Full-spectrum cell-free RAN for 6G systems: system design and experimental results

Dongming WANG[1,2], Xiaohu YOU[1,2*], Yongming HUANG[1,2], Wei XU[1,2], Jiamin LI[1,2],
Pengcheng ZHU[1], Yanxiang JIANG[1], Yang CAO[1,2*], Xinjiang XIA[2],
Ziyang ZHANG[1,2*], Qingji JIANG[1], Pan WANG[2], Dongjie LIU[2], Kang ZHENG[1],
Mengting LOU[3], Jing JIN[3], Qixing WANG[3] & Jiangzhou WANG[4]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China;*
[2]*Purple Mountain Laboratories, Nanjing 211100, China;*
[3]*China Mobile Research Institute, Beijing 100053, China;*
[4]*School of Engineering, University of Kent, Canterbury CT2 7NT, UK*

Abstract   With distributed transceivers and scalable cooperative transmission, cell-free massive multiple-input multiple-output (CF-mMIMO) technology eliminates the "cells" of traditional mobile communications, thereby significantly improving the performance of wireless networks. CF-mMIMO is considered as a major technical tool for achieving the key performance indicators of sixth generation (6G) mobile communication systems. In this paper, a scalable implementation architecture for CF-mMIMO is proposed, based on which potential key technologies such as channel information acquisition, transceiver design, dynamic user and access point association, and novel duplexing are introduced. Finally, based on low-cost commercial access points, a CF-mMIMO testbed with 64 transmit and receive antennas is developed, and the performance of the distributed receiver design is evaluated.

Keywords   6G, scalable cell-free massive MIMO, cell-free radio access network, prototype system

1   Introduction

With the deployment of commercial fifth generation (5G) mobile communication systems, international organizations and national governments now have plans to research sixth generation (6G) mobile communication systems. The International Telecommunication Union (ITU) Telecommunication Standardization Sector (ITU-T) established the ITU-T 2030 network technology focus group in July 2018 to study application scenarios and requirements, network services and technologies, architecture, and infrastructure for 2030 and beyond [1]. The ITU Radiocommunication Sector (ITU-R) started its research work on future technology trends in February 2020. In 2019, the European Union Horizon program launched a 6G technology development project, with main tasks including terahertz communication and new coding and modulation technologies. The University of Oulu in Finland initiated a 6G Flagship program and published a white paper exploring key metrics of 6G and key technical issues facing wireless transmission in 2018 [2]. The China Ministry of Science and Technology launched a kick-off meeting for 6G technology research and development in November 2019, and several universities and enterprises in China have also started 6G research, among which Purple Mountain Laboratories took the lead in laying out 6G technology research at its establishment in 2018 [3].

Currently, although there is no common definition of 6G as of yet, an initial consensus has been established on the related application scenarios, technology trends, and key indicators. In terms of application scenarios, 6G will be deeply integrated with all industries, and the application scope will be

* Corresponding author (email: xhyu@seu.edu.cn, caoyang1@pmlabs.com.cn, ziyangzhang@seu.edu.cn)

extended from the current human-to-human communication to the formation of ultradense intelligent connections for cooperative human-machine-object communication. In terms of technology trends, 6G will make full use of new spectral resources, such as millimeter-wave (mmWave) and terahertz resources, to improve the peak data rate and use ultramassive antenna arrays to improve spectral efficiency, energy efficiency, and the number of connections. 6G will integrate satellite mobile communications to achieve space-air-ground-sea information networks and will be deeply integrated with big data and artificial intelligence to support pervasive intelligent wireless communications. Regarding key technical indicators, 6G will support higher peak rates of up to 1 terabit per second (1 Tbps), a spectral efficiency of up to 1000 bits per second per hertz (Kbps/Hz), a number of connections of up to 100 per square meters, a submillisecond delay of less than 400 μs, and high reliability of greater than 99.99999% [4].

Wireless transmission is a critical technology to achieve the expected capabilities of future generations of mobile communication systems. For 5G systems, although the spectrum has been extended from below 6 GHz to the mmWave region, the problem of spectrum resource shortage is still prominent, and the spectral efficiency should be further improved [5]. Multiantenna technology and dense networking are widely used as the main techniques to improve spectral efficiency, and the typical number of transmitting antennas of a base station has increased from four for fourth generation (4G) applications to 64 or even 128 for 5G applications [6]. Simultaneously, the cell radius has shrunk from macrocells and microcells to the scale of picocells. However, because of the implementation problems encountered with an increasing number of antennas for a centralized system and the interference problems encountered with cell splitting, the spectral efficiency of 5G systems is not sustainable. Therefore, further innovations in the architecture of cellular networks are needed.

Shannon's information theory states that increasing the number of parallel channels can linearly increase the channel capacity. The multiple-input multiple-output (MIMO) technique can be used to scale up the spectral efficiency of wireless communication systems by increasing the number of parallel channels in the spatial domain. Point-to-point MIMO technology has been widely used in 4G and 5G systems. In cellular mobile communication systems, increasing the number of antennas at the base station for communication among multiple users in the same time and frequency resources, thereby forming point-to-multipoint multiuser MIMO, can further increase the number of parallel channels and thus increase the total spectral efficiency of the system. Massive MIMO for 5G is a specific application of this technology. The deployment of multiple radio frequency (RF) units brings each user closer to the corresponding RF unit and reduces the path loss, thus forming a multipoint-to-multipoint distributed MIMO system (also known as coordinated multiple points, CoMP), which can also increase the number of parallel channels and the total spectral efficiency.

However, these three methods inevitably face problems of intercell interference. The intensive deployment of small cells based on point-to-point MIMO technology or a further increase in the number of antennas at the base stations still cannot solve the performance problem for users at cell boundaries [7]. The CoMP technology introduced in 4G is still based on a cellular implementation; consequently, the advantages of CoMP are not fully exploited due to the limited capacity of intercell interactions and the limited number of coordinated points. So far, coordinated transmission with joint processing has not been implemented in commercial base stations for 4G and 5G.

When all access points (APs) deployed in a system collaborate to serve a fixed number of users, as the number of APs increases, the cell boundary effect of traditional cells can be eliminated. As a result, a "cell-free" system is formed, and the capacity and performance of the system can be significantly improved. In this scenario, the key problem is how to address the complexity and scalability of massively coordinated MIMO. For joint coordinated uplink reception, local coherent detection together with centralized combining is used, and for joint downlink transmission, coherent local precoding is used to form a distributed multiuser coordinated transmitter. Such a distributed transceiver scheme with cooperative transmission can not only reduce the complexity of massively coordinated MIMO but also allow for sustainable scaling of users and collaborating APs. Finally, it is possible to break away from the traditional cellular structure to form a cell-free massive MIMO (CF-mMIMO) system [8,9].

CF-mMIMO not only increases the spectral efficiency but also improves the system reliability with diversity and multiplexing trade-offs, thus supporting ultrareliable low-latency transmission. Extending CF-mMIMO to high-frequency bands can increase the system capacity and effectively solve the link robustness problem. Therefore, CF-mMIMO is considered as a key technology for 6G.

CF-mMIMO has paved the way for "cell-free" mobile communication systems. However, different from centralized massive MIMO, the implementation of CF-mMIMO is more complex, which includes

centralized processing [7], full dynamic cooperative processing [9], and full distributed processing [9]. Full centralized processing can achieve optimal performance, but it is difficult to support scale expansion. The performance of full dynamic cooperative processing can approach the performance of fully centralized cooperative processing on the condition that very complete signaling and data interaction are required, which is difficult to achieve. The implementation of fully distributed processing has low complexity and favourable scalability, however, it requires the number of APs far greater than the number of users and the large overhead of fronthaul transmission. Therefore, it is necessary to study the implementation architecture of the new CF-mMIMO, taking both performance and implementation complexity into account. Nevertheless, how to integrate CF-mMIMO into a 6G radio access network (RAN) and design a cell-free RAN (CF-RAN) remains an important issue.

The contributions of this paper mainly include:

(a) We propose a novel full-spectrum CF-RAN architecture. Based on the reasonable division of the physical layer, a scalable CF-mMIMO with RF unit, edge distributed unit (EDU) and user-centric distributed unit (UCDU) functions formed, in which RF unit completes RF signal receiving and transmitting in each frequency band, EDU implements distributed precoding and receiver, UCDU implements data distribution and combining. The new architecture can achieve a trade-off between the complexity and the performance of cooperative transmission.

(b) We propose the key transmission technologies for full-spectrum CF-RAN, including channel information acquisition, transceiver design, and dynamic resource allocation and finally present the experimental results of a CF-mMIMO prototype system with 64 distributed antennas.

2 A full-spectrum CF-RAN architecture

2.1 Transceiver and implementation architecture of CF-mMIMO

Here, we introduce the implementation architecture of CF-mMIMO in accordance with the uplink and downlink signal models. As shown in Figure 1, it is assumed that there are N APs, each configured with L antennas, and K single-antenna users in the system. The received signal of the nth AP can be expressed as

$$\boldsymbol{y}_{\mathrm{UL},n} = \sum_{k=1}^{K} \boldsymbol{g}_{k,n} s_k + z_n, \tag{1}$$

where $\boldsymbol{g}_{k,n}$ represents the channel from the kth user to the nth AP, s_k represents the transmitted signal of the kth user, z_n represents Gaussian noise, and $\boldsymbol{y}_{\mathrm{UL},n}$ represents the received signal of the nth AP.

Ref. [8] gave a generalized expression for a transceiver using dynamic cooperation clustering (DCC). The detection result of the kth user can be expressed as

$$\hat{s}_k = \sum_{n=1}^{N} \delta_{k,n} \boldsymbol{v}_{k,n}^{\mathrm{H}} \boldsymbol{y}_{\mathrm{UL},n}, \tag{2}$$

where $\delta_{k,n}$ represents the association coefficient between the users and the APs — that is, when the nth AP serves the kth user, $\delta_{k,n} = 1$, and otherwise, $\delta_{k,n} = 0$ and $\boldsymbol{v}_{k,n}$ represents the collective combining vector. For downlink transmission, the received signal of the kth user can be expressed as

$$y_{\mathrm{DL},k} = \sum_{n=1}^{N} \boldsymbol{g}_{k,n}^{\mathrm{H}} \left(\sum_{i=1}^{K} \boldsymbol{w}_{i,n} \delta_{i,n} s_i \right) + z_n. \tag{3}$$

Ref. [9] proved that DCC can achieve the scalable implementation conditions of CF-mMIMO and designed a scalable partial minimum mean squared error (partial MMSE, P-MMSE) receiver, meaning that as $K \to \infty$, the complexity of P-MMSE does not increase with K. We assume that the set of APs serving the kth user is represented by $\mathcal{M}_k$ ($N_k = |\mathcal{M}_k|$) and define

$$\boldsymbol{\Delta}_k = \mathrm{diag}\left(\delta_{k,i_1} \cdots \delta_{k,i_{N_k}} \right) = \boldsymbol{I}_{N_k}, \quad i_1, \ldots, i_{N_k} \in \mathcal{M}_k. \tag{4}$$
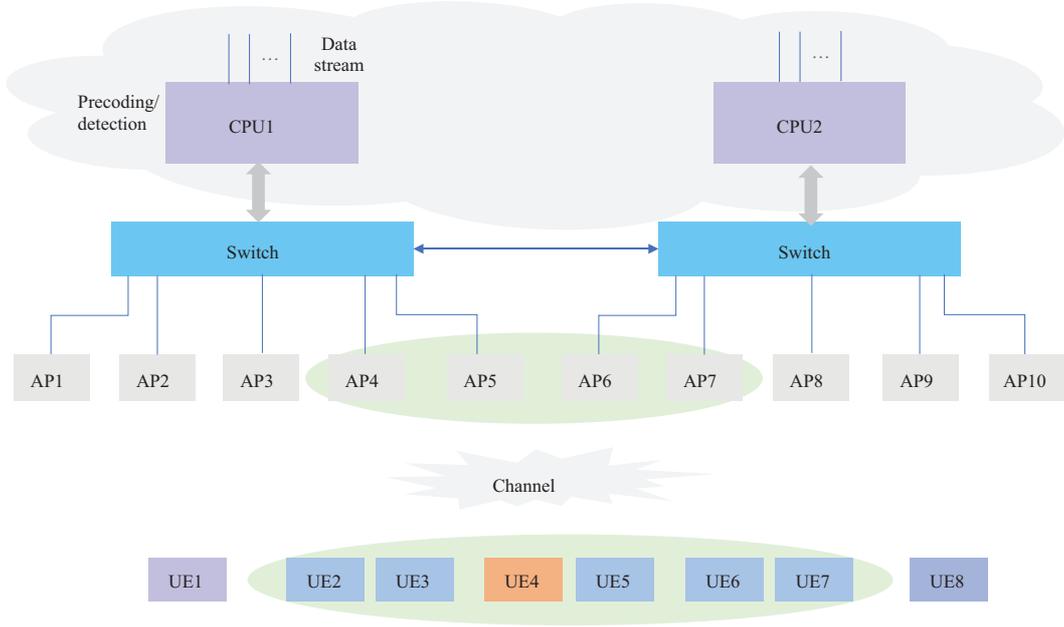
Figure 1 (Color online) DCC-based CF-mMIMO implementation architecture.

The set of other users served by any one of these N_k APs is represented by $\mathcal{P}_k$, wherein the association matrix of the ith user is $\boldsymbol{\Delta}_i$, $\boldsymbol{\Delta}_i \neq \mathbf{0}_{N_k}$. The P-MMSE receiver can be represented as

$$\boldsymbol{v}_k = \left[\boldsymbol{v}_{k,i_1}^{\mathrm{T}} \cdots \boldsymbol{v}_{k,i_{N_k}}^{\mathrm{T}}\right]^{\mathrm{T}} = p_k \left[\sum_{i \in \mathcal{P}_k} p_i \left(\boldsymbol{\Delta}_i \otimes \boldsymbol{I}_L\right) \breve{\boldsymbol{g}}_i \breve{\boldsymbol{g}}_i^{\mathrm{H}} \left(\boldsymbol{\Delta}_i \otimes \boldsymbol{I}_L\right) + \sigma_{\mathrm{UL}}^2 \boldsymbol{I}_{N_k L}\right]^{-1} \breve{\boldsymbol{g}}_k, \tag{5}$$

where $\breve{\boldsymbol{g}}_i = [\boldsymbol{g}_{i,i_1}^{\mathrm{T}} \cdots \boldsymbol{g}_{i,i_{N_k}}^{\mathrm{T}}]^{\mathrm{T}}$, $p_k = \mathrm{E}(|s_k|^2)$, and σ_{UL}^2 represents the variance of the uplink noise. In accordance with the reciprocity of the uplink and downlink channels, the downlink precoding can be obtained from the uplink combining vector [9].

DCC can be applied to realize user-centric networking. As shown in Figure 1, taking UE_4 as an example, the set of APs serving UE_4 includes AP_4 to AP_7, and these APs also partially serve UE_2 to UE_7 at the same time. When the set of APs serving users is managed by multiple central processing units (CPUs), these APs need to be able to dynamically distribute received signals to multiple CPUs, and the P-MMSE technique must be realized on the CPUs. On the one hand, this will introduce more fronthaul overhead; on the other hand, it will introduce certain difficulties for practical network deployment and implementation.

A fully distributed transceiver scheme is the simplest way to achieve the scalability of CF-mMIMO. In such a scheme, each AP performs precoding and detection independently, and each AP knows only local channel information. As shown in Figure 2, taking UE_4 as an example, the set of APs serving UE_4 includes AP_4 to AP_7, and these APs perform multiuser precoding and detection locally. The detection results for UE_4 from AP_4 to AP_7 are sent to CPU_1, which combines the detection results for UE_4 and then performs demodulation and decoding. The implementation of maximum ratio combining/transmitting (MRC/MRT) and local P-MMSE proposed in [9] can also be fully distributed. Take the uplink MRC receiver as an example, $\boldsymbol{v}_{k,n} = \boldsymbol{g}_{k,n}$. According to the simulation results in [9], the performance of the P-MMSE scheme is significantly better than that of the distributed transceiver scheme.

The fully distributed transceiver scheme can be realized in a serial manner, as in the concept of radio stripes proposed in [10]. Ref. [11] further expanded this serial method, in which not only data are transmitted serially but also the channel information of the last AP is transmitted to the next AP, thereby achieving a fully centralized realization.

2.2 An enhanced architecture for the practical implementation of CF-mMIMO

As seen from the above discussion of transceiver implementation for CF-mMIMO, in general, the baseband processing mainly involves two functions: distributed detection/precoding and combining/distribution.
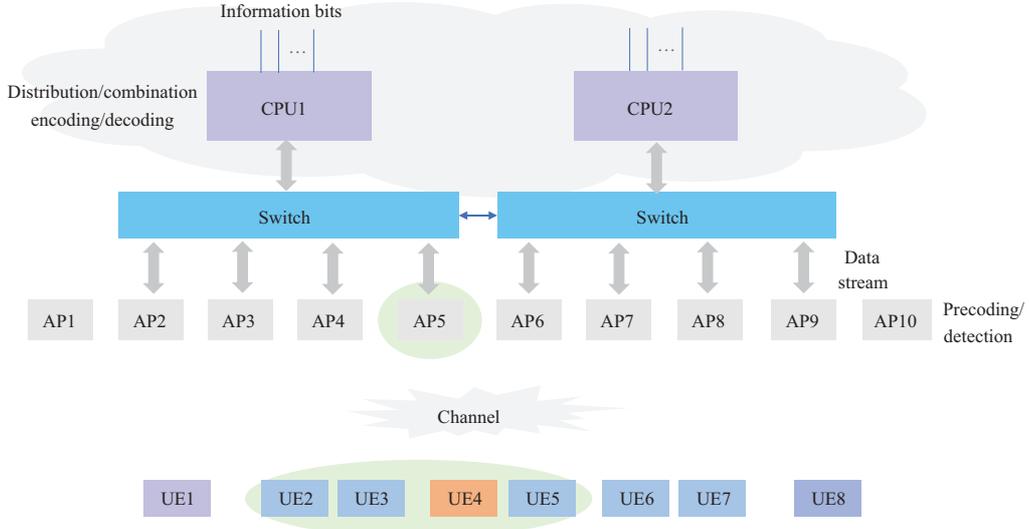
Figure 2 (Color online) Fully distributed CF-mMIMO implementation architecture.
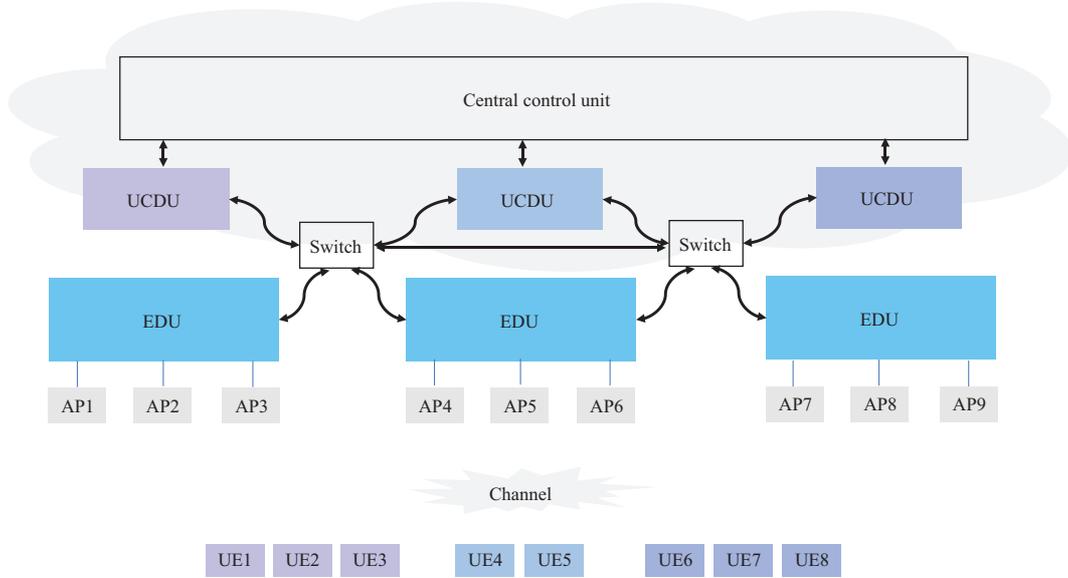

Figure 3 (Color online) A novel CF-mMIMO architecture.

Considering the cooperative processing capability, we can connect the APs in Figure 2 to a dedicated EDU, where detection and precoding are performed. In this way, we can improve the joint processing capability with more cooperative APs connected to the EDU while maintaining scalability. Based on this idea, we can develop a CF-RAN as shown in Figure 3.

Taking uplink reception as an example, after separating the data streams for multiple users, the EDUs send the streams to a UCDU. A user can be associated with multiple EDUs but with only one UCDU. In the UCDU, data streams for the same user sent from different EDUs can be combined. For downlink transmission, the user data are distributed to the associated EDUs via the UCDUs, and local multiuser precoding is performed in the EDUs.

Under this new architecture, we can also implement DCC. For example, when a user is served by APs managed by multiple EDUs, the P-MMSE scheme is implemented in each EDU, and the detection results for the user are combined in their corresponding UCDUs. From this point of view, the new implementation architecture is different from DCC. Compared with a traditional fully distributed cell-free system, the new architecture has stronger joint processing capabilities.

Theoretically, the new architecture described above is a form of implementation of (2) and (3). Suppose that the set of users associated with the mth EDU is $\mathcal{M}_m$, the EDU serves all users in this set, and each
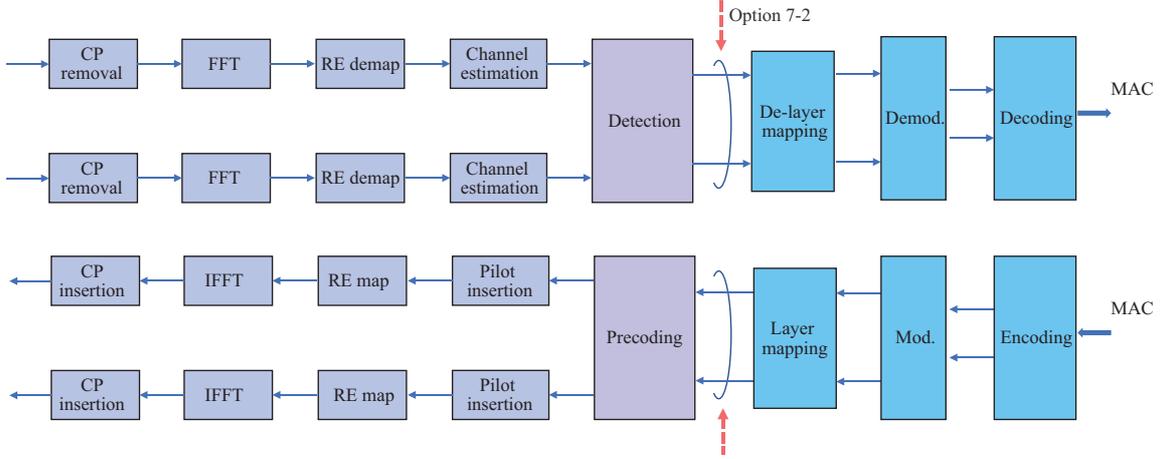
Figure 4   (Color online) Physical-layer signal processing and splitting Option 7-2 of 5G new radio (NR).

EDU manages J APs. The set of EDUs serving the kth user is represented by $\mathcal{J}_k$. The P-MMSE receiver scheme can be represented as

$$
\boldsymbol{v}_{\mathrm{EDU},m,k} = p_k \left[\sum_{i \in \mathcal{M}_m} p_i \hat{\boldsymbol{g}}_{i,m} \hat{\boldsymbol{g}}_{i,m}^{\mathrm{H}} + \sigma_{\mathrm{UL}}^2 \boldsymbol{I}_{JL} \right]^{-1} \hat{\boldsymbol{g}}_{k,m},
\tag{6}
$$

where $\hat{\boldsymbol{g}}_{i,m}$ represents the uplink channel between the mth EDU and the ith user ($i \in \mathcal{M}_m$). The detected signal of the kth user can be expressed as

$$
\hat{s}_k = \sum_{m \in \mathcal{J}_k} \hat{\delta}_{k,m} \boldsymbol{v}_{\mathrm{EDU},m,k}^{\mathrm{H}} \boldsymbol{y}_{\mathrm{EDU},m},
\tag{7}
$$

where $\hat{\delta}_{k,m} = 1$ when the mth EDU serves the kth user and $\hat{\delta}_{k,m} = 0$ otherwise. The expression for downlink precoding can be obtained similarly.

2.3   A full-spectrum CF-RAN

Based on the new scalable CF-mMIMO implementation architecture proposed in Subsection 2.2, this paper further develops a full-spectrum CF-RAN. Regarding the implementation architecture for an RAN, the 3rd Generation Partnership Project (3GPP) has defined eight split options [12] for the fronthaul network, including Options 4 to 8. Moreover, the xRAN Forum has specified more detailed options for Option 7 [13]. Figure 4 shows the implementation of Option 7-2, which has been applied to 5G massive MIMO base stations [13]. It uses hardware to efficiently implement computation-intensive functions such as precoding and detection, thus reducing the fronthaul throughput.

When fully distributed CF-mMIMO, as shown in Figure 2, is implemented based on Option 7-2, since each AP serves multiple users, the fronthaul throughput will increase significantly. For the CF-mMIMO architecture shown in Figure 3, based on Split Option 7-2, the low-level physical layer (low PHY) can be implemented in the EDU, and the high-level physical layer (high PHY) can be implemented in the cloud-based UCDU. The split between the EDU and UCDU lies at the data stream level. The spatial-domain data streams are transmitted in the same time-frequency resources or in orthogonal time-frequency resources. For uplink reception, the term 'data stream' refers to the output of a detector for multiple users, while for downlink transmission, it refers to the data symbols from multiple users.

Here, we introduce the traditional Option7-2 into scalable CF-mMIMO, developing a novel architecture for a user-centric CF-RAN. Different from the traditional RAN implementation, in this new architecture, a user's data stream is associated with an AP, an EDU and a UCDU, and multiple EDUs and their connected APs can process the uplink or downlink data streams of the same user. For uplink, the EDUs can only send the detection results for the same user (such as each term in the sum in (7)) to a single UCDU, where the combining in (7) and the subsequent high PHY processing are implemented. Similarly, the downlink data streams for the same user can only be sent from a single UCDU to multiple EDUs, and the distributed formation of downlink coherent joint transmission (CJT) among multiple EDUs is

implemented. The data exchange between the EDUs and UCDUs can be realized through the enhanced common public radio interface (eCPRI) standard.

Since the EDUs can perform channel estimation, precoding, detection and other functions independently, mmWave-related technologies, such as hybrid precoding, can be realized in the EDUs. High PHY functions that are independent of the frequency bands, such as modulation and demodulation, coding and decoding, and some medium access control (MAC)-layer functions, can be implemented in the UCDUs. Therefore, under the new architecture, it is feasible to integrate high-frequency and low-frequency systems. With a cloud-based implementation, the scalability of the UCDUs is also not a problem. Therefore, we call our proposed design a full-spectrum CF-RAN.

The full-spectrum CF-RAN can be viewed as integrating "computing power" and "networking". With the capabilities of the fronthaul network, distributed computation for the joint processing of CF-mMIMO, also known as decentralized collaboration, can be efficiently implemented, and a better compromise between performance and complexity can be achieved.

3 Key technologies for wireless transmission in a CF-RAN

To realize user-centric distributed cooperative transmission, the wireless transmission technology of the physical layer needs to be enhanced. On the one hand, channel information acquisition has always been a key challenge in CF-mMIMO systems. For distributed deployment, time-frequency synchronization and the acquisition of channel information by the transmitter are particularly important; different from the local detection in traditional CF-mMIMO, due to the large numbers of antennas and users associated with the EDUs, it is also necessary to study precoding and detection technologies to achieve better performance.

3.1 Channel information acquisition

3.1.1 *Reciprocal calibration and time-frequency synchronization among multiple APs*

CJT for CF-mMIMO relies on the acquisition of downlink channel state information (CSI). To avoid CSI feedback, it is usually assumed that the system operates with time-division multiplexing (TDD) and that the downlink CSI can be obtained from the uplink channel by using the reciprocity between the uplink and downlink channels [14, 15]. However, in a practical system, the overall reciprocity of the uplink and downlink channels is also impacted by the following factors. On the one hand, differences between the transceiver circuits can make the channels nonreciprocal. On the other hand, imperfect time and carrier frequency synchronization between APs also introduces nonreciprocity of the downlink and uplink channels. With distributed deployment, these two problems become coupled, making the implementation of reciprocity-based coherent transmission extremely challenging.

The RF parameters of transmit and receive (TR) modules usually change on a time scale of minutes. The time and frequency synchronization between APs is related to system implementation. When the APs in CF-mMIMO have a common local oscillator (LO) and the carrier frequency is synchronized, the calibration coefficient is usually stable, and reciprocal calibration is relatively simple. When the APs have a common reference clock but use independent LOs, the carrier phase of an AP will change with its LO. The phase drift of an LO is related to the implementation of its phase-locked loop (PLL). When the APs share neither a common reference clock nor an LO, each AP usually recovers its clock signal from the global positioning system (GPS), and there will be a frequency offset between APs. A more difficult issue is that this frequency offset drifts as the reference clock is repeatedly recovered from GPS.

In current commercial 5G deployment, the baseband unit (BBU) usually recovers the clock and precise timing from GPS. The precision time protocol (PTP, IEEE 1588v2) and synchronous Ethernet (SyncE) are used between the AP and the BBU to obtain accurate time and reference clock signals. Therefore, the configuration with a common reference clock and independent LOs is widely adopted. The results in [16] showed that in this case, the calibration coefficient changes with the phase drift of the LO's PLL. The phase of the calibration coefficient of an AP varies within $\pm 30°$ with respect to its mean [16]. The calibration coefficients of the antennas of an AP are relatively consistent, that is, there is only a fixed phase difference between each pair of channels. However, since the APs use independent LOs, the calibration coefficients between APs change rapidly in accordance with the LO phase drift, and this change can exceed $20°$ in 7.5 ms.

h_0

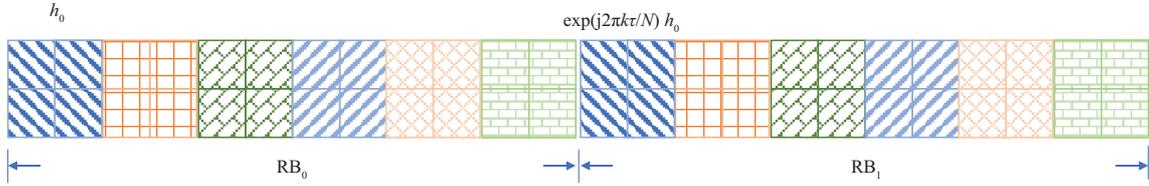$\exp(j2\pi k\tau/N)\, h_0$

RB$_0$ RB$_1$

Figure 5 (Color online) A pattern of a 24-port orthogonal DMRS.

The experimental results of [16] also showed that when joint precoding of multiple APs is adopted, the cooperative transmission performance is highly sensitive to the accuracy of the phase synchronization between APs. However, when the total number of data streams is fewer than the number of antennas of an AP (we assume that each AP has an equal number of antennas) and local precoding is employed by each AP, the performance is not sensitive to the phase drift.

To achieve accurate phase synchronization, the time interval of the calibration signals should be sufficiently small; therefore, the calibration overhead is high. In addition, if there is no common reference clock among the APs, to estimate the frequency offset, the time interval of the pilot signals should also be sufficiently small with respect to the maximum frequency offset. As pointed out in [16], we can use a larger subcarrier spacing (or a shorter orthogonal frequency-division multiplexing (OFDM) symbol duration) for this purpose; therefore, calibration/synchronization can be performed within a guard period. In addition, for distributed precoding, when the total number of data streams is small, the calibration overhead is almost negligible.

Finally, considering that there are a large number of APs in CF-mMIMO systems, to achieve dynamic cooperation, it is necessary to calibrate all APs. To achieve fast calibration, the APs can first be clustered; then, intracluster calibration is performed, followed by relative calibration between clusters. To further reduce the calibration time overhead, the APs in a cluster can be further divided into two groups, between which orthogonal time-frequency calibration signals are sent. Considering that there are many APs and the number of orthogonal calibration reference signals available is limited, for practical implementation, the allocation of the orthogonal pilot sequences can be designed in accordance with the deployment of the APs. This layered approach can achieve better calibration accuracy with a lower calibration overhead. The detailed design and performance evaluation can be found in [16].

3.1.2 *Pilot reuse and channel estimation*

In a CF-mMIMO system, when the total number of antennas is sufficiently large, due to the effect of channel hardening, it is theoretically necessary to estimate only large-scale fading parameters for demodulation. However, the number of APs and the total number of antennas in a practical system are usually not so large; thus, we still need to design demodulation reference signals (DMRSs).

To obtain accurate CSI, orthogonal DMRSs are designed in the 5G NR standard to support up to 12 data streams. To avoid pilot contamination, more orthogonal pilot resources need to be supported for CF-mMIMO, with its large amount of spatial multiplexing. Currently, a working group of 3GPP is studying 24-port orthogonal pilots. With a large number of DMRS ports, pilot reuse and channel estimation can be very simple. First, users who are far away from each other use the same DMRS port. Then, we further establish associations between the users and APs. Users assigned to the same DMRS port may not also be associated with the same AP. After DMRS port assignment and user-AP association, a simple channel estimation method is for each AP to estimate only the channels from the users associated with it and regard other users as interference (simply setting the other users' channels to null).

We can also use the statistics of the channels to further reduce pilot contamination. A 24-port orthogonal DMRS with two OFDM symbols is shown in Figure 5. Every four resource elements (REs) use scrambling code and orthogonal cover code (OCC) to form a code-division multiplexing (CDM) group with four ports. It is assumed that the coherent bandwidth is larger than two resource blocks (RBs), and the timing advance of an antenna port is assumed to be τ. After the bundling of two RBs, the channel parameters of the eight corresponding ports can be estimated. The received signals in a CDM group for two RBs can be expressed as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{n},$$

where

$$
X = \left[
\begin{array}{cc}
X_1 W & X_2 W \\
X_3 W T_1 & X_4 W T_2
\end{array}
\right],
$$

X_1 and X_2 are the scrambling matrices of the CDM group in RB_0, X_3 and X_4 are the scrambling matrices of the CDM group in RB_1, W represents the OCC matrix, and T_0 and T_1 represent the phase rotation matrices introduced by the propagation delays of the user equipment(UE) within the same CDM group, both of which are diagonal matrices. Using the large-scale statistical properties of the channels, we can obtain frequency-domain channel estimates for eight ports based on MMSE estimation [17]. Then, in turn, 48 ports can be supported for the DMRS pattern in Figure 5.

3.1.3 *Joint phase noise and channel information estimation technique*

Different from low-frequency systems, high-frequency systems are susceptible to phase noise. The current 5G NR standard adopts a phase tracking reference signal (PTRS) for phase noise estimation. Considering the overhead, only a single-port PTRS is supported in 5G NR. When multiple beams are used, there may be interference between the PTRS and the data. In addition, when multiple UEs communicate with multiple APs, the phase noise of each channel is different, and the estimation of multiport phase noise should be studied.

Therefore, to realize CF-mMIMO in the mmWave frequency band, it is necessary to further enhance the PTRS and improve the accuracy of phase noise estimation. On the one hand, the design of a multiport PTRS needs to be further enhanced. On the other hand, when the PTRS and data are multiplexed, the estimation accuracy for the phase noise should be improved considering the interference among multiple beams. Finally, a joint phase noise estimation technique, such as the tensor decomposition method in [18], can be used to estimate the multichannel phase noise.

3.2 Downlink precoding and uplink detection

A key feature of CF-mMIMO is the scalability of uplink multiuser detection and downlink multiuser precoding. For a CF-RAN, downlink multiuser precoding and uplink multiuser detection can be implemented in the EDUs. Thus, under the new decentralized cooperative architecture, the precoding and detection methods for a CF-RAN should be studied.

3.2.1 *Downlink multiuser precoding*

Considering the problem of complexity, non-codebook multi-RBs linear precoding is still the most important form of precoding implementation. Robust precoding for CF-mMIMO should be studied while considering the statistics of distributed MIMO channels.

On the one hand, in a CF-mMIMO system, the delays between different users and different APs will be different, which will introduce severe frequency selectivity. Therefore, when performing the precoding calculations, it is necessary to consider the effects of these delay differences on the performance of multi-RBs precoding. Ref. [19] proposed an optimized precoding design using time delay difference information, which can achieve better performance.

On the other hand, calibration/phase synchronization errors will exist among the multiple APs connected to an EDU; therefore, a precoding algorithm that is robust to phase drift should be studied. Considering that the phase consistency within each AP is good and the phase drift between APs varies within $\pm\theta_{\max}$ around the mean value ($\theta_{\max} = \pi/6$ in [16]), we can take advantage of these statistical characteristics to optimize the precoder to improve its robustness [20].

3.2.2 *Detection method*

As shown in (6), after multiuser detection in the EDU, the detection results and the signal-to-interference-plus-noise ratio (SINR) of each data stream are quantized and sent to the CPU. To reduce the fronthaul overhead, data compression can be used. Another option is to execute soft-output demodulation in the EDU. In the CPU, which combines the results for the same data stream from different EDUs, the high PHY function is performed.

Since the detection and combining processes are separated in the distributed receiver scheme, it is difficult to implement turbo receivers. To further improve the multiuser detection performance of the EDUs, advanced receivers can also be considered, such as soft interference cancellation receivers [21].

3.3 Dynamic association and flexible duplexing for a CF-RAN

Radio resource allocation is also very important for a CF-RAN. There have been many studies on various types of radio resource allocation for CF-mMIMO, such as power allocation or power control [22], user association [23], and energy efficiency optimization [24]. In this paper, we focus on user association and flexible duplexing in a CF-RAN.

3.3.1 *Dynamic UE-AP (EDU) and UCDU association*

In a CF-RAN, cooperative transmission should support large numbers of UEs and APs. From a UE's perspective, the associated APs change dynamically as the UE moves. Meanwhile, from a CPU's point of view, each AP is dynamically associated with multiple UEs. Since the capabilities of the EDUs are limited, to make the system scalable, each EDU can serve only those UEs that are sufficiently close to the APs connected to that EDU. In a CF-RAN, once the association of the UEs and APs has been determined, the association of the UEs and EDUs is also determined. Another issue is the association of the UEs and UCDUs for cloud implementation.

The association between the UEs and APs can be established in the process of random access. During data transmission, this association can be further improved with uplink sounding. The random access process in a CF-RAN is different from that in the existing 5G NR standard. Specifically, it should support large-scale coverage and decentralized access, allowing the establishment of an initial association between the UEs and APs. In the massive MIMO and mmWave systems of 5G NR, the association between the UEs and beams is realized by setting one beam corresponding to each random access channel (RACH) instance.

In the random access process, based on the physical RACH (PRACH) signals received by multiple APs, the initial association is established in accordance with the statistical information of the signals. More specifically, the received signal strength indicator (RSSI), time of arrival (TA) and other information can be obtained based on the PRACH signals that are received by different APs. Thus, the position of a UE can be estimated, and the initial association between the UEs and APs can be determined.

In accordance with the deployment of the APs, the APs are initially clustered into AP groups. Each AP group includes one primary AP and multiple secondary APs. The primary AP is responsible for the transmission of synchronization signal block (SSB) signals and the reception of PRACH signals during the initial access phase of the UEs. The secondary APs are responsible only for receiving the PRACH signals in the initial access phase of the UEs. The primary AP is numbered with an AP index, and this parameter is embedded in the system information block (SIB) of the SSB signal. During data transmission, the APs are regrouped in accordance with the UEs' locations, which can be obtained through uplink sounding, and more APs can be allocated. On the premise that the minimum number requirements are satisfied, the number of secondary APs can be expanded infinitely.

In practical deployment, dynamic load balancing can be implemented through the association between the UCDUs and UEs, and the mobility management should be further studied.

3.3.2 *Network-assisted full duplexing*

Flexible duplexing was introduced as a new feature of 5G NR. With the development and maturity of co-frequency co-time full-duplexing (CCFD) technology, its application in 6G has received further attention. However, in wireless networks with flexible duplexing or CCFD, cross-link interference (CLI) is inevitable [7]. CLI refers to the interference between the downlink transmission of an AP and the uplink reception of another AP as well as the interference between the uplink transmission of a UE and the downlink reception of another UE. The global collaborative capabilities of CF-mMIMO provide important support for more flexible duplexing technology.

CCFD, traditional TDD and frequency-division duplexing can all be harmonized in the CF-RAN shown in Figure 3. The main operating principles are that the uplink and downlink work simultaneously on the same frequency band; the UCDU determines the slot pattern for each AP; when there is interference from the downlink to the uplink, the UCDU sends the interference information to the EDU, and the interference

can then be eliminated in the EDU by utilizing the CSI between the APs; and the interference between the uplink transmission of a UE and the downlink reception of another UE can be reduced through multiuser scheduling in the UCDU. This method of harmonized duplexing in a CF-RAN is also called network-assisted full duplexing (NAFD) [25].

There are still many problems to be solved in NAFD, including the following:

• In a practical 5G NR system, the receiving APs and transmitting APs are not aligned in time due to the advance reception of the uplinks. For standardization, it will be necessary to consider how to solve the asynchronous interference problem.

• Eliminating CLI relies on cooperation between APs, which indicates that centralized signal processing can be beneficial to eliminating interference. The interference elimination capability that can be achieved when a distributed architecture is adopted needs further study.

• For the case in which fully dynamic control of the transmission and reception of the APs is adopted, it will be necessary to study the selection of the transmission or reception mode from a global perspective [26–28] in order to reduce interference and improve the system performance.

4 Prototype CF-RAN system

To demonstrate the performance of the proposed CF-RAN design, we have developed a prototype system. Figure 6 shows a schematic diagram of the full-spectrum CF-RAN platform. The RF unit of the AP in the system can be a low-frequency AP (also called a remote radio unit (RRU) in a 5G system) or a mmWave AP (also called an active antenna unit (AAU) in a 5G system). The APs mainly perform up- and downconversion, analog-to-digital or digital-to-analog conversion, fast Fourier transformation/inverse fast Fourier transformation, cyclic prefix addition and removal, and resource mapping and demapping. For the high-frequency band, the AAU is also capable of analog precoding. In the prototype system, the UEs are implemented with the same hardware as the AP. The baseband processing is implemented in an x86 server, which is configured with a fronthaul card and a forward error correction accelerator card. The fronthaul card provides accurate time and reference clock recovery from GPS for the RF front end. An eCPRI interface is adopted between the fronthaul card and AP, including the control plane, synchronization plane and user plane.

Figure 7 shows the considered CF-RAN scenario. There are 16 RRUs and 12 UEs in the system. Each RRU or UE is equipped with four antennas. The system operates in the 4.9 GHz frequency band and uses low-cost commercial 5G NR RRUs, with a bandwidth of 100 MHz, and the frame pattern follows the 5G NR standard. The coding and modulation of the system comply with the 5G standard, and the modulation and coding scheme (MCS) follows Table 5.1.3.1-1 [29]. To reduce pilot contamination and the complexity of channel estimation, we use a 48-port DMRS with a pattern similar to that in Figure 5 to estimate the channel parameter of a port for every two RBs. In the frequency domain, the channel gains of all subcarriers in a subband can be obtained through Wiener interpolation with the uniform power spectrum of the eight RBs.

For the test scenario shown in Figure 7, each terminal has four data streams, for a total of 48 data streams in the system. Figure 8 shows the offline analysis of the cumulative distribution function (CDF) of the total spectral efficiency using three joint receivers, including joint zero forcing (J-ZF), joint MMSE (J-MMSE) and MMSE-based successive interference cancellation (MMSE-SIC). Considering the channel estimation error [7], the average spectral efficiencies of the three receivers are approximately 319, 323.3, and 382.2 bps/Hz, respectively. In the test scenario shown in Figure 7, when 64-QAM modulation is adopted for each data stream, the receiver adopts MMSE equalization, and the average SINR of each data stream is approximately 20 to 23 dB. Each stream of terminals can reach MCS 28 (LDPC code rate 948/1024). When the overheads of the pilot and cyclic prefix are not considered, the total spectral efficiency reaches 265.7 bps/Hz. When the overheads of the pilot and cyclic prefix are considered, the total spectral efficiency of the system can reach 209.66 bps/Hz.

Next, we present an offline analysis of the performance of CF-mMIMO with the new architecture. Considering that each terminal is equipped with one antenna, the total number of data streams in the system is 12, and the total number of receive antennas is 64. We compare the total spectral efficiencies under four different receiving schemes, namely, local MRC detection in the AP, local MMSE (L-MMSE) detection in the AP, P-MMSE detection in the EDU, and J-MMSE detection. With the total number of antennas being 64, we consider three configurations, i.e., with eight, four, and two EDUs, where each
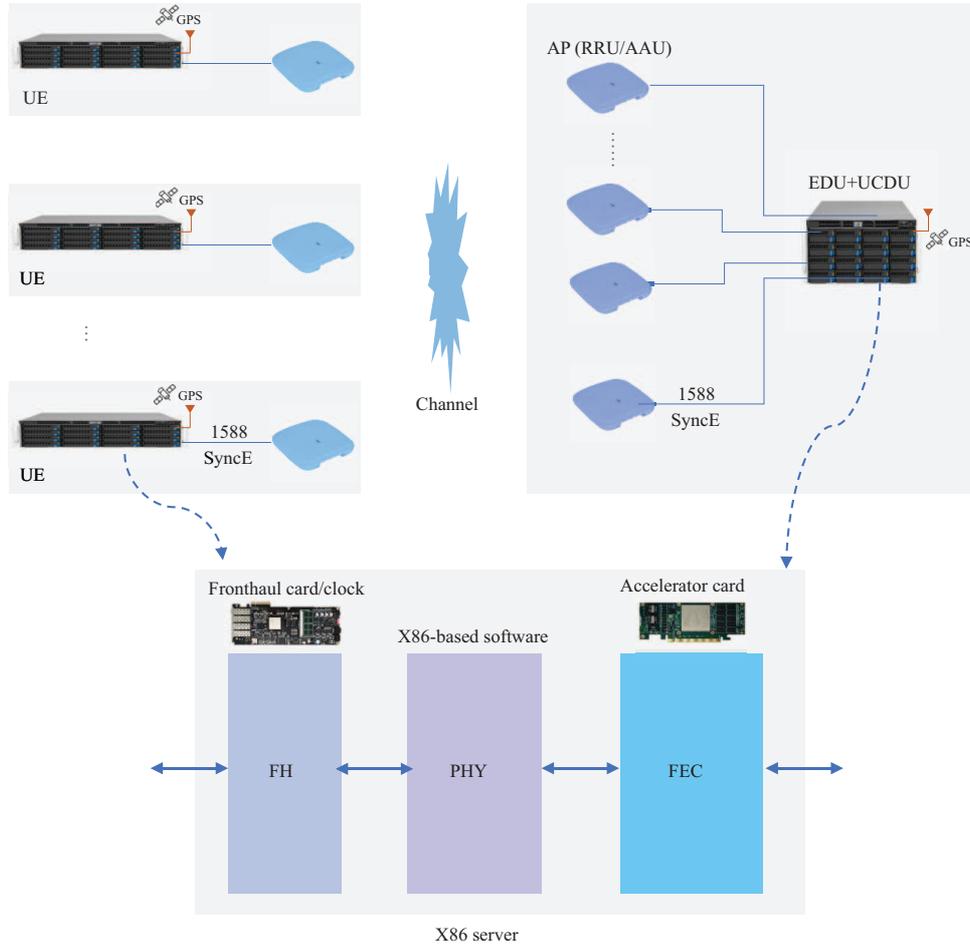
Figure 6 (Color online) Schematic diagram of the prototype full-spectrum CF-RAN system.
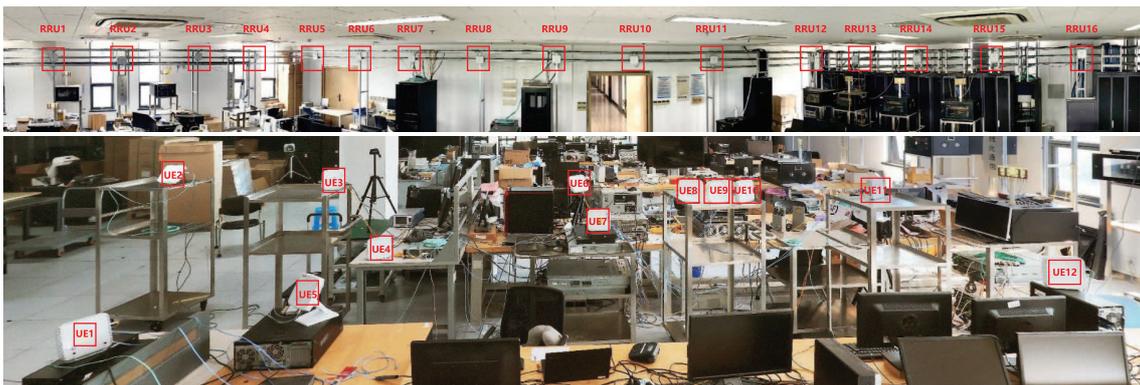

Figure 7 (Color online) Test scenario of a low-frequency CF-RAN.

EDU is configured with 8, 16, and 32 antennas, respectively. Figures 9 and 10 show the CDF of the system spectral efficiency and the average spectral efficiency, respectively. It can be seen from these figures that in the test scenario, the interference between users is strong, and the interference suppression capabilities of the local MRC and L-MMSE schemes are poor. Even so, compared with the local MRC scheme, the L-MMSE scheme still has a 38% performance gain. When the number of antennas configured in each EDU of the system is doubled, the performance improvements compared with the local MRC scheme are 82%, 175%, and 213%, respectively. When the system is configured with two EDUs, each with 32 antennas, the performance can reach 96% of that of the J-MMSE scheme. Because the total number of data streams supported in the system is 12, the improvement in the system performance is
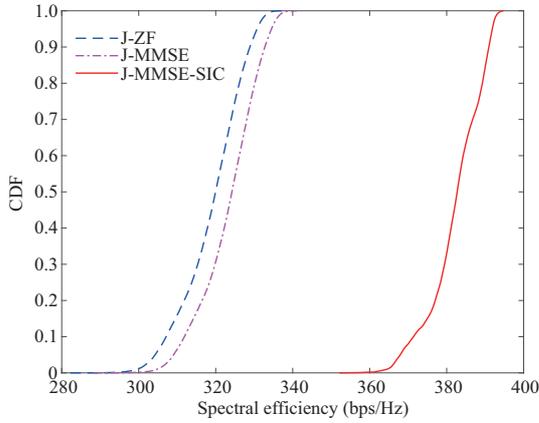
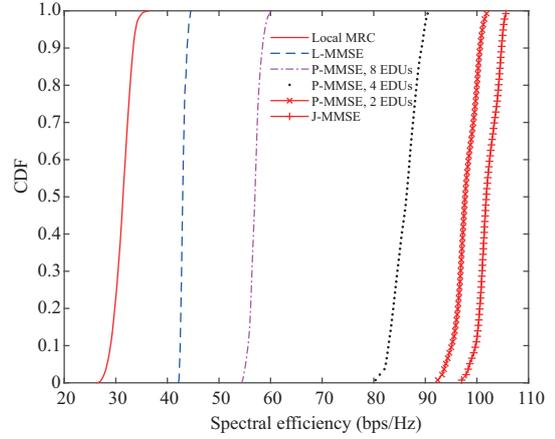**Figure 8** (Color online) Total spectral efficiency of the system with 48 data streams.

Figure 9 (Color online) CDF of the total spectral efficiency of the system with 12 data streams.
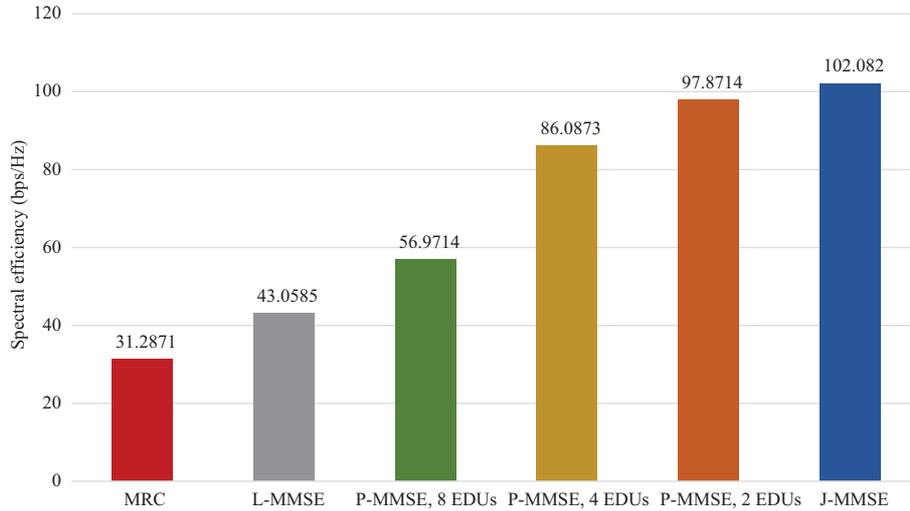

Figure 10 (Color online) Average total spectral efficiency of the system with 12 data streams.

most significant when the total number of antennas per EDU is 16.

5 Conclusion

In this paper, we have studied a novel 6G-oriented cell-free architecture for an RAN that paves the way toward user-centric networking. We first proposed a scalable CF-mMIMO implementation architecture. Accordingly, through the functional splitting of the RAN, we introduce EDUs to achieve a user-centric full-spectrum CF-RAN. We further studied wireless transmission technologies for such a CF-RAN and presented relevant experimental results. In practice, considering the power consumption of the APs, deployment costs and other factors, the ratio of the total number of antennas to the total number of space-division data streams should not be too large. In this case, the results show that compared to the implementation of a distributed transceiver scheme in the APs, the proposed architecture can achieve better performance. In future studies, we will further research and develop our CF-RAN design in the mmWave band to improve the peak rate and robustness of the system.

References

1 ITU-T. Focus group on technologies for network 2030. 2019. https://www.itu.int/en/ITU-T/focusgroups/net2030/Pages/default.aspx

2 6G-Flagship. 6G Visions on Paper. 2019. https://www.6gflagship.com/white-papers/

3 IMT-2030. 6G Typical Scenarios and Key Capabilities White Paper. IMT-2030 (6G) Promotion Group, 2022

4 You X H, Wang C X, Huang J, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. Sci China Inf Sci, 2021, 64: 110301

5 Pang J Y, Wang S B, Tang Z F, et al. A new 5G radio evolution towards 5G-Advanced. Sci China Inf Sci, 2022, 65: 191301

6 Marzetta T L, Larsson E G, Yang H, et al. Fundamentals of Massive MIMO. Cambridge: Cambridge University Press, 2016

7 You X H, Wang D M, Wang J Z. Distributed MIMO and Cell-free Mobile Communication. Berlin: Springer, 2021

8 Ngo H Q, Ashikhmin A, Yang H, et al. Cell-free massive MIMO versus small cells. IEEE Trans Wireless Commun, 2017, 16: 1834–1850

9 Bjornson E, Sanguinetti L. Scalable cell-free massive MIMO systems. IEEE Trans Commun, 2020, 68: 4247–4261

10 Interdonato G, Björnson E, Ngo H Q, et al. Ubiquitous cell-free massive MIMO communications. J Wireless Com Netw, 2019, 2019: 197

11 Shaik Z H, Bjornson E, Larsson E G. MMSE-optimal sequential processing for cell-free massive MIMO with radio stripes. IEEE Trans Commun, 2021, 69: 7775–7789

12 3GPP. Study on new radio access technology: radio access architecture and interfaces. TR.38.801. https://www.3gpp.org/ftp/Specs/archive/38_series/38.801/38801-e00.zip

13 NGMN. 5G RAN CU-DU network architecture, transport options and dimensioning. 2019. https://www.ngmn.org/wp-content/uploads/Publications/2019/190412_NGMN_RANFSX_D2a_v1.0.pdf

14 Wei H, Wang D, Zhu H, et al. Mutual coupling calibration for multiuser massive MIMO systems. IEEE Trans Wireless Commun, 2016, 15: 606–619

15 Rogalin R, Bursalioglu O Y, Papadopoulos H, et al. Scalable synchronization and reciprocity calibration for distributed multiuser MIMO. IEEE Trans Wireless Commun, 2014, 13: 1815–1831

16 Cao Y, Wang P, Zheng K, et al. Experimental performance evaluation of cell-free massive MIMO systems using COTS RRU with OTA reciprocity calibration and phase synchronization. 2022. ArXiv:2208.14048

17 Kong D J, Xia X G, Liu P, et al. MMSE channel estimation for two-port demodulation reference signals in new radio. Sci China Inf Sci, 2021, 64: 169303

18 Sokal B, Gomes P R B, Almeida A L F, et al. Tensor-based receiver for joint channel, data, and phase-noise estimation in MIMO-OFDM systems. IEEE J Sel Top Signal Process, 2021, 15: 803–815

19 Guo Y, Fan Z, Lu A, et al. Downlink transmission and channel estimation for cell-free massive MIMO-OFDM with DSDs. EURASIP J Adv Signal Process, 2022, 2022: 17

20 Han S, Yang C, Wang G, et al. Coordinated multi-point transmission strategies for TDD systems with non-ideal channel reciprocity. IEEE Trans Commun, 2013, 61: 4256–4270

21 Murillo-Fuentes J J, Santos I, Aradillas J C, et al. A low-complexity double EP-based detector for iterative detection and decoding in MIMO. IEEE Trans Commun, 2021, 69: 1538–1547

22 Guenach M, Gorji A A, Bourdoux A. Joint power control and access point scheduling in fronthaul-constrained uplink cell-free massive MIMO systems. IEEE Trans Commun, 2021, 69: 2709–2722

23 Wang J, Wang B, Fang J, et al. Millimeter wave cell-free massive MIMO systems: joint beamforming and AP-user association. IEEE Wireless Commun Lett, 2022, 11: 298–302

24 Mai T C, Ngo H Q, Tran L N. Energy efficiency maximization in large-scale cell-free massive MIMO: a projected gradient approach. IEEE Trans Wireless Commun, 2022, 21: 6357–6371

25 Wang D, Wang M, Zhu P, et al. Performance of network-assisted full-duplex for cell-free massive MIMO. IEEE Trans Commun, 2020, 68: 1464–1478

26 Xia X, Fan Z, Luo W, et al. Joint uplink power control, downlink beamforming, and mode selection for secrecy cell-free massive MIMO with network-assisted full duplexing. IEEE Syst J, 2022, doi: 10.1109/JSYST.2022.3188514

27 Mohammadi M, Vu T T, Beni B N, et al. Virtually full-duplex cell-free massive MIMO with access point mode assignment. In: Proceedings of the 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC), 2022. 1–5

28 Xia X, Zhu P, Li J, et al. Joint optimization of spectral efficiency for cell-free massive MIMO with network-assisted full duplexing. Sci China Inf Sci, 2021, 64: 182311

29 3GPP. Physical layer procedures for data. TS.38.214. https://www.3gpp.org/ftp/Specs/archive/38_series/38.214/38214-h40.zip