User-Centric C-RAN Architecture for Ultra-Dense 5G Networks: Challenges and Methodologies

Cunhua Pan, Maged Elkashlan, Jiangzhou Wang, Jinhong Yuan, and Lajos Hanzo

The authors advocate the application of dense user-centric C-RAN philosophy to UDNs, thanks to the recent development of cloud computing techniques. Under dense C-RAN, centralized signal processing can be invoked for supporting CoMP transmission. They summarize the main challenges in dense user-centric C-RANs, and highlight several promising research directions to make dense user-centric C-RAN a reality.

ABSTRACT

Ultra-dense networks (UDNs) constitute one of the most promising techniques for supporting the 5G mobile system. By deploying more small cells in a fixed area, the average distance between users and access points can be significantly reduced; hence, dense spatial frequency reuse can be exploited. However, severe interference is the major obstacle in UDNs. Most contributions deal with the interference by relying on cooperative game theory. This article advocates the application of dense user-centric C-RAN philosophy to UDNs, thanks to the recent development of cloud computing techniques. Under dense C-RAN, centralized signal processing can be invoked for supporting CoMP transmission. We summarize the main challenges in dense user-centric C-RANs. One of the most challenging issues is the requirement of the global CSI for the sake of cooperative transmission. We investigate this requirement by only relying on partial CSI, namely, on inter-cluster large-scale CSI. Furthermore, the estimation of the intra-cluster CSI is considered, including the pilot allocation and robust transmission. Finally, we highlight several promising research directions to make the dense user-centric C-RAN become a reality, with special emphasis on the application of big data techniques.

INTRODUCTION

The fifth generation (5G) wireless system is anticipated to offer substantially increased data throughput compared to the fourth generation (4G) system. To achieve this ambitious goal, ultra dense networks (UDNs) have been widely regarded as one of the most promising solutions [1]. In UDNs, the average distance between users and small cell base stations (BSs) is significantly reduced; hence, the link quality is dramatically improved, which further increases the network capacity.

However, the drastic interference generated by the neighboring small cells is a limiting factor in UDNs. The attainable network performance may even be decreased when the BS density is extremely high [2]. Hence, the interference should be carefully managed in order to reap the poten-

tial benefits of UDNs. Most existing contributions deal with the interference by designing partially distributed algorithms based on game-theoretical approaches. By adopting cooperative game theory, multiple small cell BSs exchange the necessary information for their coordination through wired backhaul (BH) links, which works well for small-scale networks. However, for UDNs, the heavy overhead of coordination and the increasing cost of deploying wired BH links will preclude the application of this approach. Apart from the interference issues, employing more small cell BSs will also increase the maintenance and operational costs. Hence, a new network architecture should be adopted to support reliable communications in UDNs.

Due to the recent developments in cloud computing, the ultra-dense cloud radio access network (C-RAN) concept has been regarded as a promising network architecture that can efficiently address the issues arising in UDNs.

The ultra-dense C-RAN architecture is shown in Fig. 1, which consists of three key components:
- Baseband unit (BBU) pool supported by the techniques of cloud computing, network functions virtualization (NFV), software-defined networks (SDNs), and so on
- Low-cost, low-power radio remote heads (RRHs) distributed over the coverage area
- Wireless fronthaul links that connect the RRHs to the BBU pool

The main characteristic of dense C-RAN is that all the baseband signal processing units of conventional small cell BSs have been incorporated in the BBU pool, where the computing resources can be shared among the BBUs. Then conventional full-functionality small cell BSs can be replaced by low-cost, low-complexity RRHs, which are only responsible for the transmission/reception. Hence, RRHs can be densely deployed at low hardware cost to provide ultra-high throughput and seamless coverage for a large number of users in tele-traffic hotspots, such as airports and shopping malls. Thanks to the centralized architecture of C-RAN, the global network information can be shared in the BBU pool and cooperative communication techniques can be realized with the aid of powerful cloud computing, such as
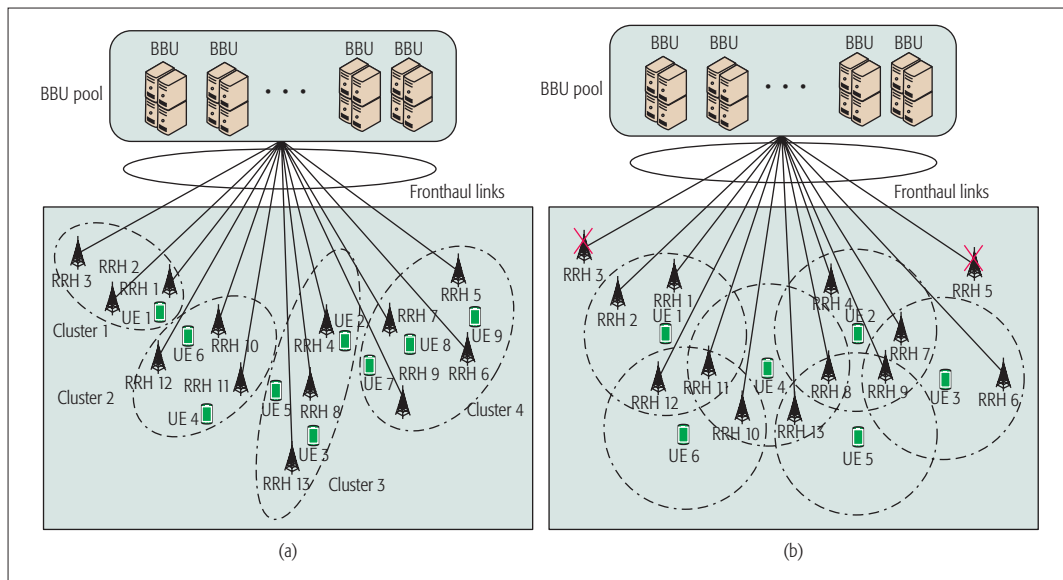
Cunhua Pan and Maged Elkashlan are with Queen Mary University of London; Jiangzhou Wang is with the University of Kent; Jinhong Yuan is with the University of New South Wales; Lajos Hanzo is with the University of Southampton.

Figure 1. Illustration of a C-RAN architecture: a) disjoint cluster, where the whole network is divided into several non-overlapped clusters, and the cluster edge users still suffer from high inter-cluster interference; b) user-centric cluster, where the cluster is formed from the user side and the cluster edge issues are eliminated.

Type of UDN	Functionality	Interference management	Connectivity	Deployment cost
Small cells	Fully functioning	Distributed algorithm	Wired backhaul	High maintenance cost
Dense C-RAN	PHY layer functioning	Centralized algorithm	Wireless fronthaul	Low hardware cost

Table 1. Different types of UDN deployment.

large-scale network coordination, global resource management, and coordinated multipoint (CoMP) processing. Hence, the cochannel interference that is a limiting factor in UDNs can be efficiently mitigated under the ultra-dense C-RAN architecture. Additionally, the operating status of RRHs and the computing resources of the BBU pool can be dynamically controlled in order to adapt to the capacity demand fluctuations of the users, which leads to significant energy and operational cost reductions. In [3], Tang *et al.* showed that the system cost, which consists of computation cost and wireless transmission cost (including both transmit power and circuit power), can be reduced by roughly 12 percent through jointly optimizing the number of active virtual machines (VMs) and operating status of RRHs and fronthaul links. Table 1 contrasts the differences between a conventional small cell network and the ultra-dense C-RAN.

However, there are some challenges that need to be handled in dense C-RAN before its successful practical deployment. The contributions of this article are summarized as follows:

• We first summarize the unique research challenges in dense C-RAN and the existing solutions for each challenge. We place an emphasis on the challenge of heavy training overhead for estimating all the channel state information (CSI) for CoMP transmission. Most of the existing works consider perfect intra-cluster CSI, which is impractical due to the limited amount of pilot resources.

• We provide a complete framework to deal with the imperfect intra-cluster CSI. One novel low-complexity pilot allocation algo-

rithm is first proposed by considering the multi-user pilot interference. Then robust beamforming vector design is considered to account for the CSI estimation error.

• Some promising research directions are highlighted, especially the cluster design with the aid of big data.

The rest of this article is organized as follows. The new challenges for dense C-RAN are presented; we provide a complete procedure to deal with the imperfect CSI, including both the pilot allocation and robust beamforming design. Simulation results are also shown. Finally, conclusions are drawn along with some future research directions.

RESEARCH CHALLENGES AND STATE-OF-THE-ART SOLUTIONS

In this section, we summarize the challenges arising in dense C-RAN along with the state-of-the-art solutions.

High Computational Complexity: In dense C-RAN, the BBU pool usually supports a large number of RRHs, and the number of variables to optimize, such as beamforming-vectors, will become excessive, even in the context of cloud computing. The most common technique of reducing the complexity is to adopt the cluster technique. In general, there are two types of cluster techniques, as shown in Fig. 1: disjoint clustering and user-centric clustering. In the disjoint cluster, all the RRHs in the network are partitioned into several non-overlapped clusters, and the RRHs in each cluster employ the CoMP technique to serve the users within the coverage area

of this cluster. However, the cluster edge users still suffer from severe inter-cluster interference. In contrast, in the user-centric cluster, each user is individually served by its nearby RRHs. The scheduled user is the center of the cluster. Different clusters may overlap with each other, which will eliminate the potential cluster edge effect. Hence, the user-centric cluster constitutes the focus of this article.

Stringent Fronthaul Capacity Requirement: In conventional C-RAN, the fronthaul links are typically wired links, such as optical fibers. However, in ultra-dense C-RAN a large number of fronthaul links are required. Laying wired links requires high operational and maintenance costs. An attractive alternative is to use wireless fronthaul links, such as millimeter-wave (mmWave) communications, which are much more scalable and cost-effective than wired links. However, the bandwidth of wireless links is much lower than that of wired links, which means that the number of users supported by each wireless link is much lower. The fronthaul capacity constraint has been extensively studied, and can be divided into two categories: the compression strategy and the data sharing strategy. In the compression strategy, the BBU pool first computes the beamforming vectors for each RRH. Then the beamformed signals are generated at the BBU pool, which are compressed and sent to the corresponding RRHs. The fronthaul capacity is related to how fine the resolution of the compressed signals is: higher resolutions require higher fronthaul capacity. Hence, the compression resolution should be optimized under the fronthaul capacity constraints. In [4], Park et al. proposed to leverage joint compression of the signals of different RRHs to better optimize the compression resolutions. In the second strategy, the beamforming vectors computed at the BBU pool are directly sent to the corresponding RRHs. Then the BBU pool shares each user's data directly with its serving cluster. The beamformed signals are generated at each RRH. In this strategy, the fronthaul capacity depends on the number of users served by each link. Hence, the user-RRH associations should be optimized under the fronthaul capacity constraints. In [5], Dai et al. considered the network utility maximization problem for the user-centric downlink C-RAN, where the per-RRH fronthaul capacity constraints are explicitly taken into account.

Huge Training Overhead for CSI Estimation: To facilitate CoMP transmission, the global CSI should be available at the BBU pool, which constitutes an excessive channel estimation overhead for dense C-RAN. Caire et al. [6] showed that the increasing amount of training overhead may outweigh the gains provided by CoMP. To reduce the overhead, a promising technique is to rely on partial CSI under the user-centric cluster, where each user only estimates the CSI of the links from the RRHs within its cluster (called intra-cluster CSI) and only tracks the path loss and shadowing outside its own cluster (called inter-cluster CSI), which are sent back to the BBU pool. These parameters are necessary for the CoMP design at the BBU pool. Indeed, the large-scale fading may be readily tracked since it changes slowly compared to the instantaneous CSI. The design of beamforming vectors under

this partial CSI case is a challenging task. Hence, there is a paucity of contributions [7, 8] based on partial CSI. To elaborate, compressive CSI acquisition was proposed in [7] for determining the set of instantaneous CSI and large-scale fading gains. However, its complexity is high, so it cannot be readily implemented in dense C-RAN. Recently, we provided a design framework in [8] to design green transmission under the partial CSI case. However, the intra-cluster CSI was assumed to be perfect in [7, 8], which is difficult to achieve in practice. Hence, in the article, we aim to consider the imperfect intra-cluster CSI case, where both the channel estimation procedure for intra-cluster CSI and robust beamforming vector design are considered in the following two sections.

TRANSMISSION SCHEME DESIGNED FOR IMPERFECT INTRA-CLUSTER CSI

For TDD C-RANs, the training pilots sent from different users to the same RRH should be mutually orthogonal so that the RRH can distinguish the channel vectors from different users. The number of pilots required scales linearly with the number of users, which becomes excessive for dense C-RANs. Hence, the number of time slots used for data transmission will be significantly reduced. To reduce the number of pilots, they may be reused by a group of users under the condition that none of the users in the group shares the same RRH with the other users. It is widely recognized that the pilot reuse scheme will impose the pilot contamination issue, which results in sizeable channel estimation errors. In the following, we propose a two-stage optimization method to optimize the transmissions for dense C-RAN. In Stage I, a novel pilot reuse scheme is proposed; in Stage II, a robust beamforming vector optimization algorithm is conceived by considering the pilot contamination incurred by Stage I.

STAGE I: NOVEL PILOT REUSE SCHEME

The pilot reuse issues have been extensively studied in massive multiple-input multiple-output (MIMO) systems [9]. The basic idea is to reuse the same pilot within the specific group of users having different angles of arrival, which is not applicable in dense C-RAN due to the limited number of antennas at each RRH. Recently, Chen et al. [10] proposed a novel pilot allocation scheme for dense C-RANs by using the classic Dsatur graph coloring algorithm, which minimizes the number of pilots required for a given set of users. However, for simplicity of implementation, the 4G LTE system suggests that the proportion of pilots designated for channel estimation should be fixed within the channel's coherence time (i.e., 1 percent for a 10 ms training period) [11]. Given the fixed number of pilots, some users may not be allocated any pilots. Hence, we aim to provide a joint user selection and pilot allocation scheme for maximizing the number of users admitted at a given number of available pilots, while satisfying the following two conditions:
- The users sharing the same RRH must not reuse the same pilot.
- There is an upper bound on how many times each pilot is reused, denoted as n_{max}.

The second constraint is imposed for guaranteeing fair use of the available pilots. In the following, a low-complexity nearly optimal algorithm is provided to deal with this problem.

For a dense C-RAN with K users, we construct a $(K \times K)$ matrix $\mathbf{B}$, where each element is given by

$$b_{k,k'} = \begin{cases} 1, \text{if } \mathcal{I}_k \cap \mathcal{I}_{k'} \neq \varnothing \text{ and } k \neq k' \\ 0, \text{otherwise} \end{cases} \quad (1)$$

where $\mathcal{I}_k$ denotes the set of RRHs that potentially serve user K. The above definitions mean that if user K and user k' share common RRHs, the corresponding element is set to $b_{k,k'} = 1$. Otherwise, the element is zero. Based on matrix $\mathbf{B}$, a unidirectional graph can be constructed for the network of Fig. 1b in Fig. 2a. Figure 2b illustrates the pilot allocation results after applying the Dsatur algorithm [10], which shows that at least three different pilots are required.

To solve the above pilot allocation problem, we first adopt the Dsatur algorithm to find the minimum number of pilots required. If this number is higher than the number of pilots available, some users should be removed. Otherwise, all users can be admitted. In the latter case, some pilots may be reused by up to n_{max} users, while some pilots are not allocated, which wastes the pilot resources. To resolve this problem, all available pilots should be used to reduce the pilot contamination. Let us denote the minimum number of pilots required by the Dsatur algorithm as n^*. In the following, we provide a detailed algorithm to deal with each case:

Case I. $n^* > \tau$

Case II. $n^* < \tau$

When $n^* = \tau$, the final solution is obtained.

Case I: $n^* > \tau$. In this case, some users should be removed. These users can be rescheduled for transmission in other time slots or frequency bands, which is beyond the scope of this article. Let us define $\bar{\mathcal{U}}$ as the set of all users, and $\theta_k \triangleq \sum_{k' \neq k, k' \in \bar{\mathcal{U}}} b_{k,k'}$ as the degree of the vertex for user k that represents the total number of users connected to this user in the constructed undirected graph. The user having the largest value of θ_k should be deleted with high priority, since many users should be allocated different pilots compared to this user. However, there may be several users with the same value of θ_k, and randomly removing one of them will lead to reduced performance. Hence, the user earmarked for deletion should be carefully selected. Intuitively, the user suffering from the highest pilot interference should be removed. To this end, we define the metric $\eta_{k,k'}$ to quantify the level of pilot interference between any two users when reusing the same pilot:

$$\eta_{k,k'} = \log\left(1 + \frac{\sum_{i \in \mathcal{I}_{k'}} \alpha_{i,k}}{\sum_{i \in \mathcal{I}_k} \alpha_{i,k}}\right) + \log\left(1 + \frac{\sum_{i \in \mathcal{I}_k} \alpha_{i,k'}}{\sum_{i \in \mathcal{I}_{k'}} \alpha_{i,k'}}\right), \quad (2)$$

where $\alpha_{i,k}$ represents the large-scale fading power from RRH i to user K. Obviously, a larger $\eta_{k,k'}$ means larger pilot interference between these two users. Then we quantify the level of pilot interference due to user K by $\xi_k = \sum_{k' \in \mathcal{K}\pi_k \setminus \{k\}} \eta_{k,k'}$ as our figure of merit, where π_k denotes the pilot index
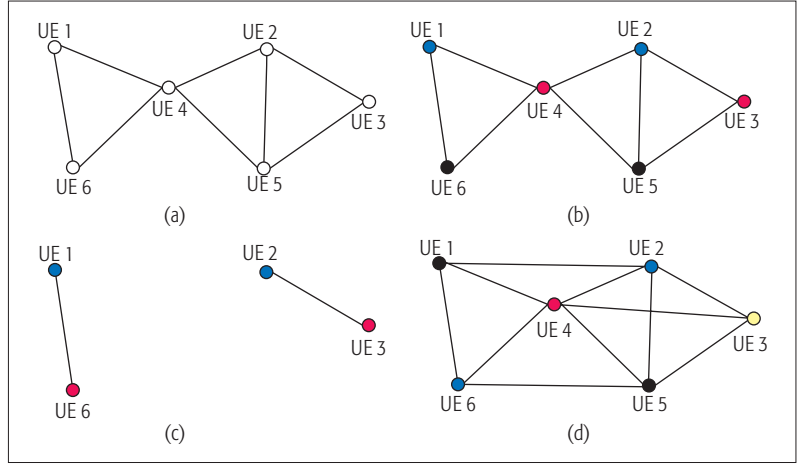

Figure 2. a) Construction of the undirected graph for the network in Fig. 1b; b) the colored graph after applying the Dsatur algorithm [10] with $n_{max} = 2$, the minimum number of required pilots is $n^* = 3$; c) the user selection and pilot allocation result after using the algorithm for Case I when $\tau = 2$ and $n_{max} = 2$; d) the pilot reallocation result after using the bisection search algorithm for Case II. In this pilot reallocation result, user 1 and user 2 are allocated different pilots to reduce the pilot interference, and the same holds for user 5 and user 6, or UE 3 and UE 4.

1: Initialize matrix $\mathbf{B}$, user set $\mathcal{U}$, initial number of required pilots n^* obtained from the Dsatur algorithm;

2: While $n^* > \tau$

3:    Find $k^* = \arg \max_{k \in \mathcal{U}} \theta_k$. If there are multiple users with the same θ_k, remove the user with the largest ξ_k;

4:    Remove user k^* from $\mathcal{U}$, i.e., $\mathcal{U} = \mathcal{U}/k^*$, and update matrix $\mathbf{B}$ with current $\mathcal{U}$;

5:    Use the Dsatur algorithm to calculate n^* with $\mathbf{B}$ and U;

Algorithm 1. User selection and pilot allocation algorithm for Case I.

---

used by user K and $\mathcal{K}\pi_k$ represents the set of users that reuses pilot π_k. The user with the largest ξ_k should be removed. Hence, we conceive the user selection and pilot allocation method formulated in Algorithm 1. By invoking it for the network in Fig. 1b, the final user selection result is shown in Fig. 2c.

Case II: $n^* < \tau$. In this case, we aim to reallocate all the available pilots to all users to additionally reduce the pilot interference. Note that in Fig. 2b, there may be measurable pilot interference between user 1 and user 2. If there are four pilots, these two users can be allocated different pilots as seen in Fig. 2d. We can reconstruct an undirected graph by introducing a threshold η_{th}. If $\eta_{k,k'} > \eta_{th}$, user K and user k' should be connected. Based on this idea, matrix $\mathbf{B}$ can be reconstructed as

$$b_{k,k'} = \begin{cases} 1, \text{if } \mathcal{I}_k \cap \mathcal{I}_{k'} \neq \varnothing \text{ and } k \neq k', \\ 1, \text{if } \eta_{k,k'} > \eta_{th}, \mathcal{I}_k \cap \mathcal{I}_{k'} = \varnothing \text{ and } k \neq k', \\ 0, \text{otherwise.} \end{cases} \quad (3)$$

As expected, a smaller value of η_{th} will result in more users becoming connected to each other; hence, more pilots are required. In the extreme case of $\eta_{th} < \min\{\eta_{k,k'}\}$, all users become connected and the number of pilots required is equal to K. On the other hand, if $\eta_{th} \geq \max\{\eta_{k,k'}\}$, the reconstructed matrix $\mathbf{B}$ reduces to the initial $\mathbf{B}$ defined in Eq. 1, where the number of pilots required is n^*. Since $\tau < K$, there must be a η_{th} value so that the number of pilots required is
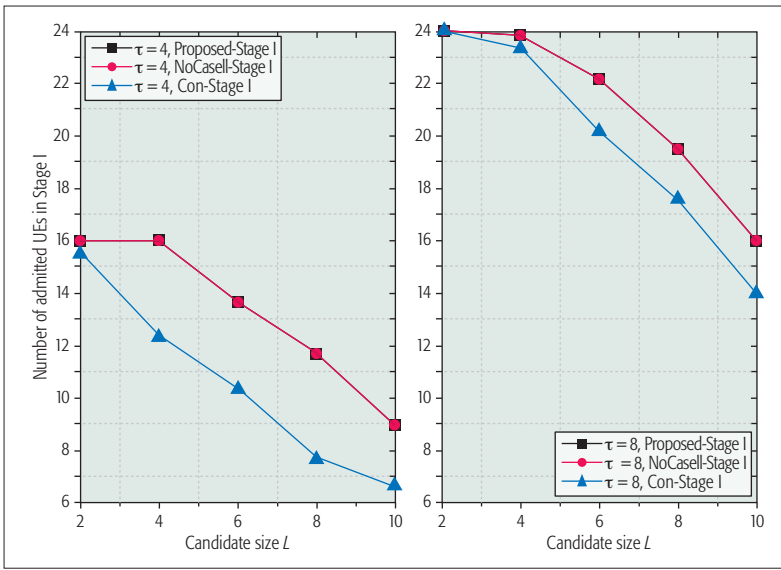
Figure 3. Number of admitted UEs in Stage I vs. the candidate size L. The left subplot corresponds to the case when the number of available pilots is $\tau = 4$, while the right one is $\tau = 8$.

equal to τ. The bisection search algorithm can be adopted to find this threshold η_{th}. Again, Fig. 2d shows the pilot allocation results after using this algorithm.

STAGE II: ROBUST BEAMFORMING-VECTOR DESIGN

In Stage II, we aim to design the beamforming vectors by considering the pilot contamination due to the pilot reuse scheme in Stage I. Specifically, we formulate a user selection problem for the dense C-RAN where each RRH is equipped with multiple antennas. The following three constraints are considered:
- Each user's data rate should be higher than its minimum requirement.
- Each fronthaul link capacity constraint is imposed.
- Each RRH has its individual power constraint.
  There are three challenges to solve this optimization problem:
- Since we consider the partial CSI case where only the inter-cluster large-scale fading parameters are available, the exact data rate of each user is difficult to obtain.
- The fronthaul capacity constraint has the non-smooth and non-differentiable indicator function, which is a mixed-integer non-linear programming (MINLP) problem that is NP-hard to solve [8].
- Due to the channel estimation error, each user suffers from residual self-interference. The conventional weighted minimum mean square error (WMMSE) method of [12], which has been successfully applied in the perfect intra-cluster CSI scenario [5, 8], cannot be used for solving the problem here.

We provide brief descriptions of the associated methods to address the above three challenges.

First, Jensen's inequality is used for finding the lower bound of the exact data rate, which is more amenable to the design of our algorithm. In [8], we have shown that for the specific scenario of non-overlapped cluster, the gap between the lower bound and the exact

data rate is within 3 percent for both sparse and dense C-RAN scenarios. Hence, a simple correction factor may be used for practical applications.

Second, the indicator function is replaced by a concave function

$$f_{\theta}(x) = \frac{x}{x+\theta} \ ,$$

where θ is a small positive value. The transformed problem is the difference of convex (d.c.) program, which can be efficiently solved by the successive convex approximation (SCA) method [8].

Finally, to deal with the final challenge, we adopt the semi-definite relaxation approach and formally prove that semidefinite relaxation is tight with a probability of 1.

SIMULATION RESULTS

We now present our simulation results for evaluating the performance of the proposed algorithms. The dense C-RAN is assumed to cover a square shaped area of 700 m × 700 m. The numbers of RRHs and users are set to $I = 36$ and $K = 24$ with densities of 73 RRHs/km^2 and 49 users/km^2, respectively. This is a typical 5G ultra-dense cellular network as stated in [13]. Both the users and RRHs are uniformly and independently distributed in this area. It is assumed that each user is potentially served by its nearest L RRHs. Each fronthaul link is assumed to only support three users, since mmWave communication is employed as the wireless fronthaul link. The maximum transmit power for each RRH is 100 mW, and the pilot power is 200 mW. The maximum pilot reuse time for each pliot is $n_{max} = 4$, and the rate requirement for each user is 4 b/s/Hz. We compare our proposed algorithm to the following algorithms:
- Orthogonal pilot allocation ("Ortho"): As the terminology implies, all users are allocated orthogonal pilot sequences; hence, the maximum number of users that can be admitted in Stage I is equal to the number of available pilots τ. These τ users are randomly selected from K users.
- No reallocation operations for Case II in Stage I ("NoCaseII"): This algorithm is similar to our proposed algorithm, except when Case II occurs, no additional operations are performed.
- Conventional pilot allocation method ("Con"): This algorithm is similar to the "NoCaseII" algorithm, except when Case I occurs, the users are randomly removed until the number of required pilots is equal to τ.
- Perfect CSI estimation ("Perfect"): This is the baseline algorithm, where the intra-cluster CSI is assumed to be perfectly known.

Figures 3 and 4 illustrate the number of users admitted vs. the candidate RRH set size L in Stage I and Stage II, respectively. It is seen from Fig. 3 that the number of admitted users monotonically decreases upon increasing the candidate set size L. This is due to the fact that with a larger candidate RRH set size for each user, more users will become connected to each other, which requires more pilots. This figure also illustrates the superior performance of our proposed algorithm over

the "Con" algorithm, highlighting the necessity of carefully considering the pilot interference when removing users.

It is interesting to observe from Fig. 4 that the number of users admitted by all algorithms (except the "Ortho" algorithm) initially increases with the candidate set size and then decreases. The reason is that when L starts to increase, a higher spatial degree of freedom is available to support more users. However, when L continues to increase, many users are rejected in Stage I, as seen in Fig. 3. This trend is different from the widely accepted concept that increasing the candidate set size will always lead to better performance. Hence, the channel estimation process should be taken into account when designing the cluster. This figure also shows the performance advantage of our proposed algorithms over the other algorithms.

CONCLUSIONS AND FUTURE RESEARCH CHALLENGES

We have advocated the application of a user-centric dense C-RAN architecture for UDNs due to its appealing features including the facilitation of centralized signal processing, low hardware cost, and so on. However, we have also identified the challenge of requiring a heavy training overhead for estimating all the CSIs for cooperative transmission. As a remedy, we have adopted the partial CSI model, where only the large-scale inter-cluster CSI is available. The channel estimation required for intra-cluster CSI has also been considered, where a novel pilot allocation scheme has been proposed. We have developed a robust transmission design by considering the effect of channel estimation errors. Our simulation results verify the performance advantages of the proposed algorithm over the existing ones.

Finally, we now highlight two promising research directions to make the user-centric dense C-RAN more amenable for practical implementations.

Dynamic Cluster Formations: In this work, the cluster for each user was assumed to be fixed; that is, each user is only connected to its nearest L RRHs. However, in practical systems, the cluster sizes for users should be adapted to the network state, such as the users' rate requirements or traffic load. Additionally, the channel estimation stage should be taken into account, as seen in Fig. 4, where the network performance may even degrade with the cluster size. How to optimize the cluster size individually for each user by jointly considering the above elements remains an inspiring research direction.

User Mobility Management: User mobility is a very challenging issue in user-centric ultra-dense C-RANs. When a user moves from one place to another, the cluster of RRHs assigned for serving this user should be adaptively changed. Explicitly, an adaptive mobility management method should be developed so that the serving cluster can follow each user's behavior (e.g., mobility and service demands) and provide data transmission without the user's involvement. Fortunately, big data, relying on machine learning, is becoming mature, which can track users' mobility and then predict their future locations.
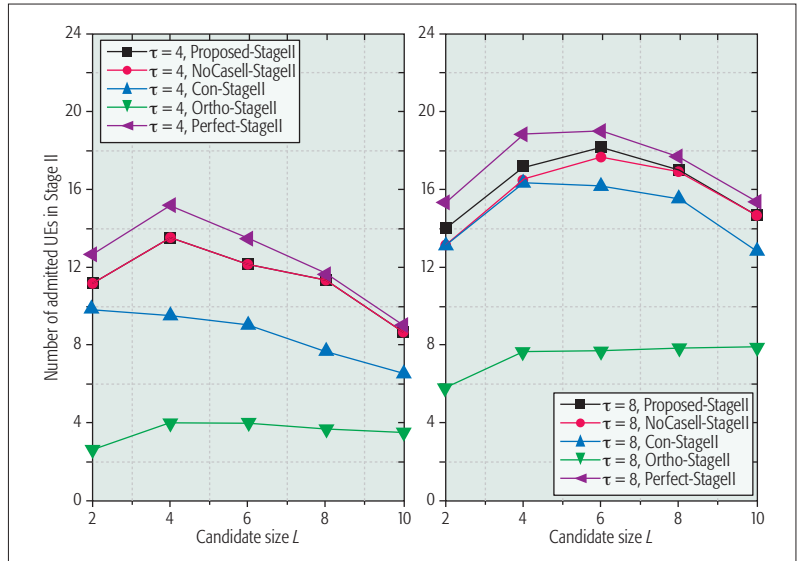

Figure 4. Number of admitted UEs in Stage II vs. the candidate size L. The left subplot corresponds to the case when the number of available pilots is $\tau = 4$, while the right one is $\tau = 8$.

By applying this, the users' serving cluster can be formed beforehand, which significantly reduces the processing time and meets the targeted quality of experience levels.

REFERENCES

[1] J. Andrews et al., "What Will 5G Be?," IEEE JSAC, vol. 32, no. 6, June 2014, pp. 1065–82.
[2] M. Ding et al., "Performance Impact of LoS and NLoS Transmissions in Dense Cellular Networks," IEEE Trans. Wireless Commun., vol. 15, no. 3, Mar. 2016, pp. 2365–80.
[3] J. Tang et al., "System Cost Minimization in Cloud RAN with Limited Fronthaul Capacity," IEEE Trans. Wireless Commun., vol. 16, no. 5, May 2017, pp. 3371–84.
[4] S. H. Park et al., "Joint Precoding and Multivariate Backhaul Compression for the Downlink of Cloud Radio Access Networks," IEEE Trans. Signal Processing, vol. 61, no. 22, Nov. 2013, pp. 5646–58.
[5] B. Dai and W. Yu, "Sparse Beamforming and User-Centric Clustering for Downlink Cloud Radio Access Network," IEEE Access, vol. 2, Oct. 2014, pp. 1326–39.
[6] G. Caire, S. A. Ramprashad, and H. C. Papadopoulos, "Rethinking Network MIMO: Cost of CSIT, Performance Analysis, and Architecture Comparisons," Info. Theory and Applications Wksp., 2010, pp. 1–10.
[7] Y. Shi, J. Zhang, and K. B. Letaief, "CSI Overhead Reduction with Stochastic Beamforming for Cloud Radio Access Networks," 2014 IEEE ICC, 2014, pp. 5154–59.
[8] C. Pan et al., "Joint User Selection and Energy Minimization for Ultra-Dense Multi-Channel C-RAN with Incomplete CSI," IEEE JSAC, vol. 35, no. 8, Aug. 2017, pp. 1809–24.
[9] H. Yin et al., "A Coordinated Approach to Channel Estimation in Large-Scale Multiple-Antenna Systems," IEEE JSAC, vol. 31, no. 2, Feb. 2013, pp. 264–73.
[10] Z. Chen, X. Hou, and C. Yang, "Training Resource Allocation for User-Centric Base Station Cooperation Networks," IEEE Trans. Vehic. Tech., vol. 65, no. 4, Apr. 2016, pp. 2729–35.
[11] G. R1-074068, "Summary of Reflector Discussions on E-UTRA UL RS," 3GPP TSG RAN WG1 Meeting 50bis, Oct. 2007.
[12] Q. Shi et al., "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," IEEE Trans. Signal Processing, vol. 59, no. 9, Sept. 2011, pp. 4331–40.

[13] X. Ge *et al.*, "5G Ultra-Dense Cellular Networks," *IEEE Wireless Commun.*, vol. 23, no. 1, Feb. 2016, pp. 72–79.

BIOGRAPHIES

CUNHUA PAN [M'15] (c.pan@qmul.ac.uk) received his B.S. and Ph.D. degrees from Southeast University, Nanjing, China, in 2010 and 2015, respectively. From 2015 to 2016, he worked as a research associate at the University of Kent, United Kingdom. He is currently a senior research fellow with Queen Mary University of London, United Kingdom. His research interests include C-RAN, mmWave communications, NOMA, D2D, massive MIMO, and cloud computing. He was a TPC member of IEEE ICC and GLOBECOM during 2015–2017.

MAGED ELKASHLAN [M'06] (maged.elkashlan@qmul.ac.uk) received his Ph.D. degree in electrical engineering from the University of British Columbia, Canada, in 2006. From 2007 to 2011, he was with the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. In 2011, he joined the School of Electronic Engineering and Computer Science at Queen Mary University of London. He currently serves as an Editor of *IEEE TWC* and *IEEE TVT*. He received the Best Paper Award at IEEE ICC 2016 and 2014, and IEEE VTC-Spring 2013.

JIANGZHOU WANG [F'17] (j.z.wang@kent.ac.uk) is currently head of the School of Engineering and Digital Arts and a professor at the University of Kent, United Kingdom. He received the Best Paper Award from IEEE GLOBECOM 2012. He is the TPC Chair of IEEE ICC 2019. He was the Executive Chair of IEEE ICC 2015 and the TPC Chair of IEEE WCNC 2013. He was an Editor for *IEEE TCOM* and was a Guest Editor for *IEEE JSAC, IEEE Communications Magazine,* and *IEEE Wireless Communications.*

JINHONG YUAN [F'16] (j.yuan@unsw.edu.au) received his B.E. and Ph.D. degrees in electronics engineering in 1991 and 1997, respectively. He is currently a professor with the School of EE&T at the University of New South Wales, Sydney, Australia. He has published two books, three book chapters, over 200 papers, and 40 industrial reports. He has co-authored winners of three Best Paper Awards and one Best Poster Award. He is currently serving as an Associate Editor for *IEEE TCOM*. His current research interests are communication theory and wireless communications.

LAJOS HANZO [F'08] (lh@ecs.soton.ac.uk), FREng, FIET, Fellow of EURASIP, D.Sc., received his degree in electronics in 1976 and his doctorate in 1983. He holds honorary doctorates from the Technical University of Budapest (2009) and the University of Edinburgh (2015). He is a member of the Hungarian Academy of Sciences and a former Editor-in-Chief of IEEE Press. He is a Governor of both IEEE ComSoc and VTS. He has 30,000+ citations.