

Low-complexity Tomlinson-Harashima Precoding Update Algorithm for Massive MIMO System

Yiming Fang, *Graduate Student Member, IEEE*, Li Chen, *Senior Member, IEEE*, Yunfei Chen, *Senior Member, IEEE*, Huarui Yin, *Member, IEEE*, and Guo Wei

Abstract—Efficient implementation of Tomlinson-Harashima precoding (THP) is crucial in massive multiple-input-multiple-output (MIMO) systems with a large number of antennas at the base station (BS) serving many user equipments (UEs). To address the high computational complexity of THP, in this paper, we first propose novel THP update algorithms that can avoid recomputing the THP filters when a new UE arrives or departs. Specifically, by using the Gram-Schmidt process and a series of Givens matrices, the THP filters are computed without full matrix operations. Then we extend the THP update algorithms to a more general scenario when multiple multi-antenna UEs arrive or depart. In this case, the proposed algorithms use both direct and iterative approaches. Moreover, the computational complexity of the proposed algorithms is derived and compared with that of the conventional THP. Finally, to further align with the practical scenario, we analyze and derive the approximate close-form expressions for the sum achievable rate of the proposed algorithms under imperfect channel state information (CSI). Simulation results are provided to illustrate the effectiveness of the proposed algorithms. The impact of quasi-static fading and slow time-varying scenarios with imperfect CSI on the communication performance of the proposed algorithms is also evaluated.

Index Terms—Givens matrix, Low complexity, massive MIMO, QR decomposition, Tomlinson-Harashima precoding.

I. INTRODUCTION

MASSIVE multiple-input-multiple-output (MIMO) has emerged as a key technology for the fifth generation (5G) and beyond wireless communication systems [1]. By utilizing a large number of antennas at the base station (BS), massive MIMO can significantly improve the energy efficiency and the spectral efficiency over the conventional MIMO [2], [3]. However, as the dimension of massive MIMO increases, the computational complexity of the conventional MIMO algorithms becomes unbearable, posing challenges to practical implementation. Thus, one crucial aspect is the design of low-complexity precoding schemes. Specifically, non-linear precoding schemes, such as dirty paper coding (DPC) [4], vector perturbation (VP) [5] and lattice-aided methods [6], have been proposed and shown to have better performance

than conventional linear schemes. Among various precoding schemes, Tomlinson-Harashima precoding (THP) [7], [8] has attracted considerable attention due to its near-capacity performance compared with linear precoding yet low complexity compared with DPC and VP [9], [10]. Meanwhile, linear precoding schemes have lower complexity and achieve good performance when the number of user equipments (UEs) is much smaller than the number of BS antennas. However, where the number of UEs is close to the number of BS antennas, linear precoding methods show poor performance and are outperformed by the non-linear precoding methods, i.e., THP. Unfortunately, the complexity of the conventional THP algorithm may still be too high as the number of antennas keeps increasing for massive MIMO systems.

To address the complexity challenge of THP for massive MIMO systems, researchers have proposed several low-complexity algorithms. Most of them focused on three aspects: hierarchical hybrid design, successive interference cancellation (SIC), and hardware architecture. For the hierarchical hybrid design, the authors in [11], [12] proposed a robust minimum maximum mean square error (MMSE) hybrid linear and THP precoder, which combines the low complexity of linear precoding with the high performance of THP. In this scheme, the UEs are divided into different groups, where linear precoding is used to mitigate the interference between groups, while THP is used to mitigate the interference between users within the same group. In [13], the channel model was extended to Line-Of-Sight (LOS) environments from [11], [12] and a novel low-complexity hybrid linear and THP precoder with max-min power control was proposed. A hybrid VP and THP precoder was proposed in [14]. The result indicated that the proposed algorithm has lower computational complexity than other nonlinear precoding schemes.

Other works focused on the optimal order of SIC in the THP structure. For instance, in [15], a low-complexity ordering algorithm based on a sorted QR decomposition was proposed. The multi-branch (MB) THP algorithm was proposed in [16], where each branch contains a THP with a predetermined ordering strategy, and a selection criterion is then applied to choose the branch that generates the best final output. Compared to the classic MMSE THP in [17], the complexity of the MB THP algorithm is lower.

Moreover, the low-complexity THP algorithm in the hybrid analog-digital (A/D) architectures due to high power consumption was studied. In [18], a low-complexity TH hybrid precoding was proposed in the full-connected structure, which utilizes an orthogonal matching pursuit algorithm to decompose the

This research was supported by the National Key Research and Development Program of China (Grant No. 2021YFB2900302), Natural Science Foundation of China (Grant No. 62071445), and Anhui Provincial Natural Science Foundation (No. 230808532). (*Corresponding author: Li Chen.*)

Y. Fang, L. Chen, H. Yin, and G. Wei are with the CAS Key Laboratory of Wireless-Optical Communications, University of Science and Technology of China, Hefei 230027, China (e-mail: fym1219@mail.ustc.edu.cn; {chenli87, yhr, wei}@ustc.edu.cn).

Y. Chen is with the Department of Engineering, University of Durham, Durham, UK, DH1 3LE (e-mail: Yunfei.Chen@durham.ac.uk).

fully digital TH precoding matrix into the product of the RF precoding matrix and the digital precoding matrix. Considering the sub-connected structure, the authors in [19] established a recurrence relation of the matrix inversion to reduce the complexity of hybrid THP. Additionally, the authors in [20] proposed a novel approach to improving the performance of hybrid A/D transceivers for mmWave systems using THP. By minimizing the MMSE of the system under channel uncertainties with realistic transmit power and unit modulus constraints, the proposed design achieved lower complexity and power cost. In [21], a novel algorithm was proposed to jointly optimize the THP and the hybrid transmit precoder of the BS with the linear digital receivers of mobile UEs.

The aforementioned low-complexity THP algorithms have tackled the complexity issue of THP from the system level, but they do not take into account the computation complexity of basic matrix operations such as matrix multiplication and matrix decomposition. Note that, as the numbers of antennas and UEs increase, the complexity of these basic matrix operations cannot be ignored and can even become dominant. As an example, in the field of machine learning, memristive neural network circuits are considered one of the potential paths to the future of artificial intelligence due to their ability to perform basic matrix operations with significantly lower energy consumption and area compared to conventional complementary metal oxide semiconductor (CMOS) circuits [22]–[24]. In [25], a memristor-based synaptic circuit was used for online gradient descent training. The circuit demonstrated significant improvements in power and area efficiency, consuming only between 2% and 8% of the static power and area of conventional CMOS-only hardware. Thus, it is important to consider basic matrix operations. Furthermore, mixed precision algorithms for basic matrix operations have already attracted a lot of attention in the field of numerical linear algebra, which can achieve a balance between the fast computation and low energy consumption of low precision arithmetic, and the accuracy of high precision arithmetic simultaneously. For the matrix multiplication, the authors in [26] proposed a mixed precision summation algorithm, which calculates the partial sums with a low precision and combines them with a high precision. Similarly, as for the matrix decomposition, a mixed precision Householder QR algorithm and round error analysis were developed in [27]. In [28], a mixed precision algorithm for eigenvalue decomposition was proposed. The algorithm involves transforming the input matrix to tridiagonal form in single precision, followed by the computation of eigenpairs in double precision.

Additionally, some linear precoding update algorithms have been proposed to reduce the complexity of basic matrix operations. One such example is the low-complexity zero-forcing (ZF) precoding proposed in [29]. This algorithm used the matrix inversion lemma to update the inverse of the Gram matrix when single UE is added or removed from the system. Another approach in [30] used two low-complexity algorithms to update the URV decomposition of channel matrices upon the arrival or departure of single UE from the cell. For the slow time-varying scenario, the authors in [31] reduced the complexity of matrix inversion by exploiting the characteristics

of channel correlation in the time domain. Other works have also studied the update of matrix decomposition in the mathematical field [32]–[34]. However, to the best of our knowledge, the reduction of complexity in basic matrix operations for non-linear precoding schemes has never been studied before.

Motivated by the above observations, in this paper, we propose low-complexity THP update algorithms for the multi-user massive MIMO downlink system. First, by using part of the computation results in the previous time slot, low-complexity THP update algorithms are proposed when a single-antenna UE arrives in or departs from the cell. Then we extend our algorithms to consider the arrival or departure of multiple multi-antenna UEs. Additionally, the impact of imperfect channel state information (CSI) is considered and closed-form approximate expressions for the achievable rate of the proposed algorithm are derived. Finally, computational complexity analysis is derived, and the impact of the quasi-static fading scenario and slow time-varying scenario on the proposed algorithms is examined through simulation. The main contributions of this paper are summarized as follows.

- **THP update algorithms for single-antenna UE.** Different from existing low-complexity THP algorithms, our focus is on the basic matrix operations in the THP structure, particularly the process of obtaining THP filters. Furthermore, we propose novel THP update algorithms for a multi-user massive MIMO downlink system that do not need to recompute THP filters every time. In particular, the Gram-Schmidt process and a series of Givens rotation matrices are utilized to update the THP filters when a UE arrives or departs.
- **THP update algorithms for multiple multi-antenna UEs.** Compared with the algorithms in [29]–[31], our proposed algorithms can be extended to a scenario where multiple multi-antenna UEs are present. This scenario is more complex than the scenario of single-antenna UE, making the update algorithm for the latter inapplicable. To address this problem, we present direct and iterative update algorithms when multiple multi-antenna UEs arrive or depart, respectively.
- **Sum Achievable rate of the proposed algorithm with imperfect CSI.** To further align with real-world feasibility, we consider the impact of imperfect CSI and derive closed-form approximate expressions for the achievable rate of the proposed algorithm in the Rayleigh channel over quasi-static fading and time-varying scenarios with imperfect CSI. These results reveal the effects of the number of antennas at BS, channel estimation errors, and time-variations on the achievable rate performance.
- **Computational complexity analysis and communication performance comparison.** We conduct a computational complexity analysis for each proposed algorithm. Additionally, from the communication performance perspective, we not only study the bit error rate (BER) performance compared to conventional THP in the quasi-static fading scenario but evaluate the BER performance compared to conventional THP in the slow time-varying scenario which shows that our algorithms have lower

complexity and only a negligible loss in performance. Moreover, we investigate the impact of imperfect CSI on the proposed algorithms.

The remainder of this paper is organized as follows. Section II introduces a multi-user massive MIMO downlink system model and formulates the key problem. In Section III, we propose a novel THP update algorithm for the scenario of single-antenna UE. In Section IV, the proposed THP update algorithm is extended to a general scenario. Section V derives closed-form approximate expressions for the proposed algorithms under quasi-static fading and time-varying scenarios with imperfect CSI. Simulation results are presented in Section VI. At last, the conclusions are provided in Section VII.

Notation: Bold uppercase letters denote matrices and bold lowercase letters denote vectors. For a matrix \mathbf{A} , \mathbf{A}^T , \mathbf{A}^H and \mathbf{A}^{-1} denote the transpose, the Hermitian transpose and inverse of \mathbf{A} , respectively. $a_{i,j}$ denotes (i,j) th entry of \mathbf{A} . $\mathbf{A}(k,:)$ and $\mathbf{A}(:,k)$ denote the k th row and the k th column of matrix \mathbf{A} . $\text{diag}(\mathbf{A})$ denotes the matrix of the diagonal elements of matrix \mathbf{A} . $\mathbb{E}\{\mathbf{A}\}$ denotes the expectation of \mathbf{A} . $\mathbb{C}^{M \times N}$ denotes the space of $M \times N$ complex matrices. For a vector \mathbf{a} , $\|\mathbf{a}\|_2$ denotes its Euclidean norm.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

As shown in Fig. 1, we consider the downlink of a multi-user massive MIMO system. The BS is equipped with M transmit antennas and communicates N data streams to N UEs with single antenna. The BS is assumed to apply a precoding matrix \mathbf{W} . Hence, the transmit signal from BS is given by $\mathbf{x} = \mathbf{W}\mathbf{s}$, where $\mathbf{s} = [s_1^T, s_2^T, \dots, s_N^T]^T \in \mathbb{C}^{N \times 1}$ is the transmit data vector. And \mathbf{s} is assumed to be i.i.d Gaussian distributed with $\mathbf{s} \sim \mathcal{CN}(0, \sigma_s^2 \mathbf{I}_N)$.

The received signals of the UEs are expressed as [35]

$$\mathbf{y} = \beta(\mathbf{H}^H \beta^{-1} \mathbf{x} + \mathbf{n}), \quad (1)$$

where $\mathbf{y} \in \mathbb{C}^{N \times 1}$, $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N] \in \mathbb{C}^{M \times N}$ is the downlink channel matrix, where $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$ is the channel between the BS and the k th UE $_k$. Moreover, β is a power normalization factor and $\mathbf{n} \in \mathbb{C}^{N \times 1}$ is the additive white Gaussian noise (AWGN) with $\mathbf{n} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_N)$.

Nonlinear precoding, such as THP, generally has better performance than linear precoding. Therefore, we choose to use the centralized THP (cTHP) structure [36], which incorporates the filter \mathbf{G} at the BS and enables joint precoding and updating, resulting in reduced complexity for the UE, and we have

$$\mathbf{x} = \mathbf{F}\mathbf{G}\mathbf{B}^{-1}(\mathbf{s} + \mathbf{d}), \quad (2)$$

where \mathbf{d} is the residual error and the three filters \mathbf{F} , \mathbf{G} , and \mathbf{B} can be obtained as [37]

$$\mathbf{F} = \mathbf{Q}, \quad (3)$$

$$\mathbf{G} = \begin{bmatrix} r_{1,1}^{-1} & & & \\ & r_{2,2}^{-1} & & \\ & & \ddots & \\ & & & r_{N,N}^{-1} \end{bmatrix}, \quad (4)$$

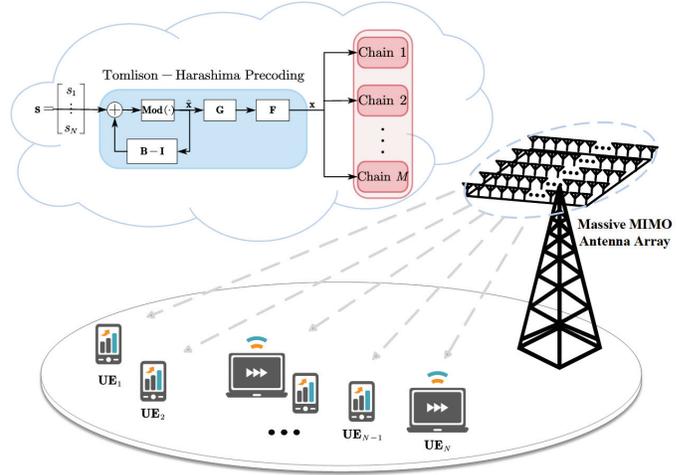


Fig. 1. Multi-user massive MIMO downlink THP system.

$$\mathbf{B} = \mathbf{R}^H \mathbf{G}, \quad (5)$$

where $r_{1,1}, r_{2,2}, \dots, r_{N_r, N_r}$ are diagonal entries of \mathbf{R} , \mathbf{Q} and \mathbf{R} is the QR decomposition to \mathbf{H} , i.e.

$$\mathbf{H} = \mathbf{Q}\mathbf{R}. \quad (6)$$

Additionally, we have $\beta = \sqrt{\sum_{i=1}^{N_r} r_{ii}^{-2} / M}$ in the cTHP structure [35].

In practice, channel estimation errors are inevitable. According to [20], [38], the channel estimation errors can be modelled as

$$\hat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}, \quad (7)$$

where $\hat{\mathbf{H}} \in \mathbb{C}^{M \times N}$ is the estimated channel matrix and $\Delta\mathbf{H} = [\Delta\mathbf{h}_1, \Delta\mathbf{h}_2, \dots, \Delta\mathbf{h}_N]$ is additive error matrix, and the entries in $\Delta\mathbf{H}$ are assumed to be i.i.d. $\mathcal{CN}(0, \sigma_e^2)$. Therefore, for the scenario of imperfect CSI, the THP filters are calculated based on $\hat{\mathbf{H}}$ but not on \mathbf{H} .

B. Problem Formulation

Considering the process to obtain the three filters \mathbf{F} , \mathbf{G} , and \mathbf{B} , we can determine that the computational complexity of cTHP is $\mathcal{O}(MN^2)$. While THP offers superior performance compared to linear precoding, its computational complexity remains a challenge, especially in scenarios where the BS has a large number of antennas or there are a large amount of UEs.

To reduce the computational complexity of THP, we propose novel THP update algorithms. The key idea of the algorithm is to leverage the correlation between the three filters \mathbf{F} , \mathbf{G} , and \mathbf{B} from the previous time slot to optimize the calculation process. Specifically, we reuse part of the computation results from \mathbf{F} , \mathbf{G} , and \mathbf{B} in the previous time slot to update them in the current time slot. By doing so, the algorithm can improve efficiency and reduce redundant calculations.

As shown in Fig. 2, there is a substantial temporal correlation between consecutive channels which are denoted as $\mathbf{H}_t, t = 1, 2, \dots$. This temporal correlation is due to the quasi-static fading property of the channel, where the change from

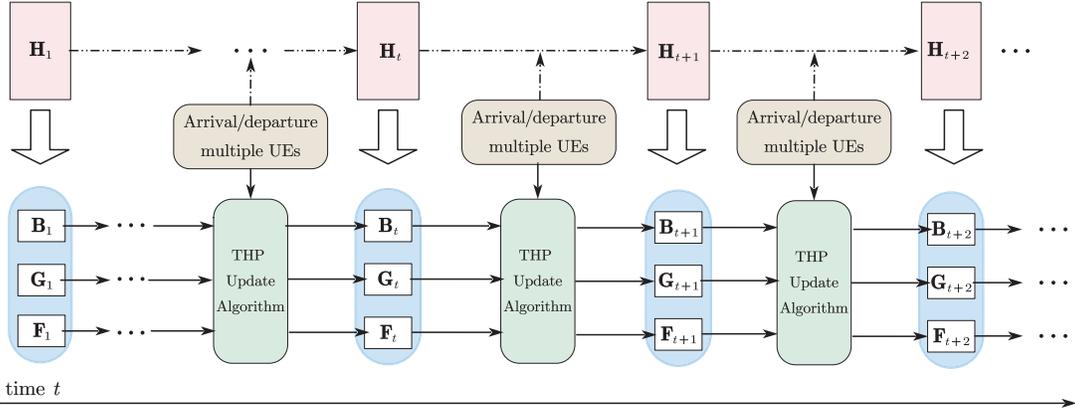


Fig. 2. Formulation of proposed update algorithm.

\mathbf{H}_t to \mathbf{H}_{t+1} is only caused by the arrival or departure of UEs. In other words, the arrival or departure of UEs from the cell is equivalent to the addition or deletion of columns from \mathbf{H}_t to \mathbf{H}_{t+1} . Therefore, our challenge is to find a way to utilize the three filters \mathbf{F}_t , \mathbf{G}_t , and \mathbf{B}_t at time t to generate the new filters \mathbf{F}_{t+1} , \mathbf{G}_{t+1} , and \mathbf{B}_{t+1} at time $t+1$ when UEs arrive in or depart from the cell at time $t+1$.

III. NEW THP UPDATE ALGORITHM FOR THE SINGLE-ANTENNA UE

In this section, we will propose a new THP update algorithm that accounts for the arrival or departure of a UE at time $t+1$. First, we discuss the scenario where a UE arrives at the cell, and then the scenario where a UE departs from the cell is discussed.

A. Single UE Arrival

Proposition 1 (Algorithm for a UE Arrival). When a new UE \mathbf{h}_p arrives in the cell at time $t+1$, the three filters \mathbf{F}_{t+1} , \mathbf{G}_{t+1} , \mathbf{B}_{t+1} and upper triangular matrix \mathbf{R}_{t+1} at the time $t+1$ can be expressed using \mathbf{F}_t , \mathbf{G}_t , \mathbf{B}_t and \mathbf{R}_t at time t as

$$\mathbf{F}_{t+1} = \bar{\mathbf{Q}}\mathbf{G}_v^H, \quad (8)$$

$$\mathbf{R}_{t+1} = \mathbf{G}_v\bar{\mathbf{R}}\mathbf{P}^H, \quad (9)$$

$$\mathbf{G}_{t+1} = \begin{bmatrix} \mathbf{G}_t(:, 1:p-1) & \\ & \text{diag}(\mathbf{R}_{t+1}^{-1}(:, p:N+1)) \end{bmatrix}, \quad (10)$$

$$\mathbf{B}_{t+1} = \begin{bmatrix} \mathbf{B}_t(1:p-1,:) & \mathbf{0} \\ \mathbf{R}_{t+1}^H(p:N+1,:) & \mathbf{G}_{t+1} \end{bmatrix}, \quad (11)$$

where p is the position of the new UE in the channel matrix, $\bar{\mathbf{Q}} = [\mathbf{F}_t, \mathbf{q}]$, $\bar{\mathbf{R}} = \begin{bmatrix} \mathbf{R}_t & \mathbf{r} \\ 0 & \alpha \end{bmatrix}$, $\mathbf{r} = \mathbf{Q}_t^H \mathbf{h}_p$, $\alpha = \|(\mathbf{I} - \mathbf{Q}_t \mathbf{Q}_t^H) \mathbf{h}_p\|_2$, $\mathbf{q} = \alpha^{-1}(\mathbf{I} - \mathbf{Q}_t \mathbf{Q}_t^H) \mathbf{h}_p$, $\mathbf{G}_v \in \mathbb{C}^{(N+1) \times (N+1)}$ is a unitary matrix and $\mathbf{P} \in \mathbb{C}^{(N+1) \times (N+1)}$ is a permutation matrix.

Proof: When a UE arrives in the cell, it is equivalent to adding a column in the channel matrix \mathbf{H}_t . Therefore, we have $\mathbf{H}_{t+1} = [\mathbf{H}_t(:, 1:p-1), \mathbf{h}_p, \mathbf{H}_t(:, p:N)]$, where $\mathbf{h}_p \in \mathbb{C}^{M \times 1}$ is the channel matrix of the new UE.

To simplify the update process, we move \mathbf{h}_p to the last column of \mathbf{H}_{t+1} . We can then define \mathbf{H}_a as the product of \mathbf{H}_{t+1} with a series of permutation matrices $\mathbf{P}_{p,p+1}, \mathbf{P}_{p+1,p+2}, \dots, \mathbf{P}_{N,N+1}$, denoted as \mathbf{P} . The operation of right multiplying $\mathbf{P}_{i,j}$ is equivalent to permuting the i th and j th columns of the channel matrix. Therefore, \mathbf{H}_a can be expressed as $\mathbf{H}_a = [\mathbf{H}_t, \mathbf{h}_p]$. Given the QR decomposition of \mathbf{H}_t , we can use the Gram-Schmidt process [33] to generate the QR decomposition of \mathbf{H}_a as follows:

$$\begin{aligned} \mathbf{H}_a &= [\mathbf{Q}_t \quad \mathbf{h}_p] \begin{bmatrix} \mathbf{R}_t & 0 \\ 0 & 1 \end{bmatrix} \\ &= [\mathbf{Q}_t \quad \mathbf{q}] \begin{bmatrix} \mathbf{I} & \mathbf{r} \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} \mathbf{R}_t & 0 \\ 0 & 1 \end{bmatrix} \\ &= \underbrace{[\mathbf{Q}_t \quad \mathbf{q}]}_{\bar{\mathbf{Q}}} \underbrace{\begin{bmatrix} \mathbf{R}_t & \mathbf{r} \\ 0 & \alpha \end{bmatrix}}_{\bar{\mathbf{R}}}, \end{aligned} \quad (12)$$

where $\mathbf{r} = \mathbf{Q}_t^H \mathbf{h}_p$, $\alpha = \|(\mathbf{I} - \mathbf{Q}_t \mathbf{Q}_t^H) \mathbf{h}_p\|_2$, $\mathbf{q} = \alpha^{-1}(\mathbf{I} - \mathbf{Q}_t \mathbf{Q}_t^H) \mathbf{h}_p$. Furthermore, \mathbf{H}_{t+1} can be denoted as

$$\mathbf{H}_{t+1} = \mathbf{H}_a \mathbf{P}^H = \bar{\mathbf{Q}} \bar{\mathbf{R}} \mathbf{P}^H = \bar{\mathbf{Q}} \bar{\mathbf{R}} \mathbf{P}_{N+1,N} \dots \mathbf{P}_{p+1,p}. \quad (13)$$

It is important to note that $\tilde{\mathbf{R}} = \bar{\mathbf{R}} \mathbf{P}^H$ is not guaranteed to be an upper triangular matrix if $p \neq N+1$. To illustrate this, an example of $\bar{\mathbf{R}}$ is provided in Case 1 of Appendix A. In such a case, a common approach to reconstructing an upper triangular matrix is to use the Givens rotation matrix. We give the definition and property of a complex Givens matrix $\tilde{\mathbf{G}}_i(a, b)$ in Appendix B.

Before using the Givens matrix, we denote $\hat{\mathbf{R}} = \bar{\mathbf{R}} \mathbf{P}_{i+1,i}^H$, which exchanges the i th and the $(i+1)$ th columns of the upper triangular matrix $\bar{\mathbf{R}}$. Thus, we have $\hat{r}_{x,y} = 0, \forall 1 \leq x < y \leq N+1$, except for $(x, y) = (i+1, i)$.

Then, using the property of Givens matrix, the entry $\hat{r}_{x,y}$ can be nullified by left multiplying $\tilde{\mathbf{G}}_i = \tilde{\mathbf{G}}_i(\hat{r}_{i,i}, \hat{r}_{i+1,i})$. Hence, $\tilde{\mathbf{G}}_i \hat{\mathbf{R}} \mathbf{P}_{i+1,i}^H$ is an upper triangular matrix. Denote $\mathbf{G}_v = \tilde{\mathbf{G}}_p \tilde{\mathbf{G}}_{p+1} \dots \tilde{\mathbf{G}}_N$ and one has

$$\mathbf{H}_{t+1} = \mathbf{H}_a \mathbf{P}^H = \bar{\mathbf{Q}} \bar{\mathbf{R}} \mathbf{P}^H = \underbrace{\bar{\mathbf{Q}} \mathbf{G}_v^H}_{\mathbf{Q}_{t+1}} \underbrace{\mathbf{G}_v \bar{\mathbf{R}} \mathbf{P}^H}_{\mathbf{R}_{t+1}}. \quad (14)$$

Algorithm 1: Update THP when a UE \mathbf{h}_p arrives in \mathbf{H}_t at position p

Input: $\mathbf{H}_t = \mathbf{Q}_t \mathbf{R}_t$, \mathbf{h}_p , \mathbf{F}_t , \mathbf{G}_t , \mathbf{B}_t
Output: \mathbf{F}_{t+1} , \mathbf{G}_{t+1} , \mathbf{B}_{t+1} , \mathbf{R}_{t+1}

- 1 $\mathbf{r} \leftarrow \mathbf{F}_t^H \mathbf{h}_p$, $\mathbf{t} \leftarrow \mathbf{h}_p - \mathbf{F}_t \mathbf{r}$, $\alpha \leftarrow \|\mathbf{t}\|_2$
- 2 $\bar{\mathbf{Q}} = \begin{bmatrix} \mathbf{F}_t & \mathbf{t} \\ & \alpha \end{bmatrix}$, $\bar{\mathbf{R}} = \begin{bmatrix} \mathbf{R}_t & \mathbf{r} \\ 0 & \alpha \end{bmatrix}$
- 3 **for** $i = N : -1 : p$ **do**
- 4 $\bar{\mathbf{R}} \leftarrow \bar{\mathbf{R}} \mathbf{P}_{i,i+1}$
- 5 $\bar{\mathbf{G}}_i \leftarrow \mathbf{G}_i(\bar{r}_{i,i}, \bar{r}_{i+1,i})$
- 6 $\bar{\mathbf{R}} \leftarrow \bar{\mathbf{G}}_i \bar{\mathbf{R}}$
- 7 $\bar{\mathbf{Q}} \leftarrow \bar{\mathbf{Q}} \bar{\mathbf{G}}_i^H$
- 8 **end**
- 9 $\mathbf{R}_{t+1} = \bar{\mathbf{R}}$, $\mathbf{F}_{t+1} = \bar{\mathbf{Q}}$
- 10 $\mathbf{G}_{t+1} = \begin{bmatrix} \mathbf{G}_t(:, 1 : p-1) & & & \\ & \bar{r}_{p,p}^{-1} & & \\ & & \ddots & \\ & & & \bar{r}_{N+1,N+1}^{-1} \end{bmatrix}$
- 11 $\mathbf{B}_{t+1} = \begin{bmatrix} \mathbf{B}_t(1 : p-1, :) & \mathbf{0} \\ \mathbf{R}_{t+1}^H(p : N+1, :) \mathbf{G}_{t+1} \end{bmatrix}$

Since \mathbf{G}_v is a unitary matrix, the columns of \mathbf{Q}_{t+1} is orthogonal. Hence, we get the QR decomposition of \mathbf{H}_{t+1} when adding a UE. Then we can easily obtain \mathbf{F}_{t+1} , \mathbf{G}_{t+1} and \mathbf{B}_{t+1} in (8), (10) and (11) from (3)-(5). Moreover, if $p = N + 1$, we can further reduce the complexity of the algorithm by noting that $\bar{\mathbf{R}}$ is an upper triangular matrix and the Givens matrices are unnecessary. We also give an example in Case 2 of Appendix A. ■

The overall procedure of the proposed algorithm is summarized in Algorithm 1. Next, we determine the computation complexity of Algorithm 1. Steps 1 ~ 2 require $\mathcal{O}(MN)$ operations to generate the matrices $\bar{\mathbf{Q}}$ and $\bar{\mathbf{R}}$. Note that multiplying a matrix with a Givens rotation matrix only requires two linear transformations of two rows or columns [39]. Similarly, right-multiplying a matrix with a permutation matrix is equivalent to exchanging two columns of the matrix, which requires $\mathcal{O}(1)$ operations. Thus, we require $\mathcal{O}((N-p+1)(M+N))$ operations from Steps 3 ~ 8. At last, Steps 10 ~ 11 require $\mathcal{O}((N-p+2)(N+p+1)/2)$ operations to generate the three filters. Therefore, the complexity of Algorithm 1 is $\mathcal{O}(MN + (N-p+1)(M+N) + (N-p+2)(N+p+1)/2)$. If $p = 1$, i.e. the new UE is at the beginning position of the channel matrix with the worst-case complexity of around $\mathcal{O}(MN + N^2)$, and if $p = N + 1$, i.e. the new UE is at the last position of the channel matrix with the best-case complexity of around $\mathcal{O}(MN)$. However, the complexity of the conventional algorithm requires $\mathcal{O}(M(N+1)^2)$, much more than that of Algorithm 1.

B. Single UE Departure

Proposition 2 (Algorithm for a UE Departure). When a UE departs from the cell at time $t + 1$, the three filters \mathbf{F}_{t+1} ,

Algorithm 2: Update THP when a UE departs from \mathbf{H}_t at position p

Input: $\mathbf{H}_t = \mathbf{Q}_t \mathbf{R}_t$, p , \mathbf{F}_t , \mathbf{G}_t , \mathbf{B}_t
Output: \mathbf{F}_{t+1} , \mathbf{G}_{t+1} , \mathbf{B}_{t+1} , \mathbf{R}_{t+1}

- 1 $\bar{\mathbf{R}} \leftarrow \mathbf{R}_t$, $\bar{\mathbf{Q}} \leftarrow \mathbf{F}_t$
- 2 **for** $i = p : N - 1$ **do**
- 3 $\bar{\mathbf{R}} \leftarrow \bar{\mathbf{R}} \mathbf{P}_{i,i+1}$
- 4 $\bar{\mathbf{G}}_i \leftarrow \mathbf{G}_i(\bar{r}_{i,i}, \bar{r}_{i+1,i})$
- 5 $\bar{\mathbf{R}} \leftarrow \bar{\mathbf{G}}_i \bar{\mathbf{R}}$
- 6 $\bar{\mathbf{Q}} \leftarrow \bar{\mathbf{Q}} \bar{\mathbf{G}}_i^H$
- 7 **end**
- 8 $\mathbf{R}_{t+1} = \bar{\mathbf{R}}(1 : N - 1, 1 : N - 1)$
- 9 $\mathbf{F}_{t+1} = \bar{\mathbf{Q}}(:, 1 : N - 1)$
- 10 $\mathbf{G}_{t+1} = \begin{bmatrix} \mathbf{G}_t(:, 1 : p-1) & & & \\ & \bar{r}_{p,p}^{-1} & & \\ & & \ddots & \\ & & & \bar{r}_{N-1,N-1}^{-1} \end{bmatrix}$
- 11 $\mathbf{B}_{t+1} = \begin{bmatrix} \mathbf{B}_t(1 : p-1, 1 : N-1) \\ \mathbf{R}_{t+1}^H(p : N-1, :) \mathbf{G}_{t+1} \end{bmatrix}$

\mathbf{G}_{t+1} , \mathbf{B}_{t+1} and upper triangular matrix \mathbf{R}_{t+1} at time $t + 1$ can be expressed using \mathbf{F}_t , \mathbf{G}_t , \mathbf{B}_t and \mathbf{R}_t at time t as

$$\mathbf{F}_{t+1} = (\mathbf{F}_t \mathbf{G}_v^H)(:, 1 : N - 1), \quad (15)$$

$$\mathbf{R}_{t+1} = (\mathbf{G}_v \mathbf{R}_t \mathbf{P})(1 : N - 1, 1 : N - 1), \quad (16)$$

$$\mathbf{G}_{t+1} = \begin{bmatrix} \mathbf{G}_t(:, 1 : p-1) & \\ & \text{diag}(\mathbf{R}_{t+1}^{-1}(:, p : N-1)) \end{bmatrix}, \quad (17)$$

$$\mathbf{B}_{t+1} = \begin{bmatrix} \mathbf{B}_t(1 : p-1, 1 : N-1) \\ \mathbf{R}_{t+1}^H(p : N-1, :) \mathbf{G}_{t+1} \end{bmatrix}, \quad (18)$$

where p is the position of the departure of UE in the channel matrix, $\mathbf{G}_v \in \mathbb{C}^{N \times N}$ is a unitary matrix and $\mathbf{P} \in \mathbb{C}^{N \times N}$ is a permutation matrix.

Proof: When a UE departs from the cell, it is equivalent to deleting a column in the channel matrix \mathbf{H}_t . Therefore, we have $\mathbf{H}_{t+1} = [\mathbf{H}_t(:, 1 : p-1), \mathbf{H}_t(:, p+1 : N)]$.

Similar to the approach in the proof of Proposition 1, we move the channel matrix of UE $_p$ \mathbf{h}_p to the last column of \mathbf{H}_t , which needs to right multiply a series of permutation matrices $\mathbf{P}_{p,p+1}$, $\mathbf{P}_{p+1,p+2}$, \dots , $\mathbf{P}_{N-1,N}$, denoted as \mathbf{P} . Then we have

$$[\mathbf{H}_{t+1} \quad \mathbf{h}_p] = \mathbf{H}_t \mathbf{P}_{p,p+1} \cdots \mathbf{P}_{N-1,N} = \mathbf{Q}_t \mathbf{R}_t \mathbf{P}. \quad (19)$$

Note that $\tilde{\mathbf{R}} = \mathbf{R}_t \mathbf{P}^H$ is not guaranteed to be an upper triangular matrix if $p \neq N$. An example is shown in Case 3 of Appendix C. However, we can use the Givens matrix to reconstruct an upper triangular matrix from $\tilde{\mathbf{R}}$. Denote $\mathbf{G}_v = \tilde{\mathbf{G}}_{N-1} \tilde{\mathbf{G}}_{N-2} \cdots \tilde{\mathbf{G}}_p$. Using this, we can rewrite (19) as:

$$[\mathbf{H}_{t+1} \quad \mathbf{h}_p] = \underbrace{\mathbf{Q}_t \mathbf{G}_v^H}_{\bar{\mathbf{Q}}} \underbrace{\mathbf{G}_v \mathbf{R}_t \mathbf{P}}_{\bar{\mathbf{R}}}. \quad (20)$$

Additionally, according to the Gram-Schmidt process, we have

$$[\mathbf{H}_{t+1} \quad \mathbf{h}_p] = [\mathbf{Q}_{t+1} \quad \mathbf{q}] \begin{bmatrix} \mathbf{R}_{t+1} & \mathbf{Q}_{t+1}^H \mathbf{h}_p \\ 0 & \alpha \end{bmatrix}. \quad (21)$$

Algorithm 4: Iterative Update THP when k UEs arrive in \mathbf{H}_t before the p th UE

Input: $\mathbf{H}_t = \mathbf{Q}_t \mathbf{R}_t, \mathbf{F}_t, \mathbf{G}_t, \mathbf{B}_t$
Output: $\mathbf{F}_{t+1}, \mathbf{G}_{t+1}, \mathbf{B}_{t+1}, \mathbf{R}_{t+1}$

- 1 $N_k = \sum_{j=1}^k N_{r,j}, N_p = \sum_{i=1}^{p-1} N_{r,i}$
- 2 **for** $i = 1 : N_k$ **do**
- 3 $\mathbf{h} = \mathbf{H}_p(:, i),$
- 4 go to the step 1 ~ 8 in Algorithm 1, which \mathbf{h} is added to the position $p,$
- 5 $p = p + 1$
- 6 **end**
- 7 $\mathbf{R}_{t+1} = \bar{\mathbf{R}}, \mathbf{F}_{t+1} = \bar{\mathbf{Q}}$
- 8 $\mathbf{G}_{t+1} =$

$$\begin{bmatrix} \mathbf{G}_t(:, 1 : N_p) & & & \\ & \bar{r}_{N_p+1, N_p+1}^{-1} & & \\ & & \ddots & \\ & & & \bar{r}_{N_r+N_k, N_r+N_k}^{-1} \end{bmatrix}$$
- 9 $\mathbf{B}_{t+1} = \begin{bmatrix} & \mathbf{B}_t(1 : N_p, :) & \mathbf{O} \\ \mathbf{R}_{t+1}^H(N_p + 1 : N_r + N_k, :) & \mathbf{G}_{t+1} \end{bmatrix}$

$\mathbf{R}_{t+1}(N_r + N_k, :)$ to obtain the economy-size QR decomposition. Then we can easily obtain $\mathbf{F}_{t+1}, \mathbf{G}_{t+1}$ and \mathbf{B}_{t+1} by (3)-(5). Moreover, if $p = N + 1$, we only need to use Givens rotations to multiply the matrix \mathbf{V} , and no operations are required on the matrix \mathbf{R}_2 . An example for $\bar{\mathbf{R}}$ also is given in Case 6 of Appendix D. ■

The overall procedure of the proposed algorithm is summarized in Algorithm 3. Then we calculate the computation complexity of Algorithm 3. Step 2 requires $\mathcal{O}(M^2 N_k)$ to generate the matrix \mathbf{V} . Since Steps 5 ~ 13 need $(M - N_p - j)\mathcal{O}(N_k + M) + (N_r + j)\mathcal{O}(N_r)$, the total computational complexity of Steps 4 ~ 14 is $\sum_{j=1}^{N_k} [(M - N_p - j)\mathcal{O}(N_k + M) + (N_r + j)\mathcal{O}(N_r)]$. At last, Steps 19 ~ 20 require $\mathcal{O}((N_r + N_k - N_p)(N_r + N_k + N_p + 1)/2)$. Therefore, the complexity of Algorithm 3 is $\mathcal{O}(N_k M^2 - N_k M N_p + N_k^2 (M + N_r - N_p) - N_k^3 (M - N_r)/2 - N_k^4 + (N_r + N_k - N_p)(N_r + N_k + N_p + 1))$. We have the worst-case complexity is around $\mathcal{O}(N_k M^2 + N_k^2 (M + N_r) - N_k^3 (M - N_r) - N_k^4)$ when $p = 1$, i.e. $N_p = 0$. However, the conventional algorithm requires $\mathcal{O}(M(N_r + N_k)^2)$.

Remark 1 (Low-complexity Iterative Update Algorithm). Note that, when $M = 64$, the complexity of Algorithm 3 is less than the complexity of the conventional algorithm. However, the complexity of Algorithm 3 will be higher than that of the conventional algorithm as the number of antennas M increases, e.g. $M = 128, 256, 512$. In this case, one possible solution is to iterate Algorithm 1 N_k times to reduce the complexity, which is summarized in Algorithm 4. Algorithm 4 only requires $\mathcal{O}(\sum_{i=1}^{N_k} [M(N_r + i - 1) + (N_r + i - 1 - N_p)(N_r + i - 1 + M)] + (N_r + N_k - N_p)(N_r + N_k + N_p + 1)/2) \approx \mathcal{O}(M N_k N_r + M N_k^2 + N_r^2 N_k) \ll \mathcal{O}(M(N_r + N_k)^2)$.

C. Multiple UEs Departure

Proposition 4 (Algorithm for Multiple UEs Departure). When k neighboring UE $_{p:p+k-1}$ depart from the cell at time

Algorithm 5: Update THP when k UEs depart from \mathbf{H}_t from the p th UE

Input: $\mathbf{H}_t = \mathbf{Q}_t \mathbf{R}_t, \mathbf{F}_t, \mathbf{G}_t, \mathbf{B}_t$
Output: $\mathbf{F}_{t+1}, \mathbf{G}_{t+1}, \mathbf{B}_{t+1}, \mathbf{R}_{t+1}$

- 1 $N_k = \sum_{i=p}^{p+k-1} N_{r,i}, N_p = \sum_{i=1}^{p-1} N_{r,i}$
- 2 $\bar{\mathbf{R}} = [\mathbf{R}_t(:, 1 : N_p) \quad \mathbf{R}_t(:, N_p + N_k + 1 : N_r)]$
- 3 $\bar{\mathbf{Q}} \leftarrow \mathbf{F}_t$
- 4 **for** $j = N_p + 1 : N_r - N_k$ **do**
- 5 **for** $i = j + N_k - 1 : -1 : j$ **do**
- 6 $\bar{\mathbf{G}}_i \leftarrow \mathbf{G}_i(\bar{r}_{i,j}, \bar{r}_{i+1,j})$
- 7 $\bar{\mathbf{R}} \leftarrow \bar{\mathbf{G}}_i \bar{\mathbf{R}}$
- 8 $\bar{\mathbf{Q}} \leftarrow \bar{\mathbf{Q}} \bar{\mathbf{G}}_i^H$
- 9 **end**
- 10 **end**
- 11 $\mathbf{R}_{t+1} = \bar{\mathbf{R}}(1 : N_r - N_k, 1 : N_r - N_k)$
- 12 $\mathbf{F}_{t+1} = \bar{\mathbf{Q}}(:, 1 : N_r - N_k)$
- 13 $\mathbf{G}_{t+1} =$

$$\begin{bmatrix} \mathbf{G}_t(:, 1 : N_p) & & & \\ & \bar{r}_{N_p+1, N_p+1}^{-1} & & \\ & & \ddots & \\ & & & \bar{r}_{N_r-N_k, N_r-N_k}^{-1} \end{bmatrix}$$
- 14 $\mathbf{B}_{t+1} = \begin{bmatrix} & \mathbf{B}_t(1 : N_p, 1 : N_r - N_k) \\ \mathbf{R}_{t+1}^H(N_p + 1 : N_r - N_k, :) & \mathbf{G}_{t+1} \end{bmatrix}$

$t + 1$, the three filters $\mathbf{F}_{t+1}, \mathbf{G}_{t+1}, \mathbf{B}_{t+1}$ and upper triangular matrix \mathbf{R}_{t+1} at time $t + 1$ can be expressed using $\mathbf{F}_t, \mathbf{G}_t, \mathbf{B}_t$ and \mathbf{R}_t at time t as

$$\mathbf{F}_{t+1} = (\mathbf{F}_t \mathbf{G}_v^H)(:, 1 : N_r - N_k), \quad (28)$$

$$\mathbf{R}_{t+1} = (\mathbf{G}_v \bar{\mathbf{R}})(1 : N_r - N_k, 1 : N_r - N_k), \quad (29)$$

$$\mathbf{G}_{t+1} = \begin{bmatrix} \mathbf{G}_t(:, 1 : N_p) & \\ & \text{diag}(\mathbf{R}_{t+1}^{-1}(N_p + 1 : N_r - N_k, :)) \end{bmatrix}, \quad (30)$$

$$\mathbf{B}_{t+1} = \begin{bmatrix} & \mathbf{B}_t(1 : N_p, 1 : N_r - N_k) \\ \mathbf{R}_{t+1}^H(N_p + 1 : N_r - N_k, :) & \mathbf{G}_{t+1} \end{bmatrix}, \quad (31)$$

where $\bar{\mathbf{R}} = [\mathbf{R}_t(:, 1 : N_p), \mathbf{R}_t(:, N_p + N_k + 1 : N_r)]$, $\mathbf{G}_v \in \mathbb{C}^{N_r \times N_r}$ is a unitary matrix and $1 \leq N_k \leq N_r$.

Proof: When k neighboring UE $_{p:p+k-1}$ depart from the cell, it is equivalent to deleting N_k columns from the $(N_p + 1)$ th column in the channel matrix, and we have the new channel matrix $\mathbf{H}_{t+1} = [\mathbf{H}_t(:, 1 : N_p), \mathbf{H}_t(:, N_p + N_k + 1 : N_r)]$.

Given the QR decomposition of \mathbf{H}_t , we have $\mathbf{H}_t = [\mathbf{H}_t(:, 1 : N_p), \mathbf{H}_p, \mathbf{H}_t(:, N_p + N_k + 1 : N_r)] = \mathbf{Q}_t [\mathbf{R}_t(:, 1 : N_p), \mathbf{R}_p, \mathbf{R}_t(:, N_p + N_k + 1 : N_r)]$, where \mathbf{H}_p is the channel matrix of UE $_{p:p+k-1}$ leaving the cell. Thus, we can get \mathbf{H}_{t+1} as follows:

$$\mathbf{H}_{t+1} = \mathbf{Q}_t [\mathbf{R}_1 \quad \mathbf{R}_2] = \mathbf{Q}_t \bar{\mathbf{R}}, \quad (32)$$

where $\mathbf{R}_1 = \mathbf{R}_t(:, 1 : N_p)$, $\mathbf{R}_2 = \mathbf{R}_t(:, N_p + N_k + 1 : N_r)$ and an example for $\bar{\mathbf{R}}$ is given in Case 7 of Appendix E. Similar to the approach in the proof Proposition 3, we use a series of Givens matrices $\mathbf{G}_i(\bar{r}_{i,j}, \bar{r}_{i+1,j})$ to eliminate $\bar{r}_{i,j}$, where $j + N_k - 1 \leq i \leq j$ and $N_p + 1 \leq j \leq N_r - N_k$.

Denote $\mathbf{G}_v = \prod_{j=N_p+1}^{N_r-N_k} \prod_{i=j+N_k-1}^j \mathbf{G}_i(\bar{r}_{i,j}, \bar{r}_{i+1,j})$ and we have

$$\mathbf{H}_{t+1} = \mathbf{Q}_t \bar{\mathbf{R}} = \underbrace{\mathbf{Q}_t \mathbf{G}_v^H}_{\bar{\mathbf{Q}}} \underbrace{\mathbf{G}_v \bar{\mathbf{R}}}_{\bar{\mathbf{R}}}. \quad (33)$$

Similar to Algorithm 2, $\mathbf{Q}_{t+1} \in \mathbb{C}^{M \times (N_r - N_k)}$ is a sub-matrix of $\bar{\mathbf{Q}}$ by deleting the last N_k columns and $\mathbf{R}_{t+1} \in \mathbb{C}^{(N_r - N_k) \times (N_r - N_k)}$ is a sub-matrix of $\bar{\mathbf{R}}$ by deleting the last N_k rows. Then we can easily obtain \mathbf{F}_{t+1} , \mathbf{G}_{t+1} and \mathbf{B}_{t+1} by (3)-(5). Furthermore, if $p + k - 1 = N$, i.e. the last k UEs leave the cell, we note that $\bar{\mathbf{R}}$ is an upper triangular matrix and Givens matrix is not required. We also show an example in Case 8 of Appendix E. ■

The overall procedure of the proposed algorithm is summarized in Algorithm 5. Now we derive the computation complexity of Algorithm 5. Since Steps 5 ~ 9 require $\mathcal{O}(N_k(M + N_r))$ operations, Steps 4 ~ 10 require $\mathcal{O}(N_k(N_r - N_k - N_p)(M + N_r))$ operations. Moreover, there are $(N_r - N_k - N_p)(N_r - N_k + N_p + 1)/2$ operations in Steps 13 ~ 14. Therefore, the complexity of Algorithm 5 is $\mathcal{O}(N_k(N_r - N_k - N_p)(M + N_r) + (N_r - N_k - N_p)(N_r - N_k + N_p + 1)/2)$. If $p = 1$, i.e. UE_{1:k} depart from the cell, the worst-case complexity is around $\mathcal{O}(N_k N_r M)$, and if $p = N - k + 1$, i.e. the last k UEs depart from the cell, the best-case complexity is around $\mathcal{O}(1)$. In contrast, the conventional algorithm requires $\mathcal{O}(M(N_r - N_k)^2)$.

Remark 2 (Low-complexity Iterative Update Algorithm). If we set $N_{r,i} = 1$ and $k = 1$ in Algorithm 5, the resulting algorithm is equivalent to Algorithm 2. In other words, Algorithm 5 is essentially equivalent to Algorithm 2 with N_k iterations. Therefore, we do not need to provide additional details about the iterative update design for k UEs departure. In practice, the number of UEs departure is generally much smaller than the total number of UEs, i.e. $N_k \ll N_r$, which means that the complexity of Algorithm 5 is much lower than that of the conventional algorithm.

V. SUM ACHIEVABLE RATE OF PROPOSED ALGORITHM WITH IMPERFECT CSI

In this section, we will derive approximate closed-form expressions for the sum achievable rate of the proposed algorithms with imperfect CSI in the Rayleigh channel. Both cases of quasi-static fading and time-varying scenarios will be considered.

A. Quasi-Static Fading Scenario

Proposition 5 (The Sum Approximate Achievable Rate of Proposed Algorithm for a UE Arrival). When a UE arrives in the cell under the quasi-static fading scenario at time $t + 1$, the sum approximate achievable rate of the proposed algorithm can be expressed as

$$C_{\text{sum}}^{\text{QA}} = (N + 1) \log_2(1 + \gamma^{\text{QA}}) \quad (34)$$

where γ^{QA} is the signal-to-interference-plus-noise ratio (SINR), i.e.,

$$\gamma^{\text{QA}} \approx \frac{\sigma_s^2}{\left(\sigma_s^2 \sigma_e^2 + \frac{\sigma_n^2}{M}\right) \sum_{i=1}^{N+1} r_{i,i}^{-2}}. \quad (35)$$

Proof: Please see in the Appendix F. ■

Proposition 6 (The Sum Approximate Achievable Rate of Proposed Algorithm for a UE Departure). When a UE departs from the cell under the quasi-static fading scenario at time $t + 1$, the sum approximate achievable rate of the proposed algorithm can be expressed as

$$C_{\text{sum}}^{\text{QD}} = (N - 1) \log_2(1 + \gamma^{\text{QD}}) \quad (36)$$

where γ^{QD} is the SINR, i.e.,

$$\gamma^{\text{QD}} \approx \frac{\sigma_s^2}{\left(\sigma_s^2 \sigma_e^2 + \frac{\sigma_n^2}{M}\right) \sum_{i=1}^{N-1} r_{i,i}^{-2}}. \quad (37)$$

Proof: The proof of Proposition 6 is similar to the proof of Proposition 5. We will not elaborate on this further. ■

B. Time-Varying Scenario

When channel is time-varying, we use the classic Gauss-Markov model [40], [41], i.e.,

$$\mathbf{H}_{t+1} = \mathbf{H}_t + \Delta \mathbf{E}, \quad (38)$$

where the entries of $\Delta \mathbf{E}$ are assumed to be i.i.d. $\mathcal{CN}(0, 2(1 - \rho))$, and ρ is the normalized correlation coefficient specified by the Jakes' model, i.e., $\rho = J_0(2\pi f_d \tau)$, with $J_0(\cdot)$ the zeroth-order Bessel function of the first kind, and the f_D and τ is the maximum Doppler frequency and channel block length, respectively.

Proposition 7 (The Sum Approximate Achievable Rate of Proposed Algorithm for a UE Arrival). When a UE arrives in the cell under the time-varying scenario at time $t + 1$, the sum approximate achievable rate of the proposed algorithm can be expressed as

$$C_{\text{sum}}^{\text{TA}} = N \log_2(1 + \gamma_o^{\text{TA}}) + \log_2(1 + \gamma_n^{\text{TA}}), \quad (39)$$

where γ_o^{TA} and γ_n^{TA} are the SINR of original N UEs and the new arrived UE, i.e.,

$$\gamma_o^{\text{TA}} \approx \frac{\sigma_s^2}{\left[\sigma_s^2 (\sigma_e^2 + 2(1 - \rho)) + \frac{\sigma_n^2}{M}\right] \sum_{i=1}^{N+1} r_{i,i}^{-2}}, \quad (40)$$

$$\gamma_n^{\text{TA}} \approx \frac{\sigma_s^2}{\left(\sigma_s^2 \sigma_e^2 + \frac{\sigma_n^2}{M}\right) \sum_{i=1}^{N+1} r_{i,i}^{-2}}. \quad (41)$$

Proof: Please see in the Appendix G. ■

Proposition 8 (The Sum Approximate Achievable Rate of Proposed Algorithm for a UE Departure). When a UE departs from the cell under the time-varying scenario at time $t + 1$, the sum approximate achievable rate of the proposed algorithm can be expressed as

$$C_{\text{sum}}^{\text{TD}} = (N - 1) \log_2(1 + \gamma^{\text{TD}}) \quad (42)$$

where γ^{TD} is the SINR, i.e.,

$$\gamma^{\text{TD}} \approx \frac{\sigma_s^2}{\left[\sigma_s^2 (\sigma_e^2 + 2(1 - \rho)) + \frac{\sigma_n^2}{M}\right] \sum_{i=1}^{N-1} r_{i,i}^{-2}}. \quad (43)$$

Proof: The proof of Proposition 8 is similar to the proof of Proposition 7. We will not elaborate on this further. ■

It is clear that the channel estimation and time-variations error have a baneful effect on the achievable rate. Then we set $\rho = 1$, (39) and (42) reduce to (34) and (36), respectively. For the multiple multi-antenna UEs scenario, we can substitute $N + 1$ or $N - 1$ with $N_r + N_k$ or $N_r - N_k$ and obtain the corresponding sum approximate achievable rate in the Rayleigh channel. Moreover, we also evaluate the performance of the proposed algorithms by simulation in the mmWave channel with imperfect CSI in Section VI-D.

VI. SIMULATION RESULTS AND DISCUSSIONS

In this section, we will evaluate the performance of the proposed THP update algorithm via simulation. We first will describe the channel models and other relevant communication parameters. Then, we will present and evaluate the communication performance of the proposed THP update algorithm of single-antenna UE and multiple multi-antenna UEs, respectively. Finally, the impact of imperfect CSI will be discussed.

A. Simulation Setup

1) *mmWave Channel*: We consider the widely used mmWave MIMO model [42] as shown below:

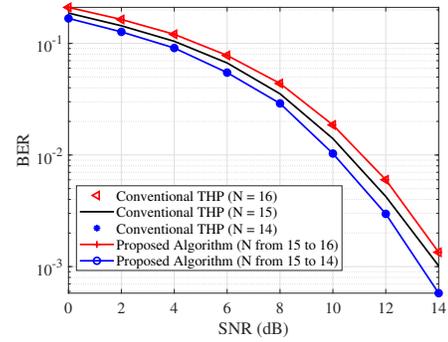
$$\mathbf{H}_k = \sqrt{\frac{MN_{r,k}}{L_k}} \sum_{i=1}^{L_k} \alpha_{k,i} \mathbf{a}_t(\phi_{k,i}^t) \mathbf{a}_r^H(\phi_{k,i}^r), \quad (44)$$

where L_k is the total number of propagation paths in \mathbf{H}_k , $\alpha_{k,i}$ is the complex channel gain of the i th path, and $\mathbf{a}_t(\phi_{k,i}^t)$ and $\mathbf{a}_r(\phi_{k,i}^r)$ is the array response vectors at the BS and at the k th UE, respectively. For a N -element linear line antenna array (ULA), the array response vector can be expressed as

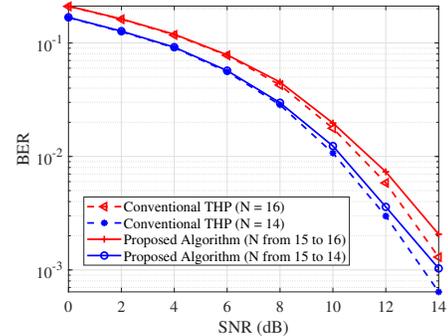
$$\mathbf{a}(\phi) = \frac{1}{\sqrt{N}} [1, e^{j2\pi \frac{d}{\lambda} \sin \phi}, \dots, e^{j2\pi \frac{d}{\lambda} (N-1) \sin \phi}]^T. \quad (45)$$

The BS and all the UEs are assumed to be equipped with ULAs with an inter-antenna spacing of half of the wavelength. We assume that $M = 64$, $N = 16$, $N_{r,k} = 1$ and $L_k = 3$ for all $k = 1, \dots, N$ in the single-antenna scenario and $M = 64$, $N = 8$, $N_{r,k} = 2$ and $L_k = 3$ for all $k = 1, \dots, N$ in the multi-antenna scenario. Additionally, the channel gain $\alpha_{k,i}$ follows the complex normal distribution $\mathcal{CN}(0, 1)$. The angle of departure (AoD) $\phi_{k,i}^t$ and angle of arrival (AoA) $\phi_{k,i}^r$ for each path follow the uniform distribution $\mathcal{U}(-\pi, \pi)$. 16QAM modulation is adopted, and the carrier frequency is 28GHz. For the time-varying scenario, we adopt a classic time-correlated model in [43], [44], i.e., $\alpha_{k,i}(t+1) = \rho \alpha_{k,i}(t) + \sqrt{1 - \rho^2} \Delta a$, $\phi_{k,i}^t(t+1) = \phi_{k,i}^t(t) + \Delta \phi$ and $\phi_{k,i}^r(t+1) = \phi_{k,i}^r(t) + \Delta \phi$, where $\Delta a \sim \mathcal{CN}(0, 1)$.

2) *Rayleigh Channel*: The channel matrix \mathbf{H}_k is adopted by Rayleigh channel, i.e., the entries of \mathbf{H}_k is i.i.d $\mathcal{CN}(0, 1)$. We set $M = 64$ and $N = 32$ in the single-antenna scenario and $M = 64$, $N = 16$ and $N_{r,k} = 2$ in the multi-antenna scenario. For the time-varying scenario, we adopt the Gauss-Markov model in (38).



(a) Quasi-static fading scenario



(b) Slow time-varying scenario with $\rho = 0.999$ and $\Delta \phi = 10^{-4}$

Fig. 3. Average BER comparison between the proposed algorithm and conventional THP in the scenario of single-antenna UE with 16QAM and mmWave channel.

B. Single-antenna UE

The performance of the proposed algorithm and conventional THP in terms of BER is compared in Fig. 3a for single-antenna UE with different SNR levels. Specifically, the BER is evaluated as the number of UEs increases from $N = 15$ to $N = 16$, indicating the arrival of a UE, or decreases from $N = 15$ to $N = 14$, indicating the departure of a UE. It can be observed that the BER of the proposed algorithm is identical to that of the conventional THP since the proposed algorithm is mathematically equivalent to the conventional THP without any approximation errors. It is important to note that these comparisons are made in a quasi-static fading scenario. Thus, the proposed algorithms can greatly reduce the computational complexity, without any penalty on the BER performance.

In Fig. 3b, the BER performance of the proposed algorithm and conventional THP is illustrated for single-antenna UE with different SNRs. The results indicate that the proposed algorithm experiences a negligible performance loss due to the slow time-varying scenario caused by the channel time-variations. Moreover, the departure of a UE causes a larger performance loss than the arrival of a UE. This is because the BS has knowledge of the channel state information for the new UE, i.e., BS has partial information about the channel at time $t + 1$. However, when a UE departs from the cell, the BS has no information about the channel at time $t + 1$.

In Fig. 4, we compare the BER performance of the proposed algorithm with conventional THP at SNR = 14 dB for single-

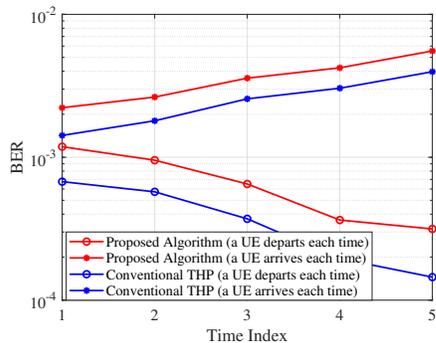
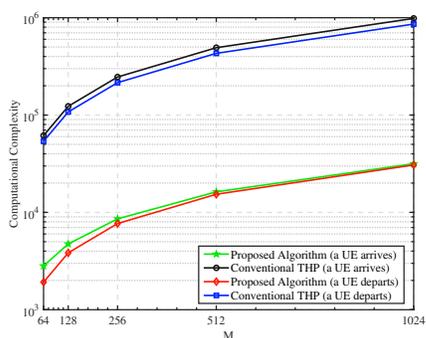
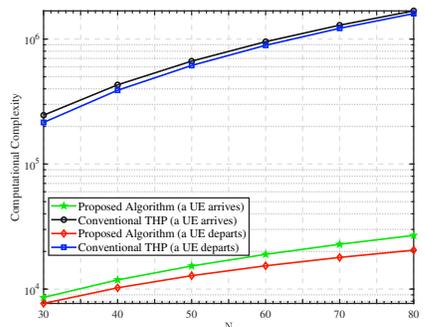


Fig. 4. Average BER as a function of the time index t for $\text{SNR} = 14$ dB for the slow time-varying scenario in the scenario of single-antenna UE with 16QAM and mmWave channel.



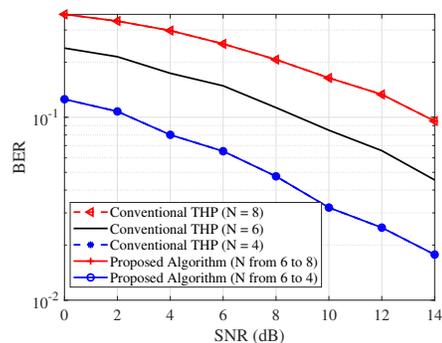
(a) Computational complexity versus M for $N = 30$ and $p = 1$



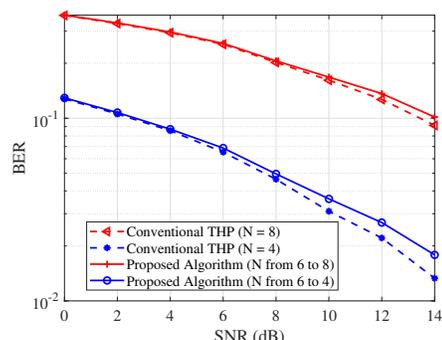
(b) Computational complexity versus N for $M = 256$ and $p = 1$

Fig. 5. Computational complexity comparison between the proposed algorithm and conventional THP with different numbers of antennas M and UEs N in the scenario of single-antenna UE.

antenna UE with different time index t . In this scenario, the number of UEs adds or decreases by one each time, starting with an initial number $N = 15$. The BER performance gap between the proposed algorithm and conventional THP gets bigger with the increase of time t . Note that when the time t is sufficiently large, the BER performance loss due to channel time-variations becomes unacceptable, and it may become necessary to recompute the three filter \mathbf{F} , \mathbf{G} , \mathbf{B} to maintain optimal performance. Furthermore, [31] provides a threshold value, which can be used to determine when to recompute the



(a) Quasi-static fading scenario



(b) Slow time-varying scenario with $\rho = 0.999$ and $\Delta\phi = 10^{-4}$

Fig. 6. Average BER comparison between the proposed algorithm and conventional THP in the scenario of multi-antenna UE with 16QAM and mmWave channel.

filters.

In Fig. 5, the computational complexity of the proposed algorithm and conventional THP is compared for single-antenna UE with different numbers of antennas M and UEs N . It can be observed that the proposed algorithm consistently exhibits lower computational complexity than conventional THP regardless of the numbers of antennas or UEs from Fig. 5a and 5b. Therefore, the proposed algorithm is a highly efficient alternative to provide benefits in practical systems where computational resources are often limited.

C. Multiple Multi-antenna UEs

In Fig. 6a, the BER performance of the proposed algorithm and conventional THP are compared for multi-antenna UE, when the number of UEs increases from $N = 6$ to $N = 8$, indicating the arrival of two UEs, or decreases from $N = 6$ to $N = 4$, indicating the departure of two UEs. Similar to the previous case in Fig. 3a, the BER performance of the proposed algorithm is indistinguishable from that of the conventional THP. Hence, complexity reduction is achieved at no cost of performance loss in this case. Similar to single-antenna UE, we evaluate the BER performance of the proposed algorithm and conventional THP in the slow time-varying scenario for multi-antenna UE in Fig. 6b. It is observed that the proposed algorithms only experience a negligible performance loss compared with conventional THP.

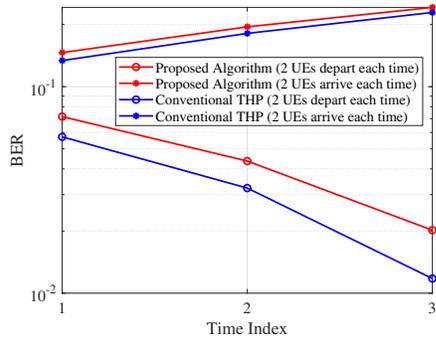
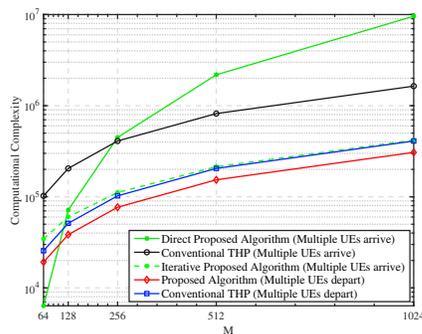
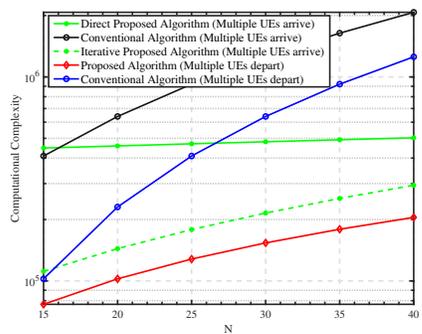


Fig. 7. Average BER as a function of the time t for SNR = 14 dB for the slow time-varying scenario in the scenario of multi-antenna UE with 16QAM and mmWave channel.



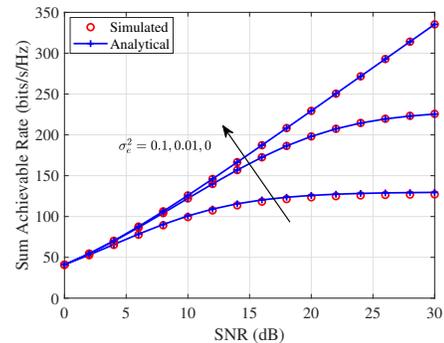
(a) Computational complexity versus M for $N = 15$, $N_{r,i} = 2$, $p = 1$ and $k = 5$



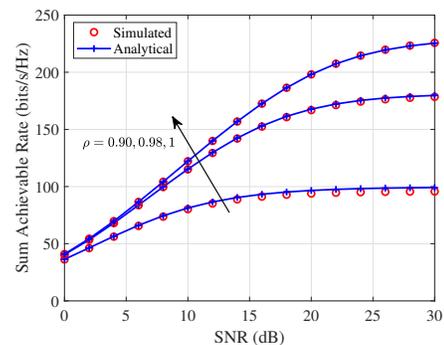
(b) Computational complexity versus N for $M = 256$, $N_{r,i} = 2$, $p = 1$ and $k = 5$

Fig. 8. Computational complexity comparison between the proposed algorithm and conventional THP with different numbers of antennas M and UEs N in the scenario of multi-antenna UE.

In Fig. 7, the BER performance of the proposed algorithm and conventional THP at SNR = 14 dB is illustrated for the scenario of multi-antenna UE versus different time t . In this scenario, the number of UEs adds or decreases by two each time, starting with an initial number $N = 8$. In comparison to Fig. 4, the BER performance loss is more pronounced in the time-varying scenario, indicating that more frequent re-computation is necessary. Nevertheless, the proposed algorithms continue to strike a good trade-off between computation complexity and BER performance in the slow time-varying scenario.



(a) Quasi-static fading scenario



(b) Time-varying scenario with $\sigma_e^2 = 0.01$

Fig. 9. Sum achievable rate of the proposed algorithms for a UE arrival against SNR with $M = 64$, N from 31 to 32, $\sigma_s^2 = 1$ and Rayleigh channel.

In Fig. 8, we compare the computational complexity of the proposed algorithm in the worst case with conventional THP for different numbers of antennas M and UEs N . The computational complexity of the iterative algorithm for the departure of multiple UEs is not shown in Fig. 8, since it is identical to that of the direct algorithm. As shown in Fig. 8a, the direct algorithm has a lower computational complexity than conventional THP and iterative algorithms when $M = 64$ for multiple UEs arrival. However, for larger M values such as $M = 128, 256, 512$ or 1024 , the computational complexity of the direct algorithm increases dramatically due to $\mathcal{O}(N_k M^2)$, which exceeds that of conventional THP. In such cases, the iterative algorithm is a better choice. To further illustrate the impact of changing the number of UEs, we fix $M = 256$ and plot the computational complexity of both algorithms as a function of N in Fig. 8b. It is worth noting that the direct algorithm when multiple UEs arrive is less affected by changes in N compared with the iterative algorithm. Therefore, when N is great large, the direct algorithm can be considered as a more efficient alternative. Additionally, as can be observed from Fig. 8, the proposed algorithm for the departure of multiple UEs always has lower complexity than that of conventional THP.

D. The impact of Imperfect CSI

We provide simulated and analytical results on the downlink sum achievable rate of the proposed algorithm for a UE arrival over quasi-static fading scenario and time-varying scenario as

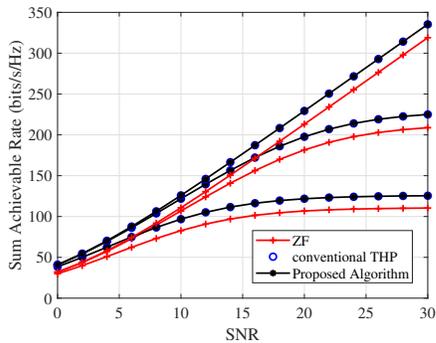
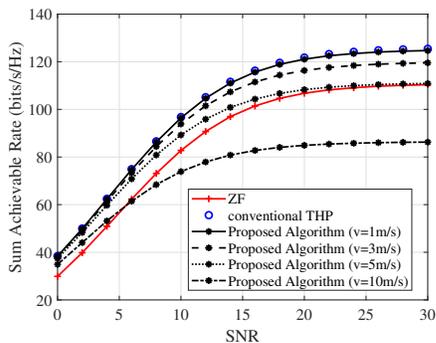
(a) Quasi-static fading scenario with $\sigma_e^2 = 0, 0.01, 0.1$ (b) Time-varying scenario with $\sigma_e^2 = 0.1$

Fig. 10. Sum achievable rate of the proposed algorithms and ZF for a UE arrival against SNR with $M = 64$, N from 31 to 32, $\sigma_s^2 = 1$ and Rayleigh channel.

an example. As is shown in Fig. 9a and 9b, the simulated and analytical curves are very tight in all considered cases, which confirms the correctness of our derived results.

Moreover, we also compare the sum achievable rate of the proposed algorithm with conventional THP and ZF in the quasi-static fading scenario and time-varying scenario under the Rayleigh channel as shown in Fig. 10a and 10b. These results show that the sum achievable rate of the proposed algorithms is identical to that of the conventional THP in the quasi-static fading scenario with imperfect CSI since the proposed algorithm is mathematically equivalent to the conventional THP without any approximation errors. In Fig. 11, we compare the BER of the proposed algorithm with conventional THP and ZF in the time-varying scenario under the mmWave channel, where $v = 0 \text{ km/h}$ ¹ means the quasi-static fading scenario. Similar with the results in the Fig. 10, the proposed algorithms continue to strike a good trade-off between computation complexity and performance compared with conventional THP and ZF in the slow time-varying scenario. Additionally, for the time-varying scenario, as the velocity of the UE increases, i.e., the channel variations are becoming increasingly severe, the performance of the proposed algorithm deteriorates. Hence, in the case of fast time-varying scenario, it may be more suitable to employ

¹In the Rayleigh channel, we set $\tau = 1 \text{ ms}$ with a carrier frequency f_c of 2.8GHz [45]. For the mmWave channel, τ is set to 0.1ms [44]. Given a fixed velocity value v , we can calculate the Doppler frequency shift as $f_d = f_c v/c$ and have $\rho = J_0(2\pi f_d \tau)$.

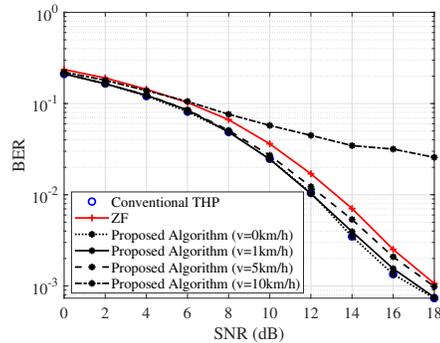


Fig. 11. Average BER comparison between the proposed algorithm with conventional THP and ZF for a UE arrival against SNR in the time-varying scenario with $M = 64$, N from 15 to 16, $\sigma_e^2 = 0.01$, 16QAM and mmWave channel.

conventional THP or ZF.

VII. CONCLUSIONS

In this paper, we have proposed low-complexity THP update algorithms for the multi-user massive MIMO downlink system. First, we have utilized the Gram-Schmidt process and a series of Givens matrices to derive the THP update algorithms for scenarios where a single-antenna UE arrives or departs. And then proposed algorithms have been extended to handle multiple multi-antenna UEs arriving or departing from the cell. Furthermore, the proposed algorithms significantly reduce the complexity of conventional THP, reducing it from $\mathcal{O}(MN^2)$ to $\mathcal{O}(MN)$ when a new UE arrives in the cell. Moreover, we have derived closed-form approximate expressions for the sum achievable rate of the proposed algorithms with imperfect CSI over the Rayleigh channel. Finally, simulation results have indicated that the proposed algorithms perform equally well as conventional THP in the quasi-static fading scenarios, and remain effective even in the slow time-varying scenarios. Future work includes investigating the impact of time-varying scenarios and determining the optimal user order for symbol encoding.

APPENDIX A

AN EXAMPLE FOR $\tilde{\mathbf{R}}$ WHEN A UE ARRIVES IN THE CELL

Case 1. When $M = 10$, $N = 5$ and $p = 2$, we have an example of $\tilde{\mathbf{R}}$ as follows:

$$\tilde{\mathbf{R}} = \begin{bmatrix} + & \otimes & + & + & + & + \\ 0 & \otimes & + & + & + & + \\ 0 & \ominus & \oplus & + & + & + \\ 0 & \ominus & 0 & \oplus & + & + \\ 0 & \ominus & 0 & 0 & \oplus & + \\ 0 & \ominus & 0 & 0 & 0 & \oplus \end{bmatrix}, \quad (46)$$

where $+$ represents the old nonzero entries to remain, \otimes represents the new added entries to remain, \ominus represents the new added entries to be eliminated and \oplus represents the zero entries to be filled.

Case 2. When $M = 10$, $N = 5$ and $p = N + 1 = 6$, we have an example of $\tilde{\mathbf{R}}$ as follows:

$$\tilde{\mathbf{R}} = \begin{bmatrix} + & + & + & + & + & \otimes \\ 0 & + & + & + & + & \otimes \\ 0 & 0 & + & + & + & \otimes \\ 0 & 0 & 0 & + & + & \otimes \\ 0 & 0 & 0 & 0 & + & \otimes \\ 0 & 0 & 0 & 0 & 0 & \otimes \end{bmatrix}, \quad (47)$$

where $+$ represents the old nonzero entries to remain and \otimes represents the new added nonzero entries to remain.

APPENDIX B

DEFINITION OF COMPLEX GIVENS MATRIX

A complex Givens matrix $\mathbf{G}_i(a, b) \in \mathbb{C}^{N \times N}$ can be expressed as follows:

$$\mathbf{G}_i(a, b) = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & c & s & \cdots & 0 \\ 0 & \cdots & -s^* & c^* & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{matrix} i \\ i+1 \end{matrix}, \quad (48)$$

where $c = a^*/\sqrt{a^2 + b^2}$ and $s = b^*/\sqrt{a^2 + b^2}$, which means that the scalar c and s have

$$\begin{bmatrix} c & s \\ -s^* & c^* \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} d \\ 0 \end{bmatrix}, \quad (49)$$

where d is a complex scalar. Therefore, the Givens matrix is an effective tool for introducing zero entries in matrix computations.

APPENDIX C

AN EXAMPLE FOR $\tilde{\mathbf{R}}$ WHEN A UE DEPARTS FROM THE CELL

Case 3. When $M = 10$, $N = 8$ and $p = 2$, we give an example of $\tilde{\mathbf{R}}$ as follows:

$$\tilde{\mathbf{R}} = \begin{bmatrix} + & + & + & + & + & + & + & + \\ 0 & + & + & + & + & + & + & + \\ 0 & \ominus & + & + & + & + & + & \oplus \\ 0 & 0 & \ominus & + & + & + & + & \oplus \\ 0 & 0 & 0 & \ominus & + & + & + & \oplus \\ 0 & 0 & 0 & 0 & \ominus & + & + & \oplus \\ 0 & 0 & 0 & 0 & 0 & \ominus & + & \oplus \\ 0 & 0 & 0 & 0 & 0 & 0 & \ominus & \oplus \end{bmatrix}, \quad (50)$$

where $+$ represents the old nonzero entries to remain, \ominus represents the old entries to be eliminated and \oplus represents the zero entries to be filled.

Case 4. When $M = 10$, $N = 8$ and $p = N = 8$, we give an example of $\tilde{\mathbf{R}}$ as follows:

$$\tilde{\mathbf{R}} = \begin{bmatrix} + & + & + & + & + & + & + & + \\ 0 & + & + & + & + & + & + & + \\ 0 & 0 & + & + & + & + & + & + \\ 0 & 0 & 0 & + & + & + & + & + \\ 0 & 0 & 0 & 0 & + & + & + & + \\ 0 & 0 & 0 & 0 & 0 & + & + & + \\ 0 & 0 & 0 & 0 & 0 & 0 & + & + \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & + \end{bmatrix}, \quad (51)$$

where $+$ represents the old nonzero entries to remain.

APPENDIX D

AN EXAMPLE FOR $\bar{\mathbf{R}}$ WHEN MULTIPLE UES ARRIVE IN THE CELL

Case 5. When $M = 10$, $N = 5$, $N_{r,i} = 1$, $p = 2$ and $k = 3$, we have an example of $\bar{\mathbf{R}}$ as follows:

$$\bar{\mathbf{R}} = \begin{bmatrix} + & \otimes & \otimes & \otimes & + & + & + & + \\ 0 & \otimes & \otimes & \otimes & + & + & + & + \\ 0 & \ominus & \otimes & \otimes & \oplus & + & + & + \\ 0 & \ominus & \ominus & \otimes & \oplus & \oplus & + & + \\ 0 & \ominus & \ominus & \ominus & \oplus & \oplus & \oplus & + \\ 0 & \ominus & \ominus & \ominus & 0 & \oplus & \oplus & \oplus \\ 0 & \ominus & \ominus & \ominus & 0 & 0 & \oplus & \oplus \\ 0 & \ominus & \ominus & \ominus & 0 & 0 & 0 & \oplus \\ 0 & \ominus & \ominus & \ominus & 0 & 0 & 0 & 0 \\ 0 & \ominus & \ominus & \ominus & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (52)$$

where $+$ represents the old nonzero entries to remain, \otimes represents the new added entries to remain, \ominus represents the new added entries to be eliminated and \oplus represents the zero entries to be filled.

Case 6. When $M = 10$, $N = 5$, $N_{r,i} = 1$, $p = N + 1 = 6$ and $k = 3$, we have an example of $\bar{\mathbf{R}}$ as follows:

$$\bar{\mathbf{R}} = \begin{bmatrix} + & + & + & + & + & \otimes & \otimes & \otimes \\ 0 & + & + & + & + & \otimes & \otimes & \otimes \\ 0 & 0 & + & + & + & \otimes & \otimes & \otimes \\ 0 & 0 & 0 & + & + & \otimes & \otimes & \otimes \\ 0 & 0 & 0 & 0 & + & \otimes & \otimes & \otimes \\ 0 & 0 & 0 & 0 & 0 & \otimes & \otimes & \otimes \\ 0 & 0 & 0 & 0 & 0 & \ominus & \otimes & \otimes \\ 0 & 0 & 0 & 0 & 0 & \ominus & \ominus & \otimes \\ 0 & 0 & 0 & 0 & 0 & \ominus & \ominus & \ominus \\ 0 & 0 & 0 & 0 & 0 & \ominus & \ominus & \ominus \end{bmatrix}, \quad (53)$$

where $+$ represents the old nonzero entries to remain, \otimes represents the new added entries to remain and \ominus represents the new added entries to be eliminated.

APPENDIX E

AN EXAMPLE FOR $\bar{\mathbf{R}}$ WHEN MULTIPLE UES DEPART FROM THE CELL

Case 7. When $M = 10$, $N = 8$, $N_{r,i} = 1$, $k = 2$ and $p = 2$, we give an example of $\bar{\mathbf{R}}$ as follows:

$$\bar{\mathbf{R}} = \begin{bmatrix} + & + & + & + & + & + & + & + \\ 0 & + & + & + & + & + & + & + \\ 0 & \ominus & + & + & + & + & + & + \\ 0 & \ominus & \ominus & + & + & + & + & + \\ 0 & 0 & \ominus & \ominus & + & + & + & + \\ 0 & 0 & 0 & \ominus & \ominus & + & + & + \\ 0 & 0 & 0 & 0 & \ominus & \ominus & + & + \\ 0 & 0 & 0 & 0 & 0 & \ominus & \ominus & + \end{bmatrix}, \quad (54)$$

where $+$ represents the old nonzero entries to remain and \ominus represents the old entries to be eliminated.

Case 8. When $M = 10$, $N = 8$, $N_{r,i} = 1$, $k = 2$ and $p = N - k + 1 = 7$, we give an example of $\bar{\mathbf{R}}$ as follows:

$$\bar{\mathbf{R}} = \begin{bmatrix} + & + & + & + & + & + \\ 0 & + & + & + & + & + \\ 0 & 0 & + & + & + & + \\ 0 & 0 & 0 & + & + & + \\ 0 & 0 & 0 & 0 & + & + \\ 0 & 0 & 0 & 0 & 0 & + \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (55)$$

where $+$ represents the old nonzero entries to remain.

APPENDIX F PROOF OF PROPOSITION 5

We assume that the channel matrix of the new UE is $\mathbf{h}_p \in \mathbb{C}^{M \times 1}$. Without loss of generality, we assume that the position of the new UE is at the end of the channel matrix. Then the channel for computing THP filters at time $t + 1$ is $\hat{\mathbf{H}}_{t+1} = [\hat{\mathbf{H}}_t, \hat{\mathbf{h}}_p]$, where $\hat{\mathbf{h}}_p = \mathbf{h}_p + \Delta\mathbf{h}$. Moreover, we denote $\hat{\mathbf{F}}, \hat{\mathbf{G}}, \hat{\mathbf{B}}$ are the three THP filters generated by the proposed algorithm under imperfect CSI.

Then, for the k th UE, through the practical equivalent channel $\mathbf{h}_{t+1}^{(k)}$ at time $t + 1$, the received signal is given by

$$\begin{aligned} y_k &= \beta(\mathbf{h}_{t+1}^{(k)H} \cdot \beta^{-1} \hat{\mathbf{F}} \hat{\mathbf{G}} \hat{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) + n_k) \\ &= \beta((\hat{\mathbf{h}}_{t+1}^{(k)} - \Delta\mathbf{h})^H \cdot \beta^{-1} \hat{\mathbf{F}} \hat{\mathbf{G}} \hat{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) + n_k) \\ &= \hat{\mathbf{h}}_{t+1}^{(k)H} \hat{\mathbf{F}} \hat{\mathbf{G}} \hat{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) - \Delta\mathbf{h}^H \hat{\mathbf{F}} \hat{\mathbf{G}} \hat{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) + \beta n_k \quad (56) \\ &= \underbrace{s_k + d_k}_{\text{desired signal}} - \underbrace{\Delta\mathbf{h}^H \hat{\mathbf{F}} \hat{\mathbf{G}} \hat{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d})}_{\text{channel estimation}} + \underbrace{\beta n_k}_{\text{noise}}, \end{aligned}$$

where $s_k + d_k$ is the desired signal for k th UE, and $-\Delta\mathbf{h}^H \hat{\mathbf{F}} \hat{\mathbf{G}} \hat{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) + \beta n_k$ is the interference-plus-noise term with its variance, i.e., power, expressed as follows:

$$\begin{aligned} r_k &= \mathbb{E}\{\Delta\mathbf{h}^H \hat{\mathbf{F}} \hat{\mathbf{G}} \hat{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) \cdot (\mathbf{s} + \mathbf{d})^H \hat{\mathbf{B}}^{-H} \hat{\mathbf{G}}^H \hat{\mathbf{F}}^H \Delta\mathbf{h}\} \\ &\quad + \beta^2 \mathbb{E}\{n_k n_k^H\} \\ &\stackrel{(a)}{\approx} \sigma_s^2 \mathbb{E}\{\Delta\mathbf{h}^H \hat{\mathbf{F}} \hat{\mathbf{G}} \hat{\mathbf{B}}^{-1} \hat{\mathbf{F}}^H \Delta\mathbf{h}\} + \beta^2 \sigma_n^2 \\ &= \text{trace}\left(\sigma_s^2 \mathbb{E}\{\Delta\mathbf{h}^H \hat{\mathbf{F}} \hat{\mathbf{G}} \hat{\mathbf{B}}^{-1} \hat{\mathbf{F}}^H \Delta\mathbf{h}\}\right) + \beta^2 \sigma_n^2 \\ &= \text{trace}\left(\sigma_s^2 \mathbb{E}\{\hat{\mathbf{G}}^H \hat{\mathbf{F}}^H \Delta\mathbf{h} \Delta\mathbf{h}^H \hat{\mathbf{F}} \hat{\mathbf{G}}\}\right) + \beta^2 \sigma_n^2 \\ &\stackrel{(b)}{=} \text{trace}\left(\sigma_s^2 \sigma_e^2 \mathbb{E}\{\hat{\mathbf{G}}^H \hat{\mathbf{G}}\}\right) + \beta^2 \sigma_n^2 \\ &\stackrel{(c)}{\approx} \text{trace}\left(\sigma_s^2 \sigma_e^2 \text{diag}(r_{1,1}^{-2}, \dots, r_{N+1,N+1}^{-2})\right) + \beta^2 \sigma_n^2 \\ &= \sigma_s^2 \sigma_e^2 \sum_{i=1}^{N+1} r_{i,i}^{-2} + \beta^2 \sigma_n^2 \\ &= \left(\sigma_s^2 \sigma_e^2 + \frac{\sigma_n^2}{M}\right) \sum_{i=1}^{N+1} r_{i,i}^{-2}, \quad 1 \leq k \leq N+1, \end{aligned} \quad (57)$$

where at (a) point, when the shaping loss [46] is neglected, $\mathbb{E}\{\hat{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) \cdot (\mathbf{s} + \mathbf{d})^H \hat{\mathbf{B}}^{-H}\} \approx \sigma_s^2 \mathbf{I}_N$ holds true. At (b) point, $\mathbb{E}\{\Delta\mathbf{h} \Delta\mathbf{h}^H\} = \sigma_e^2 \mathbf{I}_M$ and $\hat{\mathbf{F}}^H \hat{\mathbf{F}} = \mathbf{I}_N$. At (c) point, according to [35], we assume that $\hat{\mathbf{G}} \approx \mathbf{G}$.

When the modulo loss [46] is neglected, i.e., the power of $s_k + d_k$ is thought to be approximately equal to that of s_k , the SINR γ^{QA} of the k UE can be expressed as

$$\gamma^{\text{QA}} \approx \frac{\sigma_s^2}{\left(\sigma_s^2 \sigma_e^2 + \frac{\sigma_n^2}{M}\right) \sum_{i=1}^{N+1} r_{i,i}^{-2}}. \quad (58)$$

Finally, we have the sum approximate achievable rate of the proposed algorithm as shown in (34).

APPENDIX G PROOF OF PROPOSITION 7

Without loss of generality, we assume that the position of the new UE is in the end of the channel matrix. Then the practical channel matrix at time $t + 1$ is given by

$$\mathbf{H}_{t+1} = \begin{bmatrix} \underbrace{\mathbf{H}_t + \Delta\mathbf{E}}_{\text{time-varying}} & \underbrace{\mathbf{h}_p}_{\text{new UE}} \end{bmatrix}. \quad (59)$$

Note that the channel matrix for computing the three THP filters $\hat{\mathbf{F}}, \hat{\mathbf{G}}, \hat{\mathbf{B}}$ is $\hat{\mathbf{H}}_{t+1} = [\hat{\mathbf{H}}_t, \hat{\mathbf{h}}_p]$, where $\hat{\mathbf{H}}_t$ and $\hat{\mathbf{h}}_p$ are the estimated channel matrix. Therefore, we can obtain the relationship between \mathbf{H}_{t+1} and $\hat{\mathbf{H}}_{t+1}$, i.e.,

$$\mathbf{H}_{t+1} = [\hat{\mathbf{H}}_t - \Delta\mathbf{H} + \Delta\mathbf{E} \quad \hat{\mathbf{h}}_p - \Delta\mathbf{h}]. \quad (60)$$

It is observed that the initial N UEs are affected by both channel estimation errors and time-variations, whereas the newly arrived UE in the proposed algorithm is only affected by channel estimation errors.

Then, for the k th UE, where $1 \leq k \leq N$, through the practical equivalent channel $\mathbf{h}_{t+1}^{(k)}$ at time $t + 1$, the received signal is given by

$$\begin{aligned} y_k &= \beta(\mathbf{h}_{t+1}^{(k)H} \cdot \beta^{-1} \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) + n_k) \\ &= \beta((\hat{\mathbf{h}}_{t+1}^{(k)} - \Delta\mathbf{h} + \Delta\mathbf{e})^H \cdot \beta^{-1} \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) + n_k) \\ &= \hat{\mathbf{h}}_{t+1}^{(k)H} \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) - \Delta\mathbf{h}^H \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) \\ &\quad + \Delta\mathbf{e}^H \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) + \beta n_k \\ &= \underbrace{s_k + d_k}_{\text{desired signal}} - \underbrace{\Delta\mathbf{h}^H \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d})}_{\text{channel estimation}} + \underbrace{\Delta\mathbf{e}^H \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d})}_{\text{channel time-variations}} \\ &\quad + \underbrace{\beta n_k}_{\text{noise}}, \end{aligned} \quad (61)$$

where $s_k + d_k$ is the desired signal for k th UE, and $-\Delta\mathbf{h}^H \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) + \Delta\mathbf{e}^H \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) + \beta n_k$ is the interference-plus-noise term with its variance, i.e., power, expressed as follows:

$$\begin{aligned} r_k &= \mathbb{E}\{\Delta\mathbf{h}^H \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) \cdot (\mathbf{s} + \mathbf{d})^H \tilde{\mathbf{B}}^{-H} \tilde{\mathbf{G}}^H \tilde{\mathbf{F}}^H \Delta\mathbf{h}\} \\ &\quad + \mathbb{E}\{\Delta\mathbf{e}^H \tilde{\mathbf{F}} \tilde{\mathbf{G}} \tilde{\mathbf{B}}^{-1} (\mathbf{s} + \mathbf{d}) \cdot (\mathbf{s} + \mathbf{d})^H \tilde{\mathbf{B}}^{-H} \tilde{\mathbf{G}}^H \tilde{\mathbf{F}}^H \Delta\mathbf{e}\} \\ &\quad + \beta^2 \mathbb{E}\{n_k n_k^H\} \\ &= \sigma_s^2 \sigma_e^2 \sum_{i=1}^{N+1} r_{i,i}^{-2} + 2\sigma_s^2 (1 - \rho) \sum_{i=1}^{N+1} r_{i,i}^{-2} + \beta^2 \sigma_n^2 \\ &= \left[\sigma_s^2 (\sigma_e^2 + 2(1 - \rho)) + \frac{\sigma_n^2}{M}\right] \sum_{i=1}^{N+1} r_{i,i}^{-2}, \quad 1 \leq k \leq N. \end{aligned} \quad (62)$$

For the new arrived UE, we have its variance from the Proposition 5 as follows:

$$r_n = \left(\sigma_s^2 \sigma_e^2 + \frac{\sigma_n^2}{M}\right) \sum_{i=1}^{N+1} r_{i,i}^{-2}. \quad (63)$$

Then the SINR γ_o^{TA} and γ_n^{TA} can be expressed as (40) and (41), respectively.

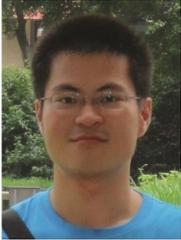
Finally, we have the sum approximate achievable rate of the proposed algorithm as shown in (39).

REFERENCES

- [1] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.
- [2] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.
- [3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.
- [4] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, 1983.
- [5] B. Hochwald, C. Peel, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiuser communication-part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, 2005.
- [6] C. Windpassinger, R. Fischer, and J. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2057–2060, 2004.
- [7] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, no. 5, pp. 138–139, 1971.
- [8] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, no. 4, pp. 774–780, 1972.
- [9] R. Wesel and J. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 824–831, 1998.
- [10] R. F. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-time transmission using Tomlinson-Harashima precoding," in *Proc. 4th ITG Conf. Source and Channel Coding*, 2002, pp. 139–148.
- [11] S. Zarei, W. Gerstacker, and R. Schober, "Low-complexity hybrid linear/Tomlinson-Harashima precoding for downlink large-scale MU-MIMO systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2016, pp. 1–7.
- [12] S. Zarei, W. H. Gerstacker, R. Weigel, M. Vossiek, and R. Schober, "Robust MSE-balancing hierarchical linear/Tomlinson-Harashima precoding for downlink massive MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7309–7324, 2018.
- [13] A. Farsaei, U. Gustavsson, A. Alvarado, and F. M. J. Willems, "A low-complexity hybrid linear and nonlinear precoder for Line-Of-Sight massive MIMO with max-min power control," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8410–8422, 2021.
- [14] R. Chen, M. Moretti, and X. Wang, "Hybrid TH-VP precoding for multiuser MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11 399–11 403, 2017.
- [15] D. Wubben, R. Bohnke, V. Kuhn, and K.-D. Kammeyer, "MMSE extension of V-BLAST based on sorted QR decomposition," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, vol. 1, 2003, pp. 508–512 Vol.1.
- [16] K. Zu, R. C. de Lamare, and M. Haardt, "Multi-branch Tomlinson-Harashima precoding design for MU-MIMO systems: Theory and algorithms," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 939–951, 2014.
- [17] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding," in *Proc. 2004 ITG Conf. Source and Channel Coding*, 2004, pp. 387–394.
- [18] T.-Y. Chang and C.-E. Chen, "A hybrid Tomlinson-Harashima transceiver design for multiuser mmwave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 118–121, 2018.
- [19] X. Bai, F. Liu, R. Du, Y. Xu, and Z. Sun, "Hybrid TH precoding and combining with sub-connected structure for mmwave systems," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1821–1824, 2020.
- [20] K. Xu, Y. Cai, M. Zhao, Y. Niu, and L. Hanzo, "MIMO-aided nonlinear hybrid transceiver design for multiuser mmwave systems relying on Tomlinson-Harashima precoding," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6943–6957, 2021.
- [21] Q. Wang, C. Xing, C. Du, L. Zhao, and L. Hanzo, "Hybrid nonlinear transceiver optimization for the RIS-aided MIMO downlink," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6441–6455, 2022.
- [22] F. Alibart, E. Zamanidoost, and D. B. Strukov, "Pattern classification by memristive crossbar circuits using ex situ and in situ training," *Nature Commun.*, vol. 4, no. 1, p. 2072, 2013.
- [23] S. Wen, R. Hu, Y. Yang, T. Huang, Z. Zeng, and Y.-D. Song, "Memristor-based echo state network with online least mean square," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 9, pp. 1787–1796, 2019.
- [24] H. Ran, S. Wen, Q. Li, Y. Yang, K. Shi, Y. Feng, P. Zhou, and T. Huang, "Memristor-based edge computing of blaze block for image recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 5, pp. 2121–2131, 2022.
- [25] D. Soudry, D. Di Castro, A. Gal, A. Kolodny, and S. Kvatinsky, "Memristor-based multilayer neural networks with online gradient descent training," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2408–2421, 2015.
- [26] P. Blanchard, N. J. Higham, and T. Mary, "A class of fast and accurate summation algorithms," *SIAM J. Sci. Comput.*, vol. 42, no. 3, pp. A1541–A1557, 2020.
- [27] L. M. Yang, A. Fox, and G. Sanders, "Rounding error analysis of mixed precision block householder QR algorithms," *SIAM J. Sci. Comput.*, vol. 43, no. 3, pp. A1723–A1753, 2021.
- [28] Y. M. Tsai, P. Luszczek, and J. Dongarra, "Mixed-precision algorithm for finding selected eigenvalues and eigenvectors of symmetric and hermitian matrices," in *Proc. IEEE/ACM Workshop Latest Adv. Scalable Algorithms Large-Scale Heterogeneous Sys. (ScalAH)*, 2022, pp. 43–50.
- [29] F. Rosário, F. A. Monteiro, and A. Rodrigues, "Fast matrix inversion updates for massive MIMO detection and precoding," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 75–79, 2016.
- [30] C.-Y. Wu, W.-J. Huang, and W.-H. Chung, "Robust update algorithms for zero-forcing detection in uplink large-scale MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 424–427, 2018.
- [31] Y. Xia, C. Jayawardena, and K. Nikitopoulos, "Reduced complexity matrix inversions in slow time-varying MIMO channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2022, pp. 456–461.
- [32] J. R. Bunch and C. P. Nielsen, "Updating the singular value decomposition," *Numer. Math.*, vol. 31, no. 2, pp. 111–129, 1978.
- [33] J. W. Daniel, W. B. Gragg, L. Kaufman, and G. W. Stewart, "Reorthogonalization and stable algorithms for updating the gram-schmidt QR factorization," *Math. Comput.*, vol. 30, no. 136, pp. 772–795, 1976.
- [34] G. W. Stewart, "Updating a rank-revealing ULV decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 14, no. 2, pp. 494–499, 1993.
- [35] M. Huang, S. Zhou, and J. Wang, "Analysis of Tomlinson-Harashima precoding in multiuser MIMO systems with imperfect channel state information," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 2856–2867, 2008.
- [36] C. Windpassinger, "Detection and precoding for multiple input multiple output channels," Ph.D. dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2004.
- [37] C. Windpassinger, R. Fischer, T. Vencel, and J. Huber, "Precoding in multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, 2004.
- [38] N. Lee, O. Simeone, and J. Kang, "The effect of imperfect channel knowledge on a MIMO system with interference," *IEEE Trans. Commun.*, vol. 60, no. 8, pp. 2221–2229, 2012.
- [39] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013.
- [40] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO wireless communications*. Cambridge university press, 2007.
- [41] D. Tsipouridou and A. P. Liavas, "On the sensitivity of the MIMO Tomlinson-Harashima precoder with respect to channel uncertainties," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2261–2272, 2010.
- [42] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.
- [43] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang, "Millimeter wave MIMO channel tracking systems," in *Proc. Globecom Workshops (GC Wkshps)*, 2014, pp. 416–421.
- [44] S. Srivastava, P. Singh, A. K. Jagannatham, A. Karandikar, and L. Hanzo, "Bayesian learning-based doubly-selective sparse channel estimation for millimeter wave hybrid MIMO-FBMC-OQAM systems," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 529–543, 2021.
- [45] H.-C. Chen and Y.-P. Lin, "Differential feedback of geometrical mean decomposition precoder for time-correlated MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3833–3845, 2017.
- [46] W. Yu, D. Varodayan, and J. Cioffi, "Trellis and convolutional precoding for transmitter-based interference presubtraction," *IEEE Trans. on Commun.*, vol. 53, no. 7, pp. 1220–1230, 2005.



Yiming Fang (Graduate Student Member, IEEE) received the bachelor's degree in communications engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2022. He is currently pursuing the master's degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China (USTC), Hefei, China. His research interests include low-complexity transceiver design, massive MIMO, and finite-precision float-point arithmetic.



Li Chen (Senior Member, IEEE) received the B.E. in electrical and information engineering from Harbin Institute of Technology, Harbin, China, in 2009 and the Ph.D. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2014. He is currently an Associate Professor with the Department of Electronic Engineering and Information Science, University of Science and Technology of China. His research interests include integrated communication and computation, integrated sensing and communication and

wireless IoT networks.



Yunfei Chen (S'02-M'06-SM'10) received his B.E. and M.E. degrees in electronics engineering from Shanghai Jiaotong University, Shanghai, P.R.China, in 1998 and 2001, respectively. He received his Ph.D. degree from the University of Alberta in 2006. He is currently working as a Professor in the Department of Engineering at the University of Durham, UK. His research interests include wireless communications, cognitive radios, wireless relaying and energy harvesting.



Huarui Yin (M'08) received the bachelor's degree and the Ph.D. degree in electronic engineering and information science from the University of Science and Technology of China (USTC), Hefei, Anhui, in 1996 and 2006, respectively. Since 2010, he has been with the Department of Electronic Engineering and Information Science, USTC, as an Associate Professor. His research interests include digital signal processing, low complexity receiver design, multiple access for massive connections, and throughput analysis of wireless networks.



Guo Wei received the B.S. degree in electronic engineering from the University of Science and Technology of China (USTC), Hefei, China, in 1983 and the M.S. and Ph.D. degrees in electronic engineering from the Chinese Academy of Sciences, Beijing, China, in 1986 and 1991, respectively. He is currently a Professor with the School of Information Science and Technology, USTC. His current research interests include wireless and mobile communications, wireless multimedia communications, and wireless information networks.



Citation on deposit: Fang, Y., Chen, L., Chen, Y., Yin, H., & Wei, G. (in press). Low-complexity Tomlinson-Harashima precoding update algorithm for massive MIMO system. IEEE Transactions on Communications

For final citation and metadata, visit Durham Research Online URL:

<https://durham-repository.worktribe.com/output/2046900>

Copyright statement: This accepted manuscript is licensed under the Creative Commons Attribution 4.0 licence.

<https://creativecommons.org/licenses/by/4.0/>