

User-Centric Cell-Free (UCCF) Wireless Systems: Principles and Optimization

Lie-Liang Yang

Abstract

User-centric cell-free (UCCF) wireless networks have a range of distinguished characteristics, which can be exploited for meeting some challenges that the conventional cellular systems are hard to. This chapter is devoted to delivering the fundamentals of wireless communications in UCCF systems, including channel modeling and estimation, uplink (UL) detection, downlink (DL) transmission, and resource optimization. Specifically, the advantages of cell-free networking are examined in contrast to the conventional cellular systems. The global and location-aware distributed UL detection are explored in the principles of minimum mean-square error (MMSE) and brief propagation. Correspondingly, the global and distributed DL transmission schemes are designed based on the MMSE precoding. The optimization of both UL and DL is analyzed with respect to system design and resource-allocation. Furthermore, some challenges for the implementation of UCCF systems in practice are identified and analyzed.

Index Terms

User-centric cell-free network, cellular network, optimization, resource optimization, time-division duplex, in-band full duplex, multicarrier-division duplex, channel training, channel estimation, multiuser detection, global detection, location-aware detection, access-point message passing detection, minimum mean-square error (MMSE), precoding, preprocessing, centralized precoding, distributed precoding, physical-layer security.

I. INTRODUCTION

In wireless communications, cellular networking has been a major breakthrough in solving the problems of spectral congestion and user capacity. Conceptually, a cellular network has the structure as shown in Fig. 1. A region covered by a cellular system is divided into cells. Each cell is centred with a base-station (BS) that is responsible for the control and signal transmissions of the users, or user equipments (UEs), within the cell. In cellular systems, each BS is allocated a fraction of the total number of channels available to the entire system, and neighbouring BSs are assigned different groups of channels to decrease the interference between them. Nonetheless, the total available channels are distributed throughout the geographic region and can be reused as many times as necessary. Owing to this spectrum

L.-L. Yang is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK. (E-mail: lly@ecs.soton.ac.uk). This document is a chapter in the book: L.-L. Yang, J. Shi, K.-T. Feng, L.-H. Shen, S.-H. Wu and T.-S. Lee, Resource Optimization in Wireless Communications: Fundamentals, Algorithms and Applications, Academic Press, USA (to be published in 2024).

reuse, a cellular system is capable of offering very high capacity by using a limited amount of spectrum resource. In a given region, a higher capacity can be obtained by dividing the region into more cells of smaller size.

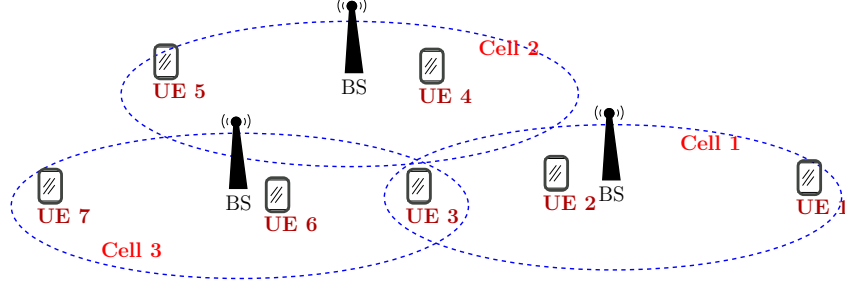


Fig. 1. Illustration of cellular network structure.

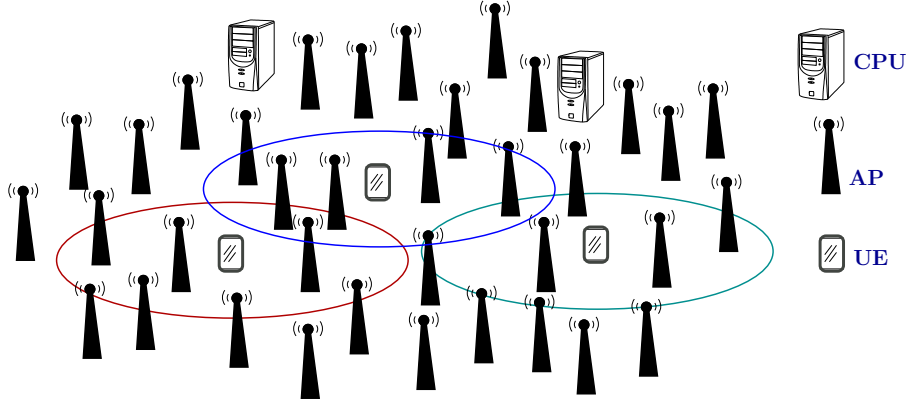


Fig. 2. Illustration of UCCF network structure.

However, when wireless communications enters the generations that a system's performance is not just measured by the performance, such as capacity, reliability, etc., of itself, but mainly measured by the qualities of services (QoS) demanded by users, the conventional cellular networks encounter some deficiencies for providing services in some applications. For example, in cellular systems, the cell-edge users far away from BS, such as UEs 1, 3, 5, 7 in Fig. 1, may have to endure the poor channels and reduce their requirements for QoS. The far away users may also suffer strong interference from the users close to BS, such as UEs 2, 4, 6 in Fig. 1, resulting in the near-far problem. Accordingly, BS may have to limit the desire of the users close to BS for the best possible QoS, so as to protect the cell-edge users. Furthermore, the spectrum reuse with a reuse factor¹ less than one limits the exploitation of the precious frequency resource, reducing the spectral-efficiency. Moreover, spectrum reuse

¹Spectrum reuse factor is defined as the rate at which the same spectrum can be used in a network. A reuse factor $1/N$ means that a cluster of N cells cannot use the same spectrum to transmit signals.

generates inter-cell interference or so-called co-channel interference, making the cell-edge users, such as UE 3 in Fig. 1, vulnerable.

To circumvent the shortcomings of cellular systems presented in some applications, the concept of user-centric cell-free (UCCF) wireless communications came in sight [1–3]. UCCF networks have the structure as shown in Fig. 2. A UCCF network may include one to several central processing units (CPUs), and many distributed access points (APs) that are connected with CPU(s) via fiber optics or dedicated radio resources. APs and CPU(s) work together to serve the UEs in the network. A UCCF network may be a heterogeneous network with the dense deployment of APs. In contrast to the cellular systems where BS is the center of a cell, in UCCF networks, every user is the center of its virtual cell², which sends signals to and receives signals from the APs located in its virtual cell.

In comparison with cellular networks, UCCF networks have their unique characteristics for meeting some challenges that cellular networks are hard to. First, by deploying many distributed APs, instead of the co-located antennas at BS, user (or system) capacity can be significantly increased for given spectrum and antenna (space) resources. Specifically, as shown in [1, 2], a distributed antenna system with user-centric detection is capable of supporting many more users than a corresponding cellular system with co-located antennas at BS. In addition to the capacity benefit brought by distributed APs, in UCCF systems, a spectrum reuse factor of one can be implemented, which also engenders the capacity advantage of UCCF systems over the conventional cellular systems.

In UCCF networks, APs are usually close to UEs. Hence, signals sent by APs to UEs or by UEs to APs do not experience severe propagation pathloss and shadowing. The chance of having line-of-sight (LoS) transmission paths between APs and UEs is high. Therefore, wireless communications in UCCF networks is high energy-efficiency. In UCCF networks, any UE is a strong user in its virtual cell, hence, there is no near-far problem, rendering that low-complexity detection is near-optimum. consequently, in UCCF networks, power-control is mainly QoS dependent, but not for the near-far problem, as done in cellular systems [4, 5].

There is no handover as that in cellular systems needed in UCCF networks. When a UE moves, its virtual cell also moves, some current APs may move out of its virtual cell, while some new APs are added in it. During the process, system can dynamically update the APs serving the UE, according to the real-time signal measurement, or simply, the prediction of the UE's movement. In UCCF networks, the sizes of UEs' virtual cells can be dynamic and variant, being set according to UEs' specific communications environments or/and their QoS requirements. For example, a UE having a higher reliability or/and rate requirement can have a virtual cell of bigger size than a UE having a lower requirement for reliability or/and rate.

UCCF networks are high-robustness. APs can be arbitrary added to a UCCF network to enhance its performance. Some random failures of APs would not result in the catastrophic

²The virtual cell of a user is defined as the certain area around the user.

effect on communications. UCCF networks are feasible for meeting the demands of various types of services, e.g., high-rate, high-reliability, low-latency. The required QoS can be met via setting an appropriate size for the virtual cell, allocating the relevant spectrum and time resources, executing the tailored signal processing, etc.

In UCCF networks, the signal processing for transmission and detection is user location aware. On uplink (UL) transmission, one UE's transmitted signal is only received by a small subset of APs that are around the UE. On downlink (DL) transmission, usually, only a few of APs surrounding a UE are dedicated to transmitting signals to the UE. As above-mentioned, this networking feature brings the advantages of, such as, energy-efficiency, efficiency of resource usage, etc. Moreover, it can be shown that the inputs and outputs of such a UCCF network are related by a sparse graph, which is beneficial to engaging the high-efficiency brief-propagation algorithms for attaining near-optimum performance in signal detection [6].

Additionally, UCCF networking may provide a promising method for the implementation of secrecy communications at physical layer, which will be explained in Section VII.

Naturally, there are challenges to meet in the design, optimization and implementation of UCCF networks. Therefore, this chapter motivates to address the fundamentals of UCCF systems and explain some possible challenges. Specifically, in Section II, the system models for UCCF networks are discussed. Section III-A focuses on the channel modeling and channel estimation, while Section IV provides some UE association methods. In Section V, the UL detection and optimization are analyzed, while Section VI considers the DL transmission and optimization. Finally, Section VII summarizes the chapter and provides some concluding remarks.

II. SYSTEM MODELS FOR UCCF WIRELESS NETWORKS

A UCCF network has the structure as shown in Fig. 2. As above-mentioned, in a UCCF network, APs are geographically distributed in an area, which are connected via fiber optics or dedicated radio resources to one or to multiple CPUs, where network control functions and/or main signal processings are implemented. In a UCCF network, each AP usually only serves a small number of UEs, responsible for their signal receiving and transmission. Each UE is usually associated with one to a few of APs. The association of UEs to APs may be simply based on the physical distances between UEs and APs or other measurements, as that to be discussed in Section IV. Depended on the practical application scenarios, a UE may employ one or several antennas or an antenna array. Similarly, an AP may be equipped with one antenna, multiple antennas or even several antenna arrays. Furthermore, in UCCF networks, various duplex techniques between UL and DL, multiuser multiplexing schemes and signaling methods may be implemented [2, 3, 7–10].

In this chapter, a UCCF network with baseband OFDM signaling is considered to analyze the UL detection and DL preprocessing, as well as their related optimization. Specifically, a UCCF network employing one CPU connected with M APs, which may be randomly or regularly distributed, is assumed to support K UEs. Each of APs and UEs is equipped with

one antenna for receiving or transmission. The number of subcarriers of OFDM is denoted by N , whose indices form a set $\mathcal{N} = \{1, 2, \dots, N\}$. After UE association, the set of UEs associated with AP m is expressed as $\mathcal{K}_m$ for $m = 1, 2, \dots, M$, and the set of APs monitoring the k th UE is expressed as $\mathcal{M}_k$ for $k = 1, 2, \dots, K$. Hence, $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cdots \mathcal{M}_K$, and $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2 \cdots \mathcal{K}_M$, where $\mathcal{M} = \{1, 2, \dots, M\}$ and $\mathcal{K} = \{1, 2, \dots, K\}$. Both UL transmissions from UEs and DL transmissions to UEs are assumed to be synchronous. Below let us first consider the channel modeling and estimation in UCCF networks.

III. CHANNEL MODELING AND ESTIMATION

In this section, the channel modeling for UCCF networks is first provided. Then, the principles of channel estimation is explained.

A. Channel Modeling

Consider a UCCF system where channels experience both the large-scale propagation pathloss and shadowing, and the small-scale fading, the channel impulse response (CIR) between UE k and AP m can be denoted as

$$\mathbf{h}_{mk} = \sqrt{g_{mk}} \tilde{\mathbf{h}}_{mk}, \quad k \in \mathcal{K}, m \in \mathcal{M} \quad (1)$$

where $\tilde{\mathbf{h}}_{mk} \in \mathcal{C}^{L_{mk} \times 1}$ accounts for the small-scale fading, L_{mk} is the number of taps of CIR and $E[\|\tilde{\mathbf{h}}_{mk}\|^2] = 1$. $\tilde{\mathbf{h}}_{mk}$ varies relatively fast, with the coherence period expressed as τ_c . In (1), g_{mk} models the large-scale propagation pathloss and shadowing effect. In practice, g_{mk} varies much slower than $\tilde{\mathbf{h}}_{mk}$. Hence, when averaging out the effect of $\tilde{\mathbf{h}}_{mk}$, the transmit power P_t of UE k and the receive power P_r of AP m from UE k have the relationship of $P_r = g_{mk} P_t$.

Typically, g_{mk} can be modelled to obey the lognormal distribution, with a PDF expressed as [2, 11]

$$f_{g_{mk}}(x) = \frac{\xi}{\sqrt{2\pi}\sigma_g x} \exp\left[-\frac{(10\log_{10} x - \mu(d_{mk}))^2}{2\sigma_g^2}\right], \quad x > 0 \quad (2)$$

where $\xi = 10/\ln 10 = 4.3429$, $\mu(d_{mk})$ (dB) and σ_g (dB) are the mean and standard deviation of $10\log_{10} g_{mk}$, respectively, d_{mk} is the distance between UE k and AP m . The mean $\mu(d_{mk})$ (dB) accounts for the propagation pathloss. In [2], the *double-slope* propagation pathloss model was employed for performance evaluation, which is represented as

$$\mu(d_{mk}) = -10\log_{10}\left[d_{mk}^a \left(1 + \frac{d_{mk}}{d_{Break}}\right)^b\right] \quad (3)$$

where a is referred to as the basic pathloss exponent, which takes a value of approximately 2, b is the additional pathloss exponent, which has a value ranging from 2 to 6, and d_{Break} is referred to as the break point of the propagation pathloss curve.

Another propagation pathloss model applied in the research on cell-free systems [12, 13] is the *triple-slope* model, which can be represented as [14]

$$\mu(d_{mk}) = \begin{cases} -L_P - 35 \log_{10}(d_{mk}), & \text{if } d_{mk} > d_1 \\ -L_P - 10 \log_{10}(d_1^{1.5} d_{mk}^2), & \text{if } d_0 < d_{mk} \leq d_1 \\ -L_P - 10 \log_{10}(d_1^{1.5} d_0^2), & \text{if } d_{mk} \leq d_0 \end{cases} \quad (4)$$

for certain d_0 and d_1 values, where L_P is given as

$$L_P = 46.3 + 33.9 \log_{10} f - 13.82 \log_{10} h_{AP} - [1.11 \log_{10} f - 0.7] h_{UE} + 1.56 \log_{10} f - 0.8 \quad (5)$$

with f defined as the carrier frequency in MHz, h_{AP} and h_{UE} the antenna heights of AP and UE.

B. Channel Estimation

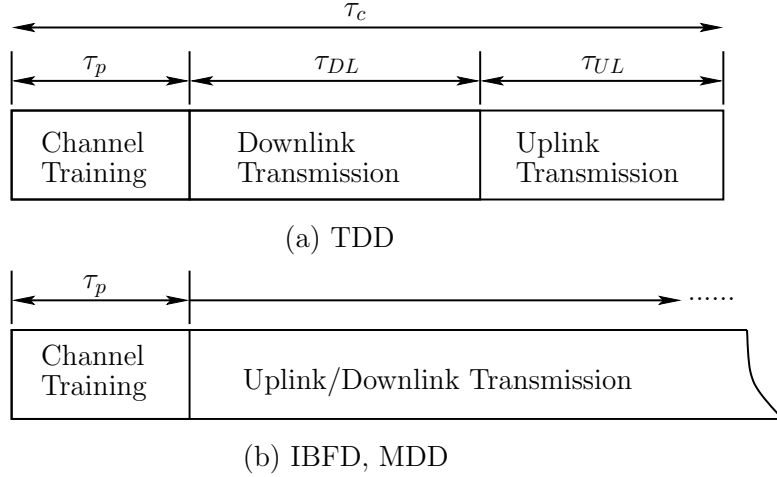


Fig. 3. Frame structures for the UCCF systems with TDD, MDD, and IBFD, respectively.

Assume a UCCF operational scenario, where channel's coherence time (or period, duration) is denoted as τ_c , as shown in Fig. 3. When time-division duplex (TDD) is employed, signal transmission over one coherence duration can be arranged as shown in Fig. 3(a). First, $\tau_p < \tau_c$ is used for UL channel training to estimate channels. Then, relying on the reciprocity between DL and UL channels, the estimated channel state information (CSI) is first applied for DL signal transmission to UEs, followed by using it for UL signal detection at APs or at CPU.

TDD is beneficial to UCCF systems for saving the overhead of CSI acquisition, in comparison with frequency-division duplex (FDD). However, in some applications, e.g., in fast time-varying communication environments, TDD may be deficient. First, when a wireless channel becomes more time variant, the channel's coherence duration τ_c becomes shorter. Hence, when given the time duration τ_p required for channel training, the duration left for

DL/UL data transmission becomes shorter, overall, resulting in a significant overhead increase of channel acquisition. Second, the CSI estimated at the beginning of a coherence duration becomes less accurate with time, yielding the channel ageing problem [7, 15]. If channel varies fast, the CSI applied for DL transmission and, especially, for UL detection may become outdated, resulting in significant performance degradation.

To mitigate the channel aging problem and save the overhead consumed for CSI acquisition under TDD, in-band full-duplex (IBFD) [16] may be employed to enable the simultaneous DL and UL transmissions on the same spectrum. Accordingly, the CSI for DL transmission and UL detection can be obtained whenever needed, by appropriately inserting UL pilots according to requirement or, simply, by employing the UL decision-directed channel estimation, which estimates channels with the aid of the reliably detected UL data. However, IBFD conflicts severe self-interference (SI), making channel estimation hard or even impossible, if SI is not sufficiently suppressed. Yet, the dilemma is that most efficient techniques for the SI suppression in IBFD are relied on CSI.

To circumvent the dilemma of IBFD but avoiding the shortcomings of TDD, multicarrier-division duplex (MDD) [7, 15, 17–19] may provide the solution. MDD belongs to an out-band full-duplex scheme, where a portion of subcarriers in a spectrum are assigned to support UL transmission, while the rest subcarriers are used to support DL transmission. Hence, as TDD, MDD enjoys the DL/UL reciprocity for CSI estimation, owing to the frequency-domain channel correlation. As IBFD, MDD possesses the properties of full-duplex, allowing simultaneous DL and UL transmissions. Hence, as IBFD, instantaneous CSI is available for DL transmission and UL detection in the MDD-supported systems, and hence, there is no channel aging problem. With regard to SI, in contrast to IBFD, which demands SI suppression in propagation, analog and digital domains, MDD only requires SI suppression in propagation and analog domains to protect the received signals to linearly pass analog-to-digital converters (ADCs). In this case, MDD enables the received signals to be free from digital-domain SI [18].

Hence, both IBFD- and MDD-aided systems may use the frame structure as shown in Fig. 3(b). During one session of communication, which may be much longer than the channel's coherence duration, channel training is only required at the beginning of the transmission session. Then, as above-mentioned, CSI can be updated whenever needed, with the aid of the appropriately inserted UL pilots or the decision-directed channel estimation. Hence, a significant overhead required for the channel acquisition by TDD can be saved, while without experiencing the channel aging problem.

Below the channel estimation during the initial channel training stage, as shown in Fig. 3(a) and (b), is analyzed. Assume that a pilot symbol block used by UE k for channel training is expressed as $\mathbf{S}_k \in \mathcal{C}^{N_k \times \tau_p}$, where N_k , $L_k \leq N_k \leq N$, is the number of pilot symbols sent per OFDM symbol duration, and τ_p is the number of OFDM symbol periods used for training. It is assumed that $E[|\mathbf{S}_k(i, j)|^2] = 1$, and $\mathbf{S}_k$ is known to the APs that UE k is associated with or/and to the CPU.

When pilot symbols are sent over the channels with the CIR expressed as (1), by following

the principles of OFDM in Chapter (OFDM), the frequency-domain observations obtained by AP m (assume that UE k is associated with AP m) can be expressed as

$$\mathbf{Y}_m = \sum_{k \in \mathcal{K}_m} \sqrt{P_k} \text{diag}(\mathbf{F}_N \mathbf{\Psi}_{mk} \mathbf{h}_{mk}) \mathbf{\Phi}_k \mathbf{S}_k + \mathbf{J}_m + \mathbf{N}_m, \quad m \in \mathcal{M} \quad (6)$$

where $\mathbf{Y}_m \in \mathcal{C}^{N \times \tau_p}$, $\mathbf{N}_m \in \mathcal{C}^{N \times \tau_p}$ is the Gaussian noise distributed with mean zero and a variance $\sigma^2/2$ per dimension, and $\mathbf{J}_m \in \mathcal{C}^{N \times \tau_p}$ is the possible interference from the other UEs not associated with AP m , which can be modelled to obey a Gaussian distribution with zero mean and a variance $\sigma_J^2/2$ per dimension. P_k is the transmit power per active subcarrier of UE k . $\mathbf{F}_N$ is the non-normalized FFT matrix satisfying $\mathbf{F}_N \mathbf{F}_N^H = \mathbf{F}_N^H \mathbf{F}_N = N \mathbf{I}_N$. $\mathbf{\Psi}_{mk}$ is a mapping matrix consisting of the first L_k columns of identity matrix $\mathbf{I}_N$, and $\mathbf{\Phi}_k$ is a mapping matrix, constructed by the N_k columns of $\mathbf{I}_N$ that correspond to the N_k subcarriers activated by UE k to send pilot symbols. $\mathbf{\Psi}_{mk}$ and $\mathbf{\Phi}_k$ are also known to AP m or/and CPU. Finally, $\text{diag}(\mathbf{a})$ yields a diagonal matrix using vector $\mathbf{a}$.

Upon vectorizing (6) and expressing $\mathbf{y}_m = \text{vec}(\mathbf{Y}_m)$, $\mathbf{n}_m = \text{vec}(\mathbf{N}_m)$ and $\mathbf{j}_m = \text{vec}(\mathbf{J}_m)$, which are $N\tau_p$ -length vectors, we have

$$\begin{aligned} \mathbf{y}_m &= \sum_{k \in \mathcal{K}_m} \sqrt{P_k} \tilde{\mathbf{S}}_k \mathbf{F}_N \mathbf{\Psi}_{mk} \mathbf{h}_{mk} + \mathbf{j}_m + \mathbf{n}_m, \\ &= \sum_{k \in \mathcal{K}_m} \mathbf{A}_{mk} \mathbf{h}_{mk} + \mathbf{j}_m + \mathbf{n}_m, \quad m = 1, 2, \dots, M \end{aligned} \quad (7)$$

where $\tilde{\mathbf{S}}_k = [\text{diag}(\mathbf{\Phi}_k \mathbf{s}_{k1}), \dots, \text{diag}(\mathbf{\Phi}_k \mathbf{s}_{k\tau_p})]^T$ with $\mathbf{s}_{ki}$ being the i th column of $\mathbf{S}_k$, and for simplicity, $\mathbf{A}_{mk} = \sqrt{P_k} \tilde{\mathbf{S}}_k \mathbf{F}_N \mathbf{\Psi}_{mk}$ is defined. Note that, $\tilde{\mathbf{S}}_k$ is a $(N\tau_p \times N)$ matrix and $\mathbf{A}_{mk}$ is $(N\tau_p \times L_k)$ -dimensional.

The objective of channel training is to estimate $\mathbf{h}_{mk}$ for all $k \in \mathcal{K}$ and their associated APs, which may be carried out locally at APs or at CPU. Assume that channel estimation is executed locally at APs. If pilot sequences are designed and assigned to UEs to make the terms in $\{\mathbf{A}_{mk}\}$ nearly orthogonal, i.e., $\mathbf{A}_{mk}^H \mathbf{A}_{nl}$ gives an almost orthogonal matrix for $k \neq l$, individual UE's channel can be estimated without considering multiuser interference (MUI) mitigation. Specifically, when the minimum mean-square-error (MMSE) relied channel estimation [20] is employed, $\mathbf{h}_{mk}$ can be estimated as

$$\hat{\mathbf{h}}_{mk} = \mathbf{C}_{mk}^{-1} \mathbf{Q}_{mk}^H \mathbf{A}_{mk}^H [\mathbf{A}_{mk} \mathbf{Q}_{mk} \mathbf{A}_{mk}^H + (\sigma_J^2 + \sigma^2) \mathbf{I}]^{-1} \mathbf{y}_m \quad (8)$$

where $\mathbf{Q}_{mk} = E[\mathbf{h}_{mk} \mathbf{h}_{mk}^H]$ is the covariance matrix of $\mathbf{h}_{mk}$, and $\mathbf{C}_{mk}$ is introduced to obtain an unbiased estimator, which is a diagonal matrix formed by the diagonal elements of $\mathbf{Q}_{mk}^H \mathbf{A}_{mk}^H [\mathbf{A}_{mk} \mathbf{Q}_{mk} \mathbf{A}_{mk}^H + (\sigma_J^2 + \sigma^2) \mathbf{I}]^{-1} \mathbf{A}_{mk}$.

However, if MUI suppression is required during channel estimation, $\mathbf{h}_{mk}$ can be estimated as

$$\hat{\mathbf{h}}_{mk} = \mathbf{C}_{mk}^{-1} \mathbf{Q}_{mk}^H \mathbf{A}_{mk}^H \left[\sum_{l \in \mathcal{K}_m} \mathbf{A}_{ml} \mathbf{Q}_{ml} \mathbf{A}_{ml}^H + (\sigma_J^2 + \sigma^2) \mathbf{I} \right]^{-1} \mathbf{y}_m \quad (9)$$

In (9), the matrix in the bracket is constructed using the pilots assigned to UEs, the mapping matrices used by UEs, and the statistics about interference and noise. In the case that these requirements cannot be satisfied, a matrix can be directly estimated from (6). First, based on (6), an autocorrelation matrix of $\mathbf{Y}_m$ can be obtained as

$$\mathbf{R}_m = \mathbf{Y}_m \mathbf{Y}_m^H / \tau_p \quad (10)$$

Note that, if some observations for data transmission following channel training are available, the estimation of $\mathbf{R}_m$ can be enhanced by invoking the data-related observations in (10). Then, using $\mathbf{R}_m$, the matrix in the bracket of (9) is replaced by $(\mathbf{I}_{\tau_p} \otimes \mathbf{R}_m)$, where $\otimes$ represents the Kronecker operation.

After the channels of UEs are respectively estimated at APs, APs can forward the estimated channels to CPU, if required. Alternatively, APs may forward the observations seen in (6) to CPU, and CPU carries out the channel estimation. In this case, the observations obtained by CPU may be expressed as

$$\mathbf{Y}'_m = a_{cm} \mathbf{Y}_m, \quad m \in \mathcal{M} \quad (11)$$

where a_{cm} is the channel gain between CPU and AP m . In ideal case, or when CPU and AP m have the ideal knowledge about a_{cm} , $a_{cm} = 1$ can be assumed. After CPU obtains $\mathbf{Y}'_m$, it can estimate the channels of the UEs associated with AP m , in the same way as done in (8) or (9).

After obtaining the estimate $\hat{\mathbf{h}}_{mk}$ to the CIR of $\mathbf{h}_{mk}$, the channel gains of N subcarriers can be obtained as

$$\hat{\mathbf{h}}_{f,mk} = \mathbf{F}_N \mathbf{\Psi}_{mk} \hat{\mathbf{h}}_{mk}, \quad m \in \mathcal{M}; k \in \mathcal{K}_m \quad (12)$$

which can be applied for UL detection and DL transmission.

It is worth noting that in $\hat{\mathbf{h}}_{f,mk} = \sqrt{\hat{g}_{mk}} \hat{\mathbf{h}}_{mk}$, the large-scale propagation pathloss and shadowing, i.e., $\hat{g}_{mk}$, is the same for all subcarriers, but the small-scale fading gains of different subcarriers in $\hat{\mathbf{h}}_{mk}$ may be different.

IV. ACCESS-POINT ASSOCIATION OF USER EQUIPMENTS

Optimal association of UEs with APs with the objective to, e.g., maximize system sum-rate, energy efficiency, etc., under various constraints, e.g., individual UEs' minimum rate and power, total AP transmit power, backhaul resources, etc., can be extremely complicated, if the number of APs and the number of UEs are relatively big. The complexity roots in that the optimization of UE association is often coupled with the other optimizations at system and link levels. Hence, in practice, simple low-complexity association methods are desired, and to achieve this, the UE association process is usually separated from the processes of system and link optimization. Below are some low-complexity UE association methods.

The first one is the simplest distance-based UE association. With this method, a distance R_{th} is initialized for the association decision making. A UE k is associated with all the APs

that have their distances from UE k not exceeding R_{th} , which form a set $\mathcal{M}_k$. If there is no AP satisfying the condition, UE k is then associated with an AP that has the minimum distance from the UE, provided that the minimum QoS required by UE k can be met via, such as, power-control or/and variable data rate transmission, etc., techniques. Otherwise, if the minimum QoS requirement of UE k is unable to be guaranteed in any way, the UE will not be associated with any APs, and may be disconnected from the network.

The second method carries out UE association based on the large-scale fading, i.e., based on g_{mk} seen in (1). To implement this method, for each UE k , $\{g_{mk}\}$ are first listed in descending order. Then, the APs are selected to be included in $\mathcal{M}_k$ according to the list, in the order from the best (largest g_{mk}) to worst (smallest g_{mk}), until the stop conditions are met. The stop conditions may be the number of APs, g_{mk} is lower than a pre-set threshold, etc.

In some UCCF networks, APs are supposed to be densely deployed. This kind of networks have the feasibility for the implementation of wireless sensing, which can benefit from the relatively easy access of LoS signals and the availability of many reference points (such as APs) for variable sensing. Furthermore, with the aid of integrated sensing and communications (ISAC), it is achievable that APs and UEs are able to know the detailed communication environments, in addition to the accurate positions of them. With this positioning and environmental information, it is possible for a UE to find the best physical paths for it to propagate signals to APs. Similarly, APs can use the best paths to send information to UEs. Hence, with the aid of wireless sensing or ISAC, UEs can collaborate with APs to settle down their associations based on the sensed information about communication environments as well as the APs and UEs themselves.

V. UPLINK DETECTION AND OPTIMIZATION

This section first provides the principles of UL detection in UCCF systems via analyzing several detection methods. Then, the UL resource optimization is discussed.

A. Uplink Detection Schemes

To study the UL detection and optimization, assume that UE k is associated with the APs in $\mathcal{M}_k$, which includes AP m . Let the frequency-domain symbol vector sent by UE k on N_k subcarriers is denoted as $\mathbf{x}_k \in \mathcal{C}^{N_k \times 1}$, which satisfies $E[\mathbf{x}_k \mathbf{x}_k^H] = \mathbf{I}_{N_k}$. Assume that the maximum transmit power of UEs is P_u , and the k th UE's transmit power is $P_k = \eta_k P_u$, where $\eta_k \leq 1$ is the power-control coefficient. Then, by following the principles of OFDM in Chapter (OFDM), the corresponding observations obtained by AP m from N subcarriers can be represented as

$$\mathbf{y}_m = \sum_{k \in \mathcal{K}} \mathbf{H}_{mk} \Phi_k \eta_k^{1/2} \mathbf{x}_k + \mathbf{n}_m, \quad m = 1, 2, \dots, M \quad (13)$$

where $\mathbf{H}_{mk} = \text{diag}(\mathbf{F}_N \Psi_{mk} \mathbf{h}_{mk})$, with the diagonal elements denoting the channel gains of corresponding subcarriers, and $\mathbf{h}_{mk}$ is given by (1), including both the large-scale propagation

pathloss and shadowing, as well as the small-scale fading gains. It can be understood that the large-scale propagation pathloss and shadowing are UE dependent but not subcarrier dependent, while the small-scale fading gains of different subcarriers of a UE may be different. Again, Φ_k executes subcarrier-allocation to choose N_k subcarriers for transmitting the information of UE k . $\eta_k = \text{diag}\{\eta_{k1}, \eta_{k2}, \dots, \eta_{kN_k}\}$, satisfying $\sum_{n=1}^{N_k} \eta_{kn} = \eta_k$, controls the power assigned to the N_k subcarriers. $\mathbf{n}_m$ is the Gaussian noise distributed with mean zero and a covariance matrix of $\gamma_u^{-1} \mathbf{I}_N$, where $\gamma_u = P_u/\sigma^2$.

Below the principles of three detection schemes are analyzed, including the global MMSE (GMMSE) detection, local MMSE (LMMSE) detection and the AP message-passing (APMP) detection. To illustrate the principles, channels are assumed to be ideally estimated, backhaul links are ideal for information exchange between CPU and APs, and the UCCF network is operated in the ideal synchronization state.

1) *Global MMSE Detection:* GMMSE detector is operated at CPU, which uses all the observations from M APs to detect any a UE's information. In other words, APs do not process their observations locally, but directly forward $\{\mathbf{y}_m\}$ in (13) to CPU. At CPU, let $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_M^T]^T$, $\mathbf{n} = [\mathbf{n}_1^T, \mathbf{n}_2^T, \dots, \mathbf{n}_M^T]^T$, $\mathbf{H}_k = [\mathbf{H}_{1k}^T, \mathbf{H}_{2k}^T, \dots, \mathbf{H}_{Mk}^T]^T$. Then, $\mathbf{y}$ can be written as

$$\mathbf{y} = \sum_{k \in \mathcal{K}} \mathbf{H}_k \Phi_k \eta_k^{1/2} \mathbf{x}_k + \mathbf{n} \quad (14)$$

To detect the information of UE k , CPU forms the decision variable in MMSE principle as

$$\hat{\mathbf{x}}_k = \mathbf{W}_k^H \mathbf{y}, \quad k \in \mathcal{K} \quad (15)$$

where $\mathbf{W}_k$ can be expressed as [17]

$$\mathbf{W}_k = \mathbf{R}_y^{-1} \mathbf{R}_{yk} \quad (16)$$

with $\mathbf{R}_y$ being autocorrelation matrix of $\mathbf{y}$, given by

$$\mathbf{R}_y = E[\mathbf{y}\mathbf{y}^H] = \sum_{k \in \mathcal{K}} \mathbf{H}_k \Phi_k \eta_k \Phi_k^T \mathbf{H}_k^H + \gamma_u^{-1} \mathbf{I}_{MN} \quad (17)$$

and $\mathbf{R}_{yk}$ being the cross-correlation matrix between $\mathbf{y}$ and $\mathbf{x}_k$, having the expression of

$$\mathbf{R}_{yk} = E[\mathbf{y}\mathbf{x}_k^H] = \mathbf{H}_k \Phi_k \eta_k^{1/2} \quad (18)$$

Hence, $\mathbf{W}_k$ is

$$\mathbf{W}_k = \left(\sum_{l \in \mathcal{K}} \mathbf{H}_l \Phi_l \eta_l \Phi_l^T \mathbf{H}_l^H + \gamma_u^{-1} \mathbf{I}_{MN} \right)^{-1} \mathbf{H}_k \Phi_k \eta_k^{1/2} \quad (19)$$

Note that, in addition to mitigating MUI, the GMMSE detector described in (15)-(19) is also able to suppress the embedded inter-carrier interference (ICI), if it exists. If subcarriers are ideally orthogonal, the GMMSE detector can be separated into N GMMSE detectors operated in parallel, each is for detecting the symbols conveyed on one subcarrier. It can also

be shown that the complexity required by the N separate GMMSE detectors is much lower than that of the joint GMMSE detector. Moreover, corresponding to one subcarrier, such as, n , the formulas for the GMMSE detector are similar as that provided in (14)-(19), only with (13) changed to

$$y_{mn} = \sum_{k \in \mathcal{K}} \sqrt{\eta_{kn}} h_{mn,k} \delta_{kn} x_{kn} + n_{mn}, n \in \mathcal{N}; m \in \mathcal{M} \quad (20)$$

where $h_{mn,k}$ is the channel gain of the n th subcarrier of UE k , x_{kn} is the symbol sent on subcarrier n by UE k , $\delta_{kn} = 1$ if UE k is assigned subcarrier n , otherwise, $\delta_{kn} = 0$.

Expressing $\mathbf{W}_k = [\mathbf{w}_{k1}, \mathbf{w}_{k2}, \dots, \mathbf{w}_{kN_k}]$. It can be shown that $\mathbf{w}_{ki}$ for detecting the i th symbol of UE k is

$$\mathbf{w}_{ki} = \sqrt{\eta_{ki}} \left(\sum_{l \in \mathcal{K}} \mathbf{H}_l \Phi_l \eta_l \Phi_l^T \mathbf{H}_l^H + \gamma_u^{-1} \mathbf{I}_{MN} \right)^{-1} \mathbf{H}_k \phi_{ki} \quad (21)$$

where ϕ_{ki} is the i th column of Φ_k . The SINR is [17]

$$\begin{aligned} \gamma_{ki}(\{\eta_l\}, \{\Phi_l\}) &= \eta_{ki} \phi_{ki}^H \mathbf{H}_k^H \mathbf{R}_{ki}^{-1} \mathbf{H}_k \phi_{ki} \\ &= \eta_{ki} \phi_{ki}^H \mathbf{H}_k^H \left(\sum_{l \in \mathcal{K}} \mathbf{H}_l \Phi_l \eta_l \Phi_l^T \mathbf{H}_l^H - \eta_{ki} \mathbf{H}_k \phi_{ki} \phi_{ki}^T \mathbf{H}_k^H + \gamma_u^{-1} \mathbf{I}_{MN} \right)^{-1} \mathbf{H}_k \phi_{ki} \end{aligned} \quad (22)$$

where $\mathbf{R}_{ki} = \sum_{l \in \mathcal{K}} \mathbf{H}_l \Phi_l \eta_l \Phi_l^T \mathbf{H}_l^H - \eta_{ki} \mathbf{H}_k \phi_{ki} \phi_{ki}^T \mathbf{H}_k^H + \gamma_u^{-1} \mathbf{I}_{MN}$ is the autocorrelation matrix of interference plus noise when detecting symbol i of UE k . Accordingly, the sum-rate of UCCF system is

$$R(\{\eta_l\}, \{\Phi_l\}) = \sum_{k \in \mathcal{K}} \sum_{i=1}^{N_k} \log_2 [1 + \gamma_{ki}(\{\eta_l\}, \{\Phi_l\})] \quad (23)$$

which is a function of the power-control coefficients in $\{\eta_l\}$, and the subcarrier mapping matrices $\{\Phi_l\}$ that implement subcarrier-allocation.

GMMSE detector has the potential to achieve promising performance in the case that $M \geq K$, and the UEs are evenly distributed in the network. However, it is hard to implement. As shown in $\mathbf{H}_k$ defined above (14), it includes not only the channels between UE k and the APs in $\mathcal{M}_k$, i.e., the APs that UE k is associated with, but also the channels in $\bar{\mathcal{M}}_k$, which represents the APs that UE k is not associated with. In practice, the channels between UE k and the APs in $\bar{\mathcal{M}}_k$ are supposed to be weak. Hence, these channels can hardly be estimated with sufficient accuracy for information detection. Furthermore, GMMSE detector needs to invert a matrix of $(MN \times MN)$ -dimensional, or N matrices of $(M \times M)$ -dimensional in the case of using N parallel detectors, with each for one subcarrier, the detection complexity may be extreme for practical implementation. To circumvent these problems, below the local detection schemes that are UEs' location-aware are considered. Here, the local detection of UE k means that its information is detected only based on the observations collected from its associated APs.

2) *Local MMSE Detection*: LMMSE detection can be operated at CPU or at APs [6]. First, when operated at CPU, let us express the inverted matrix $\mathbf{R}_y^{-1}$ in (19) as $\mathbf{R}_y^{-1} = [\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_M]$, where $\mathbf{Q}_m$ are $(MN \times N)$ matrices. Then, when only considering the APs in $\mathcal{M}_k$ to detect UE k , the weight matrix can be approximated by

$$\mathbf{W}_k \approx \sum_{m \in \mathcal{M}_k} \mathbf{Q}_m \mathbf{H}_{mk} \Phi_k \boldsymbol{\eta}^{1/2} \quad (24)$$

Note that the autocorrelation matrix $\mathbf{R}_y$ in (19) can be kept the same for the LMMSE detection, as it can be directly estimated from the observations provided by APs as

$$\mathbf{R}_y = \frac{1}{U} \sum_{u=1}^U \mathbf{y}(u) \mathbf{y}^H(u) \quad (25)$$

where $\{\mathbf{y}(u)\}$ are observations of $\mathbf{y}$ obtained in U OFDM symbol periods. The estimate to $\mathbf{R}_y$ becomes more accurate, as U increases.

Although the approximated weight matrix in (24) does not use all the columns of $\mathbf{R}_y^{-1}$, but each $\mathbf{Q}_m$ has the corresponding elements in the MN rows of $\mathbf{R}_y^{-1}$. Hence, the LMMSE detector with the $\mathbf{W}_k$ of (24) is capable of effectively suppressing MUI. In comparison with the ideal GMMSE detector in Section V-A1, the performance loss of the LMMSE detector lies in the fact that some APs may receive certain power from UE k , but they are not included in $\mathcal{M}_k$, and hence the power is not exploited for the detection of UE k . Therefore, in the UE association stage, the APs making noticeable contribution to UE k 's detection performance are expected to be included in $\mathcal{M}_k$, if the complexity for, such as, channel estimation, is allowable.

The LMMSE detector for UE k may be implemented at CPU by only considering the observations from the APs that UE k is associated with. Alternatively, it can be implemented at an AP (e.g., the best one in terms of UE k) via AP cooperation, which is supported by information exchange between APs (usually between the geographically adjacent APs). Assume that the information exchange between APs and CPU, or between APs is ideal. Let, for the $|\mathcal{M}_k|$ APs associated by UE k , define $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_{|\mathcal{M}_k|}^T]^T$, $\mathbf{n} = [\mathbf{n}_1^T, \mathbf{n}_2^T, \dots, \mathbf{n}_{|\mathcal{M}_k|}^T]^T$ and $\mathbf{H}_k = [\mathbf{H}_{1k}^T, \mathbf{H}_{2k}^T, \dots, \mathbf{H}_{|\mathcal{M}_k|k}^T]^T$. Then, we have a same representation as (14) for carrying out the LMMSE detection. Furthermore, the LMMSE detector has the same set of formulas as shown in (15)-(23), but the dimension MN of the matrices invoked in the GMMSE detector is reduced to $|\mathcal{M}_k|N$ in the LMMSE detector. Since $|\mathcal{M}_k| \ll M$, it can be expected that the LMMSE detector performs worse than the ideal GMMSE detector, due to its deteriorated MUI suppressing capability as compared to the ideal GMMSE detector. However, this LMMSE detector has significantly lower complexity than the GMMSE detector and the LMMSE detector using the weighting of (24).

In a LMMSE detection scenario where main detection processing is carried out at APs, if APs do not cooperate with each other, the MMSE processing has to be built on $\mathbf{y}_m$ of (13). In this case, the m th AP can first form a local estimator corresponding to UE k as

$$\mathbf{z}_{mk} = \mathbf{W}_{mk}^H \mathbf{y}_m, \quad m \in \mathcal{M}_k, \quad k \in \mathcal{K} \quad (26)$$

where the weight matrix $\mathbf{W}_{mk}$ can be derived based on (13), and is given by

$$\mathbf{W}_{mk} = \left(\sum_{l \in \mathcal{K}} \mathbf{H}_{ml} \Phi_l \boldsymbol{\eta}_l \Phi_l^T \mathbf{H}_{ml}^H + \gamma_u^{-1} \mathbf{I}_N \right)^{-1} \mathbf{H}_{mk} \Phi_k \boldsymbol{\eta}_k^{1/2} \quad (27)$$

Then, $\mathbf{z}_{mk}$ is forwarded to CPU, where a final decision variable vector for UE k 's N_k symbols can be formed as

$$\mathbf{z}_k = \sum_{m \in \mathcal{M}_k} \boldsymbol{\Lambda}_{mk} \mathbf{z}_{mk}, \quad k \in \mathcal{K} \quad (28)$$

where $\boldsymbol{\Lambda}_{mk} = \text{diag}\{\lambda_{mk,1}, \lambda_{mk,2}, \dots, \lambda_{mk,N_k}\}$ are the weights used by CPU to combine $\mathbf{z}_{mk}$.

Depending on the channel knowledge available to CPU, the combining weights in (28) can be set in different ways. First, when CPU has no knowledge at all, it chooses $\boldsymbol{\Lambda}_{mk} = \mathbf{I}_{N_k}/|\mathcal{M}_k|$. Second, express $\mathbf{z}_{mk}$ in (26) as $\mathbf{z}_{mk} = \mathbf{A}_{mk} \mathbf{x}_k + \mathbf{n}_{I,mk} + \mathbf{n}_{mk}$, where $\mathbf{n}_{I,mk}$ is interference. Then, when CPU knows $\mathbf{A}_{mk}$ and the covariance matrix of interference plus noise, which is expressed as $\mathbf{C}_{mk}$, it can set

$$\boldsymbol{\Lambda}_{mk} = \mathbf{C}_{mk}^{-1} \mathbf{A}_{mk}^H / |\mathcal{M}_k| \quad (29)$$

to implement maximal ratio combining (MRC), which maximize SINR [21]. To compute (29), the knowledge of both the large-scale and small-scale fading are required. Third, if CPU only has the knowledge of the large-scale fading, it may combine $\mathbf{z}_{mk}$ as

$$\mathbf{z}_k = \sum_{m \in \mathcal{M}_k} \frac{g_{mk}}{\sum_{l \in \mathcal{M}_k} g_{lk}} \mathbf{z}_{mk} \quad \text{or} \quad \mathbf{z}_k = \sum_{m \in \mathcal{M}_k} \frac{\sqrt{g_{mk}}}{\sum_{l \in \mathcal{M}_k} \sqrt{g_{lk}}} \mathbf{z}_{mk} \quad (30)$$

where g_{lk} is defined with (1).

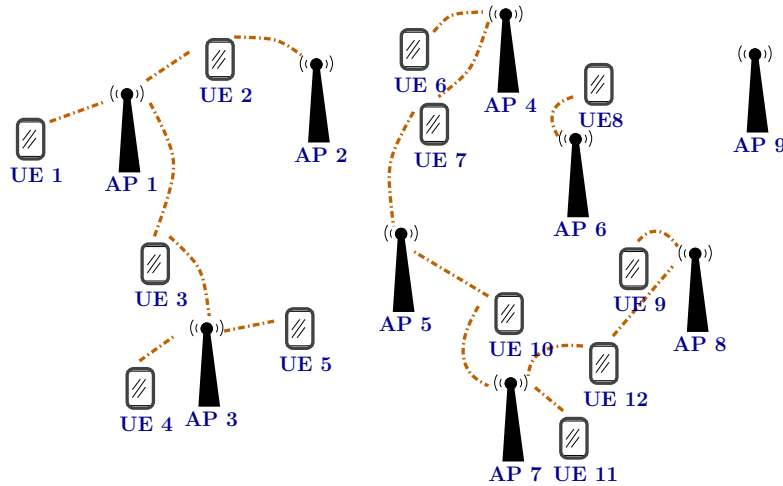


Fig. 4. An example of UCCF network to illustrate the application of the message-passing algorithm (MPA) for UL detection.

3) *Access Point Message Passing Detection*: In a densely deployed UCCF network, each AP typically serves a small number UEs, and each UE is only associated with a small number of APs close to the UE. Hence, when abstracting APs as function nodes (FNs) and UEs as

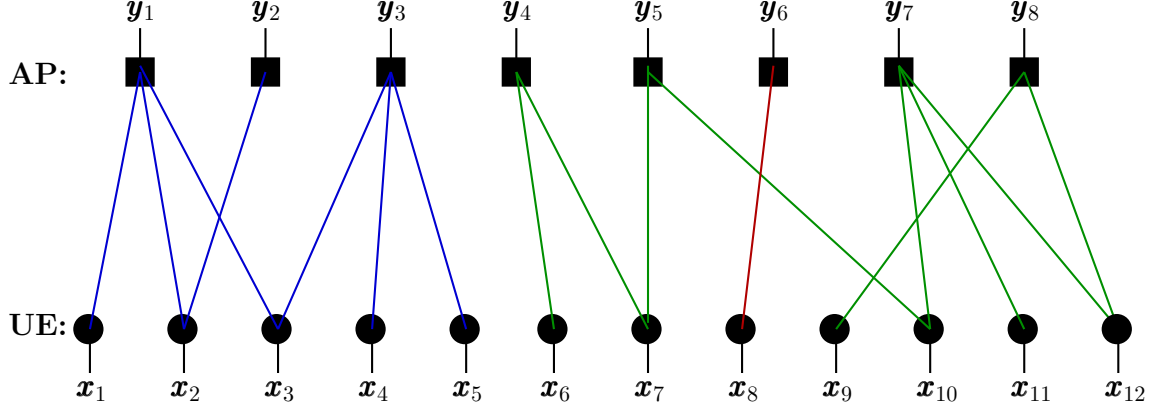


Fig. 5. Factor graph showing the relationship between APs and UEs in the UCCF network of Fig. 4.

variable nodes (VNs), a sparse factor graph [22] can be constructed to explain the relationship of APs and UEs in the UCCF network. For example, Fig. 4 shows a UCCF network with 9 APs and 12 UEs, as well as the association states of the UEs. Accordingly, the factor graph describing the relationship between APs/UEs is shown in Fig. 5, where AP i and UE j are connected by an edge, if UE j is associated with AP i . Based on Fig. 4 and Fig. 5, the factor graph describing a UCCF network has the following characteristics:

- The factor graph is sparse.
- In comparison with the well-designed regular factor graphs describing low-density parity check (LDPC) codes [22], the edges connecting FNs and VNs are irregular, which are depended on the distributions of UEs and APs, as well as the association rules applied.
- The factor graph may be divided into several sub-graphs that are independent of each other. Seen in Fig. 5, there are three sub-graphs, having the FNs $\{1, 2, 3\}$, $\{4, 5, 7, 8\}$ and $\{6\}$, respectively. This reflects that the UEs in different sub-region are associated with the isolated groups of APs, due to their geographical separation. Hence, the interference between two UEs belonging to two sub-graphs is low or no interference between them.
- It is possible that one sub-graph contains only one UE connected to one AP, e.g., UE 8 to AP 6 in Fig. 5, or to multiple APs. Furthermore, it is possible that an AP has no UEs associated with, e.g., AP 9 in Fig. 4. Hence, the corresponding FN has no edges and in the case, the AP can be switched off to save energy.

The access point message passing (APMP) detector [6] exploits the principles of message passing algorithm (MPA) for signal detection. It is operated on the factor graph, as shown in Fig. 5 for example. In detail, the APMP detection process is described as follows, where intrinsic and extrinsic information can be calculated in the principles, for example, as detailed in [6, 22–24]. To explain, let us assume an example that APs i and j are two adjacent APs, and UE k has been associated with both of them.

- 1) After UE association, APs report to CPU their associated UEs, i.e., send $\mathcal{K}_m, \forall m \in \mathcal{M}$,

to CPU.

- 2) Based on $\{\mathcal{K}_m\}$, CPU builds the factor graph and informs each AP its adjacent APs. Two APs are defined as adjacent APs, if they monitor one to several common UEs. Hence, in the example, APs i and j are adjacent APs. Later during detection, two adjacent APs will exchange the information about their commonly monitored UEs to improve the detection reliabilities of these UEs.
- 3) Initially, each of APs calculates the intrinsic information of the symbols (or bits) of the UEs associated with it, based on its locally received observations.
- 4) Adjacent APs exchange the information of their commonly monitored UEs. Considering the example, APs i and j exchange their information about UE k . AP i and AP j may also receive information about UE k from their other adjacent APs.
- 5) After receiving the information from adjacent APs, for each of its associated UEs, an AP computes, respectively, the total information of the individual symbols (or bits) of the UE.
- 6) Then, each AP feeds back the extrinsic information of symbols (bits) of an UE to its adjacent APs that also monitor the UE. Here, the extrinsic information of a symbol (or bit) passed to a specific AP is equal to the difference between the total information of the symbol (or bit) and the information of that symbol (or bit) previously received from the same AP.
Considering AP i in the example, during an iteration, it first calculates the total information of a symbol (or bit) based on its intrinsic information and the extrinsic information received from other APs. Then, the extrinsic information fed back to AP j is the difference between the total information of the symbol (or bit) and the information about the symbol (or bit) that AP i previously received from AP j .
- 7) The iteration process 4) - 6) is repeated until no increase of information is available, or the allowed number of iterations is reached.
- 8) Finally, the symbols (or bits) of a specific UE are decided by one of its APs.

From the above description, the APMP detector is mainly operated by APs, CPU is only required to build and maintain the factor graph, and update it to APs when there are changes occurred in the network. APs are only required to exchange information with their adjacent APs. Hence, it is expected that the demand on backhaul resources is moderate. In addition, it is well-known, also evidenced by the results in literature, such as references [6, 22–24], that the performance achieved by the MPA-assisted detection is near-optimum, close to that achieved by the optimum maximum likelihood detection.

B. Uplink Resource Optimization

Uplink resource optimization in the considered OFDM signaling UCCF system includes UE association, subcarrier-allocation and power-control. Optimization constraints include the available transmit power of UEs and, possibly, some service requirements, such as, minimum

rate, delay, etc., of individual UEs. Hence, the optimum solutions to maximize UL sum-rate are required to solve, for example, the problem:

$$\begin{aligned} & \{\zeta_{mk}^*\}, \{\delta_{nk}^*\}, \{\eta_{kn}^*\} \\ & = \arg \max_{\{\delta_{kn}\}, \{\zeta_{mk}\}, \{\eta_{kn}\}} \left\{ \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \delta_{kn} \log_2 (1 + \gamma_{kn}(\{\zeta_{mk}\}, \{\eta_{kn}\})) \right\} \end{aligned} \quad (31a)$$

$$s.t. \quad \zeta_{mk} \in \{0, 1\}, \quad \forall m \in \mathcal{M}, k \in \mathcal{K} \quad (31b)$$

$$\delta_{kn} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, n \in \mathcal{N} \quad (31c)$$

$$\sum_{n \in \mathcal{N}} \eta_{kn} = \eta_k \leq 1, \quad \forall k \in \mathcal{K} \quad (31d)$$

$$\sum_{n \in \mathcal{N}} \delta_{kn} \log_2 (1 + \gamma_{kn}(\{\zeta_{mk}\}, \{\eta_{kn}\})) \geq R_k^{\min}, \quad \forall k \in \mathcal{K} \quad (31e)$$

In this optimization problem, (31)(b) controls UE association, (31)(c) carries out subcarrier-allocation, (31)(d) constrains the power assigned to the active subcarriers of UEs, and finally, (31)(e) guarantees that each UE is at least provided with the minimum service rate as required. In the above formula, $\gamma_{kn}(\{\zeta_{mk}\}, \{\eta_{kn}\})$ is the SINR of subcarrier n of UE k , which is dependent on the detection method employed, it is a function of the APs that UE k is associated with, and the power allocated to the subcarriers assigned to UE k . Finally, $\{\zeta_{mk}^*\}$, $\{\delta_{nk}^*\}$ and $\{\eta_{kn}^*\}$ are the sets of solutions obtained after the optimization.

Due to the involved binary optimization from UE association and subcarrier-allocation, the problem of (31) is NP hard to solve. Hence, low-complexity and yet efficient methods are aimed at. The simplest method might be optimizing UE association, subcarrier-allocation and power-allocation in succession separately. Specifically, UE association can be executed first, for example, by applying a method discussed in Section IV. After UE association, referring to Chapters **(Subcarrier-Allocation)**, subcarrier-allocation can then be carried out by considering the rate requirements of individual UEs, the relative distributions of APs and UEs, possible interference between UEs, etc. Finally, power is allocated to the assigned subcarriers of a UE to maximize the total rate of the UE.

Instead of maximizing sum-rate, energy-efficiency can be an alternative utility function to maximize in Problem (31). In the UCCF systems as considered, the power consumption includes the circuit power by UEs, APs and CPU, the transmit power between UEs and APs, that between APs and CPU, and possibly, the circuit/transmit power for AP cooperation. Due to the fact that sum-rate is a function of power, the energy-efficiency is a fractional objective function, which is non-concave and hard to solve using standard programming methods, even only power-control is considered [25]. Hence, the optimization problem has to be divided into a range of sub-problems to be tackled, such as, by the means of the sequential optimization operated on the lower-bound of the energy-efficiency function [25]. Note that, maximizing the energy-efficiency of UCCF networks may result in that some APs are switched off to save energy.

Another optimization objective may be to maximize the minimum rate or minimum SINR of UEs, forming the max-min problem, as shown in Chapter **(Power-Allocation)**. Again, the joint optimization of UE association, subcarrier-allocation and power-control is extremely difficult to achieve. They may instead be optimized successively, following the order of UE association, subcarrier-allocation and power-control. In Chapter **(Power-Allocation)**, it is shown that the max-min optimization makes a trade-off on efficiency for fairness. However, in UCCF networks, each UE can be expected to be located close to one and even several APs. Hence, every UE is a strong UE with regard to some APs, resulting in that the greedy-based subcarrier-allocation should be near-optimum. Consequently, each UE is able to send information over high-quality channels. In this case, the followed power-allocation is relatively simple to implement, and the power-control optimized in max-min principle can also be near-optimum in the sense of maximal sum-rate or maximal energy efficiency, in addition to the fairness provided by the max-min optimization. In plain language, a UCCF network is capable of enabling the system-level efficiency and the individual-level fairness to be simultaneously near-optimum.

VI. DOWNLINK TRANSMISSION AND OPTIMIZATION

Considering the transmitter preprocessing in MMSE principle (TMMSE), in this section, the principles of DL transmission and optimization are considered. For clarity and generality, subcarrier level and OFDM symbol level processings are analyzed in parallel. Note that, when only one subcarrier is invoked, the model and analysis become the same of a DL space division multiplexing (SDM) model. By contrast, if OFDM is invoked, the model implements both SDM and frequency-division multiplexing (FDM). Note that, if an OFDM system exists ICI, the preprocessing at OFDM symbol level is required to suppress this interference.

A. Downlink Transmission with Precoding

Following the assumptions and settings applied in the UL analysis, let us assume that a subcarrier symbol or an OFDM symbol to be sent to UE k by AP m is expressed as $x_k \in \mathcal{C}$ or $\mathbf{x}_k \in \mathcal{C}^{N_k \times 1}$, which satisfies $E[|x_k|^2] = 1$ or $E[\|\mathbf{x}_k\|^2] = N_k$. Assume that the design is implemented by CPU and, for the moment, assume that CPU has the ideal channel knowledge from any of APs to any of UEs. In other words, all M APs are assumed to simultaneously transmit signals to all K UEs. Then, after preprocessing the data symbols of the K UEs, the signal sent by AP m , $m \in \mathcal{M}$, can be expressed as

$$s_m = A_0 \sum_{k \in \mathcal{K}} p_{mk} x_k \quad (32a)$$

$$\tilde{\mathbf{s}}_m = A_0 \sum_{k \in \mathcal{K}} \mathbf{P}_{mk} \Phi_k \mathbf{x}_k \quad (32b)$$

where $\mathbf{a}$ and $\tilde{\mathbf{a}}$ are used to indicate the subcarrier level and OFDM symbol level, respectively. In (32), $s_m \in \mathcal{C}$ and $\tilde{\mathbf{s}}_m \in \mathcal{C}^{N \times 1}$, $p_{mk} \in \mathcal{C}$ and $\mathbf{P}_{mk} = \text{diag}\{\mathbf{p}_{mk}\} \in \mathcal{C}^{N \times N}$ are preprocessing scalar and matrices. Again, the mapping matrices $\{\Phi_k\}$ implement DL subcarrier-allocation.

The transmission is under power constraint, to be satisfied via the precoder design and the amplification by a constant gain A_0 , which will be discussed later. The objective of DL optimization is to design the preprocessing vectors or matrices, or so-called precoders, under the constraints of APs' transmit power and/or other service requirements.

When s_m or $\tilde{s}_m$ for all $m \in \mathcal{M}$ are sent over DL channels, it can be shown that the observation received by UE k , $k \in \mathcal{K}$, can be expressed as

$$y_k = A_0 \sum_{m \in \mathcal{M}} h_{mk} s_m + n_k \quad (33a)$$

$$\tilde{\mathbf{y}}_k = A_0 \sum_{m \in \mathcal{M}} \mathbf{H}_{mk}^T \tilde{s}_m + \tilde{\mathbf{n}}_k \quad (33b)$$

where h_{mk} and $\mathbf{H}_{mk}$ are the channels from UE k to AP m , and n_k and $\tilde{\mathbf{n}}_k$ are Gaussian noise distributed with zero mean and a common variance $\sigma^2/2$ per dimension.

Substituting s_m and $\tilde{s}_m$ from (32) respectively into (33), it can be shown that

$$y_k = A_0 \mathbf{h}_k^T \sum_{l \in \mathcal{K}} \mathbf{p}_l x_l + n_k \quad (34a)$$

$$\tilde{\mathbf{y}}_k = A_0 \mathbf{H}_k^T \sum_{l \in \mathcal{K}} \mathbf{P}_l \Phi_l \mathbf{x}_l + \tilde{\mathbf{n}}_k \quad (34b)$$

where $\mathbf{h}_k = [h_{1k}, h_{2k}, \dots, h_{Mk}]^T$, $\mathbf{p}_l = [p_{1l}, p_{2l}, \dots, p_{Ml}]^T$, and $\mathbf{H}_k$ is defined above (14). $\mathbf{p}_l \in \mathcal{C}^{M \times 1}$ and $\mathbf{P}_l \in \mathcal{C}^{MN \times N}$ are the vector and matrix for preprocessing the signals sent to UE k . Corresponding to $\mathbf{H}_k$, we have $\mathbf{P}_k = [\mathbf{P}_{1k}^T, \mathbf{P}_{2k}^T, \dots, \mathbf{P}_{Mk}^T]^T$.

In (34), the preprocessing vectors (matrices) can be designed by exploiting the equivalency between the linear transmitter preprocessing and linear receiver processing (detection) [17, 26–29]. In other words, the preprocessing vector $\mathbf{p}_k$ or matrix $\mathbf{P}_k$ can be obtained from the receiver processing vector $\mathbf{w}_k$ or matrix $\mathbf{W}_k$ in the equivalent detection of

$$z_k = \mathbf{w}_k^H \left(\mathbf{y} = \sum_{l \in \mathcal{K}} \mathbf{h}_l x_l + \mathbf{n} \right) \quad (35a)$$

$$\tilde{z}_k = \mathbf{W}_k^H \left(\tilde{\mathbf{y}} = \sum_{l \in \mathcal{K}} \mathbf{H}_l \Phi_l \mathbf{x}_l + \tilde{\mathbf{n}} \right) \quad (35b)$$

yielding $\mathbf{p}_k = \sqrt{\Delta_k} \mathbf{w}_k^*$ and $\mathbf{P}_k = \mathbf{W}_k^* \Delta_k^{1/2}$, where $\{\Delta_k\}$ and the diagonal matrices $\{\Delta_k\}$, joining with A_0 , account for the power-allocation to different UEs and also to the symbols sent on different subcarriers of a UE. Hence, based on (35)(a), we have

$$\mathbf{p}_k = \sqrt{\Delta_k} \underbrace{\left(\sum_{l \in \mathcal{K}} \mathbf{h}_l^* \mathbf{h}_l^T + \sigma^2 \mathbf{I}_M \right)^{-1}}_{(\mathbf{R}_y^*)^{-1}} \mathbf{h}_k^*, \quad k \in \mathcal{K} \quad (36)$$

in MMSE principle, where $\mathbf{R}_y$ is the autocorrelation matrix of $\mathbf{y}$. The MMSE detection problem in (35)(b) has been analyzed with the GMMSE detector in Section V-A1. After ignoring the power-control coefficients in (19), we have

$$\mathbf{P}_k = \underbrace{\left(\sum_{l \in \mathcal{K}} \mathbf{H}_l^* \Phi_l \Phi_l^T \mathbf{H}_l^T + \sigma^2 \mathbf{I}_{MN} \right)}_{(\tilde{\mathbf{R}}_y^*)^{-1}}^{-1} \mathbf{H}_k^* \Delta_k^{1/2}, \quad k \in \mathcal{K} \quad (37)$$

where $\tilde{\mathbf{R}}_y$ is the autocorrelation matrix of $\tilde{\mathbf{y}}$.

Note that, in (36) and (37), the channel vectors (or matrices) are depended on the large-scale propagation pathloss and shadowing, as well as the small-scale fading. Hence, the transmitter preprocessing can naturally take their effect into account. To achieve the best possible performance in terms of mean-square error (MSE), a stronger signal will be weighed by a bigger factor, while a stronger interference draws more attention of the precoder for its suppression.

Furthermore, it is worth noting that, while the preprocessing vectors $\{\mathbf{p}_k\}$ or matrices $\{\mathbf{P}_k\}$ are computed at CPU in a global optimization way, they are however transmitted by the distributed APs under their local constraints, such as, APs' transmit power. Specifically for UE k , AP m only transmits p_{mk} in $\mathbf{p}_k$ or $\mathbf{P}_{mk}$ in $\mathbf{P}_k$. Therefore, $\{\Delta_k\}$ in (36) or $\{\Delta_k\}$ in (37) are just power-allocation coefficients. In order for $\mathbf{p}_k$ or $\mathbf{P}_k$ to maintain its properties when it goes through the channels to UE k , $\mathbf{p}_k$ or $\mathbf{P}_k$ must be linearly amplified by the M distributed APs. Specifically considering $\mathbf{p}_k$, this means that a same positive scalar A_0 should be multiplied on the M different elements of $\mathbf{p}_k$ by the M distributed APs, so as to protect $\mathbf{p}_k$ from any distortion. Without any doubt, this is highly challenging in practical implementation.

Upon substituting (36) into (34)(a), the SINR of UE k 's detection can be derived to be

$$\gamma_k(\{\Delta_k\}) = \frac{|\mathbf{h}_k^T \mathbf{p}_k|^2}{\sum_{l \neq k} |\mathbf{h}_k^T \mathbf{p}_l|^2 + \sigma^2/A_0^2} \quad (38)$$

which is a function of the power-allocation coefficients $\{\Delta_k\}$. Similarly, when specifically considering the i th symbol of UE k , from (34)(b), we can obtain the SINR of

$$\gamma_{ki}(\{\Delta_k\}, \{\Phi_l\}) = \frac{|\mathbf{h}_{ki}^T \mathbf{P}_k \phi_{ki}|^2}{\sum_{j \neq i} |\mathbf{h}_{ki}^T \mathbf{P}_k \phi_{kj}|^2 + \sum_{l \neq k} \sum_{j=1}^{N_l} |\mathbf{h}_{ki}^T \mathbf{P}_l \phi_{lj}|^2 + \sigma^2/A_0^2} \quad (39)$$

which is a function of both the subcarrier-allocation reflected by $\{\Phi_l\}$ and the power-allocation interpreted by $\{\Delta_k\}$.

Hence, the DL sum-rate of UCCF network is

$$R(\{\Delta_k\}) = \sum_{k \in \mathcal{K}} \log_2 [1 + \gamma_k(\{\Delta_k\})] \quad (40a)$$

$$R(\{\Delta_k\}, \{\Phi_l\}) = \sum_{k \in \mathcal{K}} \sum_{i=1}^{N_k} \log_2 [1 + \gamma_{ki}(\{\Delta_k\}, \{\Phi_l\})] \quad (40b)$$

The above analysis assumes an idealized network, where each AP sends signals to all UEs. In a practical UCCF network, usually only a few of APs that a UE is associated with receive signals from and send signals to the UE. Hence, (36) and (37) must be modified. Accordingly, the operations may be summarized as follows.

- 1) $\mathbf{R}_y$ in (36) or $\tilde{\mathbf{R}}_y$ in (37) can be estimated from the UL signals forwarded by APs to CPU. Alternatively, if APs only forward their estimated channels of the associated UEs to CPU, then, CPU computes an approximate for $\mathbf{R}_y$ or for $\tilde{\mathbf{R}}_y$ using the estimated channels by setting all the other unknown elements to zeros.
- 2) After $\mathbf{R}_y$ or $\tilde{\mathbf{R}}_y$ is obtained, CPU computes the preprocessing vector $\mathbf{p}_k$ using (36) or matrix $\mathbf{P}_k$ using (37) for all $k \in \mathcal{K}$, without considering the optimization of power-allocation coefficients. Note that in the calculation of $\mathbf{p}_k$ or $\mathbf{P}_k$, $\mathbf{h}_k$ or $\mathbf{H}_k$ only includes the estimated elements related to the APs that UE k is associated with.
- 3) CPU optimizes the power-allocation coefficients of $\{\Delta_k\}$ or that in $\{\mathbf{\Delta}_k\}$, which will be further discussed in Section VI-B.
- 4) CPU computes s_m using (32)(a), or $\tilde{s}_m$ using (32)(b) for all $m \in \mathcal{M}$.
- 5) For $m = 1, 2, \dots, M$, CPU sends s_m or $\tilde{s}_m$ to AP m , where s_m or $\tilde{s}_m$ is sent to DL after the same amplification by a gain A_0 informed by CPU.

As above-mentioned, the TMMSE precoder designed by the central CPU but used for signal transmission by distributed APs is challenging in practical implementation. To ease the implementation challenges, distributed preprocessing and transmission by APs may be employed. Specifically considering the subcarrier-level precoding, once the CSI is available for DL transmission, AP m can send the signals as

$$s_m = \sum_{k \in \mathcal{K}_m} \sqrt{P_{mk}} \left(\frac{h_{mk}^*}{|h_{mk}|} \right) x_k, \quad m \in \mathcal{M} \quad (41)$$

where P_{mk} is AP m 's transmit power towards UE k . When all the M APs transmit signals synchronously and coherently, the received signal by UE k is

$$\begin{aligned} y_k = & \sum_{m \in \mathcal{M}_k} \sqrt{P_{mk}} |h_{mk}| x_k + \sum_{m \in \mathcal{M}_k} \sum_{l \in \mathcal{K}_m, l \neq k} \sqrt{P_{ml}} \left(\frac{h_{mk} h_{ml}^*}{|h_{ml}|} \right) x_l \\ & + \sum_{m \in \bar{\mathcal{M}}_k} \sum_{l \in \mathcal{K}_m} \sqrt{P_{ml}} \left(\frac{h_{mk} h_{ml}^*}{|h_{ml}|} \right) x_l + n_k \end{aligned} \quad (42)$$

where on the right-hand side, the first term is the desired signal, the second term is the MUI generated by the UEs associated with the APs that UE k is also associated with, the third term is the MUI from the APs other than that in $\mathcal{M}_k$, and n_k is noise. Explicitly, the signal received by UE k experiences strong interference, provided that UE k shares some APs with other UEs.

Hence, additional resources are required to mitigate the interference seen in (42). One approach is to equip each AP with multiple antennas, the number of which is bigger than

the number of UEs associated with it in the normal operational scenarios. In this case, (41) can be modified to

$$\mathbf{s}_m = \sum_{k \in \mathcal{K}_m} \sqrt{P_{mk}} \left(\frac{\mathbf{p}_{mk}^*}{\sqrt{|\mathbf{p}_{mk}|^2}} \right) x_k, \quad m \in \mathcal{M} \quad (43)$$

Assume $U(\geq |\mathcal{K}_m| \quad \forall m)$ antennas per AP. Then, in (43), $\mathbf{p}'_{mk} = \mathbf{p}_{mk}^* / \sqrt{|\mathbf{p}_{mk}|^2}$ is a U -length preprocessing vector for transmitting x_k . Let the channel vector from UE k to AP m be expressed as $\mathbf{h}_{mk}$. Let express $\mathbf{H}_m = [\mathbf{h}_{m1}, \mathbf{h}_{m2}, \dots, \mathbf{h}_{m|\mathcal{K}_m|}]$. Then, in the principles of transmitter zero-forcing (TZF) [17], the preprocessing vector $\mathbf{p}_{mk}$ for UE k is the k th column of

$$\mathbf{P}_m = \mathbf{H}_m^* (\mathbf{H}_m^T \mathbf{H}_m^*)^{-1} \quad (44)$$

Then, corresponding to (42), the signal received by UE k is

$$y_k = \sum_{m \in \mathcal{M}_k} \sqrt{P_{mk}} x_k + \sum_{m \in \mathcal{M}_k} \sum_{l \in \mathcal{K}_m} \sqrt{P_{ml}} \left(\frac{\mathbf{h}_{mk}^T \mathbf{p}_{ml}^*}{\sqrt{|\mathbf{p}_{ml}|^2}} \right) x_l + n_k \quad (45)$$

It shows that the signals sent by AP m with $m \in \mathcal{M}_k$ do not interfere UE k . UE k still experiences MUI from the UEs that do not share any APs with UE k . However, these interfering UEs as well as the APs that these UEs are associated with should be located relative far away from UE k . Consequently, their interference on UE k is insignificant.

Instead of TZF, TMMSE-based precoding can be implemented by the distributed APs to transmit signals to their associated UEs. Accordingly, the precoding vectors for AP m to send signals to its associated UEs can be obtained from the columns of [17]

$$\mathbf{P}_m = \mathbf{H}_m^* (\mathbf{H}_m^T \mathbf{H}_m^* + \sigma^2 \mathbf{I}_{|\mathcal{K}_m|})^{-1} \quad (46)$$

where σ^2 is noise variance.

In the case that the noise variance σ^2 in (46) is not available at APs, it can be replaced by a regulation parameter ρ , which can be optimized to strike a trade-off between noise mitigation and MUI suppression [17]. Specifically, when $\rho = 0$, the precoder is reduced to the TZF precoder of (44), which may amplify noise in low SNR region, but is capable of fully removing MUI. When $\rho = \sigma^2$, it achieves the TMMSE precoding, which has the capability to suppress noise, while possibly leaving a little MUI in the detection. Furthermore, when $\rho \rightarrow \infty$, the precoder becomes a transmitter matched-filtering (TMF) assisted precoder. This can be understood from (46) that the precoding vectors are given by the columns of $\mathbf{P}_m = \mathbf{H}_m^*$, when $\rho \rightarrow \infty$. As TMMSE, TMF precoding has the capability to mitigate noise, but it treats MUI as noise and hence, can hardly achieve a satisfactory MUI suppression result, especially, in the high SNR region where MUI dominates.

B. Downlink Resource Optimization

Corresponding to the problem of UL optimization, as described in (31), a general problem for the DL optimization to maximize the sum-rate of an OFDM-UCCF network can be formulated as

$$\begin{aligned} & \{\zeta_{mk}^*\}, \{\delta_{nk}^*\}, \{\Delta_{kn}^*\} \\ & = \arg \max_{\{\zeta_{mk}\}, \{\delta_{nk}\}, \{\Delta_{kn}\}} \left\{ \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \delta_{kn} \log_2 (1 + \gamma_{kn}(\{\zeta_{mk}\}, \{\Delta_{kn}\})) \right\} \end{aligned} \quad (47a)$$

$$s.t. \quad \zeta_{mk} \in \{0, 1\}, \quad \forall m \in \mathcal{M}, k \in \mathcal{K} \quad (47b)$$

$$\delta_{kn} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, n \in \mathcal{N} \quad (47c)$$

$$\sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \Delta_{kn} \leq 1 \quad (47d)$$

$$\|\tilde{\mathbf{s}}_m\|^2 \leq P_m^{\max}, \quad \forall m \in \mathcal{M} \quad (47e)$$

$$A_0 : |\tilde{\mathbf{s}}_m(n)|^2 \leq P_m^{\max}(n), \quad \forall m \in \mathcal{M}, n \in \mathcal{N} \quad (47f)$$

$$\sum_{n \in \mathcal{N}} \delta_{kn} \log_2 (1 + \gamma_{kn}(\{\zeta_{mk}\}, \{\Delta_{kn}\})) \geq R_k^{\min}, \quad \forall k \in \mathcal{K} \quad (47g)$$

where (47)(b), (47)(c) and (47)(g) have the same meaning as (31)(b), (31)(c) and (31)(e), respectively, in (31), (47)(e) is the total power constraint on an AP, while (47)(f) guarantees that the precoding vectors computed by CPU can be linearly amplified by A_0 and sent by the distributed APs. Note that in (47)(f), $|\tilde{\mathbf{s}}_m(n)|^2$ is the power of the n th element of $\tilde{\mathbf{s}}_m$, and $P_m^{\max}(n)$ is the maximum power allowed to transmit this element. This constraint means that a common amplification gain A_0 implemented by the distributed APs should satisfy the power constraints of $|\tilde{\mathbf{s}}_m(n)|^2 \leq P_m^{\max}(n)$, $\forall m \in \mathcal{M}, n \in \mathcal{N}$. Finally, (47)(d) imposes the constraint on the power-allocation coefficients assigned to the precoding vectors.

Notice in (31) and (47) that the UE association is considered independently in both UL and DL optimization. In reality, it should be optimized by jointly considering UL and DL. Nevertheless, both (31) and (47) are NP hard to solve, and the UE association is usually optimized separately before carrying out the other UL and DL optimization.

As the UL optimization, the DL optimization also needs to be divided into some sub-problems of, such as, UE association, subcarrier-allocation and power-allocation, which may be optimized in succession by various optimization algorithms [3, 13], from classic methods [8] to the more advanced deep learning and heterogeneous graph neural network [9] methods. In comparison with UL optimization, in the UCCF networks, the power-allocation in DL optimization is more challenging, as the precoders are globally computed by CPU but locally implemented by the distributed APs. To illustrate this further, let us consider the DL optimization in the relatively simpler subcarrier-level DL transmission scheme, which has the precoding vectors as shown in (36). The optimization problem corresponding to (47) can be

formulated as

$$\{\zeta_{mk}^*, \{\Delta_k^*\} = \arg \max_{\{\zeta_{mk}\}, \{\Delta_k\}} \left\{ \sum_{k \in \mathcal{K}} \log_2 (1 + \gamma_k (\{\zeta_{mk}\}, \{\Delta_{kn}\})) \right\} \quad (48a)$$

$$s.t. \quad \zeta_{mk} \in \{0, 1\}, \quad \forall m \in \mathcal{M}, k \in \mathcal{K} \quad (48b)$$

$$\sum_{k \in \mathcal{K}} \Delta_k \leq 1 \quad (48c)$$

$$A_0 : |s_m|^2 \leq P_m^{\max}, \quad \forall m \in \mathcal{M} \quad (48d)$$

$$\log_2 (1 + \gamma_k (\{\zeta_{mk}\}, \{\Delta_k\})) \geq R_k^{\min}, \quad \forall k \in \mathcal{K} \quad (48e)$$

where (48)(c) imposes the constraints on the optimization of power-allocation coefficients, while (48)(d) explains that the common amplification gain A_0 should allow the M APs to simultaneously satisfy their power constraints. When these conditions are met, $\{p_k\}$ in (36) will be linearly amplified using a scalar gain A_0 by the distributed APs. Hence, their properties and relationships will be retained after the transmitted signals go through the DL channels to UEs, yielding the performance as expected. However, in practice, this is challenging to achieve. This is because distributed APs are operated with separated amplifiers and oscillators, which are hard to be controlled to operate with the same gain and in a nearly ideal synchronization state.

VII. CONCLUDING REMARKS

In this chapter, the principles of channel training, UL detection and DL precoding in UCCF networks have been analyzed. Specifically, on channel training, the UCCF networks operated with TDD, IBFD and MDD have been discussed, showing that a large overhead may be required by a TDD-based UCCF network for channel estimation. Furthermore, the signal transmission in TDD-based UCCF networks experiences the channel ageing problem, which may significantly degrade performance. By contrast, owing to the simultaneous UL/DL transmissions, IBFD- or MDD-aided UCCF systems are free from the channel ageing problem, while the overhead for channel training can be significantly reduced, when compared with the TDD-relied UCCF systems. Between IBFD and MDD, the SI in IBFD-based UCCF systems may limit the channel estimation to achieve high reliability, but MDD is free from this limitation.

On UL detection, the GMMSE detection operated at CPU, LMMSE detection operated at CPU or by the cooperative APs, and the APPA detection carried out by the cooperative APs with the aid of CPU have been analyzed. It can be expected that the APPA detector is a high-efficiency detection scheme: it does not require strict synchronization among APs, information exchange only occurs between adjacent APs, there is little backhaul transmission between APs and CPU, and it allows to achieve near-optimum performance.

On DL transmission, the principle of precoding has been analyzed, when the transmitter preprocessing at both the subcarrier level and OFDM symbol level is considered. Both the

global precoder design by CPU and the distributed precoder design by APs, as well as the implementation issues of DL transmission and optimization have been addressed. The analysis reveals that, while the precoder design for UCCF networks has little difference from that for the classic MIMO, in particular, SDMA or massive MIMO, systems, the implementation of DL transmission is highly challenging. This challenge comes mainly from the requirement that the centrally processed signals by CPU are transmitted by the distributed APs, which may impose random distortions, resulting in that the signals received from different APs by a UE are not coherently added together. To ease this challenge, distributed AP precoding may be an alternative, but extra resources, e.g., multiple antennas, are required by APs to mitigate MUI.

Furthermore, in this chapter, both the UL and DL optimization issues have been discussed, showing that the optimization, especially the DL optimization, is more demand than that in the conventional BS centric systems. For example, in the BS centric systems, BS carries out the DL optimization under the constraint of its own transmit power. It transmits the DL signals under the control of a same oscillator. By contrast, in a UCCF network, the DL optimization may be done by CPU but under the power constraints of distributed APs. Moreover, the optimized signals by CPU are required to be linearly amplified by the distributed APs and sent from them in nearly ideal synchronization, which are difficult to implement in practice, when considering that distributed APs are operated with independent amplifiers and oscillators.

While the implementation of DL transmission is challenging in the UCCF networks, where CPU carries out the centralized global optimization but distributed APs are responsible for signal transmission, such regularized transmission strategy does provide an interesting method for CPU to send secrecy information to UEs. With this regard, first, the information to UEs is secret to APs. This can be inferred, for example, by (36) and (32)(a), where the preprocessing vectors $\{\mathbf{p}_k\}$ in (36) are computed by CPU, but AP m sends the combination of the m th elements in $\{\mathbf{p}_k\}$ and the data symbols in $\mathcal{K}_m$. Hence, AP m only knows the combination of the m th elements in $\{\mathbf{p}_k\}$ and the data symbols in $\mathcal{K}_m$, not know the data and also the other elements in $\{\mathbf{p}_k\}$. As shown in (36), $\{\mathbf{p}_k\}$ are dependent on the channels between different UEs and different APs. Since AP m only knows the channels between it and its associated UEs, it should be very difficult for AP m to derive the information sent to its associated UEs, needless to say the information sent to the other UEs.

Second, the information sent to a UE is secret to the other UEs. This can be conceived, for example, from (34)(a). When TZF or TMMSE precoder is employed, UE k is only capable of picking up its own information. The information sent to the other UEs is either fully removed, when TZF precoding is employed, or mostly removed, when TMMSE precoding is employed.

Third, the situation cannot be better for eavesdroppers. First, eavesdroppers are hard to obtain the CSI between APs and UEs. Second, the precoders are designed to achieve coherent receiving at the locations of UEs. Since eavesdroppers usually have different locations from UEs', it is impossible for them to attain the receiving effect as (34)(a) for the legitimate UEs.

Hence, the information sent to UEs is normally secret to eavesdroppers. Perhaps, an active eavesdropper may use a powerful receiver, such as, by employing multiple receive antennas, to eavesdrop. In this case, an artificial interference may be added to the transmitted signal, in the form of

$$\mathbf{s} = \sqrt{\rho} \sum_{l \in \mathcal{K}} \mathbf{p}_l x_l + \sqrt{1 - \rho} \mathbf{p}_I n_I \quad (49)$$

where n_I is the random artificial interference, $\mathbf{p}_I$ is designed to be orthogonal with all the channel vectors from UEs to APs, i.e., $\mathbf{h}_k^T \mathbf{p}_I = 0, \forall k \in \mathcal{K}$, and $\rho \in [0, 1]$ controls the power assigned to transmit information and artificial interference, respectively, which is an optimization parameter. Then, AP m sends

$$s_m = A_0 \left(\sqrt{\rho} \sum_{l \in \mathcal{K}} p_{ml} x_l + \sqrt{1 - \rho} p_{mI} n_I \right), \quad m \in \mathcal{M} \quad (50)$$

The received signal by UE k is

$$y_k = A_0 \sqrt{\rho} \mathbf{h}_k^T \sum_{l \in \mathcal{K}} \mathbf{p}_l x_l + n_k \quad (51)$$

which, except the power scaling by ρ , is the same as (34)(a). However, owing to the added artificial interference $p_{mI} n_I$, the signals sent to UEs should become harder to decode by the distributed APs and eavesdroppers.

REFERENCES

- [1] L.-L. Yang, "Performance of mmse multiuser detection in cellular ds-cdma systems using distributed antennas," in *2006 IEEE 63rd Vehicular Technology Conference*, vol. 1, 2006, pp. 274–278.
- [2] L.-L. Yang and W. Fang, "Performance of distributed-antenna DS-CDMA systems over composite lognormal shadowing and Nakagami- m -fading channels," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2872–2883, 2009.
- [3] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "User-centric cell-free massive MIMO networks: A survey of opportunities, challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 611–652, 2022.
- [4] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communications*. New York: Addison-Wesley Publishing Company, 1995.
- [5] L. Hanzo, L.-L. Yang, E.-L. Kuan, and K. Yen, *Single- and Multi-carrier DS-CDMA*. Chichester, United Kingdom: John Wiley and IEEE Press, 2003.
- [6] P. Pan, J. Zhang, and L.-L. Yang, "Massive distributed antenna systems: Channel estimation and signal detection," *IEEE Access*, vol. 8, pp. 186 055–186 070, 2020.
- [7] B. Li, L.-L. Yang, R. G. Maunder, P. Xiao, and S. Sun, "Multicarrier-division duplex: A duplexing technique for the shift to 6G wireless communications," *IEEE Vehicular Technology Magazine*, vol. 16, no. 4, pp. 57–67, 2021.
- [8] B. Li, L.-L. Yang, R. G. Maunder, S. Sun, and P. Xiao, "Spectral-efficiency of cell-free massive MIMO with multicarrier-division duplex," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 3, pp. 3404–3418, 2023.
- [9] —, "Heterogeneous graph neural network for power allocation in multicarrier-division duplex cell-free massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 23, no. 2, pp. 962–977, 2024.
- [10] B. Li, D. Dupleich, G. Xia, H. Zhou, Y. Zhang, P. Xiao, and L.-L. Yang, "Mdd-enabled two-tier terahertz fronthaul in indoor industrial cell-free massive MIMO," *IEEE Transactions on Communications*, pp. 1–1, 2023.
- [11] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. New York: John Wiley & Sons, 2005.

- [12] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 706–709, 2017.
- [13] S. Elhoushy, M. Ibrahim, and W. Hamouda, "Cell-free massive MIMO: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 492–523, 2022.
- [14] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No.01CH37202)*, vol. 1, 2001, pp. 333–336 vol.1.
- [15] B. Li, L.-L. Yang, R. G. Maunder, S. Sun, and P. Xiao, "Multicarrier-division duplex for solving the channel aging problem in massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 1940–1954, 2023.
- [16] G. Liu, F. R. Yu, H. Ji, V. C. M. Leung, and X. Li, "In-band full-duplex relaying: A survey, research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 500–524, 2015.
- [17] L.-L. Yang, *Multicarrier Communications*. Chichester, United Kingdom: John Wiley, 2009.
- [18] B. Li, L.-L. Yang, R. G. Maunder, and S. Sun, "Self-interference cancellation and channel estimation in multicarrier-division duplex systems with hybrid beamforming," *IEEE Access*, vol. 8, pp. 160 653–160 669, 2020.
- [19] —, "Resource allocation in millimeter-wave multicarrier-division duplex systems with hybrid beamforming," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7921–7935, 2021.
- [20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, New Jersey: Prentice Hall, Inc., 1993.
- [21] J. G. Proakis, *Digital Communications*, 5th ed. McGraw Hill, 2007.
- [22] W. E. Ryan and S. Lin, *Channel Codes: Classical and Modern*. New York: Cambridge University Press, 2009.
- [23] X. Wang and H. V. Poor, *Wireless Communication Systems - Advanced Techniques for Signal Reception*. Prentice Hall, 2003.
- [24] Y. Liu, L.-L. Yang, and L. Hanzo, "Spatial modulation aided sparse code-division multiple access," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1474–1487, 2018.
- [25] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 3, pp. 651–663, 2019.
- [26] L.-L. Yang, "Capacity and error performance of reduced-rank transmitter multiuser preprocessing based on minimum power distortionless response," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4646–4655, 2008.
- [27] —, "Joint transmitter-receiver design in TDD multiuser MIMO systems: An egocentric/altruistic optimization approach," in *2007 IEEE 65th Vehicular Technology Conference - VTC2007-Spring*, 2007, pp. 2094–2098.
- [28] —, "Multiuser transmission via multiuser detection: Altruistic-optimization and egocentric-optimization," in *2007 IEEE 65th Vehicular Technology Conference - VTC2007-Spring*, 2007, pp. 1921–1925.
- [29] S. Chen and L.-L. Yang, "Downlink MBER beamforming transmitter based on uplink MBER beamforming receiver for TDD-SDMA MIMO systems," in *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*, 2009, pp. 433–436.