IEEE Open Journal of Vehicular Technology

Received 16 August 2020; accepted 26 August 2020. Date of publication 28 August 2020;
date of current version 14 September 2020. The review of this article was arranged by Associate Editor D. W. K. Ng.

Digital Object Identifier 10.1109/OJVT.2020.3020228

Channel Estimation and User Activity Identification in Massive Grant-Free Multiple-Access

JIATIAN ZHANG [1], PENG PAN [2] (Senior Member, IEEE), LIE-LIANG YANG [1] (Fellow, IEEE), AND ROBERT G. MAUNDER [1] (Senior Member, IEEE)

(Invited Paper)

School of Electronics, and Computer Science, University of Southampton, SO17 1BJ, U.K.
School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, Zhejiang, China

CORRESPONDING AUTHOR: LIE-LIANG YANG (e-mail: lly@ecs.soton.ac.uk)

This work was supported in part by the Engineering, and Physical Sciences Research Council Project EP/P034284/1, and in part by the Innovate UK's Knowledge Transfer Partnership Project KTP011036.

ABSTRACT Grant-free multiple-access (GFMA) allows to significantly reduce the overhead of granted multiple-access. However, information detection in GFMA is challenging, as it has to be executed along with the activity detection of user equipments (UEs) and channel estimation. In this paper, we study the channel estimation and propose the UE activity identification (UAI) algorithms for the massive connectivity supporting GFMA (mGFMA) systems. For these purposes, the channel estimation is studied from several aspects by assuming different levels of knowledge to the access point, and based on which five UAI approaches are proposed. We study the performance of channel estimation, the statistics of estimated channels, and the performance of UAI algorithms. Our studies show that the proposed approaches are capable of circumventing some of the shortcomings of the existing techniques designed based on compressive sensing and message passing algorithms. They are robust for operation in the mGFMA systems where the active UEs and the number of them are highly dynamic.

INDEX TERMS Eigen analysis, grant-free multiple-access, interference cancellation, internet of things, massive connectivity, massive machine-type communications, MC-mGFMA, MMSE, user identification.

I. INTRODUCTION

So far, in wireless communications systems, a wireless terminal connecting its access-point (AP) or base-station (BS) is typically achieved via granted multiple-access (GMA), where a handshaking process between wireless terminal and AP is operated to fulfill the functions of multiple-access. GMA has shown its advantages in the conventional wireless communication systems, from the 1st generation (1G) to the 4th generation (4G), where traffic is mainly activated by human being. However, starting from the 5th generation (5G), wireless systems are rendered to be device (machine) centric, with the traffic initialized by numerous devices (machines). As the result, the characteristics of the traffic in 5G and beyond wireless systems will be very different from that in the 1G to 4G wireless systems. In these future wireless systems, there may be various types of sporadic short-burst traffic generated, for example, by sensing nodes and IoT devices of many kinds, nodes for sending control commands, etc. [1]. Explicitly, transmission of these kinds of sporadic short-burst traffic using the traditional communications schemes, such as that in 4G LTE/LTE-A systems, is often inefficient, due to their reliance on a huge amount of overhead for signaling and control, yielding an extremely high overhead-to-data ratio as well as introducing extra communication latency.

In order to improve the efficiency of the transmission of sporadic traffic, save the overhead for corresponding network maintenance and to support massive connectivity, in recent years, various grant free multiple-access (GFMA) schemes have been proposed and investigated [2]–[4]. In GFMA, a user equipment (UE), which may be a device, machine, etc.,

becomes active randomly to transmit data to its AP simply under the control of an assigned random access protocol, such as slotted ALOHA, without requesting for a grant from its AP [3], [5]. GFMA has the advantage of greatly saving the resources that are otherwise consumed for the extra overhead. However, it also imposes new challenges for AP to recover the data sent by active UEs. This becomes even more significant when GFMA is employed to support massive connectivity, forming the massive GFMA (mGFMA). In mGFMA, each UE usually only has a small probability to become active at a time(-slot). Hence, active UEs as well as the number of them are highly dynamic, both of which are usually not *a-priori* knowledge to AP. Therefore, in order to recover the data sent by active UEs, an AP has to accomplish first the functions of active UE identification and the channel estimation for active UEs, or implement them in joint ways [4], [6]–[12].

The investigation of UE activity identification (UAI) in random-access scenarios was already started along with the conventional code-division multiple-access (CDMA) systems [13]–[16]. Recently, it has drawn a lot of research attention with the mGFMA in IoT networks and in particular, in massive machine-type communications (mMTC), where UAI is often jointly implemented with channel estimation and even data detection [6], [8]–[10]. As above-mentioned, in mGFMA systems, the number of potential UEs supported by an AP is usually very big, up to thousands, tens of thousands or even more, while each UE only has a very small probability that is normally less than 10% to become active. Consequently, when without the knowledge of active UEs, signal detection in mGFMA is naturally a compressive sensing (CS) relied signal recovery problem. Owing to this, in recent years, various CS-based joint UAI, channel estimation and/or multiuser detection (MUD) algorithms have been developed and investigated in the context of mGFMA, when various sparsity structures are considered [6], [7], [10], [11], [17]–[28]. While CS-based methods have some outstanding merits as claimed in references, they are not appearing for operation in the mGFMA systems where the number of active UEs is large [29]. This is because the recovery performance of CS-based methods is limited by the restricted isometric property (RIP) condition [30]. This condition implies that the number of active UEs (sparsity) should be significantly lower than the number of resource units (subcarriers, antennas, spreading factor, time-slots, etc.) of an AP, as seen, e.g., in [8], [10], [18], [22], [23], [25], and hence resulting in a low efficiency of resource usage. Furthermore, in mGFMA systems, the number of active UEs is highly dynamic, making the systems often *real-time* overload[1] and substantially violating the RIP condition. For example, for an AP using $N = 32$ resource units to support $K = 1000$ potential UEs of each having an independent activation probability of 5%, the overloading probability is $P(> N) = 0.9963$. Explicitly, the CS-based methods are not suitable for operation in this mGFMA system, and other efficient approaches are required for the UAI, channel estimation and data detection.

In mGFMA, the number of potential UEs is usually significantly higher than the number of orthogonal resource units. Hence, mGFMA is generally implemented in the principles of non-orthogonal multiple-access (NOMA). In the NOMA family, sparse-code multiple-access (SCMA) is an attractive scheme, which allows to support the multiple-access of significant overloading but still achieve near optimum performance by exploiting the message-passing algorithm (MPA)-assisted detection [31]. Hence, SCMA has been introduced to GFMA, e.g., in [2], [12], [32]–[34]. Specifically in [35], the joint UE activity detection and data detection were considered, when assuming that all UEs (both active and inactive) transmit the headers for activity detection, while the channels of all UEs are known to BS. Considering an uplink GFMA, [12] investigated the MPA-assisted receiver design for joint channel estimation and signal detection. In [32], a faster-than-Nyquist (FTN) singling SCMA was proposed to support uplink GFMA, where only very small systems (6 UEs supported by 4 resource units) were investigated. In [33], the authors considered the MPA-assisted information detection in a rateless SCMA-based GFMA systems, when assuming a constant number of active UEs, and that receiver has ideal knowledge about the active users and their channels. While one UE employs only one codebook in the above-mentioned references, the authors in [34] proposed to map a UE's consecutive data symbols to the codewords chosen from different codebooks, which achieves code diversity. In general, SCMA achieves the near optimum performance depending on the MPA-assisted detection operated on *good* factor graphs. However, when we consider the mGFMA system in which an AP supports a huge number of potential UEs, while it has no knowledge of active UEs, the number of them and of their channels, maintaining good factor graphs in such highly dynamic communication scenarios is extremely challenging. In addition to the SCMA-based GFMA, a semi-grant-free power-domain NOMA based transmission scheme was proposed in [36], where one UE is operated as a primary UE under a conventional grant-based protocol, while the other secondary UEs opportunistically access the same channel in grant-free manner.

In addition to the above-mentioned, GFMA and mGFMA have also been investigated by invoking other techniques. In [19], the authors enhanced the CS-based approach by proposing a greedy algorithm relied on the maximum a-posteriori probability (MAP) criterion, which performs UE identification and data detection jointly by exploiting the a-posteriori probabilities of each other. An expectation propagation based joint UE activity detection and channel estimation scheme was introduced in [37] for achieving GFMA in mMTC networks. In [38], the authors investigated the preamble design for UE detection and channel estimation, so as to attain a high success rate in grant-free random access in massive multiple-input multiple-output (MIMO) systems.

[1]Here, *real-time* overload means that the number of active UEs is more than the number of resource units. By contrast, the overload claimed in some references considering CS-assisted GFMA means that the number of potential UEs is higher than the number of resource units. In the rest of this paper, we mean overload the real-time overload, unless otherwise notified.

The authors of [39] compared two types of preamble structures on the performance of grant-free random access massive MIMO. Similarly, GFMA in massive MIMO systems was considered in [40], where a fixed number of UEs access one of the available channels and orthogonal preambles are transmitted for UE activity identification and channel estimation. Furthermore, the GFMA massive MIMO scenario was considered in [41], where the number of active UEs are first estimated, followed by the joint detection of active UE and their messages with the aid of a so-called ensemble independent component analysis decoding algorithm, without relying on explicit channel estimation. Reed-muller sequences were introduced in [42] to aid UE identification and channel estimation in the GFMA systems supporting massive connectivity. In [43], UE activity and signal detection were studied with mGFMA by designing preambles and exploiting the interleave-division multiple access (IDMA), when user signals are randomly and asynchronously transmitted. More recently, machine-learning (ML) approaches have been introduced to mGFMA [22], [28], [44]–[46]. To be more specific, the authors of [22] investigated the joint UE activity detection and channel estimation by formulating them as a block sparse signal recovery problem, which is solved by a block sparse Bayesian learning (BSBL) algorithm. In [44], the asynchronous sparse Bayesian learning algorithm was adopted for channel estimation while the support vector machine method was applied for UE activity detection, when the number of active users is assumed to be known to BS. To alleviate the convergence problem of the BSBL algorithm [22], a deep neural network-aided message passing-based block sparse Bayesian learning algorithm was proposed in [45] to achieve the joint UE activity detection and channel estimation in mGFMA scenarios. The joint optimization of finite-alphabet spreading sequences and multi-user detection were addressed in [46], where deep learning principle was introduced for the design of both encoder and decoder. Furthermore, in [28], deep learning was used to predict the activities of UEs, followed by a CS algorithm for data detection and further enhancing the UE activity identification. While nearly all the above-mentioned works considered the grant-free access techniques at physical layer, the authors of [47] investigated the GFMA from the MAC layer perspective in the IoT systems with sporadic traffic.

In this paper, we investigate the channel estimation and UAI in the mGFMA scenario, where the number of potential UEs may be much bigger than the number of resource units and each UE has a small activation probability. Hence, the active UEs and the number of them are highly dynamic over time, and the number of active UEs at a time may be higher than the number of resource units. Specifically, we first consider the channel estimation of both active and inactive UEs in the principle of minimum mean-square error (MMSE), when AP is assumed to have the ideal knowledge, no knowledge or the partial knowledge of active UEs. In our channel estimation, AP will make the best use of the information provided by the transmission of pilot symbols and data (payload) symbols. Our studies show that any added knowledge about the active UEs can be exploited to enhance the reliability of channel estimation. This is reflected in practical mGFMA systems that the activity patterns of UEs may be correlated or one active UE may have several packets to transmit in continuous time-slots. Then, this kind of information can be exploited to enhance the reliability of channel estimation in mGFMA systems.

Following the channel estimation, we then study the statistical properties of the estimated channels of both active and inactive UEs. We explicitly show that the estimated channels of active UEs usually have relatively good mean-square error (MSE) performance, which improves for example, with the increase of signal-to-noise ratio (SNR), of the number of pilot symbols, etc. By contrast, as there are no transmissions from inactive UEs, the estimated channels for them always have a high MSE resulted from noise and the interference from active UEs, which is also insensitive to the investment of resources, such as, transmit power, pilot symbols, etc. We also study the distribution of estimated channel's power, showing that the estimated channel's power of active UEs is usually much higher than that of inactive UEs, when SNR is in the range of practical applications.

Finally, by exploiting the channel estimation and the characteristics of estimated channels, we propose a range of UAI algorithms, in which some of them make use of the estimated channel's knowledge, while the others carry out joint channel estimation and UAI. In detail, the first one is the simplest threshold-based UAI (TB-UAI) algorithm, which however makes a sharp trade-off between false-alarm and miss identification, and is also hard to set an appropriate threshold. The second one is referred to as the eigen-analysis enhanced UAI (EAE-UAI) algorithm, which exploits the eigen-analysis to improve the performance of UAI. In order to further enhance the performance of UAI, we then propose two successive interference cancellation (SIC) assisted UAI algorithms, namely, the SIC-UAIa and SIC-UAIb, which improve the performance of channel estimation and UAI with the aid of the SIC strategies as proposed. Finally, as above-mentioned, any added knowledge about active UEs can be exploited to enhance the performance of channel estimation. Correspondingly, we propose an auto-correlation matrix evolving UAI (AME-UAI) algorithm, which makes use of the earlier identified UEs to improve the channel estimation and UAI of the UEs processed later. In the above proposed five algorithms, the last three carry out joint channel estimation and UAI. Furthermore, in comparison with the TB-UAI, the other four algorithms make no or little trade-off between false-alarm and miss, in addition to the performance enhancement.

The novelty of our work is compared with the related works in Table 1 and the main contributions can be summarized as follows:

- Channel estimation in dynamic multicarrier mGFMA (MC-mGFMA) scenarios is investigated, where the channels of both active and inactive UEs are estimated in MMSE principle, when AP is assumed to employ ideal

TABLE 1. Comparison of This Work with the Related Works in Literature

Contributions	System model	Channel model	Random No. of UEs	Dynamic UEs	Priori knowledge	Real overloading	CE	UAI method	Joint UAI&CE
This work	SP-NOMA	Correlated FS	✓	✓		✓	✓	CE	✓
[8]	SP-NOMA	Uncorrelated FS		✓	UEs' channels			CS	
[10]	SP-NOMA	Uncorrelated FS	✓	✓*	UEs' channels			CS	
[11]	mMIMO	Flat	✓	✓			✓	CS	✓
[24]	SP-NOMA	Correlated FS		✓	UEs' channels			CS	
[23]	SP-NOMA	Flat		✓*			✓	CS	UAI&CE&DD
[19]	SP-NOMA	Correlated FS	✓	✓	UEs' channels	✓		MAP	
[27]	mMIMO	flat	✓	✓			✓	CS	✓
[26]	mMIMO	flat	✓	✓	No. of UEs			CS/learning	UAI&CE&DD
[18]	SP-NOMA	Uncorrelated FS		✓	UEs' channels			CS	
[22]	SP-NOMA	Uncorrelated FS	✓	✓			✓	CS/learning	✓
[35]	SCMA	Gaussian**	✓***	✓***	UEs' channels	✓		MP	
[34]	SCMA	Correlated FS	✓	✓	No. of UEs	✓	✓	ML/MP	

Notes: ✓: yes; Empty: no; SP: spreading; FS: frequency-selective fading; CE: channel estimation; mMIMO: massive or multiuser MIMO; DD: data detection; MAP: maximum a posteriori probability; MP: message-passing; AMP: approximate MP; *: temporally correlated user activity; ML: maximum likelihood; **: Gaussian channel for performance; ***: All UEs transmit headers for UAI;

knowledge, no knowledge or partial knowledge about the active UEs. It is shown that the proposed channel estimation approaches are robust to the dynamics existing in mGFMA systems. Furthermore, we show that the added knowledge about active UEs results in the enhanced reliability of channel estimation.

- The statistical properties of the estimated channels of both active and inactive UEs are analyzed, showing that the estimated channels of active UEs and that of inactive UEs have distinct features, which can be exploited for distinguishing active UEs from inactive UEs.
- Hence, based on the studies of channel estimation and of the characteristics of estimated channels, five UAI algorithms are proposed for MC-mGFMA and their performance is studied. These UAI algorithms are listed as follows:
  1) Threshold-based UAI (TB-UAI);
  2) Eigen-analysis enhanced UAI (EAE-UAI);
  3) Successive interference cancellation (SIC) assisted UAIa (SIC-UAIa);
  4) Successive interference cancellation (SIC) assisted UAIb (SIC-UAIb);
  5) Auto-correlation matrix evolving UAI (AME-UAI).

  These UAI algorithms have their distinctive characteristics and present different performance-complexity trade-off. They however have one common feature that all of them are suitable for operation in the mGFMA systems having highly dynamic UEs in terms of the active UEs and the number of them.

The remainder of this paper is organized as follows. In Section II, we describe the mGFMA system model. Section III focuses on channel estimation, while Section IV demonstrates the performance of channel estimation. Furthermore, in Section V, we study the statistical properties of the channels estimated by the approaches in Section III. The UAI algorithms and their complexity are addressed in Section VI, and the performance of UAI is demonstrated in Section VII. Finally, the conclusions from our research are summarized in Section VIII.

II. DESCRIPTION OF MGFMA SYSTEM

We consider a single-cell multicarrier mGFMA (MC-mGFMA) system in the principles of the frequency-domain spread orthogonal frequency-division multiple-access (OFDMA) [48]. The MC-mGFMA system has N subcarriers and one access point (AP) (or base-station (BS)) located at the center of the cell. Since most mGFMA systems are motivated for low-rate IoT services, we assume only the binary phase shift keying (BPSK) modulation, although it can be readily extended to other quadrature amplitude modulations (QAM) modulation. In this MC-mGFMA system, the signature code assigned to the kth UE equipment (UE) is expressed as $\boldsymbol{c}_k = [c_k(0), c_k(1), \ldots, c_k(N-1)]^T$, where $c_k(n) \in \{+1/\sqrt{N}, -1/\sqrt{N}\}$, making $\|\boldsymbol{c}_k\|^2 = 1$. The signature $\boldsymbol{c}_k$ of UE k serves for both the spreading of transmitted signal, and the identity (ID) of UE k for AP to identify its activity during the UE identification stage. If UE k is identified to be active, $\boldsymbol{c}_k$ is also used by AP to demodulate the signals received from UE k during the data detection stage. In our considered MC-mGFMA system, IFFT/FFT techniques are introduced for subcarrier modulation/demodulation, and a sufficiently long cyclic prefix (CP) is added to avoid the interference between adjacent OFDM blocks. We assume that the MC-mGFMA system supports K potential UEs, which are mobile IoT devices randomly distributed in the cell. In massive connectivity application scenarios, we have $K >> N$. We assume that the system implements a synchronization subsystem, which enables all UEs to synchronize with the AP's clock. A UE has a small probability of P_a to become active, and an active UE starts transmission at the beginning of the next time-slot. One frame is transmitted per time-slot by each active UE. We express K_A the number of active UEs during a time-slot, which is a random variable having its average satisfying $\bar{K}_A << K$, as the result of a small P_a.

A. TRANSMITTER MODELING

The transmitter schematic diagram for the considered MC-mGFMA system is shown in Fig. 1, which follows the typical

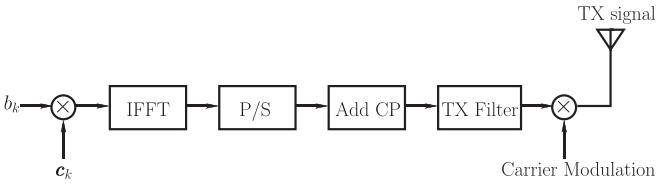


FIGURE 1. Transmitter schematic diagram for GFMA systems.

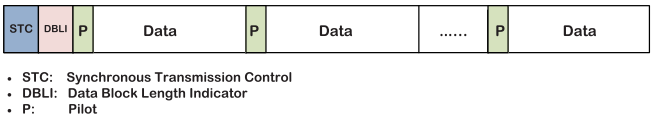


FIGURE 2. Frame structure for operation of proposed MC-mGFMA.

framework of OFDM transmitter [48]. To transmit a binary bit $b_k \in \{+1, -1\}$, it is first spread by $\boldsymbol{c}_k$, the products of which are transformed to the time-domain using IFFT. Then, after the parallel-to-serial (P/S) conversion, adding CP, transmitter (TX) filtering, and carrier modulation, the resultant signal is transmitted from the transmit antenna of UE k.

In our study, we assume block-based fading, i.e., the fading of a UE maintains constant over a frame duration, but is independent frame-by-frame. Let the frame length be expressed as $N_F = N_I + N_P + N_L$, where N_I denotes the number of information bits per frame, and N_P is the number of pilot bits per frame, inserted for channel estimation and UE identification. Furthermore, at the beginning of each frame, N_L bits are transmitted to represent the data block length indicator (DBLI), which tells AP the number of frames left to transmit. In our studies, we typically set $N_L = 2$. Hence, once a UE becomes active, the maximum number of frames to transmit in one session is 4, which is informed to AP by setting the DBLI bits in the first frame to '11'. This value reduces following the sequence '10,' '01,' until '00' in the 4th frame. Then, no matter whether this UE finishes transmission, it stops transmission. If it has more data to transmit, it activates again and starts a new transmission session. The benefit of employing the DBLI is that the AP does not need to identify every active UE, but those newly activating ones, which enhances the performance of active UE identification, channel estimation and hence of signal detection. For illustration, the frame arrangement in our MC-mGFMA is depicted in Fig. 2. As seen in Fig. 2, there is also a 'STC' domain, which is used to make the transmissions by different UEs be approximately synchronized at AP, as done in LTE/LTE-A systems [49]. To implement this, each UE measures the transmission delay between it and its AP based on the pilot signals sent from the AP. Then, the UE can adjust its starting transmission point in the 'STC' region, so that the transmissions from different UEs are approximately synchronized when they arrive at the AP.

B. REPRESENTATION OF RECEIVED SIGNAL

The receiver block diagram for the MC-mGFMA system is shown in Fig. 3, which follows the receiving principles of

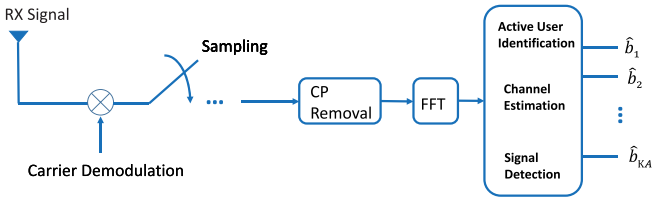


FIGURE 3. Receiver schematic diagram of MC-mGFMA system.

OFDM until after the FFT operation. Hence, after the FFT operation, the received signals from N subcarriers can be represented as

$$\boldsymbol{y}_m = \sum_{k=1}^{K} \boldsymbol{C}_k \boldsymbol{\mathcal{F}} \boldsymbol{\Psi}_k \boldsymbol{h}_k b_k^{(m)} I_k + \boldsymbol{n}_m, \tag{1}$$

$$= \sum_{k'=1}^{K_A} \boldsymbol{C}_{k'} \boldsymbol{\mathcal{F}} \boldsymbol{\Psi}_{k'} \boldsymbol{h}_{k'} b_{k'}^{(m)} + \boldsymbol{n}_m, m = 1, 2, \ldots, N_F \tag{2}$$

where $b_k^{(m)}$ is the mth bit in a frame sent from UE k, $\boldsymbol{y}_m$ and $\boldsymbol{n}_m$ are N-length observation and noise vectors, respectively, and $\boldsymbol{n}_m$ follows the complex Gaussian distribution with zero mean and covariance matrix $\sigma^2 \boldsymbol{I}_N$, expressed as $\mathcal{CN}(0, \sigma^2 \boldsymbol{I}_N)$, where $\sigma^2 = 1/SNR$ with SNR being the signal-to-noise ratio (SNR). In (1), $\boldsymbol{C}_k = \text{diag}\{c_k(0), c_k(1), \ldots, c_k(N-1)\}$, $\boldsymbol{h}_k = [h_{k,0}, \ldots, h_{k,L-1}]^T$ is the time-domain channel impulse response (CIR) of UE k, where $h_{k,l}$ obeys the distribution $\mathcal{CN}(0, 1/L)$. Correspondingly, the frequency-domain CIRs, i.e., the fading gains of N subcarriers are given by $\boldsymbol{\mathcal{F}} \boldsymbol{\Psi}_k \boldsymbol{h}_k$ [48], where $\boldsymbol{\mathcal{F}}$ is the $(N \times N)$ FFT matrix, and $\boldsymbol{\Psi}_k$ is the mapping matrix constituted by the first L columns of the identity matrix $\boldsymbol{I}_N$. Furthermore, in (1), I_k is the activity indicator, $I_k = 1$ or $I_k = 0$ indicates that UE k is active or inactive. For simplicity, let in (1) $\boldsymbol{A}_k = \boldsymbol{C}_k \boldsymbol{\mathcal{F}} \boldsymbol{\Psi}_k$, which is a $(N \times L)$ matrix. Then, we have

$$\boldsymbol{y}_m = \sum_{k=1}^{K} I_k \boldsymbol{A}_k \boldsymbol{h}_k b_k^{(m)} + \boldsymbol{n}_m, \tag{3}$$

$$= \sum_{k'=1}^{K_A} \boldsymbol{A}_{k'} \boldsymbol{h}_{k'} b_{k'}^{(m)} + \boldsymbol{n}_m, m = 1, 2, \ldots, N_F \tag{4}$$

Specifically for the pilots $p = 1, 2, \ldots, N_P$, the received signals can be expressed as

$$\boldsymbol{y}_p = \sum_{k=1}^{K} b_k^{(p)} I_k \boldsymbol{A}_k \boldsymbol{h}_k + \boldsymbol{n}_p$$

$$= \sum_{k'=1}^{K_A} b_{k'}^{(p)} \boldsymbol{A}_{k'} \boldsymbol{h}_{k'} + \boldsymbol{n}_p, p = 1, 2, \ldots, N_P \tag{5}$$

Let us collect the observations from the N_P pilot symbols to a vector $\boldsymbol{y} = [\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \ldots, \boldsymbol{y}_{N_P}^T]^T$. Then, it can be expressed as

$$\boldsymbol{y} = \sum_{k=1}^{K} I_k(\boldsymbol{p}_k \otimes \boldsymbol{A}_k)\boldsymbol{h}_k + \boldsymbol{n} \tag{6}$$

$$= \sum_{k'=1}^{K_A} (\boldsymbol{p}_{k'} \otimes \boldsymbol{A}_{k'})\boldsymbol{h}_{k'} + \boldsymbol{n} \tag{7}$$

where $\boldsymbol{p}_k = [b_k^{(1)}, b_k^{(2)}, \ldots, b_k^{(N_P)}]^T$ contains the N_P pilot bits of UE k, $\otimes$ is the Kronecker product, while $\boldsymbol{n}$ is a NN_P-length vector following the distribution of $\mathcal{CN}(0, \sigma^2\boldsymbol{I}_{NN_P})$.

III. CHANNEL ESTIMATION

In this section, we consider the channel estimation in the considered MC-mGFMA system, all based on the minimum mean-square error (MMSE) [50]. First, we derive the channel estimator by assuming that AP knows ideally the active UEs. Then, the channel estimator is obtained by assuming that AP has no knowledge at all about the active UEs. Explicitly, both the above cases are extreme cases. In practice, the situation may be that AP knows some active UEs, e.g., from the DBLI bits sent by UEs, as we considered in Section II-A, but does not know the newly activated UEs. Therefore, we also consider the channel estimation in this more practical scenario. Note that we choose MMSE instead of zero-forcing (ZF) (or least square) estimator, as ZF estimator requires to invert a large sized matrix that is dependent on the number of potential UEs, and it is well-known that MMSE estimator achieves better estimation performance than ZF estimator.

A. CHANNEL ESTIMATION WITH ACTIVE UES KNOWN TO AP - ESTIMATOR-K

When AP knows all the active UEs, the channels of these UEs can be estimated based on (7). Hence, when MMSE-relied estimation is employed, the estimated channel of the ith active UE can be expressed as

$$\hat{\boldsymbol{h}}_i = \boldsymbol{W}_i^H \boldsymbol{y}, \, i = 1, 2, \ldots, K_A \tag{8}$$

where according to the principles of MMSE, the weight matrix $\boldsymbol{W}_i$ can be expressed as [50]

$$\boldsymbol{W}_i = \boldsymbol{R}_y^{-1}\boldsymbol{R}_{y,i} \tag{9}$$

In (9), $\boldsymbol{R}_y$ is the autocorrelation matrix of $\boldsymbol{y}$, which can be derived from (7) as

$$\boldsymbol{R}_y = E\left[\boldsymbol{y}\boldsymbol{y}^H\right]$$

$$= E\left[\left(\sum_{k'=1}^{K_A}(\boldsymbol{p}_{k'} \otimes \boldsymbol{A}_{k'})\boldsymbol{h}_{k'} + \boldsymbol{n}\right)\left(\sum_{l'=1}^{K_A}(\boldsymbol{p}_{l'} \otimes \boldsymbol{A}_{l'})\boldsymbol{h}_{l'} + \boldsymbol{n}\right)^H\right]$$

$$= \sum_{k'=1}^{K_A}(\boldsymbol{p}_{k'} \otimes \boldsymbol{A}_{k'})E\left[\boldsymbol{h}_{k'}\boldsymbol{h}_{k'}^H\right](\boldsymbol{p}_{k'} \otimes \boldsymbol{A}_{k'})^H + \sigma^2\boldsymbol{I}_{N_PN} \tag{10}$$

With the aid of the relationships of $(\boldsymbol{A} \otimes \boldsymbol{B})^H = \boldsymbol{A}^H \otimes \boldsymbol{B}^H$ and $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}$ [51], we can derive that if long-term average cross many frames is imagined, we have $E[\boldsymbol{h}_{k'}\boldsymbol{h}_{k'}^H] = \boldsymbol{I}_L/L$, and (10) can be simplified to

$$\boldsymbol{R}_y = \sum_{k'=1}^{K_A}\left(\boldsymbol{p}_{k'}\boldsymbol{p}_{k'}^H \otimes \frac{\boldsymbol{A}_{k'}\boldsymbol{A}_{k'}^H}{L}\right) + \sigma^2\boldsymbol{I}_{N_PN} \tag{11}$$

By contrast, if only one frame is considered, $\boldsymbol{h}_{k'}$ is a constant vector. In this case, $\boldsymbol{R}_y$ has the form of

$$\boldsymbol{R}_y = \sum_{k'=1}^{K_A}\left(\boldsymbol{p}_{k'}\boldsymbol{p}_{k'}^H \otimes \boldsymbol{A}_{k'}\boldsymbol{h}_{k'}\boldsymbol{h}_{k'}^H\boldsymbol{A}_{k'}^H\right) + \sigma^2\boldsymbol{I}_{N_PN} \tag{12}$$

However, in MC-mGFMA systems, this $\boldsymbol{R}_y$ is inaccessible to the AP, as it is unable to construct it by exploiting the available knowledge about the UEs, while it is also incapable of estimating it using the received signals of one frame. By contrast, the AP is able to construct the $\boldsymbol{R}_y$ in (11), as it knows the pilot symbols and the spreading codes of the active UEs.

In (9), $\boldsymbol{R}_{y,i}$ is the cross-correlation matrix between $\boldsymbol{y}$ and $\boldsymbol{h}_i$, which can be shown to be

$$\boldsymbol{R}_{y,i} = \frac{1}{L}(\boldsymbol{p}_i \otimes \boldsymbol{A}_i) \tag{13}$$

Consequently, upon substituting (11) and (13) into (9), the weight matrix for estimating the ith UE's channel is given by

$$\boldsymbol{W}_i = \left[\sum_{k'=1}^{K_A}\left(\boldsymbol{p}_{k'}\boldsymbol{p}_{k'}^H \otimes \boldsymbol{A}_{k'}\boldsymbol{A}_{k'}^H\right) + \sigma^2\boldsymbol{I}_{N_PN}\right]^{-1}(\boldsymbol{p}_i \otimes \boldsymbol{A}_i)/L$$

$$= \left[\sum_{k'=1}^{K_A}\left(\boldsymbol{p}_{k'}\boldsymbol{p}_{k'}^H \otimes \boldsymbol{A}_{k'}\boldsymbol{A}_{k'}^H\right) + L\sigma^2\boldsymbol{I}_{N_PN}\right]^{-1}(\boldsymbol{p}_i \otimes \boldsymbol{A}_i) \tag{14}$$

Similarly, if (12) is considered, we have

$$\boldsymbol{W}_i = \left[\sum_{k'=1}^{K_A}\left(\boldsymbol{p}_{k'}\boldsymbol{p}_{k'}^H \otimes \boldsymbol{A}_{k'}\boldsymbol{h}_{k'}\boldsymbol{h}_{k'}^H\boldsymbol{A}_{k'}^H\right) + \sigma^2\boldsymbol{I}_{N_PN}\right]^{-1} \times (\boldsymbol{p}_i \otimes \boldsymbol{A}_i)/L \tag{15}$$

Note that although (15) is not implementable, it is insightful for deriving the other estimators, as shown in our forthcoming discourses. Furthermore, from (11) we can deduce that the $(N \times N)$ diagonal sub-matrices of $\boldsymbol{R}_y$ are given by

$$\boldsymbol{R}_y(n, n) = \boldsymbol{R}_a$$

$$= \sum_{k'=1}^{K_A}\frac{\boldsymbol{A}_{k'}\boldsymbol{A}_{k'}^H}{L} + \sigma^2\boldsymbol{I}_N, n = 1, 2, \ldots, N_P \tag{16}$$

By contrast, from (12) we can know that these diagonal sub-matrices are

$$\boldsymbol{R}_a = \sum_{k'=1}^{K_A}\boldsymbol{A}_{k'}\boldsymbol{h}_{k'}\boldsymbol{h}_{k'}^H\boldsymbol{A}_{k'}^H + \sigma^2\boldsymbol{I}_N,$$

$$n = 1, 2, \ldots, N_P \tag{17}$$

These results will be useful later.

The MSE of the channel estimation for an active UE can be derived as

$$\begin{aligned} M_{SE}(A) &= E\left[\|\boldsymbol{h}_i - \boldsymbol{W}_i^H \boldsymbol{y}\|^2\right] \\ &= \mathrm{Tr}\left(\boldsymbol{I}_L/L - \boldsymbol{R}_{y_{h_i}}^H \boldsymbol{W}_i\right) \\ &= 1 - \mathrm{Tr}\left(\boldsymbol{R}_{y,i}^H \boldsymbol{W}_i\right) \end{aligned} \tag{18}$$

If the AP mistakenly estimates the channel of a UE that is actually not active, the AP can still form $\boldsymbol{W}_i$ in the form of (14). Correspondingly, the MSE of the channel estimation for an inactive UE is

$$\begin{aligned} M_{SE}(\bar{A}) &= E\left[\| - \boldsymbol{W}_i^H \boldsymbol{y}\|^2\right] \\ &= \mathrm{Tr}\left(\boldsymbol{W}_i^H \boldsymbol{R}_y \boldsymbol{W}_i\right) \end{aligned} \tag{19}$$

Later in Section VI, we will investigate the statistics of estimated channels conditioned on that a UE is active or inactive. This statistical information will then be exploited for active UE identification.

Finally, we note that the MMSE estimator is a biased estimator [50]. In order to obtain an unbiased estimation of $\boldsymbol{h}_i$, after the MMSE estimation, we can let

$$\hat{\boldsymbol{h}}_i = \boldsymbol{Q}_i \boldsymbol{W}_i^H \boldsymbol{y},\ i = 1, 2, \ldots, K_A \tag{20}$$

where $\boldsymbol{Q}_i$ is applied to achieve the unbiased estimation, making the estimated channel satisfy $E[\hat{\boldsymbol{h}}_i] = \boldsymbol{h}_i$. Therefore, when substituting (7) into the above equation and completing the expectation, we can obtain

$$\boldsymbol{Q}_i = \left(\boldsymbol{W}_i^H (\boldsymbol{p}_i \otimes \boldsymbol{A}_i)\right)^{-1} \tag{21}$$

B. CHANNEL ESTIMATION WITH ACTIVE UES UNKNOWN TO AP - ESTIMATOR-UK

In contrast to the extreme case that AP knows ideally the active UEs, another extreme case in MC-mGFMA is that AP only knows there are K potential UEs, but does not know the active UEs and even the number of active UEs K_A. In this case, first, AP has to treat the data sent from different UEs as independent identically distributed (iid) random variables, and use $E[\boldsymbol{p}_{k'} \boldsymbol{p}_{k'}^H]$ to replace $\boldsymbol{p}_{k'} \boldsymbol{p}_{k'}^H$ in (12), where

$$E[\boldsymbol{p}_{k'} \boldsymbol{p}_{k'}^H] = \boldsymbol{I}_{N_P} \tag{22}$$

When applying this result to (12), we obtain

$$\begin{aligned} \boldsymbol{R}_y &= \sum_{k'=1}^{K_A} \left(\boldsymbol{I}_{N_P} \otimes \boldsymbol{A}_{k'} \boldsymbol{h}_{k'} \boldsymbol{h}_{k'}^H \boldsymbol{A}_{k'}^H\right) + \sigma^2 \boldsymbol{I}_{N_P N} \\ &= \left(\boldsymbol{I}_{N_P} \otimes \left[\sum_{k'=1}^{K_A} \boldsymbol{A}_{k'} \boldsymbol{h}_{k'} \boldsymbol{h}_{k'}^H \boldsymbol{A}_{k'}^H + \sigma^2 \boldsymbol{I}_N\right]\right) \\ &= \left(\boldsymbol{I}_{N_P} \otimes \boldsymbol{R}_a\right) \end{aligned} \tag{23}$$

where $\boldsymbol{R}_a$ from (17) is substituted. Note that in (23), K_A is still unknown to AP, which will be addressed later in this subsection.

When AP attempts to estimate the channel of UE i, it can use the related information (spreading code and its pilot symbols) to construct the cross-correlation matrix $\boldsymbol{R}_{y,i}$ of (13). Consequently, AP can form the weight matrix $\boldsymbol{W}_i$ as

$$\begin{aligned} \boldsymbol{W}_i &= \boldsymbol{R}_y^{-1} \boldsymbol{R}_{y_{h_i}} \\ &= \left(\boldsymbol{I}_{N_P} \otimes \boldsymbol{R}_a\right)^{-1} (\boldsymbol{p}_i \otimes \boldsymbol{A}_i)/L \\ &= \left(\boldsymbol{I}_{N_P} \otimes \boldsymbol{R}_a^{-1}\right) (\boldsymbol{p}_i \otimes \boldsymbol{A}_i)/L \\ &= \left(\boldsymbol{p}_i \otimes \boldsymbol{R}_a^{-1} \boldsymbol{A}_i/L\right) \end{aligned} \tag{24}$$

Let us express $\boldsymbol{W}_i$ as

$$\boldsymbol{W}_i = \left[\boldsymbol{W}_{i,1}^T, \boldsymbol{W}_{i,2}^T, \cdots, \boldsymbol{W}_{i,N_p}^T\right]^T \tag{25}$$

Then, it can be readily shown that

$$\boldsymbol{W}_{i,p} = b_i^{(p)} \boldsymbol{R}_a^{-1} \boldsymbol{A}_i/L,\ p = 1, 2, \ldots, N_P \tag{26}$$

Furthermore, when substituting (25) into (8), we obtain

$$\hat{\boldsymbol{h}}_i' = \boldsymbol{W}_i^H \boldsymbol{y} = \sum_{p=1}^{N_P} \boldsymbol{W}_{i,p}^H \boldsymbol{y}_p = \sum_{p=1}^{N_P} \boldsymbol{h}_i^{(p)} \tag{27}$$

where

$$\hat{\boldsymbol{h}}_i^{(p)} = \boldsymbol{W}_{i,p}^H \boldsymbol{y}_p,\ p = 1, 2, \ldots, N_P \tag{28}$$

which can be shown is the MMSE-based estimation of $\boldsymbol{h}_i$ using only the signals related to the pth pilot symbol.

Therefore, due to the application of (22), the estimation of (27) does not directly provide an estimate for $\boldsymbol{h}_i$ in MMSE sense. Instead, each component of $\hat{\boldsymbol{h}}_i^{(p)}$, as seen in (28), is the MMSE estimation of $\boldsymbol{h}_i$ based on the observation $\boldsymbol{y}_p$ corresponding to the pth pilot symbol. In other words, (27) is the sum of the N_P estimates to $\boldsymbol{h}_i$ based on the N_P pilot symbols. Hence, in order to obtain an unbiased estimate to $\boldsymbol{h}_i$, we can form the estimation as

$$\hat{\boldsymbol{h}}_i = \frac{1}{N_P} \sum_{p=1}^{N_P} \boldsymbol{Q}_i^{(p)} \boldsymbol{W}_{i,p}^H \boldsymbol{y}_p = \frac{1}{N_P} \sum_{p=1}^{N_P} \boldsymbol{Q}_i^{(p)} \hat{\boldsymbol{h}}_i^{(p)} \tag{29}$$

where, explicitly, the factor $1/N_P$ takes account of the average over N_P pilots, while $\boldsymbol{Q}_i^{(p)}$ is for obtaining the unbiased estimation of $\boldsymbol{h}_i$ based on the pth pilot symbol, which can be found to be

$$\begin{aligned} \boldsymbol{Q}_i^{(p)} &= \left(\boldsymbol{W}_{i,p}^H b_i^{(p)} \boldsymbol{A}_i\right)^{-1} \\ &= \left(\boldsymbol{A}_i^H \boldsymbol{R}_a^{-1} \boldsymbol{A}_i/L\right)^{-1},\ i = 1, 2, \ldots, N_P \end{aligned} \tag{30}$$

From (27) and (29) we can conceive that when the AP does not know the active UEs, it has to first carry out the symbol-based channel estimation, and then average the estimates obtained from the N_P pilot symbols to give the final estimation. Furthermore, while AP knows $b_i^{(p)}$ from the pilot sequences of UEs and it can also construct $\boldsymbol{A}_i$ from the UEs' spreading sequences, AP has to know $\boldsymbol{R}_a$, in order to compute $\boldsymbol{W}_{i,p}$ of (26) to fulfill the channel estimation based on (29). However,

AP does not have the knowledge about the active UEs and even the number of them, $\boldsymbol{R}_a$ is unable to be constructed using the knowledge available to AP, but has to be obtained from alternative approaches. In this paper, we propose to estimate $\boldsymbol{R}_a$ from the received signals as

$$\hat{\boldsymbol{R}}_a \approx \frac{1}{N_F}\sum_{n=1}^{N_F}\boldsymbol{y}_n\boldsymbol{y}_n^H \tag{31}$$

The above equation implies that $\boldsymbol{R}_a$ can be estimated by making use of all the signals received within a frame-duration, which is beneficial to lessening the effect of transmitted data. When applying (4) into the above equation, we obtain an approximation of

$$\hat{\boldsymbol{R}}_a \approx \sum_{k'=1}^{K_A}\boldsymbol{A}_{k'}\boldsymbol{h}_{k'}\boldsymbol{h}_{k'}^H\boldsymbol{A}_{k'}^H + \sigma^2\boldsymbol{I}_N \tag{32}$$

the righthand side of which is exactly the $\boldsymbol{R}_a$ of (17).

Hence and in summary, for AP to estimate the ith UE's channel when it has no knowledge about the active UEs, the AP first uses all the received signals over one frame to estimate $\boldsymbol{R}_a$ based on (31). Then, it constructs $\boldsymbol{A}_i$ using the spreading sequence assigned to UE i. Furthermore, with the above and the pilots of UE i, AP forms the weight matrices based on (26) and $\boldsymbol{Q}_i^{(p)}$'s based on (30). Finally, the channel of UE i is estimated based on (29).

C. CHANNEL ESTIMATION WITH ACTIVE UES PARTIALLY KNOWN TO AP - ESTIMATOR-PK

In the previous two subsections, we have considered two extreme cases. Specially in Subsection III-A, it is assumed that AP has ideal knowledge about active UEs, which should yield the upper-bound estimation performance. By contract, in III-B, AP is assumed to have no knowledge about active UEs, which results in the worst estimation performance. In practical mGFMA systems, the operational condition may be more like that AP knows in part the active UEs, as the result that some UEs have more than one frame to send, as indicated to AP by the DBLI bits. However, AP does not know the newly activated UEs, whose activities need to be identified.

Therefore, if AP knows some UEs as *a-priori*, this information can be exploited to improve the performance of channel estimation in MC-mGFMA systems. Specifically, we can construct $\boldsymbol{R}_y$ in the form of (11) or (12) as follows.

First, AP estimates for $\boldsymbol{R}_a$ using (31). Then, AP constructs $\boldsymbol{A}_k$ for all the active UEs known to AP. With the aid of this information and the pilot symbols, AP is capable of constructing a $\boldsymbol{R}_y$. In (33), shown at the bottom of this page, $\bar{\mathcal{K}}_A$ is the set of active UEs known to AP.

The cross-correlation matrices for all UEs are available to AP, which are the same as (13). Consequently, the AP can construct the weight matrices $\boldsymbol{W}_i$ as (9), and finally, estimate the channels of UEs in the same form as (20).

D. DISCUSSION

The above-considered three channel estimation schemes are all in the principle of MMSE. Hence, all of them have relatively low-complexity for practical implementation. However, the Estimator-K and Estimator-pK need to invert a relatively large size matrix (autocorrelation matrix), they have higher complexity than the Estimator-uK. Nevertheless, during one time-slot, the autocorrelation matrix is only required to be inverted once, and then can be used for the channel estimation of all UEs.[2] Therefore, its contribution to the overall complexity should be insignificant.

In terms of the delay introduced by channel estimation, if K UEs' channels are estimated successively, the delay is proportional to K. However, from the principles of these estimators, we can readily know that all the three estimators can be implemented in parallel, i.e., all the K UEs' channels can be simultaneously estimated, once the inverse of the autocorrelation matrix is computed. In this way, the delay of channel estimation is independent of the number of UEs involved.

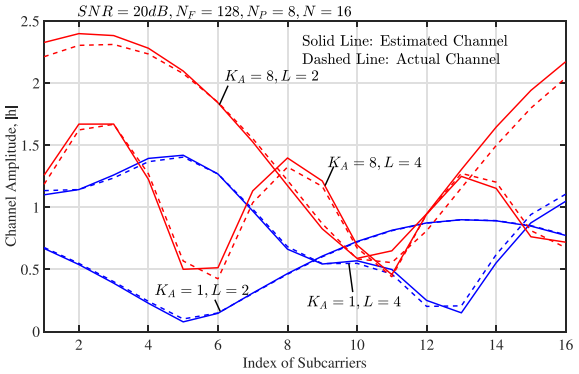
IV. PERFORMANCE OF CHANNEL ESTIMATION

In this section, we provide some results to demonstrate the performance of the respective channel estimation schemes considered in the above section. The observations from them will be exploited for UAI in Section VI.

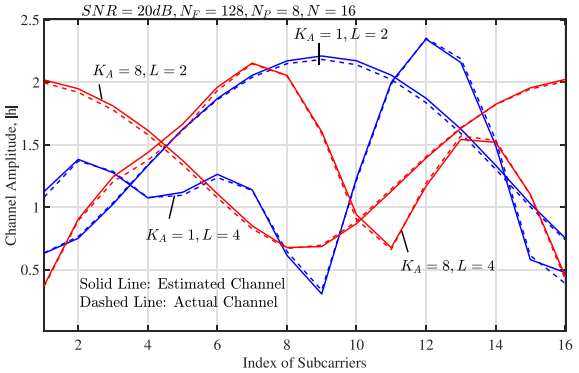
In Figs. 4(b) and 4(b), we show the snapshots of the MMSE-based channel estimation considered in Section III. In our demonstrations, we assume a MC-mGFMA system with $N = 16$ subcarriers and supporting $K_A = K = 1$ or 8 active

[2] Also, this autocorrelation matrix and its inverse can be used for the data detection of active UEs

$$\boldsymbol{R}_y = \begin{pmatrix} \boldsymbol{R}_a & \sum_{k\in\bar{\mathcal{K}}_A} b_k^{(1)}b_k^{(2)}\frac{\boldsymbol{A}_k\boldsymbol{A}_k^H}{L} & \cdots & \sum_{k\in\bar{\mathcal{K}}_A} b_k^{(1)}b_k^{(N_P)}\frac{\boldsymbol{A}_k\boldsymbol{A}_k^H}{L} \\ \sum_{k\in\bar{\mathcal{K}}_A} b_k^{(2)}b_k^{(1)}\frac{\boldsymbol{A}_k\boldsymbol{A}_k^H}{L} & \boldsymbol{R}_a & \cdots & \sum_{k\in\bar{\mathcal{K}}_A} b_k^{(2)}b_k^{(N_P)}\frac{\boldsymbol{A}_k\boldsymbol{A}_k^H}{L} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k\in\bar{\mathcal{K}}_A} b_k^{(N_P)}b_k^{(1)}\frac{\boldsymbol{A}_k\boldsymbol{A}_k^H}{L} & \sum_{k\in\bar{\mathcal{K}}_A} b_k^{(N_P)}b_k^{(2)}\frac{\boldsymbol{A}_k\boldsymbol{A}_k^H}{L} & \cdots & \boldsymbol{R}_a \end{pmatrix} \tag{33}$$

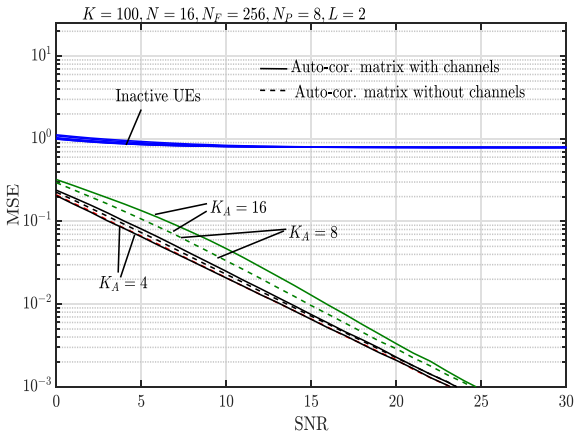


(a) Symbol-based channel estimation of (29).

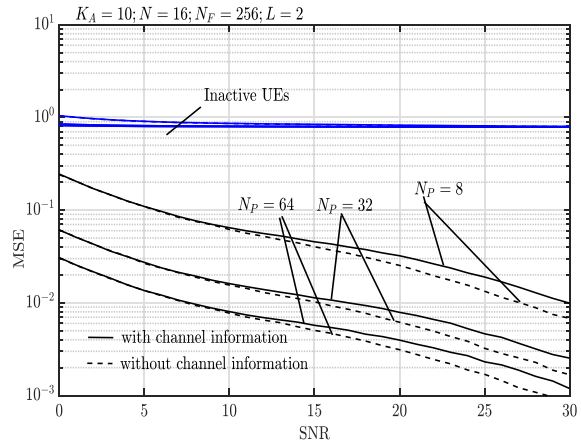


(b) Block-based channel estimation of (20).

FIGURE 4. Snapshots of estimated channels.



(a) Estimator-K.

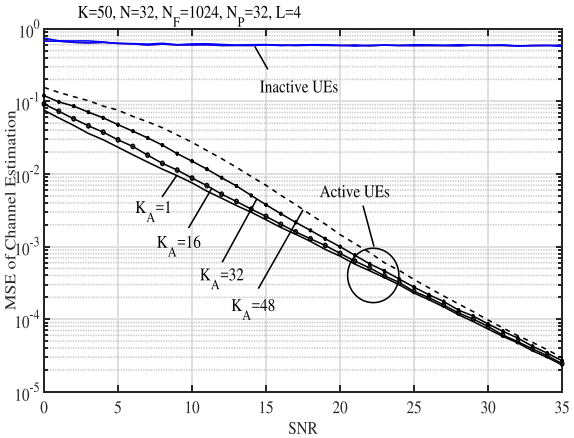


(b) Estimator-uK.

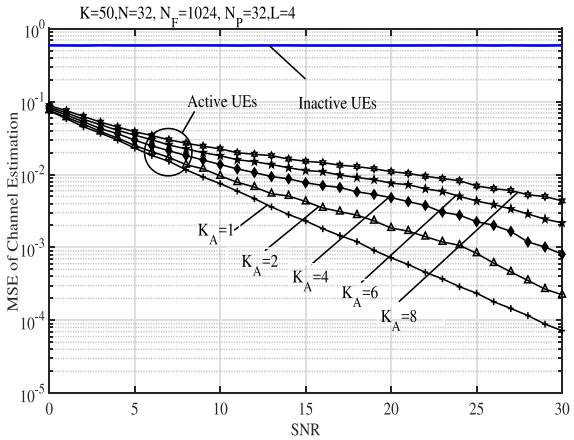
FIGURE 5. MSE performance of channel Estimator-K and Estimator-uK, when channel is invoked or not invoked in auto-correlation matrix.

UEs known to AP. The frame length is set to $N_F = 128$, in which $N_p = 8$ pilots are used for channel estimation. We assume the frequency-selective fading channels with $L = 2$ or 4. Furthermore for Fig. 4(a), we assume that the channel is estimated using the symbol-based estimator of (29), while for Fig. 4(b), the channel is estimated by the block-based estimator of (20). From the results we can have the following observations. First, when the channel becomes more frequency selective resulted from L being changed from $L = 2$ to $L = 4$, the subcarrier channels become less correlated, or the channel is more variant in the frequency domain. Second, when the number of UEs is increased from $K_A = 1$ to $K_A = 8$, the symbol-based estimator shows slight performance degradation as the result of multiuser interference. By contrast, the block-based estimator demonstrates that in both cases, the estimated channels are close to the actual ones. The reason behind is that when $N = 16$ and $N_p = 8$, the symbol-based estimator has only $N = 16$ degrees of freedom, while the block-based estimator has $NN_p = 128$ degrees of freedom available for channel estimation. Third, by comparing Fig. 4(a) with Fig. 4(b), we can explicitly see that the block-based estimator achieves more accurate estimation than the symbol-based estimator. Hence, once AP has some knowledge about active UEs, it can be exploited for improving the performance of channel estimation. This implies that among the three channel estimators, Estimator-K should outperform both Estimator-pK and Estimator-uK, while Estimator-pK outperforms Estimator-uK, as shown below.

In Section III-A, we discussed that when the auto-correlation matrix $\boldsymbol{R}_a$ is estimated using (32), and $\boldsymbol{R}_y$ is constructed based on the estimated $\boldsymbol{R}_a$, the channels would be invoked in these autocorrelation matrices. However and ideally, it is desirable that these auto-correlation matrices are free from the channels to be estimated, as seen in (11). In Fig 5(a) and Fig. 5(b), the impact of the invoked channels on the estimation performance is demonstrated. From both figures we can find that the auto-correlation matrix including the channels to be estimated only yields slight performance loss, in comparison to the auto-correlation matrix free from channels. Since the performance loss is insignificant, the estimated $\boldsymbol{R}_a$ is valid for application. Figs. 5(a) and 5(b) show that the MSE of the estimated channels for inactive UEs is usually very high and furthermore, does not change much with the increase of SNR. There will be more discussion about it in association with following figures.



(a) Estimator-K.



(b) Estimator-uK.

FIGURE 6. **MSE performance of channel Estimator-K and Estimator-uK for both active and inactive UEs.**

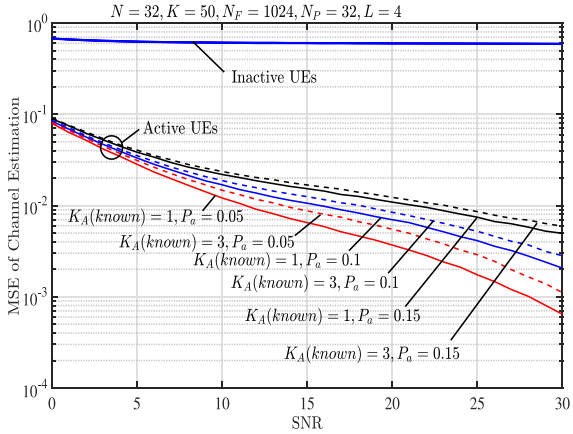


FIGURE 7. **MSE performance of channel Estimator-pK for both active and inactive UEs.**

In Fig. 6, we investigate the MSE performance of the channel estimation for both active and inactive UEs, when AP either has ideal knowledge about active UEs (Fig. 6(a)) or has no knowledge about active UEs (Fig. 6(b)) and hence, the channel estimation follows the principles in Section . The parameters used in our studies are detailed with the figures. From the results of Fig. 6 we can observe that for inactive UEs, their channels are always estimated with very high MSE. This is because in this case, the receiver at AP can only estimate interference and noise, due to the fact that there are no signals transmitted by inactive UEs. By contrast, if the channel estimation is for an active UE, the MSE presents to reduce as SNR increases. Specifically, when the Estimator-K is considered, as shown in Fig. 6(a), the MSE performance is very promising due to the simultaneous use of all the $N_P = 32$ pilots for channel estimation. As SNR increases, even for the case of $K_A = 48$, the MSE performance converges to that of a single active UE. The reason behind is that there are in total $N_P \times N = 32 \times 32 = 1024$ degrees of freedom used for channel estimation, which is significantly larger than $K_A \cdot L = 48 \times 4 = 192$ of the number of variables estimated. Furthermore, as shown in Fig. 6(a), in the low SNR region, the MSE performance slightly degrades, as the number of active UEs increases.

By contrast, as shown in Fig. 6(b) for Estimator-uK, as only the symbol-based channel estimator can be implemented, multiuser interference imposes a big effect on the MSE performance and the MSE performance significantly degrades, when the number of active UEs increases. The reason behind this observation is that the channel from each UE has $L = 4$ variables to be estimated. However, corresponding to one symbol, there are only $N = 32$ observations (degrees-of-freedom). Hence, when there is one more active UE, there are 4 extra variables to be estimated, which is a relatively big number with respect to $N = 32$ and hence, significant performance degradation is observed, as shown in Fig. 6(b). Furthermore, this can explain why the MSE performance corresponding to $K_A = 8$ is very poor. In this case, there are in total $K_A \cdot L = 32$ variables to be estimated based on $N = 32$ observations per symbol.

When comparing Fig. 6(b) with Fig. 6(a), a very important observation can be derived, i.e., the MSE performance shown in Fig. 6(a) is much better than that in Fig. 6(b) in the case of $K_A > 1$. Furthermore, as shown in Fig. 6(a), the MSE performance of the considered cases converges, if SNR is sufficiently high. These observations imply that whenever there is some information about the activity of UEs available, this information may be exploited for improving the performance of channel estimation. This is the fundamental for the Estimator-pK.

Therefore, Fig. 7 shows the MSE performance of channel estimation, when AP has partial knowledge of the active UEs. In the figure, $K_A(known) = 1$ or 3 means that the AP knows 1 or 3 active UEs, while the other UEs become active independently with the activation probability shown in the figure. As seen in the figure, the MSE performance degrades, as the activation probability per UE increases, resulting in an increase of active UEs in average. For a given activation probability, when the number of active UEs is changed from $K_A + 1$ to $K_A + 3$,

the MSE performance only slightly degrades, as AP can make use of the Estimator-pK. Again, for the inactive UEs, the MSE of channel estimation is very high.

V. STATISTICS OF ESTIMATED CHANNELS

In the above section, we have considered the channel estimation in MC-mGFMA systems. In order to achieve reliable signal detection of UEs, it is essential for the AP receiver to know which UEs are active. Before we detail the UE activity identification (UAI) schemes, let us first have a look of the statistical properties of the estimated channels corresponding to the active and inactive UEs.

Following the above section, the time-domain CIR of the ith UE is $\boldsymbol{h}_i$, which is assumed to have L taps. After the MMSE-assisted channel estimation in Section III, the estimate to $\boldsymbol{h}_i$ can be written as

$$\hat{\boldsymbol{h}}_i = \begin{cases} \boldsymbol{h}_i + \boldsymbol{n}_i, & \text{if UE } i \text{ is active} \\ \boldsymbol{n}_i', & \text{if UE } i \text{ is inactive} \end{cases} \tag{34}$$

where $\boldsymbol{n}_i$ is the channel estimation error of an active UE, which can be Gaussian approximated with a PDF of $\mathcal{CN}(\boldsymbol{0}, \sigma_1^2 \boldsymbol{I}_L)$, where σ_1^2 is the variance of channel estimation error. By contrast, $\boldsymbol{n}_i'$ is the estimate to an inactive UE, which can also be approximated to have the Gaussian distribution with a PDF of $\mathcal{CN}(\boldsymbol{0}, \sigma_0^2 \boldsymbol{I}_L)$, where σ_0^2 may be different from σ_1^2.

Let us now consider the statistics of $|\hat{\boldsymbol{h}}_i|^2$ on the condition that UE i is either active or inactive. First, when UE i is active, the elements of $\hat{\boldsymbol{h}}_i$ are iid complex Gaussian random variables, each of which has zero mean and a variance of $1/L + \sigma_1^2$, when Rayleigh fading channels are assumed, and when all component paths have the same power of $1/L$. Therefore, $|\hat{\boldsymbol{h}}_i|^2$ obeys the centre χ^2-distributions with $2L$ degrees of freedom, and the PDF is [52]

$$f_{|\hat{\boldsymbol{h}}_i|^2}(\boldsymbol{y}|I_i = 1, K_A) = \frac{1}{(L-1)!(1/L + \sigma_1^2(K_A))^L} \boldsymbol{y}^{L-1} \times \exp\left(-\frac{\boldsymbol{y}}{1/L + \sigma_1^2(K_A)}\right), y \geq 0 \tag{35}$$

where we explicitly show that $\sigma_1^2(K_A)$ is related to the number of active UEs. By contrast, when the ith UE is inactive, it can be shown that $|\hat{\boldsymbol{h}}_i|^2$ follows the centre χ^2-distribution of

$$f_{|\hat{\boldsymbol{h}}_i|^2}(\boldsymbol{y}|I_i = 0, K_A) = \frac{1}{(L-1)!\sigma_0^{2L}(K_A)} \boldsymbol{y}^{L-1} \times \exp\left(-\frac{\boldsymbol{y}}{\sigma_0^2(K_A)}\right), y \geq 0 \tag{36}$$

Both (35) and (36) are central χ^2-distributions. For given K_A, we in general have $1/L + \sigma_1^2(K_A) > \sigma_0^2(K_A)$. Hence, from the properties of χ^2-distribution we can know that the $|\hat{\boldsymbol{h}}_i|^2$ in (35) corresponding to active UE is usually larger than that in (36) corresponding to inactive UE, as shown below.

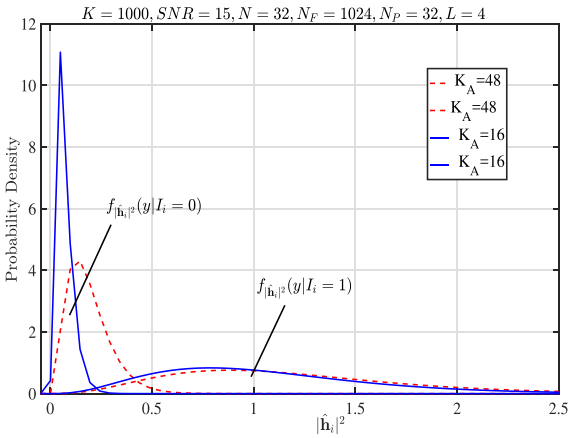


FIGURE 8. PDFs of $|\hat{\boldsymbol{h}}_i|^2$ on condition that UE i is active ($I_i = 1$) or inactive ($I_i = 0$), when channels are estimated based on the assumption that AP has no knowledge about the active UEs.

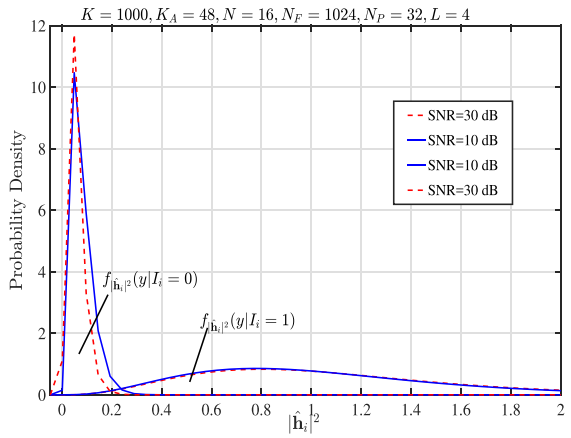


FIGURE 9. PDFs of $|\hat{\boldsymbol{h}}_i|^2$ on condition that UE i is active ($I_i = 1$) or inactive ($I_i = 0$), when channels are estimated based on the assumption that AP has no knowledge about the active UEs.

Below are some exemplified PDFs obtained from simulations, which are shown in Figs. 8–12, when different settings and assumptions are considered. Note that, the PDFs of $|\hat{\boldsymbol{h}}_i|^2$ on condition that UE i is active or inactive can be directly computed from (35) or (36), if we know σ_1^2 or σ_0^2. However, these variances need to be obtained from simulations. Therefore, we directly generate the PDFs from simulations. Specifically, in Figs. 8–10, the PDFs are obtained under the assumption that AP has no knowledge about the active UEs, corresponding to the channel estimation in Section III-B. By contrast, in Figs. 11 and 12, the AP is assumed to know all the active UEs, which can hence jointly exploit the N_P pilot symbols used for channel estimation, as shown in Section III-A. From these PDFs, we can explicitly conceive that if a UE is active, its estimated channel's magnitude, i.e., $|\hat{\boldsymbol{h}}_i|^2$, usually has a relatively high value. By contrast, if a UE is inactive, the value of $|\hat{\boldsymbol{h}}_i|^2$ is usually low and distributed mainly in the region close to zero. This distribution information of $|\hat{\boldsymbol{h}}_i|^2$ on condition of UE i being active or inactive can be exploited for

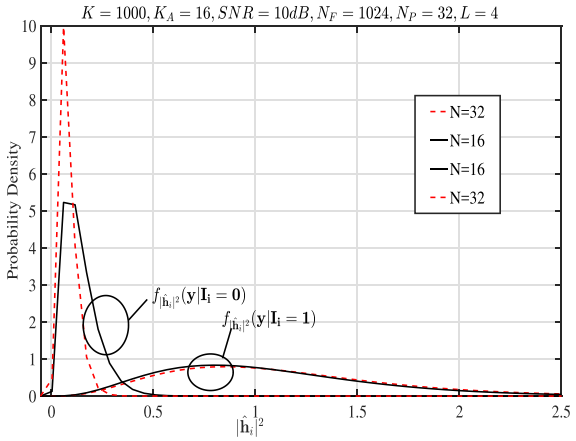


FIGURE 10. PDFs of $|\hat{\boldsymbol{h}}_i|^2$ on condition that UE i is active ($I_i = 1$) or inactive ($I_i = 0$), when channels are estimated based on the assumption that AP has no knowledge about the active UEs.

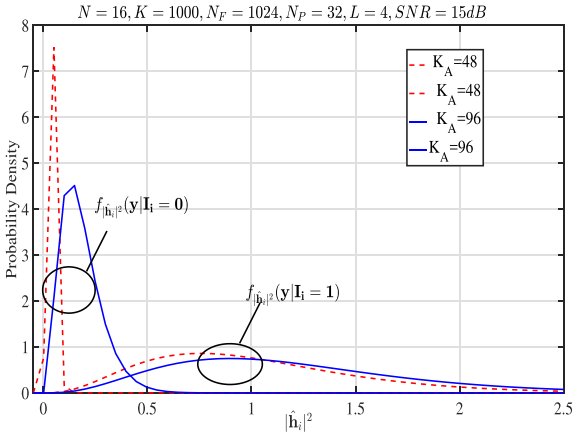


FIGURE 11. PDFs of $|\hat{\boldsymbol{h}}_i|^2$ on condition that UE i is active ($I_i = 1$) or inactive ($I_i = 0$), when channels are estimated based on the assumption that AP knows the active UEs.

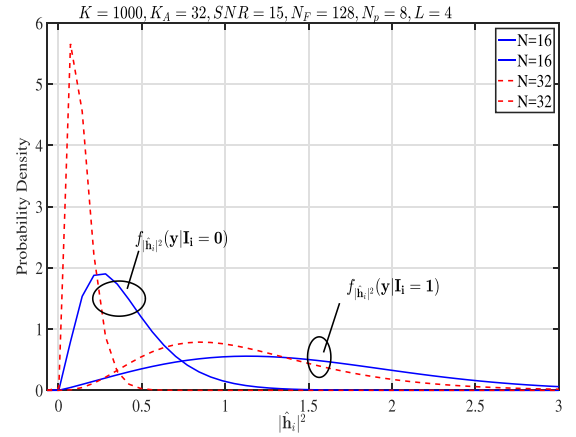


FIGURE 12. PDFs of $|\hat{\boldsymbol{h}}_i|^2$ on condition that UE i is active ($I_i = 1$) or inactive ($I_i = 0$), when channels are estimated based on the assumption that AP knows the active UEs.

the design of the UAI schemes with low-complexity, as shown in the next section.

VI. ACTIVE UE IDENTIFICATION

In this section, we propose a range of UAI schemes designed based on the signal and channel characteristics, and with the aid of the channel estimation approaches considered in Section III. Furthermore, the UAI schemes are analyzed by considering some practical issues, and their performance is investigated and compared in Section VII. Let us first consider the threshold-based UAI (TB-UAI).

A. THRESHOLD-BASED UAI (TB-UAI)

From the statistical analysis in Section V we can explicitly see that $|\hat{\boldsymbol{h}}_i|^2$ usually has a small value, if UE i is inactive. By contrast, if UE i is active, $|\hat{\boldsymbol{h}}_i|^2$ usually takes a relatively big value. These observations straightforwardly imply the threshold-based UAI approach.

Let T_h be a threshold set for the TB-UAI. Then, when given the estimated CIR of $|\hat{\boldsymbol{h}}_i|^2$ for UE i, the TB-UAI identifies the ith UE's state according to

$$\hat{I}_i = \begin{cases} 1, & \text{if} |\hat{\boldsymbol{h}}_i|^2 \geq T_h, \\ 0, & \text{else} \end{cases} \tag{37}$$

where $\hat{I}_i = 1$ or 0 represents UE i is active or inactive.

Note that in the case that AP has no knowledge about the active UEs, the channels used in (37) are estimated using the Estimator-uK. If AP can obtain partial information about the active UEs, the channels in (37) can be estimated by the Estimator-pK.

With the aid of the statistics of $|\hat{\boldsymbol{h}}_i|^2$ given in Section V, we can derive the miss and false-alarm probabilities of the TB-UAI as follows. Firstly, the miss probability conditioned on K_A active UEs is

$$P_M(K_A) = \int_0^{T_h} (\boldsymbol{y}|I_i = 1, K_A) dy \tag{38}$$

Upon substituting (35) into (38) and completing the integration, we obtain

$$P_M(K_A) = 1 - \exp\left(-\frac{LT_h}{1 + L\sigma_1^2(K_A)}\right) \times \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{LT_h}{1 + L\sigma_1^2(K_A)}\right)^k \tag{39}$$

Secondly, the false-alarm probability conditioned on K_A active UEs is

$$P_F(K_A) = \int_{T_h}^{\infty} f_{|\hat{\boldsymbol{h}}_i|^2}(\boldsymbol{y}|I_i = 0, K_A) dy \tag{40}$$

When substituting (36) into (40), we obtain

$$P_F(K_A) = \exp\left(-\frac{T_h}{\sigma_0^2(K_A)}\right) \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{T_h}{\sigma_0^2(K_A)}\right)^k \tag{41}$$

Furthermore, the probability of erroneous identification conditioned on K_A active UEs is

$$P_E(K_A) = P_a P_M(K_A) + (1 - P_a)P_F(K_A) \tag{42}$$

where P_a is the activation probability of a UE.

Assume that UEs become active independently and all have the same activation probability P_a. Then, the number of active UEs K_A at a given time-slot obeys the binomial distribution with the probability mass function (PMF) of

$$P(K_A) = \binom{K}{K_A} P_a^{K_A}(1 - P_a)^{K-K_A}, \quad K_A = 0, 1, \ldots, K \tag{43}$$

With the aid of this distribution, the average miss probability is then given by

$$P_M = \sum_{K_A=1}^{K} P(K_A)P_M(K_A) \tag{44}$$

and the average false-alarm probability is

$$P_F = \sum_{K_A=0}^{K-1} P(K_A)P_F(K_A) \tag{45}$$

Finally, the average probability of erroneous identification is

$$P_E = P_a P_M + (1 - P_a)P_F \tag{46}$$

From (39) and (41) we can know that when T_h increases, the false-alarm probability of (41) reduces, while the miss probability of (39) increases. Therefore, the miss and false-alarm probabilities make trade-off against each other. In the TB-UAI algorithm, the threshold T_h may be chosen according to different requirements. First, it may be selected to minimize the erroneous probability P_E. Second, according to the practical requirement, it is sometimes desirable to have a small false-alarm probability guaranteed. In this case, the threshold can be chosen as $T_h = \arg_{T_h'}\{\max\{P_F(T_h')\} \leq \bar{P}_F\}$, where $\bar{P}_F$ is the maximum false-alarm probability allowed. Furthermore, in some applications, the miss of an active UE may be critical, but some false-alarms can be tolerated. In this case, the threshold can be selected to protect a small miss probability. Hence, when given a small miss probability of $\bar{P}_M$, the threshold is chosen as $T_h = \arg_{T_h'}\{\max\{P_M(T_h')\} \leq \bar{P}_M\}$.

However, in any of the above cases, determining an appropriate threshold for attaining near-optimum performance is highly challenging, as the distribution of $|\hat{\boldsymbol{h}}_i|^2$ is dependent on K_A of the number of active UEs, or on their activation probability P_a, if the average miss and false-alarm probabilities are considered, as shown in Figs. 8–12, and the formulas above in Section VI-A. In order to circumvent these challenges, below we propose the alternative UAI schemes, which are not dependent on threshold setting.

B. EIGEN-ANALYSIS ENHANCED UAI (EAE-UAI)

It is well-known that with the aid of eigen-analysis, an AP may acquire the information about the number of active transmissions [13]. This approach can be introduced to enhance the UAI in our MC-mGFMA system, if the activation probability per UE is small or/and the number of potential UEs is small, resulting in that the number of active UEs has a small probability to exceed N. In detail, the EAE-UAI is operated as follows.

First, the AP estimates the autocorrelation matrix based on (31), which is a $(N \times N)$ symmetric matrix. Since this autocorrelation matrix is estimated by exploiting both pilot symbols and data symbols, i.e., exploiting all the N_F symbols in one frame, it should be sufficiently accurate for deriving the observations via the eigen-analysis on it. Then, according to the matrix theory [53], $\hat{\boldsymbol{R}}_a$ can be decomposed into a signal subspace having the dimensions of K_A if $K_A \leq N$, or of N if $K_A \geq N$. Let us execute the eigenvalue decomposition (EVD) on $\hat{\boldsymbol{R}}_a$, we obtain

$$\hat{\boldsymbol{R}}_a = \boldsymbol{U}_s \boldsymbol{\Lambda}_s \boldsymbol{U}_s^H + \boldsymbol{U}_n \boldsymbol{\Lambda}_n \boldsymbol{U}_n^H \tag{47}$$

where $\boldsymbol{U}_s$ gives the signal subspace and has K_A columns, if $K_A \leq N$, while $\boldsymbol{U}_n$ is the noise (or null) subspace having $(N - K_A)$ columns, also if $K_A < N$. Hence, if $K_A < N$, both $\boldsymbol{\Lambda}_s$ and $\boldsymbol{\Lambda}_n$ in (47) are diagonal matrices with real eigenvalues being the diagonal elements. Let us assume that the N eigenvalues of $\hat{\boldsymbol{R}}_a$ are expressed as $\lambda_1, \lambda_2, \ldots, \lambda_N$. Then, according to the principles of EVD, ideally, we have

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{K_A} > \lambda_{K_A+1} = \lambda_{K_A+2} = \cdots = \lambda_N \tag{48}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_{K_A}$ are the eigenvalues corresponding to the signal subspace or the diagonal elements in $\boldsymbol{\Lambda}_s$. By contrast, $\lambda_{K_A+1}, \ldots, \lambda_N$ are the eigenvalues corresponding to the noise subspace, which form the diagonal elements of $\boldsymbol{\Lambda}_n$. Furthermore, according to the principles of EVD, these eigenvalues can be expressed as

$$\lambda_i = \begin{cases} P_{S_i} + \sigma^2, & \text{if } i \leq K_A \\ \sigma^2, & \text{if } i > K_A \end{cases} \tag{49}$$

where P_{S_i} is the signal power of the ith component, and σ^2 is the noise power.

Hence, with the aid of the eigenvalues obtained from the EVD of $\hat{\boldsymbol{R}}_a$, the number of active UEs K_A can be identified as the number of eigenvalues whose values are larger than the noise power.

In practice, the border between the signal and noise subspaces might not be very clear. In this case, AP may set a threshold based on the noise variance. Any eigenvalues larger than the threshold are identified as corresponding to the active UEs.

Once AP has the knowledge of K_A of the number of active UEs, the EAE-UAI can identify the K_A active UEs as the K_A UEs, whose channel magnitudes $|\hat{\boldsymbol{h}}_i|^2$ are the largest among the K UEs in the system. In summary, the EAE-UAI algorithm is stated Algorithm 1.

Algorithm 1: Eigen-Analysis Enhanced Active UE Identification (EAE-UAI) Algorithm.

Input: $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{N_F}$, $\{\boldsymbol{c}_k\}$, pilot symbols.

1) *Estimate the autocorrelation matrix using (31).*
2) *Derive K_A using the EVD in Section VI-B.*
3) *Estimate the channels of K UEs following Section III (using either Estimator-uK or Estimator-pK, depending on the UE activity knowledge available to AP).*
4) *Order the channels according to their power as* $|\hat{\boldsymbol{h}}_{1'}|^2 \geq |\hat{\boldsymbol{h}}_{2'}|^2 \geq \cdots \geq |\hat{\boldsymbol{h}}_{K'}|^2$, and identify the number of active UEs K_A.
5) *UE Identification*:
if $K_A < N$, the UEs corresponding to the K_A largest terms of $|\hat{\boldsymbol{h}}_{K'}|^2$ are identified as the active UEs;
if $K_A \geq N$, then the threshold-based UAI is used to determine the active UEs.

Note that at the last step concerning $K_A \geq N$, the AP may simply select a few of UEs more than N as the active UEs, under the constraint that the outage probability of

$$P_{out} = \sum_{k=K_A+1}^{K} \binom{K}{k} P_a^k (1-P_a)^{K-k} \tag{50}$$

is below a pre-set value, such as, 10^{-5}. In this way, the probability that active UEs are missed is small. However, there are possibly some false-alarm UEs, which should be readily removed after their data detection and with the aid of the pilots and/or error detection/correction. This is because signal detection of a false-alarm UE should have a bit error rate (BER) of about 0.5, which can be easily identified by the embedded error control coding [54]. Alternatively, the AP receiver can compare the detected pilot symbols of a UE with its actual pilots, and rendering the UE a false-alarm, if the number of erroneously detected pilot symbols is higher than a given value.

Nevertheless, whenever there is a false-alarm UE, the AP receiver has to contribute time and complexity for its signal detection. In order to further improve the UAI in our MC-mGFMA, below we propose the successive interference cancellation assisted UAI (SIC-UAI).

C. SUCCESSIVE INTERFERENCE CANCELLATION ASSISTED UAI (SIC-UAI)

One of the motivations for introducing SIC to the UAI process is to solve the problem that in the EVE-UAI, AP is unable to know the number of active UEs, when $K_A \geq N$, and hence not sure when the identification process should end. Below we propose two SIC-UAI approaches for achieving the UAI objective. With the first approach, EVD is employed to determine whether the UAI process is completed. The second approach does not make use of EVD, instead, it implements pilot detection to check when the UAI process can be finished.

The fundamental principle behind the two approaches is the involvement of the SIC operation. In detail, whenever a UE is identified to be active and its channel is estimated with relatively high reliability, its contribution to the received signals corresponding to pilots can then be removed with the aid of the information available to the AP. In this way, the following UEs' identification would become more reliable.

Specifically, let the ith UE be the identified active UE, and its estimated channel be expressed as $\hat{\boldsymbol{h}}_i$. Since AP knows the pilot symbols sent by UE i and also the spreading codes of UE i, based on (5), AP can carry out the SIC as

$$\boldsymbol{y}'_p = \boldsymbol{y}_p - b_i^{(p)} \boldsymbol{A}_i \hat{\boldsymbol{h}}_i,\ p = 1, 2, \ldots, N_P \tag{51}$$

After the SIC of UE i, the autocorrelation matrix can be updated for the identification of other active UEs.

Initially, let AP use the pilot signals to estimate the autocorrelation matrix as

$$\hat{\boldsymbol{R}}_a = \frac{1}{N_P} \sum_{p=1}^{N_P} \boldsymbol{y}_p \boldsymbol{y}_p^H \tag{52}$$

Note that, in contrast to (31) that can use all the received signals of a frame to estimate $\boldsymbol{R}_a$, here only pilot symbols are used, due to the employment of SIC. Therefore, we can be conceived that this kind of algorithms require that there are sufficient number of pilots, so that $\boldsymbol{R}_a$ can be estimated with sufficient accuracy. As our simulation results show, having 16 pilots is already capable of providing reasonable estimation.

Note furthermore that if there are active UEs known to AP, their contributions to observations can be cancelled by, first, estimating their channels based on the principles of Estimator-pK, and then, carrying out the interference cancellation based on (51). In the following two SIC-UAI schemes, we assume that all active UEs are unknown to AP, and therefore, only the Estimator-uK can be implemented.

With the above preparations, the SIC-UAI algorithms corresponding to the two approaches can now be stated as Algorithm 2 and Algorithm 3, respectively, as follows. In these algorithms, I is the maximum number of iterations allowed, which can be set by assuming a small outage probability, such as 10^{-6}, base on (50).

Note that the SIC-UAIa algorithm may be extended by allowing to cancel several UE signals simultaneously at Step 3). This will accelerate the identification process, but at the cost of some performance loss.

The SIC-UAIa algorithm makes use of the properties of EVD to terminate the identification process. However, if the number of pilots is insufficient, making the estimation of $\boldsymbol{R}_a$ noisy, EVD might not be able to provide a clear boundary between signals and noise. This would degrade the UAI performance. Hence, we propose the SIC-UAI approach-b (SIC-UAIb) as Algorithm-3, which does not invoke EVD. Instead, whenever a UE is identified to be active, after its channel estimation, AP tries to detect the pilot symbols sent by the UE. If the number of erroneously detected pilot symbols is below a threshold value, such as 1/10 of the pilot symbols, the activity

Algorithm 2: Successive Interference Cancellation Assisted UE Identification – Approach-a (SIC-UAIa).

Initialization: $\boldsymbol{y}_1^{(0)} = \boldsymbol{y}_1, \boldsymbol{y}_2^{(0)} = \boldsymbol{y}_2, \cdots \boldsymbol{y}_{N_P}^{(0)} = \boldsymbol{y}_{N_P}$; $\{\boldsymbol{A}_k\}, \{b_{k,n}\}, I$.

Compute autocorrelation matrix $\hat{\boldsymbol{R}}_a^{(0)}$ *using (52).*

For $i = 1, 2, \ldots, I$:

1) *EVD on* $\hat{R}_a^{(i-1)}$ *to obtain N eigenvalues, which are ordered in descending order as* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$.
2) *If it is identified that* $K_A < N$, *active UEs are identified by following the EAE-UAI algorithm. Then, UE identification finishes.*
3) *If* $K_A > N$ *active UEs are implied from the EVD, executing the following operations:*
   - *Estimate the channels of the* $(K - i + 1)$ *unidentified UEs;*
   - *Find the UE with the maximum of* $|\hat{\boldsymbol{h}}_i|^2$, *expressed as* $k' = \arg\max_k\{|\hat{\boldsymbol{h}}_{1'}|^2, |\hat{\boldsymbol{h}}_{2'}|^2, \ldots, |\hat{\boldsymbol{h}}_{(K-i+1)'}|^2\}$;
   - *Interference cancellation of UE* k' *by following (51);*
   - *Update:* $\hat{\boldsymbol{R}}_a^{(i)} = \hat{\boldsymbol{R}}_a^{(i-1)} - \boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}\hat{\boldsymbol{h}}_{k'}^H\boldsymbol{A}_{k'}^H$.

Outputs: Active UEs and their estimated channels.

Algorithm 3: Successive Interference Cancellation Assisted UE Identification – Approach-b (SIC-UAIb).

Initialization: $\boldsymbol{y}_1^{(0)} = \boldsymbol{y}_1, \boldsymbol{y}_2^{(0)} = \boldsymbol{y}_2, \cdots \boldsymbol{y}_{N_P}^{(0)} = \boldsymbol{y}_{N_P}$; $\{\boldsymbol{A}_k\}, \{b_{k,n}\}, I$.

Compute autocorrelation matrix $\hat{\boldsymbol{R}}_a^{(0)}$ *using (52), and* $(\hat{\boldsymbol{R}}_a^{(0)})^{-1}$.

For $i = 1, 2, \ldots, I$:

1) *Estimate the channels of the* $(K - i + 1)$ *unidentified UEs, and find the most reliable active UE as* $k' = \arg\max_k\{|\hat{\boldsymbol{h}}_{1'}|^2, |\hat{\boldsymbol{h}}_{2'}|^2, \ldots, |\hat{\boldsymbol{h}}_{(K-i+1)'}|^2\}$.
2) *Use the estimated channel* $\hat{\boldsymbol{h}}_{k'}$ *and* $\hat{\boldsymbol{R}}_a^{(i-1)}$ *to form* $\boldsymbol{w}_{k'}^{(i-1)}$ *for detecting the* N_P *pilot symbols of UE* k':
   - **if***the number of erroneous pilot symbols is higher than a preset threshold* N_C, *the UAI process finishes;*
   - **if***the number of erroneous pilot symbols is lower than or equal to* N_C, *UE* k' *is identified as active, and execute the following operations:*

   a) Interference cancellation*:* $\boldsymbol{y}_n^{(i)} = \boldsymbol{y}_n^{(i-1)} - b_{k,'n}\boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}$, for$n = 1, 2, \ldots, N_P$;

   b) *Update:* $\hat{\boldsymbol{R}}_a^{(i)} = \hat{\boldsymbol{R}}_a^{(i-1)} - \boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}\hat{\boldsymbol{h}}_{k'}^H\boldsymbol{A}_{k'}^H$, *and* $(\hat{\boldsymbol{R}}_a^{(i)})^{-1} = (\hat{\boldsymbol{R}}_a^{(i-1)})^{-1} + \dfrac{\boldsymbol{w}_{k'}^{(i-1)}(\boldsymbol{w}_{k'}^{(i-1)})^H}{1 - \hat{\boldsymbol{h}}_{k'}^H\boldsymbol{A}_{k'}^H\boldsymbol{w}_{k'}^{(i-1)}}$

   c) *Return 1).*

Outputs: Active UEs and their estimated channels.

of the UE is confirmed. Otherwise, if the error ratio is high, such as near 0.5, the UE is then rendered inactive.

Note that, as shown in Algorithm-3, in Step 2), the SIC-UAIb algorithm detects the most reliably identified UE. In MMSE detection principle, during the ith iteration, the decision variables corresponding to the N_P pilots can be formed as

$$z_{k'}(n) = \left(\boldsymbol{w}_{k'}^{(i-1)}\right)^H \boldsymbol{y}_n^{(i-1)}, n = 1, 2, \ldots, N_P \tag{53}$$

where $\boldsymbol{w}_{k'}^{(i-1)} = (\boldsymbol{R}_a^{(i-1)})^{-1}\boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}$. Then, the decision is made as $\hat{b}_{k,'n} = 1$, if $\Re\{z_{k'}(n)\} > 0$, and $\hat{b}_{k,'n} = -1$, otherwise.

Note furthermore that at Step 2) when the most reliable UE k' is cancelled from the received signals, both the autocorrelation matrix and its inverse are updated. Here, the inverse is obtained by applying the *matrix inverse lemma* on $(\hat{\boldsymbol{R}}_a^{(i-1)} - \boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}\hat{\boldsymbol{h}}_{k'}^H\boldsymbol{A}_{k'}^H)^{-1}$, yielding

$$\left(\hat{\boldsymbol{R}}_a^{(i)}\right)^{-1} = \left(\hat{\boldsymbol{R}}_a^{(i-1)}\right)^{-1} + \frac{\left(\boldsymbol{R}_a^{(i-1)}\right)^{-1}\boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}\left[\left(\boldsymbol{R}_a^{(i-1)}\right)^{-1}\boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}\right]^H}{1 - \hat{\boldsymbol{h}}_{k'}^H\boldsymbol{A}_{k'}^H\left(\boldsymbol{R}_a^{(i-1)}\right)^{-1}\boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}} \tag{54}$$

Substituting $\boldsymbol{w}_{k'}^{(i-1)} = (\boldsymbol{R}_a^{(i-1)})^{-1}\boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}$ into the above equation, we obtain the updating formula in the algorithm. In this way, the algorithm only needs to compute the inverse of autocorrelation matrix once, which allows to significantly reduce complexity.

Algorithm-3 assumes that each iteration identifies only one UE, which results in an UAI delay proportional to the actual number of active UEs. In order to shorten the UAI process, several *most reliable* UEs can be simultaneously identified and processed during one iteration. In this way, the UAI delay can be reduced, but at the cost of some performance loss. Furthermore, in order to mitigate the performance loss as above-mentioned, the number of UEs identified in one iteration can be set in an adaptive way. At first, it may be set to a relatively large value, when the *most reliable* UEs can be easily identified. Then, the number is gradually reduced to improve the reliability of identification of the relatively weak UEs.

D. AUTO-CORRELATION MATRIX EVOLVING UAI (AME-UAI)

As the channel estimation performance results in Section IV suggest, whenever AP employs some knowledge about active UEs, this knowledge can be exploited by the Estimator-pK to improve channel estimation. This property can also be exploited to enhance the performance of UAI, yielding the AME-UAI algorithm, which is described as Algorithm 4.

As shown in Algorithm 4, it first estimates the autocorrelation matrix $\hat{\boldsymbol{R}}_a$. Since there is no interference cancellation operation, this estimation can exploit all the received

Algorithm 4: Auto-Correlation Matrix Evolving UAI (AME-UAI).

Initialization: $\boldsymbol{y}_1^{(0)} = \boldsymbol{y}_1, \boldsymbol{y}_2^{(0)} = \boldsymbol{y}_2, \cdots \boldsymbol{y}_{N_P}^{(0)} = \boldsymbol{y}_{N_P}$; $\{\boldsymbol{A}_k\}$, $\{b_{k,n}\}$, I.

1) Compute autocorrelation matrix $\hat{\boldsymbol{R}}_a$ using (31); 2) Construct $\hat{\boldsymbol{R}}_y^{(0)}$ of (33) with $\hat{\boldsymbol{R}}_a$ being the diagonal elements and zero for all other elements; 3) Compute $(\hat{\boldsymbol{R}}_y^{(0)})^{-1}$.

For $i = 1, 2, \ldots, I$:

1) *Estimate the channels of the $(K - i + 1)$ unidentified UEs based on Estimator-pK in Section III-C, and find the most reliable active UE as $k' = \arg\max_k\{|\hat{\boldsymbol{h}}_{1'}|^2, |\hat{\boldsymbol{h}}_{2'}|^2, \ldots, |\hat{\boldsymbol{h}}_{(K-i+1)'}|^2\}$.*
2) *Use the estimated channel $\hat{\boldsymbol{h}}_{k'}$ and $\hat{\boldsymbol{R}}_a$ to form $\boldsymbol{w}_{k'}^{(i-1)}$ for detecting the N_P pilot symbols of UE k':*
   - **if***the number of erroneous pilot symbols is higher than a preset threshold N_C, the UAI process finishes;*
   - **if***the number of erroneous pilot symbols is lower than or equal to N_C, UE k' is identified as active. Then, execute the following operations:*
     a) *Updating $\hat{\boldsymbol{R}}_y^{(i-1)}$ to $\hat{\boldsymbol{R}}_y^{(i)}$ following (33) by adding the non-diagonal component contributed by UE k';*
     b) *Updating $(\hat{\boldsymbol{R}}_y^{(i-1)})^{-1}$ to $(\hat{\boldsymbol{R}}_y^{(i)})^{-1}$;*
     c) *Return 1).*

Outputs: Active UEs and their estimated channels.

signals in one frame, including both the pilot signals and data carrying signals, i.e., $\hat{\boldsymbol{R}}_a$ can be estimated based on (31). Based on $\hat{\boldsymbol{R}}_a$, the initial auto-correlation matrix $\hat{\boldsymbol{R}}_y^{(0)}$ considering all the N_P pilots can be constructed, with $\hat{\boldsymbol{R}}_a$ being the diagonal block matrices, while all the other elements being zero. Then, the Algorithm forwards to the UAI stage, which iteratively identifies the active UEs and updates the auto-correlation matrix $\hat{\boldsymbol{R}}_y$. In detail, given $\hat{\boldsymbol{R}}_y^{(i-1)}$ obtained at the $(i-1)$-th iteration, the $(K - i + 1)$ unidentified UEs' channels are estimated by following the principles of Estimator-pK. The corresponding weight matrix for the k-th unidentified UE is expressed as

$$\boldsymbol{W}_k = \left(\hat{\boldsymbol{R}}_y^{(i-1)}\right)^{-1} (\boldsymbol{p}_k \otimes \boldsymbol{A}_k)/L \tag{55}$$

Then, the UE has the highest reliability being an active UE is identified. Furthermore, following the identification and with the aid of the estimated channels, the pilot symbols of the identified UE are detected and compared with the pilots symbols of the UE. If the number of errors is high, such as, close to 0.5, the UAI process can be rendered completed. Otherwise, if the number of errors is relatively low, and in favor of a positive identification, the algorithm updates $\hat{\boldsymbol{R}}_y^{(i-1)}$ to $\hat{\boldsymbol{R}}_y^{(i)}$, and prepares to forward to the next iteration.

E. COMPLEXITY ANALYSIS AND DISCUSSION

Above five UAI algorithms have been proposed to carry out UAI (TB-UAI and EAE-UAI) or jointly execute UAI and channel estimation (SIC-UAIa, SIC-UAIb and AME-UAI). In this subsection, we analyze their complexity. In our analysis, we assume that estimating the channel of one UE needs the computations expressed as C_e, so as to focus our attention on the complexity of UAI schemes. Note that in practice, the channel estimation of all UEs can be conducted in parallel, if latency is a critical consideration.

First, for the TB-UAI scheme, estimating K UEs channels requires KC_e computations. Then, computing $|\hat{\boldsymbol{h}}_i|^2$ for K UEs requires $2KL$ real multiplications and $K(L-1)$ real additions. Finally, there are K comparisons required to compare with the threshold. Hence in total, the number of operations required by the TB-UAI scheme is $KC_e + 3KL$. Furthermore, we can be implied from Section III that $C_e >> 3L$. Hence, the complexity of the TB-UAI scheme is determined by KC_e.

Second, for the EAE-UAI algorithm, the autocorrelation matrix has been obtained during the channel estimation stage. Based on the QR algorithm, the EVD algorithm has a complexity of $\mathcal{O}(N^3)$ [55]. Sorting the N eigenvalues using the quick-sort algorithm [56] needs in average $2N \ln N$ comparisons. Again, estimating K UEs channels requires KC_e computations, and computing $|\hat{\boldsymbol{h}}_i|^2$ of K UEs requires $2KL$ real multiplications and $K(L-1)$ real additions. Sorting the K number of $|\hat{\boldsymbol{h}}_i|^2$ needs in average $2K \ln K$ comparison. Hence in total, the complexity of the EAE-UAI algorithm is proportional to $KC_e + N^3 + 2N \ln N + 3KL + 2K \ln K$.

Third, for the SIC-UAIa algorithm, let us assume that in average $\bar{I}$ SIC iterations are implemented, and $K >> I$, due to the fact that the number of potential UEs is usually significantly larger than the number of active UEs. Then, after ignoring those insignificant operations, it can be shown that during these iterations, the complexity is proportional to $\bar{I}(KC_e + N^3 + 2N^2 + NL)$. At the final iteration without SIC, the EAE-UAI algorithm is operated, which has the complexity proportional to $KC_e + N^3 + 2N \ln N + 3KL + 2K \ln K$. Hence, the total number of operations is proportional to $(\bar{I}+1)KC_e + (\bar{I}+1)N^3 + 2N \ln N + 3KL + 2K \ln K + 2\bar{I}N^2 + \bar{I}NL$.

For the SIC-UAIb algorithm, again, we assume that the average number of iterations is $\bar{I}$, and $K >> I$. First, estimating the channels and finding the most reliable one need KC_e and K operations, respectively. Second, computing the weight vector $\boldsymbol{w}_{k'}^{(i-1)}$ needs $2N(N+L)$ multiplications and additions. Detecting N_P pilots needs about $2NN_P$ multiplications. Decision making, finding the number of errors and interference cancellation are all insignificant, and their contribution to complexity can be ignored. Finally, since $\boldsymbol{A}_{k'}\hat{\boldsymbol{h}}_{k'}$ has been computed when computing $\boldsymbol{w}_{k'}^{(i-1)}$, it can be shown that updating the autocorrelation matrix needs about $2N^2$ operations. Hence, the complexity of the SIC-UAIb algorithm is proportional to $\bar{I}(KC_e + 2N(2N + L + N_P))$.

Finally for the AME-UAI algorithm, the complexity is without any doubt dominated by the channel estimation and the inverse of $\hat{\boldsymbol{R}}_y^{(i)}$, which has the dimensions of $(NN_P \times$

TABLE 2. Complexity of UAI Schemes

UAI scheme	Complexity
TB-UAI	KC_e
EAE-UAI	$KC_e + N^3 + 2N\ln N + 3KL + 2K\ln K$
SIC-UAIa	$(\bar{I}+1)KC_e + (\bar{I}+1)N^3 + 2N\ln N$ $+3KL + 2K\ln K + 2\bar{I}N^2 + \bar{I}NL$
SIC-UAIb	$\bar{I}\left(KC_e + 2N(2N + L + N_P)\right)$
AME-UAI	$\bar{I}\left(KC_e + N^3N_P^3\right)$

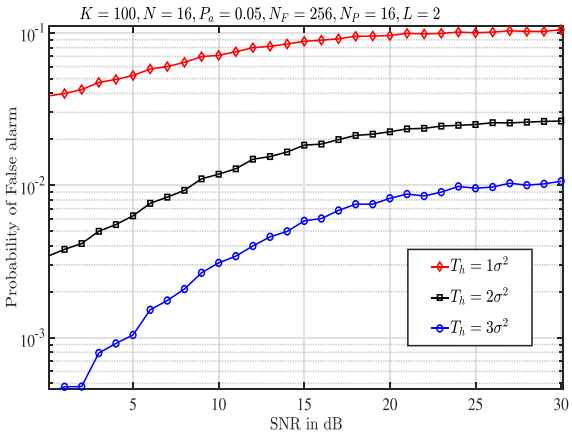
NN_P). Hence, when assuming that the average number of iterations is $\bar{I}$, the computation amount of KC_e for channel estimation per iteration, and the computation requirement of $\mu N^3 N_P^3$ for the inverse of $\hat{\boldsymbol{R}}_y^{(i)}$, the total computations of AME-UAI is proportional to $\bar{I}(KC_e + N^3N_P^3)$.

In summary, the complexity of the UAI schemes are listed in Table 2. Obviously, TB-UAI algorithm has the lowest complexity. The complexity of SIC-UAIa algorithm is higher than that of the EAE-UAI algorithm, as SIC-UAIa algorithm carries out SIC and several EVD operations may be executed during one identification process. When comparing SIC-UAIa with SIC-UAIb algorithms, SIC-UAIa algorithm needs to execute EVD, the number of iterations required by SIC-UAIa algorithm is usually significantly less than that required by SIC-UAIb algorithm. Finally, the AME-UAI has the highest complexity, as it needs a similar number of iterations as the SIC-UAIb algorithm. Furthermore, SIC-UAIb only needs to update the matrix $\hat{R}_a^{(i-1)}$ to $\hat{R}_a^{(i)}$, which is size $(N \times N)$ and can also be achieved using the existing low-complexity algorithm. By contrast, in the AME-UAI algorithm, we did not find an existing algorithm to efficiently update $\hat{R}_a^{(i-1)}$ to $\hat{R}_a^{(i)}$, which has the dimensions of $(NN_P \times NN_P)$.

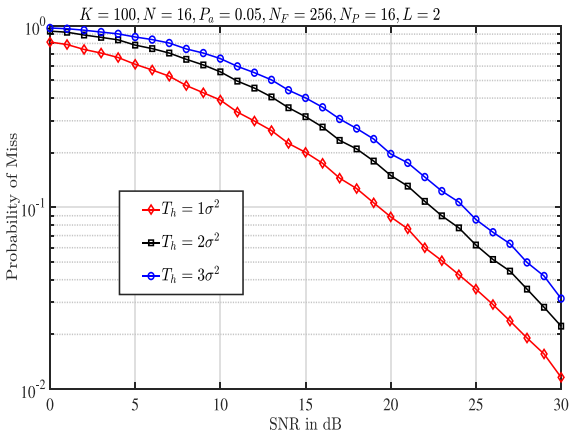
VII. PERFORMANCE OF ACTIVE UE IDENTIFICATION

In this section, the performance of the different UAI algorithms considered in the last section is investigated. We emphasize both the miss and false-alarm probabilities.

In Figs. 13(a) and 13(b), we demonstrate the probabilities of false-alarm and miss of the TB-UAI algorithm, when a MC-mGFMA system with $N = 16$ subcarriers and supports $K = 100$ potential UEs of each having the activation probability of $P_a = 0.05$. Furthermore, as shown in the figures, the frame length is $N_F = 256$ bits, among which $N_P = 16$ pilots are used, and the frequency-selective fading channel has $L = 2$ paths. Additionally, the thresholds are set to $T_h = \sigma^2, 2\sigma^2$ or $3\sigma^2$, where $\sigma^2 = 1/SNR$. The results show that the miss probability increases and simultaneously, the false-alarm probability reduces, as the threshold increases. This implies that there is a trade-off between the miss and false-alarm probabilities, as shown in Fig. 14. Hence, when given a false-alarm (or miss) probability, a suitable threshold may be found to minimize the miss (or false-alarm) probability. However, near optimum threshold is hard to derive, as it is sensitive to the SNR and in particular, to the number of active UEs, which is dynamic in mGFMA systems. The results in Figs. 13(a) and 13(b) also show that for a given $T_h = \sigma^2$, $2\sigma^2$ or $3\sigma^2$,



(a) Probability of false-alarm



(b) Probability of miss

FIGURE 13. Probability of false-alarm and miss, when active UEs are identified using the TB-UAI algorithm.

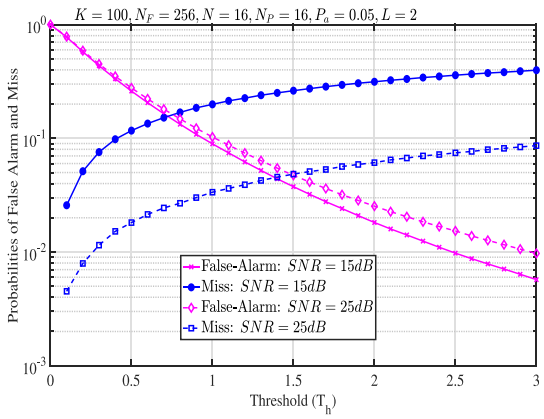
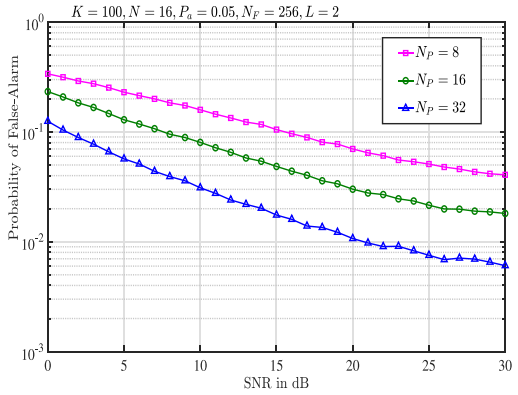
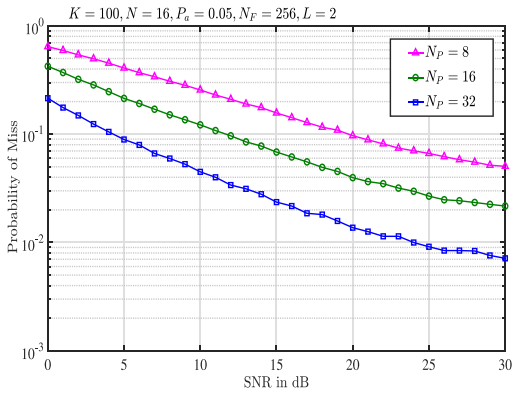


FIGURE 14. Trade-off between the probabilities of false-alarm and miss resulted by the TB-UAI algorithm.

the false-alarm probability increases and the miss probability decreases, as SNR increases. This can be explained with the aid of the statistics of estimated channels, for example, as shown in Fig. 9. As shown in this figure, when SNR is changed

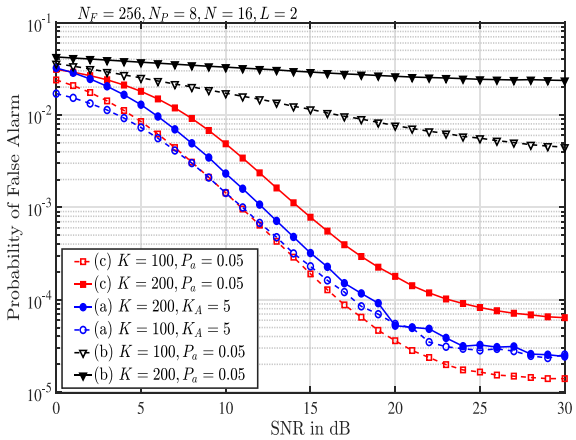


(a) Probability of false-alarm

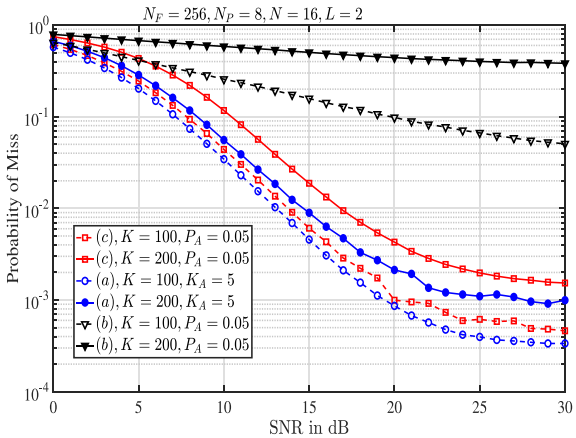


(b) Probability of miss

FIGURE 15. Probability of false-alarm and miss, when active UEs are identified using the EAE-UAI algorithm.



(a) Probability of false-alarm



(b) Probability of miss

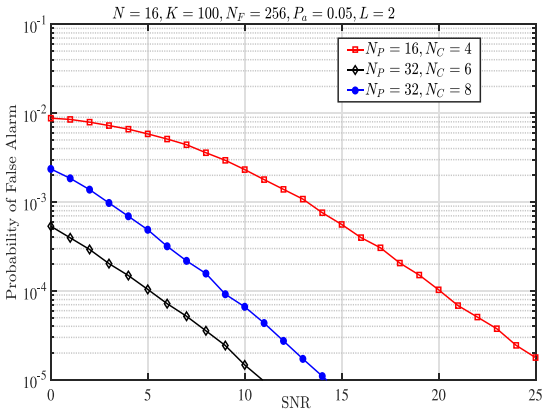
FIGURE 16. Probabilities of false alarm and miss, when active UEs are identified using the SIC-UAIa algorithm in the case: (a) the number of active UEs is known to AP, (b) the number of active UEs is not known to AP, or (c) active UEs are partially known to AP.

from 10 dB to 30 dB, the distributions do not appear significant differences. However, when SNR increases, the threshold $T_h = \sigma^2 = 1/SNR$ significantly reduces. Specifically, the threshold is reduced from 0.1 to 0.001, when SNR is changed from 10 dB to 30 dB. Due to these effects, the false-alarm probability increases and the miss probability decreases, when SNR increases.

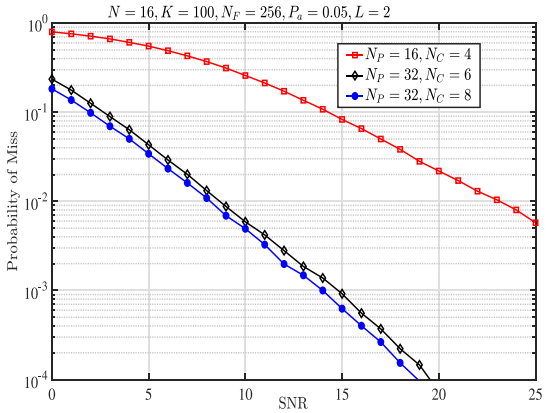
In Fig. 15, we show the probabilities of false-alarm and miss generated by the EAE-UAI algorithm, with the parameters as detailed with the figure. From the results we observe that both the false-alarm and miss probabilities decrease, when SNR increases. This is the result that the EVD in EAE-UAI algorithm allows more confident distinction between signals and noise, when SNR becomes higher. Furthermore, as seen in Fig. 15, both the false-alarm and miss probabilities reduce, when N_P increases. This is because when N_P increases, a more accurate estimation to the autocorrelation matrix can be attained, which in turn enhances the distinction between signals and noise.

In Fig. 16, we depict the probabilities of false-alarm and miss yielded by the SIC-UAIa algorithm, when it is operated with the MC-mGFMA systems with their parameters as shown in the figures. For the SIC-UAIa algorithm, three operational cases are considered, which are: (a) the number of active UEs is known to AP; (b) the number of active UEs is unknown to AP, or (c) the active UEs are partially known to AP. Hence in the case of (a), the AP only needs to identify which are the active UEs. By contrast, in the case of (b), the AP requires to identify both the number of active UEs and who they are. Finally, in the context of the case of (c), we assume that any newly activated UE at a time has the probabilities of P_1, P_2, P_3 and P_4 to transmit 1, 2, 3, or 4 frames. We assume that these probabilities obey the exponential distribution with $P_i = Ce^{-\alpha i}$, $i = 1, 2, 3, 4$, where $C = 0.641$ and $\alpha = 0.4196$. Hence, except the first frame, the AP knows a part of active UEs during all the following frames, which can be exploited to enhance the UAI performance, as analyzed previously in Section VI.

The results of Fig. 16 validate the prediction for the performance of the SIC-UAIa algorithm operated in the case of (c). The partial information about the active UEs can be exploited to significantly enhance the performance of UAI. In general, the performance of case (a) is better than that of case (c), and the performance of case (b) is the worst. However, as seen in Fig. 16(a) for the false-alarm probability, the performance in

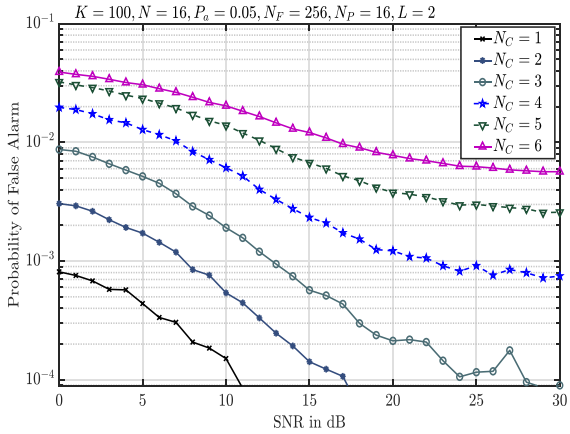


(a) Probability of false-alarm

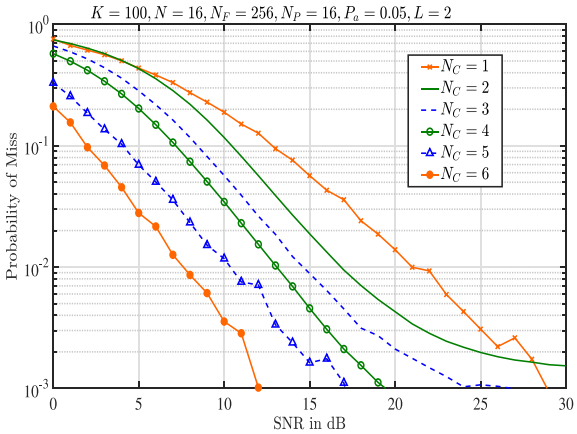


(b) Probability of miss

FIGURE 17. **Probabilities of false alarm and miss, when UEs are identified using the SIC-UAIb algorithm.**



(a) Probability of false-alarm



(b) Probability of miss

FIGURE 18. **Probabilities of false alarm and miss, when UEs are identified using the AME-UAI algorithm.**

the cases of (a) and (c) for $K = 100$ appears crossing. This is the result that the average number of active UEs in the case of (c) is low, and it also has the partial information for the active UEs (known who they are). By contrast, in the case of (a), AP only knows the number of UEs, but not know who they are. Finally, as seen in Fig. 16, all the performance curves present error floor. This is because of channel estimation errors as well as the interference cancellation resulted from incorrect UAI.

In Figs. 17(a) and 17(b), we demonstrate the probabilities of false alarm and miss, respectively, generated by the SIC-UAIb algorithm, when AP has no knowledge about both the active UEs and the number of active UEs. It is shown that when given the ratio of N_C/N_P, transmitting more pilot symbols improve the performance of UAI. By contrast, given the number of pilot symbols to $N_P = 32$, if N_C is reduced from 8 to 6, the miss probability only slightly increases, while the false-alarm probability reduces significantly. Hence, the performance trade-off (between false-alarm and miss) problem of the SIC-UAIb algorithm is much less severer than that of the TB-UAI algorithm.

Finally, Figs. 18(a) and 18(b) depict the false-alarm and miss probabilities of the MC-mGFMA systems employing the AME-UAI algorithm, when the parameters are as shown with the figure. Here we demonstrate the impact of the value of N_c on the false-alarm and miss probabilities. Explicitly, as N_c increases, the false-alarm probability increases, while the miss probability decreases. From the results we can be implied that $N_c = N_P/4$ is a reasonable value for practical operation. As shown in the figure, when $N_c = 4 = N_P/4$, the false-alarm probability is about 10^{-3} at the SNR of 20 dB. Simultaneously, at this SNR, the miss probability is below 10^{-3}. Again, the false-alarm probability is more sensitive to the value of N_C than the miss probability. However, the performance trade-off problem is severer than that of the SIC-UAIb algorithm. In contrast to the TB-UAI algorithm, the desired value of N_C in the AME-UAI algorithm can be relatively easy to find, which can be obtained such as based on numerical simulations.

VIII. CONCLUSION

We have investigated the joint channel estimation and UAI in MC-mGFMA systems. First, the channel estimation is addressed by assuming that AP has the full knowledge, no knowledge or the partial knowledge about active UEs. The studies allow us to reveal the fact that any added knowledge about active UEs can be exploited for enhancing the channel

estimation and furthermore, for the design of novel UAI algorithms. Then, the statistics of the estimated channels of a UE is studied on the condition that the UE is active or inactive. It is shown that the estimated channels of active UEs and inactive UEs present distinctive features, explaining that the estimated channels of UEs can be utilized for UAI. Finally, based on the studies of channel estimation and their statistics, five UAI algorithms have been proposed and their performance has been studied. Our studies show that while TB-UAI has the lowest complexity, it yields a sharp trade-off between false-alarm and miss probabilities, making it hard to be implemented in the practical high-dynamic mGFMA environments. By contrast, the other four UAI algorithms are capable of providing significant performance enhancement in comparison with TB-UAI. Furthermore, the EAE-UAI and SIC-UAIa algorithms impose no trade-off between false-alarm and miss probabilities and the SIC-UAIb algorithm only makes little trade-off between false-alarm and miss probabilities. Although the AME-UAI algorithm brings in some trade-off between false-alarm and miss probabilities, the related parameter N_c can be relatively easily set in mGFMA systems.

In comparison with existing approaches, the UAI algorithms proposed in this paper do not experience the RIP constraint of CS and do not depend on the factor graphs required by the MPA-based approaches. Instead, the signatures considered are random sequences and the number of active UEs can be highly dynamic, making the actual number of active UEs sometimes be significantly higher than the number of resource units of the system. Therefore, it can be expected that our proposed approaches are robust for operation in mGFMA systems.

Our future research will address the joint channel estimation, UAI and the information detection of active UEs, in order to further enhance the performance of mGFMA systems.

REFERENCES

[1] C. Pan, H. Mehrpouyan, Y. Liu, M. Elkashlan, and A. Nallanathan, "Joint pilot allocation and robust transmission design for ultra-dense user-centric TDD C-RAN with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2038–2053, Mar. 2018.

[2] J. Zhang *et al.*, "PoC of SCMA-based uplink grant-free transmission in UCNC for 5G," *IEEE J. on Sel. Areas in Commun.*, vol. 35, no. 6, pp. 1353–1362, June 2017.

[3] M. Masoudi, A. Azari, E. A. Yavuz, and C. Cavdar, "Grant-free radio access IoT networks: Scalability analysis in coexistence scenarios," in *Proc. IEEE Int. Conf. Commun.*, 2018, pp. 1–7.

[4] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surv. Tut.*, pp. 1–1, 2020.

[5] H. Huawei, "R1-164036 multiple access for UL small packets transmission," in *3GPP TSG RAN WG1 Meeting 85*. 3GPP, 2016.

[6] A. Bayesteh, E. Yi, H. Nikopour, and H. Baligh, "Blind detection of SCMA for uplink grant-free multiple-access," in *Proc. 11th Int. Symp. Wireless Commun. Syst.*, Aug. 2014, pp. 853–857.

[7] B. Wang, L. Dai, Y. Zhang, T. Mir, and J. Li, "Dynamic compressive sensing-based multi-user detection for uplink grant-free NOMA," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2320–2323, Nov. 2016.

[8] B. Wang, L. Dai, T. Mir, and Z. Wang, "Joint user activity and data detection based on structured compressive sensing for NOMA," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1473–1476, 2016.

[9] Z. Zhang, X. Wang, Y. Zhang, and Y. Chen, "Grant-free rateless multiple access: A novel massive access scheme for Internet of Things," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2019–2022, 2016.

[10] Y. Du *et al.*, "Efficient multi-user detection for uplink grant-free NOMA: Prior-information aided adaptive compressive sensing perspective," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2812–2828, Dec. 2017.

[11] L. Liu and W. Yu, "Massive connectivity with massive MIMO-Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, 2018.

[12] F. Wei, W. Chen, Y. Wu, J. Ma, and T. A. Tsiftsis, "Message-passing receiver design for joint channel estimation and data decoding in uplink grant-free SCMA systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 167–181, Jan. 2019.

[13] Wei-Chiang Wu and Kwang-ChengChen, "Identification of active users in synchronous CDMA multiuser detection," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 9, pp. 1723–1735, Dec. 1998.

[14] A. Haghighat and M. R. Soleymani, "A subspace scheme for blind user identification in multiuser DS-CDMA," in *Proc. IEEE Wireless Commun. Netw.*, vol. 1, 2003, pp. 688–692, vol. 1.

[15] A. Haghighat and M. R. Soleymani, "A blind music-based algorithm for user identification in multiuser DS-CDMA," in *Proc. IEEE Global Telecommun. Conf. (IEEE Cat. No.03CH37489)*, vol. 5, 2003, pp. 2447–2452 vol. 5.

[16] D. D. Lin and Teng Joon Lim, "Subspace-based active user identification for a collision-free slotted ad hoc network," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 612–621, Apr. 2004.

[17] G. Hannak, M. Mayer, A. Jung, G. Matz, and N. Goertz, "Joint channel estimation and activity detection for multiuser communication systems," in *Proc. IEEE Int. Conf. Commun. Workshop*, 2015, pp. 2086–2091.

[18] C. Wei, H. Liu, Z. Zhang, J. Dang, and L. Wu, "Approximate message passing-based joint user activity and data detection for NOMA," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 640–643, 2017.

[19] B. K. Jeong, B. Shim, and K. B. Lee, "MAP-based active user and data detection for massive machine-type communications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8481–8494, 2018.

[20] A. C. Cirik, N. Mysore Balasubramanya, and L. Lampe, "Multi-user detection using ADMM-based compressive sensing for uplink grant-free NOMA," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 46–49, Feb. 2018.

[21] O. O. Oyerinde, "Compressive sensing algorithms for multiuser detection in uplink grant free NOMA systems," in *Proc. IEEE 89th Veh. Technol. Conf. (VTC2019-Spring)*, 2019, pp. 1–6.

[22] Y. Zhang, Q. Guo, Z. Wang, J. Xi, and N. Wu, "Block sparse Bayesian learning based joint user activity detection and channel estimation for grant-free NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9631–9640, Oct. 2018.

[23] Y. Du, B. Dong, W. Zhu, P. Gao, Z. Chen, X. Wang, and J. Fang, "Joint channel estimation and multiuser detection for uplink grant-free NOMA," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 682–685, Aug. 2018.

[24] J. Zhang, Y. Pan, and J. Xu, "Compressive sensing for joint user activity and data detection in grant-free NOMA," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 857–860, Jun. 2019.

[25] A. T. Abebe and C. G. Kang, "Joint channel estimation and MUD for scalable grant-free random access," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2229–2233, Dec. 2019.

[26] T. Ding, X. Yuan, and S. C. Liew, "Sparsity learning-based multiuser detection in grant-free massive-device multiple access," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3569–3582, Jul. 2019.

[27] G. Chen, Y. Cui, H. V. Cheng, F. Yang, and L. Ding, "Analysis and optimization of successful symbol transmission rate for grant-free massive access with massive MIMO," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2381–2385, Dec. 2019.

[28] X. Miao, D. Guo, and X. Li, "Grant-free NOMA with device activity learning using long short-term memory," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 981–984, Jul. 2020.

[29] W. Kim, Y. Ahn, and B. Shim, "Deep neural network-based active user detection for grant-free NOMA systems," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2143–2155, Apr. 2020.

[30] S. Theodoridis, *Machine Learning: A Bayesian Optimisation Perspective*. London, U.K.: Academic, 2015.

[31] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Magn.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[32] W. Yuan, N. Wu, A. Zhang, X. Huang, Y. Li, and L. Hanzo, "Iterative receiver design for FTN signaling aided sparse code multiple access," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 915–928, Feb. 2020.

[33] Z. Pan, J. Lei, W. Liu, J. Luo, and C. Tang, "Grant-free rateless SCMA for cellular Internet of Things networks," *IEEE Access*, vol. 7, pp. 147 954–147 961, 2019.

[34] A. T. Abebe and C. G. Kang, "Grant-free uplink transmission with multi-codebook-based sparse code multiple access (MC-SCMA)," *IEEE Access*, vol. 7, pp. 169 853–169 864, 2019.

[35] J. Kim, K. Lee, J. Kim, H. Wang, M. Na, and D. Hong, "A novel SCMA system for coexistence of active users and inactive users," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2730–2733, Dec. 2017.

[36] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4464–4478, Jun. 2019.

[37] J. Ahn, B. Shim, and K. B. Lee, "EP-based joint active user detection and channel estimation for massive machine-type communications," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5178–5189, Jul. 2019.

[38] H. Jiang, D. Qu, J. Ding, and T. Jiang, "Multiple preambles for high success rate of grant-free random access with massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4779–4789, Oct. 2019.

[39] J. Ding and J. Choi, "Comparison of preamble structures for grant-free random access in massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 166–170, Feb. 2020.

[40] J. Ding, D. Qu, H. Jiang, and T. Jiang, "Success probability of grant-free random access with massive MIMO," *IEEE Int. Things J.*, vol. 6, no. 1, pp. 506–516, Feb. 2019.

[41] H. Han, Y. Li, W. Zhai, and L. Qian, "A grant-free random access scheme for M2M communication in massive MIMO systems," *IEEE Int. Things J.*, vol. 7, no. 4, pp. 3602–3613, Apr. 2020.

[42] J. Wang, Z. Zhang, and L. Hanzo, "Joint active user detection and channel estimation in massive access systems exploiting Reed-Muller sequences," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 739–752, Jun. 2019.

[43] S. Kim, H. Kim, H. Noh, Y. Kim, and D. Hong, "Novel transceiver architecture for an asynchronous grant-free IDMA system," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4491–4504, Sep. 2019.

[44] J. Fu, G. Wu, Y. Zhang, L. Deng, and S. Fang, "Active user identification based on asynchronous sparse Bayesian learning with SVM," *IEEE Access*, vol. 7, pp. 108 116–108 124, 2019.

[45] Z. Zhang, Y. Li, C. Huang, Q. Guo, C. Yuen, and Y. L. Guan, "DNN-aided block sparse Bayesian learning for user activity detection and channel estimation in grant-free non-orthogonal random access," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12 000–12 012, Dec. 2019.

[46] H. Yu, Z. Fei, Z. Zheng, and N. Ye, "Finite-alphabet signature design for grant-free NOMA: A quantized deep learning approach," *IEEE Trans. Veh. Technol.*, pp. 1–1, 2020.

[47] T. Kim and B. C. Jung, "Performance analysis of grant-free multiple access for supporting sporadic traffic in massive IoT networks," *IEEE Access*, vol. 7, pp. 166 648–166 656, 2019.

[48] L.-L. Yang, *Multicarrier Communications*. Chichester, U.K.: John Wiley & Sons, 2009.

[49] E. Dahlman, S. Parkvall, and J. Skold, *4G LTE/LTE-Advanced for Mobile Broadband*, 2nd ed. Oxford, UK: Academic, 2013.

[50] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1996.

[51] H. V. Henderson and S. R. Searle, "The VEC-permutation matrix, the VEC operator and Kronecker products: A review," *Linear and Multilinear Algebra*, vol. 9, no. 4, pp. 271–288, 1981.

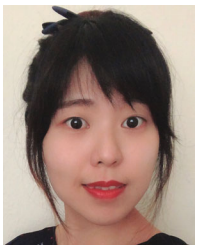
[52] J. G. Proakis and M. Salehi, *Digital Communications*, vol. 4, New York, NY, USA: McGraw-Hill, 2001.

[53] R. A. Brualdi *et al.*, *Combinatorial Matrix Theory*, vol. 39, Berlin, Germany: Springer, 1991.

[54] S. Lin and D. Costello, *Error Control Coding: Fundamentals Appl.*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[55] G. Golub and C. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

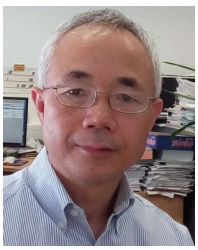
[56] W. Press *et al.*, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed., Cambridge, U.K.: Cambridge Univ. Press, 1992.



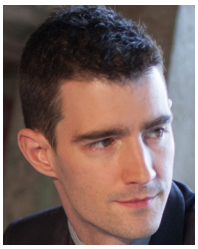
JIATIAN ZHANG received the M.Sc. degree in electrical and telecommunications engineering from Shandong University, Jinan, China, in 2012. Since 2017, she has been working toward the Ph.D. degree with the Next Generation of Wireless Group, University of Southampton, Southampton, U.K. From 2013 to 2015, she worked with the Operation Department, ZTE Corporation as a Senior Engineer across many countries globally. Then, during 2015 and 2017, she worked as a Pre-5G Project Wireless Solution Manager with ZTE Japan. Her current research interests include grant free multiple access for massive machine-type networks and distributed massive MIMO systems.

PENG PAN (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in electronic engineering from Beihang University, Beijing, China, in 2005 and 2011, respectively.

From December 2007 to December 2008, he was a Visiting Ph.D. Student with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K. He is currently an Associate Professor with the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China. His research interests include MIMO systems, multiuser detection, and performance evaluation of wireless systems.



LIE-LIANG YANG (Fellow, IEEE) received the B.Eng. degree in communications engineering from Shanghai TieDao University, Shanghai, China, in 1988, and the M.Eng. and Ph.D. degrees in communications and electronics from Northern (Beijing) Jiaotong University, Beijing, China, in 1991 and 1997, respectively. From June 1997 to December 1997, he was a Visiting Scientist with the Institute of Radio Engineering and Electronics, Academy of Sciences of the Czech Republic. Since December 1997, he has been with the University of Southampton, U.K., where he is currently the Professor of wireless communications with the School of Electronics and Computer Science. His research interest include wireless communications, wireless networks and signal processing for wireless communications, as well as molecular communications and nano-networks. He has authored or coauthored more than 390 research papers in journals and conference proceedings, three books, and also published several book chapters. The details about his research publications can be found at https://www.ecs.soton.ac.uk/people/llyang. He is a fellow of the IET, and was a Distinguished Lecturer of the IEEE VTS. He served as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and *Journal of Communications and Networks* (JCN), and is currently an Associate Editor for the IEEE ACCESS and a subject editor to the *Electronics Letters*.



ROBERT G. MAUNDER (Senior Member, IEEE) has studied with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K., since October 2000. He received the B.Eng. degree (first class Hons.) in electronic engineering in July 2003, and the Ph.D. degree in telecommunications in December 2007. He began a lectureship in November 2007 and was promoted to Associate Professor in March 2013 and to Professor in August 2017. He was awarded Senior Member status of the IEEE in December 2012, Chartered Engineer status of the IET in November 2013 and Fellow status of the IET in January 2017. His research interests include joint source/channel coding and the holistic design of algorithms and hardware implementations for wireless communications. He has published a number of IEEE papers in these areas. He is the Founder and CTO of AccelerComm Ltd, which is commercializing his research as soft-IP.