EAI Endorsed Transactions
on Industrial Networks and Intelligent Systems
Research Article EAI.EU

Uplink Performance of Cell-Free Massive MIMO with Access Point Selections

Toan X. Doan[1,*], Long D. Nguyen[2]

[1]Thu Dau Mot University, Binh Duong, Vietnam
[2]Queen's University Belfast

Abstract

Cell-free massive multiple-input multiple-output (MIMO), in which a massive number of access points (APs) distributed over a large area serve a smaller number of users in the same time and frequency resources, inherits advantages from conventional massive MIMO (i.e. favourable propagation and channel hardening), and distributed system (i.e. macro diversity gain). As a result, cell-free massive MIMO can provide a great spectral efficiency, high capacity and offer uniformly great service for all users. To contribute to this great concept, an uplink and downlink performance of cell-free massive MIMO are investigated in this work. Novel access point selection and signal detection schemes are proposed to reduce the requirements of backhaul links connecting the APs and the central processing unit, and to improve the system performance in terms of the achievable rate. Note that most of signal detection schemes for cell-free massive MIMO in the literature rely on the channel hardening property, with results in less accuracy for small and moderate number of APs. Firstly, closed-form expressions for the achievable rate of the downlink and uplink are derived. Then, performance comparisons between the proposed signal detection scheme and the conventional scheme are exploited. The result shows that the proposed scheme (with the novel AP selection and signal detection) outperforms the conventional scheme in terms of the achievable rate and the amount of data load exchanging over the backhaul links.

Keywords: class file, LaTeX 2_ε, *EAI Endorsed Transactions*

Received on 13 November 2018; accepted on 20 November 2018; published on 29 November 2018

doi:10.4108/eai.29-11-2018.155998

1. Introduction

The fast-growing and unlimited requirement of wireless network capacity with the limited resources is the most challenging for networks providers. Improving spectral efficiency is one of the feasible approaches. Particularly, massive multiple-input multiple-output (MIMO), first introduced in [1], can provide a great spectral efficiency by employing massive number of antennas at base station. Interestingly, a new kind of massive MIMO systems has been proposed in [2–4], in which antennas are spread over a large area instead of a compact area as in the conventional massive MIMO system, namely cell-free (CF) massive MIMO.

Similarly to the conventional massive MIMO, CF, which is simultaneously serves much smaller number of users compared to the number of access points (APs) in the same time and spectrum resources, has many advantages. Owing to the macro diversity gain property, CF can offer an uniformly great service for all users [2], which has been an issue for cell-edge users' performance in the conventional cellular systems. Additionally, CF is also a good solution for energy efficiency due to the short distance between users and APs. This aspect was considered in [5, 6], i.e., the relationship between total energy efficiency and network's parameters was studied, and a power allocation algorithm is also proposed. Another important advantage of CF is spectral efficiency, which was investigated in [7, 8]. Particularly, the authors in [3, 8] derived the expressions for achievable uplink and downlink rate, and compared the performance between CF and small cells. The results showed that CF

*Corresponding author. Email: toandx@tdmu.edu.vn

significantly outperformed small cell system in terms of throughput.

In CF, all APs are connected to a central processing unit (CPU) via backhaul links, where most of signal processing is performed. To minimise data exchanging over the backhaul link, conjugate beamforming or maximum-ratio processing is suggested for both uplink and downlink [8]. The reason behind the suggestion is that there is no requirement of sharing the instantaneous channel state information (CSI) in the system. To be more specific, APs can perform channel estimation locally, and the information that CPU needs to control transmit power of APs is large-scale fading which slowly changes over time. Thus, the payload data is the main load travelling on the backhaul link.

Along with the aforementioned research works, CF had been studied under many other aspects. The performance of multi-group multi-cast, where many users need the same information are grouped together, was investigated in CF [9]. In addition to this work, short-term power constraint was applied to beamforming information in the downlink. The work in [10] considered an optimal problem which aims to maximise the minimum uplink rate of user under limited backhaul constraints. Then, the work [11] proposed another solution to improve the max-min SINR which was proposed in [10].

1.1. Discussion on the requirement of backhaul links

By observing the works [5, 7–11], to detect desired information from any user (uplink data), CPU requires the received signal of all APs, which means that a large payload is transmitted through the backhaul links, and the CPU has to do a large number of computations. Moreover, the contributions of different APs on the signal-to-noise and interference ratio (SINR) at the CPU are not identical. Some APs have strong multi-user interference which can lead to a decrement in SINR value. Motivated from that, a new APs selection scheme is proposed in this work to deal with aforementioned issue by reducing the number of APs involved in the signal detection at the CPU.

Note that our proposed AP selection schemes is different from that in [5]. More precisely, two proposed schemes in [5] aim to improve the total energy efficiency, where the first scheme is based on power control coefficients with high computational complexity, and the second one chooses the AP that have strongest large-scale fading values. More importantly, in [5, 10] and [11], after implementation of the APs selection, the signal detection is still based on channel hardening, which turns less accurate as the number of selected APs is small. This problem is presented more details in the sequel.

1.2. Discussion on the case of moderate APs number

Channel hardening, a result of the law of large numbers in statistical theory, is one of the useful properties in massive MIMO, which can help the system reduce complexity, such as it does not need pilot training process and simplicity in signal detection algorithm [12**?** , 13]. Different from massive MIMO, where the large-scale fading between antennas and a user is nearly equal due to the fact that all antennas are located in a compact area, the large-scale fading of the links between a user and APs in CF massive MIMO is different. Hence, the convergence speed of the law of large numbers in this case must be slower. This problem was addressed in [14], which shows that the channel hardening strongly depends on the number of antennas at a AP and the density of APs. However, in a certain aspect, the deployment of multiple antennas at AP is impractical due to a high implementation cost.

As discussed earlier, the channels have a lack of hardening for moderate and small number of APs, leading to the fact that the signal detection algorithm based on the knowledge of channel statistics becomes less accurate. We thus propose another signal detection in this work, which does not rely on the channel hardening property. The proposed scheme can be applied for general case of AP number.

1.3. Contributions

The contributions of this chapter are summarised as

- A novel AP selection scheme is introduced. Result shows that the proposed approach improves the throughput per user, provides a solution with less computational complexity and less requirement in backhaul links.

- A new signal detection based on the knowledge of estimated channel is considered to deal with the issues of lack of channel hardening, as a consequence of deploying the AP selection. The analysis and simulation results show that our scheme can enhance the system performance, particularly for moderate and small number of AP.

- The closed-form expressions for instantaneous and ergodic achievable rate of the uplink, given the location of APs and users and the estimated CSI, are derived the first time, to the best of my knowledge.

2. System model

A CF massive MIMO system is considered, in which M APs simultaneously serve K active users in the same time and frequency resources. Both APs and users are

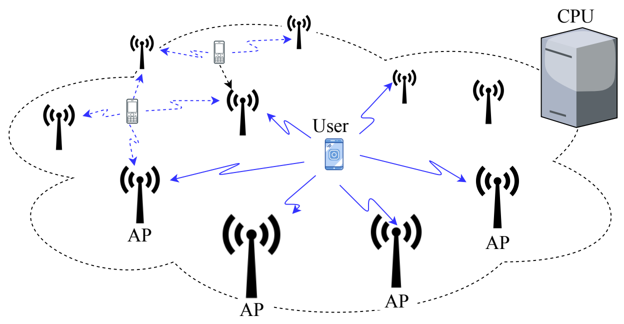

Figure 1. System model

uniformly distributed over a large area. It is assumed that all APs and users are equipped with single antenna, and all APs are connected to a central processing unit (CPU) via a perfect backhaul link.

In the first phase of the time-division duplex protocol, users send pilot sequences to all APs. The APs then estimate the channels to users and share these information with a central processing unit (CPU). In the payload uplink data transfer, users send data to all APs. APs receive the transmitted signal, only selected APs forward their received signal to CPU. The selected APs is chosen by CPU as Algorithm 1, which is discussed in the next section. Next, a signal detection is performed at the CPU based on the estimated information in the first phase.

The detail of all aforementioned processes in a coherence time is described as follows.

The channel coefficient between AP m and TU_x is denoted by

$$g_{mk} = \beta_{mk}^{1/2} h_{mk}, \tag{1}$$

where $h_{mk} \sim \mathcal{CN}(0,1)$ is small scale fading and β_{mk} is large scale fading which is modelled as in [8], following the Hata-COST231 and a three-slope path loss model, as follows:

$$PL = \begin{cases} -L - 34\log_{10}(d_{mk}) & , d_{mk} > d_1 \\ -L - 15\log_{10}(d_1) - 20\log_{10}(d_{mk}) & , d_0 < d_{mk} \le d_1 \\ -L - 15\log_{10}(d_1) - 20\log_{10}(d_0) & , d_{mk} \le d_0, \end{cases} \tag{2}$$

where

$$L \triangleq 46.3 + 33.9\log_{10}(f_c) - 13.82\log_{10}(h_{AP}) - (1.1\log_{10}(f_c) - 0.7)h_u + (1.56\log_{10}(f_c) - 0.8), \tag{3}$$

in which f_c is the carrier frequency, h_{AP} and h_u are the height of AP's antenna and user's antenna, d_{mk} is the distance from user k to AP m, d_0 and d_1 are given distances.

Uplink Training Phase. Similar to the previous works [2, 5, 7–11, 14], the uplink training phase aims to help APs obtain the channel state information. Let $\varphi_{\mathrm{ul},k} \in \mathbb{C}^{\tau_{\mathrm{ul}}^{(\mathrm{p})} \times 1}$ and $\tau_{\mathrm{ul}}^{(\mathrm{p})}$ be the pilot signal and its length (in symbols) that is sent by user k, the received pilot signal at AP m is

$$\mathbf{y}_m^{(\mathrm{p})} = \sqrt{\tau_{\mathrm{ul}}^{(\mathrm{p})} \rho_{\mathrm{ul}}^{(\mathrm{p})}} \sum_{k \in \mathcal{U}} g_{mk} \varphi_{\mathrm{ul},k} + \mathbf{w}_m, \tag{4}$$

where ρ_{ul} is the normalised transmit power, defined as the ratio of transmit power of the pilot signal over noise power, $\mathcal{U}$ is the set of users, $\mathbf{w}_m \in \mathbb{C}^{\tau_{\mathrm{ul}}^{(\mathrm{p})} \times 1}$ is the additive white Gaussian noise (AWGN) vector and its elements are independent and identically distributed (i.i.d.) random variables $\mathcal{CN}(0,1)$. For the general case of non-orthogonal pilot sequences, the greedy pilot assignment had been proposed in [8]. The pilot sequences in this work are assumed to be pairwise orthogonal. Then, the received pilot signal of user k at AP m, $\tilde{\mathbf{y}}_{mk}^{(\mathrm{p})}$, can be extracted by projection the received signal $\mathbf{y}_m^{(\mathrm{p})}$ onto $\varphi_{\mathrm{ul},k}^H$ as

$$\tilde{\mathbf{y}}_{mk}^{(\mathrm{p})} = \sqrt{\tau_{\mathrm{ul}}^{(\mathrm{p})} \rho_{\mathrm{ul}}^{(\mathrm{p})}} \, g_{mk} + \varphi_{\mathrm{ul},k}^H \mathbf{w}_m. \tag{5}$$

AP m uses the information from (5) and minimum mean square error (MMSE) approach to estimate the channel g_{mk}. Let $\hat{g}_{mk}$ denote the estimation of g_{mk}, it is given by [8, 9]

$$\hat{g}_{mk} = \frac{\sqrt{\tau_{\mathrm{ul}}^{(\mathrm{p})} \rho_{\mathrm{ul}}^{(\mathrm{p})}} \beta_{mk}}{1 + \tau_{\mathrm{ul}}^{(\mathrm{p})} \rho_{\mathrm{ul}}^{(\mathrm{p})} \beta_{mk}} \tilde{\mathbf{y}}_{mk}^{(\mathrm{p})} \overset{(a)}{=} \mu_{mk}^{1/2} z_{mk}, \tag{6}$$

where (a) is obtained due to $\tilde{\mathbf{y}}_{mk}^{(\mathrm{p})}$ being Gaussian distributed, $\mu_{mk} \triangleq \frac{\tau_{\mathrm{ul}}^{(\mathrm{p})} \rho_{\mathrm{ul}}^{(\mathrm{p})} \beta_{mk}^2}{1+\tau_{\mathrm{ul}}^{(\mathrm{p})} \rho_{\mathrm{ul}}^{(\mathrm{p})} \beta_{mk}}$, $z_{mk} \in \mathcal{CN}(0,1)$.

Let $\tilde{\varepsilon}_{mk}$ represent the channel estimation error, it is defined by

$$\tilde{\varepsilon}_{mk} \triangleq g_{mk} - \hat{g}_{mk}. \tag{7}$$

Then, from the MMSE channel estimation property [15], $\tilde{\varepsilon}_{mk}$ and $\hat{g}_{mk}$ are uncorrelated [10]. Since $\tilde{\varepsilon}_{mk}$ and $\hat{g}_{mk}$ are Gaussian distributed, they are independent. As a result, we obtain

$$\mathbb{E}\left\{|\tilde{\varepsilon}_{mk}|^2\right\} = \beta_{mk} - \mu_{mk}. \tag{8}$$

Uplink Data Transfer.

Let q_k, $\mathbb{E}\left\{|q_k|^2\right\} = 1$, be the transmit symbol of user k and ρ_{ul} be the normalized signal-to-noise ratio. The received signal at AP m is

$$y_m = \sqrt{\rho_{\text{ul}}} \sum_{k\in\mathcal{U}} g_{mk} q_k + w_m \tag{9}$$

Next, APs forward the received signals y_m, $m \in \mathcal{A}$, to CPU via perfect back-haul link. To detect the desired information that user k sent, the summation of y_m, $m \in \mathcal{A}$, is weighted by $\frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|}$ as follows:

$$y_k^{(\text{e})} = \sum_{m\in\mathcal{A}} \left(\sqrt{\rho_{\text{ul}}} \sum_{k'\in\mathcal{U}} \frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|} g_{mk'} q_{k'} + \frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|} w_m \right). \tag{10}$$

Remark 1. The equation (10) is obtained for the case of CPU using the information of all APs. For the case that CPU only selects some APs to involve the signal detection, m is in $\mathcal{A}^{\text{sel}}$, where $\mathcal{A}^{\text{sel}}$ defined in Algorithm 1 is a subset of $\mathcal{A}$.

3. Achievable Uplink Rate

Given the location of APs and users, the expression for achievable uplink rate of an arbitrary user k based on the channel hardening property is tight when the number of APs is very large. In this section, by applying the information from estimation phase into the signal detection process, the derived expression can be applied to general case of number of APs.

3.1. Detect desired signal based on channel information estimation

For this scheme, the desired signals are extracted from the received signal expressed in (10), and the estimated CSI defined in (6). Thus, APs need to share this information with the CPU.

Firstly, we can rewrite (10) as

$$y_k^{(\text{e})} = \sum_{m\in\mathcal{A}} \sqrt{\rho_{\text{ul}}} \frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|} g_{mk} q_k + \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \sqrt{\rho_{\text{ul}}} \frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|} g_{mk'} q_{k'} + \sum_{m\in\mathcal{A}} \frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|} w_m, \tag{11}$$

where $\mathcal{U}^k$ represents all users except user k.

Next, substituting (7) into (11), the expression for received signal at the CPU is given as

$$y_k^{(\text{e})} = \sum_{m\in\mathcal{A}} \sqrt{\rho_{\text{ul}}} |\hat{g}_{mk}| q_k + \sum_{m\in\mathcal{A}} \sqrt{\rho_{\text{ul}}} \frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|} \tilde{\varepsilon}_{mk} q_k + \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \sqrt{\rho_{\text{ul}}} \frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|} g_{mk'} q_{k'} + \sum_{m\in\mathcal{A}} \frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|} w_m \tag{12}$$

Observing (12), the first term is the desired signal, the second term is the estimation error, the third term represents the multi-users interference, and the last is the AWGN. Similarly to [8], the third and the fourth term can be treated as the effective noise. Using [16], the expression for SINR is given by equation (13) in the next page.

Then, the result of equation (13) is presented in the following lemma.

Lemma 1. The SINR of CPU in relation to the user k at CPU given estimated channel information is expressed by equation (14) in the next page.

Remark 2. By observing (14), the term $\sum_{m\in\mathcal{A}} \sum_{k\in\mathcal{U}} \rho_{\text{ul}}(\beta_{mk} - \mu_{mk}) + M$, representing estimation error and AWGN, is proportional to M. It means that the more APs join to the signal detection, the stronger this term becomes.

Proof. See Appendix A. □

The value of SINR in (14) changes over various coherence intervals, as $\hat{g}_{mk}$ and $\hat{g}_{mk'}$ keep constant in a coherence interval. To characterise the SINR, the cumulative distribution function (CDF) of SINR is derived in Lemma 2.

Lemma 2. The CDF of SINR $\gamma^{(\text{e})}_{\text{ul},k}$ is derived as follows:

$$\mathcal{F}_{\gamma^{(\text{e})}_{\text{ul},k}}(\gamma_{\text{th}}) = \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \left[\prod_{n\in\mathcal{A}} \prod_{k''\in\mathcal{U}^k} \left(1 - \frac{\mu_{nk''}}{\mu_{mk'}}\right)^{-1} \right]_{(n,k'')\neq(m,k')} \times \exp\left\{\frac{a}{\mu_{mk'}}\right\} \left(1 + \frac{\Omega_{\mathcal{X}}}{\Theta_{\mathcal{X}} \mu_{mk'} \gamma_{\text{th}}}\right)^{-\Theta_{\mathcal{X}} M}, \tag{15}$$

where $\Omega_{\mathcal{X}}$ and $\Theta_{\mathcal{X}}$ are defined in (B.5) and (B.6), respectively.

Proof. See Appendix B. □

From Lemma 2, the outage probability and ergodic achievable rate of a user k for the uplink are presented in Theorem 1 and Theorem 2 in the sequel.

Theorem 1. The outage probability, defined as the probability that the uplink rate of user k is below a given threshold R_{th}, is given as

$$P_{out}(R_{\text{th}}) = \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \left[\prod_{n\in\mathcal{A}} \prod_{k''\in\mathcal{U}^k} \left(1 - \frac{\mu_{nk''}}{\mu_{mk'}}\right)^{-1} \right]_{(n,k'')\neq(m,k')} \times \exp\left\{\frac{a}{\mu_{mk'}}\right\} \left(1 + \frac{\Omega_{\mathcal{X}}}{\Theta_{\mathcal{X}} \mu_{mk'} \bar{R}_{\text{th}}}\right)^{-\Theta_{\mathcal{X}} M}, \tag{16}$$

where $\bar{R}_{\text{th}} = 2^{R_{\text{th}}} - 1$.

$$\gamma_{\mathrm{ul},k}^{(\mathrm{e})} = \frac{\mathbb{E}\left\{\left|\sum_{m\in\mathcal{A}} \sqrt{\rho_{\mathrm{ul}}}|\hat{g}_{mk}|\right|^2\right\}}{\mathbb{V}\mathrm{ar}\left\{\sum_{m\in\mathcal{A}} \sqrt{\rho_{\mathrm{ul}}}\frac{\hat{g}_{mk}^*}{|\hat{g}_{mk}|}\tilde{\varepsilon}_{mk}\right\} + \mathbb{V}\mathrm{ar}\left\{\sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^k} \rho_{\mathrm{ul}}\frac{\hat{g}_{mk}^*}{|\hat{g}_{mk}|}g_{mk'}\right\} + \mathbb{V}\mathrm{ar}\left\{\sum_{m\in\mathcal{A}}\frac{\hat{g}_{mk}^*}{|\hat{g}_{mk}|}w_m\right\}} \tag{13}$$

$$\gamma_{\mathrm{ul},k}^{(\mathrm{e})} = \frac{\left(\sum_{m\in\mathcal{A}} \sqrt{\rho_{\mathrm{ul}}}|\hat{g}_{mk}|\right)^2}{\sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^k}\rho_{\mathrm{ul}}|\hat{g}_{mk'}|^2 + \sum_{m\in\mathcal{A}}\sum_{k\in\mathcal{U}}\rho_{\mathrm{ul}}(\beta_{mk}-\mu_{mk}) + M} \tag{14}$$

Theorem 2. The ergodic achievable uplink rate of user k is shown as

$$\begin{aligned}\mathcal{R}^{(\mathrm{e})} = &\sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^k}\left[\prod_{n\in\mathcal{A}}\prod_{k''\in\mathcal{U}^k}\left(1-\frac{\mu_{nk''}}{\mu_{mk'}}\right)^{-1}\right]_{(n,k'')\neq(m,k')} \\ &\times \exp\left\{\frac{a}{\mu_{mk'}}\right\}\frac{M\Theta_{\mathcal{X}}}{\ln(2)(M\Omega_{\mathcal{X}}-1)}\left[\frac{\Theta_{\mathcal{X}}\mu_{mk'}}{\Omega_{\mathcal{X}}(M\Theta_{\mathcal{X}}-2)}\right. \\ &\times {}_2F_1\left(1,1,3-\Theta_{\mathcal{X}}M,\frac{\Theta_{\mathcal{X}}\mu_{mk'}}{\Omega_{\mathcal{X}}}\right) + \epsilon - \pi Csc\left(\pi M\Theta_{\mathcal{X}}\right) \\ &\times \left.\left(\frac{\Omega_{\mathcal{X}}}{\Theta_{\mathcal{X}}\mu_{mk'}}-1\right)^{1-M\Theta_{\mathcal{X}}} + \ln\left(\frac{\Omega_{\mathcal{X}}}{\Theta_{\mathcal{X}}\mu_{mk'}}\right) + \quad (M\Omega_{\mathcal{X}}-1)\right],\end{aligned} \tag{17}$$

where ϵ is the Euler's constant, $\psi(.)$ is Poly gamma function and $Csc(.)$ is Cosecant function.

Proof.

$$\begin{aligned}\mathcal{R}^{(\mathrm{e})} &= \mathbb{E}\left\{\log_2\left(1+\gamma_{\mathrm{ul},k}^{(\mathrm{e})}(x)\right)\right\} \\ &= \int_0^\infty \log_2\left(1+\gamma_{\mathrm{ul},k}^{(\mathrm{e})}(x)\right) \mathrm{f}_{\gamma_{\mathrm{ul},k}^{(\mathrm{e})}}(x)dx,\end{aligned} \tag{18}$$

in which the PDF of $\gamma_{\mathrm{ul},k}^{(\mathrm{e})}(x)$ is

$$\begin{aligned}&\mathrm{f}_{\gamma_{\mathrm{ul},k}^{(\mathrm{e})}}(x) = \\ &\sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^k}\left[\prod_{n\in\mathcal{A}}\prod_{k''\in\mathcal{U}^k}\left(1-\frac{\mu_{nk''}}{\mu_{mk'}}\right)^{-1}\right]_{(n,k'')\neq(m,k')} \\ &\times \exp\left\{\frac{a}{\mu_{mk'}}\right\}\frac{M\Omega_{\mathcal{X}}}{\mu_{mk'}x^2}\left(1+\frac{\Omega_{\mathcal{X}}}{\Theta_{\mathcal{X}}\mu_{mk'}x}\right)^{-\Theta_{\mathcal{X}}M-1}\end{aligned} \tag{19}$$

By substituting $\mathrm{f}_{\gamma_{\mathrm{ul},k}^{(\mathrm{e})}}(x)$ into (18), (18) is re-written as

$$\begin{aligned}\mathcal{R}^{(\mathrm{e})} = &\sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^k}\left[\prod_{n\in\mathcal{A}}\prod_{k''\in\mathcal{U}^k}\left(1-\frac{\mu_{nk''}}{\mu_{mk'}}\right)^{-1}\right]_{(n,k'')\neq(m,k')} \\ &\times \exp\left\{\frac{a}{\mu_{mk'}}\right\}\int_0^\infty \log_2(1+x)\frac{M\Omega_{\mathcal{X}}}{\mu_{mk'}x^2} \\ &\times\left(1+\frac{\Omega_{\mathcal{X}}}{\Theta_{\mathcal{X}}\mu_{mk'}x}\right)^{-\Theta_{\mathcal{X}}M-1} dx\end{aligned} \tag{20}$$

By resolving the integral above, the final result is found. □

3.2. APs Selection Scheme

Observing (14), the term $\sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^k}\rho_{\mathrm{ul}}|\hat{g}_{mk'}|^2$ represents the summation of interference caused by user $k' \in \mathcal{U}^k$ on user k. To improve SINR, an AP selection scheme is proposed which is based on a new criterion defined in Algorithm 1.

Algorithm 1 APs Selection

1: APs send μ_{mk} ($m \in \mathcal{A}$, $k \in K$) to CPU
2: For each m in $\mathcal{A}$, calculate a weight factor corresponding to user k, which is defined by $w_{m,k} = \frac{\mu_{mk}}{\sum_{k\in\mathcal{U}^k}\mu_{mk}}$.
3: Order the set of $w_{m,k}$ in the descending order.
4: CPU chooses N APs that have largest weight factors, denoted by $\mathcal{A}^{\mathrm{sel}}$.

Remark 3. 1) By reducing the number of APs involving in the signal detection, the number of computations at CPU significantly decreases leading to gaining better performance in terms of high speed processing, energy efficiency and requirement of backhaul links; 2)The APs selection depends on the large scale fading, which slowly changes over time so that the selection does not require to be performed every coherence time; 3) The requirement of information of the large scale fading is not only for APs selection but also for power allocation. This means that there is no more data travelling over the backhaul links.

The results in Lemma 1, Lemma 2, Theorem 1 and Theorem 2 are also to be applied to the case of CPU deploying APs selection scheme with a substitution of $\mathcal{A}$ by $\mathcal{A}^{\mathrm{sel}}$.

3.3. Detect desired signal based on statistical knowledge of the channel

Considering the case of CPU using the statistical knowledge of channels to detect the desired information, the received signal at CPU is re-written as

$$y_k^{(h)} = \mathbb{E}\left\{\sum_{m\in\mathcal{A}} \sqrt{\rho_{ul}} \frac{\hat{g}_{mk}^*}{|\hat{g}_{mk}|} g_{mk}\right\} q_k + \left(\sum_{m\in\mathcal{A}} \sqrt{\rho_{ul}} \frac{\hat{g}_{mk}^*}{|\hat{g}_{mk}|} g_{mk} - \mathbb{E}\left\{\sum_{m\in\mathcal{A}} \sqrt{\rho_{ul}} \frac{\hat{g}_{mk}^*}{|\hat{g}_{mk}|} g_{mk}\right\}\right) q_k + \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \sqrt{\rho_{ul}} \frac{\hat{g}_{mk}^*}{|\hat{g}_{mk}|} g_{mk'} q_{k'} + \sum_{m\in\mathcal{A}} \frac{\hat{g}_{mk}^*}{|\hat{g}_{mk}|} w_m. \quad (21)$$

Then, the expression for SINR is shown as

Using similar approach of finding the achievable downlink rate in [9], the achievable uplink rate for the considered system model is derived by the following theorem.

Theorem 3. The achievable uplink rate of user k for the case of using statistical knowledge of channel is given

$$\mathcal{R}^{(h)} = \log_2\left(1 + \frac{\frac{\pi}{4}\left(\sum_{m\in\mathcal{A}} \sqrt{\rho_{ul}\mu_{mk}}\right)^2}{\sum_{m\in\mathcal{A}} \sum_{k\in K_m} \rho_{ul}\beta_{mk} - \frac{\pi}{4}\sum_{m\in\mathcal{A}} \rho_{ul}\mu_{mk} + M}\right) \quad (23)$$

4. Numerical Results

To evaluate the performance of the proposed system, a net throughput of user k metric is used, and it is defined as follows:

$$S_{m_k} = B\frac{\tau_c - \tau_{ul}^p}{\tau_c}\mathcal{R}^x \quad (24)$$

where $B = 20$ (Mhz) is the bandwidth, $\tau_c = 200$ is the coherence interval in samples, τ_{ul}^p is equal to the number of users, and $x = \{e, h\}$.

he system parameters are chosen as follows [8]: APs and users are randomly, uniformly located within a circle cell of radius $R = 1000$ m, pilot signals and symbol signals are modulated at the carrier frequency $f_c = 1.9$ GHz, using the three slopes path loss which is as shown in (2) with the standard deviation of the shadowing of $\sigma_{sh} = 8$ dB. The height of users and APs are $h_u = 1.65$ m, $h_{AP} = 15$ m, respectively, noise power $N_0 = -90$ dBm, and the transmit power of pilot and information signal are $\rho_{ul}^{(p)} = \rho_{ul} = 100$ (mW).

4.1. The CDF of SINR

The CDF of SINR with various users and APs location are shown in the figure 2. It shows that the simulation results and analysis results are identical.

4.2. Net throughput per user

The aims of this subsection is to figure out the differences in systems performance between with and without deploying AP selection, between two signal detection approaches, i.e. based on the knowledge of statistical channels and based on the estimated channel knowledge. As can be clearly observed from Fig. 3a and Fig. 4a, we can achieve the followings

- For a large number of APs joining the process (without applying AP selection), the lines for the CDFs of per-user uplink rate of two signal detections are approximate to each other. This is because the channels experiences hardening, which results in their values being close to their expectation, close to the estimated values. However, there is a gap between the lines,$d1$ and $d2$, for perfect CSI and the rests. This is the result from estimation error and additive noise, which become stronger for large M. This also reflects the discussion in Lemma 1.

- For moderate and small number of APs joining the process, the distance $d2$ from the line representing the case of using statistical CSI to the line representing the case of perfect CSI is quite large, because the channels are less hardening. More importantly, the lines for using estimated channel is approximate to the line for perfect ISI. Thus, by using the estimated channel information in this case can gain better performance compared with using the channel hardening property.

Following are some remarks on the affects of deploying the proposed APs selection scheme:

- Fig. 3b shows that the CDF of per-user uplink rate with AP selection outperforms that without AP selection in 95%-likely performance, i.e., the 95%-likely net throughput of the case with AP selection and proposed signal detection is around 6.2 Mbits/s (for $N = 20$), which is twice as high as that of the case without AP selection (around 3.1 Mbits/s). Particularly, this improvement is more significantly with higher value of likely net throughput. This is because the system discards APs that have strong inter-user interference.

- The difference of Fig. 3 and Fig. 4 is the number of APs, e.g., $M = 200$ and $M = 100$, respectively. In other words, the density of APs in Fig. 3 is two-time higher that in Fig. 4. From those two figures, with the same number of selected APs, the CDF of per-user uplink rate with higher density case is better. This can be explained as more APs generate more options for APs selection, resulting better performance.

$$\gamma_{\mathrm{ul},k}^{(\mathrm{h})} = \frac{\left|\mathbb{E}\left\{\sum_{m\in\mathcal{A}}\sqrt{\rho_{\mathrm{ul}}}\frac{\hat{g}_{mk}^{*}}{|\hat{g}_{mk}|}g_{mk}\right\}\right|^{2}}{\mathbb{V}\mathrm{ar}\left\{\sum_{m\in\mathcal{A}}\sqrt{\rho_{\mathrm{ul}}}\frac{\hat{g}_{mk}^{*}}{|\hat{g}_{mk}|}g_{mk}-\mathbb{E}\left\{\sum_{m\in\mathcal{A}}\sqrt{\rho_{\mathrm{ul}}}\frac{\hat{g}_{mk}^{*}}{|\hat{g}_{mk}|}g_{mk}\right\}\right\}+\mathbb{V}\mathrm{ar}\left\{\sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^{k}}\rho_{\mathrm{ul}}\frac{\hat{g}_{mk}^{*}}{|\hat{g}_{mk}|}g_{mk'}\right\}+\mathbb{V}\mathrm{ar}\left\{\sum_{m\in\mathcal{A}}\frac{\hat{g}_{mk}^{*}}{|\hat{g}_{mk}|}w_{m}\right\}} \quad (22)$$

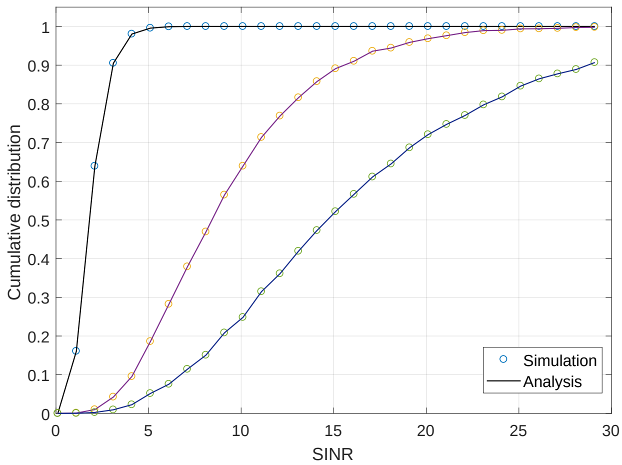

Figure 2. The CDF of SINR with various Users' and APs' locations

5. Conclusions and Discussion

A novel AP selection scheme is proposed and using estimated channel information to the signal detection is suggested for CF massive MIMO. Based on the proposed protocols, a new closed-form expression for ergodic uplink rate is derived. The results show that the system performance with the proposed schemes significantly improve. In addition to this advantage, by reducing the number of APs joining the signal detection, CPU can notably decrease the number of computations which leads to the enhancement of the energy efficiency and higher processing performance at the CPU.

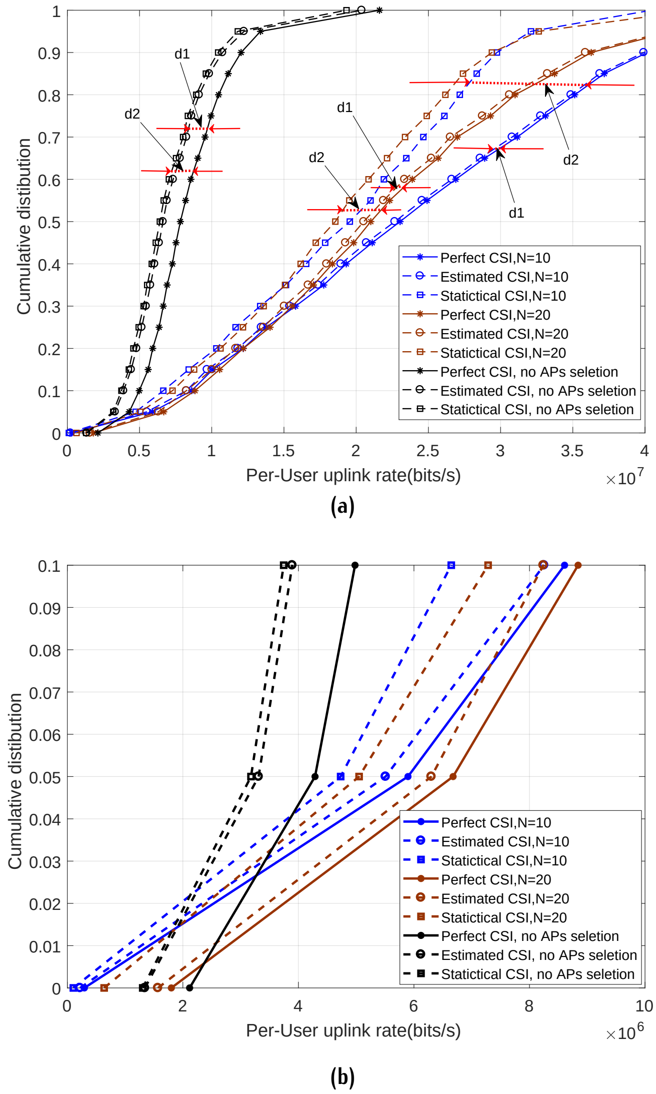

Figure 3. The cumulative distribution of the per-user uplink net throughput for the case of $M = 200$ and $K = 20$. (3b) is the zoom

Appendix A. Proof of Lemma 1

It is clear that $\mathbb{E}\left\{\left|\sum_{m\in\mathcal{A}}\sqrt{\rho_{\mathrm{ul}}}|\hat{g}_{mk}|\right|^{2}\right\} = \left|\sum_{m\in\mathcal{A}}\sqrt{\rho_{\mathrm{ul}}}|\hat{g}_{mk}|\right|^{2}$ as $\hat{g}_{mk}$ is known. The remainder of this proof is to calculate the variances of terms which are considered as the effective noise as follow:

$$\mathbb{V}\mathrm{ar}\left\{\sum_{m\in\mathcal{A}}\sqrt{\rho_{\mathrm{ul}}}\frac{\hat{g}_{mk}^{*}}{|\hat{g}_{mk}|}\tilde{\varepsilon}_{mk}\right\} = \sum_{m\in\mathcal{A}}\rho_{\mathrm{ul}}\mathbb{E}\left\{|\tilde{\varepsilon}_{mk}|^{2}\right\} \quad (A.1)$$

$$\mathbb{V}\mathrm{ar}\left\{\sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^{k}}\rho_{\mathrm{ul}}\frac{\hat{g}_{mk}^{*}}{|\hat{g}_{mk}|}g_{mk'}\right\} = \sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^{k}}\rho_{\mathrm{ul}}\mathbb{E}\left\{\left|\hat{g}_{mk'}+\tilde{\varepsilon}_{mk'}\right|^{2}\right\}$$

$$= \sum_{m\in\mathcal{A}}\sum_{k'\in\mathcal{U}^{k}}\rho_{\mathrm{ul}}\left[\mathbb{E}\left\{\left|\hat{g}_{mk'}\right|^{2}\right\}+\mathbb{E}\left\{|\tilde{\varepsilon}_{mk'}|^{2}\right\}\right], \quad (A.2)$$

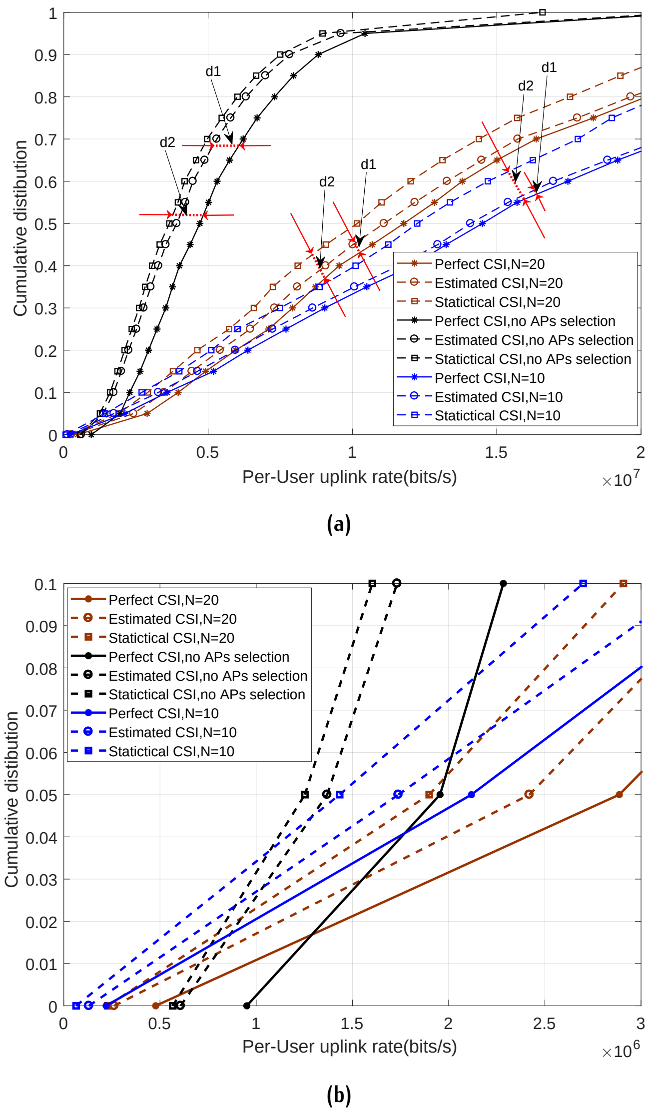

Figure 4. The cumulative distribution of the per-user uplink net throughput for the case of $M = 100$ and $K = 20$. (4b) is the zoom

$$\mathrm{Var}\left\{\sum_{m\in\mathcal{A}} \frac{\hat{g}^*_{mk}}{|\hat{g}_{mk}|} w_m\right\} = \sum_{m\in\mathcal{A}} \mathbb{E}\left\{|w_m|^2\right\} = M. \qquad \text{(A.3)}$$

By using $\mathbb{E}\left\{|\tilde{\varepsilon}_{mk'}|^2\right\} = \beta_{mk'} - \mu_{mk'}$ and substituting (A.1), (A.2) and (A.3) into (13), the expression for achievable uplink rate is obtained.

Appendix B. Proof of Lemma 2

$$\mathcal{F}_{\gamma^{(e)}_{\mathrm{ul},k}}(\gamma_{\mathrm{th}}) = P\left\{\gamma^{(e)}_{\mathrm{ul},k} < \gamma_{\mathrm{th}}\right\} = P\left\{\frac{\mathcal{X}}{\mathcal{Y}+a} < \gamma_{\mathrm{th}}\right\}$$

$$= 1 - P\left\{\mathcal{Y} < \left(\frac{\mathcal{X}}{\gamma_{\mathrm{th}}} - a\right)\right\}$$

$$= 1 - \int_0^{\infty} \mathcal{F}_{\mathcal{Y}}\left(\frac{x}{\gamma_{\mathrm{th}}} - a\right) \mathrm{f}_{\mathcal{X}}(x)dx \qquad \text{(B.1)}$$

where $\mathcal{F}_T(.)$ and $\mathrm{f}_T(.)$ are the CDF and probability density function (PDF) of random variable T, respectively, $\mathcal{X} \triangleq \left(\sum_{m\in\mathcal{A}} \sqrt{\rho_{\mathrm{ul}}}|\hat{g}_{mk}|\right)^2$, $\mathcal{Y} \triangleq \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \rho_{\mathrm{ul}}|\hat{g}_{mk'}|^2$ and $a \triangleq \sum_{m\in\mathcal{A}} \sum_{k\in\mathcal{U}} \rho_{\mathrm{ul}} \mathbb{E}\left\{|\tilde{\varepsilon}_{mk}|^2\right\} + M$. The CDF of $\mathcal{Y}$, with the assumption of all $\mu_{mk}, m = 1...M$ being distinct or no APs having the same location, can be expressed as [17, eq.37]

$$\mathcal{F}_{\mathcal{Y}}(y) = 1 - \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \left[\prod_{n\in\mathcal{A}} \prod_{k''\in\mathcal{U}^k} \left(1 - \frac{\mu_{nk''}}{\mu_{mk'}}\right)^{-1}\right]_{(n,k'')\neq(m,k')} \exp\left\{\frac{-y}{\mu_{mk'}}\right\}, \qquad \text{(B.2)}$$

Then, the equation (B.1) is re-written as

$$\mathcal{F}_{\gamma^{(e)}_{\mathrm{ul},k}}(\gamma_{\mathrm{th}}) = \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \left[\prod_{n\in\mathcal{A}} \prod_{k''\in\mathcal{U}^k} \left(1 - \frac{\mu_{nk''}}{\mu_{mk'}}\right)^{-1}\right]_{(n,k'')\neq(m,k')} \times \int_0^{\infty} \exp\left\{-\frac{\left(\frac{x}{\gamma_{\mathrm{th}}} - a\right)}{\mu_{mk'}}\right\} \mathrm{f}_{\mathcal{X}}(x)dx$$

$$= \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \left[\prod_{n\in\mathcal{A}} \prod_{k''\in\mathcal{U}^k} \left(1 - \frac{\mu_{nk''}}{\mu_{mk'}}\right)^{-1}\right]_{(n,k'')\neq(m,k')} \times \exp\left\{\frac{a}{\mu_{mk'}}\right\} \mathcal{M}_{\mathrm{f}_{\mathcal{X}}(x)}\left\{\frac{1}{\gamma_{\mathrm{th}}\mu_{mk'}}\right\}, \qquad \text{(B.3)}$$

where $\mathcal{M}_T\{s\}$ is the moment generating function (MGF).

It is noted that a closed-form of MGF of a square of a summation of independent and non-identically distributed Rayleigh random variables does not exist (to the best of my knowledge). Thus, it is approximated as [18, eq.5]

$$\mathcal{M}_{\mathrm{f}_{\mathcal{X}}(x)}\{s\} = \left(1 - \frac{\Omega_{\mathcal{X}}}{\Theta_{\mathcal{X}}} s\right)^{-\Theta_{\mathcal{X}} M}, \qquad \text{(B.4)}$$

where

$$\Omega_{\mathcal{X}} = \frac{1}{M} \sum_{m\in\mathcal{A}} \mu_{mk} + \frac{\pi}{2M} \sum_{m\in\mathcal{A}} \sum_{n=m+1}^{M} \sqrt{\mu_{nk''}\mu_{mk}} \qquad \text{(B.5)}$$

and

$$\Theta_{\mathcal{X}} = \frac{M\Omega^2_{\mathcal{X}}}{\mathbb{E}\{X^2\} - M^2\Omega^2_{\mathcal{X}}}, \qquad \text{(B.6)}$$

$$\mathbb{E}\left\{X^2\right\} = 2\sum_{m\in\mathcal{A}} \mu_{mk}^2 + \sum_{m\in\mathcal{A}} \sum_{n=m+1}^{M} \Big(6\mu_{mk}\mu_{nk''} + \frac{3\pi}{2}\sqrt{\mu_{mk}^3\mu_{nk''}} + \frac{3\pi}{2}\sqrt{\mu_{mk}\mu_{nk''}^3}\Big) + 3\pi\sum_{m\in\mathcal{A}} \sum_{n=m+1}^{M} \sum_{t=n+1}^{M} \Big(\sqrt{\mu_{mk}^2\mu_{nk''}\mu_{tk}} + \sqrt{\mu_{mk}\mu_{nk''}^2\mu_{tk}} + \sqrt{\mu_{mk}\mu_{nk''}\mu_{tk}^2}\Big) + \frac{3\pi^2}{2}\sum_{m\in\mathcal{A}} \sum_{n=m+1}^{M} \sum_{t=n+1}^{M} \sum_{l=t+1}^{M} \sqrt{\mu_{mk}\mu_{nk''}\mu_{tk}\mu_{lk}}. \tag{B.7}$$

Finally, the CDF of SINR is computed as

$$\mathcal{F}_{\gamma_{\mathrm{ul},k}^{(\mathrm{e})}}(\gamma_{\mathrm{th}}) = \sum_{m\in\mathcal{A}} \sum_{k'\in\mathcal{U}^k} \left[\prod_{n\in\mathcal{A}} \prod_{k''\in\mathcal{U}^k} \left(1 - \frac{\mu_{nk''}}{\mu_{mk'}}\right)^{-1}\right]_{(n,k'')\neq(m,k')} \times \exp\left\{\frac{a}{\mu_{mk'}}\right\} \left(1 + \frac{\Omega_{\mathcal{X}}}{\Theta_{\mathcal{X}}\mu_{mk'}\gamma_{\mathrm{th}}}\right)^{-\Theta_{\mathcal{X}}M}. \tag{B.8}$$

Acknowledgements

This work was supported by a Research Environment Links grant, ID 339568416, under the Newton Programme Vietnam partnership. The grant is funded by the UK Department of Business, Energy and Industrial Strategy (BEIS) and delivered by the British Council. For further information, please visit www.newtonfund.ac.uk./

References

[1] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] H. Q. Ngo and A. Ashikhmin and H. Yang and E. G. Larsson and T. L. Marzetta, "Cell-Free Massive MIMO: Uniformly great service for everyone," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, vol. , no. , Jun. 2015, pp. 201–205.

[3] T. C. Mai, H. Q. Ngo, M. Egan, and T. Q. Duong, "Pilot power control for cell-free massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11 264–11 268, Nov 2018.

[4] T. M. Hoang, H. Q. Ngo, T. Q. Duong, H. D. Tuan, and A. Marshall, "Cell-free massive MIMO networks: Optimal power control against active eavesdropping," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4724–4737, Oct 2018.

[5] H. Q. Ngo and L. N. Tran and T. Q. Duong and M. Matthaiou and E. G. Larsson, "On the Total Energy Efficiency of Cell-Free Massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.

[6] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, Aug 2017.

[7] E. Nayebi and A. Ashikhmin and T. L. Marzetta and H. Yang, "Cell-Free Massive MIMO systems," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, vol. , no. , Nov. 2015, pp. 695–699.

[8] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[9] T. X. Doan and H. Q. Ngo and T. Q. Duong and K. Tourki, "On the Performance of Multigroup Multicast Cell-Free Massive MIMO," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2642–2645, Dec 2017.

[10] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-free Massive MIMO with limited backhaul," in *in Proc. IEEE ICC*, vol. , no. , May. 2018, pp. 1–7.

[11] M. Bashar, K. Cumanan, A. G. Burr, , M. Debbah, and H. Q. Ngo, "Enhanced max-min SINR for uplink cell-free Massive MIMO systems," in *in Proc. IEEE ICC*, vol. , no. , May. 2018, pp. 1–6.

[12] J. Jose and A. Ashikhmin and 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.

[13] J. Hoydis and S. ten Brink and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[14] Z. Chen and E. Bjoernson, "Can We Rely on Channel Hardening in Cell-Free Massive MIMO?" in *2017 IEEE Globecom Workshops (GC Wkshps)*, vol. , no. , Dec. 2017, pp. 1–6.

[15] S. M. Kay, *Fundamentals of statistical signal processing, volume I: Estimation theory*. Prentice Hall, 1993.

[16] T. L. Marzetta, E. G. Larsson, H. Yang and H. Q. Ngo, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[17] A. Bletsas and H. Shin and M. Z. Win, "Cooperative Communications with Outage-Optimal Opportunistic Relaying," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3450–3460, Sept. 2007.

[18] N. Zlatanov and Z. Hadzi-Velkov and G. K. Karagiannidis, "An efficient approximation to the correlated Nakagami-m sums and its application in equal gain diversity receivers," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 302–310, Jan. 2010.