A Multiuser Detection Algorithm for Random Access Procedure With the Presence of Carrier Frequency Offsets in LTE Systems

Qiwei Wang, Guangliang Ren, *Member, IEEE*, and Jueying Wu

***Abstract*—In the LTE system, uplink synchronization can be established through the random access channel, by which timing and frequency offsets between transceivers can be estimated and adjusted. As the single carrier frequency division multiple access (SC-FDMA) structure is utilized in LTE uplink, preambles for the random access channel should be generated and detected in the transform-domain at user equipments and the base station, respectively. However, in available literature, an explicit transform-domain signal model with respect to timing and frequency offsets has not been provided, and the impact of frequency offsets on the transform-domain signal is not analyzed either. Therefore, suffering from unknown interference among multiple user equipments, existing multiuser detection algorithms provide unsatisfied multiuser detection and estimation performance. In this paper, an explicit transform-domain signal model is first derived with respect to both timing and frequency offsets. The characteristic of transform-domain frequency offsets is then analyzed, which indicates that transform-domain frequency offsets behave differently with time-domain frequency offsets. After that, a multi-steps hybrid multiuser detection algorithm is proposed to improve the multiuser detection and estimation performance, and the Cramër-Rao bound to the accuracy of parameters estimation for the random access procedure is finally obtained. Simulation results show that the proposed algorithm is able to significantly improve the multiuser detection and estimation performance compared to existing ones.**

***Index Terms*—Long term evolution, random access procedure, multiuser detection and estimation, carrier frequency offset.**

I. Introduction

In modern orthogonal frequency division multiple access (OFDMA) systems, synchronization, especially uplink synchronization, is the very priority in establishing stabilized communication links, and enables the transceivers to eliminate timing and frequency uncertainties between them. With accurate estimation of timing offsets, uplink synchronization could also lower the requirement for the length of the cyclic prefix in suppressing inter-symbol interference, thus improving the efficiency of resources utilization.

As for the LTE system, a physical random access channel (PRACH) is designated for user equipments (UEs) to establish uplink synchronization [1]. During the random access (RA) procedure, each UE should notify its network entry by sending a preamble, which is randomly selected from a Zadoff-Chu (ZC) code set [2]–[4] due to the perfect orthogonality/cross-correlation property. As the single carrier frequency division multiple access (SC-FDMA) is adopted in LTE uplink, the PRACH also employs the SC-FDMA structure, by which ZC codes are actually generated in the transform-domain (TrD) [5] and transferred into the time-domain (TD) for transmission. At the receiver, the received RA signal, which is the superposition of transmission signals of all active UEs after channel propagation, should be transferred back into the TrD, and the base station (BS) is required to detect active preambles and estimate their corresponding parameters including timing offsets, frequency offsets and power levels. After that, both the timing and frequency offsets could be eliminated such that uplink synchronization is established. With the rapid development of the internet-of-things, a large amount of UEs, not only user terminals but also radio nodes, need to access into the network [6], [7], indicating that it is a challenging task to improve the capacity and performance of the RA procedure.

In available literature [8]–[11], an implicit assumption has been adopted that frequency uncertainties between transceivers are small enough to be neglected, and that the BS is only required to detect possible preambles and estimate corresponding timing offsets and power levels. By treating other preambles as interference, a single user detection (SUD) algorithm is utilized in the TrD by exploiting a bank of match filters, which correlate the received RA signal with local preambles [8], [9]. If a peak correlation value exceeds a pre-defined threshold, the preamble is deemed as active. It is revealed in [10] that the TrD correlation processing of the SUD can be equivalently realized by operating one inverse discrete Fourier transform (IDFT) to lower the complexity. As the threshold for the SUD is obtained empirically, an adaptive one is proposed in [11] to make the SUD a constant false-alarm rate detector.

Unfortunately, in practical LTE systems, the carrier frequency offset (CFO) in the PRACH is almost inevitable. On one hand, the maximum tolerable frequency uncertainty at the UE and BS are 0.1 and 0.05 ppm, respectively, leading at 2.5 GHz to at most 375 Hz cumulated error. As the PRACH

Manuscript received January 9, 2015; revised May 24, 2015; accepted July 17, 2015. Date of publication July 23, 2015; date of current version September 3, 2015. This work was supported by the National Natural Science Foundation of China (61401321 and 61072102) and the National Key Basic Research Program of China (973 Program, 2014CB340205, 2014CB340206). The associate editor coordinating the review of this paper and approving it for publication was H. Steendam.

The authors are with the State Key Laboratory of ISN, Xidian University, Xi'an 710071, China (e-mail: merling870113@163.com).

Digital Object Identifier 10.1109/TCOMM.2015.2460241

subcarrier spacing is only 1.25 kHz for format 0 to 3, the CFO actually corresponds to 30% of the PRACH subcarrier spacing, indicating that the CFO always takes place unless oscillators with high-precision are employed. On the other hand, in high mobility scenarios, a Rice channel model with a large Rice factor is always assumed so that the Doppler frequency shift on the dominant path becomes the most challenging issue for the PRACH [7]. Therefore, the CFO issue in the PRACH is almost inevitable and needs to be dealt with. In literature [12]–[19], the impact of the CFOs on the RA procedure is studied, indicating that the CFO does have a profound impact on the property of ZC codes, hence a further impact on the performance of the SUD. In [12] and [13], the impact of CFOs on ZC codes is preliminarily analyzed via numerical simulation, which shows that the CFO also leads to a cyclic-shift to ZC codes like timing offsets. When correlating a CFO affected ZC code with the local one, besides the main correlation peak, false correlation peaks also take place. In [14], [15], it is mathematically derived that the cyclic-shift of a CFO affected ZC code corresponds to its root value, which facilitates the determination of locations of false correlation peaks. A restricted set of ZC codes for bearing the impact of CFOs is designed in [16], [17] by taking into account cyclic-shifts which result because of CFOs. Although cyclic-shifts of ZC codes are introduced by CFOs, non-overlapped ZC codes could be generated. Therefore, with the aid of the restricted code set, correlation peaks of different codes are supposed to appear at different locations without confusing, and a multiple window SUD (MW-SUD) algorithm is thus utilized to detect possible codes and estimate corresponding timing offsets, CFOs, and power levels according to the locations and magnitudes of both the main correlation peak and false ones [13], [16]–[19].

However, in literature [8]–[19], there exist three problems that restrict the development of the LTE RA procedure. Firstly, analyses are only derived for simplicity on the basis of the code division multiple access (CDMA) structure, which is inconsistent with the SC-FDMA structure in LTE uplink, leading to a wide gap between analyses and practical implementations. Secondly, because of the lack of an explicit TrD RA signal model, the characteristic of TrD CFOs that are totally different with the TD CFO is not illustrated either. Finally, due to the existence of channel impulse responses (CIRs), timing offsets, and CFOs, the correlation property of ZC codes would be damaged, giving rise to serious interference among them. As the MW-SUD treats other active preambles as interference, its multiuser detection and estimation performance is supposed to be highly deteriorated.

In order to solve the problems stated above, the approaches and contributions of this paper are summarised as follows.

- A TrD RA signal model is derived with respect to TrD CIRs, timing offsets, and CFOs, based on which the characteristic of TrD CFOs is illustrated, showing that TrD CFOs behave totally different with TD CFOs;
- To improve the multiuser detection and estimation performance by suppressing interference among UEs, a multi-steps hybrid multiuser detection (MS-HMUD) algorithm is proposed by exploiting in a hybrid fashion both the methodology of successive interference cancellation and the concept of iterative parameters estimation that is within the framework of the space alternating generalized expectation-maximization (SAGE) algorithm [18];
- The Cramër-Rao bound (CRB) to the accuracy of parameters estimation is also derived.

As both the methodology of interference cancellation and the SAGE algorithm are able to deal with complicated multiuser maximum likelihood estimation (MLE) problems, existing researches in [21]–[28] have exploited them to improve the transmission of multiuser signals in CDMA and OFDMA systems. However, without the assistance of established synchronization and channel estimation, none of them can be applied into the semi-blind multiuser detection scenario, especially that relating to the TrD of the SC-FDMA system in LTE uplink.

This paper is arranged as follows. Section II introduces the signal model of the PRACH in LTE systems briefly. Section III derives the TrD RA signal model, and analyzes the effect of TrD CFOs. In Section IV, the MS-HMUD algorithm is proposed with some issues in detail discussed, and the CRB is also derived. Simulation results are shown in Section V along with discussion, and the conclusion is presented in Section VI.

The following notations are utilized throughout this paper. Matrices and vectors are denoted by symbols in boldface, and variables in italic. $\mathbf{I}_M$ stands for an $M \times M$ identity matrix, and $\mathbf{0}_{M\times N}$ represents an $M \times N$ all-zero matrix. $\mathbf{A} = diag\{a(n), n = 0, 1, \ldots, N-1\}$ denotes an $N \times N$ diagonal matrix with entries $a(n)$ along its main diagonal. Notations $(\cdot)^T$, $(\cdot)^\dagger$, $(\cdot)^H$ are used for transpose, conjugate, and Hermitian transposition, respectively. $((\cdot))_M$ stands for the module operation, and $\mathbf{x}((\tau))_M$ denotes a vector $\mathbf{x}$ having a cyclic shift by (τ mod M) points. $\lceil x \rceil$ is the smallest integer that is not smaller than x, and $\lfloor x \rfloor$ is the largest integer that is smaller than or equal to x. $\varnothing$ is defined as an empty set, and $\mathrm{sinc}(x) = \frac{sin(x)}{x}$ is defined as the Sinc function.

II. Signal Model

The PRACH is taken into consideration according to the LTE standard [1]. The number of total OFDMA subcarriers is N, among which M continuous subcarriers are allocated for UEs to transmit code-division multiplexed preambles. After downlink synchronization with the BS, each UE should send a preamble to notify its network entry. The preamble is randomly selected from a ZC code set $\mathbb{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_v, \ldots, \mathbf{c}_V]$ that is generated in the TrD, and the code $\mathbf{c}_v$ could be represented as

$$\mathbf{c}_v = [c_v(0), \ldots, c_v(m), \ldots, c_v(M-1)]^T, \tag{1}$$

where $c_v(m)$ is the m^{th} element of the code $\mathbf{c}_v$. Note that the code set is generated by using one or more root ZC sequences and their cyclic-shift versions, and that the cyclic-shift versions generated from one root sequence are ideally orthogonal with each other.

Assuming there are B active UEs in the current PRACH, and that the code $\mathbf{c}_{v_b}$ is randomly selected by the b^{th} UE, where $b = 1, \ldots, B$ and $B < V$, this code is then pre-coded by using

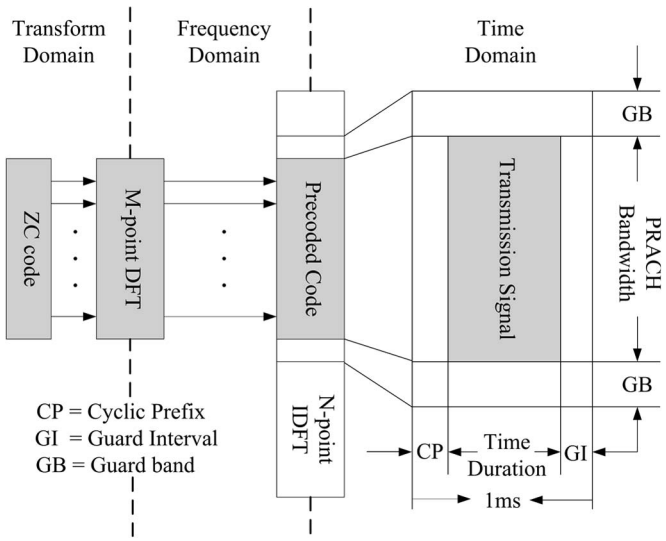

Fig. 1. An illustrative diagram of the PRACH at UE sides.

an M-point discrete Fourier transform (DFT) to form the pre-coded code in the frequency-domain (FD) as

$$\mathbf{C}_{v_b} = diag\left\{C_{v_b}(k), k = 0, 1, \ldots, M-1\right\}. \tag{2}$$

The pre-coded code is then mapped onto designated resources of the PRACH. By utilizing an N-point inverse DFT (IDFT), the transmission signal in the TD could be obtained as

$$s_{v_b}(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{M-1} C_{v_b}(k)\, e^{j2\pi \frac{\mathcal{J}_k}{N} n}, \tag{3}$$

where $0 \leqslant n \leqslant N-1$, and $\{\mathcal{J}_k = \mathcal{J}_0 + k, k = 0, 1, \ldots, M-1\}$ stand for subcarrier indexes corresponding to the bandwidth of the PRACH with $\mathcal{J}_0$ representing the starting index that is designated by the system.

After that, the transmission signal of each UE is extended with a cyclic prefix and a guard interval to occupy a time duration of one sub-frame (i.e., 1 ms). An illustrative diagram about the generation of the transmission signal is shown in Fig. 1. As the PRACH is non-orthogonal but frequency-division multiplexed with the physical uplink share channel (PUSCH), a filter is required to separate them such that signals of PUSCH are not taken into consideration. After propagating through their respective multi-path channels, the transmission signals then arrive at the BS with corresponding round-trip delays (RTDs). By discarding the cyclic prefix and guard interval, the received RA signal in the TD is expressed as

$$y(n) = \sum_{b=1}^{B} \left(e^{j2\pi \frac{n}{N}\epsilon_{v_b} + j\varphi_{v_b}} \sum_{l=d_{v_b}}^{d_{v_b}+L-1} h_{v_b}\left(n, l-d_{v_b}\right) s_{v_b}(n-l) \right) + \omega(n), \tag{4}$$

where d_{v_b}, ϵ_{v_b}, and φ_{v_b} are the RTD, CFO, and initial phase of the b^{th} UE, and $w(n)$ accounts for the additive white Gaussian noise (AWGN) in the TD. Only the integrate part of the RTD is taken into account, while the fractional part is incorporated into the TD CIR. The maximum value of RTDs is given as $d_{\max} = \lceil 2Rf_s/v_c \rceil$, where R is the radius of the cell, f_s is the sampling rate, and v_c is the speed of light. The TD CIR is assumed to to be time-varying with respect to n and have a maximum delay spread of $L-1$ samples such that $h_{v_b}(n, l) = 0$ for $l < 0$ and $l \geqslant L$. The received RA signal in the TD could be represented in the matrix form as

$$\begin{aligned}\mathbf{y} &= \sum_{b=1}^{B} \mathbf{\Gamma}(\epsilon_{v_b}) \left(\mathbf{F}^H \mathbf{C}_{v_b} \mathbf{\Phi}\left(d_{v_b}\right) \mathbf{H}_{v_b} + \overline{\mathbf{F}}^H \boldsymbol{\aleph} \right) + \boldsymbol{\omega} \\ &= \sum_{b=1}^{B} \mathbf{\Gamma}\left(\epsilon_{v_b}\right) \mathbf{F}^H \mathbf{C}_{v_b} \mathbf{\Phi}\left(d_{v_b}\right) \mathbf{H}_{v_b} + \mathbf{w}, \end{aligned} \tag{5}$$

where

- $\mathbf{y} = [y(0), \ldots, y(n), \ldots, y(N-1)]^T$;
- $\boldsymbol{\omega} = [\omega(0), \ldots, \omega(n), \ldots, \omega(N-1)]^T$;
- $\mathbf{\Gamma}(\epsilon_{v_b}) = diag\left\{e^{j2\pi \frac{n}{N}\epsilon_{v_b}}, n = 0, 1, \ldots, N-1\right\}$ is defined as the TD CFO matrix of the current UE;
- $\mathbf{\Phi}(d_{v_b}) = diag\left\{e^{-j2\pi \frac{\mathcal{J}_k}{N} d_{v_b}}, k = 0, 1, \ldots, M-1\right\}$ is defined as the FD phase shift matrix which results because of the RTD d_{v_b};
- $\overline{\mathbf{F}}$ is an N-point DFT matrix, $\mathbf{F}$ takes M rows of $\overline{\mathbf{F}}$ with row indexes of $\{\mathcal{J}_0, \ldots, \mathcal{J}_{M-1}\}$, and the entry (k, n) of $\mathbf{F}$ is represented as $\mathcal{F}(k, n) = \frac{1}{\sqrt{N}} e^{-j2\pi \frac{\mathcal{J}_k}{N} n}$, $k = 0, 1, \ldots, M-1$, $n = 0, 1, \ldots, N-1$.
- $\boldsymbol{\aleph} = [\aleph(0), \ldots, \aleph(f), \ldots, \aleph(N-1)]^T$ is defined as the vector of inter carrier interference (ICI) which results because of the time varying channel, and f stands for the index of all subcarriers;
- $\aleph(f) = \frac{1}{N} \sum\limits_{k=0, f \neq \mathcal{J}_k}^{M-1} C_{v_b}(k) \sum\limits_{n=0}^{N-1} H_{v_b}(n, \mathcal{J}_k) e^{j2\pi \frac{\mathcal{J}_k - f}{N} n}$ is the ICI on the f^{th} subcarrier with $H_{v_b}(n, \mathcal{J}_k) = \sum_{l=d_{v_b}}^{d_{v_b}+L-1} h_{v_b}(n, l-d_{v_b}) e^{j2\pi \frac{\mathcal{J}_k}{N} l}$;
- $\boldsymbol{\xi}_{v_b} = \left[h_{v_b}^{\text{ave}}(0), \ldots, h_{v_b}^{\text{ave}}(l), .., h_{v_b}^{\text{ave}}(L-1)\right]$ represents the average of the TD CIR vector over the time duration of $0 \leq n \leq N-1$, and $h_{v_b}^{\text{ave}}(l) \triangleq \frac{1}{N} \sum_{n=0}^{N-1} h_{v_b}(n, l)$;
- $\mathbf{h}_{v_b} = e^{j\varphi_{v_b}} \left[\mathbf{0}_{1\times d_{v_b}}\ \boldsymbol{\xi}_{v_b}\ \mathbf{0}_{1\times(N-L-d_{v_b})}\right]^T$ stands for the equivalent time-invariant TD CIR of the current UE;
- $\mathbf{H}_{v_b} = [H_{v_b}(\mathcal{J}_0), \ldots, H_{v_b}(\mathcal{J}_k), \ldots, H_{v_b}(\mathcal{J}_{M-1})]^T$ is defined as the vector of equivalent channel frequency responses with $\mathbf{\Phi}(d_{v_b})\mathbf{H}_{v_b} = \sqrt{N} \cdot \mathbf{F}\mathbf{h}_{v_b}$;
- $\mathbf{w} = \sum_{b=1}^{B} \mathbf{\Gamma}(\epsilon_{v_b}) \overline{\mathbf{F}}^H \boldsymbol{\aleph} + \boldsymbol{\omega}$ is the equivalent TD noise, which contains both the AWGN and interference that is resulted by the channel variation, CFO and RTD;

As we mainly focus on the PRACH with the presence of CFOs in medium mobility scenarios, the averages of channel taps over the time duration of $0 \leq n \leq N-1$ are solely taken into account in (5), while the ICI which results because of time-variation of channel taps is neglected and incorporated into the equivalent noise vector $\mathbf{w}$ for simplicity. Note that $h_{v_b}^{\text{ave}}(l)$

only means a time averaging and is different from its ensemble average [29]. It is also worth noticing that in high mobility scenarios (e.g., high-speed trains), a Rice channel model is usually taken into account [7]. In this case, the main challenging issue would also be a large Doppler shift on the dominant path rather than the Doppler spread on scattering paths, indicating that this work is applicable for high mobility scenarios as well.

In this paper, possible collisions when several UEs select the same preamble are neglected because collisions could be detected by a scheduled transmission after the multiuser detection stage. Corresponding to each detected code, a UE should send its device identification on allocated resources in the PUSCH in a manner of hybrid automatic repeat request [30]. In case a collision happens, the colliding UEs would send their device identifications on the same resources such that the BS is able to detect one of the colliding UEs, while other UEs should then re-initiate new RA procedures.

III. Derivation and Analysis of Transform-Domain Random Access Signal and Impact of CFO

As the ZC codes are actually generated in the TrD according to the SC-FDMA structure, while the TrD has similar properties with the TD, the received RA signal in (5) should therefore be transferred back into the TrD for multiuser detection and estimation. In this section, the TrD RA signal model is derived, and the impact of TrD CFOs is then analyzed accordingly.

A. Transform-Domain RA Signal Model

Defining $\mathbf{Q}$ as an $M \times M$ DFT matrix, its entry (k, m) is given as $\mathcal{Q}(k, m) = \frac{1}{\sqrt{M}} e^{-j2\pi \frac{k}{M} m}$, $k = 0, 1, \ldots, M-1$, $m = 0, 1, \ldots, M-1$. As the TrD and FD are related by $\mathbf{Q}$, by using the property that the IDFT of the FD multiplication is equivalent to the TrD cyclic convolution, we have

$$\mathbf{Q}^H \mathbf{C}_{v_b} \mathbf{\Phi}\left(d_{v_b}\right) \mathbf{H}_{v_b} = \mathbf{c}_{v_b} \odot \boldsymbol{g}_{v_b} = \mathbf{\Omega}_{v_b} \boldsymbol{g}_{v_b}, \tag{6}$$

where $\boldsymbol{g}_{v_b} = \sqrt{1/M} \cdot \mathbf{Q}^H \mathbf{\Phi}(d_{v_b}) \mathbf{H}_{v_b} = \sqrt{N/M} \cdot \mathbf{Q}^H \mathbf{F} \mathbf{h}_{v_b} = \sqrt{N/M} \cdot \mathbf{G} \mathbf{h}_{v_b} = [g_{v_b}(0), \ldots, g_{v_b}(m), \ldots, g_{v_b}(M-1)]^T$ is the TrD CIR, $\odot$ stands for the discrete cyclic convolution operation, $\mathbf{\Omega}_{v_b}$ is a Toeplitz matrix defined as $\mathbf{\Omega}_{v_b} = [\boldsymbol{c}_{v_b}((0))_M \quad \boldsymbol{c}_{v_b}((m))_M \quad \boldsymbol{c}_{v_b}((M-1))_M]$, and $\boldsymbol{c}_{v_b}((m))_M = [c_{v_b}(M-m), \ldots, c_{v_b}(0), \ldots, c_{v_b}(M-m-1)]^T$ is a cyclic-shift version of $\boldsymbol{c}_{v_b}$ by m points. $\mathbf{G} \triangleq \mathbf{Q}^H \mathbf{F}$ is defined as an $M \times N$ transfer matrix from the TD to TrD with its entry (m, n) given as [31]

$$\begin{aligned} \mathcal{G}(m, n) &= \sum_{k=0}^{M-1} \mathcal{Q}^{\dagger}(k, m) \mathcal{F}(k, n) \\ &= \frac{1}{\sqrt{M^3 N}} e^{-j2\pi \frac{\mathcal{J}_0}{N} n} e^{j\pi \frac{M-1}{M}\left(m - \frac{M}{N} n\right)} \cdot \chi(m, n), \end{aligned} \tag{7}$$

where

$$\chi(m, n) = \frac{\operatorname{sinc}\left[\pi\left(m - \frac{M}{N} n\right)\right]}{\operatorname{sinc}\left[\frac{\pi}{M}\left(m - \frac{M}{N} n\right)\right]}. \tag{8}$$

As $0 \leqslant m \leqslant M-1$ and $0 \leqslant n \leqslant N-1$, we have $-M + \frac{M}{N} \leqslant m - \frac{M}{N} n \leqslant M-1$. When $n = m = 0$, $\chi(0, 0) = M$. When $m \neq 0$ or $n \neq 0$, $\operatorname{sinc}\left[\frac{\pi}{M}\left(m - \frac{M}{N} n\right)\right] > 0$ is within its main lobe. Therefore, $\chi(m, n)$ is actually a sinc-like function for each possible value of m with its peak value located at $m = \frac{M}{N} n$. By using the transfer function $\mathbf{G}$ as well as the scaling factor M/N, it can be known that the energy of the TrD CIR mainly concentrates on the interval $\mathfrak{L}_0 = \{m = 0, 1, \ldots, \tau_{\max}\}$, where $\mathfrak{L}_0$ is named as the detection interval for the current code, $\tau_{\max} \triangleq \lceil M/N(d_{\max} + L - 1) \rceil \leq M_{cs}$ is deemed as the maximum delay spread of the TrD CIR, and M_{cs} is defined as the length of the cyclic-shift of each code.

According to the derivations above, the TD RA signal in (5) could then be represented in an equivalent way as

$$\mathbf{y} = \sum_{b=1}^{B} \mathbf{\Gamma}\left(\epsilon_{v_b}\right) \mathbf{G}^H \mathbf{\Omega}_{v_b} \boldsymbol{g}_{v_b} + \mathbf{w}. \tag{9}$$

By using the transfer function $\mathbf{G}$ and the TD RA signal, the TrD RA signal is eventually derived as

$$\mathbf{r} = \mathbf{G}\mathbf{y} = \sum_{b=1}^{B} \mathbf{G}\mathbf{\Gamma}\left(\epsilon_{v_b}\right) \mathbf{G}^H \mathbf{\Omega}_{v_b} \boldsymbol{g}_{v_b} + \boldsymbol{\varpi}, \tag{10}$$

where $\boldsymbol{\varpi} = \mathbf{G}\mathbf{w} = [\varpi(0), \ldots, \varpi(m), \ldots, \varpi(M-1)]^T$ is the TrD AWGN. In what follows, both the TD and TrD signal models in (9) and (10) would be utilized for the following analyses and derivations.

B. Analysis of Transform-Domain CFO

In (10), we define $\mathbf{\Psi}_{v_b} \triangleq \mathbf{G}\Gamma(\epsilon_{v_b})\mathbf{G}^H$ as the TrD CFO matrix of the b^{th} UE, and its entry (m, m') is given as

$$\begin{aligned} \Psi_{v_b}(m, m') &= \sum_{n=0}^{N-1} \mathcal{G}(m, n) \mathcal{G}^{\dagger}(m', n) e^{j2\pi \frac{n}{N} \epsilon_{v_b}} \\ &= \frac{1}{M^3 N} \sum_{n=0}^{N-1} e^{j\pi \frac{M-1}{M}(m-m')} \chi(m, n) \chi(m', n) e^{j2\pi \frac{n}{N} \epsilon_{v_b}}. \end{aligned} \tag{11}$$

As $\mathbf{\Psi}_{v_b}$ is not a diagonal matrix, extra interference is supposed to be introduced into the TrD signal. To illustrate the characteristic of the TrD CFO matrix, diagonal elements of $\mathbf{\Psi}_{v_b}$ are first analyzed as follows

$$\Psi_{v_b}(m, m) = \frac{1}{M^3 N} \sum_{n=0}^{N-1} \chi^2(m, n) e^{j2\pi \frac{n}{N} \epsilon_{v_b}}. \tag{12}$$

It is seen that each diagonal entry of $\mathbf{\Psi}_{v_b}$ is a summation of elements of a linear TD CFO weighted by $\chi^2(m, n)$, and the weighting function $\chi^2(m, n)$ is given as

$$\chi^2(m, n) = \frac{\operatorname{sinc}\left[\pi\left(m - \frac{M}{N} n\right)\right]^2}{\operatorname{sinc}\left[\frac{\pi}{M}\left(m - \frac{M}{N} n\right)\right]^2}. \tag{13}$$

As diagonal entries of the TrD and TD CFO matrices are related by $\chi^2(m,n)$, where m stands for TrD samples and n TD samples, we can replace TD samples n with a continuous variable $t \in [0, N-1]$, which makes $\chi^2(m,n)$ time-continuous in the TD, i.e.,

$$\chi^2(m,t) = \frac{\operatorname{sinc}\left[\pi\left(m - \frac{M}{N}t\right)\right]^2}{\operatorname{sinc}\left[\frac{\pi}{M}\left(m - \frac{M}{N}t\right)\right]^2}. \tag{14}$$

In order to fully represent the characteristic of the TrD CFO matrix, we re-sample $\chi^2(m,t)$ by slightly changing the sampling rate on n. Assuming the sampling interval is $\Delta = \frac{N}{M\varrho}$, where $\varrho \geq 1$ is an integer, there are total $M\varrho$ samples, and the sampling points are expressed as $t_{n_\kappa} = n_\kappa\Delta = \frac{N}{M}\cdot\frac{n_\kappa}{\varrho}$ with $n_\kappa = 0, 1, \ldots, M\varrho - 1$. We represent $n_\kappa/\varrho = \kappa + \iota/\varrho$, where $\kappa = \lfloor n_\kappa/\varrho \rfloor \in \{0, 1, \ldots, M-1\}$ stands for the integrate part, and $\iota \in \{0, 1, \ldots, \varrho - 1\}$ is a multiplier of $\frac{1}{\varrho}$ to represent the fractional part. Therefore, we have $t_{n_\kappa} = n_\kappa\Delta = \frac{N}{M}\cdot\frac{n_\kappa}{\varrho} = \frac{N}{M}(\kappa + \frac{\iota}{\varrho})$. When $n_\kappa/\varrho \in \{0, 1, \ldots, M-1\}$, i.e., $t_{n_\kappa} \in \frac{N}{M}\{0, 1, \ldots, M-1\}$, $\chi^2(m, t_{n_\kappa})$ achieves its peak or zero-crossing points with $\varrho - 1$ samples located between any two neighboring zero-crossing points or between the peak and its adjacent zero-crossing points. By using the resampled signals, the diagonal entries of $\mathbf{\Psi}_{v_b}$ could then be re-expressed as

$$\begin{aligned}\Psi_{v_b}(m,m) &= \frac{1}{M^3N}\sum_{n_k=0}^{M\varrho-1}\chi^2(m, n_k\Delta)e^{j2\pi\frac{n_k\Delta}{N}\epsilon_{v_b}} \\ &= \frac{1}{M^3N}\sum_{\kappa=0}^{M-1}\sum_{\iota=0}^{\varrho-1}\chi^2\left(m, \frac{N}{M}\left(\kappa + \frac{\iota}{\varrho}\right)\right)e^{j2\pi\frac{1}{M}\left(\kappa+\frac{\iota}{\varrho}\right)\epsilon_{v_b}}.\end{aligned} \tag{15}$$

Inspection of (15) shows that for any value of m the diagonal entry $\Psi_{v_b}(m,m)$ is a summation of $M\varrho$ quantities. By only taking into account the re-sampled diagonal entries $\Psi_{v_b}(m,m)$, a new diagonal matrix is formed as

$$\mathbf{D}_{v_b} = diag\left\{\Psi_{v_b}(m,m), m = 0, 1, \ldots, M-1\right\}. \tag{16}$$

To proceed further the derivations above, it is worth noting that the peak value of $\chi^2\left(m, \frac{N}{M}(\kappa + \frac{\iota}{\varrho})\right)$ actually relates to the value of m. Corresponding to any value of m, we can further express $\kappa + \iota/\varrho = ((m + \kappa' + \iota'/\varrho))_M$, where κ' and ι' are also chosen from $\{0, 1, \ldots, M-1\}$ and $\{0, 1, \ldots, \varrho - 1\}$ respectively, and the matrix $\mathbf{D}_{v_b}$ could thus be decomposed as

$$\begin{aligned}\mathbf{D}_{v_b} = \frac{1}{M^3N}\sum_{\kappa'=0}^{M-1}\sum_{\iota'=0}^{\varrho-1}\chi^2\left(m, \frac{N}{M}\left((m + \kappa' + \iota'/\varrho)\right)_M\right) \\ \times\, \boldsymbol{\xi}_{v_b}(\kappa', \iota'),\end{aligned} \tag{17}$$

where

$$\boldsymbol{\xi}_{v_b}(\kappa', \iota') = diag\left\{e^{j2\pi\frac{\left((m+\kappa'+\iota'/\varrho)\right)_M}{M}\epsilon_{v_b}}, m = 0, \ldots, M-1\right\}. \tag{18}$$

Similar to the TD CFO matrix $\mathbf{\Gamma}(\epsilon_{v_b})$, $\boldsymbol{\xi}_{v_b}(\kappa', \iota')$ is actually the TrD CFO matrix with respect to m and has an initial phase relating to both κ' and ι'. Therefore, $\mathbf{D}_{v_b}$ could be decomposed into a summation of $M\varrho$ TrD CFOs that are impacted by different initial phases and weighting coefficients of $\chi^2(m, \frac{N}{M}((m + \kappa' + \iota'/\varrho))_M)$, indicating that $\mathbf{D}_{v_b}$ only results in an impact of TrD CFOs rather than extra interference.

On the other hand, defining $\mathbf{E}_{v_b} = \mathbf{\Psi}_{v_b} - \mathbf{D}_{v_b}$, inspection of (11) shows that the peak value of any $\chi(m, n_k\Delta)$ corresponds to a zero-crossing point of $\chi(m', n_k\Delta)$ when $m \neq m'$. Therefore, the quantities $\Psi_{v_b}(m,m')$ are supposed to be much smaller than that of diagonal entries $\Psi_{v_b}(m,m)$, and $\mathbf{E}_{v_b}$ is deemed as the CFO interference matrix. Furthermore, our analyses are based on the resampled quantities of $\chi^2(m, t_{n_\kappa})$ and $e^{j2\pi\frac{t_{n_\kappa}}{N}\epsilon_{v_b}}$ with respect to the fractional sampling interval Δ. As $n \in \mathbb{Z}$ and $n_\kappa \neq n$, the sampling interval should be properly selected to make $\chi^2(m, t_{n_\kappa})$ approach $\chi^2(m,n)$ as close as possible, and the quantization error would also take place due to the mismatch of the sampling intervals, leading to a much more complicated problem. Note that the quantization error is also incorporated into $\mathbf{E}_{v_b}$ as interference, and $\mathbf{D}_{v_b}$ in (17) is utilized for the following analyses.

IV. Proposed Multi-Steps Hybrid Multiuser Detection Algorithm

Generally, during the RA procedure, since the cardinality of the code set is always much larger than the number of active codes and activities of preambles are also blind to the BS, the multiuser detector at the BS is required to be designed in a semi-blind way. In this case, we define the set of all possible code indexes as $\mathbf{\Theta} = \{1, 2, \ldots, V\}$, and the TD RA signal model in (9) could be equivalently re-expressed as

$$\mathbf{y} = \sum_{v\in\mathbf{\Theta}}\mathbf{\Gamma}(\epsilon_v)\mathbf{G}^H\mathbf{\Omega}_v\boldsymbol{g}_v + \mathbf{w}, \tag{19}$$

where the activity, RTD, and power of the code $\mathbf{c}_v$ are hidden in $\boldsymbol{g}_v$. If a code is inactive, its TrD CIR is deemed as $\boldsymbol{g}_v = \mathbf{0}_{M\times 1}$; Otherwise, the activity and parameters of this code could be detected and estimated by exploiting $\boldsymbol{g}_v$.

To detect the activities and estimate the parameters of all possible codes simultaneously, the MLE could be utilized in theory. For a code $\mathbf{c}_v$, its parameters that are required to be estimated are ϵ_v and $\boldsymbol{g}_v$. Define $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_v, \ldots, \epsilon_V]$ and $\boldsymbol{g} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_v, \ldots, \boldsymbol{g}_V]$ as the sets that contain CFOs and TrD CIRs of all possible codes, respectively. The MLE for all possible codes is given as

$$\{\widehat{\boldsymbol{\epsilon}}\ \ \widehat{\boldsymbol{g}}\} = \arg\min\left\|\mathbf{y} - \sum\nolimits_{v\in\mathbf{\Theta}}\mathbf{\Gamma}(\epsilon_v)\mathbf{G}^H\mathbf{\Omega}_v\boldsymbol{g}_v\right\|^2, \tag{20}$$

where $\widehat{\boldsymbol{\epsilon}}$ and $\widehat{\boldsymbol{g}}$ are estimates of $\boldsymbol{\epsilon}$ and $\boldsymbol{g}$, respectively. However, the MLE operation requires an exhaustive search over the multi-dimensional space spanned by $\boldsymbol{\epsilon}$ and $\boldsymbol{g}$. If there are $N_{\widetilde{\epsilon}}$ possible values of ϵ, and $N_{\widetilde{g}}$ possible values for each $g_v(m)$, the total complexity would be $N_{\widetilde{g}}^{VM}N_{\widetilde{\epsilon}}^{V}$, which shows that the MLE is prohibitively complex for practical implementation.

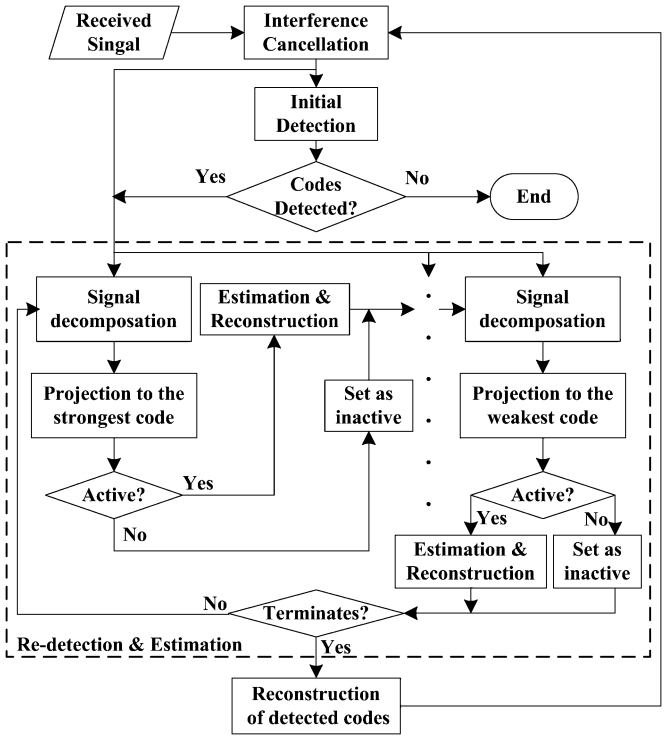

Fig. 2. A flow chart of the proposed MS-HMUD algorithm.

To circumvent this tractable problem, the MW-SUD is usually utilized [13], [16]–[19], which is actually the MLE for each possible code as follows

$$\{\widehat{\epsilon}_v \quad \widehat{\boldsymbol{g}}_v\} = \arg\min \left\| \mathbf{y} - \boldsymbol{\Gamma}(\epsilon_v)\mathbf{G}^H \boldsymbol{\Omega}_v \boldsymbol{g}_v \right\|^2, \tag{21}$$

where $\widehat{\epsilon}_v$ and $\widehat{\boldsymbol{g}}_v$ are estimates of ϵ_v and $\boldsymbol{g}_v$, respectively. It can be seen that the SUD is operated by deeming the signals of other preambles as interference, which leads to unsatisfied detection and estimation performance. In this case, the MS-HMUD algorithm is proposed to deal with this problem.

A. Multi-Steps Hybrid Multiuser Detection

As there is no extra information but local preambles that can assist in the multiuser detection and estimation, the proposed MS-HMUD algorithm is designed into an adaptive way that consists of semi-blind multiuser detection, hybrid parameters estimation, and interference cancellation. This leads to that the proposed MS-HMUD algorithm consists of *loops* and *steps*, and three steps constitute one loop. A flow chart of the proposed MS-HMUD algorithm is illustrated in Fig. 2, and the processing in each step is summarized as follows.

- At the first step, initial detection is first implemented by using the MW-SUD algorithm [13], [16]–[19], and V_α codes are assumed to be detected. Note that at the first step only a group of codes that are strong enough could be detected, and that the detection results are not guaranteed such that these detected codes should be re-detected by suppressing mutual interference among them.
- At the second step, re-detection and estimation are implemented within the framework of the SAGE algorithm. This further leads to *iterations* and *stages* at the second step, and V_α stages make an iteration with one code re-detected and estimated at each stage. Estimates of parameters are updated over iterations, and the iterations proceed until no more significant changes is observed.
- At the third step, according to the detection and estimation results, interference cancellation could be realized by cancelling the reconstructed signals from the received signal, which makes it possible for weak codes to be detected. The interference cancelled signal is then utilized for the next loop, and the loops proceed until no more codes could be detected by the initial detection.

1) Initial Detection: Define the loop index as i, and that a set of indexes of codes that have been detected in previous loops as $\boldsymbol{\Theta}^{(i-1)} = \{u_a^{(i-1)}, a = 1, \ldots, U^{(i-1)}\} \subset \boldsymbol{\Theta}$, where $U^{(i-1)}$ is the number of detected codes in previous loops. Note that u is utilized to represent indexes of detected codes and v undetected ones, while U is used to stand for the number of detected codes and V undetected ones. With a little distortion in using the mathematical expression, the set of indexes of undetected codes in the i^{th} loop is defined as $\boldsymbol{\Theta} \setminus \boldsymbol{\Theta}^{(i-1)}$ with its cardinality being $V - U^{(i-1)}$. Defining $\widehat{\epsilon}_u$ and $\widehat{\boldsymbol{g}}_u$, $u \in \boldsymbol{\Theta}^{(i-1)}$ as the estimation results of detected codes, the TD interference-cancelled signal in the i^{th} loop could be represented as

$$\begin{aligned}\mathbf{y}^{(i)} &= \mathbf{y} - \sum_{u \in \boldsymbol{\Theta}^{(i-1)}} \boldsymbol{\Gamma}(\widehat{\epsilon}_u)\mathbf{G}^H \boldsymbol{\Omega}_u \widehat{\boldsymbol{g}}_u \\ &= \sum_{v \in \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}^{(i-1)}} \boldsymbol{\Gamma}(\epsilon_v)\mathbf{G}^H \boldsymbol{\Omega}_v \boldsymbol{g}_v + \mathbf{w}^{(i)},\end{aligned} \tag{22}$$

where $\mathbf{w}^{(i)}$ stands for the equivalent TD noise that contains the AWGN and residual interference of cancelled codes, i.e.,

$$\mathbf{w}^{(i)} = \mathbf{w} + \sum_{u \in \boldsymbol{\Theta}^{(i-1)}} \left\{ \boldsymbol{\Gamma}(\epsilon_u)\mathbf{G}^H \boldsymbol{\Omega}_u \boldsymbol{g}_u - \boldsymbol{\Gamma}(\widehat{\epsilon}_u)\mathbf{G}^H \boldsymbol{\Omega}_u \widehat{\boldsymbol{g}}_u \right\}. \tag{23}$$

As the TrD signal relates to the TD signal by the transfer function $\mathbf{G}$, and the TrD CFO matrix $\boldsymbol{\Psi}_v$ could be decomposed into $\mathbf{D}_v$ and $\mathbf{E}_v$, in the i^{th} loop, the TrD interference-cancelled signal could then be expressed as

$$\begin{aligned}\mathbf{r}^{(i)} = \mathbf{G}\mathbf{y}^{(i)} &= \sum_{v \in \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}^{(i-1)}} \boldsymbol{\Psi}_v \boldsymbol{\Omega}_v \boldsymbol{g}_v + \mathbf{G}\mathbf{w}^{(i)} \\ &= \sum_{v \in \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}^{(i-1)}} \mathbf{D}_v \boldsymbol{\Omega}_v \boldsymbol{g}_v + \boldsymbol{\varpi}^{(i)},\end{aligned} \tag{24}$$

where $\boldsymbol{\varpi}^{(i)}$ stands for the equivalent TrD noise that contains not only the AWGN and residual interference of cancelled codes but also the interference which results because of the interference part of the TrD CFO matrix, i.e.,

$$\boldsymbol{\varpi}^{(i)} = \sum_{v \in \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}^{(i-1)}} \mathbf{E}_v \boldsymbol{\Omega}_v \boldsymbol{g}_v + \mathbf{G}\mathbf{w}^{(i)}. \tag{25}$$

In (24), as it is obvious that $\mathbf{D}_v \boldsymbol{\Omega}_v \boldsymbol{g}_v = \mathbf{D}_v(\mathbf{c}_v \odot \boldsymbol{g}_v) = \sum_{m=0}^{M-1} g_v(m)\mathbf{D}_v \boldsymbol{c}_v((m))_M$, the signal of each possible code can be deemed as a summation of its replicas which are weighted

by both TrD CIRs and TrD CFOs that are further impacted by different initial phases and weighting coefficients.

Assuming the root value of the code $\mathbf{c}_v$ is γ and the length of its cyclic-shift is qM_{cs} with $q \in \{0, 1, \ldots, \lfloor M/M_{cs} \rfloor\}$, the expression of the code is given as

$$c_v(m) = e^{-j\pi\gamma \frac{(m+qM_{cs})(m+qM_{cs}+1)}{M}}. \tag{26}$$

In order to represent the property of ZC codes, we take into consideration one CFO affected ZC code, which is shown in [14], [15] that

$$\begin{aligned}\widetilde{c}_v(m) &\triangleq c_v(m) e^{j2\pi \frac{m}{M}\epsilon_v} \\ &= \psi e^{-j\pi\gamma \frac{(m+qM_{cs}\pm\gamma^{-1})(m+qM_{cs}\pm\gamma^{-1}+1)}{M}} e^{j2\pi \frac{m}{M}(\epsilon_v \pm p)}, \end{aligned} \tag{27}$$

where $\psi = e^{j\frac{\pi}{M}(\pm 2qM_{cs}p + p^2\gamma^{-1} \pm p)}$ is a constant, and $p \in \mathbb{Z}, p > 0$ is introduced to represent an integral multiplier of γ^{-1}.

When $\epsilon_v = \mp p \in \mathbb{Z}, p > 0$, the code $\mathbf{c}_v$ becomes another code $\widetilde{\mathbf{c}}_v$, meaning that the CFO of $\mp p$ causes a cyclic-shift of $\tau_\gamma^{(\pm p)} \triangleq ((\pm p\gamma^{-1}))_M$. Otherwise, when $\epsilon_v \neq \mp p \in \mathbb{Z}, p > 0$, $e^{j2\pi \frac{m}{M}(\epsilon_v \pm p)} \neq 1$ such that the resulted new code is also affected by a CFO of $e^{j2\pi \frac{m}{M}(\epsilon_v \pm p)}$. In other words, if $\epsilon_v \neq 0$, false correlation peaks are supposed to take place in other intervals when correlating a CFO affected code with its local version. In this case, the detection intervals of each possible code become $\mathfrak{L}_{\pm p} = \{m = ((\tau_\gamma^{(\pm p)}))_M, \ldots, ((\tau_\gamma^{(\pm p)} + \tau_{\max}))_M\}$. When $\epsilon_v < p$, $2p + 1$ correlation peaks are possible to take place at $\mathfrak{L}_{p'}, p' = 0, \pm 1, \ldots, \pm p$.

According to the analyses above, the MW-SUD algorithm could be utilized to detect the CFO affected codes in possible intervals $\mathfrak{L}_{p'}, p' = 0, \pm 1, \ldots, \pm p$. By deeming other codes as interference, the MW-SUD is also implemented by exploiting a bank of TrD correlators, and each correlation processing is defined based on the TrD signal model in (24) as

$$\begin{aligned}\mathcal{R}\left(\mathbf{\Omega}_v | \mathbf{r}^{(i)}\right) &= \left[\mathcal{R}_v^{(i)}(0), \ldots, \mathcal{R}_v^{(i)}(m), \ldots, \mathcal{R}_v^{(i)}(M-1)\right]^T \\ &= \mathbf{\Omega}_v^H \cdot \mathbf{r}^{(i)} = \sum_{v \in \mathbf{\Theta} \setminus \mathbf{\Theta}^{(i-1)}} \mathbf{\Omega}_v^H \mathbf{D}_v \mathbf{\Omega}_v \boldsymbol{g}_v + \mathbf{\Omega}_v^H \boldsymbol{\varpi}^{(i)}. \end{aligned} \tag{28}$$

As $\mathbf{D}_v$ is a combination of TrD CFOs that are impacted by different initial phases and weighting coefficients, possible correlation peaks should also appear at $\mathfrak{L}_{p'}, p' = 0, \pm 1, \ldots, \pm p$. The detection metric of the MW-SUD is therefore given as

$$\Upsilon_v^{(i)} = \arg\max_{m \in \mathfrak{L}_{p'}} \frac{1}{\widehat{\sigma}_v^{(i)}} \left| \mathcal{R}_v^{(i)}(m) \right|, p' = 0, \pm 1, \ldots, \pm p, \tag{29}$$

where $\widehat{\sigma}_v^{(i)2}$ is the estimation of noise variance, i.e.,

$$\widehat{\sigma}_v^{(i)2} = \frac{1}{|\Xi_v|} \sum_{m \in \Xi_v} \left| \mathcal{R}_v^{(i)}(m) \right|^2, \tag{30}$$

and Ξ_v is defined as the interval for the v^{th} code to estimate the variance of noises and interference, $|\Xi_v|$ is defined as the cardinality of Ξ_v. Note that Ξ_v is a set of TrD sample indexes and it can be obtained by excluding from indexes of $\{0, 1, \ldots, M-1\}$ all the indexes of cyclic-shift regions of codes that are generated from the same root as the v^{th} code's.

If $\Upsilon_v^{(i)} > \lambda$, this code is deemed as active; Otherwise, it is inactive. As the CFO matrix $\mathbf{D}_v$ is pretty complicated, in available literature [19], estimates of TrD CIRs are unable to be derived, while CFOs, RTDs and power levels are only coarsely estimated according to the locations and magnitudes of peak correlation values. This leads to a fact that the MW-SUD algorithm is unable to adopt the interference cancellation processing due to missing or inaccurate parameters estimation, thus suffering from an inverse impact of serious interference. Therefore, at the second step of the MS-HMUD, a joint multiuser detection and estimation algorithm within the framework of the SAGE algorithm is proposed to re-check the detected codes and estimate their parameters precisely, while suppressing mutual interference among them.

2) Re-Detection and Parameters Estimation: In the i^{th} loop, assuming $V_\alpha^{(i)}$ codes have been detected by the initial detection at the first step, their indexes are defined as a set $\boldsymbol{\vartheta}^{(i)} = \{v_a^{(i)}, a = 1, 2, \ldots, V_\alpha^{(i)}\}$. In order to avoid error propagation, the codes should be re-detected and estimated in an order from strong to weak. As the second step is proceeded over iterations, at the first iteration, the code indexes should be re-ordered as $\boldsymbol{\vartheta}^{(i)} = \{v_a^{(i)}, a = 1, 2, \ldots, V_\alpha^{(i)}\} \rightarrow \boldsymbol{\vartheta}^{(i)} = \{v_\alpha^{(i)(1)}, \alpha = 1, 2, \ldots, V_\alpha^{(i)}\}$ by re-arranging the detection metrics in (29) in a descend order, where $\{v_\alpha^{(i)(j)}, j = 1, \ldots, j_{\max}\}$ are defined as re-ordered code indexes at the j^{th} iteration, and $j_{\max}$ is the maximum number of iterations. Furthermore, as the estimates of all detected codes are updated from one stage to another according to the re-ordered indexes, at the j^{th} iteration, for a code $\mathbf{c}_v$ with $v \in \boldsymbol{\vartheta}^{(i)}$, we define a set of indexes of updated codes as $\boldsymbol{\vartheta}_v^{(i)(j)}$, and that of un-updated codes as $\boldsymbol{\vartheta}_v^{(i)(j-1)}$. Note that $v \notin \boldsymbol{\vartheta}_v^{(i)(j)}$ or $\boldsymbol{\vartheta}_v^{(i)(j-1)}$, that $\boldsymbol{\vartheta}_v^{(i)(j)} \subset \boldsymbol{\vartheta}^{(i)}$ and $\boldsymbol{\vartheta}_v^{(i)(j-1)} \subset \boldsymbol{\vartheta}^{(i)}$, and that $\boldsymbol{\vartheta}_v^{(i)(j)} \cap \boldsymbol{\vartheta}_v^{(i)(j-1)} = \varnothing$.

Unlike the MW-SUD algorithm, since both the TrD CIRs and the CFOs could be estimated and updated alternately by using SAGE algorithm, interference among detected codes could be mitigated to facilitate the re-detection and estimation for each code. Defining the signal of each possible code as $\mathbf{z}_v = \mathbf{\Gamma}(\epsilon_v)\mathbf{G}^H \mathbf{\Omega}_v \boldsymbol{g}_v$ with its estimation at the j^{th} iteration given as $\widehat{\mathbf{z}}_v^{(j)} = \mathbf{\Gamma}(\widehat{\epsilon}_v^{(j)})\mathbf{G}^H \mathbf{\Omega}_v \widehat{\boldsymbol{g}}_v^{(j)}$, the second step is initialized by setting the iteration number $j = 1$ and $\widehat{\mathbf{z}}_v^{(0)} = \mathbf{0}_{N \times 1}$. The expectation step (E-step) and maximization step (M-step) of the MW-HMUD algorithm are presented as follows.

E-Step: At the j^{th} iteration, compute

$$\mathbf{y}_v^{(j)} = \mathbf{y}^{(i)} - \sum_{u \in \boldsymbol{\vartheta}_v^{(i)(j)}} \widehat{\mathbf{z}}_u^{(j)} - \sum_{u \in \boldsymbol{\vartheta}_v^{(i)(j-1)}} \widehat{\mathbf{z}}_u^{(j-1)} = \mathbf{z}_v + \boldsymbol{\eta}_v^{(j)}, \tag{31}$$

where

$$\boldsymbol{\eta}_v^{(j)} = \sum_{u \in \boldsymbol{\vartheta}_v^{(i)(j)}} \left[\mathbf{z}_u - \widehat{\mathbf{z}}_u^{(j)}\right] + \sum_{u \in \boldsymbol{\vartheta}_v^{(i)(j-1)}} \left[\mathbf{z}_u - \widehat{\mathbf{z}}_u^{(j-1)}\right] + \mathbf{w}^{(i)}. \tag{32}$$

Signal decomposition in (31) is first implemented to separate a complicated ML problem into $V_\alpha^{(i)}$ simpler ones. Note that

$\boldsymbol{\eta}_v^{(j)}$ is a disturbance term that consists of AWGN and residual interference at the j^{th} iteration, and entries of $\boldsymbol{\eta}_v^{(j)}$ are nearly complex Gaussian distributed with zero-mean and variance of $\sigma_v^{(j)2}\mathbf{I}_N$ [27]. According to the terminology of the SAGE algorithm, at the j^{th} iteration, $\mathbf{y}_v^{(j)}$ is viewed as the observed data, $\boldsymbol{g}_v$ is deemed as the missing data, and $[\mathbf{y}_v^{(j)}\ \boldsymbol{g}_v]$ is regarded as the complete data. A conditional log-likelihood function of each code is then constructed as follows [27].

$$\Lambda\left(\epsilon_v|\widehat{\epsilon}_v^{(j-1)}\right) \triangleq E_{\boldsymbol{g}_v|\mathbf{y}_v^{(j)},\widehat{\epsilon}_v^{(j-1)}}\left\{ln\left[\mathcal{P}\left(\mathbf{y}_v^{(j)}|\boldsymbol{g}_v,\epsilon_v\right)\right]\right\}. \quad (33)$$

M-Step: Compute

$$\widehat{\epsilon}_v^{(j)} = \arg\max\left|\Lambda\left(\epsilon_v|\widehat{\epsilon}_v^{(j-1)}\right)\right|. \quad (34)$$

As $\mathbf{G}$ is a transfer function and elements of $\boldsymbol{h}_v$ follow complex Gaussian distribution, $\boldsymbol{g}_v = \sqrt{N/M}\cdot\mathbf{G}\boldsymbol{h}_v$ leads to a fact that elements of $\boldsymbol{g}_v$ also follow complex Gaussian distribution. Defining the covariance matrix of $\boldsymbol{g}_v$ as $\mathbf{V}_{\boldsymbol{g}_v} = E\{\boldsymbol{g}_v\boldsymbol{g}_v^H\}$, the probability density function (PDF) of $\boldsymbol{g}_v$ is given as

$$\mathcal{P}(\boldsymbol{g}_v) = \frac{1}{\pi^M det(\mathbf{V}_{\boldsymbol{g}_v})}e^{-\boldsymbol{g}_v^H\mathbf{V}_{\boldsymbol{g}_v}\boldsymbol{g}_v}. \quad (35)$$

In (33), $ln\left[\mathcal{P}\left(\mathbf{y}_v^{(j)}|\boldsymbol{g}_v,\epsilon_v\right)\right]$ and $\mathcal{P}\left(\mathbf{y}_v^{(j)}|\boldsymbol{g}_v,\widehat{\epsilon}_v^{(j-1)}\right)$ have PDFs as follows

$$ln\left[\mathcal{P}\left(\mathbf{y}_v^{(j)}|\boldsymbol{g}_v,\epsilon_v\right)\right] = -Nln\left(\pi\sigma_v^{(j)2}\right) - \frac{1}{\sigma_v^{(j)2}}\left\|\mathbf{y}_v^{(j)} - \mathbf{z}_v\right\|^2 \quad (36)$$

$$\mathcal{P}\left(\mathbf{y}_v^{(j)}|\boldsymbol{g}_v,\widehat{\epsilon}_v^{(j-1)}\right) = \frac{1}{\left(\pi\sigma_v^{(j)2}\right)^N}e^{-\frac{1}{\left(\pi\sigma_v^{(j)2}\right)^N}\left\|\mathbf{y}_v^{(j)} - \mathbf{\Gamma}\left(\widehat{\epsilon}_v^{(j-1)}\right)\mathbf{G}^H\boldsymbol{\Omega}_v\boldsymbol{g}_v\right\|^2}, \quad (37)$$

By using these equations above, it can be obtained that

$$\mathcal{P}\left(\boldsymbol{g}_v|\mathbf{y}_v^{(j)},\widehat{\epsilon}_v^{(j-1)}\right) \propto \mathcal{P}\left(\mathbf{y}_v^{(j)}|\boldsymbol{g}_v,\widehat{\epsilon}_v^{(j-1)}\right)\mathcal{P}(\boldsymbol{g}_v), \quad (38)$$

and after some algebra we can obtain that

$$\mathcal{P}\left(\boldsymbol{g}_v|\mathbf{y}_v^{(j)},\widehat{\epsilon}_v^{(j-1)}\right) \sim \mathcal{N}\left(\boldsymbol{\mu}_v^{(j)},\boldsymbol{\Sigma}_v^{(j)}\right), \quad (39)$$

where

$$\begin{cases}\boldsymbol{\mu}_v^{(j)} &= \frac{1}{M\sigma_v^{(j)2}}\boldsymbol{\Sigma}_v^{(i-1)}\boldsymbol{\Omega}_v^H\mathbf{G}\mathbf{\Gamma}^H\left(\widehat{\epsilon}_v^{(j-1)}\right)\mathbf{y}_v^{(j)}\\ \boldsymbol{\Sigma}_v^{(j)} &= \left(\mathbf{V}_{\boldsymbol{g}_v}^{-1} + \frac{1}{\sigma_v^{(j)2}}\boldsymbol{I}_M\right)^{-1}.\end{cases} \quad (40)$$

Note that the TrD CIR $\boldsymbol{g}_v$ has a conditional expectation of $\boldsymbol{\mu}_v^{(j)}$, which is actually the minimum mean square error (MMSE) estimation of $\boldsymbol{g}_v$ conditioned by $\widehat{\epsilon}_v^{(j-1)}$. However, as the BS is unable to obtain both $\mathbf{V}_{\boldsymbol{g}_v}$ and $\sigma_v^{(j)}$, the least square (LS) estimation of $\boldsymbol{g}_v$ could be utilized instead as follows

$$\widehat{\boldsymbol{g}}_v^{(j)} = \frac{1}{M}\boldsymbol{\Omega}_v^H\left(\mathbf{G}\mathbf{\Gamma}^H\left(\widehat{\epsilon}_v^{(j-1)}\right)\mathbf{y}_v^{(j)}\right) \quad (41)$$

The conditional log-likelihood function is then derived as

$$\Lambda\left(\widetilde{\epsilon}\mid\widehat{\epsilon}_v^{(j-1)}\right) \propto 2\Re\left\{\left(\mathbf{y}_v^{(j)}\right)^H\mathbf{\Gamma}(\widetilde{\epsilon})\left(\mathbf{G}^H\boldsymbol{\Omega}_v\widehat{\boldsymbol{g}}_v^{(j)}\right)\right\}, \quad (42)$$

where $\widetilde{\epsilon}\in\mathfrak{C}$ is a trial of possible CFOs, and $\mathfrak{C}$ is the set of possible CFOs. The estimate of the CFO $\widehat{\epsilon}_v^{(j)}$ at the current iteration could then be obtained by maximizing the conditional log-likelihood function above. Inspection of (41) shows that the TrD CIR is estimated by compensating the TD CFO and correlating the CFO compensated TrD signal with local codes. Also, it is shown in (42) that the CFO could be estimated in a grid-search fashion by correlating the re-constructed TD signal with each possible CFO compensated TD signal.

However, three major problems arise, impeding the implementation of the derived E-Step in (41) and M-Step in (42). Firstly, as the PRACH has no assistance of pilots or training sequences in initial estimation, none of the TrD CIR and CFO could be estimated preferentially due to their mutual dependency as well as the complicated TrD signal model. Secondly, incorrectly compensated CFO in (41) would lead to incorrect TrD CIR estimation, which inversely impacts the re-detection operation. Finally, since the CFO is estimated in a grid-search fashion with $N \gg M$, the TD correlation processing would result in much more computational burden than the TrD correlation processing. In this case, to alleviate the computational complexity, the mutual dependency between the TrD CIR and the CFO is utilized to realize the joint estimation on them, and the E-Step and M-Step could then be re-expressed as follows.

E-Step: At the j^{th} iteration, compute $\mathbf{y}_v^{(j)} = \mathbf{z}_v + \boldsymbol{\eta}_v^{(j)}$ according to (31). The log-likelihood function conditioned by possible values of the CFO is represented as

$$ln\left[\mathcal{P}\left(\mathbf{y}_v^{(j)}|\boldsymbol{g}_v,\widetilde{\epsilon}\right)\right] \propto -\left\|\mathbf{y}_v^{(j)} - \mathbf{\Gamma}(\widetilde{\epsilon})\mathbf{G}^H\boldsymbol{\Omega}_v\boldsymbol{g}_v\right\|^2, \quad (43)$$

According to (41), compute the LS estimation of $\boldsymbol{g}_v$ conditioned by different $\widetilde{\epsilon}$ as

$$\begin{aligned}\widehat{\boldsymbol{g}}_v^{(j)}(\widetilde{\epsilon}) &= \left[\widehat{g}_v^{(j)}(\widetilde{\epsilon},m), m=0,1,\ldots,M-1\right]^T\\ &= \arg\max\left\{ln\left[\mathcal{P}\left(\mathbf{y}_v^{(j)}|\boldsymbol{g}_v,\widetilde{\epsilon}\right)\right]\right\}\\ &= \frac{1}{M}\boldsymbol{\Omega}_v^H\left(\mathbf{G}\mathbf{\Gamma}^H(\widetilde{\epsilon})\mathbf{y}_v^{(j)}\right).\end{aligned} \quad (44)$$

M-Step: Compute

$$\widehat{\epsilon}_v^{(j)} = \arg\max_{\widetilde{\epsilon}\in\mathfrak{C}}\left\{\max_{m\in\mathfrak{L}_0}\left|\widehat{g}_v^{(j)}(\widetilde{\epsilon},m)\right|\right\}, \quad (45)$$

In the M-Step, once the CFO is precisely compensated, only one correlation peak would appear and would only appear in the main interval $\mathfrak{L}_0$. Therefore, estimates of the TrD CIR in $\mathfrak{L}_0$

TABLE I
COMPARISON OF COMPLEXITIES IN TERMS OF THE NUMBER OF REAL MULTIPLICATIONS

MS-HMUD	Complexity	MW-SUD	Complexity
First Step	$\mathcal{O}\left\{\sum_{i=1}^{i_{\max}} 4\left[Nlog_{10}N+\left(V-V^{(i-1)}\right)Mlog_{10}M\right]\right\}$		$\mathcal{O}\left\{4Nlog_{10}N+4VMlog_{10}M\right\}$
Second Step	$\mathcal{O}\left\{\sum_{i=1}^{i_{\max}} 4V_{\alpha}^{(i)} j_{\max}(N_{\tilde{\epsilon}}+1)\left[Nlog_{10}N+Mlog_{10}M\right]\right\}$		
Third Step	$\mathcal{O}\left\{\sum_{i=1}^{i_{\max}} 4V_{\beta}^{(i)}\left[Nlog_{10}N+Mlog_{10}M\right]\right\}$		

are only utilized for the estimation of the CFO in (45), and the detection metric of the current code is then given as

$$\Upsilon_v^{(j)} = \frac{1}{\widehat{\sigma}_v^{(j)}} \max_{m\in\mathfrak{L}_0} \left|\widehat{g}_v^{(j)}\left(\widehat{\epsilon}_v^{(j)}, m\right)\right|, \tag{46}$$

where

$$\widehat{\sigma}_v^{(j)2} = \frac{1}{|\Xi_v|}\sum_{m\in\Xi_v}\left|\widehat{g}_v^{(j)}\left(\widehat{\epsilon}_v^{(j)}, m\right)\right|^2. \tag{47}$$

The estimation of the TrD CIR is finally given as

$$\widehat{\boldsymbol{g}}_v^{(j)} = \begin{cases} \boldsymbol{\Pi}\cdot\widehat{\boldsymbol{g}}_v^{(j)}\left(\widehat{\epsilon}_v^{(j)}\right), & \text{if } \Upsilon_v^{(j)} > \lambda \\ \mathbf{0}_{M\times 1}, & \text{if } \Upsilon_v^{(j)} < \lambda. \end{cases} \tag{48}$$

where $\boldsymbol{\Pi} = diag\{\overbrace{1,1,\ldots,1}^{\tau_{\max}}\ \overbrace{0,0,\ldots,0}^{M-\tau_{\max}}\}$ is a weighting matrix, meaning that only the TrD channel taps in the main interval $\mathfrak{L}_0$ are reserved as the TrD CIR estimation due to fact that the energy of the TrD CIR mainly concentrates in the main interval $\mathfrak{L}_0$. It can be seen in (45)–(48) that both the CFO and the TrD CIR could be jointly estimated, and the detection threshold is also introduced in (48) to implement re-detection for each code. At the end of each iteration, the code indexes should be re-ordered from strong to weak according to (46) as $\boldsymbol{\vartheta}^{(i)} = \{v_{\alpha}^{(i)(j+1)}, \alpha = 1, 2, \ldots, V_{\alpha}^{(i)}\}$.

After $j_{\max}$ iterations, estimates of CFOs, TrD CIRs, RTDs, and power levels of the codes in $\boldsymbol{\vartheta}^{(i)}$ are finally given as

$$\begin{cases} \widehat{\epsilon}_v &= \widehat{\epsilon}_v^{(j_{\max})} \\ \widehat{\boldsymbol{g}}_v &= \widehat{\boldsymbol{g}}_v^{(j_{\max})} \\ \widehat{d}_v &= \arg\max_{m\in\mathfrak{L}_0}\left|\widehat{g}_v^{(j_{\max})}\left(\widehat{\epsilon}_v^{(j_{\max})}, m\right)\right| \\ \widehat{P}_v &= \sum_{m\in\mathfrak{L}_0}\left|\widehat{g}_v^{(j_{\max})}\left(\widehat{\epsilon}_v^{(j_{\max})}, m\right)\right|^2 \end{cases} \tag{49}$$

3) Interference Cancellation and Termination: During the i^{th} loop, assuming $U_{\alpha}^{(i)} (\leq V_{\alpha}^{(i)})$ codes have been deemed as active at the second step, their indexes are defined as $\boldsymbol{\theta}^{(i)} = \{u_{\alpha}^{(i)}, \alpha = 1, 2, \ldots, U_{\alpha}^{(i)}\} \subseteq \boldsymbol{\vartheta}^{(i)}$. According to the estimation results in (49), the signal of each detected code is then reconstructed as $\widehat{\mathbf{z}}_u^{(i)} = \boldsymbol{\Gamma}\left(\widehat{\epsilon}_u\right)\mathbf{G}^H\boldsymbol{\Omega}_u\widehat{\boldsymbol{g}}_u$, $u \in \boldsymbol{\theta}^{(i)}$, and the interference cancellation processing is implemented for the $(i+1)^{\text{th}}$ loop as follows

$$\mathbf{y}^{(i+1)} = \mathbf{y}^{(i)} - \sum_{u\in\boldsymbol{\theta}^{(i)}}\widehat{\mathbf{z}}_u^{(i)} \tag{50}$$

The interference-cancelled signal $\mathbf{y}^{(i+1)}$ is transferred to the first step for the next loop, and the set of detected code indexes up to the $(i+1)^{\text{th}}$ loop is then given as $\boldsymbol{\Theta}^{(i)} = \boldsymbol{\Theta}^{(i-1)} \cup \boldsymbol{\theta}^{(i)}$ with its cardinality becoming $U^{(i)} = U^{(i-1)} + U_{\alpha}^{(i)}$.

The MS-HMUD should operate until no more new codes can be detected by the initial detection. Assuming the MS-HMUD terminates during the $i_{\max}^{\text{th}}$ loop, indexes of detected codes are finally represented as $\boldsymbol{\Theta}^{(i_{\max}-1)}$, and estimates of their parameters are given according to (49) in each loop.

B. Remarks

1) Intuitively speaking, iterations at the second step should terminate when there is no more significant variation observed in the log-likelihood function in (43). To reduce the system complexity, a simpler stopping criterion is used that the iterations should stop at a pre-designed maximum number of iterations $j_{\max}$, and the maximum number of iterations $j_{\max}$ would be further analyzed via simulation results in Section V.
2) The threshold λ should be designed to make the MS-HMUD a constant false-alarm rate detector. If a code is invalid, the detection metrics in (29) and (46) could be deemed as normalized Rayleigh distributed variables. A false-alarm occurs when the detection metric exceeds the threshold, which is given as $\lambda = \mathbb{F}_{\text{Ray}}^{-1}(1-\mathcal{P}_{fa})$, where $\mathbb{F}_{\text{Ray}}^{-1}(\cdot)$ is defined as the normalized inverse cumulative distribution function of a Rayleigh variable, and $\mathcal{P}_{fa}$ is the probability of the desired false-alarm rate.
3) It is worth noting that the processing in (28), (41), and (44) is actually the TrD correlation. However, as the TrD correlation processing also requires too much computational resources, by using the equations in (6), the TrD correlation processing could be equivalently implemented by applying an IDFT on the conjugate-multiplication of one DFT-precoded code with the FD signal [10]. Further, since the inverse fast Fourier transform (IFFT) could significantly reduce the complexity of the IDFT, the FD signal is usually padded to make M a power of 2, and an M-point IFFT is therefore utilized to reduce the complexity. Defining the number of loops as $j_{\max}$ and that the cardinality of $\mathfrak{C}$ as $N_{\tilde{\epsilon}}$, the approximate complexity of the proposed MS-HMUD is given in Table I. Because a real multiplication takes much more hardware resources than that of a real addition, we present the complexity by using the number of real multiplications. The value of $i_{\max}$ and the comparison of complexity between the algorithms would be further analyzed by simulation results in Section V.
4) In Table I, it can be observed that the main computational burden of the proposed algorithm depends on the second

step, at which the number of possible CFOs is the main factor that determines the complexity. This indicates that the number of CFO searches should be managed to reduce the computational burden. To deal with this, we could first coarsely estimate the CFO with a coarse searching step, and then precisely estimate it within its possible range with a precise searching step.

C. Cramër-Rao Bound for LTE Random Access Procedure

In this subsection, as the limiting performance of parameters estimation for the LTE RA procedure has not been provided before, the CRB for the estimation of the CFO, RTD, and power level is derived. To obtain the "best-case" performance and for simplicity, interference among active UEs is neglected, and we only take into account the CRB for a single active UE by simplifying the signal model in (5) as $\mathbf{y} = \mathbf{z}_v + \mathbf{w} = \mathbf{\Gamma}(\epsilon_v)\mathbf{F}^H\mathbf{C}_v\mathbf{\Phi}(d_v)\mathbf{H}_v + \mathbf{w}$. The real and imaginary parts of $\mathbf{H}_v$ are called $\mathbf{H}_v^{\Re}$ and $\mathbf{H}_v^{\Im}$, respectively. Denoting by $\boldsymbol{\zeta} = [\mathbf{H}_v^{\Re}\ \ \mathbf{H}_v^{\Im}\ \ d_v\ \ \epsilon_v]$ the vector of parameters of one active UE, the Fisher information matrix (FIM) is represented as

$$[\mathbf{F}_{IM}]_{i,j} = -E\left\{\frac{\partial^2 ln\mathcal{P}(\mathbf{y}|\boldsymbol{\zeta})}{\partial\zeta_i\partial\zeta_j}\right\}, 1 \le i,j \le 2M+2, \quad (51)$$

where ζ_ℓ is the ℓ^{th} element of $\boldsymbol{\zeta}$. After some common computations, the FIM is given by

$$\mathbf{F}_{IM} = \frac{2}{\sigma_n^2}\begin{pmatrix}\mathbf{I}_{2M} & \mathbf{u}\\ \mathbf{u}^T & \mathbf{U}\end{pmatrix}, \quad (52)$$

where

$$\mathbf{u} = \begin{pmatrix}\Re\{\mathbf{T}^H\mathbf{H}_v\} & \Re\left\{\mathbf{\Phi}^H(d_v)\mathbf{C}_v^H\mathbf{F}\mathbf{L}\mathbf{F}^H\mathbf{C}_v\mathbf{\Phi}(d_v)\mathbf{H}_v\right\}\\ \Im\{\mathbf{T}^H\mathbf{H}_v\} & \Im\left\{\mathbf{\Phi}^H(d_v)\mathbf{C}_v^H\mathbf{F}\mathbf{L}\mathbf{F}^H\mathbf{C}_v\mathbf{\Phi}(d_v)\mathbf{H}_v\right\}\end{pmatrix}$$

$$\mathbf{U} = \begin{pmatrix}\|\mathbf{T}^H\mathbf{H}_v\|^2 & \mu_{d,\epsilon}\\ \mu_{d,\epsilon} & \left\|\mathbf{L}^H\mathbf{F}^H\mathbf{C}_v\mathbf{\Phi}(d_v)\mathbf{H}_v\right\|^2\end{pmatrix}$$

$$\mu_{d,\epsilon} = \Re\left\{\mathbf{H}_v^H\mathbf{\Phi}^H(d_v)\mathbf{C}_v^H\mathbf{F}\mathbf{L}\mathbf{F}^H\mathbf{C}_v\mathbf{\Phi}(d_v)\mathbf{T}\mathbf{H}_v\right\}$$

$$\mathbf{T} = diag\left\{-j2\pi\frac{\mathcal{J}_k}{N}, k = 0, 1, \ldots, M-1\right\}$$

$$\mathbf{L} = diag\left\{j2\pi\frac{n}{N}, n = 0, 1, \ldots, N-1\right\}.$$

As it is known that $\mathbf{H}_v$ also follows complex Gaussian distribution with zero-mean and a variance matrix of $\mathbf{V}_{\mathbf{H}_v} = E\{\mathbf{H}_v\mathbf{H}_v^H\}$, the CRB could be loosened by integrating the FIM over the distribution of $\mathbf{H}_v$, i.e.,

$$\mathbf{F}_{IM} = \frac{2}{\sigma_n^2}\begin{pmatrix}\mathbf{I}_{2M} & \mathbf{0}_{2M\times 2}\\ \mathbf{0}_{2\times 2M} & \mathbf{U}'\end{pmatrix}, \quad (53)$$

where

$$\mathbf{U}' = \begin{pmatrix}tr\left(\mathbf{V}_{\mathbf{H}_v}\mathbf{T}^2\right) & E_{\mathbf{H}_v}\{\mu_{d,\epsilon}\}\\ E_{\mathbf{H}_v}\{\mu_{d,\epsilon}\} & tr\left(\mathbf{\Phi}^H(d_v)\mathbf{C}_v^H\mathbf{F}\mathbf{L}^2\mathbf{F}^H\mathbf{C}_v\mathbf{\Phi}(d_v)\mathbf{V}_{\mathbf{H}_v}\right)\end{pmatrix}. \quad (54)$$

The inverse of the FIM is found to be

$$\mathbf{F}_{IM}^{-1} = \frac{\sigma_n^2}{2}\begin{pmatrix}\mathbf{I}_{2M} & \mathbf{0}_{2M\times 2}\\ \mathbf{0}_{2\times 2M} & \mathbf{U}'^{-1}\end{pmatrix}. \quad (55)$$

Therefore, the CRB to the accuracy of d_v and ϵ_v is given as

$$CRB(d_v) \approx \frac{\sigma_n^2}{2|\mathbf{U}'|}tr\left(\mathbf{\Phi}^H(d_v)\mathbf{C}_v^H\mathbf{F}\mathbf{L}^2\mathbf{F}^H\mathbf{C}_v\mathbf{\Phi}(d_v)\mathbf{V}_{\mathbf{H}_v}\right), \quad (56)$$

$$CRB(\epsilon_v) \approx \frac{\sigma_n^2}{2|\mathbf{U}'|}tr\left(\mathbf{V}_{\mathbf{H}_v}\mathbf{T}^2\right). \quad (57)$$

As it known that $P_v = \frac{1}{M}\|\mathbf{H}_v\|^2$, the CRB to the accuracy of P_v is given as

$$CRB(P_v) \approx \frac{\sigma_n^2}{2}\boldsymbol{\delta}\boldsymbol{\delta}^T = \frac{\sigma_n^2}{2M^2}tr\left(\mathbf{V}_{\mathbf{H}_v}\right), \quad (58)$$

where $\boldsymbol{\delta} = \frac{1}{M}\left[(\mathbf{H}_v^{\Re})^T\ (\mathbf{H}_v^{\Im})^T\ \mathbf{0}_{1\times 2}\right]$ is the gradient of P_v over $\boldsymbol{\zeta}$.

V. Simulation Results

A. Simulation Parameters

According to LTE standards [1], the number of total subcarriers for the PRACH is 12288 with a cyclic prefix composed of 1584 samples, and the sampling frequency is 15.36 MHz. The total bandwidth of the LTE system is 10 MHz, and the carrier frequency is 2.5GHz. The format of the PRACH is 0, and the subcarrier spacing of the PRACH is 1.25 kHz. The bandwidth of the PRACH is 1.08 MHz corresponding to 864 subcarriers, among which 839 are utilized to transmit DFT-precoded ZC codes. The cardinality of the code set is 64, in which 54 codes are utilized in the simulation with other 10 codes preserved for non-contention based RA. The ZC codes are generated according to the restricted code set with the length of the cyclic-shift given as 22, and the cell radius is 2.3 km such that the maximum RTD is 235 samples. Throughout the simulation, the false-alarm rate is kept around 1×10^{-3}, and the channel model is the Extended-ITU Typical Urban channel with the maximum vehicle speed being 60 km/h [3].

In LTE systems, although the initial power control mechanisms has been employed to compensate the path-loss [30], power differences among UEs (i.e., the near-far effect) always exist due to the existence of power estimation errors, shadow fading, small-scale fading, and the power ramping mechanism, etc. In order to represent the near-far effect, the signal-to-noise ratio (SNR) of each UE is uniformly distributed in [0, 10] dB or [0, 20] dB, and the SNR in [0, 20] dB represents more severe near-far effect than the SNR in [0, 10] dB. As stated in the introduction, we consider a CFO of 375 Hz, corresponding to which the maximum CFO in the PRACH is 0.3 (< 1) of the subcarrier spacing. In this case, we have $p = 1$ and $p' = 0, \pm 1$ such that only three intervals are utilized for the initial detection of the MS-HMUD. The coarse searching step is chosen as 0.1 and the precise searching step is chosen as 0.05. In this case, $\widetilde{\epsilon} = \pm 0.3, \pm 0.2, \pm 0.1, 0$ should first be used for a coarse estimation of the CFO, and then the CFO is precisely estimated in possible ranges of $[\widetilde{\epsilon} - 0.05, \widetilde{\epsilon} + 0.05]$ with the

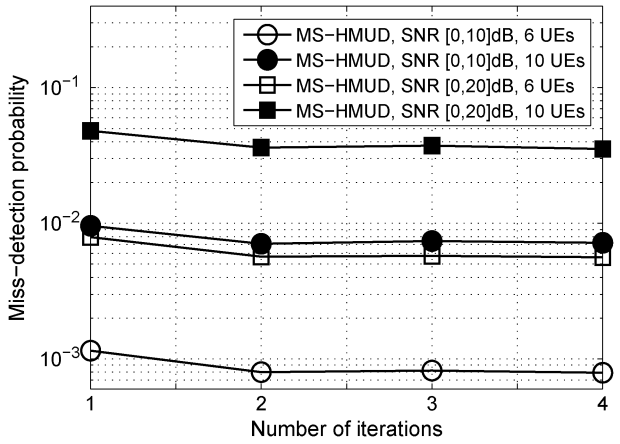

Fig. 3. Miss-detection performance of the MS-HMUD with respect to different numbers of iterations $j_{\max}$ with 6 or 10 active UEs.

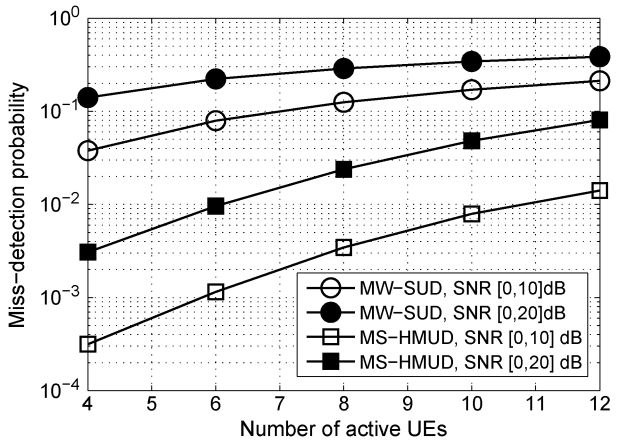

Fig. 5. Comparison of miss-detection performance with $j_{\max} = 1$ and the number of active UEs ranging from 4 to 12.

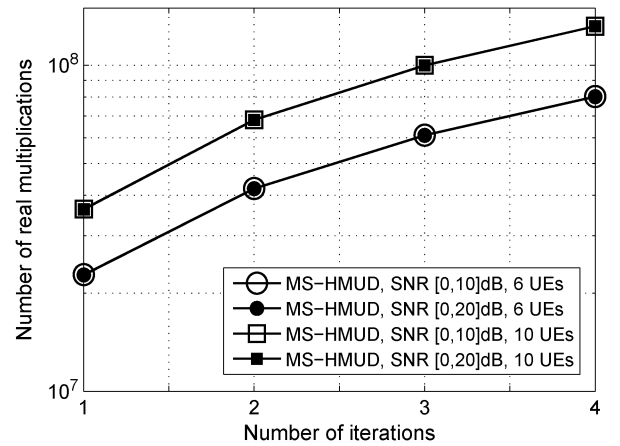

Fig. 4. Complexity of the MS-HMUD with respect to different numbers of iterations $j_{\max}$ with 6 or 10 active UEs.

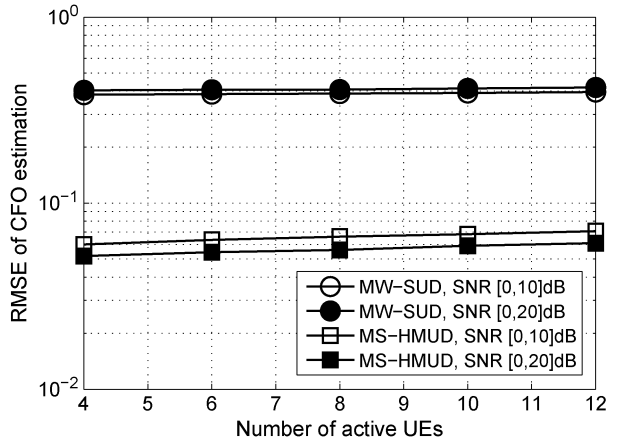

Fig. 6. Comparison of normalized CFO estimation with $j_{\max} = 1$ and the number of active UEs ranging from 4 to 12.

precise searching step being 0.05. Therefore, the number of possible CFOs is $N_{\widetilde{\epsilon}} = 8$.

B. Performance Evaluation

Case 1: Selection of Maximum Number of Iterations for the Second Step of the MS-HMUD: Figs. 3 and 4 show the miss-detection performance and complexity of the proposed MS-HMUD algorithm with respect to different numbers of iterations. The number of active UEs is selected as 6 and 10 with the SNR in [0, 10] dB and [0, 20] dB, respectively, representing general RA scenarios.

It can be seen in Fig. 3 that with the number of iterations increasing from 1 to 2, there is a slight improvement in the detection performance. While no further improvement is brought with the number of iterations increasing from 2 to 4, indicating that the maximum number of iterations should be selected less than or equal to 2. On the other hand, it can be observed in Fig. 4 that the complexity increases significantly with the increase in the number of iterations, which indicates that the iterative processing at the second step leads to much computational burden. Further, it is worth noting that the complexity with the SNR in [0, 10] dB is almost the same as that in [0, 20] dB due to the reason that the complexity of the MS-HMUD actually comes from the grid-search operation on the CFO as well as the number of RSSs. Finally, by combining the simulation results in both Figs. 3 and 4, in what follows, the maximum number of iterations for the second step of the MS-HMUD is chosen as $j_{\max} = 1$, which is actually a trade-off between the detection performance and complexity.

Case 2: Comparison of Proposed MS-HMUD Algorithm With Conventional MW-SUD Algorithm: Figs. 5–9 show the comparison of the proposed MS-HMUD algorithm with the conventional MW-SUD algorithm in terms of miss-detection, CFO estimation, RTD estimation, power estimation, and complexity. The number of active UEs is set from 4 to 12, and the SNR of UEs is uniformly distributed in [0, 10] dB or [0, 20] dB.

Fig. 5 shows the comparison of miss-detection performance between both algorithms. As the SNR in [0, 20] dB represents a more severe near-far effect than that in [0, 10] dB, both algorithms with the SNR in [0, 20] dB perform worse detection performance than that in [0, 10] dB. With the increase in the number of active UEs, the performance of both algorithms degrades, indicating that interference among active UEs is the

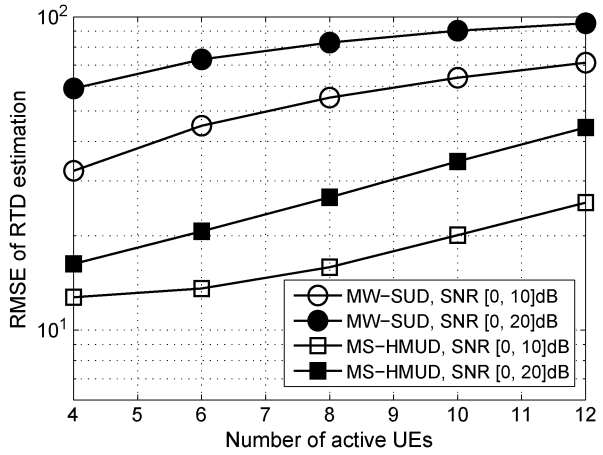

Fig. 7. Comparison of normalized RTD estimation with $j_{\max} = 1$ and the number of active UEs ranging from 4 to 12.

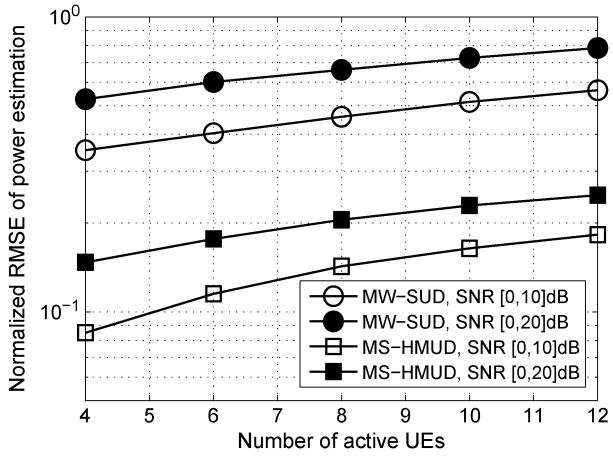

Fig. 8. Comparison of normalized power estimation with $j_{\max} = 1$ and the number of active UEs ranging from 4 to 12.

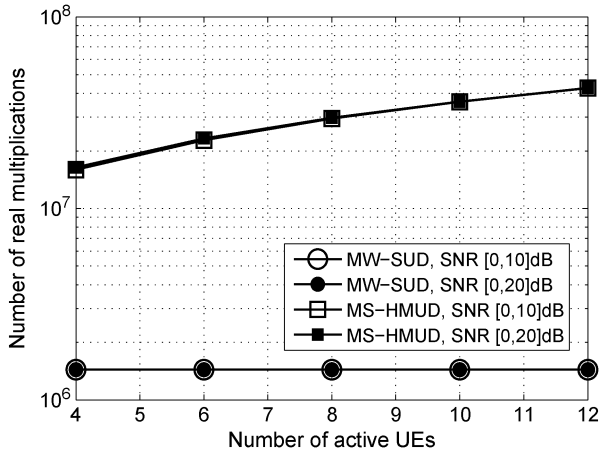

Fig. 9. Comparison of complexity with $j_{\max} = 1$ and the number of active UEs ranging from 4 to 12.

main reason that causes performance degradation. It can be seen that the performance of the proposed algorithm outperforms that of the MW-SUD algorithm with 1 to 2 orders of magnitude. This further indicates that interference among UEs would greatly inversely impact the performance of the MW-SUD algorithm, while the adoption of interference cancellation of the MS-HMUD is able to achieve better performance by suppressing interference among UEs.

Fig. 6 shows the comparison of the CFO estimation performance between both algorithms, and the CFO estimation is measured in terms of root mean square error (RMSE). As the MW-SUD could only coarsely estimate the CFO on the basis of magnitudes of both the main peak and false ones, the RMSE of the MW-SUD is almost larger than 0.3, indicating that the CFO cannot be accurately estimated by using the conventional MW-SUD algorithm. This is because the CFO estimation of the MW-SUD algorithm is only derived according to the CDMA structure [19] that is not in accordance with the SC-FDMA structure in LTE uplink. Once the SC-FDMA structure is adopted, both the inter-carrier interference in the FD and the CFO interference matrix $\mathbf{E}_v$ in the TrD would be introduced by the CFO, along with the TrD CIR, greatly impacting the TrD correlation processing and leading to inaccurate estimation results. On the other hand, for the MS-HMUD algorithm, as the TD CFO is much easier to be compensated than the TrD CFO, the CFO could therefore be estimated with enough precision. It can be seen that the absolute CFO estimation error of the MS-HMUD algorithm is less than 0.07, which guarantees the estimation results of TrD CIRs and further facilitates the interference cancellation processing for the detection of weak codes.

Figs. 7 and 8 show the comparison of the RTD and power estimation performance between both algorithms. It is obviously seen that the MS-HMUD algorithm also greatly outperforms the MW-SUD algorithm in two-fold. For one thing, as a false detection and estimation would lead to significant biases on the RTD estimation results, the MW-SUD algorithm cannot provide satisfied RTD estimation performance that is the very task of the RA procedure. For another, the normalized RMSE of power estimation of the MS-HMUD is less than 0.28, while that of the MW-SUD is always larger than 0.32. Combining with the simulation results in Fig. 6, it is obvious that the MS-HMUD algorithm is able to provide parameters estimation with more precision than the MW-SUD algorithm, facilitating the adoption of interference cancellation in suppressing interference among UEs.

Fig. 9 shows the comparison of complexity between both algorithms. Note that as the complexity of the MS-HMUD mainly depends on the numbers of possible CFOs and active UEs, as stated above, the MS-HMUD algorithm requires almost the same complexity with the SNR in both [0, 10] dB and [0, 20] dB. According the simulation results, the typical value of $i_{\max}$ is usually 2 or 3, meaning that more weak codes could be detected by spending more loops. It can be seen that the complexity of the MS-HMUD is approximate 12 to 26 times that of the MW-SUD algorithm. The extra complexity is paid off to improve the system efficiency and user experience.

VI. Conclusion

For the LTE random access procedure with the presence of frequency offsets, an explicit transform-domain random access

signal model is first provided, and the impact of transform-domain frequency offsets is theoretically analyzed. After that, a multiuser detection and estimation algorithm is proposed, which employs both the methodology of successive interference cancellation and the framework of iterative parameters estimation that is within the framework of the well-known SAGE algorithm. In addition, the Cramër-Rao bound to the accuracy of parameters estimation is also performed, which actually indicates the "best-case" estimation performance for the LTE random access procedure. Analyses and simulation results show that with the presence of a large carrier frequency offset, the proposed MS-HMUD algorithm is able to offer improved multiuser detection and estimation performance and user experience compared to the existing MW-SUD algorithm.

REFERENCES

[1] 3rd Generation Partnership Project (3GPP), "Evolved Unversal Terrestrial Radio Access (E-UTRA), User equipment radio transmission and reception," Sophia Antipolis, France, Tech. Spec. 36.211, v. 11.4.0, Sep. 2013.

[2] D. C. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 531–532, Jul. 1972.

[3] S. Sesia, I. Toufik, and M. Baker, *The UMTS Long Term Evolution–From Theory to Practice*. New York, NY, USA: Wiley, 2009.

[4] M. M. Mansour, "Optimized architecture for computing Zadoff-Chu sequences with application to LTE," in *Proc. IEEE GLOBECOM*, Nov. 2009, pp. 1–6.

[5] L. Yang, G. Ren, and B. Yang, "Fast time-varying channel estimation technique for LTE uplink in HST environment," *IEEE Trans. Veh. Technol.*, vol. 61, no. 9, pp. 4009–4019, Nov. 2012.

[6] A. Laya, L. Alonso, and J. Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 4–16, 1st Quart. 2013.

[7] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 86–93, Jun. 2013.

[8] B. Liang and Z. He, "The research on random access signal detection algorithm in LTE systems," in *Proc. IEEE Int. Symp. MAPE Wireless Commun.*, Oct. 2013, pp. 115–118.

[9] X. Yang and A. O. Fapojuwo, "Enhanced preamble detection for PRACH in LTE," in *Proc. IEEE WCNC*, Apr. 2013, pp. 3306–3311.

[10] Y. Chen, X. Wen, Z. Wei, and L. Xinqi, "Random access algorithm of LTE TDD system based on frequency domain detection," in *Proc. Int. Conf. Semantics, Knowl. Grid*, Oct. 2009, pp. 725–728.

[11] Y. Hu *et al.*, "A method of PRACH detection threshold setting in LTE TDD femtocell system," in *Proc. CHINACOM Conf. Netw.*, Aug. 2012, pp. 408–413.

[12] 3GPP TSG RAN, "Frequency offset effects on RACH preambles detectors," LG Electron., Seoul, Korea, WG1#47 R1-063161, Nov. 2006.

[13] 3GPP TSG RAN, "Using restricted preamble set for RACH in high mobility environments," Samsung, Seoul, Korea, WG1#49 R1-072234, Jun. 2007.

[14] M. Hua *et al.*, "Analysis of the frequency offset effect on random access signals," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4728–4740, Nov. 2013.

[15] M. Hua *et al.*, "Analysis of the frequency offset effect on Zadoff-Chu sequence timing performance," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4024–4039, Nov. 2014.

[16] 3GPP TSG RAN, "RACH in support of high speed UEs," Tallinn, Estonia, WG1#46 R1-062387, Sep. 2006.

[17] 3GPP TSG RAN, "Restricted sets of RACH preamble signatures for environments with high doppler shifts," WG1#49 R1-070377, Nokia, Espoo, Finland, Jan. 2007.

[18] 3GPP TSG RAN, "Specification of restricted set of cyclic shifts of root Zadoff-Chu sequences," WG1#49 R1-072898, Huawei, Shenzhen, China, Nov. 2009.

[19] A. Cao, P. Xiao, and R. Tafazolli, "Frequency offset estimation based on PRACH preambles in LTE," in *Proc. IEEE ISWCS*, Aug. 2014, pp. 22–26.

[20] J. Fessler and A. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.

[21] J. G. Andrews and T. H.-Y. Meng, "Performance of multicarrier CDMA with successive interference cancellation in a multipath fading channel," *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 811–822, May 2004.

[22] F. Lin and L. B. Milstein, "Successive interference cancellation in multicarrier DS/CDMA," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1530–1540, Sep. 2000.

[23] C. Shi and R. Narasimhan, "Soft interference cancellation receiver for SC-FDMA uplink in LTE," in *Proc. IEEE WCNC*, Apr. 2013, pp. 3318–3322.

[24] N. I. Niridakis and D. D. Vergados, "A survey on the successive interference cancellation performance for single-antenna and multiple-antenna OFDM systems," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 312–335, 1st Quart. 2013.

[25] N. Benvenuto, G. Carnevale, and S. Tomasin, "MC-CDMA with SIC: Power control by discrete stochastic approximation and comparison with OFDMA," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2006, pp. 5715–5720.

[26] A. Ruegg and A. Tarable, "Iterative SIC receiver scheme for non-orthogonally superimposed signals on top of OFDMA," in *Proc. IEEE PIMRC*, Dec. 2011, pp. 156–161.

[27] M. Pun, M. Morelli, and C.-C. J. Kuo, "Iterative detection and frequency synchronization for OFDMA uplink transmissions," in *IEEE Trans. Wireless Commun.*, vol. 6, no. 2 pp. 629–639, Feb. 2007.

[28] N. Kabaoglu, "SAGE based suboptimal receiver for downlink MC-CDMA system," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1381–1383, Dec. 2011.

[29] Y. Mostofi and D. C. Cox, "ICI mitigation for pilot-aided OFDM mobile systems," in *IEEE Trans. Wireless Commun.*, vol. 4, no. 2 pp. 765–774, Mar. 2005.

[30] 3rd Generation Partnership Project (3GPP), "Evolved universal terrestrial radio access (E-UTRA), medium access control (MAC) protocol specification," Sophia Antipolis, France, Tech. Spec. 36.321, v. 12.1.0, Mar. 2014.

[31] D. D. Falconer, "Linear precoding of OFDMA signals to minimize their instantaneous power variance," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 1154–1162, Apr. 2011.

Qiwei Wang was born in Henan, China, in 1987. He received the B.S. degree in communication engineering from Harbin University of Science and Technology, Harbin, China, in 2009, the M.S. degree in communication and information systems from Xidian University, Xi'an, China, in 2012, and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2015.

His current research interests include the wireless communication and digital signal processing, including the random access procedures in 4G communication systems, multi-user detection and parameters estimation, iterative receivers, etc.

Guangliang Ren (M'06) was born in Jiangsu, China, in 1971. He received the B.S. degree in communication engineering from Xidian University, Xi'an, China, in 1993, the M.S. degree in signal processing from the Academy of China Ordnance, Beijing, China, in 1996, and the Ph.D. degree in communication and information systems from Xidian University, in 2006.

He is currently a Professor with the School of Telecommunications Engineering, Xidian University. He is the author of more than 40 research papers in journals and conference proceedings, such as IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and an author or coauthor of three books. His research interests include wireless communication and digital signal processing, particularly multiple-input-multiple-output systems, WiMax, LTE, etc.

Jueying Wu was born in Guangxi, China, in 1991. She received the B.S. degree in communication engineering from Xidian University of Telecommunication Engineering, Xi'an, China, in 2013. She is currently working toward the M.S. degree in communication and information systems with Xidian University.

Her current research interests include uplink synchronization in OFDMA systems, including the random access procedure in LTE systems, multiuser detection and parameters estimation, coordinated multipoint, etc.