

Large-Scale Cloud Radio Access Networks: Fundamental Asymptotic Analysis

Kyung Jun Choi and Kwang Soon Kim

Abstract: Large-scale cloud radio access network (LS-CRAN) is a highly promising next-generation cellular network architecture whereby many base stations (BSs) equipped with a massive antenna array are connected to a cloud-computing based central processor unit via digital front/backhaul links. This paper studies the asymptotic behavior of downlink (DL) performance of a LS-CRAN. As an asymptotic performance measure, the scaling exponent of the signal-to-interference-plus-noise-ratio (SINR) is derived for interference-free (IF), maximum-ratio transmission (MRT), and zero-forcing (ZF) operations. Our asymptotic analysis reveals four fundamental operating regimes and the performances of both MRT and ZF operations are fundamentally limited by the UL transmit power for estimating user's channel state information, not the DL transmit power. We obtain the conditions that MRT or ZF operation becomes interference-free, i.e., order-optimal. As higher UL transmit power is provided, more users can be associated and the data rate per user can be increased simultaneously while keeping the order-optimality.

Index Terms: Asymptotic performance, cloud-RAN, massive-MIMO, ultra-dense network.

I. INTRODUCTION

RECENTLY, mobile data traffic is explosively and continuously rising due to smart phone and tablet users and it is expected that next-generation (i.e., the 5th generation) cellular networks will offer a 1000x increase in network capacity as well as a 1000x increase in energy-efficiency in the following decade to meet such an excessively high user demand [1]. The most prominent method to increase the network capacity is the network densification [2] by either implementing more antennas at base stations (BSs) called large-scale antenna system (LSAS) [3] or adding more small cells at hot spot areas called ultra-dense network (UDN) [4]. By using a lot of antennas at each BS, the LSAS can exploit massive spatial dimension to generate a sharp beam and thus it can provide higher spectral efficiency and also lower power consumption [5]. However, the performance of the LSAS is highly limited by the accuracy of channel state information (CSI) and varies over the geographical

locations of users. On the other hand, an UDN exploits spatial reuse obtained from deploying more small BSs so that it can offer geographically uniform performance to each user due to the reduction of the access distance between users and BSs [6], [7]. However, its performance is limited by uncoordinated out-of-cell interference so that an efficient interference management is a key challenge for an UDN, while managing the costs in terms of the front/backhaul overhead and computational complexity for a network operator.

Large-scale cloud radio access network (LS-CRAN) is recently regarded as a novel wireless cellular network architecture to unify the above two architectures and is able to implement the interference handling mechanisms introduced for the long-term evolution (LTE) or LTE-Advanced, such as the enhanced inter-cell interference coordination [8] and the coordinated multi-point transmission [9], [10]. In the LS-CRAN, lots of BSs each with a massive antenna array are connected to a virtualized central processing unit (CPU) via dedicated front/backhaul link and some of the baseband processing functionality of each BS is migrated to the CPU. As a result, it is expected that the performance of the LS-CRAN is much higher than that of a conventional network.

Although the performance of an LS-CRAN has been widely investigated in literature [9], [11] via simulations, an intuitive analytic result considering practical limitations is not available yet. The exact distribution of the signal-to-interference-plus-noise ratio (SINR) of the MRT or ZF operation is derived in [12] for the case of two distributed BSs and its Laplace approximated distribution is derived in [13] for a general LS-CRAN. In [14], a simpler form of the SINR distribution is derived by approximating the SINR as a Gamma random variable. As parallel works, the average achievable rate of the MRT or ZF operation is derived in [15] for a single BS case, and in [16] for the case of fully-distributed or fully-colocated antennas as a function of the BS and user locations. In order to understand the network behavior more intuitively, stochastic geometry has been recently adopted to remove such dependency and describe important network metrics in a probabilistic way. In [17]–[19], the SINR distributions of cooperative transmission schemes are provided by adopting stochastic geometry, but the analysis resorts on multiple numerical integrals. In [20], an asymptotic analysis on the tail of the SINR distribution is provided by applying large-deviation theory and stochastic geometry. In [21], the average achievable rate of the MRT and ZF successive interference cancellation is investigated and its scaling laws are obtained in a dense random network with multiple receive antennas.

The main objective of this paper is to characterize the asymptotic behavior of the LS-CRAN. The scaling exponents of the signal-to-interference-plus-noise ratio (SINR) defined and

Manuscript received December 3, 2018; approved for publication by EIC, Division I Editor, December 5, 2019.

This work was supported in part by ICT R&D program of MSIT/IITP {2014-0-00552, Next Generation WLAN System with High Efficient Performance, and 2018-0-00923, Scalable Spectrum Sensing for Beyond 5G Communication}, and in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2014R1A2A2A01007254, NRF-2015K2A3A1000189).

K.J.Choi and K.S.Kim are with the Department of Electrical and Electronic Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Korea, email:{kjchoi,ks.kim}@yonsei.ac.kr.

K.S.Kim is the corresponding author.

Digital Object Identifier: 10.1109/JCN.2019.000008

1229-2370/19/\$10.00 © 2019 KICS

Creative Commons Attribution-NonCommercial (CC BY-NC).

This is an Open Access article distributed under the terms of Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided that the original work is properly cited.

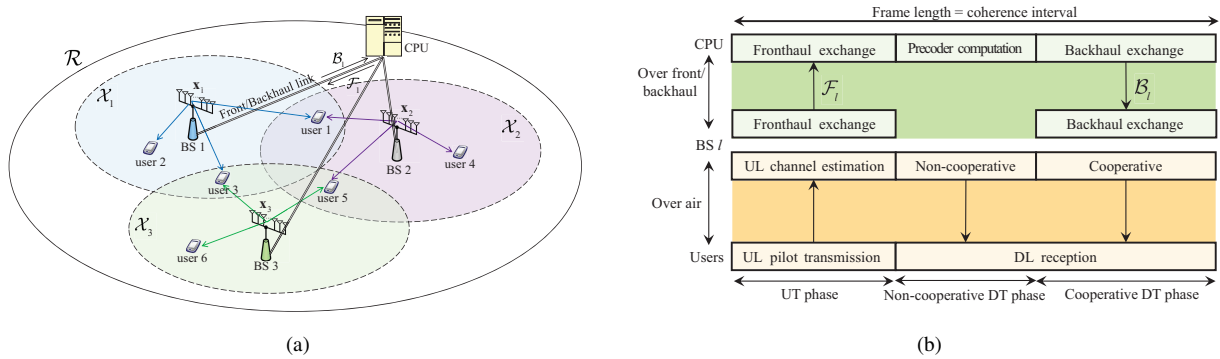


Fig. 1. System model and DL frame structure for LS-CRAN: (a) System model and (b) DL frame structure.

driven as a function of the numbers of BS, BS antennas, and single-antenna users, and the UL/DL transmit power for the interference-free (IF), MRT, and ZF operations, which are stated in Theorems 1–3. Based on the derived scaling exponents, four fundamental regimes are distinguished according to the UL transmit power (or the CSI acquisition error): (i) The extremely high (UL transmission) power regime (EH), (ii) the high power regime (H), (iii) the medium power regime (M), and (iv) the low power regime (L). The scaling exponents of the three operations are derived according to the four regimes with insightful discussions.

The remainder of this paper is organized as follows. In Section II, the LS-CRAN system model is described and the scaling exponent of the SINR is defined as a network performance measure. Section III provides some candidate operations suitable for practical implementation and the corresponding scaling exponents are analyzed to provide an intuitive understanding on the LS-CRAN in Section IV. Finally, conclusion is given in Section V.

Matrices and vectors are respectively denoted by boldface uppercase and lowercase characters. The superscript $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote the conjugate, transpose and conjugate transpose, respectively. $|\cdot|$, $\mathbb{E}[\cdot]$, $\text{Tr}(\cdot)$, and $\delta(n)$ stand for the cardinality of a set, statistical expectation, trace of a square matrix, and Kronecker delta function, respectively. Also, $\mathbf{1}_{A \times 1}$ and $\mathbf{0}_{A \times 1}$ are the $A \times 1$ all one vector and all zero vector, respectively. Further, $\text{diag}(a_1, a_2, \dots, a_n)$ denotes the diagonal matrix whose (k, k) th element is a_k , and $\mathcal{CN}(\mu, \sigma^2)$ denotes the circularly symmetric complex Gaussian distribution with mean μ and variance σ^2 . For a $C \times 1$ vector \mathbf{x}_{ab} , $[\mathbf{x}_{ab}]_{a=1, b=1}^{A, B}$ denotes the $AC \times B$ matrix $[[\mathbf{x}_{11}^H, \dots, \mathbf{x}_{A1}^H]^H, \dots, [\mathbf{x}_{1B}^H, \dots, \mathbf{x}_{AB}^H]^H]$.

In this paper, we only deal with a scaling exponent, s , of a random measure, $f(n)$, as $n \rightarrow \infty$ in a probabilistic sense, which is mathematically defined as follows. The scaling exponent of $f(n)$ is s in probability if

$$\lim_{n \rightarrow \infty} \Pr \left(\left| \frac{\log f(n)}{\log n} - s \right| < \epsilon \right) = 1, \quad (1)$$

holds for any positive finite ϵ . Also $s = \pm\infty$ is used if $\lim_{n \rightarrow \infty} \Pr \left(\left| \frac{\log f(n)}{\log n} \right| > t \right) = 1$ for any finite t . Let s and t

be the scaling exponents of $f(n)$ and $g(n)$, respectively. Then, the following modified order notations are used through this paper.

- 1) $f(n) = o(g(n))$ or $f(n) \ll g(n)$, if $s < t$,
- 2) $f(n) = O(g(n))$, if $s \leq t$,
- 3) $f(n) = \omega(g(n))$ or $f(n) \gg g(n)$, if $s > t$,
- 4) $f(n) = \Omega(g(n))$, if $s \geq t$, and
- 5) $f(n) = \Theta(g(n))$ or $f(n) \asymp g(n)$, if $s = t$.

II. SYSTEM MODEL

Consider an LS-CRAN system as illustrated in Fig. 1 (a). Suppose that L BSs with M antennas and K users with a single antenna are uniformly distributed on a finite region \mathcal{R} . The sets of BSs and users are denoted as $\mathcal{X} = \{X_1, X_2, \dots, X_L\}$ and $\mathcal{U} = \{U_1, U_2, \dots, U_K\}$, respectively, and they are assumed to be independent homogeneous Poisson point processes (PPPs) over \mathcal{R} .¹ With a slight abuse of notations, X_l and U_k are used as the locations of BS l and user k , respectively. It is assumed that each user is associated with neighboring BSs and the set of the BSs serving user k is defined as

$$\mathcal{X}_k = \{X_l \in \mathcal{X} \mid |U_k - X_l| \leq R_{\text{th}}\}, \quad (2)$$

where R_{th} is the association range and the set of users associated with BS l is denoted as $\mathcal{U}_l = \{U_k \in \mathcal{U} \mid X_l \in \mathcal{X}_k\}$. To guarantee each user is served by at least one BS, \mathcal{X}_k needs to be not empty for all k . Note that \mathcal{U}_l and $\mathcal{U}_{l'}$ are not necessarily disjoint and typical users are associated with multiple BSs for being served cooperatively. In this paper, a network is called *fully associated*, if $\mathcal{X}_k = \mathcal{X}$ for all k . Otherwise, the network is called *partially associated*.

The BSs are connected to the CPU via high-speed dedicated front/backhaul link for enabling a cooperative transmission operation. For inter-signaling between the BSs and the CPU, the sets of front/backhaul information are defined as \mathcal{F}_l and \mathcal{B}_l ,

¹Typical BS deployment in practice would be better represented by a point process with a mutual interaction. However, such a mutual interaction typically becomes weaker as a deployment becomes denser and more random so that an asymptotic analysis using a PPP assumption is not only mathematically tractable but also enough for capturing good insights on the asymptotic behaviors of large networks.

where \mathcal{F}_l is the fronthaul information set (from CPU to BS l) and \mathcal{B}_l is the backhaul information set (from BS l to CPU). The elements of \mathcal{F}_l and \mathcal{B}_l are closely related according to a specific transmission operation, which will be described in Section III.

The DL frame structure operating in time-division duplex (TDD) mode is shown in Fig. 1(b) and has three phases as follows. UL training (UT) phase, non-cooperative data transmission (DT) phase, and cooperative DT-phase. In the UT-phase, the instantaneous CSI of each user is estimated by receiving user's pilot (or reference) signal at each associated BS independently. After the UT-phase, the non-cooperative DT-phase is performed first and followed by the cooperative DT-phase. In the non-cooperative DT-phase, each BS separately transmits the data symbols to the associated users without any front/backhaul exchange. In the cooperative DT-phase, the BSs jointly transmit the data symbols to the associated users with the aid of the exchanged information via the front/backhaul link. For simplicity, the cooperative DT-phase is only focused in this paper.

Let \mathbf{g}_{lk}^H denote the $1 \times M$ flat-fading DL channel vector from BS l to user k , which can be written as²

$$\mathbf{g}_{lk}^H = \sqrt{\beta_{lk}} \mathbf{h}_{lk}^H, \quad (3)$$

where $\mathbf{h}_{lk} \in \mathbb{C}^{M \times 1}$ is the short-term CSI whose elements are independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ and $\beta_{lk} (\geq 0)$ is the long-term CSI depending on the path-loss and shadowing. The long-term CSI between BS l and user k is modeled as $\beta_{lk} = |X_l - U_k|^{-\alpha}$, where $\alpha (> 2)$ is the wireless channel path-loss exponent.³

It is assumed that the short-term CSI of each user remains constant within a given frame but independent across different frames, while the long-term CSI does not vary during a much longer interval. Further, it is assumed that the long-term CSIs among all BSs and users are perfectly known at the CPU through an infrequent feedback with a negligible overhead. Additionally, we assume perfect TDD reciprocity calibration so that the UL channel vector is just a transpose of the DL channel vector (i.e., the UL channel vector from user k to BS l is denoted as \mathbf{g}_{lk}^*).

Let $\mathbf{s} = [s_1, \dots, s_K]^T$ be the $K \times 1$ information symbol vector with $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_K$, where s_k denotes the information symbol for user k . Let $\mathbf{x} = [\mathbf{x}_1^T, \dots, \mathbf{x}_L^T]^T$ be the $LM \times 1$ global transmitted signal vector, where \mathbf{x}_l is the local transmitted signal vector of BS l , given by

$$\mathbf{x} = \mathbf{F}^\Upsilon (\mathbf{Q}^\Upsilon)^{\frac{1}{2}} \mathbf{s}, \quad (4)$$

where $\mathbf{Q}^\Upsilon \triangleq \text{diag}(Q_1^\Upsilon, \dots, Q_K^\Upsilon)$ denotes the $K \times K$ power allocation matrix and $Q_k^\Upsilon (\geq 0)$ denotes the power allocated to user k , $\mathbf{F}^\Upsilon = [\mathbf{f}_{lk}^\Upsilon]_{l=1, k=1}^{L, K}$ denotes the $LM \times K$ precoding matrix and \mathbf{f}_{lk}^Υ denotes the $M \times 1$ precoding vector of user k for BS l , and Υ denotes the cooperative transmission operation used in the LS-CRAN. Since the BS serves the associated users only, $\mathbf{f}_{lk}^\Upsilon = \mathbf{0}_{M \times 1}$ for $\forall X_l \notin \mathcal{X}_k$ (or equivalently $\forall U_k \notin \mathcal{U}_l$). Note that the transmitted signal vector of BS l can be written

²A narrow-band flat-fading channel is assumed because wideband frequency-selective channels may be decomposed into multiple narrow-band channels using modulation schemes such as the orthogonal frequency division multiplexing.

³By using the random displacement theorem similarly as in [22], this model can include the shadowing effect.

by $\mathbf{x}_l = \sum_{U_j \in \mathcal{U}_l} \mathbf{f}_{lj}^\Upsilon \sqrt{Q_j^\Upsilon} s_j$. Then, the DL transmit power for user j , P_j^{dl} , is given as

$$P_j^{\text{dl}} = Q_j^\Upsilon \sum_{X_l \in \mathcal{X}_j} \|\mathbf{f}_{lj}^\Upsilon\|^2, \quad (5)$$

and the total DL transmit power is given by $P_\Sigma^{\text{dl}} \triangleq \sum_{U_j \in \mathcal{U}} P_j^{\text{dl}}$.

Let $\mathbf{y} \triangleq [y_1, \dots, y_K]^T$ be the $K \times 1$ aggregated received signal vector, given by

$$\begin{aligned} \mathbf{y} &= \mathbf{G}\mathbf{x} + \mathbf{n} \\ &= \mathbf{G}\mathbf{F}^\Upsilon (\mathbf{Q}^\Upsilon)^{\frac{1}{2}} \mathbf{s} + \mathbf{n}, \end{aligned} \quad (6)$$

where $\mathbf{G} = \left([\mathbf{g}_{lk}]_{l=1, k=1}^{L, K} \right)^H$ denotes the $K \times LM$ channel matrix among all users and all BSs and $\mathbf{n} \triangleq [n_1, \dots, n_K]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ denotes the $K \times 1$ noise vector. Then, the received signal at user k can be expressed as

$$y_k = \psi_{kk}^\Upsilon \sqrt{Q_k^\Upsilon} s_k + \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} \psi_{kj}^\Upsilon \sqrt{Q_j^\Upsilon} s_j + n_k, \quad (7)$$

where $\psi_{kj}^\Upsilon = \sum_{X_l \in \mathcal{X}_j} \mathbf{g}_{lk}^H \mathbf{f}_{lj}^\Upsilon$ is the effective channel seen at user k .

III. LS-CRAN OPERATIONS AND PERFORMANCE MEASURE

In this section, the ideal IF operation is introduced as a reference system and practical cooperative operations are reviewed. Note that a comprehensive review on the operations is beyond the scope of this paper so that two well-known practical operations, MRT operation [23] and ZF operation [24], are focused.

A. Cooperative Transmission Operations

A.1 Ideal IF Operation

As a reference, the ideal IF operation is considered, where the interference term in (7) is removed by Genie perfectly without any cost while the desired signal power is maximized by using the IF precoding matrix,

$$\mathbf{F}^{\text{if}} = \widehat{\mathbf{G}}^H. \quad (8)$$

Obviously, the ideal IF operation provides an upper-bound on the performance of any practical operation. Since this operation cannot be realizable, $\mathcal{B}_l^{\text{if}}$ and $\mathcal{F}_l^{\text{if}}$ are not defined.

A.2 MRT Operation

MRT operation tries to maximize the received signal power of a desired user without considering the effect of interference to undesired users [23]. This operation is regarded as a good candidate as the number of BS antennas increases due to its low-computational complexity and low-overhead requirement [25]. The precoder for MRT operation is given by

$$\mathbf{F}^{\text{mrt}} = \widehat{\mathbf{G}}^H. \quad (9)$$

Since the precoder of MRT operation does not need information exchange among BSs, the backhaul and fronthaul information sets can be expressed as $\mathcal{B}_l^{\text{mrt}} = \emptyset$, and $\mathcal{F}_l^{\text{mrt}} = \left\{ \sqrt{Q_j^{\text{mrt}}} s_j | U_j \in \mathcal{U}_l \right\}$, respectively.

A.3 ZF Operation

ZF operation can cancel the interference term in (7) (perfectly, provided that the perfect channel estimation is available at the CPU) at the expense of the desired signal power loss [24]. The precoder for ZF operation is given by

$$\mathbf{F}^{\text{zf}} = \widehat{\mathbf{G}}^H \left(\widehat{\mathbf{G}} \widehat{\mathbf{G}}^H \right)^{-1}. \quad (10)$$

Note that ZF operation can be used only when the number of antennas in the system is larger than or equal to that of users, i.e., $LM \geq K$ and the estimated channel matrix, $\widehat{\mathbf{G}}$, has full-rank. To construct the ZF precoder, the CPU requires to know the estimated channel matrix $\widehat{\mathbf{G}}$ so that the short-term CSIs of all associated users at each BS need to be conveyed to the CPU via the dedicated front/backhaul link. Thus, the backhaul and fronthaul information sets can be expressed as $\mathcal{B}_l^{\text{zf}} = \left\{ \widehat{\mathbf{h}}_{lj} | U_j \in \mathcal{U}_l \right\}$ and $\mathcal{F}_l^{\text{zf}} = \left\{ \mathbf{f}_{lj}^{\text{zf}} \sqrt{Q_j^{\text{zf}}} s_j | U_j \in \mathcal{U}_l \right\}$, respectively.

Note that it is well-known that the performance of ZF operation is better than that of MRT operation if the network has sufficient transmit power. But, if not, ZF operation can be inferior to MRT operation. In this paper, we will show that ZF operation is always superior or identical to MRT operation in the viewpoint of the scaling exponent of the SINR regardless of the operating downlink or uplink transmit power.

B. Pilot Allocation and Channel Estimation

The UL channel is estimated during the dedicated UT-phase and then the DL channel is obtained by the TDD channel reciprocity. It is assumed that the length of the UT-phase is T and there are T orthonormal pilot signals, denoted as $\psi_i \in \mathbb{C}^{T \times 1}$, $i = 1, 2, \dots, T$, where $\psi_i^H \psi_j = \delta(i - j)$. Then, user j transmits $\sqrt{P_j^{\text{ul}}} \psi_{\pi_j}^T$ during the UT-phase of length T , where P_j^{ul} is the UL transmit power of user j and π_j is the index of the pilot signal allocated to user j . The total UL transmit power is denoted as $P_{\Sigma}^{\text{ul}} = \sum_{U_j \in \mathcal{U}} P_j^{\text{ul}}$. Then, the $M \times T$ received signal matrix at BS l during the UT-phase can be written as

$$\mathbf{Y}_l = \sum_{U_j \in \mathcal{U}} \sqrt{P_j^{\text{ul}}} \beta_{lj} \mathbf{h}_{lj}^* \psi_{\pi_j}^T + \mathbf{V}_l, \quad (11)$$

where \mathbf{V}_l denotes the $M \times T$ noise matrix whose elements are i.i.d. $\mathcal{CN}(0, 1)$. Using the minimum mean-square error (MMSE) channel estimator [26], the estimated short-term CSI of user $k \in \mathcal{U}_l$ at BS l can be written as

$$\begin{aligned} \widehat{\mathbf{h}}_{lk} &= \vartheta_{lk} \mathbf{Y}_l^* \psi_{\pi_k} \\ &= \phi_{lk} \mathbf{h}_{lk} + \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} \phi_{lj} \mathbf{h}_{lj} \delta(\pi_k - \pi_j) + \vartheta_{lk} \widetilde{\mathbf{v}}_{lk}, \end{aligned} \quad (12)$$

where

$$\vartheta_{lk} = \frac{\sqrt{P_k^{\text{ul}}} \beta_{lk}}{\sum_{U_i \in \mathcal{U}} P_i^{\text{ul}} \beta_{li} \delta(\pi_k - \pi_i) + 1},$$

$$\phi_{lj} = \frac{\sqrt{P_j^{\text{ul}}} \beta_{lj} \beta_{lk}}{\sum_{U_i \in \mathcal{U}} P_i^{\text{ul}} \beta_{li} \delta(\pi_k - \pi_i) + 1},$$

$\widetilde{\mathbf{v}}_{lk} = \mathbf{V}_l^* \psi_{\pi_k}$ with $[\widetilde{\mathbf{v}}_{lk}]_m \sim \mathcal{CN}(0, 1)$. In (12), the first term is the desired user's channel, the second term is the leakage from the other users' channels, called the *pilot contamination* (PC) effect due to the pilot signal reuse, and the third term is the noise part. Invoking the orthogonality principle of the MMSE estimator [26], \mathbf{h}_{lk} can be decomposed as $\mathbf{h}_{lk} = \widehat{\mathbf{h}}_{lk} + \widetilde{\mathbf{h}}_{lk}$, where $\widehat{\mathbf{h}}_{lk} \sim \mathcal{CN}(\mathbf{0}, \phi_{lk} \mathbf{I}_M)$ and $\widetilde{\mathbf{h}}_{lk} \sim \mathcal{CN}(\mathbf{0}, (1 - \phi_{lk}) \mathbf{I}_M)$ are mutually independent. Note that the estimated version of \mathbf{g}_{lk} at BS l is given as $\widehat{\mathbf{g}}_{lk} = \sqrt{\beta_{lk}} \widehat{\mathbf{h}}_{lk}$ for $\forall k \in \mathcal{U}_l$ or $\widehat{\mathbf{g}}_{lk} = \mathbf{0}_{M \times 1}$ for $\forall k \notin \mathcal{U}_l$ and thus the estimated version of \mathbf{G} is given as $\widehat{\mathbf{G}} = \left([\widehat{\mathbf{g}}_{lk}]_{l=1, k=1}^{L, K} \right)^H$.

C. Performance Measure

When operation Υ is used, the signal-to-noise ratio (SNR), SNR_k^{Υ} , the signal-to-interference ratio (SIR), SIR_k^{Υ} , and the signal-to-interference-plus-noise ratio (SINR), SINR_k^{Υ} , are respectively defined as

$$\begin{aligned} \text{SNR}_k^{\Upsilon} &= Q_k^{\Upsilon} |\psi_{kk}^{\Upsilon}|^2, \\ \text{SIR}_k^{\Upsilon} &= \frac{Q_k^{\Upsilon} |\psi_{kk}^{\Upsilon}|^2}{\sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\Upsilon} |\psi_{kj}^{\Upsilon}|^2}, \\ \text{SINR}_k^{\Upsilon} &= \frac{Q_k^{\Upsilon} |\psi_{kk}^{\Upsilon}|^2}{\sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\Upsilon} |\psi_{kj}^{\Upsilon}|^2 + 1}. \end{aligned}$$

Obviously, SNR_k^{Υ} , SIR_k^{Υ} , and SINR_k^{Υ} are random variables depending on the realization of the short-term fading and the long-term fading (i.e., realization of locations of users and BSs).

One of the main objectives of this paper is to characterize the asymptotic behavior of SINR_k^{Υ} when the key network parameters such as the number of BSs L , the number of users K , and the number of BS antennas M are scaled up. To do this, we define an auxiliary parameter $N = LM$ as the network size (or equivalently, the total number of antennas in the network) and make the following relations:

$$L = \Theta(N^{\eta_{\text{bs}}}), \quad M = \Theta(N^{\eta_{\text{ant}}}), \quad K = \Theta(N^{\eta_{\text{user}}}), \quad (13)$$

where η_{bs} , η_{user} , and η_{ant} denote the scaling exponents of the numbers of BSs, users, and BS antennas, respectively.⁴ Note that we only consider the case where $0 \leq \eta_{\text{bs}}, \eta_{\text{ant}}, \eta_{\text{user}} \leq 1$ and $\eta_{\text{bs}} + \eta_{\text{ant}} = 1$ by definition. Then, the asymptotic performance of the network can be characterized as follows.

⁴Note that the network behavior when the network size (N) is sufficiently large is of interest in this paper.

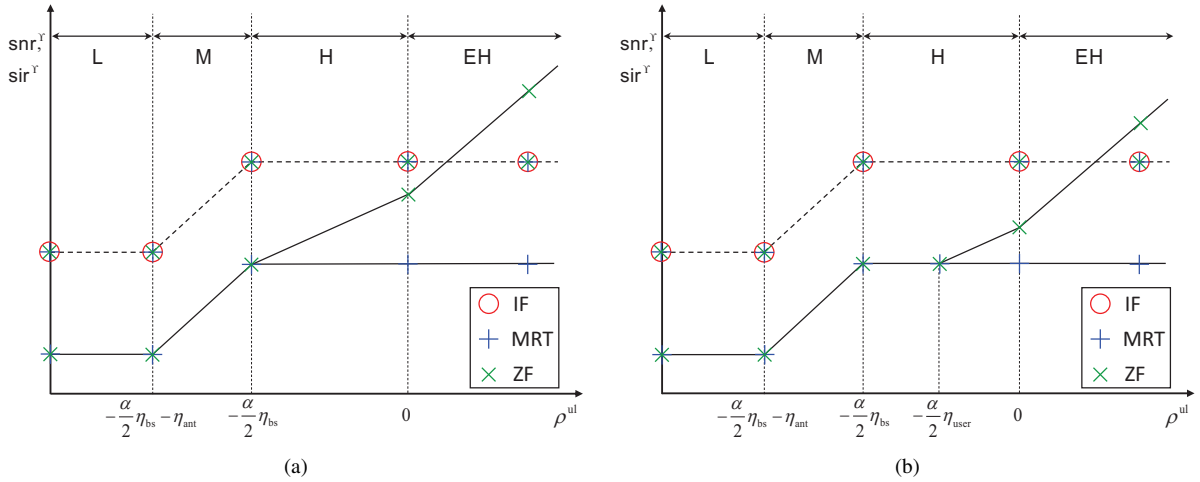


Fig. 2. SNR (dashed line) or SIR (solid line) scaling exponents for IF (o-marker), MRT (+-marker), or ZF (x-marker) operations according to ρ^{ul} : (a) $\eta_{\text{user}} > \eta_{\text{bs}}$ and (b) $\eta_{\text{user}} \leq \eta_{\text{bs}}$.

Definition 1 (Performance Measure) The scaling exponent of the SINR of operation Υ , sinr^Υ , is the order of growth of the SINR of a randomly selected user as N increases such that, for any $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} \Pr \left(\left| \frac{\log \text{SINR}_k^\Upsilon}{\log N} - \text{sinr}^\Upsilon \right| < \epsilon \right) = 1. \quad (14)$$

Manipulating (14), we can also obtain

$$\lim_{N \rightarrow \infty} \Pr \left(N^{\text{sinr}^\Upsilon - \epsilon} < \text{SINR}_k^\Upsilon < N^{\text{sinr}^\Upsilon + \epsilon} \right) = 1,$$

which implies that the SINR of a randomly selected user is bounded by $[N^{\text{sinr}^\Upsilon - \epsilon}, N^{\text{sinr}^\Upsilon + \epsilon}]$ as N increases. Invoking our order notations, we can simply write $\text{SINR}_k^\Upsilon = \Theta(N^{\text{sinr}^\Upsilon})$ and the sum-rate served by a network can also be written as $C_\Sigma = \Theta(N^{\eta_{\text{user}} \log_2(1 + N^{\text{sinr}^\Upsilon})})$. Similarly, we define snr^Υ and sir^Υ as the scaling exponents of the SNR and SIR, respectively. Note that it is sufficient to find snr^Υ and sir^Υ and then $\text{sinr}^\Upsilon = \min\{\text{snr}^\Upsilon, \text{sir}^\Upsilon\}$ is obtained since SINR_k^Υ is equal to the harmonic mean of SNR_k^Υ and SIR_k^Υ , i.e., $(\text{SINR}_k^\Upsilon)^{-1} = (\text{SNR}_k^\Upsilon)^{-1} + (\text{SIR}_k^\Upsilon)^{-1}$.

IV. SCALING EXPONENTS OF THE SINR

The major merit of the LS-CRAN is that it can decrease the transmit power consumption so that it is suitable for an energy-efficient communication system. So, the total transmit power is a key constraint in a future cellular system. In order to limit the total transmit power of uplink or downlink, similarly as in (13), we make an additional asymptotic relationship as

$$P_j^{\text{ul}} = \Theta(N^{\rho^{\text{ul}}}) \text{ for UL and } P_j^{\text{dl}} = \Theta(N^{\rho^{\text{dl}}}) \text{ for DL,} \quad (15)$$

where ρ^{ul} and ρ^{dl} denote the scaling exponents of the UL and DL transmit powers, respectively. Also, the total transmit power is $P_\Sigma = \sum_{j=1}^K (P_j^{\text{ul}} + P_j^{\text{dl}}) = \Theta(N^{\eta_{\text{user}} + \max\{\rho^{\text{ul}}, \rho^{\text{dl}}\}})$ ⁵.

⁵Note that the power constraint used in this paper can include both the total power constraint and the per-node power constraint so that the derived asymptotic result is valid even if tighter per-node power constraint is assumed.

In order to quantify the effect of the CSI accuracy according to the UL transmit power, we consider the following four regimes.

- Case EH ($\rho^{\text{ul}} \geq 0$): the UL transmit power is sufficiently high so that every BS can acquire the accurate CSI of a randomly selected users.
- Case H ($-\frac{\alpha}{2}\eta_{\text{bs}} \leq \rho^{\text{ul}} < 0$): each BS can acquire the accurate CSI of a randomly selected users within a distance of $\Theta(N^{\rho^{\text{ul}}/\alpha})$.
- Case M ($-\frac{\alpha}{2}\eta_{\text{bs}} - \eta_{\text{ant}} \leq \rho^{\text{ul}} < -\frac{\alpha}{2}\eta_{\text{bs}}$): randomly selected user's CSI is erroneous even at the nearest BS but is still meaningful for providing an array gain.
- Case L ($\rho^{\text{ul}} < -\frac{\alpha}{2}\eta_{\text{bs}} - \eta_{\text{ant}}$): randomly selected user's CSI becomes quite poor even at the nearest BS so that no array gain can be provided.

Theorem 1: Suppose that IF operation is used with a full association and no pilot reuse. Then, the scaling exponents are respectively given by

$$\text{snr}^{\text{if}} = \rho^{\text{dl}} + \frac{\alpha}{2}\eta_{\text{bs}} + \Xi, \quad (16)$$

$$\text{sir}^{\text{if}} = \infty, \quad (17)$$

$$\text{sinr}^{\text{if}} = \text{snr}^{\text{if}}, \quad (18)$$

where $\Xi = (\rho^{\text{ul}} + \frac{\alpha}{2}\eta_{\text{bs}} + \eta_{\text{ant}})^+ - (\rho^{\text{ul}} + \frac{\alpha}{2}\eta_{\text{bs}})^+$ denotes the array gain and $(x)^+ = \max\{x, 0\}$.

Proof: Please see Appendix B. \square

Theorem 1 is also illustrated in Figs. 2 and 3 according to ρ^{ul} and η_{bs} , respectively. Intuitively, the SNR of IF operation is composed of the three parts as

$$\text{SNR}_k^{\text{if}} \asymp \underbrace{N^{\rho^{\text{dl}}}}_{\text{DL transmit power}} \times \underbrace{N^{\frac{\alpha}{2}\eta_{\text{bs}}}}_{\text{Densification gain}} \times \underbrace{N^{\Xi}}_{\text{Array gain}}, \quad (19)$$

where the first part is the DL transmit power, the second part is the densification gain which comes from the decrease of the access distance of $\Theta(N^{-\frac{1}{2}\eta_{\text{bs}}})$ and the last part is the array gain

of a coherent transmission which depends on the CSI accuracy and thus the UL transmit power.

Remark 1 (SNR behavior of IF operation) Fig. 2 reveals how the UL transmit power affects on the SNR behavior. In EH and H, the full array gain ($\Xi = \eta_{\text{ant}}$) is achieved so that an additional UL transmit power does not improve the quality of DL service in the network, i.e., is wasteful. In M, a partial array gain ($0 \leq \Xi < \eta_{\text{ant}}$) depending on the UL transmit power is obtained so that the network total power needs to be consumed by considering both the DL transmit power and the CSI accuracy. In L, no array gain ($\Xi = 0$) is obtained due to poor CSI accuracy so that the quality of DL service becomes irrelevant to the UL transmit power and its performance is identical to the random beamforming without small-scale CSIs in [27].

Fig. 3 shows the SNR behavior according to the BS scaling exponent. It turns out that additional BSs (even with smaller number of BS antennas at each BS) are always beneficial but the slope of snr^{if} (vs. η_{bs}) varies according to ρ^{ul} . The slope becomes $\alpha/2$ in L, $\alpha - 1$ in M, and $\alpha/2 - 1$ in H or EH, which implies that additional BSs (while keeping the network size fixed) become the most effective in M because the additional BSs improve not only the densification gain but also the array gain and the least effective in H or EH because only the densification gain is improved. ■

Theorem 2: Suppose that MRT operation is used with a full association and no pilot reuse. Then, the scaling exponents are respectively given by

$$\text{snr}^{\text{mrt}} = \text{snr}^{\text{if}}, \quad (20)$$

$$\text{sir}^{\text{mrt}} = \text{snr}^{\text{mrt}} - \Delta^{\text{mrt}}, \quad (21)$$

$$\text{sinr}^{\text{mrt}} = \text{snr}^{\text{mrt}} - (\Delta^{\text{mrt}})^+, \quad (22)$$

where $\Delta^{\text{mrt}} = \rho^{\text{dl}} + \frac{\alpha}{2} \min\{\eta_{\text{bs}}, \eta_{\text{user}}\} + (\eta_{\text{user}} - \eta_{\text{bs}})^+$.

Proof: Please see Appendix C. □

Theorem 2 is also illustrated in Figs. 2 and 3 according to ρ^{ul} and η_{bs} , respectively. Interestingly, although snr^{mrt} is identical to snr^{if} , the gap between sir^{mrt} and snr^{mrt} , denoted by Δ^{mrt} , changes according to the sign of $\eta_{\text{bs}} - \eta_{\text{user}}$. From the definition of Δ^{mrt} , the interference caused in MRT operation at a randomly selected user k can be represented as

$$I_k \asymp N^{\Delta^{\text{mrt}}} = \underbrace{N^{\rho^{\text{dl}}}}_{\text{DL transmit power}} \times \underbrace{N^{\frac{\alpha}{2} \min\{\eta_{\text{bs}}, \eta_{\text{user}}\}}}_{\text{Received power of a dominant interferer}} \times \underbrace{N^{(\eta_{\text{user}} - \eta_{\text{bs}})^+}}_{\text{\# of dominant interferers}}, \quad (23)$$

where the second and third parts depend on the sign of $\eta_{\text{user}} - \eta_{\text{bs}}$. When $\eta_{\text{user}} \geq \eta_{\text{bs}}$, one BS should simultaneously serve $\Theta(N^{\eta_{\text{user}} - \eta_{\text{bs}}})$ users apart by $\Theta(N^{-\frac{1}{2}\eta_{\text{bs}}})$ so that there are $\Theta(N^{\eta_{\text{user}} - \eta_{\text{bs}}})$ interferer whose received power is $\Theta(N^{\frac{\alpha}{2}\eta_{\text{bs}}})$. When $\eta_{\text{user}} < \eta_{\text{bs}}$, the dominant interference comes from the BS apart by $\Theta(N^{-\frac{1}{2}\eta_{\text{user}}})$ so that the received power of the dominant interference is $\Theta(N^{\frac{\alpha}{2}\eta_{\text{user}}})$.

Remark 2 (Asymptotical optimality of MRT operation) In order for MRT operation to behave as IF operation asymptotically

(i.e., $\Delta^{\text{mrt}} \leq 0$), the DL transmit power should be limited as

$$\rho^{\text{dl}} \leq \begin{cases} -\frac{\alpha}{2}\eta_{\text{user}}, & \text{if } \eta_{\text{bs}} \geq \eta_{\text{user}}, \\ -\frac{\alpha}{2}\eta_{\text{bs}} - (\eta_{\text{user}} - \eta_{\text{bs}}), & \text{if } \eta_{\text{bs}} < \eta_{\text{user}}, \end{cases} \quad (24)$$

which gives us the following insights.

- For MRT operation, the IF optimality condition depends only on the numbers of users (η_{user}) and BSs (η_{bs}) and the DL transmit power (ρ^{dl}), but is independent to the number of BS antennas (η_{ant}) and the UL transmit power (ρ^{ul}).⁶
- When the number of antennas in a BS is much larger than the total number of users in a multi-cell network, MRT operation becomes interference-free asymptotically as expected in literature [3]. However, this is of little interest because the network size is too large (or the number of users is too small). Instead, Theorem 2 and (24) gives more insightful IF optimality condition for MRT operation being asymptotically interference-free in a multi-cell network.
- The above asymptotical optimality may become meaningful (i.e., positive SNR scaling exponent) in case of an ultra-dense deployment of BS ($\eta_{\text{bs}} \geq \eta_{\text{user}}$ or $\rho^{\text{ul}} + (\alpha/2)\eta_{\text{bs}} \geq 0$). ■

Remark 3 (SIR behavior of MRT operation) As can be seen from Fig. 3, additional BSs are beneficial (while fixing N) in most cases and the slope of sir^{mrt} (vs. η_{bs}) can be 0, 1, $\alpha/2 - 1$ and $\alpha/2$. Interestingly, when $-(2/\alpha)\rho^{\text{ul}} \leq \eta_{\text{bs}} \leq \eta_{\text{user}}$, the slope of sir^{mrt} becomes 0, which means that additional BSs are wasteful as long as η_{bs} is within that interval. On the other hand, the slope becomes $\alpha/2$ in M and in L if $\eta_{\text{bs}} > \eta_{\text{user}}$, in which additional BSs are the most effective. ■

Theorem 3: Suppose that ZF operation is used with a full association and no pilot reuse. Then, the scaling exponents are respectively given by

$$\text{snr}^{\text{zf}} = \text{snr}^{\text{if}}, \quad (25)$$

$$\text{sir}^{\text{zf}} = \text{snr}^{\text{zf}} - \Delta^{\text{zf}} \quad (26)$$

$$\text{sinr}^{\text{zf}} = \text{snr}^{\text{zf}} - (\Delta^{\text{zf}})^+, \quad (27)$$

where $\Delta^{\text{zf}} = \Delta^{\text{mrt}} - (1 - (2/\alpha)) ((\alpha/2) \min\{\eta_{\text{bs}}, \eta_{\text{user}}\} + \rho^{\text{ul}})^+ - \frac{2}{\alpha} (\rho^{\text{ul}})^+$.

Proof: Please see Appendix D. □

Theorem 3 is also illustrated in Figs. 2 and 3 according to ρ^{ul} and η_{bs} , respectively. Although the SNR exponent is identical to that in MRT operation, the interference is reduced as the CSI accuracy improves so that the gap between snr^{zf} and sir^{zf} , Δ^{zf} , varies as ρ^{ul} increases. Again, from the definition of Δ^{zf} , the interference in MRT operation at a randomly selected user k for

⁶Such an independence on the number of BS antennas or the UL transmit power comes from the fact that the array gain is already reflected both on snr^{mrt} and sir^{mrt} .

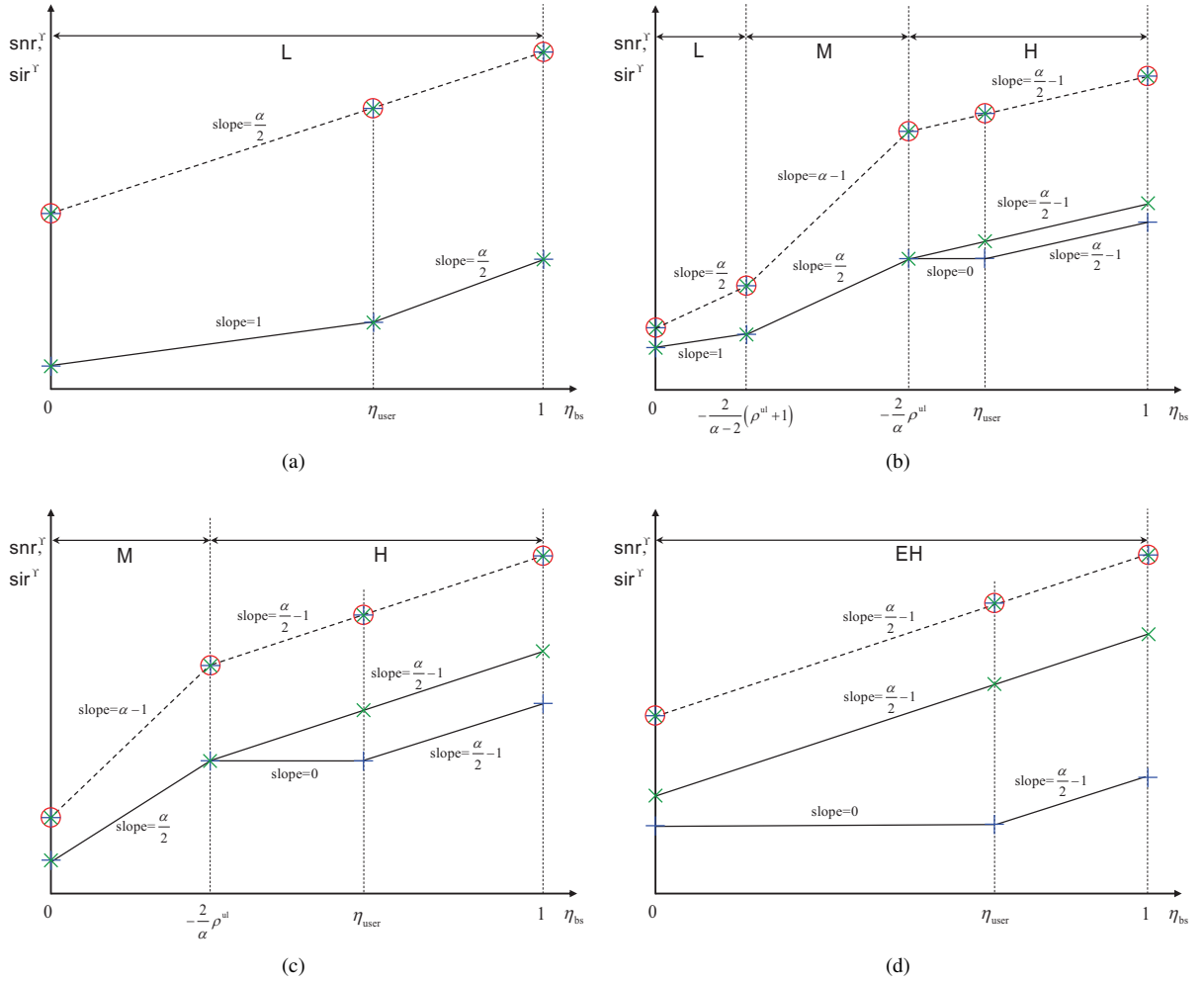


Fig. 3. SNR (dashed line) or SIR (solid line) scaling exponents for IF (o-marker), MRT (+-marker), or ZF (x-marker) operations according to η_{bs} : (a) $\rho^{ul} < -\frac{\alpha}{2}$, (b) $-\frac{\alpha}{2} \leq \rho^{ul} < -1$, (c) $-1 \leq \rho^{ul} < 0$, and (d) $0 \leq \rho^{ul}$.

$\eta_{user} \geq \eta_{bs}$ can be represented as

$$I_k \asymp N^{\Delta^{zf}}$$

$$= \begin{cases} \underbrace{N^{\rho^{dl}}}_{\text{DL transmit power}} \times \underbrace{N^{\frac{\alpha}{2}\eta_{bs}}}_{\text{Received power of a dominant interferer}} \times \underbrace{N^{\eta_{user} - \eta_{bs}}}_{\text{\# of dominant interferers}}, & \text{L, M,} \\ N^{\rho^{dl}} \times N^{-\rho^{ul}} \times N^{\frac{2}{\alpha}\rho^{ul} + \eta_{user}}, & \text{H,} \\ \underbrace{N^{\rho^{dl}}}_{\text{DL transmit power}} \times \underbrace{N^{-\rho^{ul}}}_{\text{Received power of a dominant interferer}} \times \underbrace{N^{\eta_{user}}}_{\text{\# of dominant interferers}}, & \text{EH.} \end{cases} \quad (28)$$

The CSI of the users whose access distance is $o(N^{\rho^{ul}/\alpha})$ is accurately estimated at the nearest BS, while the CSI of the users whose access distance is $\Omega(N^{\rho^{ul}/\alpha})$ is poor. So, when $\rho^{ul} < -(\alpha/2)\eta_{bs}$, i.e., L or M, the CSIs of all users are poorly estimated at all BSs because the access distance is $\Omega(N^{-\frac{1}{2}\eta_{bs}})$. Thus, the interference cancellation is not effective so that the interference in ZF operation is asymptotically the same to that in MRT operation. However, when $\rho^{ul} \geq -(\alpha/2)\eta_{bs}$, i.e., H or EH, accurate CSI is available so that the cancellation operation becomes effective. In H, the dominant interference comes from the BS apart by $\Theta(N^{\rho^{ul}/\alpha})$ so that the received power of

the dominant interference is $\Theta(N^{-\rho^{ul}})$. The dominant interferers are located in the doughnut of radii of $\Theta(N^{\rho^{ul}/\alpha})$ and $\Theta(N^{\rho^{ul}/\alpha + \epsilon})$ with an arbitrarily small ϵ so that the number of interferers is $\Theta(N^{\frac{2}{\alpha}\rho^{ul} + \eta_{user}})$. In EH, the number of dominant interferers becomes $\Theta(N^{\eta_{user}})$, while the received power of them is still $\Theta(N^{-\rho^{ul}})$ which can be explained as follows. Once a BS acquires sufficiently accurate CSI of a user, the interference caused from the user becomes proportional to the channel estimation error, i.e., inversely proportional to the UL transmit power. Thus, in H and EH, the slopes of the reduction in ZF operation are $(1 - 2/\alpha)$ and 1 with respect to ρ^{ul} , respectively, as shown in Fig. 3. In EH, every BS can acquire all user's CSI with a sufficiently good accuracy so that the slope becomes 1. In H, however, BSs far from each user (outside the circle with radius of $\Theta(N^{\rho^{ul}/\alpha})$) cannot obtain its CSI accurately so that ZF operation does not reduce the interference from those BSs. Note that sir^{zf} does not improve as ρ^{ul} increases within $[-(\alpha/2)\eta_{bs}, -(\alpha/2)\eta_{user}]$ in H or EH if $\eta_{user} \geq \eta_{bs}$ because no actual interference reduction happens during that interval due to the low user density.

Remark 4 (Asymptotical optimality of ZF operation) In order for ZF operation to behave as IF operation asymptotically

(i.e., $\Delta^{\text{zf}} \leq 0$), the DL transmit power should be limited as follows.

If $\eta_{\text{bs}} \geq \eta_{\text{user}}$,

$$\rho^{\text{dl}} \leq \begin{cases} \rho^{\text{ul}} - \eta_{\text{user}}, & \text{if EH,} \\ -\frac{\alpha}{2}\eta_{\text{user}} + \left(1 - \frac{2}{\alpha}\right) \left(\frac{\alpha}{2}\eta_{\text{user}} + \rho^{\text{ul}}\right)^+, & \text{if H,} \\ -\frac{\alpha}{2}\eta_{\text{user}}, & \text{if M or L,} \end{cases}$$

or if $\eta_{\text{bs}} < \eta_{\text{user}}$

$$\rho^{\text{dl}} \leq \begin{cases} \rho^{\text{ul}} - \eta_{\text{user}}, & \text{if EH,} \\ -\eta_{\text{user}} + \left(1 - \frac{2}{\alpha}\right) \rho^{\text{ul}}, & \text{if H,} \\ -\frac{\alpha}{2}\eta_{\text{bs}} - (\eta_{\text{user}} - \eta_{\text{bs}}), & \text{if M or L,} \end{cases}$$

which gives us the following insights.

- Unlike MRT operation, the IF condition of ZF operation depends on the UL transmit power (ρ^{ul}) as well as η_{bs} , η_{user} , and ρ^{dl} , but is still independent to the number of BS antennas (η_{ant}).
- $\Delta^{\text{zf}} = \Delta^{\text{mrt}}$ in M and L while $\Delta^{\text{zf}} \leq \Delta^{\text{mrt}}$ in EH or H, which implies that multi-cell cooperation using ZF operation is useless without a sufficient CSI accuracy. As the UL transmit power increases, ZF operation begins to further reduce the interference caused from other BSs' users as shown in Fig. 2.

Remark 5 (SIR behavior of ZF operation) As can be seen again from Fig. 3, additional BSs (while fixing N) are always beneficial and although the slope of sir^{zf} vs. η_{bs} is identical to that of sir^{mrt} vs. η_{bs} in L or M, it remains $\alpha/2 - 1$ in H or EH, unlike MRT operation. ■

V. CONCLUSION

This paper presented a comprehensive and rigorous asymptotic analysis on the performance of the large-scale cloud radio access network (LS-CRAN). As a main performance measure, the scaling exponent of the signal-to-interference-plus-noise ratio (SINR) was defined and derived as a function of key network parameters, 1) the number of BS, $L = \Theta(N^{\eta_{\text{bs}}})$, 2) the number of BS antennas $M = \Theta(N^{\eta_{\text{ant}}})$, 3) the number of single-antenna users $K = \Theta(N^{\eta_{\text{user}}})$, 4) the uplink (UL) transmit power, $P_k^{\text{ul}} = \Theta(N^{\rho^{\text{ul}}})$, 5) the DL transmit power, $P^{\text{dl}} = \Theta(N^{\rho^{\text{dl}}})$, and the distance-dependent pathloss exponent α , when interference-free (IF), maximum ratio transmission (MRT), or zero-forcing (ZF) operation is applied.

Then, we show that when MRT or ZF operation becomes interference-free, i.e., order-optimal with the three practical constraints. By considering limited transmit power only, MRT operation is shown to become interference-free only when the DL transmit power is less than a threshold, but the threshold is too low to be meaningful. Also, ZF operation is shown to become interference-free at meaningful DL transmit power so that as higher UL transmit power is provided, more users can be associated and the data rate per user can be increased simultaneously while keeping the order-optimality.

APPENDIX A PRELIMINARY RESULTS

Here, we provide key preliminary results to prove theorems in this paper. First, we state the method to derive the scaling exponent of a random measure given by a sum of infinitely many i.i.d. random variables.

Lemma 1. Let x_1, x_2, \dots be a sequence of i.i.d. real-valued random variables with common distribution function of $F(x)$ and $h(x)$ be a non-negative integrable function. Suppose that $h(n)$ has a finite scaling exponent. Then, as $n \rightarrow \infty$, we have

$$\begin{aligned} \sum_{k=1}^n h(x_k) &= \Theta \left(\log n \int_0^1 n^t h(F^{-1}(n^{-1+t})) dt \right) \\ &= \Theta \left(n \int_{F^{-1}(\frac{1}{n})}^{F^{-1}(1)} h(u) f(u) du \right). \end{aligned} \quad (29)$$

where $F^{-1}(x)$ is an inverse (or quantile) function of $F(x)$ and $f(x) = \frac{d}{dx}F(x)$ is the probability density function.

Proof: Suppose that y_1, y_2, \dots, y_n are i.i.d. uniform random variables on $[0, 1]$. Then, for any positive small ϵ , the following holds:

$$\begin{aligned} \Pr \left(\min_{1 \leq k \leq n} y_k > \frac{1}{n^{1+\epsilon}} \right) &= \left(1 - \frac{1}{n^{1+\epsilon}} \right)^n \\ &\rightarrow 1, \text{ as } n \rightarrow \infty, \end{aligned} \quad (30)$$

which implies that all of y_1, \dots, y_n are included in the interval $(n^{-1-\epsilon}, 1]$, but not included in the interval $[0, n^{-1-\epsilon}]$ in probability. By dividing the interval $(n^{-1-\epsilon}, 1]$ into $\lceil g(n) \rceil$ intervals, $\mathcal{S}_1 = (n^{-1-\epsilon}, n^{u_1}]$, $\mathcal{S}_2 = (n^{u_1}, n^{u_2}]$, \dots , $\mathcal{S}_{\lceil g(n) \rceil} = (n^{u_{\lceil g(n) \rceil-1}}, 1]$, where $u_i = -1 + \frac{i}{\lceil g(n) \rceil}$ for $i = 1, 2, \dots, \lceil g(n) \rceil$, $\lceil \cdot \rceil$ denotes the ceiling operation, and $g(n)$ is a positive function of n . Let $\mathcal{Y}_i = \{y_j | y_j \in \mathcal{S}_i\}$ for $i = 1, 2, \dots, \lceil g(n) \rceil$. Then, for $g(n) = \Theta(\log(n))$,

$$|\mathcal{Y}_i| = \Theta(n^{1+u_i}) = \Theta \left(n^{\frac{i}{\lceil g(n) \rceil}} \right) \quad (31)$$

and

$$\begin{aligned} \frac{1}{\lceil g(n) \rceil} \sum_{i=1}^{\lceil g(n) \rceil} |\mathcal{Y}_i| &= \frac{n}{\lceil g(n) \rceil} \\ &= \Theta \left(\frac{n}{\log n} \right). \end{aligned} \quad (32)$$

On the other hand, the left side hand of (32) can be rewritten as

$$\frac{1}{\lceil g(n) \rceil} \sum_{i=1}^{\lceil g(n) \rceil} |\mathcal{Y}_i| \asymp \frac{1}{\lceil g(n) \rceil} \sum_{i=1}^{\lceil g(n) \rceil} n^{\frac{i}{\lceil g(n) \rceil}}, \quad (33)$$

where the right hand side may be considered as a Riemann sum (although not guaranteed because the integrand is also a function of n) and if it is, it becomes $\int_0^1 n^t dt = \Theta \left(\frac{n}{\log n} \right)$, which coincides with (32). Thus, in this particular case, the scaling exponent of the left hand side of (32) can be obtained by first

changing the summation into an integral form as if it is a Riemman sum and then taking the scaling exponent from the integral form. Also, it is easily seen that this property still holds when $q(n)$ i.i.d. random variables are considered instead of n i.i.d. random variables as long as $q(n)$ has a finite scaling exponent. For example, when n^r i.i.d. random variables are considered, the right hand side of (32) becomes $\Theta\left(n^{r-1+\frac{1}{\lceil g(n) \rceil}}\right)$ so that both the direct calculation as in (32) and the above method yield the same result, $\Theta\left(\frac{n^r}{\log n}\right)$.

Now, assume that $h(x)$ is a monotonically decreasing non-negative integrable function and $h(n)$ has a finite scaling exponent. First, we consider the upper-bound:

$$\begin{aligned} \frac{1}{\lceil g(n) \rceil} \sum_{k=1}^n h(y_k) &= \frac{1}{\lceil g(n) \rceil} \sum_{i=1}^{\lceil g(n) \rceil} \sum_{y_k \in \mathcal{Y}_i} h(y_k) \\ &\leq \frac{1}{\lceil g(n) \rceil} \sum_{i=1}^{\lceil g(n) \rceil} |\mathcal{Y}_i| h(n^{u_i-1}) \\ &= \Theta\left(\frac{n^{\frac{1}{\lceil g(n) \rceil}}}{\lceil g(n) \rceil} \sum_{i=1}^{\lceil g(n) \rceil} n^{\frac{i-1}{\lceil g(n) \rceil}} h\left(n^{-1+\frac{i-1}{\lceil g(n) \rceil}}\right)\right) \\ &= \Theta\left(n^{\frac{1}{\lceil g(n) \rceil}} \int_0^1 n^t h(n^{-1+t}) dt\right), \end{aligned} \quad (34)$$

where the last equality comes from the discussion in the previous paragraph. Similarly, we obtain the lower-bound:

$$\frac{1}{\lceil g(n) \rceil} \sum_{k=1}^n h(y_k) \geq \Theta\left(\int_0^1 n^t h(n^{-1+t}) dt\right) \quad (35)$$

so that the two bounds are asymptotically tight (tight in terms of the scaling exponent) because $1/\lceil g(n) \rceil$ is ignorable for $g(n) = \Theta(\log n)$. Finally, by using the continuous mapping theorem [28] and changing the variable $u = F^{-1}(n^{-1+t})$, we obtain (29). \square

Lemma 2. Let Φ and Ψ be two independent homogeneous PPPs over a finite region $\mathcal{R} \subset \mathbb{R}^2$ with densities λ and $\mu = \Theta(\lambda^\delta)$ for $\delta > 0$, respectively. Assume that an arbitrarily chosen point $Y_0 \in \Psi$ is given. Then, the following holds for $\Phi_p(Z) \triangleq \{X \in \Phi \mid |X - Z| = O(\lambda^p)\}$,

1) For $\alpha > 2$ and $-1/2 < p \leq 0$, as $\lambda \rightarrow \infty$,

$$\sum_{X \in \Phi_p(Y_0)} |X - Y_0|^{-\alpha} = \Theta(\lambda^{\frac{\alpha}{2}}). \quad (36)$$

2) For $\alpha_0, \alpha_1 > 2$, as $\lambda \rightarrow \infty$,

$$\begin{aligned} \sum_{Y \in \Psi \setminus \{Y_0\}} \sum_{X \in \Phi \setminus (\Phi_p(Y_0) \cup \Phi_p(Y))} |X - Y_0|^{-\alpha_0} |X - Y|^{-\alpha_1} \\ = \Theta(\lambda^s), \end{aligned} \quad (37)$$

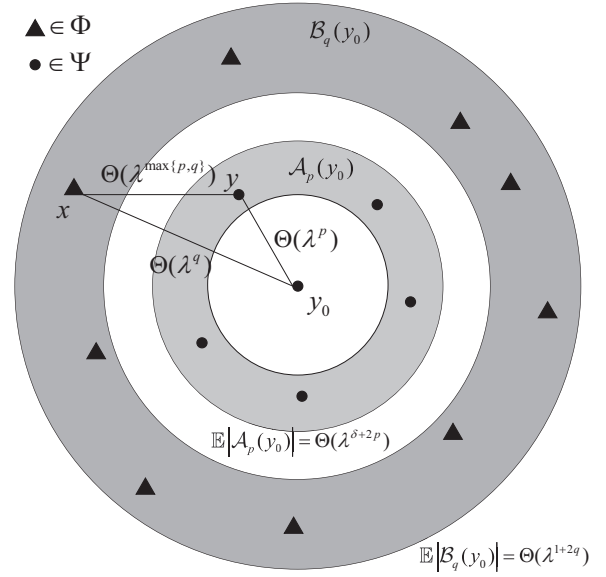


Fig. 4. Conceptual illustration for the proof of Lemma 2.

where s is given by

$$s = \begin{cases} \frac{\alpha_{\max}}{2} + \frac{\alpha_{\min}}{2} \delta, & \text{if } z < -\frac{1}{2}, \\ 1 + (2 - \alpha_{\max})z + \frac{\alpha_{\min}}{2} \delta, & \text{if } -\frac{1}{2} \leq z < -\frac{1}{2} \delta, \\ 1 + \delta + (4 - \alpha_{\max} - \alpha_{\min})z, & \text{if } z \geq -\frac{1}{2} \delta, \end{cases}$$

for $\delta < 1$, or

$$s = \begin{cases} \frac{\alpha_{\max} + \alpha_{\min}}{2} + \delta - 1, & \text{if } z < -\frac{1}{2}, \\ 1 + \delta + (4 - \alpha_{\max} - \alpha_{\min})z, & \text{if } z \geq -\frac{1}{2}, \end{cases}$$

for $\delta \geq 1$, and

$$\sum_{Y \in \Psi \setminus \{Y_0\}} \sum_{X \in \Phi_p(Y_0) \cap \Phi_p(Y)} |X - Y_0|^{-\alpha_0} |X - Y|^{-\alpha_1} = \Theta(\lambda^t), \quad (38)$$

where t is given by

$$t = \begin{cases} \frac{\alpha_{\max}}{2} + \frac{\alpha_{\min}}{2} \delta, & \text{if } z > -\frac{1}{2} \delta > -\frac{1}{2}, \\ \frac{\alpha_{\max} + \alpha_{\min}}{2} + \delta - 1, & \text{if } z > -\frac{1}{2} > -\frac{1}{2} \delta, \\ \frac{\alpha_{\max} + \alpha_{\min}}{2} + \delta + 2z, & \text{if } -\frac{1}{2} \delta < z < -\frac{1}{2}, \\ -\infty, & \text{otherwise,} \end{cases}$$

for $\alpha_{\max} = \max\{\alpha_0, \alpha_1\}$ and $\alpha_{\min} = \min\{\alpha_0, \alpha_1\}$.

Proof: 1) To prove the first part of Lemma 2, we construct the lower-bound and upper-bound as follows.

$$\begin{aligned} \max_{X \in \mathcal{A}_p(Y_0)} |X - Y_0|^{-\alpha} &\leq \sum_{X \in \mathcal{A}_p(Y_0)} |X - Y_0|^{-\alpha} \\ &\leq \sum_{X \in \Phi} |X - Y_0|^{-\alpha}. \end{aligned}$$

Here, the lower-bound is obviously $\Theta(\lambda^{\alpha/2})$ because the minimum distance is $\min_{X \in \mathcal{A}_p(Y_0)} |X - Y_0| = \Theta(\lambda^{-1/2})$. To complete the proof of the first part, it is sufficient to prove $\sum_{X \in \Phi} |X - Y_0|^{-\alpha} = \Theta(\lambda^{\alpha/2})$. Let $h(x) = x^{-\alpha}$. For a randomly selected $X \in \Phi$, its contact distribution function is given by

$$F(x; Y_0) = \Pr(|X - Y_0| < x) = \frac{v(b(Y_0, X) \cap \mathcal{R})}{v(\mathcal{R})},$$

where $b(a, r) = \{b \in \mathbb{R}^2 \mid |a - b| \leq r\}$ and $v(\cdot)$ denotes the Lebesgue measure (i.e. area measure). Since $v(\mathcal{R})$ is finite and independent to x , $F(x) = c'x^2 + o(x^2)$ and thus $h(F^{-1}(x)) = cx^{-\frac{\alpha}{2}} + o(x^{-\frac{\alpha}{2}})$, where c is independent to x . From Lemma 1, we have

$$\begin{aligned} \sum_{k=1}^n h(x_k) &= \Theta \left(\log n \int_0^1 n^{(1-\frac{\alpha}{2})t + \frac{\alpha}{2}} dt \right) \\ &= \Theta(n) + \Theta(n^{\frac{\alpha}{2}}) \\ &= \Theta(n^{\frac{\alpha}{2}}), \end{aligned} \quad (39)$$

where the last equality comes from $\alpha > 2$ and the fact $n = \Theta(\lambda)$ completes the proof of the first part.

2) Define two sets $\mathcal{A}_p(Z) \triangleq \{X \in \Phi \mid |X - Z| \stackrel{p}{\leq} \Theta(\lambda^p)\}$ and $\mathcal{B}_q(Z) \triangleq \{y \in \Psi \mid |Y - Z| \stackrel{q}{\leq} \Theta(\lambda^q)\}$ for a given point Z . Since Φ and Ψ are PPPs with density λ and μ , respectively, $\mathbb{E}|\mathcal{A}_p(Z)| = \Theta(\lambda^{1+2p})$ and $\mathbb{E}|\mathcal{B}_q(Z)| = \Theta(\lambda^{\delta+2q})$. Note that $\Pr\{|\mathcal{A}_p(Z)| > 0\} \rightarrow 1$ if and only if $-\frac{1}{2} < p \leq 0$ and $\Pr\{|\mathcal{B}_q(Z)| > 0\} \rightarrow 1$ if and only if $-\frac{\delta}{2} < q \leq 0$. Assume that $Y_1 \in \Psi$ is given with $|Y_0 - Y_1| = \Theta(\lambda^d)$ and $\alpha_0 \geq \alpha_1$. For any $X \in \mathcal{A}_p(Y_0)$, $|X - Y_1| = \Theta(\lambda^{\max\{p, d\}})$ so that $|X - Y_0|^{-\alpha_0} |X - Y_1|^{-\alpha_1} = \Theta(\lambda^{-\alpha_0 p - \alpha_1 \max\{p, d\}})$. Then, we have

$$\begin{aligned} &\sum_{Y \in \mathcal{B}_d(Y_1)} \sum_{X \in \mathcal{A}_p(Y_0) \setminus (\Phi_p(Y_0) \cup \Phi_p(Y))} |X - Y_0|^{-\alpha_0} |X - Y_1|^{-\alpha_1} \\ &= \Theta \left(\lambda^{\delta+2d+1+2p-\alpha_0 p - \alpha_1 \max\{d, p\}} \right), \end{aligned} \quad (40)$$

with the constraints of $\max\{z, -\frac{1}{2}\} < p \leq 0$ and $-\frac{\delta}{2} \leq d \leq 0$. Similarly, we have

$$\begin{aligned} &\sum_{Y \in \mathcal{B}_d(Y_1)} \sum_{X \in \mathcal{A}_p(Y_0) \cap (\Phi_p(Y_0) \cap \Phi_p(Y))} |X - Y_0|^{-\alpha_0} |X - Y_1|^{-\alpha_1} \\ &= \Theta \left(\lambda^{\delta+2d+1+2p-\alpha_0 p - \alpha_1 \max\{d, p\}} \right), \end{aligned} \quad (41)$$

with the constraints of $-\frac{1}{2} < p \leq \min\{z, 0\}$ and $-\frac{\delta}{2} \leq d \leq 0$. Then, the second part is easily given by obtaining the supremums of (40) and (41) under the above constraints.

APPENDIX B PROOF OF THEOREM 1

Since the interference is removed by the Genie without any cost, $\text{sir}^{\text{if}} = \infty$. Inserting $\mathbf{F}^{\text{if}} = \hat{\mathbf{G}}^H$ into (7), $\psi_{kk}^{\text{mrt}} =$

$\sum_{X_l \in \mathcal{X}_k} \beta_{lk} \mathbf{h}_{lk}^H \hat{\mathbf{h}}_{lk}$ and the square of which is given by

$$\begin{aligned} |\psi_{kk}^{\text{if}}|^2 &\asymp \sum_{X_l \in \mathcal{X}_k} \beta_{lk}^2 |\mathbf{h}_{lk}^H \hat{\mathbf{h}}_{lk}|^2 \\ &= \sum_{X_l \in \mathcal{X}_k} \left(\frac{P_k^{\text{ul}}}{P_k^{\text{ul}} \beta_{lk} + 1} \right)^2 \beta_{lk}^4 |\mathbf{h}_{lk}^H \mathbf{h}_{lk}|^2 \\ &\quad + \sum_{X_l \in \mathcal{X}_k} \left(\frac{\sqrt{P_k^{\text{ul}}}}{P_k^{\text{ul}} \beta_{lk} + 1} \right)^2 \beta_{lk}^3 |\mathbf{h}_{lk}^H \tilde{\mathbf{v}}_{lk}|^2, \end{aligned} \quad (42)$$

where the first asymptotic equality comes from the fact that the cross term is asymptotically ignorable and the last equality comes from inserting (12).

Define \mathcal{L}_k as the set of BSs sufficiently near user k , given by

$$\mathcal{L}_k = \{X_l \in \mathcal{X}_k \mid P_k^{\text{ul}} \beta_{lk} = \Omega(1)\}. \quad (43)$$

Since $P_k^{\text{ul}} \beta_{lk} = \Omega(1) \Leftrightarrow |X_l - U_k| = O(N^{\rho^{\text{ul}}/\alpha})$, $|\mathcal{X}_k| = O(N^{\eta_{\text{bs}} + \frac{2}{\alpha} \rho^{\text{ul}}})$, which means that \mathcal{L}_k is a non-empty set if and only if $\rho^{\text{ul}} \geq -\frac{\alpha}{2} \eta_{\text{bs}}$ and $\mathcal{L}_k = \mathcal{X}_k$ if $\rho^{\text{ul}} \geq 0$. Then, we have

$$\begin{aligned} |\psi_{kk}^{\text{if}}|^2 &= \sum_{X_l \in \mathcal{L}_k} \beta_{lk}^2 |\mathbf{h}_{lk}^H \mathbf{h}_{lk}|^2 + \frac{1}{P_k^{\text{ul}}} \sum_{X_l \in \mathcal{L}_k} \beta_{lk} |\mathbf{h}_{lk}^H \tilde{\mathbf{v}}_{lk}|^2 \\ &\quad + (P_k^{\text{ul}})^2 \sum_{X_l \in \mathcal{X}_k \setminus \mathcal{L}_k} \beta_{lk}^4 |\mathbf{h}_{lk}^H \mathbf{h}_{lk}|^2 \\ &\quad + P_k^{\text{ul}} \sum_{X_l \in \mathcal{X}_k \setminus \mathcal{L}_k} \beta_{lk}^3 |\mathbf{h}_{lk}^H \tilde{\mathbf{v}}_{lk}|^2. \end{aligned} \quad (44)$$

Case EH ($\rho^{\text{ul}} \geq 0$): in this case, $\mathcal{L}_k = \mathcal{X}_k$ and $\mathcal{X}_k \setminus \mathcal{L}_k = \emptyset$. Then, we have

$$\begin{aligned} |\psi_{kk}^{\text{if}}|^2 &= \sum_{X_l \in \mathcal{X}_k} \beta_{lk}^2 |\mathbf{h}_{lk}^H \mathbf{h}_{lk}|^2 + \frac{1}{P_k^{\text{ul}}} \sum_{X_l \in \mathcal{X}_k} \beta_{lk} |\mathbf{h}_{lk}^H \tilde{\mathbf{v}}_{lk}|^2 \\ &\asymp N^{\alpha \eta_{\text{bs}} + 2\eta_{\text{ant}}} + N^{-\rho^{\text{ul}} + \frac{\alpha}{2} \eta_{\text{bs}} + \eta_{\text{ant}}} \\ &\asymp N^{\alpha \eta_{\text{bs}} + 2\eta_{\text{ant}}}, \end{aligned} \quad (45)$$

where the second asymptotic equality comes from Lemma 2 and the last asymptotic equality comes from the fact that the first term is always dominant for $\rho^{\text{ul}} \geq 0$.

Cases M and L ($\rho^{\text{ul}} < -\frac{\alpha}{2} \eta_{\text{bs}}$): in this case, $\mathcal{L}_k = \emptyset$ and $\mathcal{X}_k \setminus \mathcal{L}_k = \mathcal{X}_k$, we have

$$\begin{aligned} |\psi_{kk}^{\text{if}}|^2 &= (P_k^{\text{ul}})^2 \sum_{X_l \in \mathcal{X}_k} \beta_{lk}^4 |\mathbf{h}_{lk}^H \mathbf{h}_{lk}|^2 + P_k^{\text{ul}} \sum_{X_l \in \mathcal{X}_k} \beta_{lk}^3 |\mathbf{h}_{lk}^H \tilde{\mathbf{v}}_{lk}|^2 \\ &\asymp N^{2\rho^{\text{ul}} + 2\alpha \eta_{\text{bs}} + 2\eta_{\text{ant}}} + N^{\rho^{\text{ul}} + \frac{3\alpha}{2} \eta_{\text{bs}} + \eta_{\text{ant}}} \\ &\asymp \begin{cases} N^{2\rho^{\text{ul}} + 2\alpha \eta_{\text{bs}} + 2\eta_{\text{ant}}}, & \text{if M,} \\ N^{\rho^{\text{ul}} + \frac{3\alpha}{2} \eta_{\text{bs}} + \eta_{\text{ant}}}, & \text{if L,} \end{cases} \end{aligned} \quad (46)$$

where the second asymptotic equality comes from Lemma 2 and the third asymptotic equality comes from the fact that the first term is only dominant if $\rho^{\text{ul}} > -\frac{\alpha}{2} \eta_{\text{bs}} - \eta_{\text{ant}}$.

Case H ($-\frac{\alpha}{2}\eta_{\text{bs}} \leq \rho^{\text{ul}} < 0$): The scaling is simply derived by combining the results in (45) and (46) so that it is given by

$$|\psi_{kk}^{\text{if}}|^2 \asymp N^{\alpha\eta_{\text{bs}}+2\eta_{\text{ant}}}. \quad (47)$$

Define an auxiliary variable ν such that $Q_k^{\text{if}} = \Theta(N^\nu)$ for all k . The final step is to find the scaling exponent of the DL transmit power for user k , which is given by

$$\begin{aligned} P_k^{\text{dl}} &= Q_k^{\text{if}} \sum_{X_l \in \mathcal{X}_k} \beta_{lk} \|\widehat{\mathbf{h}}_{lk}\|^2 \\ &\asymp M Q_k^{\text{if}} \sum_{X_l \in \mathcal{X}_k} \frac{P_k^{\text{ul}} \beta_{lk}^2}{P_k^{\text{ul}} \beta_{lk} + 1} \\ &\asymp \begin{cases} N^{\nu+\frac{\alpha}{2}\eta_{\text{bs}}+\eta_{\text{ant}}}, & \text{if EH or H,} \\ N^{\nu+\rho^{\text{ul}}+\alpha\eta_{\text{bs}}+\eta_{\text{ant}}}, & \text{if M or L.} \end{cases} \end{aligned} \quad (48)$$

By using $P_k^{\text{dl}} = \Theta(N^{\rho^{\text{dl}}})$ for all k , and the results (45), (46), and (47), we obtain (16).

APPENDIX C PROOF OF THEOREM 2

Inserting $\mathbf{F}^{\text{mrt}} = \widehat{\mathbf{G}}^H$ into (7), $\psi_{kj}^{\text{mrt}} = \sum_{X_l \in \mathcal{X}_j} \sqrt{\beta_{lk}\beta_{lj}} \mathbf{h}_{lk}^H \widehat{\mathbf{h}}_{lj}$,

and the square of which is given by

$$\begin{aligned} |\psi_{kj}^{\text{mrt}}|^2 &\asymp \sum_{X_l \in \mathcal{X}_j} \beta_{lk}\beta_{lj} \left| \mathbf{h}_{lk}^H \widehat{\mathbf{h}}_{lj} \right|^2 \\ &= \sum_{X_l \in \mathcal{X}_j} \left(\frac{P_j^{\text{ul}}}{P_j^{\text{ul}} \beta_{lj} + 1} \right)^2 \beta_{lk}\beta_{lj}^3 \left| \mathbf{h}_{lk}^H \mathbf{h}_{lj} \right|^2 \\ &\quad + \sum_{X_l \in \mathcal{X}_j} \left(\frac{\sqrt{P_j^{\text{ul}}}}{P_j^{\text{ul}} \beta_{lj} + 1} \right)^2 \beta_{lk}\beta_{lj}^2 \left| \mathbf{h}_{lk}^H \widetilde{\mathbf{v}}_{lj} \right|^2, \end{aligned} \quad (49)$$

where the first asymptotic equality comes from the fact that the cross term is asymptotically ignorable and the third equality is obtained by inserting (12). Note that $\psi_{kk}^{\text{mrt}} = \psi_{kk}^{\text{if}}$ so that the SNR scaling exponent of MRT operation is the same as that of IF operation.

The interference power, denoted as I_k , can be written as

$$\begin{aligned} I_k &= \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\text{mrt}} |\psi_{kj}^{\text{mrt}}|^2 \\ &= \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\text{mrt}} \sum_{X_l \in \mathcal{X}_j} \beta_{lk}\beta_{lj} \left| \mathbf{h}_{lk}^H \mathbf{h}_{lj} \right|^2 \\ &\quad + \frac{1}{P_k^{\text{ul}}} \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\text{mrt}} \sum_{X_l \in \mathcal{L}_j} \beta_{lk} \left| \mathbf{h}_{lk}^H \widetilde{\mathbf{v}}_{lj} \right|^2 \\ &\quad + (P_k^{\text{ul}})^2 \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\text{mrt}} \sum_{X_l \in \mathcal{X}_j \setminus \mathcal{L}_j} \beta_{lk}\beta_{lj}^3 \left| \mathbf{h}_{lk}^H \mathbf{h}_{lj} \right|^2 \\ &\quad + P_k^{\text{ul}} \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\text{mrt}} \sum_{X_l \in \mathcal{X}_j \setminus \mathcal{L}_j} \beta_{lk}\beta_{lj}^2 \left| \mathbf{h}_{lk}^H \widetilde{\mathbf{v}}_{lj} \right|^2, \end{aligned} \quad (50)$$

where \mathcal{L}_j is defined in (43).

Define an auxiliary variable ν such that $Q_k^{\text{mrt}} = \Theta(N^\nu)$ for all k and consider the four regimes EH, H, M, and L.

Case EH ($\rho^{\text{ul}} \geq 0$): in this case, $\mathcal{L}_j = \mathcal{X}_j$ and $\mathcal{X}_j \setminus \mathcal{L}_j = \emptyset$ for all j . Then, we have

$$\begin{aligned} I_k &= \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\text{mrt}} \sum_{X_l \in \mathcal{X}_j} \beta_{lk}\beta_{lj} \left| \mathbf{h}_{lk}^H \mathbf{h}_{lj} \right|^2 \\ &\quad + \frac{1}{P_k^{\text{ul}}} \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\text{mrt}} \sum_{X_l \in \mathcal{X}_j} \beta_{lk} \left| \mathbf{h}_{lk}^H \widetilde{\mathbf{v}}_{lj} \right|^2 \\ &\asymp N^{\nu+\frac{\alpha}{2}(\eta_{\text{bs}}+\eta_{\text{user}})+(1-\frac{\alpha}{2})(\eta_{\text{user}}-\eta_{\text{bs}})^++\eta_{\text{ant}}} \\ &\quad + N^{\nu-\rho^{\text{ul}}+\frac{\alpha}{2}\eta_{\text{bs}}+\eta_{\text{user}}+\eta_{\text{ant}}} \\ &\asymp N^{\nu+\frac{\alpha}{2}(\eta_{\text{bs}}+\eta_{\text{user}})+(1-\frac{\alpha}{2})(\eta_{\text{user}}-\eta_{\text{bs}})^++\eta_{\text{ant}}}, \end{aligned} \quad (51)$$

where the second asymptotic equality comes from Lemma 2 and the third asymptotic equality comes from the fact that the first term is always dominant due to $(\frac{\alpha}{2}-1)((\eta_{\text{user}}-\eta_{\text{bs}})^+-\eta_{\text{user}}) \leq 0$.

Cases M and L ($\rho^{\text{ul}} < -\frac{\alpha}{2}\eta_{\text{bs}}$): In this case, $\mathcal{L}_j = \emptyset$ and $\mathcal{X}_j \setminus \mathcal{L}_j = \mathcal{X}_j$. Then, we have

$$\begin{aligned} I_k &= (P_k^{\text{ul}})^2 \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\text{mrt}} \sum_{X_l \in \mathcal{X}_j} \beta_{lk}\beta_{lj}^3 \left| \mathbf{h}_{lk}^H \mathbf{h}_{lj} \right|^2 \\ &\quad + P_k^{\text{ul}} \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} Q_j^{\text{mrt}} \sum_{X_l \in \mathcal{X}_j} \beta_{lk}\beta_{lj}^2 \left| \mathbf{h}_{lk}^H \widetilde{\mathbf{v}}_{lj} \right|^2 \\ &\asymp N^{\nu+2\rho^{\text{ul}}+\frac{3\alpha}{2}\eta_{\text{bs}}+\frac{\alpha}{2}\eta_{\text{user}}+(1-\frac{\alpha}{2})(\eta_{\text{user}}-\eta_{\text{bs}})^++\eta_{\text{ant}}} \\ &\quad + N^{\nu+\rho^{\text{ul}}+\alpha\eta_{\text{bs}}+\frac{\alpha}{2}\eta_{\text{user}}+(1-\frac{\alpha}{2})(\eta_{\text{user}}-\eta_{\text{bs}})^++\eta_{\text{ant}}} \\ &\asymp N^{\nu+\rho^{\text{ul}}+\alpha\eta_{\text{bs}}+\frac{\alpha}{2}\eta_{\text{user}}+(1-\frac{\alpha}{2})(\eta_{\text{user}}-\eta_{\text{bs}})^++\eta_{\text{ant}}}, \end{aligned} \quad (52)$$

where the second equality comes from Lemma 2 and the third asymptotic equality comes from the fact that the first term is always dominant due to $\rho^{\text{ul}} < -\frac{\alpha}{2}\eta_{\text{bs}}$.

Case H ($-\frac{\alpha}{2}\eta_{\text{bs}} \leq \rho^{\text{ul}} < 0$): The scaling exponent can be simply derived as

$$I_k \asymp N^{\nu+\frac{\alpha}{2}(\eta_{\text{bs}}+\eta_{\text{user}})+(1-\frac{\alpha}{2})(\eta_{\text{user}}-\eta_{\text{bs}})^++\eta_{\text{ant}}}, \quad (53)$$

because the result in (52) at $\rho^{\text{ul}} = -\frac{\alpha}{2}\eta_{\text{bs}}$ is the same to the result in (51). Thus, by combining (51)-(53), the scaling exponent of the interference power is given by

$$I_k \asymp \begin{cases} N^{\nu+\frac{\alpha}{2}(\eta_{\text{bs}}+\eta_{\text{user}})+(1-\frac{\alpha}{2})(\eta_{\text{user}}-\eta_{\text{bs}})^++\eta_{\text{ant}}}, & \text{if EH or H,} \\ N^{\nu+\rho^{\text{ul}}+\alpha\eta_{\text{bs}}+\frac{\alpha}{2}\eta_{\text{user}}+(1-\frac{\alpha}{2})(\eta_{\text{user}}-\eta_{\text{bs}})^++\eta_{\text{ant}}}, & \text{if M or L,} \end{cases} \quad (54)$$

and combining it with (16) and (48) results in (21), which completes the proof.

APPENDIX D PROOF OF THEOREM 3

Inserting $\mathbf{F}^{\text{zf}} = \widehat{\mathbf{G}}^H (\widehat{\mathbf{G}} \widehat{\mathbf{G}}^H)^{-1}$ into (7), $\psi_{kj}^{\text{zf}} = \delta(k-j) + [\widetilde{\mathbf{G}}]_{k,:} \mathbf{F}^{\text{zf}} \mathbf{e}_j$, where \mathbf{e}_j denotes the j th column of the identity

matrix \mathbf{I}_K , and the square of which is given by

$$\begin{aligned} |\psi_{kj}^{\text{zf}}|^2 &= \delta(k-j) + 2\text{Re} \left\{ [\tilde{\mathbf{G}}]_{k,:}^H \mathbf{F}^{\text{zf}} \mathbf{e}_j \right\} \\ &\quad + \text{Tr} \left((\mathbf{F}^{\text{zf}})^H \mathbf{R}_k \mathbf{F}^{\text{zf}} \mathbf{e}_j \mathbf{e}_j^H \right) \\ &\asymp \delta(k-j) + \text{Tr} \left((\mathbf{F}^{\text{zf}})^H \mathbf{R}_k \mathbf{F}^{\text{zf}} \mathbf{e}_j \mathbf{e}_j^H \right), \end{aligned} \quad (55)$$

where $\mathbf{R}_k = [\tilde{\mathbf{G}}]_{k,:}^H [\tilde{\mathbf{G}}]_{k,:}$ and the last asymptotic equality comes from the fact that $(a+b)^2 \asymp a^2 + b^2$. Also, the DL transmit power consumed for user k is given by

$$P_k^{\text{dl}} = Q_k^{\text{zf}} \text{Tr} \left((\mathbf{F}^{\text{zf}})^H \mathbf{F}^{\text{zf}} \mathbf{e}_k \mathbf{e}_k^H \right). \quad (56)$$

Note that we have

$$\begin{aligned} [\widehat{\mathbf{G}}\widehat{\mathbf{G}}^H]_{i,j} &= \sum_{X_l \in \mathcal{X}} \sqrt{\beta_{li}\beta_{lj}} \widehat{\mathbf{h}}_l^H \widehat{\mathbf{h}}_l \\ &= \begin{cases} M \sum_{X_l \in \mathcal{X}} \frac{P_j^{\text{ul}} \beta_{lj}^2}{P_j^{\text{ul}} \beta_{lj} + 1}, & \text{if } i = j, \\ \sqrt{M} \sum_{X_l \in \mathcal{X}} \sqrt{\frac{P_i^{\text{ul}} \beta_{li}^2}{P_i^{\text{ul}} \beta_{li} + 1} \frac{P_j^{\text{ul}} \beta_{lj}^2}{P_j^{\text{ul}} \beta_{lj} + 1}}, & \text{if } i \neq j, \end{cases} \end{aligned} \quad (57)$$

and consider the four regimes.

Case EH ($\rho^{\text{ul}} \geq 0$): In this case, $\mathcal{L}_k = \mathcal{X}_k$ so that $\mathbf{R}_k \asymp \frac{1}{P_k^{\text{ul}}} \mathbf{I}_N$ and

$$[\widehat{\mathbf{G}}\widehat{\mathbf{G}}^H]_{i,j} \asymp \begin{cases} M \sum_{X_l \in \mathcal{X}} \beta_{lj}, & \text{if } i = j, \\ \sqrt{M} \sum_{X_l \in \mathcal{X}} \sqrt{\beta_{li}\beta_{lj}}, & \text{if } i \neq j, \end{cases} \quad (58)$$

Inserting (58) into (55), we have

$$\begin{aligned} |\psi_{kk}^{\text{zf}}|^2 &\asymp 1 + \frac{1}{MP_k^{\text{ul}}} \left(\sum_{X_l \in \mathcal{X}_k} \beta_{lk} \right)^{-1} \\ &\asymp 1 + N^{-\rho^{\text{ul}} - \frac{\alpha}{2} \eta_{\text{bs}} - \eta_{\text{ant}}} \\ &\asymp 1 \end{aligned} \quad (59)$$

and

$$\begin{aligned} \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} |\psi_{kj}^{\text{zf}}|^2 &\asymp \frac{1}{MP_k^{\text{ul}}} \\ &\quad \cdot \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} \left(\sum_{X_l \in \mathcal{X}_k} \beta_{lk} \right) \left(\sum_{X_l \in \mathcal{X}_j} \beta_{lj} \right)^{-2} \\ &\asymp N^{-\rho^{\text{ul}} - \frac{\alpha}{2} \eta_{\text{bs}} - \eta_{\text{ant}} + \eta_{\text{user}}}, \end{aligned} \quad (60)$$

where the last asymptotic equality comes from Lemma 2 and $\rho^{\text{ul}} \geq 0$. Defining an auxiliary variable ν such that $Q_k = \Theta(N^\nu)$, the DL transmit power can be written as

$$\begin{aligned} P_k^{\text{dl}} &= \frac{Q_k^{\text{zf}}}{M} \left(\sum_{X_l \in \mathcal{X}_k} \beta_{lk} \right)^{-1} \\ &\asymp N^{\nu - \frac{\alpha}{2} \eta_{\text{bs}} - \eta_{\text{ant}}}. \end{aligned} \quad (61)$$

For a reasonable DL transmit power $P_k^{\text{dl}} = \Theta(N^{\rho^{\text{dl}}})$, $\nu = \rho^{\text{dl}} + \frac{\alpha}{2} \eta_{\text{bs}} + \eta_{\text{ant}}$ and thus

$$\begin{aligned} \text{snr}^{\text{zf}} &= \rho^{\text{dl}} + \frac{\alpha}{2} \eta_{\text{bs}} + \eta_{\text{ant}}, \\ \text{sir}^{\text{zf}} &= \rho^{\text{ul}} + \frac{\alpha}{2} \eta_{\text{bs}} + \eta_{\text{ant}} - \eta_{\text{user}}. \end{aligned} \quad (62)$$

Case H ($-\frac{\alpha}{2} \eta_{\text{bs}} \leq \rho^{\text{ul}} < 0$): Define $\mathcal{U}_k^{(d)} = \{U_j \in \mathcal{U} \mid |U_k - U_j| = \Theta(N^d)\}$. Since $\mathbf{R}_k = \text{diag}\{\theta_{1k} \mathbf{I}_M, \dots, \theta_{Lk} \mathbf{I}_M\}$, where $\theta_{lk} \asymp 1/P_k^{\text{ul}}$, if $X_l \in \mathcal{X}_k$, or $\theta_{lk} \asymp \beta_{lk}$, if $X_l \in \mathcal{X} \setminus \mathcal{X}_k$, and $[\widehat{\mathbf{G}}\widehat{\mathbf{G}}^H]_{i,j}$ is given as in (63). Then, $|\psi_{kj}^{\text{zf}}|^2$ is given as in (64), where the last asymptotic equality comes from the fact that

$$\begin{aligned} \sum_{X_l \in \mathcal{L}_k} \beta_{lk} + P_k^{\text{ul}} \sum_{X_l \in \mathcal{X}_k \setminus \mathcal{L}_k} \beta_{lk}^2 &\asymp N^{\frac{\alpha}{2} \eta_{\text{bs}}} + N^{\eta_{\text{bs}} + (\frac{\alpha}{2} - 1) \rho^{\text{ul}}} \\ &\asymp N^{\frac{\alpha}{2} \eta_{\text{bs}}}. \end{aligned} \quad (65)$$

Here, t_1, t_2, t_3 , and t_4 can be further derived from Lemma 2 as follows. Suppose that $\eta_{\text{bs}} \geq \eta_{\text{user}}$ and $-\frac{\alpha}{2} \eta_{\text{bs}} \leq \rho^{\text{ul}} < -\frac{\alpha}{2} \eta_{\text{user}}$. In this case, $\mathcal{L}_k \cap \mathcal{L}_j = \emptyset$ for all $U_j \in \mathcal{U}$ in probability so that, by using Lemma 2, we obtain

$$\begin{aligned} t_1 &= 0, \\ t_2 &\asymp N^{\eta_{\text{bs}} + \alpha \eta_{\text{user}} + \frac{2}{\alpha} \rho^{\text{ul}}}, \\ t_3 &\asymp N^{\frac{\alpha}{2} \eta_{\text{bs}} + \frac{\alpha}{2} \eta_{\text{user}}}, \\ t_4 &\asymp N^{\eta_{\text{bs}} + \frac{\alpha}{2} \eta_{\text{user}} + (\frac{2}{\alpha} - 1) \rho^{\text{ul}}}. \end{aligned}$$

Since $\frac{\alpha}{2} \eta_{\text{bs}} + \frac{\alpha}{2} \eta_{\text{user}} - (\eta_{\text{bs}} + \alpha \eta_{\text{user}} + \frac{2}{\alpha} \rho^{\text{ul}}) = (\frac{\alpha}{2} - 1) (\eta_{\text{bs}} - \eta_{\text{user}}) - \frac{2}{\alpha} (\rho^{\text{ul}} + \frac{\alpha}{2} \eta_{\text{user}}) \geq 0$ and $\frac{\alpha}{2} \eta_{\text{bs}} + \frac{\alpha}{2} \eta_{\text{user}} - (\eta_{\text{bs}} + \frac{\alpha}{2} \eta_{\text{user}} + (\frac{2}{\alpha} - 1) \rho^{\text{ul}}) = (1 - \frac{\alpha}{2}) (\frac{\alpha}{2} \eta_{\text{bs}} + \rho^{\text{ul}}) \geq 0$, t_3 becomes always dominant and thus

$$\sum_{U_j \in \mathcal{U} \setminus \{U_k\}} |\psi_{kj}^{\text{zf}}|^2 \asymp N^{-\frac{\alpha}{2} \eta_{\text{bs}} + \frac{\alpha}{2} \eta_{\text{user}} - \eta_{\text{ant}}}. \quad (66)$$

Now, suppose that $\eta_{\text{bs}} \geq \eta_{\text{user}}$ and $\rho^{\text{ul}} \geq -\frac{\alpha}{2} \eta_{\text{user}}$. In this case, $\mathcal{L}_j = \mathcal{X}_j$ if $d \leq \rho^{\text{ul}}/\alpha$ and $\mathcal{X}_j \cap \mathcal{L}_j = \emptyset$ if $d > \rho^{\text{ul}}/\alpha$ for all $U_j \in \mathcal{U}_k^{(d)}$ in probability so that, by using Lemma 2, we obtain

$$\begin{aligned} t_1 &\asymp N^{\frac{\alpha}{2} \eta_{\text{bs}} + \eta_{\text{user}} + (\frac{2}{\alpha} - 1) \rho^{\text{ul}}}, \\ t_2 &\asymp N^{\eta_{\text{bs}} + \eta_{\text{user}} + 2(\frac{2}{\alpha} - 1) \rho^{\text{ul}}}, \\ t_3 &\asymp N^{\frac{\alpha}{2} \eta_{\text{bs}} + \eta_{\text{user}} + (\frac{2}{\alpha} - 1) \rho^{\text{ul}}}, \\ t_4 &\asymp N^{\eta_{\text{bs}} + \eta_{\text{user}} + 2(\frac{2}{\alpha} - 1) \rho^{\text{ul}}}. \end{aligned}$$

Since $\frac{\alpha}{2} \eta_{\text{bs}} + \eta_{\text{user}} + (\frac{2}{\alpha} - 1) \rho^{\text{ul}} - (\eta_{\text{bs}} + \eta_{\text{user}} + 2(\frac{2}{\alpha} - 1) \rho^{\text{ul}}) = (1 - \frac{2}{\alpha}) (\frac{\alpha}{2} \eta_{\text{bs}} + \rho^{\text{ul}}) \geq 0$, t_1 or t_3 becomes dominant and thus

$$\sum_{U_j \in \mathcal{U} \setminus \{U_k\}} |\psi_{kj}^{\text{zf}}|^2 \asymp N^{(\frac{2}{\alpha} - 1) \rho^{\text{ul}} - \frac{\alpha}{2} \eta_{\text{bs}} + \eta_{\text{user}} - \eta_{\text{ant}}}. \quad (67)$$

$$[\widehat{\mathbf{G}}\widehat{\mathbf{G}}^H]_{i,j} \asymp \begin{cases} M \left(\sum_{X_l \in \mathcal{L}_j} \beta_{lj} + P_j^{\text{ul}} \sum_{X_l \in \mathcal{X}_j \setminus \mathcal{L}_j} \beta_{lj}^2 \right), & \text{if } i = j, \\ \sqrt{M} \left(\begin{array}{l} \sum_{X_l \in \mathcal{L}_i \cap \mathcal{L}_j} \sqrt{\beta_{li}\beta_{lj}} + \sqrt{P_j^{\text{ul}}} \sum_{X_l \in \mathcal{L}_i \setminus \mathcal{L}_j} \sqrt{\beta_{li}\beta_{lj}^2} \\ + \sqrt{P_i^{\text{ul}}} \sum_{X_l \in \mathcal{L}_j \setminus \mathcal{L}_i} \sqrt{\beta_{li}^2\beta_{lj}} + \sqrt{P_i^{\text{ul}}P_j^{\text{ul}}} \sum_{X_l \in \mathcal{X}_i \cap \mathcal{X}_j \setminus \mathcal{L}_i \cup \mathcal{L}_j} \sqrt{\beta_{li}^2\beta_{lj}^2} \end{array} \right), & \text{if } i \neq j. \end{cases} \quad (63)$$

$$\begin{aligned} \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} |\psi_{kj}^{\text{zf}}|^2 &\asymp \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} \frac{\frac{1}{P^{\text{ul}}} \sum_{X_l \in \mathcal{X}_k \cap \mathcal{X}_j} \beta_{lj} + \sum_{X_l \in \mathcal{X}_k \setminus \mathcal{X}_j} \beta_{lj}^2 + \sum_{X_l \in \mathcal{X}_j \setminus \mathcal{X}_k} \beta_{lk}\beta_{lj} + P^{\text{ul}} \sum_{X_l \in \mathcal{X} \setminus (\mathcal{X}_k \cup \mathcal{X}_j)} \beta_{lk}\beta_{lj}^2}{M \left(\sum_{l \in \mathcal{X}_j} \beta_{lj} + P^{\text{ul}} \sum_{X_l \in \mathcal{X}_j} \beta_{lj}^2 \right)^2} \\ &\asymp \frac{1}{N^{\alpha\eta_{\text{bs}} + \eta_{\text{ant}}}} \left(\underbrace{\frac{1}{P^{\text{ul}}} \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} \sum_{X_l \in \mathcal{X}_k \cap \mathcal{X}_j} \beta_{lj}}_{t_1} + \underbrace{\sum_{U_j \in \mathcal{U} \setminus \{U_k\}} \sum_{X_l \in \mathcal{X}_k \setminus \mathcal{X}_j} \beta_{lj}^2}_{t_2} \right. \\ &\quad \left. + \underbrace{\sum_{U_j \in \mathcal{U} \setminus \{U_k\}} \sum_{X_l \in \mathcal{X}_j \setminus \mathcal{X}_k} \beta_{lk}\beta_{lj}}_{t_3} + \underbrace{P^{\text{ul}} \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} \sum_{X_l \in \mathcal{X} \setminus (\mathcal{X}_k \cup \mathcal{X}_j)} \beta_{lk}\beta_{lj}^2}_{t_4} \right), \end{aligned} \quad (64)$$

When $\eta_{\text{bs}} \leq \eta_{\text{user}}$, the similar derivation can be done and we obtain

$$\sum_{U_j \in \mathcal{U} \setminus \{U_k\}} |\psi_{kj}^{\text{zf}}|^2 \asymp N^{\left(\frac{2}{\alpha}-1\right)\rho^{\text{ul}} - \frac{\alpha}{2}\eta_{\text{bs}} + \eta_{\text{user}} - \eta_{\text{ant}}}. \quad (68)$$

Also,

$$\begin{aligned} |\psi_{kk}^{\text{zf}}|^2 &\asymp 1 + \frac{\frac{1}{P^{\text{ul}}} \sum_{X_l \in \mathcal{L}_k} \beta_{lk} + P^{\text{ul}} \sum_{X_l \in \mathcal{X}_k \setminus \mathcal{L}_k} \beta_{lk}^3}{M \left(\sum_{X_l \in \mathcal{L}_k} \beta_{lk} + P_k^{\text{ul}} \sum_{X_l \in \mathcal{X}_k \setminus \mathcal{L}_k} \beta_{lk}^2 \right)^2} \\ &\asymp 1 + N^{-\rho^{\text{ul}} - \frac{\alpha}{2}\eta_{\text{bs}} - \eta_{\text{ant}}} \\ &\asymp 1, \end{aligned} \quad (69)$$

where the second and last asymptotic equalities come from Lemma 2 and the fact $-\frac{\alpha}{2}\eta_{\text{bs}} \leq \rho^{\text{ul}} < 0$, respectively. Since the DL transmit power can be written as

$$P_k^{\text{dl}} = \frac{Q_k^{\text{zf}}}{M} \left(\sum_{X_l \in \mathcal{L}_k} \beta_{lk} + P^{\text{ul}} \sum_{X_l \in \mathcal{X}_k \setminus \mathcal{L}_k} \beta_{lk}^2 \right)^{-1} \quad (70) \quad \text{and}$$

$$\asymp N^{\nu - \frac{\alpha}{2}\eta_{\text{bs}} - \eta_{\text{ant}}}$$

so that $\nu = \rho^{\text{dl}} + \frac{\alpha}{2}\eta_{\text{bs}} + \eta_{\text{ant}}$ for a reasonable DL transmit power,

$$\text{snr}^{\text{zf}} = \rho^{\text{dl}} + \frac{\alpha}{2}\eta_{\text{bs}} + \eta_{\text{ant}}, \quad (71)$$

and sir^{zf} is given as in (72).

Cases M and L ($\rho^{\text{ul}} < -\frac{\alpha}{2}\eta_{\text{bs}}$): In these cases, $\mathbf{R}_k \asymp \text{diag}\{\beta_{1k}\mathbf{I}_M, \beta_{2k}\mathbf{I}_M, \dots, \beta_{Lk}\mathbf{I}_M\}$, and

$$[\widehat{\mathbf{G}}\widehat{\mathbf{G}}^H]_{i,j} \asymp \begin{cases} MP_j^{\text{ul}} \sum_{X_l \in \mathcal{X}_j} \beta_{lj}^2, & \text{if } i = j, \\ \sqrt{MP_i^{\text{ul}}P_j^{\text{ul}}} \sum_{X_l \in \mathcal{X}_i \cap \mathcal{X}_j} \sqrt{\beta_{li}^2\beta_{lj}^2}, & \text{if } i \neq j, \end{cases} \quad (73)$$

Inserting (73) into (55), we have

$$|\psi_{kj}^{\text{zf}}|^2 \asymp \mathbb{1}_{kj} + \frac{1}{MP_j^{\text{ul}}} \left(\sum_{X_l \in \mathcal{X}_j} \beta_{lj}^2 \right)^{-2} \sum_{X_l \in \mathcal{X}_j} \beta_{lj}^2 \beta_{lk}, \quad (74)$$

which is obtained in a similar way as before as

$$\begin{aligned} |\psi_{kk}^{\text{zf}}|^2 &\asymp 1 + N^{-\rho^{\text{ul}} - \frac{\alpha}{2}\eta_{\text{bs}} - \eta_{\text{ant}}} \\ &\asymp \begin{cases} 1, & \text{if M,} \\ N^{-\rho^{\text{ul}} - \frac{\alpha}{2}\eta_{\text{bs}} - \eta_{\text{ant}}}, & \text{if L,} \end{cases} \end{aligned} \quad (75)$$

$$\begin{aligned} &\sum_{U_j \in \mathcal{U} \setminus \{U_k\}} |\psi_{kj}^{\text{zf}}|^2 \\ &\asymp \frac{1}{MP^{\text{ul}}} \sum_{U_j \in \mathcal{U} \setminus \{U_k\}} \left(\sum_{X_l \in \mathcal{X}_j} \beta_{lj}^2 \right)^{-2} \sum_{X_l \in \mathcal{X}_j} \beta_{lj}^2 \beta_{lk} \\ &\asymp N^{-\rho^{\text{ul}} - \left(\frac{\alpha}{2}+1\right)\eta_{\text{bs}} - \eta_{\text{ant}} + \eta_{\text{user}} - \left(\frac{\alpha}{2}-1\right)(\eta_{\text{bs}} - \eta_{\text{user}})^+}. \end{aligned} \quad (76)$$

$$\text{sir}^{\text{zf}} = \begin{cases} (1 - \frac{2}{\alpha}) (\rho^{\text{ul}} + \frac{\alpha}{2} \eta_{\text{user}})^+ + \frac{\alpha}{2} (\eta_{\text{bs}} - \eta_{\text{user}}) + \eta_{\text{ant}}, & \text{if } \eta_{\text{bs}} \geq \eta_{\text{user}}, \\ (1 - \frac{2}{\alpha}) \rho^{\text{ul}} + \frac{\alpha}{2} \eta_{\text{bs}} - \eta_{\text{user}} + \eta_{\text{ant}}, & \text{if } \eta_{\text{bs}} < \eta_{\text{user}}. \end{cases} \quad (72)$$

Since the DL transmit power can be written as

$$P_k^{\text{dl}} = \frac{Q_k^{\text{zf}}}{M P_k^{\text{ul}}} \left(\sum_{X_i \in \mathcal{X}_k} \beta_{ik}^2 \right)^{-1} \asymp N^{\nu - \rho^{\text{ul}} - \alpha \eta_{\text{bs}} - \eta_{\text{ant}}} \quad (77)$$

so that $\nu = \rho^{\text{dl}} + \rho^{\text{ul}} + \alpha \eta_{\text{bs}} + \eta_{\text{ant}}$ for a reasonable DL transmit power, we can obtain

$$\text{snr}^{\text{zf}} = \begin{cases} \rho^{\text{dl}} + \rho^{\text{ul}} + \alpha \eta_{\text{bs}} + \eta_{\text{ant}}, & \text{if M,} \\ \rho^{\text{dl}} + \frac{\alpha}{2} \eta_{\text{bs}}, & \text{if L,} \end{cases} \quad (78)$$

and

$$\text{sir}^{\text{zf}} = \begin{cases} \rho^{\text{ul}} + \frac{\alpha}{2} \eta_{\text{bs}} + \eta_{\text{ant}} \\ + (\eta_{\text{bs}} - \eta_{\text{user}}) + (\frac{\alpha}{2} - 1) (\eta_{\text{bs}} - \eta_{\text{user}})^+, & \text{if M,} \\ (\eta_{\text{bs}} - \eta_{\text{user}}) + (\frac{\alpha}{2} - 1) (\eta_{\text{bs}} - \eta_{\text{user}})^+, & \text{if L,} \end{cases} \quad (79)$$

which concludes the proof. \square

REFERENCES

- [1] J. G. Andrews et al., "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1065–1082, June 2014.
- [2] N. Bhushan et al., "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.
- [3] F. Rusek et al., "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–46, Jan. 2013.
- [4] A. Gotsis, S. Stefanatos, and A. Alexiou, "Ultra dense networks: The new wireless frontier for enabling 5G access," *IEEE Veh. Technol. Mag.*, vol. 11, no. 2, pp. 71–78, June 2016.
- [5] 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.
- [6] I. Hwang, B. Song, and S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, pp. 20–27, June 2013.
- [7] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, pp. 37–43, Mar. 2011.
- [8] H. Huang, M. Trivellato, A. Hottinen, M. Sha, P. Smith, and R. Valenzuela, "Increasing DL cellular throughput with limited network MIMO coordination," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 2983–2989, June 2009.
- [9] R. Irmer et al., "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.
- [10] A. Checko et al., "Cloud RAN for mobile networks – A technology overview," *IEEE Commun. Surv. Tut.*, vol. 17, no. 1, 1 Quarter 2015, pp. 405–426.
- [11] P. Marsch, M. Grieger, and G. Fettweis, "Large scale field trial results on different UL coordinated multi-point (CoMP) concepts in an urban environment," in *Proc. IEEE WCNC*, pp. 1858–1863.
- [12] D. A. Basnayaka, P. J. Smith, and P. A. Martin, "Performance analysis of dual-user macrodiversity MIMO systems with linear receivers in flat Rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4394–4404, Dec. 2012.
- [13] D. A. Basnayaka, P. J. Smith, and P. A. Martin, "Performance analysis of macrodiversity MIMO systems with MMSE and ZF receivers in flat Rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, May 2013.
- [14] D. B. Cheikh, J.-M. Kelif, M. Coupechoux, and P. Godlewski, "Analytical joint processing multi-point cooperation performance in Rayleigh fading," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, Aug. 2012.
- [15] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.
- [16] J. Wang and L. Dai, "Asymptotic rate analysis of DL multi-user systems with co-located and distributed antennas," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3046–3058, June 2015.
- [17] R. Tanboug, S. Singh, J. G. Andrew, and F. K. Jondral, "A tractable model for noncoherent joint-transmission base station cooperation," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4959–4973, Sept. 2014.
- [18] G. Nigam, P. Minero, and M. Haenggi, "Coordinated multipoint joint transmission in heterogeneous networks," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4134–4146, Nov. 2014.
- [19] A. Giovanidis and F. Baccelli, "A stochastic geometry framework for analyzing pairwise-cooperative cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 794–808, Feb. 2015.
- [20] K. Huang and J. G. Andrews, "An analytical framework for multicell cooperation via stochastic geometry and large deviations," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2501–2516, Apr. 2013.
- [21] N. Lee, F. Baccelli, and R.W. Heath, "Spectral efficiency scaling laws in dense random wireless networks with multiple receive antennas," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1344–1359, Mar. 2016.
- [22] H. S. Dhillon and J. G. Andrews, "Downlink rate distribution in heterogeneous cellular networks under generalized cell selection," *IEEE Wireless Commun. Lett.*, vol. 3, no. 1, pp. 42–45, Feb. 2014.
- [23] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, pp. 1458–1461, Oct. 1999.
- [24] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for DL spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
- [25] C. Shepard et al., "Argos: Practical many-antenna base stations," in *Proc. MobiCom*, 2012, pp. 53–64.
- [26] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1993.
- [27] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.
- [28] P. Billingsley, *Convergence of Probability Measures*, 1968, Wiley.



Kyung Jun Choi was born in Busan, Korea, on 7 February, 1987. He received the B.S. and Ph.D. 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.

He 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 hoc networks.