Optimal Uplink Pilot-Data Power Allocation for Large-Scale Antenna Array Aided OFDM Systems

Xinying Guo, Jiankang Zhang, *Senior Member, IEEE*, Sheng Chen, *Fellow, IEEE*, Chunhua Zhu, and Jing Yang

***Abstract*—The effective aggregate rate (AR) performance of large-scale multiple-input multiple-output aided orthogonal frequency division multiplexing systems based on time division duplexing protocol critically depends on uplink (UL) pilot-data power allocations for all users. We derive an asymptotic closed-form expression of the system's achievable AR for typical channel estimation schemes, which is a function of individual users' UL power allocation factors. We prove that the effective AR is a convex function of each UL power allocation factor, when all the other factors are given. The globally optimal UL pilot-data power allocation that maximizes the system's effective AR can easily be obtained using an iterative procedure, by solving an univariate convex optimization with the golden section method for individual UL power allocation factors one by one in each iteration. Owing to the piecewise convexity of the effective AR, this iterative algorithm guarantees to find the unique globally optimal solution with only one iteration. The simulation results confirm that the achievable effective AR is significantly enhanced through the proposed optimization of UL pilot-data power allocation.**

***Index Terms*—Multi-cell systems, massive multiple-input multiple-output, orthogonal frequency division multiplexing, uplink power allocation, channel estimation, aggregation rate**

I. Introduction

Owing to its high spectral efficiency (SE) and energy efficiency (EE), the massive multiple-input multiple-output (MIMO) technology [1]–[7] has attracted considerable attention from both academia and industry. By deploying a large-scale antenna array at the base station (BS), it is capable of serving multiple terminals with the same time-frequency resource block. As the number of BS antennas increase, the effects of uncorrelated intra-cell interference as well as noise diminish, and they completely disappear at the limit case of infinite many antennas. Therefore, when a single-cell setup is considered, it can always recover from low signal-to-noise ratio (SNR) conditions with simple linear processing by employing a sufficient number of antennas [5]. However, in the realistic multi-cell scenario, owing to the inevitable reuse of the same pilot sequences among neighboring cells, pilot contamination (PC) may occur, which may result in the BS being unable to reliably estimate the true channels between its serving users and itself, even when an infinite number of antennas are deployed. PC constitutes a serious impairment that limits the system's achievable performance, and hence considerable research efforts have been focused on mitigating or eliminating PC [8]–[17]. Among various PC mitigating/eliminating schemes [8]–[17], the scheme of [17] has been shown to be most effective and efficient. This is because the PC elimination schemes [8]–[16] require an excessively long channel coherence time to work, and they can no longer be used if the channel coherent time is insufficiently large[1]. By contrast, the scheme [17] is capable of completely eliminating PC under a much shorter coherence time, and even for extremely short channel coherent time, it can still be implemented to significantly reduce PC.

For the fairness of service to all the users, typically, the network operator divides the total downlink (DL) power equally among all the users. For the same reason, the uplink (UL) power allocated to each user is identical too. Most massive MIMO systems adopt the time division duplexing (TDD) protocol, and the training only occurs at the UL. So far, all the existing literature for pilot decontamination in multi-cell TDD based massive MIMO systems assume that the UL power for each user is divided equally between its UL pilot training and data transmission [8]–[17]. Although these pilot decontamination schemes are able to improve the system's SE, they miss an excellent opportunity to further improve the system's performance by UL pilot and data power allocation optimization. Clearly, if the UL pilot power is too small, the accuracy of the channel estimation (CE) will be poor, and this will have adverse effects on the achievable UL data rate as well as DL data rate. On the other hand, if more UL power is assigned to the CE, less power is used for UL data transmission, which will reduce the achievable UL data rate. Therefore, given the same total UL power for every user, the system's achievable aggregate rate (AR), i.e., the total rate of the UL and DL sum rates, can be maximized by optimizing the UL pilot and data power allocation for every user.

Power allocation has been an important issue, dating back to single-antenna systems. Recently, power allocation in massive MIMO systems has become a hot topic. In [18], numerical optimization is applied to find the optimal UL pilot-data power allocation by maximizing the effective UL sum rate. However, the effect of UL pilot-data power allocation on the DL sum rate is not considered. The study [19] considers UL pilot power allocation to improve the CE quality, and it reveals that an imbalance power between pilot and data is beneficial for cell-edge users. Specifically, more power should be allocated to pilot training than to data transmission to improve the CE

X. Guo (guoxinying@haut.edu.cn), C. Zhu (zhuchunhua@haut.edu.cn) and J. Yang (yangjing@haut.edu.cn) are with College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450000, China.

J. Zhang (jz09v@ecs.soton.ac.uk) and S. Chen (sqc@ecs.soton.ac.uk) are with Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK. S. Chen is also with King Abdulaziz University, Jeddah 21589, Saudi Arabia.

This work was supported by the National Natural Science Foundation of China (Grants 61901159, 61571401, 61601170 and 61871176) and the Innovative Talent of Colleges and University of Henan Province under grant 18HASTIT021.

[1]The scheme of [9] for example can eliminate the PC completely but only with the help of the second-order statistics of all the UL channels. The second-order statistics depend on the user distribution that can change. Even with a fixed user distribution, the acquisition of such a large amount of second-order statistics at the BSs is time-consuming and, moreover, sharing them requires a huge amount of back-haul transmissions.

quality. But this is hardly surprising. Moreover, improving the CE does not necessarily enhance the UL sum rate. Chien *et al.* [20] consider joint pilot signal design and UL pilot-data power allocation, but the effect of UL pilot-data power allocation to the DL sum rate is not taken into account. For the single-cell massive MIMO system, the work [21] applies a pilot and data power allocation to minimize the total UL and DL transmit power under the per-user signal to interference-plus-noise ratio (SINR) constraint. Cheng *et al.* [22] derive the optimal pilot-data power allocation for the single-cell massive MIMO UL only. Yang and Marzetta [23] perform the data power optimization for guaranteeing uniformly high user throughput in a multi-cell wireless network. To achieve the optimal UL system's EE, Guo *et al.* [24] propose an UL pilot and data power control to minimize the sum transmit power of all users subject to the per-user SINR and per-user power constraints in multi-cell massive MIMO systems.

To the best of our knowledge, no prior work analyzes the impact of the power allocation between UL pilot training and UL data transmission to the system's effective AR performance per-cell under the per-user power constraints for the multi-cell massive MIMO aided orthogonal frequency division multiplexing (OFDM) system, which motivates the work in this paper. The main contributions of this paper are summarized as follows.

- We consider typical CE schemes that are capable of eliminating PC given power allocation, and derive the common expressions for the estimated channel state information (CSI) as the functions of the UL pilot-data power allocations for all the individual users. Based on the estimated CSI, we derive the asymptotic closed-form expressions of the achievable UL and DL sum rates for these CE schemes. This leads to the asymptotic closed-form expression of the effective AR per-cell, which is a function of all the UL pilot-data power allocation factors.
- We prove that the effective AR is a convex function of each UL power allocation factor, when all the other factors are given. Therefore, the globally maximum effective AR can be attained by an iterative procedure, which solves an univariate convex optimization by the golden section method for individual UL power allocation factor one by one at each iteration. Owing to the piecewise convexity, the iterative procedure converges to the unique globally optimal UL power allocation solution with only one iteration.

In this paper, boldface upper-case symbols denote matrices, e.g., $\mathbf{X}$, and underlined boldface upper-case symbols denote column vectors, e.g., $\underline{\mathbf{X}}$, while real vectors are also denoted by underlined boldface lower-case symbols, e.g., $\underline{\varepsilon}$. The transpose and Hermitian transpose operators are denoted by $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$, respectively, while $\mathrm{diag}\{\underline{\mathbf{X}}\}$ is the diagonal matrix with the elements of $\underline{\mathbf{X}}$ as its diagonal entries, and $\mathrm{Tr}\{\cdot\}$ is the matrix trace operator. $[\mathbf{X}]_{(i,j)}$ is the (i,j)-th sub-matrix of $\mathbf{X}$, while $\widehat{X}$ represents the estimate of X. The $(K\times K)$ identity matrix is denoted by $\mathbf{I}_K$, and $\mathbf{0}_K$ is the $(K\times K)$ zero matrix. $\mathbb{E}\{\cdot\}$ is the expectation operator and $vec(\cdot)$ is the column stacking operator, while $\lfloor\cdot\rfloor$ is the integer floor operator.

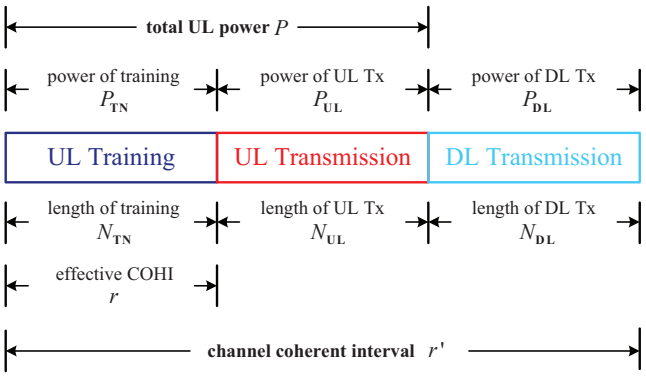

Fig. 1. TDD protocol frame structure and its relationship to the channel's coherence interval and power allocation.

II. Multi-cell TDD OFDM System

Consider a cellular network composed of L hexagonal cells, labelled by $l=1,2,\cdots,L$, where the BS of each cell employs an array of Q antennas to serve the U single-antenna mobile stations (MSs) using the same resource block. It is assumed that $Q \gg U$ and OFDM is employed, where all BSs and MSs are synchronized, relying on a TDD protocol with unity frequency reuse. Fig. 1 details the TDD protocol frame structure and its relationship to the channel coherence interval (COHI) and power allocation. The COHI r' specifies the maximum number of OFDM symbols during the duration of which the channel impulse responses (CIRs) remain near constant. The pilot length of UL training is N_{TN}, which is equal to the effective COHI r, while the data lengths of UL and DL transmissions are N_{UL} and N_{DL}, respectively. Clearly, $r'=N_{\mathrm{TN}}+N_{\mathrm{UL}}+N_{\mathrm{DL}}$. In other words, the COHI must be larger than the UL training duration N_{TN}. Clearly, a block-type pilot arrangement is adopted.

The DL transmission power per antenna of each BS at each subcarrier is P_{DL}, and the total UL power of each user at each subcarrier is P which consists of the UL training power P_{TN} and the UL transmission power P_{UL}, that is, $P=P_{\mathrm{TN}}+P_{\mathrm{UL}}$. For fairness, every user is assigned with the same P. Define

$$P_{\mathrm{TN}} = \varepsilon_l^u P, \tag{1}$$

$$P_{\mathrm{UL}} = (1-\varepsilon_l^u) P, \tag{2}$$

where $0<\varepsilon_l^u<1$ is referred to as the UL pilot power allocation factor of user u $(1\le u\le U)$ in cell l. Then the UL power allocation of user u in cell l between pilot and data becomes determining the value of ε_l^u. Note that the existing works [8]–[17] set all $\varepsilon_l^u=0.5$.

III. UL Training

In order to explore the relationship between the effective AR and the UL pilot-data power allocation, we consider the three typical CE schemes, the conventional simultaneous CE scheme, the successive CE scheme [14] and our user-grouping based CE scheme [17]. The channel models for these three CE schemes are explicitly explained in [17]. Specifically, the first two CE schemes consider the signals received by all the target BS's antennas on an individual OFDM subcarrier, which is less efficient in terms of training duration. Specifically, for the conventional simultaneous CE scheme, in order to have

TABLE I
BASIC PARAMETERS FOR DIFFERENT CE SCHEMES GIVEN r', P AND $P_{\rm DL}$, ASSUMING THAT $N_{\rm UL}=N_{\rm DL}$

Schemes \ Parameters	$r=N_{\rm TN}$	$N_{\rm UL}=\frac{r'-N_{\rm TN}}{2}$	$N_{\rm DL}=\frac{r'-N_{\rm TN}}{2}$	$\rho_{\rm TN}=\frac{P}{N_{\rm TN}}$	$\rho_{\rm UL}=\frac{P}{N_{\rm UL}}$	$p_{\rm DL}=\frac{P_{\rm DL}}{N_{\rm DL}}$
Conventional simultaneous CE	LU	$\frac{r'-LU}{2}$	$\frac{r'-LU}{2}$	$\frac{P}{LU}$	$\frac{P}{\frac{r'-LU}{2}}$	$\frac{P_{\rm DL}}{\frac{r'-LU}{2}}$
Successive CE [14]	$(L+1)U$	$\frac{r'-(L+1)U}{2}$	$\frac{r'-(L+1)U}{2}$	$\frac{P}{(L+1)U}$	$\frac{P}{\frac{r'-(L+1)U}{2}}$	$\frac{P_{\rm DL}}{\frac{r'-(L+1)U}{2}}$
Our user grouping based CE [17]	f	$\frac{r'-f}{2}$	$\frac{r'-f}{2}$	$\frac{P}{f}$	$\frac{P}{\frac{r'-f}{2}}$	$\frac{P_{\rm DL}}{\frac{r'-f}{2}}$

LU orthogonal pilots, the length of pilot sequences must be LU and, therefore, the effective COHI required is $r=N_{\rm TN}=LU$. For the successive CE scheme [14], which consists of $L+1$ successive UL pilot training phases, the effective COHI required is $r=N_{\rm TN}=(L+1)U$. By contrast, our user grouping based CE scheme [17] considers the signals received for all the OFDM subcarriers together on an individual BS antenna basis, and the effective COHI required is equal to the number of groups f, i.e., $r=N_{\rm TN}=f$, which is much smaller than LU. The basic parameters of these three CE schemes are given in Table I, where $\varepsilon_l^u\rho_{\rm TN}$ is the average UL training power of user u in the cell l at each subcarrier of each OFDM symbol, $(1-\varepsilon_l^u)\rho_{\rm UL}$ is the average UL transmission power of user u in the cell l at each subcarrier of each OFDM symbol, and $p_{\rm DL}$ is the average transmission power per antenna of the BS at each subcarrier of each OFDM symbol.

A. Conventional Simultaneous CE

The COHI must be larger than $N_{\rm TN}=LU$. At the commencement of a frame, all users of all cells synchronously transmit $N_{\rm TN}$ OFDM pilot symbols (PSs) to their serving BSs. The mth frequency domain (FD) PS of user u in the lth cell is given by $\underline{\mathbf{S}}_l^u[m]=\left[S_l^u[m,1]\ S_l^u[m,2]\cdots S_l^u[m,N]\right]^{\rm T}$, $1\le m\le N_{\rm TN}$, where N is the number of subcarriers and the power of each $S_l^u[m,n]$ is unity. Let $H_{l,l',q}^{u,n}[m]$ be the UL FD channel transfer function (FDCHTF) linking the uth user in cell l to the qth antenna of the l'th cell's BS, at the nth subcarrier of the mth OFDM symbol. Denote $Y_{l',q}[m,n]$ as the signal received by the qth receive antenna of the l'th BS at the nth subcarrier of the mth OFDM symbol, and let the FD representation of the corresponding UL channel's additive white Gaussian noise (AWGN) be $W_{l',q}[m,n]$, which has the power of σ_w^2, that is, $W_{l',q}[m,n]\sim\mathcal{CN}(0,\sigma_w^2)$. Next let $\underline{\mathbf{Y}}_{l'}[m,n]\in\mathbb{C}^Q$ and $\underline{\mathbf{W}}_{l'}[m,n]\in\mathbb{C}^Q$ be the two vectors hosting $Y_{l',q}[m,n]$ and $W_{l',q}[m,n]$ for $1\le q\le Q$, respectively, while defining $\underline{\mathbf{S}}_{l'}[m,n]\in\mathbb{C}^U$ and $\underline{\mathbf{S}}_l[m,n]\in\mathbb{C}^U$ as the two vectors hosting $S_{l'}^u[m,n]$ and $S_l^u[m,n]$ for $1\le u\le U$, respectively. Furthermore, define $\mathbf{H}_{l',l'}^n[m]\in\mathbb{C}^{Q\times U}$ and $\mathbf{H}_{l,l'}^n[m]\in\mathbb{C}^{Q\times U}$ as the two FDCHTF matrices having their (q,u)th elements given by $H_{l',l',q}^{u,n}[m]$ and $H_{l,l',q}^{u,n}[m]$, respectively. Then, we have

$$\underline{\mathbf{Y}}_{l'}[m,n]=\underbrace{\sqrt{\rho_{\rm TN}}\mathbf{H}_{l',l'}^n[m]\boldsymbol{\Delta}_{l'}\underline{\mathbf{S}}_{l'}[m,n]}_{\text{Desired term}}+\underbrace{\sqrt{\rho_{\rm TN}}\sum_{l=1,l\neq l'}^{L}\mathbf{H}_{l,l'}^n[m]\boldsymbol{\Delta}_l\underline{\mathbf{S}}_l[m,n]}_{\text{Inter-cell interference}}+\underline{\mathbf{W}}_{l'}[m,n],\quad(3)$$

where $\rho_{\rm TN}=\frac{P}{LU}$ and $\boldsymbol{\Delta}_l=\text{diag}\{\underline{\boldsymbol{\varepsilon}}_l^{\frac{1}{2}}\}$ with $\underline{\boldsymbol{\varepsilon}}_l^{\frac{1}{2}}=\left[\sqrt{\varepsilon_l^1}\ \sqrt{\varepsilon_l^2}\cdots\sqrt{\varepsilon_l^U}\right]^{\rm T}$. Since the channel is time-invariant for the duration of COHI, $\mathbf{H}_{l,l'}^n[m]=\mathbf{H}_{l,l'}^n\in\mathbb{C}^{Q\times U}$ for $1\le m\le N_{\rm TN}$ in (3), and the (q,u)th element of $\mathbf{H}_{l,l'}^n$ is $H_{l,l',q}^{u,n}$ with zero mean and a variance of $\beta_{l,l'}^u$, where $\beta_{l,l'}^u$ denotes the path-loss coefficient of the link between the uth user in cell l to any antenna of the l'th cell's BS, which is a constant with respect to both frequency and BS's antenna index since the geometric and shadow fading only change very slowly over space. As a total of $N_{\rm TN}=LU$ OFDM symbols are entirely dedicated for pilots subcarriers, the received signals associated with the LU OFDM symbols, namely, (3) over $1\le m\le LU$, can be expressed as

$$\mathbf{Y}_{l'}[n]=\sqrt{\rho_{\rm TN}}\mathbf{H}_{l',l'}^n\boldsymbol{\Delta}_{l'}\mathbf{S}_{l'}[n]+\sqrt{\rho_{\rm TN}}\sum_{l=1,l\neq l'}^{L}\mathbf{H}_{l,l'}^n\boldsymbol{\Delta}_l\mathbf{S}_l[n]+\mathbf{W}_{l'}[n],\ 1\le n\le N,\quad(4)$$

with $\mathbf{Y}_{l'}[n]\in\mathbb{C}^{Q\times LU}$, $\mathbf{S}_l[n]\in\mathbb{C}^{U\times LU}$ and $\mathbf{W}_{l'}[n]\in\mathbb{C}^{Q\times LU}$. Clearly, $\mathbf{S}_{l'}[n]\mathbf{S}_{l'}^{\rm H}[n]=LU\mathbf{I}_U$, and $\mathbf{S}_l[n]\mathbf{S}_{l'}^{\rm H}[n]=\mathbf{0}_U$ for $l\neq l'$, since we have a total of LU orthogonal pilot sequences. Also $\mathbb{E}\{vec(\mathbf{H}_{l',l'}^n)\}=\underline{\mathbf{0}}_{QU}$ and the covariance matrix $\mathbf{R}_{l',l'}^n\in\mathbb{C}^{QU\times QU}$ of $vec(\mathbf{H}_{l',l'}^n)$ is given by

$$\mathbf{R}_{l',l'}^n=\mathbb{E}\left\{vec(\mathbf{H}_{l',l'}^n)\left(vec(\mathbf{H}_{l',l'}^n)\right)^{\rm H}\right\}=\begin{bmatrix}\beta_{l',l'}^1\mathbf{I}_Q & \mathbf{0}_Q & \cdots & \mathbf{0}_Q\\ \mathbf{0}_Q & \beta_{l',l'}^2\mathbf{I}_Q & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}_Q\\ \mathbf{0}_Q & \cdots & \mathbf{0}_Q & \beta_{l',l'}^U\mathbf{I}_Q\end{bmatrix}.\quad(5)$$

The minimum mean square error (MMSE) estimate $\widehat{\mathbf{H}}_{l',l'}^n$ of $\mathbf{H}_{l',l'}^n$ is given by

$$\begin{aligned}vec\left(\widehat{\mathbf{H}}_{l',l'}^n\right)&=\mathbf{R}_{l',l'}^n\left(\frac{\sigma_w^2}{P}\widetilde{\boldsymbol{\Delta}}_{l'}+\mathbf{R}_{l',l'}^n\right)^{-1}\\&\quad\times\left(vec(\mathbf{H}_{l',l'}^n)+\frac{1}{\sqrt{LUP}}vec\left(\mathbf{W}_{l'}[n]\mathbf{S}_{l'}^{\rm H}[n]\boldsymbol{\Delta}_{l'}^{-1}\right)\right)\\&=\boldsymbol{\Upsilon}_{l',l'}^n\left(vec(\mathbf{H}_{l',l'}^n)+\frac{1}{\sqrt{LUP}}vec\left(\mathbf{W}_{l'}[n]\mathbf{S}_{l'}^{\rm H}[n]\boldsymbol{\Delta}_{l'}^{-1}\right)\right),\quad(6)\end{aligned}$$

where

$$\widetilde{\boldsymbol{\Delta}}_{l'}=\begin{bmatrix}[\widetilde{\boldsymbol{\Delta}}_{l'}]_{(1,1)} & \mathbf{0}_Q & \cdots & \mathbf{0}_Q\\ \mathbf{0}_Q & [\widetilde{\boldsymbol{\Delta}}_{l'}]_{(2,2)} & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}_Q\\ \mathbf{0}_Q & \cdots & \mathbf{0}_Q & [\widetilde{\boldsymbol{\Delta}}_{l'}]_{(U,U)}\end{bmatrix},\quad(7)$$

$$\boldsymbol{\Upsilon}^n_{l',l'}=\begin{bmatrix}[\boldsymbol{\Upsilon}^n_{l',l'}]_{(1,1)} & \mathbf{0}_Q & \cdots & \mathbf{0}_Q\\ \mathbf{0}_Q & [\boldsymbol{\Upsilon}^n_{l',l'}]_{(2,2)} & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}_Q\\ \mathbf{0}_Q & \cdots & \mathbf{0}_Q & [\boldsymbol{\Upsilon}^n_{l',l'}]_{(U,U)}\end{bmatrix}. \quad (8)$$

in which

$$[\widetilde{\boldsymbol{\Delta}}_{l'}]_{(u,u)}=\frac{1}{\varepsilon^u_{l'}}\mathbf{I}_Q,\ 1\le u\le U, \quad (9)$$

$$[\boldsymbol{\Upsilon}^n_{l',l'}]_{(u,u)}=\frac{\beta^u_{l',l'}}{\frac{\sigma^2_w}{\varepsilon^u_{l'}P}+\beta^u_{l',l'}}\mathbf{I}_Q,\ 1\le u\le U. \quad (10)$$

It is well-known that the distribution of the MMSE estimator is $vec(\widehat{\mathbf{H}}^n_{l',l'})\sim\mathcal{CN}(\underline{\mathbf{0}}_{QU},\boldsymbol{\Phi}^n_{l',l'})$, that is, $vec(\widehat{\mathbf{H}}^n_{l',l'})$ is an unbiased estimate of $vec(\mathbf{H}^n_{l',l'})$ with the estimation accuracy specified by the covariance matrix $\boldsymbol{\Phi}^n_{l',l'}$, which is given by

$$\boldsymbol{\Phi}^n_{l',l'}=\mathbf{R}^n_{l',l'}\left(\frac{\sigma^2_w}{P}\widetilde{\boldsymbol{\Delta}}_{l'}+\mathbf{R}^n_{l',l'}\right)^{-1}\mathbf{R}^n_{l',l'}$$
$$=\begin{bmatrix}[\boldsymbol{\Phi}^n_{l',l'}]_{(1,1)} & \mathbf{0}_Q & \cdots & \mathbf{0}_Q\\ \mathbf{0}_Q & [\boldsymbol{\Phi}^n_{l',l'}]_{(2,2)} & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}_Q\\ \mathbf{0}_Q & \cdots & \mathbf{0}_Q & [\boldsymbol{\Phi}^n_{l',l'}]_{(U,U)}\end{bmatrix} \quad (11)$$

with

$$\left[\boldsymbol{\Phi}^n_{l',l'}\right]_{(u,u)}=\frac{(\beta^u_{l',l'})^2}{\frac{\sigma^2_w}{\varepsilon^u_{l'}P}+\beta^u_{l',l'}}\mathbf{I}_Q,\ 1\le u\le U. \quad (12)$$

The channel estimation error $vec(\widetilde{\mathbf{H}}^n_{l',l'})=vec(\mathbf{H}^n_{l',l'})-vec(\widehat{\mathbf{H}}^n_{l',l'})$ has the covariance matrix

$$\boldsymbol{\Xi}^n_{l',l'}=\mathbf{R}^n_{l',l'}-\boldsymbol{\Phi}^n_{l',l'}$$
$$=\begin{bmatrix}[\boldsymbol{\Xi}^n_{l',l'}]_{(1,1)} & \mathbf{0}_Q & \cdots & \mathbf{0}_Q\\ \mathbf{0}_Q & [\boldsymbol{\Xi}^n_{l',l'}]_{(2,2)} & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}_Q\\ \mathbf{0}_Q & \cdots & \mathbf{0}_Q & [\boldsymbol{\Xi}^n_{l',l'}]_{(U,U)}\end{bmatrix}, \quad (13)$$

where

$$[\boldsymbol{\Xi}^n_{l',l'}]_{(u,u)}=\left(\beta^u_{l',l'}-\frac{(\beta^u_{l',l'})^2}{\frac{\sigma^2_w}{\varepsilon^u_{l'}P}+\beta^u_{l',l'}}\right)\mathbf{I}_Q,\ 1\le u\le U. \quad (14)$$

B. Successive CE [14]

This scheme consists of $(L+1)$ pilot transmission phases. In the first phase, denoted as phase 0, all users in all cells transmit their assigned pilot sequences to their associated BSs synchronously. At phase l, where $1\le l\le L$, the users in cell l stay silent, while all the other users in the other $(L-1)$ cells transmit their pilot sequences synchronously.

Phase 0: All the users of all the cells transmit the U OFDM PSs to their serving BSs. The mth FD PS of user u in cell l is given by $\underline{\mathbf{S}}^u_l[m]=\left[S^u_l[m,1]\ S^u_l[m,2]\cdots S^u_l[m,N]\right]^{\mathrm{T}}$, $1\le m\le U$, and the power of $S^u_l[m,n]$ is unity. Since the length of pilot sequences is U, the U orthogonal pilots are reused in all the L cells. That is, $\{\underline{\mathbf{S}}^u_l[m]\}$ and $\{\underline{\mathbf{S}}^u_{l'}[m]\}$ are identical for $l\ne l'$. Similarly to (4), the l'th BS's received signal matrix $\mathbf{Y}^{(0)}_{l'}[n]\in\mathbb{C}^{Q\times U}$ can be expressed by

$$\mathbf{Y}^{(0)}_{l'}[n]=\sqrt{\rho_{\mathrm{TN}}}\mathbf{H}^n_{l',l'}\boldsymbol{\Delta}_{l'}\mathbf{S}[n]$$
$$+\sqrt{\rho_{\mathrm{TN}}}\sum_{l=1,l\ne l'}^{L}\mathbf{H}^n_{l,l'}\boldsymbol{\Delta}_l\mathbf{S}[n]+\mathbf{W}^{(0)}_{l'}[n],\ 1\le n\le N, \quad (15)$$

where $\rho_{\mathrm{TN}}=\frac{P}{(L+1)U}$ and $\mathbf{S}_{l'}[n]=\mathbf{S}_l[n]=\mathbf{S}[n]\in\mathbb{C}^{U\times U}$, while the elements of $\mathbf{W}^{(0)}_{l'}[n]\in\mathbb{C}^{Q\times U}$ are AWGNs obeying the distribution $\mathcal{CN}(0,\sigma^2_w)$. Clearly, $\mathbf{S}[n]\mathbf{S}^{\mathrm{H}}[n]=U\mathbf{I}_U$.

Phase $1\le l\le L$: The users of the lth cell stay silent, while the users of the other cells transmit their pilots. Thus, the l'th BS's received signal matrix $\mathbf{Y}^{(l)}_{l'}[n]\in\mathbb{C}^{Q\times U}$ is given by

$$\mathbf{Y}^{(l)}_{l'}[n]=\sqrt{\rho_{\mathrm{TN}}}\sum_{\bar{l}=1,\bar{l}\ne l}^{L}\mathbf{H}^n_{\bar{l},l'}\boldsymbol{\Delta}_{\bar{l}}\mathbf{S}[n]+\mathbf{W}^{(l)}_{l'}[n],1\le n\le N, \quad (16)$$

where the elements of $\mathbf{W}^{(l)}_{l'}[n]\in\mathbb{C}^{Q\times U}$ are AWGNs obeying the distribution $\mathcal{CN}(0,\sigma^2_w)$.

During the $(L$–$1)$ phases of $1\le l\le L$ and $l\ne l'$, $\mathbf{Y}^{(l)}_{l'}[n]$ does not contain the interference from the users in the lth cell. The BS of the l'th cell can compute the aggregated observation

$$\mathbf{Y}_{l'}[n]=\mathbf{Y}^{(0)}_{l'}[n]+\sum\nolimits_{l=1,l\ne l'}^{L}\mathbf{Y}^{(l)}_{l'}[n]-(L-1)\mathbf{Y}^{(l')}_{l'}[n]$$
$$=\sqrt{\rho_{\mathrm{TN}}}\mathbf{H}^n_{l',l'}\boldsymbol{\Delta}_{l'}\widetilde{\mathbf{S}}[n]+\widetilde{\mathbf{W}}_{l'}[n], \quad (17)$$

where $\widetilde{\mathbf{S}}[n]=L\mathbf{S}[n]$ with $\widetilde{\mathbf{S}}[n]\widetilde{\mathbf{S}}^{\mathrm{H}}[n]=L^2U\mathbf{I}_U$, and $\widetilde{\mathbf{W}}_{l'}[n]=\mathbf{W}^{(0)}_{l'}[n]+\sum_{l=1,l\ne l'}^{L}\mathbf{W}^{(l)}_{l'}[n]-(L-1)\mathbf{W}^{(l')}_{l'}[n]$ whose elements are white Gaussian variables obeying the distribution $\mathcal{CN}\left(0,(L^2-L+1)\sigma^2_w\right)$. Clearly, $\mathbf{Y}_{l'}[n]$ is PC free. The MMSE estimate $\widehat{\mathbf{H}}^n_{l',l'}$ of $\mathbf{H}^n_{l',l'}$ is given by

$$vec(\widehat{\mathbf{H}}^n_{l',l'})=\boldsymbol{\Upsilon}^n_{l',l'}\Bigg(vec(\mathbf{H}^n_{l',l'})$$
$$+\frac{1}{\sqrt{\frac{L^4UP}{(L+1)}}}vec\left(\widetilde{\mathbf{W}}_{l'}[n]\widetilde{\mathbf{S}}^{\mathrm{H}}[n]\boldsymbol{\Delta}^{-1}_{l'}\right)\Bigg), \quad (18)$$

where $\boldsymbol{\Upsilon}^n_{l',l'}\in\mathbb{C}^{QU\times QU}$ takes the same form of (8), but its U sub-matrices are given by

$$[\boldsymbol{\Upsilon}^n_{l',l'}]_{(u,u)}=\frac{\beta^u_{l',l'}}{\frac{(L+1)(L^2-L+1)\sigma^2_w}{L^2\varepsilon^u_{l'}P}+\beta^u_{l',l'}}\mathbf{I}_Q,1\le u\le U. \quad (19)$$

The distribution of this MMSE estimator is $vec(\widehat{\mathbf{H}}^n_{l',l'})\sim\mathcal{CN}(\underline{\mathbf{0}}_{QU},\boldsymbol{\Phi}^n_{l',l'})$. The covariance matrix of this estimator takes the same form of (11), but the U sub-matrices of $\boldsymbol{\Phi}^n_{l',l'}$ are given by

$$[\boldsymbol{\Phi}^n_{l',l'}]_{(u,u)}=\frac{(\beta^u_{l',l'})^2}{\frac{(L+1)(L^2-L+1)\sigma^2_w}{L^2\varepsilon^u_{l'}P}+\beta^u_{l',l'}}\mathbf{I}_Q,1\le u\le U. \quad (20)$$

The covariance matrix $\boldsymbol{\Xi}^n_{l',l'}$ of the channel estimation error $vec(\widetilde{\mathbf{H}}^n_{l',l'})=vec(\mathbf{H}^n_{l',l'})-vec(\widehat{\mathbf{H}}^n_{l',l'})$ has the same form as

given in (13), and its U sub-matrices are given by

$$\left[\mathbf{\Xi}^n_{l',l'}\right]_{(u,u)} = \left(\beta^u_{l',l'} - \frac{(\beta^u_{l',l'})^2}{\frac{(L+1)(L^2-L+1)\sigma_w^2}{L^2\varepsilon^u_{l'}P} + \beta^u_{l',l'}}\right)\mathbf{I}_Q, \quad 1 \le u \le U. \tag{21}$$

C. Our grouping-based CE [17]

The UL pilot transmission consists of f time-shifted groups, and the MSs of all the cells transmit an OFDM PS to their serving BSs group-by-group. The FD PS of user u in the lth cell is denoted by $\underline{\mathbf{X}}^u_l = \left[X^u_l[1]\ X^u_l[2]\cdots X^u_l[N]\right]^{\mathrm{T}}$, where the power of each pilot $X^u_l[n]$ is unity. Let $H^{u,n}_{l,l',q}$ be the UL FDCHTF linking the uth user in cell l to the qth antenna of the l'th cell's BS, at the nth subcarrier. The signal $Y^{\bar{f}}_{l',q}[n]$ received by the qth receive antenna of the l'th BS at the nth subcarrier and received from the users in the $\bar{f}$th group is given by

$$Y^{\bar{f}}_{l',q}[n] = \sqrt{\rho_{\mathrm{TN}}} \underbrace{\sum\nolimits_{l=1}^{L}\sum\nolimits_{u=1}^{U}}_{(l,u)\in\bar{f}} \sqrt{\varepsilon^u_l} H^{u,n}_{l,l',q} X^u_l[n] + W^{\bar{f}}_{l',q}[n], \quad 1 \le l' \le L, 1 \le q \le Q, \tag{22}$$

where $\rho_{\mathrm{TN}} = \frac{P}{f}$, $(l,u) \in \bar{f}$ indicates that user u in cell l must belong to the group $\bar{f}$, and $1 \le \bar{f} \le f$, while $W^{\bar{f}}_{l',q}[n] \sim \mathcal{CN}(0,\sigma_w^2)$ is the FD AWGN. Collecting $Y^{\bar{f}}_{l',q}[n]$ of (22) for $1 \le n \le N$ leads to

$$\underline{\mathbf{Y}}^{\bar{f}}_{l',q} = \sqrt{\rho_{\mathrm{TN}}} \underbrace{\sum\nolimits_{l=1}^{L}\sum\nolimits_{u=1}^{U}}_{(l,u)\in\bar{f}} \sqrt{\varepsilon^u_l} \mathbf{X}^u_l \underline{\mathbf{H}}^u_{l,l',q} + \underline{\mathbf{W}}^{\bar{f}}_{l',q}, \quad 1 \le l' \le L, 1 \le q \le Q, \tag{23}$$

where $\mathbf{X}^u_l = \mathrm{diag}\{\underline{\mathbf{X}}^u_l\}$, while $\underline{\mathbf{Y}}^{\bar{f}}_{l',q} \in \mathbb{C}^N$, $\underline{\mathbf{H}}^u_{l,l,q} \in \mathbb{C}^N$ and $\underline{\mathbf{W}}^{\bar{f}}_{l',q} \in \mathbb{C}^N$ are the three vectors hosting $Y^{\bar{f}}_{l',q}[n]$, $H^{u,n}_{l,l,q}$ and $W^{\bar{f}}_{l',q}[n]$ for $1 \le n \le N$, respectively. It is worth emphasizing that unlike the conventional simultaneous CE case (4) and the successive CE case (17), the signal (23) is collected over all the N OFDM subcarriers for an individual BS antenna.

Denote the symbol-rate CIR for the channel linking the uth user in cell l to the qth antenna of the l'th cell's BS by

$$\underline{\mathbf{G}}^u_{l,l',q} = \left[G^u_{l,l',q}[1]\ G^u_{l,l',q}[2]\cdots G^u_{l,l',q}[K]\right]^{\mathrm{T}} \in \mathbb{C}^K, \tag{24}$$

where K is the maximum delay spread of the dispersive channel, and in practice $K \ll N$ [17]. Then $\underline{\mathbf{H}}^u_{l,l',q}$ is the N-point fast Fourier transform (FFT) of $\underline{\mathbf{G}}^u_{l,l',q}$ given by $\underline{\mathbf{H}}^u_{l,l',q} = \mathbf{F}\underline{\mathbf{G}}^u_{l,l',q}$, where $\mathbf{F} \in \mathbb{C}^{N\times K}$ is the FFT matrix, whose elements are given by $F_{n,k} = \frac{1}{\sqrt{K}} e^{-\mathrm{j}2\pi(n-1)(k-1)/N}$ for $1 \le n \le N$ and $1 \le k \le K$. Clearly, (23) can be rewritten as

$$\underline{\mathbf{Y}}^{\bar{f}}_{l',q} = \sqrt{\rho_{\mathrm{TN}}} \underbrace{\sum\nolimits_{l=1}^{L}\sum\nolimits_{u=1}^{U}}_{(l,u)\in\bar{f}} \sqrt{\varepsilon^u_l} \mathbf{X}^u_l \mathbf{F}\underline{\mathbf{G}}^u_{l,l',q} + \underline{\mathbf{W}}^{\bar{f}}_{l',q}. \tag{25}$$

We have the following theorem from [17].

Theorem 1: Design a FD PS matrix set for all the MSs in all the cells according to [25]

$$\mathbf{P} = \left\{\mathbf{X}^u_l, 1 \le u \le U, 1 \le l \le L\right\} = \left\{\mathbf{P}[i], 1 \le i \le LU\right\} = \left\{\mathbf{X}^1_1, \mathbf{X}^1_2, \cdots, \mathbf{X}^1_L; \mathbf{X}^2_1, \cdots, \mathbf{X}^2_L; \cdots; \mathbf{X}^U_1, \cdots, \mathbf{X}^U_L\right\}, \tag{26}$$

which contains the LU diagonal PS matrices, indexed by $1 \le i \le LU$ for the LU users:

$$\mathbf{P}[i] = \mathbf{P}[(u-1)L + l] = \mathbf{X}^u_l, \quad i = (u-1)L + l, 1 \le u \le U, 1 \le l \le L. \tag{27}$$

The ith element of this PS matrix set is generated from a reference $\mathbf{P}[1] = \mathbf{X}^1_1$ according to

$$\mathbf{P}[i] = \mathbf{\Phi}[i]\mathbf{P}[1],\ 1 \le i \le LU, \tag{28}$$

which shifts $\mathbf{P}[1]$ in phase by the diagonal matrix $\mathbf{\Phi}[i]$ parameterized by a positive integer ζ:

$$\mathbf{\Phi}[i] = \mathrm{diag}\left\{e^{\mathrm{j}2\pi\frac{(i-1)\zeta 0}{N}}, e^{\mathrm{j}2\pi\frac{(i-1)\zeta 1}{N}}, \cdots, e^{\mathrm{j}2\pi\frac{(i-1)\zeta(N-1)}{N}}\right\}. \quad 1 \le i \le LU. \tag{29}$$

If $\zeta = \lfloor \frac{N}{LU} \rfloor \ge K$, this FD PS matrix set offers the desired orthogonality as follows:

$$\left(\mathbf{P}[(u_1-1)L + l_1]\mathbf{F}\right)^{\mathrm{H}}\left(\mathbf{P}[(u_2-1)L + l_2]\mathbf{F}\right) = \begin{cases} \mathbf{0}_K, & l_1 \ne l_2 \cup u_1 \ne u_2, \\ \frac{N}{K}\mathbf{I}_K, & l_1 = l_2 \cap u_1 = u_2, \end{cases} \tag{30}$$

where $1 \le u_1, u_2 \le U$ and $1 \le l_1, l_2 \le L$. Based on this FD PS matrix set, we can always group all the LU users into the f groups, and the pilots assigned to the users of each group are orthogonal [17]. Moreover, $f \ll LU$, and in particular, if $KLU \le N$, we have $f = 1$.

Proof: See [17]. ∎

Clearly, $\mathbb{E}\left\{\underline{\mathbf{G}}^{u'}_{l',l',q}\right\} = \underline{\mathbf{0}}_K$, and the covariance matrix $\mathbf{R}^{u'}_{l',l',q} \in \mathbb{C}^{K\times K}$ of $\underline{\mathbf{G}}^{u'}_{l',l',q}$ is given by

$$\mathbf{R}^{u'}_{l',l',q} = \mathbb{E}\left\{\underline{\mathbf{G}}^{u'}_{l',l',q}\left(\underline{\mathbf{G}}^{u'}_{l',l',q}\right)^{\mathrm{H}}\right\} = \beta^{u'}_{l',l'}\mathbf{I}_K, \tag{31}$$

where $\beta^{u'}_{l',l'}$ is the link path-loss coefficient. The MMSE estimate $\widehat{\underline{\mathbf{G}}}^{u}_{l',l',q}$ of $\underline{\mathbf{G}}^{u}_{l',l',q}$ is given by

$$\begin{aligned}\widehat{\underline{\mathbf{G}}}^{u'}_{l',l',q} &= \beta^{u'}_{l',l'}\mathbf{I}_K\left(\frac{fK\sigma_w^2}{N\varepsilon^{u'}_{l'}P}\mathbf{I}_K + \beta^{u'}_{l',l'}\mathbf{I}_K\right)^{-1} \times \left(\underline{\mathbf{G}}^{u'}_{l',l',q} + \frac{K}{N\sqrt{\frac{\varepsilon^{u'}_{l'}P}{f}}}\left(\mathbf{X}^{u'}_{l'}\mathbf{F}\right)^{\mathrm{H}}\underline{\mathbf{W}}^{\bar{f}}_{l',q}\right) \\ &= \frac{\beta^{u'}_{l',l'}}{\frac{fK\sigma_w^2}{N\varepsilon^{u'}_{l'}P} + \beta^{u'}_{l',l'}}\left(\underline{\mathbf{G}}^{u'}_{l',l',q} + \frac{K}{N\sqrt{\frac{\varepsilon^{u'}_{l'}P}{f}}}\left(\mathbf{X}^{u'}_{l'}\mathbf{F}\right)^{\mathrm{H}}\underline{\mathbf{W}}^{\bar{f}}_{l',q}\right).\end{aligned} \tag{32}$$

Thus the estimate $\widehat{\underline{\mathbf{H}}}^{u'}_{l',l',q}$ of the FDCHTF vector $\underline{\mathbf{H}}^{u'}_{l',l',q}$ is

$$\widehat{\underline{\mathbf{H}}}^{u'}_{l',l',q} = \mathbf{F}\widehat{\underline{\mathbf{G}}}^{u'}_{l',l',q} = \frac{\beta^{u'}_{l',l'}}{\frac{fK\sigma_w^2}{N\varepsilon^{u'}_{l'}P} + \beta^{u'}_{l',l'}}\left(\underline{\mathbf{H}}^{u'}_{l',l',q} + \frac{K}{N\sqrt{\frac{\varepsilon^{u'}_{l'}P}{f}}}\mathbf{F}\left(\mathbf{X}^{u'}_{l'}\mathbf{F}\right)^{\mathrm{H}}\underline{\mathbf{W}}^{\bar{f}}_{l',q}\right). \tag{33}$$

The MMSE estimates $\widehat{\mathbf{H}}^n_{l',l'} \in \mathbb{C}^{Q\times U}$ for $1 \le n \le N$ can then be obtained given all the MMSE estimates $\widehat{\underline{\mathbf{H}}}^{u'}_{l',l',q}$ for

$1 \le u' \le U$ and $1 \le q \le Q$. More specifically,

$$vec\Big(\widehat{\mathbf{H}}^n_{l',l'}\Big) = \mathbf{\Upsilon}^n_{l',l'}\bigg(vec(\mathbf{H}^n_{l',l'}) + \frac{1}{\sqrt{\frac{N}{fK}P}} vec\Big(\overline{\mathbf{W}}_{l'}[n]\mathbf{\Delta}_{l'}^{-1}\Big)\bigg), \tag{34}$$

where the elements of $\overline{\mathbf{W}}_{l'}[n] \in \mathbb{C}^{Q\times U}$ are white and obey the distribution $\mathcal{CN}(0,\sigma_w^2)$, while $\mathbf{\Upsilon}^n_{l',l'} \in \mathbb{C}^{QU\times QU}$ takes the same form of (8) and its U sub-matrices are given by

$$[\mathbf{\Upsilon}^n_{l',l'}]_{(u,u)} = \frac{\beta^u_{l',l'}}{\frac{fK\sigma_w^2}{N\varepsilon^u_{l'}P} + \beta^u_{l',l'}}\mathbf{I}_Q,\ 1 \le u \le U. \tag{35}$$

The distribution of this MMSE estimator is $vec\big(\widehat{\mathbf{H}}^n_{l',l'}\big) \sim \mathcal{CN}\big(\underline{\mathbf{0}}_{QU}, \mathbf{\Phi}^n_{l',l'}\big)$, and the covariance matrix $\mathbf{\Phi}^n_{l',l'}$ has the same form of (11) with its U sub-matrices given by

$$[\mathbf{\Phi}^n_{l',l'}]_{(u,u)} = \frac{\big(\beta^u_{l',l'}\big)^2}{\frac{fK\sigma_w^2}{N\varepsilon^u_{l'}P} + \beta^u_{l',l'}}\mathbf{I}_Q,\ 1 \le u \le U. \tag{36}$$

Finally, the covariance matrix $\mathbf{\Xi}^n_{l',l'}$ of the channel estimation error $vec\big(\widetilde{\mathbf{H}}^n_{l',l'}\big) = vec\big(\mathbf{H}^n_{l',l'}\big) - vec\big(\widehat{\mathbf{H}}^n_{l',l'}\big)$ also takes the form of (13) and its U sub-matrices are given by

$$[\mathbf{\Xi}^n_{l',l'}]_{(u,u)} = \bigg(\beta^u_{l',l'} - \frac{\big(\beta^u_{l',l'}\big)^2}{\frac{fK\sigma_w^2}{N\varepsilon^u_{l'}P} + \beta^u_{l',l'}}\bigg)\mathbf{I}_Q,\ 1 \le u \le U. \tag{37}$$

D. Summary

For these three PC-free CE methods, the MMSE estimate $\widehat{\mathbf{H}}^n_{l',l'}$ of $\mathbf{H}^n_{l',l'}$ takes the form

$$vec\Big(\widehat{\mathbf{H}}^n_{l',l'}\Big) = \mathbf{\Upsilon}^n_{l',l'}\bigg(vec\big(\mathbf{H}^n_{l',l'}\big) + \frac{1}{\sqrt{\varrho_1 P}} vec\Big(\check{\mathbf{W}}_{l'}[n]\mathbf{\Delta}_{l'}^{-1}\Big)\bigg), \tag{38}$$

where the elements of $\check{\mathbf{W}}_{l'}[n] \in \mathbb{C}^{Q\times U}$ are white Gaussian variables obeying the distribution $\mathcal{CN}\left(0, \varrho_2\sigma_w^2\right)$, and $\mathbf{\Upsilon}^n_{l',l'} \in \mathbb{C}^{QU\times QU}$ takes the form of (8) with its U sub-matrices given by

$$[\mathbf{\Upsilon}^n_{l',l'}]_{(u,u)} = \frac{\beta^u_{l',l'}}{\varrho_3\frac{\sigma_w^2}{\varepsilon^u_{l'}P} + \beta^u_{l',l'}}\mathbf{I}_Q,\ 1 \le u \le U. \tag{39}$$

The distribution of the MMSE estimator is $vec\big(\widehat{\mathbf{H}}^n_{l',l'}\big) \sim \mathcal{CN}(\underline{\mathbf{0}}_{QU}, \mathbf{\Phi}^n_{l',l'})$, and the covariance matrix $\mathbf{\Phi}^n_{l',l'} \in \mathbb{C}^{QU\times QU}$ takes the form of (11) with the U sub-matrices specified by

$$[\mathbf{\Phi}^n_{l',l'}]_{(u,u)} = \frac{\big(\beta^u_{l',l'}\big)^2}{\varrho_3\frac{\sigma_w^2}{\varepsilon^u_{l'}P} + \beta^u_{l',l'}}\mathbf{I}_Q,\ 1 \le u \le U. \tag{40}$$

The covariance matrix $\mathbf{\Xi}^n_{l',l'} \in \mathbb{C}^{QU\times QU}$ of the channel estimation error $vec\big(\widetilde{\mathbf{H}}^n_{l',l'}\big) = vec\big(\mathbf{H}^n_{l',l'}\big) - vec\big(\widehat{\mathbf{H}}^n_{l',l'}\big)$ takes the form of (13) with its U sub-matrices given by

$$[\mathbf{\Xi}^n_{l',l'}]_{(u,u)} = \bigg(\beta^u_{l',l'} - \frac{\big(\beta^u_{l',l'}\big)^2}{\varrho_3\frac{\sigma_w^2}{\varepsilon^u_{l'}P} + \beta^u_{l',l'}}\bigg)\mathbf{I}_Q,\ 1 \le u \le U. \tag{41}$$

For these three estimators, we have

$$\varrho_1 = \begin{cases} LU, & \text{simultaneous CE,} \\ \frac{L^4U}{(L+1)}, & \text{successive CE,} \\ \frac{N}{fK}, & \text{our grouping based CE,} \end{cases} \tag{42}$$

$$\varrho_2 = \begin{cases} LU, & \text{simultaneous CE,} \\ L^2U(L^2-L+1), & \text{successive CE,} \\ 1, & \text{our grouping based CE,} \end{cases} \tag{43}$$

and furthermore $\varrho_3 = \frac{\varrho_2}{\varrho_1}$.

IV. Uplink and Downlink Transmissions

A. Uplink Transmission

Each BS performs the maximum ratio combining (MRC) on its received signals subcarrier-by-subcarrier. Thus we can drop the subcarrier superscript n in $\widehat{\mathbf{H}}^n_{l',l'}$, $\mathbf{\Phi}^n_{l',l'}$, $\mathbf{R}^n_{l',l'}$ and $\mathbf{\Xi}^n_{l',l'}$. At the subcarrier considered, let the tth information-bearing symbols transmitted by the U users of cell l be $\underline{\mathbf{C}}_l[t] = \big[C^1_l[t]\ C^2_l[t]\cdots C^U_l[t]\big]^{\mathrm{T}}$, $1 \le t \le N_{\mathrm{UL}}$, where $\mathbb{E}\big\{\underline{\mathbf{C}}_l[t]\big\} = \mathbf{0}_U$ and $\mathbb{E}\big\{\underline{\mathbf{C}}_l[t]\underline{\mathbf{C}}_l^{\mathrm{H}}[t]\big\} = \mathbf{I}_U$. Further denote $\mathbf{C}_l = \big[\underline{\mathbf{C}}_l[1]\ \underline{\mathbf{C}}_l[2]\cdots\underline{\mathbf{C}}_l[N_{\mathrm{UL}}]\big] \in \mathbb{C}^{U\times N_{\mathrm{UL}}}$, and define

$$\gamma^u_l = 1 - \varepsilon^u_l,\ 1 \le l \le L,\ 1 \le u \le U. \tag{44}$$

Then the signal matrix $\mathbf{Z}^{\mathrm{UL}}_{l'} \in \mathbb{C}^{Q\times N_{\mathrm{UL}}}$ received by the l'th cell's BS can be expressed as

$$\mathbf{Z}^{\mathrm{UL}}_{l'} = \sqrt{\rho_{\mathrm{UL}}}\sum\nolimits_{l=1}^{L}\mathbf{H}_{l,l'}\mathbf{\Gamma}_{\boldsymbol{l}}\mathbf{C}_l + \mathbf{V}_{l'},\ 1 \le l' \le L, \tag{45}$$

where $\rho_{\mathrm{UL}} = \frac{P}{N_{\mathrm{UL}}}$, and $\mathbf{\Gamma}_l = \mathrm{diag}\Big\{\big[\sqrt{\gamma^1_l}\ \sqrt{\gamma^2_l}\cdots\sqrt{\gamma^U_l}\big]^{\mathrm{T}}\Big\}$, while $\mathbf{V}_{l'} \in \mathbb{C}^{Q\times N_{\mathrm{UL}}}$ is the FD UL channel AWGN vector, whose elements obey the distribution $\mathcal{CN}(0,\sigma_v^2)$. The l'th BS performs the MRC on $\mathbf{Z}^{\mathrm{UL}}_{l'}$ with the multi-user detection matrix $\mathbf{A}_{l'} = \widehat{\mathbf{H}}^{\mathrm{H}}_{l',l'} \in \mathbb{C}^{U\times Q}$ according to:

$$\mathbf{Y}_{l'} = \mathbf{A}_{l'}\mathbf{Z}^{\mathrm{UL}}_{l'} = \sqrt{\rho_{\mathrm{UL}}}\sum\nolimits_{l=1}^{L}\mathbf{A}_{l'}\mathbf{H}_{l,l'}\mathbf{\Gamma}_{\boldsymbol{l}}\mathbf{C}_l + \mathbf{A}_{l'}\mathbf{V}_{l'}. \tag{46}$$

Denote the u'th row of $\mathbf{A}_{l'}$, i.e., the u'th row of $\widehat{\mathbf{H}}^{\mathrm{H}}_{l',l'}$, by $\underline{\mathbf{A}}^{u'}_{l'}$, the uth column of $\mathbf{H}_{l,l'}$ by $\underline{\mathbf{H}}^u_{l,l'}$, and the tth column of $\mathbf{V}_{l'}$ by $\underline{\mathbf{V}}^t_{l'}$. Then the tth signal of the u'th MS in cell l' received by the l'th cell's BS is given by

$$Y^{u'}_{l'}[t] = \sqrt{\rho_{\mathrm{UL}}}\sum_{l=1}^{L}\sum_{u=1}^{U}\underline{\mathbf{A}}^{u'}_{l'}\underline{\mathbf{H}}^u_{l,l'}\sqrt{\gamma^u_l}C^u_l[t] + \underline{\mathbf{A}}^{u'}_{l'}\underline{\mathbf{V}}^t_{l'}, \tag{47}$$

which can be expressed as (48)

$$\begin{aligned} Y^{u'}_{l'}[t] = {} & \sqrt{\rho_{\mathrm{UL}}}\underline{\mathbf{A}}^{u'}_{l'}\widehat{\underline{\mathbf{H}}}^{u'}_{l',l'}\sqrt{\gamma^{u'}_{l'}}C^{u'}_{l'}[t] + \underbrace{\sqrt{\rho_{\mathrm{UL}}}\underline{\mathbf{A}}^{u'}_{l'}\widetilde{\underline{\mathbf{H}}}^{u'}_{l',l'}\sqrt{\gamma^{u'}_{l'}}C^{u'}_{l'}[t]}_{\text{Interference by estimation error}} \\ & \underbrace{+\sqrt{\rho_{\mathrm{UL}}}\sum\nolimits_{u=1,u\neq u'}^{U}\underline{\mathbf{A}}^{u'}_{l'}\underline{\mathbf{H}}^u_{l',l'}\sqrt{\gamma^u_{l'}}C^u_{l'}[t]}_{\text{Intra-cell interference}} \\ & \underbrace{+\sqrt{\rho_{\mathrm{UL}}}\sum_{l=1,l\neq l'}^{L}\sum_{u=1}^{U}\underline{\mathbf{A}}^{u'}_{l'}\underline{\mathbf{H}}^u_{l,l'}\sqrt{\gamma^u_l}C^u_l[t]}_{\text{Inter-cell interference}} + \underline{\mathbf{A}}^{u'}_{l'}\underline{\mathbf{V}}^t_{l'}. \end{aligned} \tag{48}$$

$$\text{SINR-UL}_{l'}^{u'}=\frac{\mathbb{E}\Big\{\Big|\sqrt{\gamma_{l'}^{u'}}\underline{\mathbf{A}}_{l'}^{u'}\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\Big|^2\Big\}}{\mathbb{E}\Big\{\Big|\sqrt{\gamma_{l'}^{u'}}\underline{\mathbf{A}}_{l'}^{u'}\widetilde{\underline{\mathbf{H}}}_{l',l'}^{u'}\Big|^2\Big\}+\sum\limits_{u=1,u\neq u'}^{U}\mathbb{E}\Big\{\Big|\sqrt{\gamma_{l'}^{u}}\underline{\mathbf{A}}_{l'}^{u'}\underline{\mathbf{H}}_{l',l'}^{u}\Big|^2\Big\}+\sum\limits_{l=1,l\neq l'}^{L}\sum\limits_{u=1}^{U}\mathbb{E}\Big\{\Big|\sqrt{\gamma_{l}^{u}}\underline{\mathbf{A}}_{l'}^{u'}\underline{\mathbf{H}}_{l,l'}^{u}\Big|^2\Big\}+\frac{1}{\rho_{\mathrm{UL}}}\mathbb{E}\Big\{\Big|\underline{\mathbf{A}}_{l'}^{u'}\underline{\mathbf{V}}_{l'}^{t}\Big|^2\Big\}}. \quad (49)$$

where $\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}$ and $\widetilde{\underline{\mathbf{H}}}_{l',l'}^{u'}$ are the u'th columns of $\widehat{\mathbf{H}}_{l',l'}^{u'}$ and $\widetilde{\mathbf{H}}_{l',l'}^{u'}$, respectively. The UL SINR of the u'th user in the l'th cell is therefore given by (49) at the top of the next page.

Theorem 2: The asymptotical $\text{SINR-UL}_{l'}^{u'}$ is a function of $\underline{\boldsymbol{\varepsilon}}\in\mathbb{R}^{LU}$ given by

$$\begin{aligned}\text{SINR-UL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon}})=&\frac{\frac{Q\left(\beta_{l',l'}^{u'}\right)^2}{\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}}}{\frac{\overline{\beta}}{1-\varepsilon_{l'}^{u'}}-\frac{\left(\beta_{l',l'}^{u'}\right)^2}{\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}}+\frac{\mu}{1-\varepsilon_{l'}^{u'}}}\\=&\frac{f_{1,l',u'}^{\mathrm{UL}}\left(\varepsilon_{l'}^{u'}\right)}{f_{2,l',u'}^{\mathrm{UL}}(\underline{\boldsymbol{\varepsilon}})},\end{aligned} \quad (50)$$

where ε_l^u for $1\le l\le L$ and $1\le u\le U$ are the elements of $\underline{\boldsymbol{\varepsilon}}$, and

$$\varrho=\varrho_3\frac{\sigma_w^2}{P}, \quad (51)$$

$$\mu=\frac{\sigma_v^2N_{\mathrm{UL}}}{P}, \quad (52)$$

$$\overline{\beta}=\sum\nolimits_{l=1}^{L}\sum\nolimits_{u=1}^{U}(1-\varepsilon_l^u)\beta_{l,l'}^u. \quad (53)$$

.

Proof: See Appendix B. ∎

Then the asymptotically achievable UL rate is defined as

$$\text{R-UL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon}})=\frac{N_{\mathrm{UL}}}{\frac{1}{2}N_{\mathrm{TN}}+N_{\mathrm{UL}}}\log_2\Big(1+\text{SINR-UL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon}})\Big). \quad (54)$$

B. Downlink Transmission

Let the tth information-bearing symbols at the subcarrier considered transmitted to the U users of cell l be $\underline{\mathbf{D}}_l[t]=\left[D_l^1[t]\ D_l^2[t]\cdots D_l^U[t]\right]^{\mathrm{T}}$, $1\le t\le N_{\mathrm{DL}}$, where $\mathbb{E}\{\underline{\mathbf{D}}_l[t]\}=\underline{\mathbf{0}}_U$ and $\mathbb{E}\{\underline{\mathbf{D}}_l[t]\underline{\mathbf{D}}_l^{\mathrm{H}}[t]\}=\mathbf{I}_U$. The lth cell's BS first performs the transmit precoding (TP) on $\underline{\mathbf{D}}_l[t]$ with the TP matrix $\mathbf{B}_l\in\mathbb{C}^{Q\times U}$. It then transmits the signal $\mathbf{B}_l\underline{\mathbf{D}}_l[t]$ in the DL to its U MSs. The signals $\underline{\mathbf{Z}}_{l'}[t]\in\mathbb{C}^U$ received by the U MSs in cell l' is given by

$$\underline{\mathbf{Z}}_{l'}[t]=\sqrt{p_{\mathrm{DL}}}\sum\nolimits_{l=1}^{L}\mathbf{H}_{l',l}^{\mathrm{H}}\mathbf{B}_l\underline{\mathbf{D}}_l[t]+\underline{\mathbf{J}}_{l'}[t], \quad (55)$$

where $\underline{\mathbf{J}}_{l'}[t]\in\mathbb{C}^U$ is the FD DL channel AWGN vector, whose elements obey the distribution $\mathcal{CN}(0,\sigma_J^2)$. Thus, the signal received by the u'th MS in cell l' is given by (56) at the bottom of this page, where $\underline{\mathbf{B}}_l^u$ is the uth column of $\mathbf{B}_l$ and $J_{l'}^{u'}[t]$ is the u'th element of $\underline{\mathbf{J}}_{l'}[t]$.

Assume that the TP is based on the matched filter (MF) criterion, i.e., $\mathbf{B}_{l'}=\sqrt{\lambda_{l'}}\widehat{\mathbf{H}}_{l',l'}$ with

$$\lambda_{l'}=\frac{U}{\mathbb{E}\{\mathrm{Tr}\{\widehat{\mathbf{H}}_{l',l'}(\widehat{\mathbf{H}}_{l',l'})^{\mathrm{H}}\}\}}. \quad (57)$$

The DL SINR of the u'th user in the l'th cell is given by (58) at the bottom of this page, where $\mathrm{var}\{\cdot\}$ is the variance operator, $\text{SP-DL}_{l'}^{u'}$ and $\text{INP-DL}_{l'}^{u'}$ are the DL desired signal power and the interference plus noise power, respectively.

Theorem 3: The asymptotical $\text{SINR-DL}_{l'}^{u'}$ is a function of $\underline{\boldsymbol{\varepsilon_{l'}}}\in\mathbb{R}^U$ given by

$$\begin{aligned}\text{SINR-DL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon_{l'}}})=&\frac{\bar{\lambda}_{l'}\left(\frac{Q\left(\beta_{l',l'}^{u'}\right)^2}{\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}}\right)^2}{U\sum_{l=1}^{L}\beta_{l',l}^{u'}-\frac{\bar{\lambda}_{l'}Q(\beta_{l',l'}^{u'})^4}{\left(\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}\right)^2}+\frac{\sigma_J^2}{p_{\mathrm{DL}}}}\\=&\frac{\text{SP-DL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon_{l'}}})}{\text{INP-DL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon_{l'}}})},\end{aligned} \quad (59)$$

where $\varepsilon_{l'}^u$ for $1\le u\le U$ are the elements of $\underline{\boldsymbol{\varepsilon_{l'}}}$, and $\bar{\lambda}_{l'}$ is the asymptotic value of $\lambda_{l'}$, which is defined by

$$\bar{\lambda}_{l'}=\frac{1}{\frac{Q}{U}\sum_{u=1}^{U}\frac{\left(\beta_{l',l'}^{u}\right)^2}{\varrho(\varepsilon_{l'}^{u})^{-1}+\beta_{l',l'}^{u}}}. \quad (60)$$

Proof: See Appendix C. ∎

Given $\text{SINR-DL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon_{l'}}})$, the asymptotically achievable DL rate is readily formulated as

$$\text{R-DL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon_{l'}}})=\frac{N_{\mathrm{UL}}}{\frac{1}{2}N_{\mathrm{TN}}+N_{\mathrm{UL}}}\log_2\Big(1+\text{SINR-DL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon_{l'}}})\Big). \quad (61)$$

$$\begin{aligned}Z_{l'}^{u'}[t]=&\sqrt{p_{\mathrm{DL}}}\sum_{l=1}^{L}\sum_{u=1}^{U}\left(\underline{\mathbf{H}}_{l',l}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_l^uD_l^u[t]+J_{l'}^{u'}[t]=\sqrt{p_{\mathrm{DL}}}\mathbb{E}\left\{\left(\underline{\mathbf{H}}_{l',l'}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}\right\}D_{l'}^{u'}[t]+\underbrace{\sqrt{p_{\mathrm{DL}}}\left(\left(\underline{\mathbf{H}}_{l',l'}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}-\mathbb{E}\left\{\left(\underline{\mathbf{H}}_{l',l'}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}\right\}\right)D_{l'}^{u'}[t]}_{\text{Self interference}}\\&+\underbrace{\sqrt{p_{\mathrm{DL}}}\sum_{u=1,u\neq u'}^{U}\left(\underline{\mathbf{H}}_{l',l'}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u}D_{l'}^{u}[t]}_{\text{Intra-cell interference}}+\underbrace{\sqrt{p_{\mathrm{DL}}}\sum_{l=1,l\neq l'}^{L}\sum_{u=1}^{U}\left(\underline{\mathbf{H}}_{l',l}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_{l}^{u}D_{l}^{u}[t]}_{\text{Inter-cell interference}}+J_{l'}^{u'}[t],\end{aligned} \quad (56)$$

$$\text{SINR-DL}_{l'}^{u'}=\frac{\left|\mathbb{E}\left\{\left(\underline{\mathbf{H}}_{l',l'}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}\right\}\right|^2}{\mathrm{var}\left\{\left(\underline{\mathbf{H}}_{l',l'}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}\right\}+\sum\limits_{u=1,u\neq u'}^{U}\mathbb{E}\left\{\left|\left(\underline{\mathbf{H}}_{l',l'}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u}\right|^2\right\}+\sum\limits_{l=1,l\neq l'}^{L}\sum\limits_{u=1}^{U}\mathbb{E}\left\{\left|\left(\underline{\mathbf{H}}_{l',l}^{u'}\right)^{\mathrm{H}}\underline{\mathbf{B}}_{l}^{u}\right|^2\right\}+\frac{\sigma_J^2}{p_{\mathrm{DL}}}}=\frac{\text{SP-DL}_{l'}^{u'}}{\text{INP-DL}_{l'}^{u'}}, \quad (58)$$

Algorithm 1 Find Optimal UL Power Allocation Solution $\underline{\boldsymbol{\varepsilon}}^{\star}$

1: Set iteration index Ite $= 0$ and initial $\underline{\boldsymbol{\varepsilon}} = \{\underline{\boldsymbol{\varepsilon}}[i], 1 \le i \le LU\} = \{\varepsilon_l^u = 0.5, 1 \le u \le U, 1 \le l \le L\}$.
2: **While** Ite $<$ Ite$_{\max}$ **do**
3: $\quad \underline{\boldsymbol{\varepsilon}}^{\star} = \underline{\boldsymbol{\varepsilon}}$.
4: $\quad$ **for** $i = 1$ to LU **do**
5: $\quad\quad$ Find optimal individual $\underline{\boldsymbol{\varepsilon}}[i]$ using golden section.
6: $\quad$ **If** $\underline{\boldsymbol{\varepsilon}} == \underline{\boldsymbol{\varepsilon}}^{\star}$
7: $\quad\quad$ Break.

C. Convexity Analysis of Aggregation Rate

The average effective AR per cell is a function of $\underline{\boldsymbol{\varepsilon}}$, which is given by

$$\mathrm{AR}(\underline{\boldsymbol{\varepsilon}}) = \frac{N_{\mathrm{UL}}}{\left(\frac{1}{2}N_{\mathrm{TN}} + N_{\mathrm{UL}}\right)L} \sum_{l'=1}^{L} \sum_{u'=1}^{U} \Big(\log_2\left(1 + \text{SINR-UL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon}})\right) + \log_2\left(1 + \text{SINR-DL}_{l'}^{u'}(\underline{\boldsymbol{\varepsilon}_{l'}})\right)\Big). \quad (62)$$

Theorem 4: Given ε_l^u for $1 \le u \le U$ and $1 \le l \le L$ with $u \ne u^*$ and $l \ne l^*$, we have: *a).* when $u^* = u'$ and $l^* = l'$, $\text{SINR-UL}_{l'}^{u'}(\varepsilon_{l^*}^{u^*})$ is a convex function of $\varepsilon_{l^*}^{u^*}$ for $0 < \varepsilon_{l^*}^{u^*} < 1$, and *b).* when $u^* \ne u'$ and/or $l^* \ne l'$, $\text{SINR-UL}_{l'}^{u'}(\varepsilon_{l^*}^{u^*})$ is an increasing function of $\varepsilon_{l^*}^{u^*}$ for $0 < \varepsilon_{l^*}^{u^*} < 1$.

Proof: See Appendix D. ∎

Theorem 5: Given ε_l^u for $1 \le u \le U$ and $1 \le l \le L$ with $u \ne u^*$ and $l \ne l^*$, we have: *a).* when $u^* = u'$ and $l^* = l'$, $\text{SINR-DL}_{l'}^{u'}(\varepsilon_{l^*}^{u^*})$ is an increasing function of $\varepsilon_{l^*}^{u^*}$ for $0 < \varepsilon_{l^*}^{u^*} < 1$, and *b).* when $u^* \ne u'$ and/or $l^* \ne l'$, $\text{SINR-DL}_{l'}^{u'}$ does not depend on $\varepsilon_{l^*}^{u^*}$.

Proof: See Appendix E. ∎

Theorem 6: Given ε_l^u for $1 \le u \le U$ and $1 \le l \le L$ with $u \ne u^*$ and $l \ne l^*$, $\mathrm{AR}(\varepsilon_{l^*}^{u^*})$ of (62) is a convex function of $\varepsilon_{l^*}^{u^*}$ for $0 < \varepsilon_{l^*}^{u^*} < 1$.

Proof: Because the logarithmic and summation functions do not change the convexity and monotonicity, for $0 < \varepsilon_{l^*}^{u^*} < 1$, $\mathrm{AR}(\varepsilon_{l^*}^{u^*})$ is obviously a convex function of $\varepsilon_{l^*}^{u^*}$ based on Theorems 4 and 5. ∎

The optimal UL power allocation solution that maximizes the effective AR is defined as

$$\underline{\boldsymbol{\varepsilon}}^{\star} = \arg \max_{\underline{\boldsymbol{\varepsilon}}:\ 0 < \varepsilon_l^u < 1, 1 \le u \le U, 1 \le l \le L} \mathrm{AR}(\underline{\boldsymbol{\varepsilon}}). \quad (63)$$

Since the closed-form expression for $\mathrm{AR}(\underline{\boldsymbol{\varepsilon}})$ is available and, furthermore, owing to its piecewise convexity property as proved in Theorem 6, the unique global optimal solution $\underline{\boldsymbol{\varepsilon}}^{\star}$ can be obtained very fast using the iterative procedure given in Algorithm 1, which involves the golden section method to find the optimal individual ε_l^u one by one in each iteration. In fact, Algorithm 1 is capable of converging to $\underline{\boldsymbol{\varepsilon}}^{\star}$ in one iteration. This is because physically, the optimal UL pilot-data power allocation factors $\left(\varepsilon_l^u\right)^{\star}$ are unlikely to be very small (near 0) or very large (near 1), since the former degrades the estimation accuracy too much and the latter leaves too little power for UL transmission, both being detrimental to the achievable AR. In other words, $\underline{\boldsymbol{\varepsilon}}^{\star}$ is not far away from $\left(\underline{\boldsymbol{\varepsilon}}\right)^{(0)} = \underline{\boldsymbol{\varepsilon}}^{\dagger}$. This together with the piecewise convexity property of the objective ensures that after just one iteration, $\left(\underline{\boldsymbol{\varepsilon}}\right)^{(1)}$ must be at least very close to $\underline{\boldsymbol{\varepsilon}}^{\star}$. In all our extensive investigation experiments, Algorithm 1 always finds $\underline{\boldsymbol{\varepsilon}}^{\star}$ with only one iteration.

TABLE II
DEFAULT PARAMETERS USED IN THE SIMULATED MULTIPLE-ANTENNA AIDED AND TDD BASED OFDM NETWORK

Number of cells L	7
Radius of each cell r_c	500 m
Radius of cell hole r_h	50 m
Number of MSs per-cell U	8
Number of antennas at each BS Q	100
Total UL power of each user at each subcarrier P	10 dB
DL transmission power per BS antenna at each subcarrier P_{DL}	20 dB
Path loss exponent τ	3
Mean of path AOAs θ	90°
Standard deviation of path AOAs σ_{AOA}	90°
Antenna spacing D	$\frac{\lambda}{2}$
Length of CIRs K	54
Number of subcarriers N	1024
COHI r'	84

V. Simulation Results

The default values of the parameters in our simulated network are listed in Table II. Unless otherwise specified, these default parameters are used. In the simulation, a uniformly spaced linear antenna array is assumed at each BS, and the CIR vector $\underline{\mathbf{G}}_{l,l',q}^{u}$ of (24) is specified by

$$G_{l,l',q}^{u}[k] = \alpha_{l,l',q,k}^{u} e^{-\mathrm{j}\frac{2\pi(q-1)D}{\lambda}} \cos\left(\theta_{l,l',q,k}^{u}\right), 1 \le k \le K, \quad (64)$$

where D and λ are the antenna spacing and the carrier's wavelength, respectively, and all the angle of arrivals (AOAs) $\theta_{l,l',q,k}^{u}$ are Gaussian random variables with mean $\bar{\theta}$ and standard deviation σ_{AOA}, while $\alpha_{l,l',q,k}^{u} = \sqrt{\beta_{l,l'}^{u}} e^{-\mathrm{j}\varphi_{l,l',q,k}^{u}}$, in which all the $\varphi_{l,l',q,k}^{u}$ are random variables uniformly distributed in $[0,\ 2\pi)$. MSs are randomly located in each cell. The path-loss coefficient is defined by $\beta_{l,l'}^{u} = \left(\frac{d_{l,l'}^{u}}{r_h}\right)^{-\tau}$ [5], where τ is the path-loss exponent and $d_{l,l'}^{u}$ denotes the distance from user u in cell l to the BS of cell l', while r_h is the radius of 'cell hole' around BS. Clearly, this is a system with $N < KLU$, and our 4-group based CE can completely eliminate the PC with $N_{\mathrm{TN}} = f = 4$ [17]. This leads to $N_{\mathrm{UL}} = N_{\mathrm{DL}} = 40$. For the conventional simultaneous CE, to completely eliminate the PC requires $N_{\mathrm{TN}} = LU = 56$ and this leaves $N_{\mathrm{UL}} = N_{\mathrm{DL}} = 14$. By comparison, the successive CE [14] requires $N_{\mathrm{TN}} = (L+1)U = 64$, leaving only $N_{\mathrm{UL}} = N_{\mathrm{DL}} = 10$. We define the SNR as E_s/N_0, where E_s denotes the energy per symbol and N_0 denotes the power of the AWGN. The UL training SNR is set to equal to the UL transmission SNR in all the simulations.

We compare the maximum effective AR performance $\mathrm{AR}(\underline{\boldsymbol{\varepsilon}}^{\star})$ obtained by solving the optimization (63) using Algorithm 1 with the non-optimal $\mathrm{AR}(\underline{\boldsymbol{\varepsilon}}^{\dagger})$ achieved by the standard UL pilot-data power allocation that divides the UL power equally between UL training and transmission. Hence, all the elements of $\underline{\boldsymbol{\varepsilon}}^{\dagger} \in \mathbb{R}^{LU}$ are 0.5. The results obtained for the three CE schemes under various SNR conditions are given in Tables III to V and Figs. 2 to 4. Specifically, Table III to V compere the optimal $\mathrm{AR}(\underline{\boldsymbol{\varepsilon}}^{\star})$ and the non-optimal $\mathrm{AR}(\underline{\boldsymbol{\varepsilon}}^{\dagger})$ for the three CE schemes, respectively, while Fig. 2 to 4 depict the

TABLE III
COMPARISON OF OPTIMAL $\mathrm{AR}(\underline{\varepsilon}^{\star})$ AND NON-OPTIMAL $\mathrm{AR}(\underline{\varepsilon}^{\dagger})$ FOR OUR USER-GROUPING BASED CE [17]

UL SNR \ DL SNR	0 dB	10 dB	20 dB	30 dB
0 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 11.205$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 12.995$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 13.348$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 13.389$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 11.535$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 13.614$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 14.067$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 14.122$
10 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 19.894$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 23.140$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 24.052$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 24.181$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 20.278$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 23.471$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 24.447$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 24.596$
20 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 27.840$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 31.954$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 33.469$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 33.719$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 30.497$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 34.447$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 35.867$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 36.100$
30 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 30.115$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 34.436$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 36.127$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 36.415$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 35.572$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 39.825$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 41.460$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 41.737$

TABLE IV
COMPARISON OF OPTIMAL $\mathrm{AR}(\underline{\varepsilon}^{\star})$ AND NON-OPTIMAL $\mathrm{AR}(\underline{\varepsilon}^{\dagger})$ FOR THE CONVENTIONAL SIMULTANEOUS CE

UL SNR \ DL SNR	0 dB	10 dB	20 dB	30 dB
0 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 3.498$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 3.633$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 3.651$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 3.653$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 3.625$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 3.813$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 3.839$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 3.842$
10 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 7.434$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 7.881$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 7.951$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 7.959$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 7.461$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 7.935$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 8.014$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 8.023$
20 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 10.417$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 11.242$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 11.423$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 11.445$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 11.355$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 12.164$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 12.343$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 12.365$
30 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 11.311$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 12.362$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 12.635$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 12.671$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 13.087$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 14.101$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 14.360$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 14.393$

TABLE V
COMPARISON OF OPTIMAL $\mathrm{AR}(\underline{\varepsilon}^{\star})$ AND NON-OPTIMAL $\mathrm{AR}(\underline{\varepsilon}^{\dagger})$ FOR THE SUCCESSIVE CE [14]

UL SNR \ DL SNR	0 dB	10 dB	20 dB	30 dB
0 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 1.419$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 1.444$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 1.447$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 1.447$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 1.514$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 1.547$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 1.551$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 1.552$
10 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 3.800$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 3.893$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 3.905$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 3.906$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 3.939$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 4.053$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 4.068$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 4.069$
20 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 6.470$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 6.758$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 6.802$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 6.807$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 6.987$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 7.299$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 7.349$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 7.354$
30 dB	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 7.879$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 8.411$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 8.519$	$\mathrm{AR}(\underline{\varepsilon}^{\dagger}) = 8.532$
	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 8.874$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 9.397$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 9.505$	$\mathrm{AR}(\underline{\varepsilon}^{\star}) = 9.517$

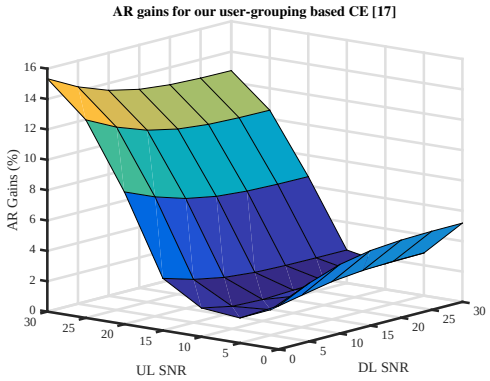

Fig. 2. Per cell effective AR gains of optimal $\mathrm{AR}(\underline{\varepsilon}^{\star})$ over non-optimal $\mathrm{AR}(\underline{\varepsilon}^{\dagger})$ for our user-grouping based CE [17].

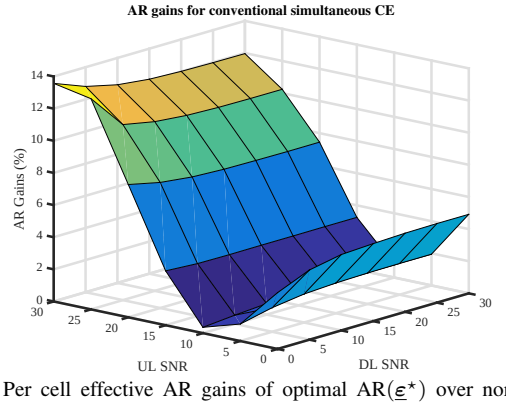

Fig. 3. Per cell effective AR gains of optimal $\mathrm{AR}(\underline{\varepsilon}^{\star})$ over non-optimal $\mathrm{AR}(\underline{\varepsilon}^{\dagger})$ for the conventional simultaneous CE.

achievable per cell effective AR gains of the optimal $\mathrm{AR}(\underline{\varepsilon}^{\star})$ over the non-optimal $\mathrm{AR}(\underline{\varepsilon}^{\dagger})$, defined as

$$\text{AR Gain} = \frac{\mathrm{AR}(\underline{\varepsilon}^{\star}) - \mathrm{AR}(\underline{\varepsilon}^{\dagger})}{\mathrm{AR}(\underline{\varepsilon}^{\star})}, \tag{65}$$

for the three CE schemes respectively.

Observe from Fig. 2 that the gains in the effective AR achieved by the proposed UL power allocation optimization range from 1.41% to 15.34% for our grouping based CE. From Fig. 3, it is seen that the gains in the effective AR achieved by the UL power allocation optimization are in 0.36% to 13.57% for the conventional simultaneous CE. From Fig. 4, we observe that the gains in the effective AR are in 3.53%

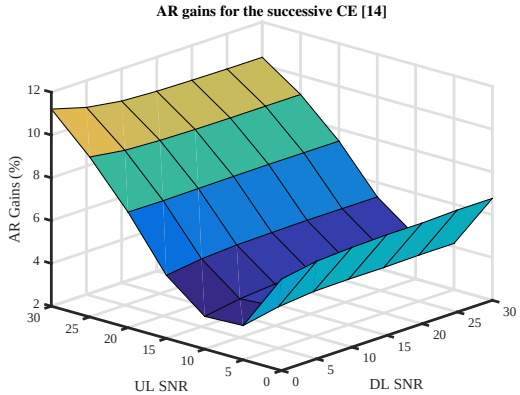

Fig. 4. Per cell effective AR gains of optimal $\mathrm{AR}(\underline{\varepsilon}^{\star})$ over non-optimal $\mathrm{AR}(\underline{\varepsilon}^{\dagger})$ for the successive CE [14].

to 11.21% for the successive CE. As expected, our grouping based CE dramatically outperforms the other two schemes because our scheme imposes a much shorter training duration of $N_{\mathrm{TN}} = 4$. The simultaneous CE achieves higher effective AR than the successive CE, also because it imposes a shorter training duration than the latter. To be specific, when the UL SNR is around 10 dB, the effective AR gain is the smallest. In particular, at this UL SNR, the optimal AR($\underline{\varepsilon}^{\star}$) is almost the same as the non-optimal AR($\underline{\varepsilon}^{\dagger}$) for the conventional and successive schemes. But when the UL SNR is relatively low or high, the effective AR is significantly enhanced through the proposed optimization of UL power allocation. Intriguingly, Figs. 2 to 4 reveal that the DL SNR has little impact on the achievable AR gains but by contrast the UL SNR has clear influence on the achievable AR gains. Physically, this is easily understood. The UL SNR has considerably influence on the channel estimation accuracy, which in turn impacts on both the UL sum rate and DL sum rate. By contrast, since there is no DL training, the DL SNR has no impact at all to the accuracy of the channel estimate and it has no influence at all on the UL sum rate. Consequently, its influence to the DL sum rate is actually less than the UL SNR. These two factors mean that the DL SNR has little influence on the achievable AR gain.

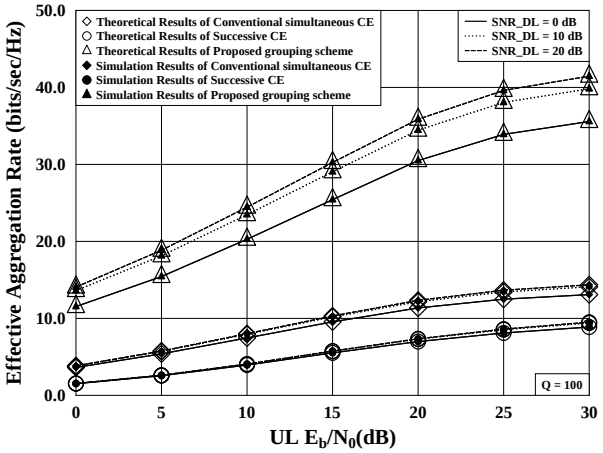

Fig. 5. Achievable per cell effective AR($\underline{\varepsilon}^{\star}$) performance by the proposed UL power allocation optimization as the functions of the UL SNR for three estimators under three different DL SNRs.

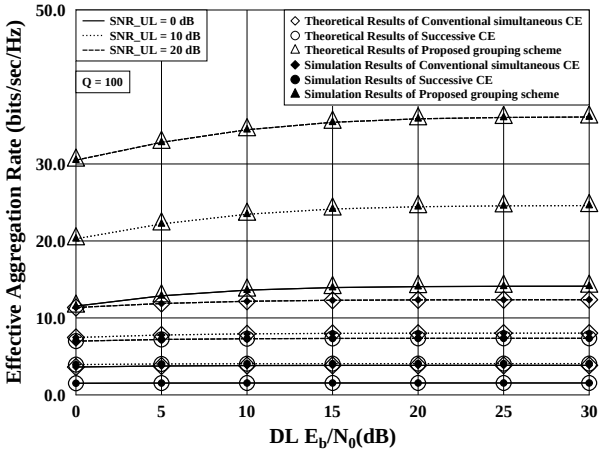

Fig. 6. Achievable per cell effective AR($\underline{\varepsilon}^{\star}$) performance by the proposed UL power allocation optimization as the functions of the DL SNR for three estimators under three different UL SNRs.

This phenomenon also has a clear mathematical explanation as it actually corresponds to the influence of the UL SNR and the DL SNR on the AR gain, respectively, according to (50), (59) and (62).

Similarly, the results of Tables III to V also reveal that the achievable effective AR performance mainly depends on the UL SNR, and the DL SNR has little effect on the effective AR. This is further confirmed by Figs. 5 and 6. In these two figures, the curves labeled with 'Theoretical Results' indicate that they are calculated using the closed-form asymptotic effective AR expression (62), while the curves labeled with 'Simulation Results' are calculated using simulation by averaging over 100 channel realizations. It can be see that the simulated effective AR agrees well with the theoretical result. This verifies the accuracy of our asymptotic analysis for SINR-UL$_{l'}^{u'}$ and SINR-DL$_{l'}^{u'}$ given in Theorems 2 and 3, respectively. Fig. 7 further compares the maximum effective AR($\underline{\varepsilon}^{\star}$) performance with the non-optimal AR($\underline{\varepsilon}^{\dagger}$) for the three CE schemes by varying the number of antennas Q deployed at each BS. As expected, the achievable effective AR increases with Q, since both SINR-UL$_{l'}^{u'}$ of (50) and SINR-DL$_{l'}^{u'}$ of (59) increase with

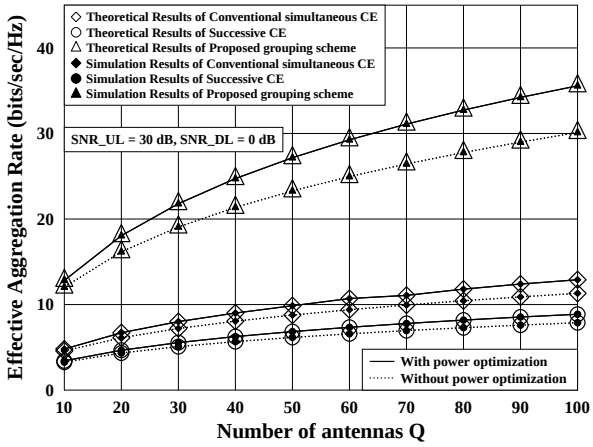

Fig. 7. Achievable per cell optimal effective AR($\underline{\varepsilon}^{\star}$) and non-optimal effective AR($\underline{\varepsilon}^{\dagger}$) as the functions of the number of antennas Q for three estimators. The UL SNR is 30 dB and DL SNR is 0 dB.

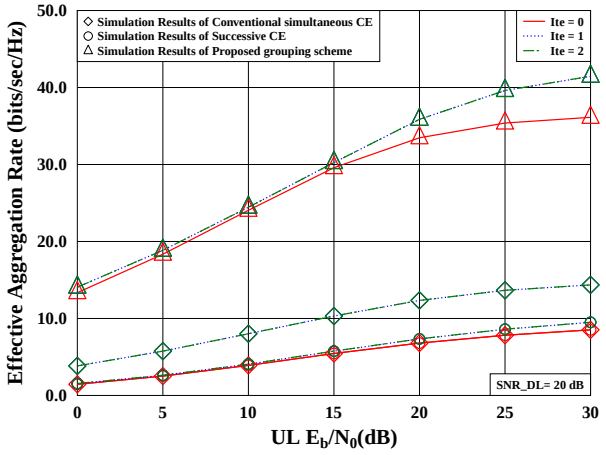

Fig. 8. Achievable per cell effective AR($\underline{\varepsilon}^{\star}$) performance by the proposed UL power allocation optimization as the functions of the UL SNR for three estimators under different iterations. All the elements of $\underline{\varepsilon}$ are 0.5 at iteration 0.

Q. Given the DL SNR 20 dB, Fig. 8 shows the achievable per cell effective AR($\underline{\varepsilon}^{\star}$) performance by the proposed UL power allocation optimization as the functions of the UL SNR for three estimators, at iteration 0, 1 and 2, where all the initial elements of $\underline{\varepsilon} \in \mathbb{R}^{LU}$ are set to 0.5 at iteration 0. From Fig. 8, we readily conclude that the proposed iterative algorithm guarantees to find the unique globally optimal solution with only one iteration.

VI. Conclusions

In order to maximize the effective AR of large-scale MIMO-OFDM systems, we have proposed to optimize the power allocation between UL pilot and data for every user. Specifically, we have derived the asymptotic closed-form expression of the achievable effective AR per cell as a function of the UL pilot-data power allocation for three typical channel estimators, and we have verified that the system's effective AR is a piecewise convex function of individual user's UL power allocation factor. Based on this closed-form asymptotic effective AR expression, the UL pilot-data power allocation optimization problem has been formulated, which can be solved efficiently using an iterative algorithm involving the golden section at each iteration. Our simulation results have demonstrated that the proposed optimization of UL pilot-data power allocation is capable of dramatically enhancing the effective AR, and the proposed iterative algorithm guarantees to attain the unique global maximum effective AR with only one iteration.

In this work, the BS performs the MRC for UL transmission, and the precoding matrix is designed using the MF criterion for DL transmission. We future work will explore the extension to other linear processing methods, such as the zero-forcing (ZF) method or the regularized ZF method.

Appendix

A. List of Useful Lemmas

Lemma 1: (Lemma 12 in [26]). Let $\check{\mathbf{A}} \in \mathbb{C}^{Q\times Q}$ and $\underline{\check{\mathbf{X}}} \sim \mathcal{CN}(\underline{\mathbf{0}}_Q, \frac{1}{Q}\mathbf{I}_Q)$. $\check{\mathbf{A}}$ has uniformly bounded spectral norm with respect to Q and it is independent of $\underline{\check{\mathbf{X}}}$. Then

$$\mathbb{E}\left\{\left|\left(\underline{\check{\mathbf{X}}}^{\mathrm{H}}\check{\mathbf{A}}\underline{\check{\mathbf{X}}}\right)^2 - \left(\frac{1}{Q}\mathrm{Tr}\{\check{\mathbf{A}}\}\right)^2\right|\right\} \xrightarrow[Q\to\infty]{\text{a.s.}} 0. \tag{66}$$

Lemma 2: Let $\check{\mathbf{A}} \in \mathbb{C}^{Q\times Q}$, and $\underline{\check{\mathbf{X}}} \sim \mathcal{CN}(\underline{\mathbf{0}}_Q, \frac{1}{Q}\mathbf{\Omega})$ with the covariance matrix $\frac{1}{Q}\mathbf{\Omega} \in \mathbb{C}^{Q\times Q}$. $\check{\mathbf{A}}$ has a uniformly bounded spectral norm with respect to Q and $\underline{\check{\mathbf{X}}}$ is independent of $\check{\mathbf{A}}$. Then

$$\lim_{Q\to\infty} \underline{\check{\mathbf{X}}}^{\mathrm{H}}\check{\mathbf{A}}\underline{\check{\mathbf{X}}} = \mathrm{Tr}\Big\{\frac{1}{Q}\mathbf{\Omega}\check{\mathbf{A}}\Big\}. \tag{67}$$

Proof: The proof is straightforward according to the trace lemma of [26]. ∎

Lemma 3: For the channel $\mathbf{H}_{l',l'} \in \mathbb{C}^{Q\times U}$ with the covariance matrix $\mathbf{R}_{l',l'}$ defined by (5), the MMSE CE $\widehat{\mathbf{H}}_{l',l'} \in \mathbb{C}^{Q\times U}$ with the covariance matrix $\mathbf{\Phi}_{l',l'}$ specified by (11), and the TP matrix $\mathbf{B}_{l'} = \sqrt{\lambda_{l'}}\widehat{\mathbf{H}}_{l',l'}$ with $\lambda_{l'}$ given by (57), we have

$$\mathrm{var}\Big\{\big(\underline{\mathbf{H}}_{l',l'}^{u'}\big)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}\Big\} = \bar{\lambda}_{l'}Q\bigg(\beta_{l',l'}^{u'} - \frac{\big(\beta_{l',l'}^{u'}\big)^2}{\varrho(\varepsilon_{l'}^{u'})^{-1} + \beta_{l',l'}^{u'}}\bigg) \times \frac{\big(\beta_{l',l'}^{u'}\big)^2}{\varrho(\varepsilon_{l'}^{u'})^{-1} + \beta_{l',l'}^{u'}}, \tag{68}$$

where $\bar{\lambda}_{l'}$ is defined in (60).

Proof: According to the distribution (11), we have $\mathbb{E}\Big\{\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\big(\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\big)^{\mathrm{H}}\Big\} = \big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}$. By setting $\underline{\check{\mathbf{X}}} = \widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}$ and $\check{\mathbf{A}} = \mathbf{I}_Q$ in Lemma 2, we have $\lim\limits_{Q\to\infty}\big(\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\big)^{\mathrm{H}}\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'} = \mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}\Big\}$. Hence, we have $\big(\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\big)^{\mathrm{H}}\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'} \approx \mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}\Big\}$ for a sufficiently large Q. From the covariance matrix of $\widetilde{\mathbf{H}}_{l',l'}$ given in (13), $\mathbb{E}\Big\{\widetilde{\underline{\mathbf{H}}}_{l',l'}^{u'}\big(\widetilde{\underline{\mathbf{H}}}_{l',l'}^{u'}\big)^{\mathrm{H}}\Big\} = \big[\mathbf{\Xi}_{l',l'}\big]_{(u',u')}$. Therefore, we arrive at (69) given at the top of this page. This completes the proof. ∎

B. Proof of Theorem 2

Proof: Upon recalling the definition $\mathbf{\Phi}_{l',l'}$ of (11) and Lemma 1 of Appendix A, we have

$$\mathbb{E}\Big\{\Big|\sqrt{\gamma_{l'}^{u'}}\underline{\mathbf{A}}_{l'}^{u'}\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\Big|^2\Big\} = \Big(\sqrt{\gamma_{l'}^{u'}}\mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}\Big\}\Big)^2 = \left(\frac{\sqrt{\gamma_{l'}^{u'}}Q(\beta_{l',l'}^{u'})^2}{\varrho_3\frac{\sigma_w^2}{\varepsilon_{l'}^{u'}P} + \beta_{l',l'}^{u'}}\right)^2. \tag{70}$$

Recalling the definition $\mathbf{\Phi}_{l',l'}$ of (11) and the definition $\mathbf{\Xi}_{l',l'}$ of (13), we have

$$\begin{aligned}\mathbb{E}\Big\{\Big|\sqrt{\gamma_{l'}^{u'}}\underline{\mathbf{A}}_{l'}^{u'}\widetilde{\underline{\mathbf{H}}}_{l',l'}^{u'}\Big|^2\Big\} &= \gamma_{l'}^{u'}\mathbb{E}\Big\{\big(\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\big)^{\mathrm{H}}\widetilde{\underline{\mathbf{H}}}_{l',l'}^{u'}\big(\widetilde{\underline{\mathbf{H}}}_{l',l'}^{u'}\big)^{\mathrm{H}}\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\Big\} \\ &= \gamma_{l'}^{u'}\mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}\big[\mathbf{\Xi}_{l',l'}\big]_{(u',u')}\Big\} \\ &= Q\gamma_{l'}^{u'}\left(\beta_{l',l'}^{u'} - \frac{(\beta_{l',l'}^{u'})^2}{\varrho_3\frac{\sigma_w^2}{\varepsilon_{l'}^{u'}P} + \beta_{l',l'}^{u'}}\right)\frac{(\beta_{l',l'}^{u'})^2}{\varrho_3\frac{\sigma_w^2}{\varepsilon_{l'}^{u'}P} + \beta_{l',l'}^{u'}}.\end{aligned} \tag{71}$$

Similarly, recalling (5) and (11), we have

$$\mathbb{E}\Big\{\Big|\sqrt{\gamma_{l'}^{u}}\underline{\mathbf{A}}_{l'}^{u'}\underline{\mathbf{H}}_{l',l'}^{u}\Big|^2\Big\} = \gamma_{l'}^{u}\mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}\big[\mathbf{R}_{l',l'}\big]_{(u,u)}\Big\} = \frac{Q\gamma_{l'}^{u}\big(\beta_{l',l'}^{u'}\big)^2\beta_{l',l'}^{u}}{\varrho_3\frac{\sigma_w^2}{\varepsilon_{l'}^{u'}P} + \beta_{l',l'}^{u'}}. \tag{72}$$

Additionally, we have the following two expressions

$$\mathbb{E}\Big\{\Big|\sqrt{\gamma_{l}^{u}}\underline{\mathbf{A}}_{l'}^{u'}\underline{\mathbf{H}}_{l,l'}^{u}\Big|^2\Big\} = \gamma_{l}^{u}\mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}\big[\mathbf{R}_{l,l'}\big]_{(u,u)}\Big\} = \frac{Q\gamma_{l}^{u}\big(\beta_{l',l'}^{u'}\big)^2\beta_{l,l'}^{u}}{\varrho_3\frac{\sigma_w^2}{\varepsilon_{l'}^{u'}P} + \beta_{l',l'}^{u'}}, \tag{73}$$

$$\begin{aligned}\mathrm{var}\Big\{\big(\underline{\mathbf{H}}_{l',l'}^{u'}\big)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}\Big\} &= \lambda_{l'}\mathbb{E}\Big\{\Big|\big(\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'} + \widetilde{\underline{\mathbf{H}}}_{l',l'}^{u'}\big)^{\mathrm{H}}\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'} - \mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}\Big\}\Big|^2\Big\} \approx \lambda_{l'}\mathbb{E}\Big\{\Big|\big(\widetilde{\underline{\mathbf{H}}}_{l',l'}^{u'}\big)^{\mathrm{H}}\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\Big|^2\Big\} \\ &= \bar{\lambda}_{l'}\mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}\big[\mathbf{\Xi}_{l',l'}\big]_{(u',u')}\Big\} = \bar{\lambda}_{l'}Q\bigg(\beta_{l',l'}^{u'} - \frac{(\beta_{l',l'}^{u'})^2}{\varrho(\varepsilon_{l'}^{u'})^{-1} + \beta_{l',l'}^{u'}}\bigg)\frac{\big(\beta_{l',l'}^{u'}\big)^2}{\varrho(\varepsilon_{l'}^{u'})^{-1} + \beta_{l',l'}^{u'}}.\end{aligned} \tag{69}$$

$$\mathbb{E}\Big\{\Big|\underline{\mathbf{A}}_{l'}^{u'}\underline{\mathbf{V}}_{l'}^{t}\Big|^2\Big\}=\mathbb{E}\Big\{\Big|\big(\widehat{\underline{\mathbf{H}}}_{l',l'}^{u'}\big)^{\mathrm{H}}\underline{\mathbf{V}}_{l'}^{t}\Big|^2\Big\}=\frac{Q\big(\beta_{l',l'}^{u'}\big)^2\sigma_v^2}{\varrho_3\frac{\sigma_w^2}{\varepsilon_{l'}^{u'}P}+\beta_{l',l'}^{u'}}. \quad (74)$$

Substituting (70) to (74) into (49), we asymptotically arrive at (50). This completes the proof. ∎

C. Proof of Theorem 3

Proof: Recalling (11) and (57), we have

$$\mathbb{E}\Big\{\big(\underline{\mathbf{H}}_{l',l'}^{u'}\big)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}\Big\}=\frac{\mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u',u')}\Big\}}{\sqrt{\mathrm{Tr}\Big\{\frac{1}{U}\sum_{u=1}^{U}\big[\mathbf{\Phi}_{l',l'}\big]_{(u,u)}\Big\}}}$$
$$=\sqrt{\bar{\lambda}_{l'}}\frac{Q\big(\beta_{l',l'}^{u'}\big)^2}{\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}}. \quad (75)$$

Thus, the desired signal power is given by

$$\text{SP-DL}_{l'}^{u'}=\Big|\mathbb{E}\Big\{\big(\underline{\mathbf{H}}_{l',l'}^{u'}\big)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}\Big\}\Big|^2=\bar{\lambda}_{l'}\left(\frac{Q\big(\beta_{l',l'}^{u'}\big)^2}{\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}}\right)^2. \quad (76)$$

From Lemma 3 of Appendix A, as $Q\to\infty$, we have

$$\mathrm{var}\Big\{\big(\underline{\mathbf{H}}_{l',l'}^{u'}\big)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u'}\Big\}=\bar{\lambda}_{l'}Q\left(\beta_{l',l'}^{u'}-\frac{(\beta_{l',l'}^{u'})^2}{\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}}\right)$$
$$\times\frac{(\beta_{l',l'}^{u'})^2}{\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}}. \quad (77)$$

Additionally, recalling (5) and (11), we have

$$\mathbb{E}\Big\{\Big|\big(\underline{\mathbf{H}}_{l',l'}^{u'}\big)^{\mathrm{H}}\underline{\mathbf{B}}_{l'}^{u}\Big|^2\Big\}=\lambda_{l'}\mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l',l'}\big]_{(u,u)}\big[\mathbf{R}_{l',l'}\big]_{(u',u')}\Big\}$$
$$=\bar{\lambda}_{l'}\frac{Q\big(\beta_{l',l'}^{u}\big)^2\beta_{l',l'}^{u'}}{\varrho(\varepsilon_{l'}^{u})^{-1}+\beta_{l',l'}^{u}}, \quad (78)$$

$$\mathbb{E}\Big\{\Big|\big(\underline{\mathbf{H}}_{l',l}^{u'}\big)^{\mathrm{H}}\underline{\mathbf{B}}_{l}^{u}\Big|^2\Big\}=\lambda_{l}\mathrm{Tr}\Big\{\big[\mathbf{\Phi}_{l,l}\big]_{(u,u)}\big[\mathbf{R}_{l',l}\big]_{(u',u')}\Big\}$$
$$=\bar{\lambda}_{l}\frac{Q\big(\beta_{l,l}^{u}\big)^2\beta_{l',l}^{u'}}{\varrho(\varepsilon_{l}^{u})^{-1}+\beta_{l,l}^{u}}. \quad (79)$$

From (77) to (79) and noting (60), we arrive at

$$\text{INP-DL}_{l'}^{u'}=U\sum_{l=1}^{L}\beta_{l',l}^{u'}-\frac{\bar{\lambda}_{l'}Q\big(\beta_{l',l'}^{u'}\big)^4}{\Big(\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}\Big)^2}+\frac{\sigma_J^2}{p_{\mathrm{DL}}}. \quad (80)$$

Using (76) and (80) completes the proof. ∎

D. Proof of Theorem 4

Proof: a). $u^*=u'$ and $l^*=l'$: It can be proved that the unique solution exists by solving $\frac{\partial\text{SINR-UL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}=0$. Moreover it can be shown that $\frac{\partial^2\text{SINR-UL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial^2\varepsilon_{l'}^{u'}}<0$. Therefore, $\text{SINR-UL}_{l'}^{u'}(\varepsilon_{l'}^{u'})$ is convex.

b). $u^*\neq u'$ and/or $l^*\neq l'$: It can be shown that $\Big(f_{2,l',u'}^{\mathrm{UL}}(\varepsilon_{l'}^{u'})\Big)^2>0$. $\frac{\partial\text{SINR-UL}_{l'}^{u'}(\varepsilon_{l^*}^{u^*})}{\partial\varepsilon_{l^*}^{u^*}}>0$, and $\text{SINR-UL}_{l'}^{u'}(\varepsilon_{l^*}^{u^*})$ increases with $\varepsilon_{l^*}^{u^*}$.

The detailed proof is omitted due to space limitation. ∎

E. Proof of Theorem 5

Proof: a). $u^*=u'$ and $l^*=l'$: The derivation of $\text{SP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})$ with respect to $\varepsilon_{l'}^{u'}$ is given in (81), and the derivation of $\text{INP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})$ with respect to $\varepsilon_{l'}^{u'}$ is

$$\frac{\partial\text{INP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}=-\frac{1}{Q}\frac{\partial\text{SP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}. \quad (82)$$

As a result, we have (83), since $\frac{\partial\text{SP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}>0$. Obviously, $\big(\text{INP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})\big)^2>0$. Therefore, $\frac{\partial\text{SINR-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}>0$, and we conclude that for $0<\varepsilon_{l'}^{u'}<1$, $\text{SINR-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})$ is an increasing function of $\varepsilon_{l'}^{u'}$.

b). $u^*\neq u'$ and/or $l^*\neq l'$: From (59) and (60), it is obvious that $\text{SINR-DL}_{l'}^{u'}$ is not a function of $\varepsilon_{l^*}^{u^*}$. ∎

$$\frac{\partial\text{SP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}=\frac{\frac{2Q^2\left(\beta_{l',l'}^{u'}\right)^4\varrho(\varepsilon_{l'}^{u'})^{-2}}{\left(\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}\right)^3}\sum_{u=1}^{U}\frac{\left(\beta_{l',l'}^{u}\right)^2}{\varrho(\varepsilon_{l'}^{u})^{-1}+\beta_{l',l'}^{u}}-\frac{Q^2\left(\beta_{l',l'}^{u'}\right)^4}{\left(\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}\right)^2}\frac{\left(\beta_{l',l'}^{u'}\right)^2\varrho(\varepsilon_{l'}^{u'})^{-2}}{\left(\varrho(\varepsilon_{l'}^{u'})^{-1}+\beta_{l',l'}^{u'}\right)^2}}{\frac{Q}{U}\left(\sum_{u=1}^{U}\frac{\left(\beta_{l',l'}^{u}\right)^2}{\varrho(\varepsilon_{l'}^{u})^{-1}+\beta_{l',l'}^{u}}\right)^2}, \quad (81)$$

$$\frac{\partial\text{SINR-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}\big(\text{INP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})\big)^2=\frac{\partial\text{SP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}\text{INP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})-\frac{\partial\text{INP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}\text{SP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})$$
$$=\frac{\partial\text{SP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}\Big(\text{INP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})+\frac{1}{Q}\text{SP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})\Big)=\frac{\partial\text{SP-DL}_{l'}^{u'}(\varepsilon_{l'}^{u'})}{\partial\varepsilon_{l'}^{u'}}\left(\sum_{l=1}^{L}\frac{\sum_{u=1}^{U}\frac{Q\left(\beta_{l,l}^{u}\right)^2\beta_{l',l}^{u'}}{\varrho(\varepsilon_{l}^{u})^{-1}+\beta_{l,l}^{u}}}{\frac{Q}{U}\sum_{u=1}^{U}\frac{\left(\beta_{l,l}^{u}\right)^2}{\varrho(\varepsilon_{l}^{u})^{-1}+\beta_{l,l}^{u}}}+\frac{\sigma_J^2}{p_{\mathrm{DL}}}\right)>0, \quad (83)$$

REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] L. Zhao, K. Li, K. Zheng, and M. O. Ahmad, "An analysis of the trade-off between the energy and spectrum efficiencies in an uplink massive MIMO-OFDM system," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 62, no. 3, pp. 291–295, Mar. 2015.

[3] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[4] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Tech. J.*, vol. 20, pp. 11–22, 2015.

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

[6] 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.

[7] H. 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. 2013.

[8] J. Jose, A. Ashikhmin, T. L. 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] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimations for multicell multiuser MIMO systems with very large antenna arrays," in *Proc. ICASSP 2012* (Kyoto, Japan), Mar. 25-30, 2012, pp. 3249–3252.

[11] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[12] Z. Wang, P. Zhao, C. Qian, and S. Chen, "Location-aware channel estimation enhanced TDD based massive MIMO," *IEEE Access*, vol. 4, pp. 7828–7840, 2016.

[13] 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.

[14] T. X. Vu, T. A. Vu, and T. Q. S. Quek, "Successive pilot contamination elimination in multiantenna multicell networks," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 617–620, Dec. 2014.

[15] A. Ashikhmin, and T. L. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. ISIT 2012* (Cambridge, MA), Jul. 1-6, 2012, pp. 1137–1141.

[16] X. Guo, *et al.*, "Two-stage time-domain pilot contamination elimination in large-scale multiple-antenna aided and TDD based OFDM systems," *IEEE Access*, vol. 5, pp. 8629–8641, 2017.

[17] X. Guo, *et al.*, "Optimal pilot design for pilot contamination elimination/reduction in large-scale multiple-antenna aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7229–7243, Nov. 2016.

[18] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 605–608, Dec. 2014.

[19] P. Liu, *et al.*, "Pilot power allocation through user grouping in multi-cell massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1561–1574, Apr. 2017.

[20] T. V. Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, Mar. 2018.

[21] Y. Zhang and W. Zhu, "Energy-efficient pilot and data power allocation in massive MIMO communication systems based on MMSE channel estimation," in *Proc. ICASSP 2016* (Shanghai, China), Mar. 20-25, 2016, pp. 3571–3575.

[22] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell Massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2363–2378, May 2017.

[23] H. Yang and T. L. Marzetta, "A macro cellular wireless network with uniformly high user throughputs," in *Proc. VTC2014-Fall* (Vancouver, Canada), Sep. 14-17, 2014, pp. 1–5.

[24] K. Guo, Y. Guo, G. Fodor, and G. Ascheid, "Uplink power control with MMSE receiver in multi-cell MU-massive-MIMO systems," in *Proc. ICC 2014* (Sydney, Australia), Jun. 10-14, 2014, pp. 5184–5190.

[25] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 67–75, Jan. 2002.

[26] J. Hoydis, *Random Matrix Theory for Advanced Communication Systems*. Ph.D. dissertation, Supélec, France, 2012.