} = 0$ and we obtain the following relation

$$\log\left(1+\frac{AL^\star}{E+FL^\star}\right)=\frac{(N-L^\star)\,AE}{(E+FL^\star)\,(E+(A+F)\,L^\star)}. \tag{24}$$

Note that the near-optimal training length to maximize the energy efficiency becomes $L^\star = N/2$ for $\overline{p} \ll 1$ and $L^\star = 1$ for $\overline{p} \gg 1$.

Now, we consider the energy efficiency maximization problem (P2$''$) at given values of (L, K) which is not necessarily optimal and can be simplified as

$$(\text{P2}'') \max_{\overline{p}>0}\overline{p}^{-1}\log(1+\gamma_k(\overline{p})),$$

where $\gamma_k(\overline{p}) = \gamma_k$ is given in (18). Let $\overline{p}^\star$ be the solution of (P2$''$) for given values of (L, K). To find $\overline{p}^\star$, we use the fact that $\frac{d}{d\overline{p}}\overline{p}^{-1}\log(1+\gamma_k(\overline{p}))|_{\overline{p}=\overline{p}^\star} = 0$, and obtain the following relation:

$$\overline{p}^\star\gamma_k'\left(\overline{p}^\star\right)=\left(1+\gamma_k\left(\overline{p}^\star\right)\right)\log\left(1+\gamma_k\left(\overline{p}^\star\right)\right), \tag{25}$$

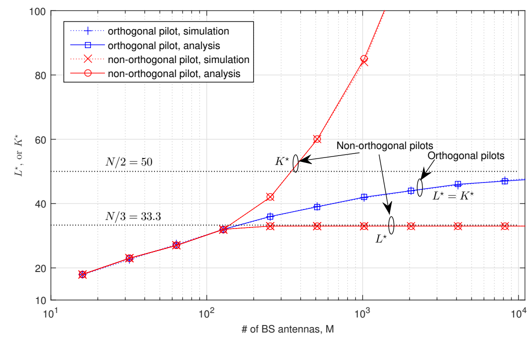
where $\gamma_k'(\overline{p}^\star) = \frac{d}{d\overline{p}}\gamma_k(\overline{p})|_{\overline{p}=\overline{p}^\star}$. Then, the near-optimal solution of (P2) can be obtained by solving (25) and (26) iteratively.

Remark 5 (Optimal value of $\overline{p}$)**.** *Since the energy efficiency is generally maximized at a small value of $\overline{p}^\star$ and a very small value of γ_k in general, we approximate $\log\left(1+\gamma_k\left(\overline{p}^\star\right)\right) \approx \gamma_k\left(\overline{p}^\star\right)$ so that we can obtain an approximated version of $\overline{p}^\star$ as*

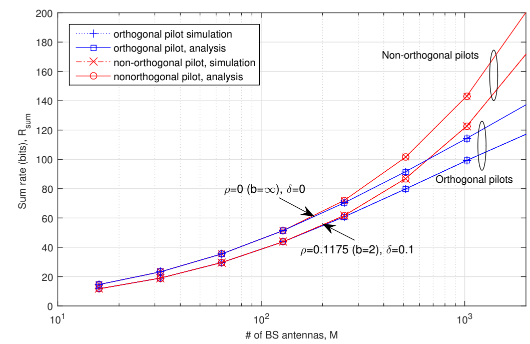
$$\overline{p}^\star\approx\left(\left((1-\rho)\,M+\rho\right)\left(1+\delta^2\right)L+\left(\left(1+\delta^2\right)K\right)^2\right)^{-1/2}, \tag{26}$$

which is an increasing function of ρ but is a decreasing function of δ. Thus, higher value of $\overline{p}$ is preferred when lower resolution ADCs are employed, and a lower value of $\overline{p}$ is preferred when less accurate transmit RF is utilized in order to improve the energy efficiency. Furthermore, for given values of (L, K), ***the near-optimal target received signal power is given asymptotically as $\overline{p}^\star = \Theta(1/\sqrt{M})$ from the perspective of energy efficiency maximization****.*

[7]Here, we show that single-user transmission is optimal in the energy efficiency sense. When the computational power of BS is taken into account, we conclude that this result is still valid. To capture the difference of the computation power (also the computational complexity) between the orthogonal pilot scheme and the non-orthogonal pilot scheme, we need to consider the MMSE channel estimation step. To compute the MMSE channel estimator, the pseudo-inverse operation is required and its computational complexity depends only on the size of the wireless channel matrix. As a result, we can obtain the computational complexity of $O(K^t)$ per a user for some small $t > 0$. It implies the computation power for the non-orthogonal pilot scheme is larger than that for non-orthogonal pilot scheme.



(a) Optimal System Parameters with $\rho = 0.1175(\mathrm{b}=2)$, $\delta = 0.1$, and $p = 10$

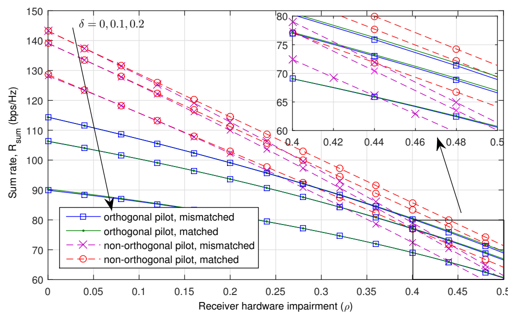


(b) Sum Rate

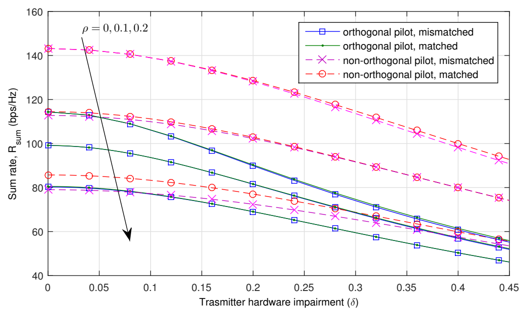
Fig. 2. Optimal system parameters and corresponding sum rate as a function of M.

VI. Simulation Results

Fig. 2 depicts both the optimal system parameters and the corresponding sum rate as functions of M when $N = 100$, $\mathrm{b} = 2$ ($\rho = 0.1175$), $\delta = 0.1$, and $p = 10$dB. The analytical results are obtained from Theorems 1 and 2, which are shown to be nearly the same as those obtained from simulations for a wide range of M. From the results, it is shown that using only orthogonal pilots is sufficient for a not-so-large value of M ($M \leq 128$) but the use of non-orthogonal pilots becomes critical and provides a non-negligible gain as M increases ($M \geq 256$). For example, the optimal frame configuration of using only orthogonal pilots is $(L^\star, K^\star) = (42, 42)$ and $R_{\text{sum}} = 99.27$ at $M = 1024$, but the optimal frame configuration when non-orthogonal pilots are employed is $(L^\star, K^\star) = (33, 85)$ at $M = 1024$ and $R_{\text{sum}} = 122.7$ so that the use of non-orthogonal pilots can provide 23.6% higher connectivity. In addition, such an improvement in the connectivity becomes more significant as M increases further as shown in Fig. 2(b), which confirms the asymptotic analysis. Note that about hundreds of bits are too high value to be achieved in a real-network. In fact, such a large bits can be achieved by the use of extremely many BS antennas, which may be impractically large in the common sense of network deployments. Consequently, the use of non-orthogonal pilots can dramatically increase the connectivity in an large-scale antenna system, even one with dirty-RFs at both transmitter and receiver. Additionally, we can observe the trend that $L^\star$ converges to $\lceil N/2 \rfloor = 50$ for the case of using orthogonal pilots only or $L^\star$ converges to $\lceil N/3 \rfloor = 33$ when non-

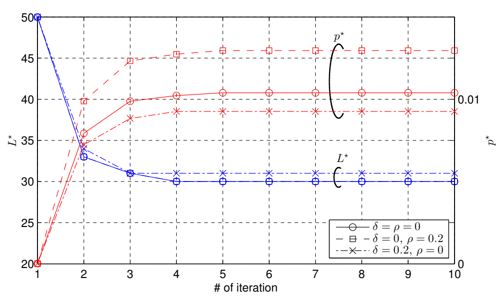


(a) Receiver hardware impairment (ρ)

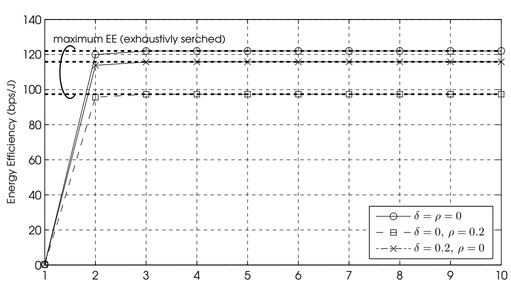


(b) Transmitter hardware impairment (δ)

Fig. 3. Sum rate with matched/mismatched optimization as a function of ρ or δ with $M = 1024$ and $p = 10$dB.



(b) Optimal System Parameters



(a) Energy Efficiency

Fig. 4. Optimal energy efficiency and corresponding system parameters with $M = 1024$ and the initial value of p at 0.0001.

orthogonal pilots are employed as the value of M becomes large, which is consistent with the asymptotic analysis.

Fig. 3 shows the effects of the hardware impairment ρ and δ on the optimal system parameters when $M = 1024$ and $\overline{p} = 10$dB. Here, the term 'matched' denotes the use of optimized values of $(L^{\star}, K^{\star})$ for the particular hardware impairment, and the term 'mismatched' denotes the use of optimized values of $(L^{\star}, K^{\star})$ without considering the hardware impairment (i.e., assuming that $\rho = \delta = 0$). From Fig. 3, it is shown that the sum rate achieved by both allowing non-orthogonal pilots and using the matched optimization increase by 25.2% at $(\rho, \delta) = (0, 0)$, 15.5% at $(\rho, \delta) = (0.2, 0)$, 43.0% at $(\rho, \delta) = (0, 0.2)$ or 9.8% at $(\rho, \delta) = (0.2, 0.2)$ over those achieved when using only orthogonal pilots with mismatched optimization, respectively. In fact, the matched optimization improvement is negligible when only orthogonal pilots are allowed. However, it becomes more significant as non-orthogonal pilots are allowed for better connectivity. Note that the use of non-orthogonal pilots without also using matched optimization can become unfavorable to the use of only orthogonal pilots at high values of ρ (e.g., $\rho > 0.3$ in Fig. 3(a)). Thus, we conclude properly estimating the hardware impairment with matched optimization is important for massive connectivity.

Fig. 4 shows both the optimal energy efficiency and the corresponding optimal system parameters according to the numbers of iterations starting with an initial value of p at 0.0001 when $M = 1024$. Clearly, the iterative solution in Section IV provides a near-optimal energy efficiency after only a few iterations for wide range of ρ and δ. In addition, as discussed in Remark 4, simulation results confirm that a higher ρ value results in a higher value of $p^{\star}$, but a higher δ value results in a lower value of $p^{\star}$. In addition, it is shown that the optimal training length is nearly $N/3 = 33$ in the presence of the transmitter hardware impairment, which is also consistent with the asymptotic analysis.

Finally, Fig. 5 shows the connectivity-energy efficiency tradeoff (achievable region) when $M = 1024$. The line for the connectivity maximization is obtained by solving (P1) for different values of p and the line for the energy efficiency point is obtained by solving (P2$'$) for different values of p. The line marked out with (+) symbols between the two points is the optimal tradeoff curve between the connectivity and the energy efficiency, which is obtained by solving the following problem using an exhaustive search

$$\max_{(L,K),p} \ \mathsf{EE}, \ \text{subject to } R_{\Sigma} \geq \text{const}.$$

Fig. 5 shows that the use of non-orthogonal pilots can enlarge the optimal connectivity-energy efficiency tradeoff and that the gain achieved by using non-orthogonal pilots as well as the matched optimization by considering hardware impairment is significant over a wide range of energy efficiency (in the low-to-mid energy efficiency region). The results demonstrate that the use of non-orthogonal pilots in a training-based large-scale antenna system with the matched optimization can provide not only high connectivity but also high energy efficiency in general, even when dirty-RFs are used.

VII. Concluding Remark

In this paper, the connectivity in an uplink training-based large-scale antenna system employing non-orthogonal pilots and dirty-RFs was derived on both transmitter and receiver sides. An accurate and simple approximation of the average

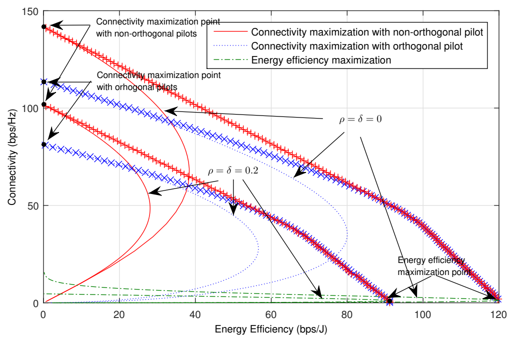


Fig. 5. Connectivity-Energy Efficiency Trade-off (achievable region) with $M = 1024$.

user rate was derived when an MMSE channel estimator and an MRC receiver were both employed, which reveals the effects of the hardware impairment on the effective SINR. Based on this result, the connectivity maximization problem was considered by appropriately selecting system parameters such as the training length and the number of users served simultaneously at a given target received signal power, and its closed-form optimal solution was provided. The optimal solution showed that using non-orthogonal pilots is beneficial for massive connectivity in general and becomes more critical as dirtier-RFs are used at the BS, the coherence time becomes shorter, or stricter power saving is applied. In order to obtain more intuitive insights, an asymptotic analysis with power saving, $p = \Theta(M^{-\alpha})$, was provided and it revealed that the use of non-orthogonal pilots can increase the connectivity from $\Theta(1)$ to $\Theta(M^{1/2-(\alpha-1/2)^+})$ while keeping the total energy consumption constant. In addition, the energy efficiency was also investigated and the optimal target received signal power was given asymptotically as $p = \Theta(1/\sqrt{M})$ in the perspective of energy efficiency maximization. Moreover, simulation results confirmed the analysis and showed the importance of not only employing non-orthogonal pilots but also of using a matched optimization of system parameters by considering the hardware impairment from the significantly improved achievable region of the connectivity and the energy efficiency, which can be considered as a cornerstone for achieving massive connectivity in the forthcoming 5G cellular systems.

Appendix A
Derivation of $\mathbf{C}_{\mathrm{agg},n}$

For simplicity, drop the index n. Inserting (3) into (4), we have

$$\rho = \frac{\mathbb{E}\left|\rho[\mathbf{y}]_m - [\mathbf{v}_{\mathrm{rx}}]_m\right|^2}{\mathbb{E}\left|[\mathbf{y}]_m\right|^2} = \rho^2 + \frac{\mathbb{E}\left|[\mathbf{v}_{\mathrm{rx}}]_m\right|^2}{\mathbb{E}\left|[\mathbf{y}]_m\right|^2}, \tag{27}$$

where the last equality comes from the assumption that $\mathbf{y}$ and $\mathbf{v}_{\mathrm{rx}}$ are independent. Also, inserting (3) into $\mathbb{E}|[\mathbf{y}]_m|^2$ yields

$$\mathbb{E}|[\mathbf{y}]_m|^2 = \mathbb{E}|[\mathbf{G}]_m^H(\overline{\mathbf{x}} + \mathbf{\Delta})|^2 = (1+\delta^2)\mathbb{E}[[\mathbf{G}]^H\mathbf{P}^{\Upsilon}[\mathbf{G}]_m], \tag{28}$$

where $[\mathbf{G}]_m$ denotes the mth row of $\mathbf{G}$ and the last equality comes from the fact $\mathbb{E}[\overline{\mathbf{x}}\overline{\mathbf{x}}^H] = \mathbf{P}^{\Upsilon}$ and $\mathbb{E}[\mathbf{\Delta}\mathbf{\Delta}^H] = \delta^2\mathbf{P}^{\Upsilon}$. Inserting (28) into (27), we have

$$\mathbb{E}|[\mathbf{v}_{rx}]_m|^2 = (1-\rho)\rho(1+\delta^2)\mathbb{E}[[\mathbf{G}]_m^H\mathbf{P}^{\Upsilon}[\mathbf{G}]_m],$$

from which we can derive $\mathbf{C}_{\mathrm{agg}}$.

Appendix B
Proof of Theorem 1

Proof: Using Lemma 1 in [16], the average user rate can be approximated as $R_k \approx (1-L/N)\log_2(1+\gamma_k)$, where

$$\gamma_k = \frac{p_k^{\mathrm{dt}}\beta_k\Xi_{kk}}{\sum\limits_{i\neq k} p_i^{\mathrm{dt}}\beta_i\Xi_{ki} + \sigma_{\mathrm{ch}}^2\mathbb{E}\|\widehat{\mathbf{h}}_k\|^2\sum\limits_{j=1}^{K} p_j^{\mathrm{dt}}\beta_j + \frac{\mathbb{E}\left[\widehat{\mathbf{h}}_k^H\mathbf{C}_{\mathrm{agg},n}^{\mathrm{dt}}\widehat{\mathbf{h}}_k\right]}{(1-\rho)^2}}, \tag{29}$$

with $\Xi_{ki} = \mathbb{E}\left[|\widehat{\mathbf{h}}_k^H\widehat{\mathbf{h}}_i|^2\right]$. Using Lemma 3 in [16] and Lemma 1, we obtain $\mathbb{E}\|\widehat{\mathbf{h}}_k\|^2 = (1-\sigma_{\mathrm{ch}}^2)M$ and

$$\Xi_{ki} = \begin{cases} \left(1-\sigma_{\mathrm{ch}}^2\right)^2 M, & \text{if } i \neq k, \\ \left(1-\sigma_{\mathrm{ch}}^2\right)^2 M(M+1), & \text{if } i = k. \end{cases} \tag{30}$$

The remaining step is to find $\mathbb{E}\left[\widehat{\mathbf{h}}_k^H\mathbf{C}_{\mathrm{agg},n}^{\mathrm{dt}}\widehat{\mathbf{h}}_k\right]$ which can be written as

$$\mathbb{E}\left[\widehat{\mathbf{h}}_k^H\mathbf{C}_{\mathrm{agg},n}^{\mathrm{dt}}\widehat{\mathbf{h}}_k\right] = (1-\rho)\left(\begin{array}{l} \left(1+\left(\rho+\delta^2\right)\sigma_{\mathrm{ch}}^2\sum\limits_{j=1}^{K}p_j^{\mathrm{dt}}\beta_j\right)\mathbb{E}||\widehat{\mathbf{h}}_k||^2 + \\ (1-\rho)\,\delta^2\mathbb{E}\left[\widehat{\mathbf{h}}_k^H\mathbf{\Phi}\widehat{\mathbf{h}}_k\right] + \\ \rho\left(1+\delta^2\right)\mathbb{E}\left[\widehat{\mathbf{h}}_k^H\overline{\mathbf{\Phi}}\widehat{\mathbf{h}}_k\right] \end{array}\right), \tag{31}$$

where the two expectations are obtained by applying Lemma 3 in [16], given as

$$\mathbb{E}\left[\widehat{\mathbf{h}}_k^H\mathbf{\Phi}\widehat{\mathbf{h}}_k\right] = \left(1-\sigma_{\mathrm{ch}}^2\right)^2 M\left(\sum_{j=1}^{K}p_j^{\mathrm{dt}}\beta_j + Mp_k^{\mathrm{dt}}\beta_k\right), \tag{32}$$

$$\mathbb{E}\left[\widehat{\mathbf{h}}_k^H\overline{\mathbf{\Phi}}\widehat{\mathbf{h}}_k\right] = \left(1-\sigma_{\mathrm{ch}}^2\right)^2 M\left(\sum_{j=1}^{K}p_j^{\mathrm{dt}}\beta_j + p_k^{\mathrm{dt}}\beta_k\right). \tag{33}$$

Inserting (30)-(33) into (29) completes the proof. ∎

Appendix C
Proof of Theorem 2

Proof: Suppose that non-orthogonal pilots are employed, i.e., K is greater than L. From Theorem 1, the average sum rate can be written as

$$\nu(L,K) = \left(1-\frac{L}{N}\right)K\log_2\left(1+\frac{AL}{BL+(CK+1)^2}\right). \tag{34}$$

In order to find the optimal values of (L,K), we first make a relaxation by replacing the integers (L,K) with real numbers (λ,κ), where $1 \leq \lambda \leq N$ and $\kappa > \lambda$. Since $\nu(\lambda,\kappa)$ is unimodal, it is maximized at the point $(\lambda^\star,\kappa^\star)$ such that

$\frac{\partial}{\partial\lambda}\nu(\lambda,\kappa)\big|_{\lambda=\lambda^\star}=0$ and $\frac{\partial}{\partial\kappa}\nu(\lambda,\kappa)\big|_{\kappa=\kappa^\star}=0$. Thus, we have

$$\log\left(1+\frac{\frac{A\lambda^\star}{(C\kappa^\star+1)^2}}{1+\frac{B\lambda^\star}{(C\kappa^\star+1)^2}}\right)=\frac{\frac{A(N-\lambda^\star)}{(C\kappa^\star+1)^2}}{\left(1+\frac{B\lambda^\star}{(C\kappa^\star+1)^2}\right)\left(1+\frac{(A+B)\lambda^\star}{(C\kappa^\star+1)^2}\right)}, \tag{35}$$

$$\log\left(1+\frac{\frac{A\lambda^\star}{(C\kappa^\star+1)^2}}{1+\frac{B\lambda^\star}{(C\kappa^\star+1)^2}}\right)=\frac{\frac{2AC\lambda^\star\kappa^\star}{(C\kappa^\star+1)^3}}{\left(1+\frac{B\lambda^\star}{(C\kappa^\star+1)^2}\right)\left(1+\frac{(A+B)\lambda^\star}{(C\kappa^\star+1)^2}\right)}. \tag{36}$$

From (35) and (36), we arrive at $(N-3\lambda^\star)\,C\kappa^\star+N-\lambda^\star=0$, which reveals that $\lambda^\star>N/3$ (and if not, the left hand side is strictly greater than 0) so that

$$\kappa^\star=\frac{N-\lambda^\star}{(3\lambda^\star-N)\,C}. \tag{37}$$

The non-orthogonal pilots become optimal if and only if $\kappa^\star>\lambda^\star$ or equivalently $\lambda^\star<\theta N/3$. Inserting (37) into (35), we obtain the relation $g_{A,B}(\lambda^\star)=0$. Suppose that there exists $\varphi\in(N/3,\theta N/3)$ such that $g_{A,B}(\varphi)=0$. Then, the use of non-orthogonal pilots is required and the optimal system parameters are given as $(L^\star,K^\star)=(\varphi,\frac{N-\varphi}{(3\varphi-N)C})$.

Now, suppose that only orthogonal pilots are allowed, i.e., $K=L$. Similar to those above, the system parameters are obtained by inserting $\lambda^\star=\kappa^\star$ into (35), which completes the proof. ∎

Appendix D
Proof of Theorem 3

Proof: The proof can be done similarly as in the proof of Theorem 2. As $M\to\infty$, we insert $p=P/M^\alpha$ into (14) and derive the asymptotic sum rate denoted by $R_{\text{sum}}^{M\to\infty}$. Then, the optimal number of simultaneously served users is set to $K=cM^t+o(M^t)$ for some value of $c>0$ and is inserted into in $R_{\text{sum}}^{M\to\infty}$. The final step is finding $t^\star$ and then $L^\star$ and $c^\star$ by equating $\frac{\partial}{\partial L}R_{\text{sum}}^{M\to\infty}|_{L=L^\star}=0$ and $\frac{\partial}{\partial c}R_{\text{sum}}^{M\to\infty}|_{c=c^\star}=0$.

Suppose that non-orthogonal pilots are employed. By inserting $p=P/M^\alpha$ into (14) and letting $M\to\infty$, we obtain

$$R_{\text{sum}}^{M\to\infty}=\left(1-\frac{L}{N}\right)K\log_2\left(1+\frac{1}{\frac{((1+\delta^2)PKM^{-\alpha}+1)^2}{L(1-\rho)^2P^2M^{1-2\alpha}}+\delta^2}\right), \tag{38}$$

which can be dealt according to α as follows.

1) $0\le\alpha<1/2$: In this case, (38) is maximized at $K=cM^{1/2}+o(M^{1/2})$ for some value of $c>0$. Inserting $K=cM^{1/2}$ into (38) by ignoring the non-dominant term $o(M^{1/2})$ provides

$$R_{\text{sum}}^{M\to\infty}=\left(1-\frac{L}{N}\right)c\sqrt{M}\log_2\left(1+\frac{1}{\frac{(1+\delta^2)^2c^2}{L(1-\rho)^2}+\delta^2}\right). \tag{39}$$

2) $\alpha=1/2$: Similarly as in the above case, (38) is maximized at $K=cM^{1/2}+o(M^{1/2})$ for some value of $c>0$. Inserting $K=cM^{1/2}$ into (38) provides

$$R_{\text{sum}}^{M\to\infty}=\left(1-\frac{L}{N}\right)c\sqrt{M}\log_2\left(1+\frac{1}{\frac{((1+\delta^2)Pc+1)^2}{L(1-\rho)^2P^2}+\delta^2}\right) \tag{40}$$

3) $\alpha>1/2$: In this case, (38) is maximized at $K=cM^\alpha+o(M^\alpha)$ for some value of $c>0$. Inserting $K=cM^\alpha$ into (38) provides

$$\begin{aligned}R_{\text{sum}}^{M\to\infty}&=\left(1-\frac{L}{N}\right)cM^\alpha\log_2\left(1+\frac{L(1-\rho)^2P^2M^{1-2\alpha}}{((1+\delta^2)\,Pc+1)^2}\right)\\&=\left(1-\frac{L}{N}\right)\frac{L(1-\rho)^2P^2M^{1-\alpha}}{((1+\delta^2)\,Pc+1)^2}\frac{c}{\log 2}.\end{aligned} \tag{41}$$

Since each of (39)-(41) is in a form similar to that of (34), $L^\star$ and $c^\star$ can be obtained similar to the way they were obtained in the proof of Theorem 2. Their values are determined after some mathematical manipulations and presented in Tables I and II. ∎

References

[1] A. Osseiran *et al.*, "Scenarios for the 5G Mobile and Wireless Communications: The Vision of the METIS Project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

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

[3] N. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2012.

[4] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[5] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[6] C. Studer, M. Wenk and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *Proc. ITG/IEEE Workshop Smart Antennas (WSA)*, Bremen, Germany, Feb. 2010, pp. 189–196.

[7] E. Björnson, P. Zetterberg, M. Bengtsson, and B. Ottersten, "Capacity limits and multiplexing gains of MIMO channels with transceiver impairments," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 91–94, Jan. 2013.

[8] X. Zhang, M. Matthaiou, M. Coldrey, and E. E. Björnson "Impact of Residual Transmit RF Impairments on Training-Based MIMO Systems," *IEEE Trans. Common.*, vol. 63, no. 8, pp. 2899–2911, Aug. 2015.

[9] X. Xia, D. Zhang, K. Xu and W. Ma, "Hardware impairments aware transceiver for full-duplex massive MIMO relaying," *IEEE Trans. on Signal Process.*, vol. 63, no. 24, pp. 6565 - 6580, Aug. 2015.

[10] U. Gustavsson *et al.*, "On the impact of hardware impairments on massive MIMO," *in proc. 2014 IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, 2014, pp. 294-300.

[11] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.

[12] J. Zhang, L. Dai, S. Sun, Z. Wang, "On the spectral efficiency of massive MIMO systems with low-resolution ADCs," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 842–845, May 2016.

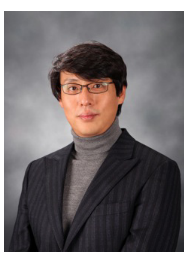
[13] J. Zhang, L. Dai, X. Zhang, E. Bjornson, and Z. Wang, "Achievable rate of Rician large-scale MIMO channels with transceiver hardware impairments," *to appear in IEEE Trans. Veh. Technol.*, 2016.

[14] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," *in Proc. IEEE Vehicular Technology Conference (VTC Spring13)*, Dresden, Germany, Jun. 2013, pp. 1–5.

[15] H. Nikopour and H. Baligh, "Sparse code multiple access," *in Proc. IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC13)*, London, U.K., Sep. 2013, pp. 332–336.

[16] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, ”Power scaling of uplink massive MIMO systems with arbitrary-rank channel means,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, Oct. 2014.
[17] H. Wang, W. Zhang, Y. Liu, Q. Xu, and P. Pan, ”On design of nonorthogonal pilot signals for a multi-cell massive MIMO system,” *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 129–132, Apr. 2015.
[18] J.-C. Shen, J. Zhang, and K. B. Letaief, ”Downlink user capacity of massive MIMO user pilot contamination,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3183–3193, Jun. 2015.
[19] K. J. Choi and K. S. Kim, ”Optimal Semi-Persistent Uplink Scheduling Policy for Large-Scale Antenna Systems,” *to be submitted to IEEE Wireless Commun.*, Apr. 2017.
[20] H. Wang, W. Zhang, Y. Liu, Q. Xu, and P. Pan, ”On design of non-orthogonal pilot signals for a multi-cell massive MIMO system,” *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 129–132, Apr. 2015.
[21] J. W. Kang, Y. Whang, B. H. Ko, and K. S. Kim, ”Generalized cross-correlation properties of Chu sequences,” *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 438–444, Jan. 2012.
[22] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, ”Systematic design of unitary space-time constellation,” *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1962–1973, Sep. 2000.
[23] P. Xia, S. Zhou, and G. B. Giannakis, ”Achieving the Welch bound with difference sets,” *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1900–1907, May 2005.

Kyung Jun Choi was born in Busan, Korea, on 7 February. 1987. He received the B.S. and PhD degrees in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea in February 2010 and February 2017, respectively. He is currently working with WILUS Inc., Seoul, Korea. His current research interests include massive MIMO transceiver design, network performance analysis via stochastic geometry and random matrix theory.



Kwang Soon Kim (S’95, M’99, SM’04) was born in Seoul, Korea, on September 20, 1972. He received the B.S. (summa cum laude), M.S.E., and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in February 1994, February 1996, and February 1999, respectively.

From March 1999 to March 2000, he was with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA, U.S.A., as a Postdoctoral Researcher. From April 2000 to February 2004, he was with the Mobile Telecommunication Research Laboratory, Electronics and Telecommunication Research Institute, Daejeon, Korea as a Senior Member of Research Staff. Since March 2004, he has been with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea, now is an Associate Professor.

Prof. Kim is a Senior Member of the IEEE, served as an Editor of the Journal of the Korean Institute of Communications and Information Sciences (KICS) from 2006-2012, as the Editor-in-Chief of the journal of KICS since 2013, as an Editor of the Journal of Communications and Networks (JCN) since 2008, as an Editor of the IEEE Transactions on Wireless Communications since 2009. He was a recipient of the Postdoctoral Fellowship from Korea Science and Engineering Foundation (KOSEF) in 1999. He received the Outstanding Researcher Award from Electronics and Telecommunication Research Institute (ETRI) in 2002, the Jack Neubauer Memorial Award (Best system paper award, IEEE Transactions on Vehicular Technology) from IEEE Vehicular Technology Society in 2008, and LG R&D Award: Industry-Academic Cooperation Prize, LG Electronics, 2013. His research interests are in signal processing, communication theory, information theory, and stochastic geometry applied to wireless heterogeneous cellular networks, wireless local area networks, wireless D2D networks and wireless ad doc networks.