This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2025.3558930

Randomized Iterative Algorithms for Distributed Massive MIMO Detection

Zheng Wang, *Senior Member, IEEE*, Cunhua Pan,Yongming Huang, *Fellow, IEEE*, Shi Jin, *Fellow, IEEE* and Giuseppe Caire, *Fellow, IEEE*

***Abstract*—Distributed detection over decentralized baseband architectures has emerged as an important problem in the uplink massive MIMO systems. In this paper, the classic Kaczmarz method is fully investigated to facilitate the distributed detection for massive MIMO. First of all, a more general iteration performance result about the traditional randomized block Kaczmarz (RBK) method is derived, which paves the way for conceiving the conditional randomized block Kaczmarz (CRBK) algorithm. By customizing RBK with the concept of conditional sampling, CRBK achieves faster convergence and smaller error bound than RBK. To further exploit the potential of conditional sampling, multi-step conditional randomized block Kaczmarz (MCRBK) algorithm is proposed, which can be readily adopted as a flexible, scalable, low-complexity distributed detection scheme to suit various decentralized baseband architectures in massive MIMO. Moreover, to eliminate the convergence error bound of MCRBK, a novel dynamic step-size mechanism is proposed for the iteration update of MCRBK. Theoretical demonstration shows that the proposed distributed MCRBK detection with the optimized dynamic step-size not only converges exponentially to the solution of linear detection schemes but also enjoys the global convergence to well suit different practical scenarios of massive MIMO.**

***Index Terms*—Distributed signal detection, decentralized baseband detection, massive MIMO detection, low-complexity detection, large-scale MIMO systems.**

I. Introduction

As a promising extension of multiple-input multiple-output (MIMO), massive MIMO plays an important role in the beyond fifth-generation (B5G) and sixth-generation (6G) wireless communication systems due to its high spectral efficiency, excellent energy efficiency, and strong transmission reliability [1]–[4]. However, the dramatically increased system dimension of massive MIMO also imposes a pressing challenge upon the centralized signal processing, rendering distributed algorithms based on decentralized baseband architectures highly demanded [5]–[8]. Typically, a decentralized architecture allows the partial data to be locally processed in a distributed fashion, leading to lower complexity, interconnection data bandwidth and so on. Towards this end, a number of distributed signal detection schemes with customized decentralized baseband architectures have been proposed for the uplink massive MIMO systems [9]–[17].

Specifically, in [9], the decentralized baseband processing (DBP) architecture was proposed for distributed detection in massive MIMO systems. DBP partitions the antennas at base station (BS) into different distributed units (DUs) so that local channel estimation and signal detection are carried out within each DU (also called as cluster in some literature) to yield the extracted information, which is further exchanged among DUs as the consensus data. After that, a fusion node is applied in DBP to process these consensus data for the detection output. Based on DBP, the classic alternating direction method of multipliers (ADMM) and conjugate gradient (CG) methods were adopted for distributed detection to alleviate the baseband obstacles in both computational complexity and interconnection throughput [10]. Moreover, to avoid the exchange of the consensus data among DUs, feedforward architectures like partially decentralized (PD) and fully decentralized (FD) were given respectively to minimize the latency issue of DBP without sacrificing spectral efficiency [11]–[13]. Instead of the local detection within each DU, PD performs the local preprocessing and then sends the results to the central unit for the final decision. Different from it, local detection is still employed by FD while only the refined information is issued by each DU, followed by a feedforward fusion to output the detection results. Based on architectures of PD and FD, several methods have been proposed to mitigate the computation and interconnection bottlenecks [14], [15].

In [18], a fully decentralized architecture that obeys the *daisy-chain* structure was given, where the traditional recursive methods were employed for the distributed detection. In particular, given the received detection estimates from the last DU, every DU in daisy-chain computes the new estimates based on its local observations, and then conveys its estimates unidirectionally to the next DU [19], [20]. In a nutshell, the sequential implementation of the iterations in recursive methods naturally accounts for forwarding the detection estimates from one DU to another, thus providing a pipeline mechanism with low interconnection data bandwidth and great scalability (i.e., plug and play). Regarding daisy-chain, the classic decentralized topologies like ring, star, tree and so on can be further adopted. For example, in [21], with respect to ring and star topologies, a decentralized Newton (DN) scheme was proposed for low-complexity distributed detection. However, in order to benefit from the *channel hardening* property, it requires the stringent condition that the number of antennas in each DU (denoted

This work was supported in part by the National Natural Science Foundation of China under Grant 62371124, 61720106003, 62225107, in part by the Fundamental Research Funds for the Central Universities 2242022k60002, in part by Jiangsu Outstanding Youth Fund BK20240071 (*Corresponding author: Zheng Wang; Yongming Huang*.)

Z. Wang (wznuaa@gmail.com), C. Pan, Y. Huang and S. Jin are with School of Information Science and Engineering, Southeast University, Nanjing 210096, China. G. Caire is with Electrical Engineering and Computer Science Department, Technische Universität Berlin, Berlin, Germany (caire@tu-berlin.de).

by q) is sufficiently larger than that of the user side (denoted by K), making it limited in practice. Besides, the traditional message passing (MP) methods with modified decentralized architecture were also introduced to the distributed detection for massive MIMO, where the computed log-likelihood ratio (LLR) is sent to DUs as a priori information [22], [23]. A range of other distributed detection schemes for massive MIMO can be found in [24]–[26]. In addition, with the development of massive MIMO, distributed detection also plays an important role in extra-large scale MIMO (XL-MIMO), cell-free (CF) massive MIMO networks and so on [27]–[35].

On the other hand, as a famous iterative method to solve the systems of linear equations with low-complexity cost, the Kaczmarz method, which is also known as algebraic reconstruction technique (ART) in the realm of image processing, is widely applied in many different research domains ranging from computed tomography (CT) to signal processing [36]–[39]. Specifically, in [40], the traditional Kaczmarz method was adopted into massive MIMO systems for low-complexity detection or precoding to circumvent the computationally expensive operations of matrix inversion. A further step has been made in [41], where a Kaczmarz-based distributed detection scheme was proposed for massive MIMO systems. Unfortunately, it only allows one single antenna at each DU, which severely erodes its applications. Meanwhile, since the convergence analysis of the Kaczmarz method has always been hard to perform, further optimization and enhancement with respect to its distributed detection counterpart cannot be carried out. Although the randomized Kaczmarz (RK) method provides an analytic way for the convergence analysis, it fails to work for distributed detection in massive MIMO due to its random processing mechanism [42]. Similarly, such an issue also happens to the randomized block Kaczmarz (RBK) method [43]. Moreover, as for the inconsistent linear systems, there is an inevitable error bound in the convergence of the Kaczmarz method (including RK, RBK and so on), resulting in considerable performance loss in applications like distributed massive MIMO detection. For these reasons, the potential of Kaczmarz method for distributed detection in massive MIMO systems has not been well exploited.

In this paper, based on Kaczmarz, RK and RBK methods, the multi-step conditional randomized block Kaczmarz (MCRBK) algorithm with dynamic step-size is proposed for the distributed detection in massive MIMO systems. Our main contributions are listed in the following:

- We carefully reexamine the iteration performance of the RBK method, and obtain a new iteration performance result under general sampling probability. This forms the basis for the following research framework.
- Based on the derived results of RBK, a novel conditional randomized block Kaczmarz (CRBK) algorithm is proposed. By taking the sampling result in the last iteration into account, it not only achieves a faster convergence but also enjoys a smaller error bound than RBK. Subsequently, by considering more sampling results in the previous multiple iterations, further system gains in both convergence and error bound can be attained. This observation leads to the proposed multi-step conditional randomized block Kaczmarz (MCRBK) algorithm.
- We show that MCRBK can operate sequentially in a fixed order, which can be smoothly applied to the distributed detection in massive MIMO. Specifically, based on daisy-chain, ring or star topology, the distributed MCRBK detection can be readily realized via the scalable DUs (each iteration update in MCRBK accounts for the computation in each modular DU in the distributed detection) with multiple antennas $q \geq 1$ in each DU. Meanwhile, the proposed distributed MCRBK detection also entails low computational cost with complexity order less than $O(K^2)$, which is quite competitive compared to other distributed detection schemes in massive MIMO. In addition, the impact of different choices of the number of antennas q in each DU is also studied for the practical implementation.
- Finally, to mitigate the convergence obstacle of MCRBK incurred by the inherent error bound, we propose a novel iteration update mechanism for MCRBK, which is modified by an optimized step-size α_t. Specifically, we prove that the error bound in the convergence gradually diminishes with the number of iterations so that the proposed distributed MCRBK detection with dynamic step-size will exactly converge to the linear detection solution $\mathbf{x}^*$ (i.e., zero forcing (ZF) or minimum mean-square error (MMSE)) of massive MIMO systems. In this way, the convergence issue that bothers the Kaczmarz-based distributed detection is completely removed. Note that besides the exponential convergence rate, distributed MCRBK detection with dynamic step-size also enjoys the global convergence, making it rather appealing for different practical scenarios of massive MIMO.

The contributions of this paper with respect to the current literature on the Kaczmarz method are summarized in Table I. In Table II we present a comparison between the proposed approach and other distributed detection schemes for massive MIMO systems.

The organization of this paper is as follows. Section II describes the system model of the uplink detection in massive MIMO and briefly introduces the state of the art of the Kaczmarz method. Then, the more general iteration performance results about RBK are given in Section III. In Section IV, the CRBK algorithm and its upgrade MCRBK are proposed. In Section V, the application of MCRBK to scalable distributed detection for massive MIMO is explicitly investigated. In Section VI, the bespoke mechanism of dynamic step-size for MCRBK is presented, and the optimized step-size of each iteration is investigated in detail. In Section VII, the applications of the proposed MCRBK scheme to XL-MIMO and CF massive MIMO networks are also shortly discussed. Section VIII shows some simulation results illustrating the performance and complexity of the proposed scheme. Finally, Section IX concludes the paper.

Notation: Matrices and column vectors are denoted by upper and lowercase boldface letters, and the conjugate transpose, inverse, pseudoinverse of a matrix $\mathbf{B}$ by $\mathbf{B}^H, \mathbf{B}^{-1}$, and $\mathbf{B}^\dagger$, respectively. We use $\mathbf{b}_i$ for the ith column of the matrix $\mathbf{B}$, $b_{i,j}$ for the entry in the ith row and jth column of the matrix $\mathbf{B}$. $\|\cdot\|$

TABLE I

A BRIEF COMPARISON OF THE RELATED LITERATURE OF KACZMARZ METHODS FOR DISTRIBUTED DETECTION IN MASSIVE MIMO SYSTEMS.

	Deterministic processing order for distributed architecture	Flexible number of antennas in each DU	Explicit iteration performance	Convergent to the desired $\mathbf{x}^*$
Kaczmarz [36]	✓	×	×	×
RK [42]	×	×	✓	×
RBK under uniform distribution [43]	×	✓	✓	×
RBK under general distribution (this work)	×	✓	✓	×
CRBK (this work)	×	✓	✓	×
MCRBK with $f = r - 1$ (this work)	✓	✓	✓	×
MCRBK with $f = r - 1$ and α_t (this work)	✓	✓	✓	✓

denotes the *Euclidean* norm of a matrix (i.e., $\|\cdot\|_2$) and $\|\cdot\|_F$ is the standard *Frobenius* norm. $\mathbf{I}$ is the identity matrix and $\text{Tr}(\cdot)$ denotes the matrix trace. $\lambda_{\text{min}}(\cdot)$ and $\lambda_{\text{max}}(\cdot)$ represent the minimum and maximum eigenvalues of a matrix while $\sigma_{\text{min}}(\cdot)$ and $\sigma_{\text{max}}(\cdot)$ indicate the minimum and maximum singular values of a matrix. In this paper, the computational complexity is measured by the number of complex multiplications while the interconnection data bandwidth is evaluated by the matrix size of the conveyed data.

II. PRELIMINARY

In this section, the linear signal detection for uplink massive MIMO systems is reviewed, followed by the background of Kaczmarz method including RK and RBK.

A. Linear Uplink Signal Detection in Massive MIMO

Considering the signal detection in a massive MIMO system with Rayleigh fading channels. The base station is equipped with N antennas and multiple user equipments (UEs) with K antennas in all are served simultaneously, $K \le N$. Let $\mathbf{x}$ represent the $K \times 1$ transmit signal from the UEs, where the i-th element of $\mathbf{x}$ (i.e., x_i) is a symbol drawn from a QAM constellation $\mathcal{X}$. Then, given the full column rank channel matrix $\mathbf{H} \in \mathbb{C}^{N \times K}$, the $N \times 1$ received signal $\mathbf{y}$ at BS can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{n}$ is an $N \times 1$ AWGN noise vector with elements obeying $\mathcal{CN}(0, \sigma^2)$. To restore the transmitted signal $\mathbf{x}$ in (1), the optimal maximum likelihood (ML) detection aims to solve the following integer least square (ILS) problem[1]

$$\widehat{\mathbf{x}}_{\text{ml}} = \arg \min_{\mathbf{x} \in \mathcal{X}^K} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2, \tag{2}$$

which is NP-hard theoretically. Note that a full column rank channel matrix $\mathbf{H}$ is a default configuration for MIMO detection[2], otherwise the solution for $\mathbf{x}$ will not be unique.

[1] It also amounts to solving the closest vector problem (CVP) in lattice decoding [44].

[2] For example, any two UEs can not have the same channel responses in scale, e.g., $\mathbf{h}_i \neq \alpha \mathbf{h}_j$, α is a constant.

Fortunately, it has been demonstrated in [45] that if $N \gg K$ the channel matrix $\mathbf{H}$ will become near orthogonal as a benefit of *favorable propagation*, which implies the ILS problem in (2) can be well approximated by the following least square (LS) problem

$$\mathbf{x}_{\text{linear}} = \arg \min_{\mathbf{x} \in \mathbb{C}^K} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2. \tag{3}$$

As a result, the optimal ML detection solution in (2) can be well approximated by the solutions of the traditional linear detection schemes like ZF or MMSE (i.e., $\widehat{\mathbf{x}}_{\text{zf}}$ and $\widehat{\mathbf{x}}_{\text{mmse}}$) with

$$\mathbf{x}_{\text{zf}} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y} \text{ and } \mathbf{x}_{\text{mmse}} = \left(\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y}, \tag{4}$$

where the final decisions $\widehat{\mathbf{x}}_{\text{zf}}$ and $\widehat{\mathbf{x}}_{\text{mmse}}$ are determined by quantizing $\mathbf{x}_{\text{zf}}$ and $\mathbf{x}_{\text{mmse}}$ according to the modulation constellation $\mathcal{X}^K$, i.e.,

$$\widehat{\mathbf{x}}_{\text{zf}} = \lceil \mathbf{x}_{\text{zf}} \rfloor_Q \in \mathcal{X}^K \text{ and } \widehat{\mathbf{x}}_{\text{mmse}} = \lceil \mathbf{x}_{\text{mmse}} \rfloor_Q \in \mathcal{X}^K. \tag{5}$$

Nevertheless, due to the matrix inversion with computational complexity $\mathcal{O}(K^3)$, either ZF or MMSE detection turns out to be prohibitive especially for large dimensional systems. Besides the need of low-complexity implementation, there is also a high demand for the distributed detection with decentralized baseband processing architecture in the uplink massive MIMO systems [18]–[21].

B. Distributed Detection based on Classic Kaczmarz Method

To be specific, in the distributed detection for massive MIMO systems, the received signal $\mathbf{y}$ at the BS is partitioned into r blocks of q antennas each (i.e., $N = rq$). This corresponds to dividing the channel matrix $\mathbf{H}$ into r stacked vertically blocks, namely,

$$\mathbf{H} = [\mathbf{H}_1; \mathbf{H}_2; \ldots; \mathbf{H}_r] \text{ or } \mathbf{H} = [\mathbf{H}_1^H \mathbf{H}_2^H \ldots \mathbf{H}_r^H]^H, \tag{6}$$

where $\mathbf{H}_i \in \mathbb{C}^{q \times K}$ is the channel matrix at the i-th group of antennas. Generally, these groups of antennas are known as clusters or DUs, and various decentralized baseband architectures can be designed to unify these DUs for the distributed detection. Note that these distributed detection schemes based

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2025.3558930

on its own decentralized baseband architectures can be easily extended to some specific scenarios of massive MIMO. For example, the ADMM or CG detection based on DBP architecture can be easily introduced to cloud radio access networks (C-RAN) to effectively reduce the interconnection bandwidth between the remote radio head (RRH) and baseband unit (BBU) of C-RAN [5], [9], [10]. For this reason, in the work we focus our attention on the distributed detection based on the decentralized baseband architectures, while further extensions to specific application scenarios of massive MIMO (i.e., cell-free, XL-MIMO) will be briefly discussed based on the proposed algorithms.

On the other hand, as one of the most popular methods for solving overdetermined linear systems, the Kaczmarz method can date back to 1930's [36]. Specifically, regarding to a consistent linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ with full column rank matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$, Kaczmarz method iteratively computes the estimate by

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \frac{b_i - \mathbf{A}_{i,:}\mathbf{x}^{t-1}}{\|\mathbf{A}_{i,:}\|^2}\mathbf{A}_{i,:}^H, \tag{7}$$

where $\mathbf{A}_{i,:} \in \mathbb{C}^{1\times n}$ denotes the i-th row of matrix $\mathbf{A}$ and t is the iteration index. Clearly, each iteration of Kaczmarz contains a single orthogonal projection, which is cyclically performed with coordinate $i = (t-1) \mod m + 1$. However, although Kaczmarz method has been widely applied in numerous research fields from computer tomography to image processing [46], its convergence analysis is rather challenging, rendering it limited in comparison with other competitive schemes.

As for the distributed detection for massive MIMO, since Kaczmarz method only involves the row computation of matrix $\mathbf{A}$, it has been adopted to massive MIMO detection in a distributed way [41], i.e.,

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \frac{y_i - \mathbf{H}_{i,:}\mathbf{x}^{t-1}}{\|\mathbf{H}_{i,:}\|^2}\mathbf{H}_{i,:}^H, \tag{8}$$

where each distributed unit (DU) contains only one single antenna (i.e., $r = N, q = 1$). In this condition, the channel matrix $\mathbf{H}_i$ in (6) of each DU can be expressed as $\mathbf{H}_{i,:}$. Unfortunately, as for the inconsistent linear system $\mathbf{A}\mathbf{x}+\mathbf{e} = \mathbf{b}$ that is corrupted by noise, the convergence of Kaczmarz method is not guaranteed [47]. Hence, there is a substantial performance degradation of Kaczmarz method when it is employed for distributed detection in massive MIMO systems.

C. Randomized Kaczmarz Method

In terms of the mean square error, the exponential convergence rate of Kaczmarz method for consistent linear systems $\mathbf{A}\mathbf{x} = \mathbf{b}$ was explicitly given [42], namely,

$$E\|\mathbf{x}^L - \mathbf{x}^\flat\|^2 \leq \left(1 - \frac{\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}\right)^L \|\mathbf{x}^0 - \mathbf{x}^\flat\|^2, \tag{9}$$

where the coordinate i at each iteration of (7) is determined by randomly sampling according to the following distribution

$$p_i^{\text{norm}} = \frac{\|\mathbf{A}_{i,:}\|^2}{\|\mathbf{A}\|_F^2}. \tag{10}$$

Here, $\mathbf{x}^\flat = \mathbf{A}^\dagger\mathbf{b}$ and L denotes the number of iterations. This leads to randomized Kaczmarz (RK) method [42]. In other words, the expectation in (9) is with respect to the random sampling in (10).

Moreover, regarding to the inconsistent linear system $\mathbf{A}\mathbf{x}+\mathbf{e} = \mathbf{b}$, it has been demonstrated in [48] that RK converges to the ordinary least squares (LS) solution $\mathbf{x}^* = \mathbf{A}^\dagger\mathbf{b}$ within a specified error bound, i.e.,

$$E\|\mathbf{x}^L-\mathbf{x}^*\|^2 \leq \left(1-\frac{\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}\right)^L \|\mathbf{x}^0-\mathbf{x}^*\|^2 + \frac{\|\mathbf{A}\|_F^2}{\sigma_{\min}^2(\mathbf{A})}\gamma^2 \tag{11}$$

with $\gamma = \max_i \frac{|\mathbf{e}_i|}{\|\mathbf{A}_{i,:}\|}$ and $\mathbf{e} = \mathbf{A}\mathbf{x}^* - \mathbf{b}$. Theoretically speaking, this explains why the Kaczmarz-based distributed massive MIMO detection fails to converge to the desired linear solution $\mathbf{x}^*$, since the system model of massive MIMO detection naturally belongs to the inconsistent linear system.

As a variant of RK, the randomized block Kaczmarz algorithm (RBK) was given in [43]. In particular, it firstly partitions the row indices of $\mathbf{A}$ into different subsets $\mathcal{Q}_j, 1 \leq j \leq r$ as

$$\mathcal{Q}_1 \cup \ldots \cup \mathcal{Q}_r = \{1, \ldots, m\} \tag{12}$$

with $\mathcal{Q}_i \cap \mathcal{Q}_j = \emptyset$, $1 \leq i \neq j \leq r$. For simplicity, there are q elements in each subset $\mathcal{Q}_j$, i.e., $|\mathcal{Q}_j| = q$, which accounts for $r = m/q$ subsets in total. After that, at each iteration, a subset $\mathcal{Q}_j$ is randomly sampled from the partition $\mathcal{T} = \{\mathcal{Q}_1, \ldots, \mathcal{Q}_r\}$ by the uniform distribution

$$p_{\mathcal{Q}}^{\text{uniform}} = \frac{1}{r}. \tag{13}$$

This yields the corresponding row matrix partition $\mathbf{A}_{\mathcal{Q}_j,:} \in \mathbb{C}^{q\times n}$, and then the iteration is carried out by projecting the current result onto the solution space of $\mathbf{b}_{\mathcal{Q}_j} \in \mathbb{C}^q$, namely,

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \mathbf{A}_{\mathcal{Q}_j,:}^\dagger(\mathbf{b}_{\mathcal{Q}_j} - \mathbf{A}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1}), \tag{14}$$

where $\mathbf{b}_{\mathcal{Q}_j}$ denotes the subvector of $\mathbf{b}$ with elements indexed by $\mathcal{Q}_j$. Regarding to the convergence of RBK for inconsistent linear system, the following Theorem is recalled[3].

Theorem 1 ([43]). *For an inconsistent linear system* $\mathbf{A}\mathbf{x}+\mathbf{e} = \mathbf{b}$*, RBK under uniform distribution is convergent by*

$$E_{p_{\mathcal{Q}}^{\text{uniform}}}\|\mathbf{x}^L-\mathbf{x}^*\|^2 \leq \left(1-\frac{\sigma_{\min}^2(\mathbf{A})}{\beta r}\right)^L \|\mathbf{x}^0-\mathbf{x}^*\|^2 + \frac{\beta}{\alpha}\cdot\frac{\|\mathbf{e}\|^2}{\sigma_{\min}^2(\mathbf{A})}, \tag{15}$$

where $\alpha \leq \lambda_{\min}(\mathbf{A}_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:}^H) = \sigma_{\min}^2(\mathbf{A}_{\mathcal{Q}_j,:})$ *and* $\sigma_{\max}^2(\mathbf{A}_{\mathcal{Q}_j,:}) = \lambda_{\max}(\mathbf{A}_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:}^H) \leq \beta$ *for each* $\mathcal{Q}_j \in \mathcal{T}$*(i.e.,* $\alpha \leq \min_j \lambda_{\min}(\mathbf{A}_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:}^H)$ *and* $\beta \geq \max_j \lambda_{\max}(\mathbf{A}_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:}^H)$*),* $\mathbf{e} = \mathbf{A}\mathbf{x}^* - \mathbf{b}$.

III. RBK With General Sampling Probability

In this section, we reexamine the iteration performance of RBK method but with a general sampling probability $p_{\mathcal{Q}}$ (not a specified distribution like uniform distribution or following other sampling criterions), which provides a more general iteration performance.

[3]From Theorem 1, the convergence of RBK is partially determined by the choice of the partition $\mathcal{T}$ (see details given in [43]).

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2025.3558930

Theorem 2. *For an inconsistent linear system $\mathbf{Ax}+\mathbf{e}=\mathbf{b}$, given the general sampling probability $p_{\mathcal{Q}}$ over a fixed partition $\mathcal{T}$, RBK is convergent by*

$$E_{p_{\mathcal{Q}}}[\|\mathbf{x}^L-\mathbf{x}^*\|^2]\le(1-\eta_{\text{RBK}})^L\|\mathbf{x}^0-\mathbf{x}^*\|^2+\frac{q\|\mathbf{e}\|^2}{\eta_{\text{RBK}}m\varphi_{\text{RBK}}}, \quad (16)$$

with

$$0<\eta_{\text{RBK}}\triangleq\frac{\inf_{1\le j\le r}\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}{\sup_{1\le j\le r}\sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:})}\le 1 \quad (17)$$

and

$$\varphi_{\text{RBK}}\triangleq\inf_{1\le j\le r}\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})>0. \quad (18)$$

Proof. Given the sampling subset at the t-th iteration, i.e., $\mathcal{Q}^t=\mathcal{Q}_j$, we express its corresponding row matrix partition as $\mathbf{A}_{\mathcal{Q}_j,:}=\mathbf{I}_{\mathcal{Q}_j,:}\mathbf{A}$, where $\mathbf{I}_{\mathcal{Q}_j,:}\in\mathbb{C}^{q\times m}$ denotes a row concatenation containing q rows indexed by $\mathcal{Q}_j$ in an $m\times m$ identity matrix $\mathbf{I}$. Then, according to the iteration shown in (14), we have

$$\begin{aligned}\mathbf{x}^t&=\mathbf{x}^{t-1}+\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{I}_{\mathcal{Q}_j,:}(\mathbf{b}-\mathbf{Ax}^{t-1})\\&=\mathbf{x}^{t-1}+\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{I}_{\mathcal{Q}_j,:}(\mathbf{Ax}^*+\mathbf{e}-\mathbf{Ax}^{t-1})\\&=\mathbf{x}^{t-1}+\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:}\mathbf{x}^*-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1}+\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{e}_{\mathcal{Q}_j,:}, \quad (19)\end{aligned}$$

where equation (19) can be further reformatted as

$$\mathbf{x}^t-\mathbf{x}^*=(\mathbf{I}-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})(\mathbf{x}^{t-1}-\mathbf{x}^*)+\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{e}_{\mathcal{Q}_j,:}. \quad (20)$$

Next, since the term $(\mathbf{I}-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})$ is orthogonal to $\mathbf{A}^\dagger_{\mathcal{Q}_j,:}$ (i.e., $(\mathbf{I}-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})\mathbf{A}^\dagger_{\mathcal{Q}_j,:}=\mathbf{0}$), by *Pythagorean Theorem*, it follows that

$$\begin{aligned}\|\mathbf{x}^t-\mathbf{x}^*\|^2&=\|(\mathbf{I}-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})(\mathbf{x}^{t-1}-\mathbf{x}^*)\|^2+\|\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{e}_{\mathcal{Q}_j,:}\|^2\\&\overset{(a)}{\le}\|(\mathbf{I}-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})(\mathbf{x}^{t-1}-\mathbf{x}^*)\|^2+\|\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\|^2\|\mathbf{e}_{\mathcal{Q}_j,:}\|^2\\&=\|(\mathbf{I}-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})(\mathbf{x}^{t-1}-\mathbf{x}^*)\|^2+\sigma^2_{\max}(\mathbf{A}^\dagger_{\mathcal{Q}_j,:})\|\mathbf{e}_{\mathcal{Q}_j,:}\|^2\\&=\|(\mathbf{I}-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})(\mathbf{x}^{t-1}-\mathbf{x}^*)\|^2+\frac{\|\mathbf{e}_{\mathcal{Q}_j,:}\|^2}{\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}. \quad (21)\end{aligned}$$

Here, the inequality (a) comes from $\|\mathbf{AB}\|\le\|\mathbf{A}\|\|\mathbf{B}\|$.

Therefore, based on the general sampling probability $p_{\mathcal{Q}}$, by taking the expectation on both sides of (21), we can arrive at the following results

$$E_{p_{\mathcal{Q}}}[\|\mathbf{x}^t-\mathbf{x}^*\|^2]\le E_{p_{\mathcal{Q}}}[\|(\mathbf{I}-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})(\mathbf{x}^{t-1}-\mathbf{x}^*)\|^2]+E_{p_{\mathcal{Q}}}\left[\frac{\|\mathbf{e}_{\mathcal{Q}_j,:}\|^2}{\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}\right]. \quad (22)$$

Then, given the sampling probability $p_{\mathcal{Q}}$, we try to upper bound these two terms on the right-hand side (RHS) of (22), respectively. On one hand, it follows that

$$\begin{aligned}&E_{p_{\mathcal{Q}}}[\|(\mathbf{I}-\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})(\mathbf{x}^{t-1}-\mathbf{x}^*)\|^2]\\&\overset{(b)}{=}\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2-E_{p_{\mathcal{Q}}}[\|\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:}(\mathbf{x}^{t-1}-\mathbf{x}^*)\|^2]\\&\overset{(c)}{\le}\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2-E_{p_{\mathcal{Q}}}[\sigma^2_{\min}(\mathbf{A}^\dagger_{\mathcal{Q}_j,:}\mathbf{A}_{\mathcal{Q}_j,:})\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2]\\&\overset{(d)}{\le}\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2-E_{p_{\mathcal{Q}}}[\sigma^2_{\min}(\mathbf{A}^\dagger_{\mathcal{Q}_j,:})\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2]\\&\overset{(e)}{=}\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2-E_{p_{\mathcal{Q}}}\left[\frac{\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}{\sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:})}\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2\right]\\&=\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2-E_{p_{\mathcal{Q}}}\left[\sum_{1\le j\le r}p_{\mathcal{Q}_j}\frac{\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}{\sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:})}\right]\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2\\&\le\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2-\inf_{1\le j\le r}\left[\frac{\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}{\sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:})}\right]\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2\\&\le\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2-\frac{\inf_{1\le j\le r}\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}{\sup_{1\le j\le r}\sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:})}\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2\\&=\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2-\frac{\inf_{1\le j\le r}\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}{\beta}\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2\\&=(1-\eta_{\text{RBK}})\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2. \quad (23)\end{aligned}$$

Here, equality (b) follows *Pythagorean theorem*, inequality (c) holds due to

$$\min_{\mathbf{x}\ne\mathbf{0}}\frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}=\sigma_{\min}(\mathbf{A}) \quad (24)$$

so that $\|\mathbf{Ax}\|\ge\sigma_{\min}(\mathbf{A})\|\mathbf{x}\|$ for $\mathbf{x}\ne\mathbf{0}$, inequality (d) holds because of

$$\begin{aligned}\sigma_{\min}(\mathbf{AB})=\min_{\mathbf{x}\ne\mathbf{0}}\frac{\|\mathbf{ABx}\|}{\mathbf{x}}&\ge\min_{\mathbf{x}\ne\mathbf{0}}\frac{\|\mathbf{ABx}\|}{\mathbf{Bx}}\frac{\|\mathbf{Bx}\|}{\mathbf{x}}\\&\ge\min_{\mathbf{x}\ne\mathbf{0}}\frac{\|\mathbf{Ay}\|}{\mathbf{y}}\frac{\|\mathbf{Bx}\|}{\mathbf{x}}\\&=\sigma_{\min}(\mathbf{A})\sigma_{\min}(\mathbf{B}), \quad (25)\end{aligned}$$

and equality (e) holds by

$$\sigma^2_{\min}(\mathbf{A}^\dagger_{\mathcal{Q}_j,:})=\sigma^{-2}_{\max}(\mathbf{A}_{\mathcal{Q}_j,:}). \quad (26)$$

Moreover, because of $\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})=\lambda_{\min}(\mathbf{A}_{\mathcal{Q}_j,:}\mathbf{A}^H_{\mathcal{Q}_j,:})$ and $\sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:})=\lambda_{\max}(\mathbf{A}_{\mathcal{Q}_j,:}\mathbf{A}^H_{\mathcal{Q}_j,:})$, it is clear to see that

$$0<\eta_{\text{RBK}}=\frac{\inf_{1\le j\le r}\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}{\sup_{1\le j\le r}\sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:})}\le 1 \quad (27)$$

On the other hand, since the noise $\mathbf{e}$ is uniformly distributed, which is essentially independent of $p_{\mathcal{Q}}$, we can obtain

$$\begin{aligned}E_{p_{\mathcal{Q}}}\left[\frac{\|\mathbf{e}_{\mathcal{Q}_j,:}\|^2}{\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}\right]&=E_{p_{\mathcal{Q}}}\left[\frac{1}{\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}\right]\cdot\frac{q\|\mathbf{e}\|^2}{m}\\&=\frac{q\|\mathbf{e}\|^2}{m}\cdot\sum_{1\le j\le r}p_{\mathcal{Q}}\cdot\frac{1}{\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}\\&\le\frac{q\|\mathbf{e}\|^2}{m}\cdot\frac{1}{\inf_{1\le j\le r}\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}. \quad (28)\end{aligned}$$

Therefore, by casting (23) and (28) into (22), we have

$$E[\|\mathbf{x}^t-\mathbf{x}^*\|^2]\le(1-\eta_{\text{RBK}})\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2+\frac{q\|\mathbf{e}\|^2}{m\varphi_{\text{RBK}}}. \quad (29)$$

To sum up, by iteratively applying the recursive relation in (29) under the full expectation, we can arrive at

$$\begin{aligned}E_{p_{\mathcal{Q}}}[\|\mathbf{x}^L-\mathbf{x}^*\|^2]&\le(1-\eta_{\text{RBK}})^L\|\mathbf{x}^0-\mathbf{x}^*\|^2+\sum_{j=0}^{L-1}(1-\eta_{\text{RBK}})^j\frac{q\|\mathbf{e}\|^2}{m\varphi_{\text{RBK}}}\\&\overset{(f)}{\le}(1-\eta_{\text{RBK}})^L\|\mathbf{x}^0-\mathbf{x}^*\|^2+\frac{q\|\mathbf{e}\|^2}{\eta_{\text{RBK}}m\varphi_{\text{RBK}}}, \quad (30)\end{aligned}$$

where inequality (f) holds due to the fact $\sum_{i=0}^{\infty}(1-x)^i=$

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2025.3558930

$1/x$. □

IV. Multi-step Conditional Randomized Block Kaczmarz Algorithm

In RBK, the iteration driven by the random sampling of $\mathcal{Q}_j$ has a latent risk — the same choice of $\mathcal{Q}_j$ may be repeatedly sampled during a very small number of iterations. In this case, the sampling diversity is not well exploited so that the convergence may slow down.

To alleviate this problem, given the general sampling probability $p_{\mathcal{Q}}$, we introduce the concept of conditional sampling into RBK. Specifically, sampling at iteration t is performed with the following conditional distribution:

$$\overline{p}_{\mathcal{Q}} = \frac{p_{\mathcal{Q}}}{1 - p_{\mathcal{Q}}(\mathcal{Q}^{t-1})}, \tag{31}$$

which gives rise to the conditional randomized block Kaczmarz (CRBK) algorithm. Here, $\overline{p}_{\mathcal{Q}}$ denotes the conditional sampling probability about $\mathcal{Q}$ at iteration t (denoted by $\mathcal{Q}^t$), and it removes the sampling choice $\mathcal{Q}_k$ of the last iteration $t-1$ (i.e., $\mathcal{Q}^{t-1} = \mathcal{Q}_k$) from the current sampling space, i.e., $\mathcal{Q}^t \subseteq \mathcal{T} \backslash \{\mathcal{Q}_k\}$.

To make it more specific, here we let $\mathcal{Q}_j$ indicate the conditional sampling result of $\mathcal{Q}^t$, then we can have the following result

$$\mathcal{Q}^t = \mathcal{Q}_j \neq \mathcal{Q}_k = \mathcal{Q}^{t-1}. \tag{32}$$

By doing this, it is clear that the risk of sampling repetitions in any two consecutive iterations is eliminated, which leads to the better iteration performance than the traditional RBK.

A. Iteration Performance of CRBK

Similar to the proof in Theorem 2, we can readily attain the following iteration results about CRBK.

Theorem 3. *For an inconsistent linear system $\mathbf{Ax} + \mathbf{e} = \mathbf{b}$, given the conditional sampling probability $\overline{p}_{\mathcal{Q}}$ in (31) over a fixed partition $\mathcal{T}$, the convergence of the CRBK algorithm is described by*

$$E_{\overline{p}_{\mathcal{Q}}}[\|\mathbf{x}^t - \mathbf{x}^*\|^2] \leq (1 - \eta_{CRBK})\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \frac{q\|\mathbf{e}\|^2}{m\varphi_{CRBK}} \tag{33}$$

with the conditional parameters

$$0 < \eta_{\text{CRBK}} \triangleq \frac{\inf_{1\leq j\leq r, j\neq k} \sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}{\sup_{1\leq j\leq r, j\neq k} \sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:})} \leq 1 \tag{34}$$

and

$$\varphi_{\text{CRBK}} \triangleq \inf_{1\leq j\leq r, j\neq k} \sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:}) > 0. \tag{35}$$

Based on the iteration performance in (33) and (16), the gains in both convergence and error bound of the proposed CRBK algorithm over RBK can be easily confirmed due to

$$\inf_{1\leq j\leq r, j\neq k} \sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:}) \geq \inf_{1\leq j\leq r} \sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:}) \tag{36}$$

and

$$\sup_{1\leq j\leq r, j\neq k} \sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:}) \leq \sup_{1\leq j\leq r} \sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:}), \tag{37}$$

which leads to the following result.

Corollary 1. *For an inconsistent linear system $\mathbf{Ax} + \mathbf{e} = \mathbf{b}$, the proposed CRBK algorithm attains faster convergence with smaller error bound than RBK.*

$$\eta_{\text{CRBK}} \geq \eta_{\text{RBK}} \text{ and } \varphi_{\text{CRBK}} \geq \varphi_{\text{RBK}}. \tag{38}$$

According to (33) in Theorem 3, by iterations, the proposed CRBK algorithm converges exponentially to the least squares solution $\mathbf{x}^*$ within a specified error bound. Note that a well conditioned matrix $\mathbf{A}$ (i.e., a small value of condition number $\kappa = \sigma_{\max}/\sigma_{\min}$) naturally yields a better convergence.

B. Upgrade to MCRBK Algorithm

Inspired by the introduced gains in both convergence and error bound, we now extend the conditional sampling $\overline{p}_{\mathcal{Q}}$ in (31) to the multi-step conditional sampling probability $\overline{p}^f_{\mathcal{Q}}, 0 \leq f \leq r-1$, namely,

$$\overline{p}^f_{\mathcal{Q}} = \frac{p_{\mathcal{Q}}}{1 - p_{\mathcal{Q}}(\mathcal{Q}^{t-1}) - \ldots - p_{\mathcal{Q}}(\mathcal{Q}^{t-f})}, \tag{39}$$

which leads to the proposed multi-step conditional randomized block Kaczmarz (MCRBK) algorithm. Typically, compared to the conditional sampling probability $\overline{p}_{\mathcal{Q}}$ in (31), $\overline{p}^f_{\mathcal{Q}}$ takes the previous f iterations into account by removing their sampling results from the current sampling state space. Specifically, let $\mathcal{Q}_j$, $\mathcal{Q}_k$, and $\mathcal{Q}_l$ indicate the multi-step conditional sampling results of $\mathcal{Q}^t$, $\mathcal{Q}^{t-1}$, and $\mathcal{Q}^{t-f}$ respectively, then we have $\mathcal{Q}^t \subseteq \mathcal{T}\backslash\{\mathcal{Q}_k, \ldots, \mathcal{Q}_l\}$, namely,

$$\mathcal{Q}_j \nsubseteq \{\mathcal{Q}_k, \ldots, \mathcal{Q}_l\}. \tag{40}$$

In other words, $\overline{p}_{\mathcal{Q}}$ and $p_{\mathcal{Q}}$ are the special cases of $\overline{p}^f_{\mathcal{Q}}$ with $f = 1$ and $f = 0$, respectively.

Intuitively, similar to Theorem 3, the iteration performance of MCRBK can be attained as follows, where the proof is omitted here.

Theorem 4. *For an inconsistent linear system $\mathbf{Ax} + \mathbf{e} = \mathbf{b}$, given the multi-step conditional sampling probability $\overline{p}^f_{\mathcal{Q}}$ in (39) over a fixed partition $\mathcal{T}$, the proposed MCRBK algorithm is convergent by*

$$E_{\overline{p}^f_{\mathcal{Q}}}[\|\mathbf{x}^t - \mathbf{x}^*\|^2] \leq (1 - \eta^f)\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \frac{q\|\mathbf{e}\|^2}{m\varphi^f} \tag{41}$$

with

$$0 < \eta^f \triangleq \frac{\inf_{1\leq j\leq r, j\neq k,\ldots,l} \sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}{\sup_{1\leq j\leq r, j\neq k,\ldots,l} \sigma^2_{\max}(\mathbf{A}_{\mathcal{Q}_j,:})} \leq 1. \tag{42}$$

and

$$\varphi^f \triangleq \inf_{1\leq j\leq r, j\neq k,\ldots,l} \sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:}) > 0. \tag{43}$$

Clearly, given η^f and φ^f in (42) and (43), the advantages of MCRBK over CRBK can be demonstrated in a straightforward way, so that we can arrive at the following result.

Corollary 2. *For an inconsistent linear system $\mathbf{Ax} + \mathbf{e} = \mathbf{b}$, the proposed MCRBK algorithm attains faster convergence with smaller error bound than CRBK due to*

$$\eta^f \geq \eta_{CRBK} \text{ and } \varphi^f \geq \varphi_{CRBK} \tag{44}$$

for $f \geq 1$.

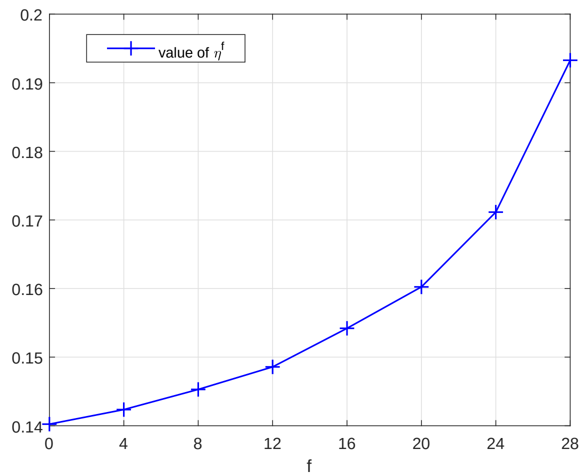

Fig. 1. Illustration of η^f in 128×16 massive MIMO systems with $q = 4$.

Intuitively, based on Theorem 4 and Corollary 2, the maximum choice $f = r - 1$ is highly recommended in the proposed MCRBK algorithm to achieve the best convergence performance and error bound by

$$\eta^{f=r-1} \geq \eta^{f=r-2} \geq \cdots \geq \eta^{f=1} = \eta_{\text{CRBK}} \geq \eta^{f=0} = \eta_{\text{RBK}} \quad (45)$$

and

$$\varphi^{f=r-1} \geq \varphi^{f=r-2} \geq \cdots \geq \varphi^{f=1} = \varphi_{\text{CRBK}} \geq \varphi^{f=0} = \varphi_{\text{RBK}}. \quad (46)$$

Here, for a better understanding, in Fig. 1 η^f's with $f = 0, 4, 8, \ldots, 28$ are presented as an example by *Monte Carlo* (MC) methods, where the 128×16 uncoded massive MIMO systems are applied with Rayleigh fading channels and $q = 4$. As expected, the convergence coefficient η^f improves gradually with the increment of f, which is in accordance to the derived results in (45).

V. Scalable Distributed Realization of MCRBK For Massive MIMO Detection

We now commence the discussion to bridge the proposed MCRBK algorithm and distributed detection for massive MIMO systems.

Besides the enhanced convergence performance and error bound, most interestingly, once $f = r - 1$ is applied in MCRBK, the sampling about the subset $\mathcal{Q}^t$ will become round robin deterministic when the iteration goes up to $t > r - 1$. This is straightforward to understand since the $r - 1$ sampling options among the total r options have been removed from the sampling state space. In this way, there is only one sampling option left for each iteration of $t > r - 1$, i.e.,

$$\mathcal{Q}^t \subseteq \mathcal{T} \backslash \{\mathcal{Q}^{t-1} = \mathcal{Q}_k, \ldots, \mathcal{Q}^{t-r+1} = \mathcal{Q}_l\} \quad (47)$$

with $|\mathcal{T} \backslash \{\mathcal{Q}^{t-1} = \mathcal{Q}_k, \ldots, \mathcal{Q}^{t-r+1} = \mathcal{Q}_l\}| = 1$ for $t > r - 1$, which corresponds to a derandomized sampling process. More precisely, for iterations $t > r - 1$, the order of the subset $\mathcal{Q}_j$ has been fixed, which is determined by the sampling results of the first $r - 1$ iterations. Because of this, the choice of the sampling distribution of $p_{\mathcal{Q}}$ only has the limit impact upon a few iterations in MCRBK with $f = r - 1$, making it less important compared to their roles in random cases. To this end, uniform sampling is highly preferred in practice due to simplicity.

Generally speaking, the deterministic choices of $\mathcal{Q}_j$ in a fixed order greatly facilitate the proposed MCRBK algorithm with $f = r - 1$ in practice as the sampling operations can be avoided without any performance loss, thus leading to a more efficient implementation. Meanwhile, because each iteration of MCRBK only concerns the partial rows of matrix $\mathbf{A}$, i.e., $\mathbf{A}_{\mathcal{Q}_j,:}$, it is applicable to work for the distributed signal detection in uplink massive MIMO systems, which will be introduced in what follows.

A. Distributed MCRBK Detection

Specifically, let $\mathcal{Q}_j$ with $1 \leq j \leq r$ denote the index of a distributed unit (DU) that contains q received antennas, and there are r DUs with $rq = N$ antennas in total. Accordingly, at each DU $\mathcal{Q}_j$, $\mathbf{H}_{\mathcal{Q}_j,:}$ accounts for the local channel matrix while $\mathbf{y}_{\mathcal{Q}_j}$ corresponds to the local received signal. Then, based on the proposed MCRBK algorithm with $f = r - 1$, the distributed MCRBK detection for uplink massive MIMO works by

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \mathbf{H}^{\dagger}_{\mathcal{Q}_j,:}(\mathbf{y}_{\mathcal{Q}_j} - \mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1}) \quad (48)$$

with $j = (t - 1) \mod r + 1$. Typically, along with the increment of t, every DU computes its estimates about the target signal $\mathbf{x}$ by (48) and then convey the estimates to the next DU via dedicated links for further processing. Intuitively, such an iteration mechanism enables the great scalability of the proposed distributed MCRBK detection, where each DU working as a system module can be freely added or removed. Note that there is a weak assumption here, namely, the samplings in the first $r-1$ iterations of MCRBK (with $f = r-1$) is also sequentially ordered, i.e., $\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_{r-1}$. Of course this would not alter the iteration convergence because the convergence holds for any possible order of $\mathcal{Q}_j$.

To make it more specific, Fig. 2 is given to illustrate the operations of the proposed distributed MCRBK detection for massive MIMO systems. Here, we point out that the ring topology is considered throughout the context but other topology schemes like daisy-chain or star can also be applied as well. The difference between ring and daisy-chain lies in the fact that there is only one iteration loop among the DUs in daisy-chain while the ring topology allows more flexible iteration loops. As for the ring topology, $t = n \cdot r$ accounts for the n-th iteration loop, where the parameter $n \geq 1$ is flexible to adjust. According to (41) in Theorem 4, the more iterations, the better convergence, and so as to the better detection performance until reaching the underlying error bound. On the other hand, accordingly, given the modular DU in MCRBK, other decentralized baseband processing architectures can be established in a flexible way. For example, the star topology shown in Fig. 3 can be mimicked as a sequential processing order like: $\mathcal{Q}_1 \to \mathcal{Q}_2 \to \mathcal{Q}_1 \to \mathcal{Q}_3 \to \mathcal{Q}_1 \to \mathcal{Q}_4 \to \cdots$. However, such a processing order is inferior to the sequential order of ring topology (i.e., $\mathcal{Q}_1 \to \mathcal{Q}_2 \to \mathcal{Q}_3 \to \mathcal{Q}_4 \to \cdots$)

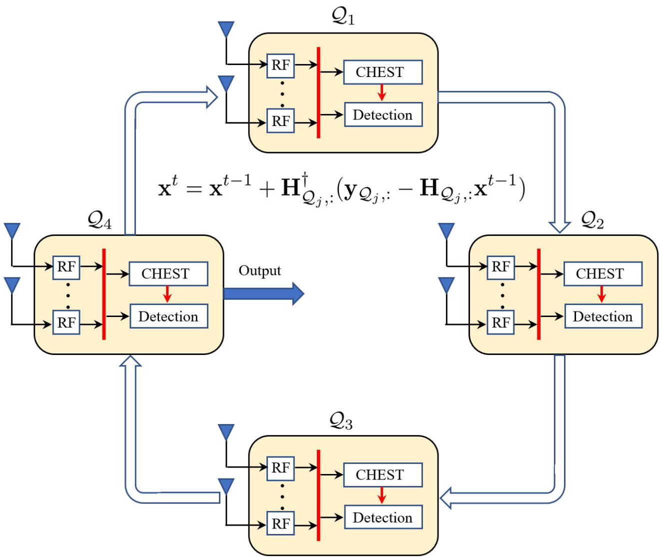

Fig. 2. Illustration of the distributed MCRBK detection for uplink massive MIMO, where a ring topology with 4 DUs is employed. Each DU is associated with local radio-frequency (RF) processing and channel estimation (CHEST).

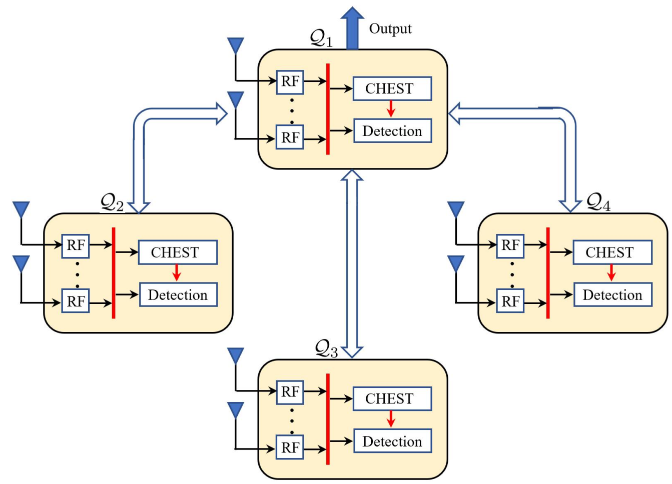

Fig. 3. Illustration of a star topology formed by the modular DUs of MCRBK, where 4 DUs are applied for uplink massive MIMO.

in terms of the processing diversity, which results in a reduced detection performance for uplink MIMO.

We now go through the computational complexity of the distributed MCRBK detection, which is measured by the required number of complex multiplications. In particular, with respect to the iteration in (48), the computational complexity of obtaining the pseudo-inverse $\mathbf{H}^{\dagger}_{\mathcal{Q}_j,:}$ is $2q^2K + q^3$; the complexity of computing $\mathbf{y}_{\mathcal{Q}_j} - \mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1}$ is qK; multiplying $\mathbf{H}^{\dagger}_{\mathcal{Q}_j,:}$ with $\mathbf{y}_{\mathcal{Q}_j} - \mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1}$ costs qK. Therefore, the total computational complexity of the proposed distributed massive MIMO detection at each iteration is $2q^2K + q^3 + 2qK$. More specifically, with $q \leq \sqrt{K}$, its overall complexity is no more than $O(K^2)$, which is rather competitive compared to other distributed detection schemes. On the other hand, the proposed distributed MCRBK detection is also appealing in terms of the data bandwidth cost, as only the vector $\mathbf{x}^t \in \mathbb{C}^K$ needs to be unidirectionally conveyed through DUs. More details about these comparisons can be found in Table II. Note that as shown in Theorem 4 the partition way of $\mathcal{T} = \{\mathcal{Q}_1, \ldots, \mathcal{Q}_r\}$ also plays a role in determining the convergence performance and the error bound of MCRBK. However, the construction of DUs in massive MIMO systems are physically predefined, which can not be altered in practice. For this reason, this point is out of the scope of this work and we leave it as one of the works in future.

Additionally, note that in the application of massive MIMO detection, the least squares solution of MCRBK $\mathbf{x}^*$ corresponds to the detection result of ZF in (4). Nevertheless, $\mathbf{x}^*$ can also be the detection result of MMSE shown in (4) via some simple transformations. One possible way is to build an extended system model as follows

$$\overline{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \sigma \mathbf{I}_K \end{bmatrix} \quad \text{and} \quad \overline{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_K \end{bmatrix}, \tag{49}$$

so that the MMSE detection result will be attained by the least squares solution as [49]

$$\mathbf{x}^* = (\overline{\mathbf{H}}^H \overline{\mathbf{H}})^{-1} \overline{\mathbf{H}}^H \overline{\mathbf{y}} = \left(\mathbf{H}^H\mathbf{H} + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{H}^H \mathbf{y}. \tag{50}$$

B. *Impact of the Choice* q

Next, let us investigate the impact of the number of antennas at each DU (i.e., $q = |\mathcal{Q}_j|$) upon the detection performance of the proposed distributed MCRBK detection. Counterintuitively, given the fixed N antennas, we point out a smaller q (this corresponds to a larger number of DUs $r = N/q$) is preferred due to the faster convergence and the smaller error bound. Theoretically, let $q_B < q_A \leq K$ and assume the local channel matrices $\mathbf{H}^{q_B}_{\mathcal{Q}_j} \in \mathbb{C}^{q_B \times K} \subseteq \mathbf{H}^{q_A}_{\mathcal{Q}_j} \in \mathbb{C}^{q_A \times K}$, then this point can be easily validated by the following derivations

$$\begin{aligned} \eta^f_{q_B} &= \frac{\inf_{1\leq j\leq r, j\neq k,\ldots,l} \sigma^2_{\min}(\mathbf{H}^{q_B}_{\mathcal{Q}_j,:})}{\sup_{1\leq j\leq r, j\neq k,\ldots,l} \sigma^2_{\max}(\mathbf{H}^{q_B}_{\mathcal{Q}_j,:})} \\ &\overset{(f)}{\geq} \frac{\inf_{1\leq j\leq r, j\neq k,\ldots,l} \sigma^2_{\min}(\mathbf{H}^{q_A}_{\mathcal{Q}_j,:})}{\sup_{1\leq j\leq r, j\neq k,\ldots,l} \sigma^2_{\max}(\mathbf{H}^{q_A}_{\mathcal{Q}_j,:})} \\ &= \eta^f_{q_A}. \end{aligned} \tag{51}$$

Here, inequality (f) comes from the fact that [50]

$$\sigma_i(\mathbf{A}) \geq \sigma_i(\mathbf{B}) \geq \sigma_{i+k}(\mathbf{A}), \tag{52}$$

where $\mathbf{B}$ denotes an $(m-k) \times n$ submatrix of $\mathbf{A}$ and $\sigma_{\max} = \sigma_1 \geq \sigma_2 \geq \ldots \sigma_{\min}$. From it, we can have

$$\sigma_{q_B}(\mathbf{H}^{q_B}_{\mathcal{Q}_j,:}) = \sigma_{\min}(\mathbf{H}^{q_B}_{\mathcal{Q}_j,:}) \geq \sigma_{\min}(\mathbf{H}^{q_A}_{\mathcal{Q}_j,:}) = \sigma_{q_A}(\mathbf{H}^{q_A}_{\mathcal{Q}_j,:}) \tag{53}$$

and

$$\sigma_1(\mathbf{H}^{q_B}_{\mathcal{Q}_j,:}) = \sigma_{\max}(\mathbf{H}^{q_B}_{\mathcal{Q}_j,:}) \leq \sigma_{\max}(\mathbf{H}^{q_A}_{\mathcal{Q}_j,:}) = \sigma_1(\mathbf{H}^{q_A}_{\mathcal{Q}_j,:}), \tag{54}$$

which leads to

$$\begin{aligned} \varphi^f_{q_B} &= \inf_{1\leq j\leq r, j\neq k,\ldots,l} \sigma^2_{\min}(\mathbf{H}^{q_B}_{\mathcal{Q}_j,:}) \\ &\geq \inf_{1\leq j\leq r, j\neq k,\ldots,l} \sigma^2_{\min}(\mathbf{H}^{q_A}_{\mathcal{Q}_j,:}) \\ &= \varphi^f_{q_A}. \end{aligned} \tag{55}$$

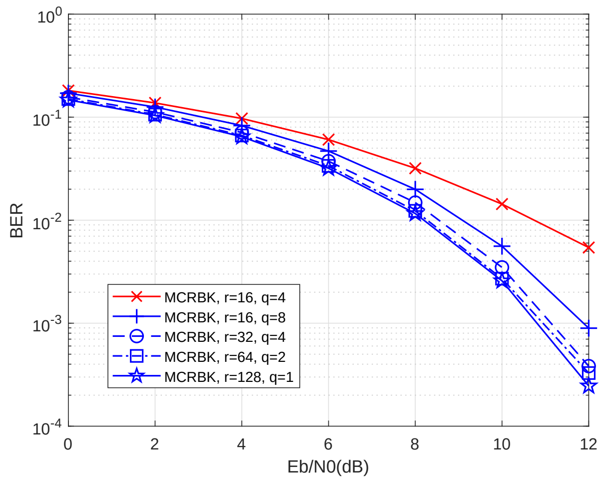

Fig. 4. BER Illustrations of MCRBK with different r and q in 128×32 massive MIMO systems using 16-QAM.

To sum up, we arrive at the following result.

Corollary 3. *Given the total number of antennas N, the proposed MCRBK algorithm achieves faster convergence with smaller error bound with the decrement of $q \leq K$ (or increment of r as $N = qr$).*

Here, for a better understanding, Fig. 4 is presented to illustrate the impacts of q and r on the bit error rate (BER) detection performance of the proposed distributed MCRBK detection in 128×32 uncoded massive MIMO systems with Rayleigh fading channels. However, given the fixed total number of antennas $N = qr$, although smaller q is desired, its corresponding requirement of the more DUs (i.e., larger r) also turns out to be an issue for massive MIMO systems in practice. Besides, the increment of DUs also naturally introduces more latency into the signal detection, which may be unbearable in practical implementation. Therefore, there is a latent trade-off with respect to the size of q in the proposed distributed MCRBK detection, and how to set it chiefly depends on the actual needs and the hardware setups in practice. On the other hand, undoubtedly, if the number of DUs r is fixed, then more number of antennas at each DU (i.e., larger q) is encouraged because of the extra received diversity gain introduced by the increased total number of antennas $N = qr$.

VI. Further Improvement by Dynamic Step-Sizes

Although the iteration performance of RBK has been greatly improved by the proposed MCRBK algorithm with $f = r - 1$, the error bound still does exist as shown in Theorem 4. This indeed hurts the estimation accuracy of MCRBK, which significantly degenerates the performance of the distributed MCRBK detection in massive MIMO systems.

To overcome this obstacle for further performance improvement, we introduce the step-size $\alpha^t > 0$ into the iterations of MCRBK. By dynamic choices of α^t at each iteration, we demonstrate that the two terms of convergence and error bound can merge together, which gradually diminishes along with the iterations. In a nutshell, the proposed MCRBK algorithm is able to return the approximation of $\mathbf{x}^*$ with arbitrary accuracy as t increases.

A. MCRBK Algorithm with Dynamic step-size

Here, for notational simplicity, we perform the following descriptions based on the formation in (48) with multi-step conditional sampling probability $\overline{p}^f_{\mathcal{Q}^t_j}$ in (39). In particular, the iteration of MCRBK with dynamic step-size $\alpha_t \in (0, 1)$ turns out to be [51]

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \alpha_t \mathbf{H}^\dagger_{\mathcal{Q}_j,:}(\mathbf{y}_{\mathcal{Q}_j} - \mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1}) \tag{56}$$

with $j = (t-1) \mod r + 1$. As for the selection of α_t, its optimization is carefully treated in section VI.B. Then, with respect to (56), we commence the analysis to specify its convergence at each iteration.

Theorem 5. *For massive MIMO system $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, given the multi-step conditional sampling probability $\overline{p}^f_{\mathcal{Q}}$ in (39) over a fixed partition $\mathcal{T}$, the distributed MCRBK detection is convergent by*

$$E[\|\mathbf{x}^t - \mathbf{x}^*\|^2] \leq \left[1 - (2\alpha_t - \alpha_t^2)\eta^f\right] \|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \alpha_t^2 \frac{q\sigma^2}{\varphi^f}. \tag{57}$$

Proof. To start with, regarding to $\mathbf{x}^t$, we define the auxiliary vector $\mathbf{z}^t$ as

$$\mathbf{z}^t = \mathbf{x}^{t-1} + \mathbf{H}^\dagger_{\mathcal{Q}_j,:}(\mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^* - \mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1}). \tag{58}$$

Then, by *Pythagorean Theorem*, the following relationship holds

$$\|\mathbf{x}^t - \mathbf{x}^*\|^2 = \|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 - \|\mathbf{z}^t - \mathbf{x}^{t-1}\|^2 + \|\mathbf{z}^t - \mathbf{x}^t\|^2. \tag{59}$$

With respect to the term $\|\mathbf{z}^t - \mathbf{x}^t\|^2$ in (59), we have

$$\begin{aligned}
\|\mathbf{z}^t - \mathbf{x}^t\|^2 &= \|\mathbf{z}^t - \mathbf{x}^{t-1} - \alpha_t \mathbf{H}^\dagger_{\mathcal{Q}_j,:}(\mathbf{y}_{\mathcal{Q}_j} - \mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1})\|^2 \\
&= \|\mathbf{z}^t - \mathbf{x}^{t-1} - \alpha_t \mathbf{H}^\dagger_{\mathcal{Q}_j,:}(\mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^* + \mathbf{n}_{\mathcal{Q}_j} - \mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1})\|^2 \\
&= \|\mathbf{z}^t - \mathbf{x}^{t-1} - \alpha_t \mathbf{H}^\dagger_{\mathcal{Q}_j,:}(\mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^* - \mathbf{H}_{\mathcal{Q}_j,:}\mathbf{x}^{t-1}) - \alpha_t \mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{n}_{\mathcal{Q}_j}\|^2 \\
&= \|\mathbf{z}^t - \mathbf{x}^{t-1} - \alpha_t(\mathbf{z}^t - \mathbf{x}^{t-1}) - \alpha_t \mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{n}_{\mathcal{Q}_j}\|^2 \\
&= \|(1 - \alpha_t)(\mathbf{z}^t - \mathbf{x}^{t-1}) - \alpha_t \mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{n}_{\mathcal{Q}_j}\|^2 \\
&= (1 - \alpha_t)^2 \|\mathbf{z}^t - \mathbf{x}^{t-1}\|^2 + \mathbf{S}^{t-1},
\end{aligned} \tag{60}$$

where

$$\mathbf{S}^{t-1} \triangleq -2(\alpha_t - \alpha_t^2)\langle \mathbf{z}^t - \mathbf{x}^{t-1}, \mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{n}_{\mathcal{Q}_j}\rangle + \alpha_t^2 \|\mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{n}_{\mathcal{Q}_j}\|^2. \tag{61}$$

After that, by substituting (60) into (59), we can obtain that

$$\begin{aligned}
\|\mathbf{x}^t - \mathbf{x}^*\|^2 &= \|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 - (2\alpha_t - \alpha_t^2)\|\mathbf{z}^t - \mathbf{x}^{t-1}\|^2 + \mathbf{S}^{t-1} \\
&= \left[1 - (2\alpha_t - \alpha_t^2)\frac{\|\mathbf{z}^t - \mathbf{x}^{t-1}\|^2}{\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2}\right] \|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \mathbf{S}^{t-1} \\
&= \left[1 - (2\alpha_t - \alpha_t^2)\frac{\|\mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{H}_{\mathcal{Q}_j,:}(\mathbf{x}^* - \mathbf{x}^{t-1})\|^2}{\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2}\right] \|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \mathbf{S}^{t-1} \\
&\overset{(g)}{\leq} \left[1 - (2\alpha_t - \alpha_t^2)\frac{\sigma^2_{\min}(\mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{H}_{\mathcal{Q}_j,:})\|\mathbf{x}^* - \mathbf{x}^{t-1}\|^2}{\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2}\right] \|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \mathbf{S}^{t-1} \\
&= \left[1 - (2\alpha_t - \alpha_t^2)\sigma^2_{\min}(\mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{H}_{\mathcal{Q}_j,:})\right] \|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \mathbf{S}^{t-1},
\end{aligned} \tag{62}$$

where the inequality $\|\mathbf{AB}\| \geq \sigma_{\min}(\mathbf{A})\|\mathbf{B}\|$ is applied in (g).

Regarding to (62), by taking the expectation on the both sides with $\overline{p}^f_{\mathcal{Q}^t_j}$ in (39), it follows that

$$\begin{aligned}
E[\|\mathbf{x}^t - \mathbf{x}^*\|^2] &\leq \Big[1 - (2\alpha_t - \alpha_t^2)E[\sigma^2_{\min}(\mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{H}_{\mathcal{Q}_j,:})]\Big]\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 \\
&\qquad + E[\mathbf{S}^{t-1}] \\
&\leq \left[1 - (2\alpha_t - \alpha_t^2)\eta^f\right]\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + E[\mathbf{S}^{t-1}] \\
&\overset{(h)}{=} \left[1 - (2\alpha_t - \alpha_t^2)\eta^f\right]\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \alpha_t^2 E[\|\mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{n}_{\mathcal{Q}_j}\|^2] \\
&\overset{(i)}{\leq} \left[1 - (2\alpha_t - \alpha_t^2)\eta^f\right]\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \alpha_t^2 E[\|\mathbf{H}^\dagger_{\mathcal{Q}_j,:}\|^2]E[\|\mathbf{n}_{\mathcal{Q}_j}\|^2] \\
&\overset{(j)}{=} \left[1 - (2\alpha_t - \alpha_t^2)\eta^f\right]\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \alpha_t^2 E\left[\frac{1}{\sigma^2_{\min}(\mathbf{A}_{\mathcal{Q}_j,:})}\right] q\sigma^2 \\
&\leq \left[1 - (2\alpha_t - \alpha_t^2)\eta^f\right]\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \alpha_t^2 \frac{q\sigma^2}{\varphi^f}. \qquad (63)
\end{aligned}$$

Here, $E[\mathbf{S}^{t-1}] = \alpha_t^2 E[\|\mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{n}_{\mathcal{Q}_j}\|^2]$ in (h) is attributed to the fact that $\mathbf{n}$ is independent of $\mathbf{H}$ with $\mathbf{0}$ mean so that

$$E\langle \mathbf{z}^t - \mathbf{x}^{t-1}, \mathbf{H}^\dagger_{\mathcal{Q}_j,:}\mathbf{n}_{\mathcal{Q}_j}\rangle = \mathbf{0}. \qquad (64)$$

Meanwhile, the inequality (i) comes from the fact $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$, and (j) is due to

$$\begin{aligned}
E[\|\mathbf{n}_{\mathcal{Q}_j}\|^2] &= E[(\mathbf{n}_{\mathcal{Q}_j} - E[\mathbf{n}_{\mathcal{Q}_j}])^H(\mathbf{n}_{\mathcal{Q}_j} - E[\mathbf{n}_{\mathcal{Q}_j}])] \\
&= \mathrm{Var}(\mathbf{n}_{\mathcal{Q}_j}) \\
&= q\sigma^2. \qquad (65)
\end{aligned}$$

□

According to Theorem 5, by adopting the step-size into the iterations, the proposed MCRBK algorithm for distributed massive MIMO detection is convergent to a specific error bound. Meanwhile, according to (45) and (46), the multi-step conditional sampling probability $\overline{p}^f_{\mathcal{Q}}$ with $f = r-1$ is still preferred for the best convergence (i.e., η^{r-1}) and error bound (i.e., φ^{r-1}) in (57). Therefore, we can readily arrive at the following result, where the proof is omitted due to simplicity.

Corollary 4. *As for the proposed MCRBK algorithm with step-size α_t, the multi-step $f = r-1$ achieves the fastest convergence and smallest error bound due to*

$$\eta^{r-1} \geq \eta^f \text{ and } \varphi^{r-1} \geq \varphi^f \qquad (66)$$

for $0 \leq f \leq r-1$.

As can be seen from (57), if only the convergence term $\left[1 - (2\alpha_t - \alpha_t^2)\eta^f\right]\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2$ is considered, the choice $\alpha_t = 1$ should be taken to achieve the best convergence. By contrast with it, because of the existence of the error term $\alpha_t^2 \frac{q\sigma^2}{\varphi^f}$, α_t is also encouraged to be as small as possible, thus resulting in an inherent trade-off between the convergence and the error bound. To this end, we try to systematically merge these two terms by optimizing the choice of α_t at each iteration.

B. Optimization with respect to Dynamic step-size α_t

According to (42) and (43), both η^f and φ^f are varying along with the iterations given the different previous sampling results $\mathcal{Q}^{t-1} = \mathcal{Q}_k, \ldots, \mathcal{Q}^{t-f} = \mathcal{Q}_l$, which is not friendly to the upcoming convergence analysis. Therefore, both the $\eta^{f=0}$ (i.e., η_{RBK}) and $\varphi^{f=0}$ (i.e., φ_{RBK}) are employed here as a compromise, which remain the same during the whole iterations because they are independent of any previous sampling results.

Theorem 6. *For massive MIMO system $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$, given the multi-step conditional sampling probability $\overline{p}^f_{\mathcal{Q}}$ in (39) over a fixed partition $\mathcal{T}$, with the choice of the dynamic step-size*

$$\alpha_t = \frac{\eta^{f=0}g(t-1)}{\eta^{f=0}g(t-1)+1} < 1, \qquad (67)$$

where

$$g(t) = g(t-1)(1 - \eta^{f=0}\alpha_t) \qquad (68)$$

and

$$g(0) = \|\mathbf{x}^0 - \mathbf{x}^*\|^2 \frac{\varphi^{f=0}}{q\sigma^2}, \qquad (69)$$

the distributed MCRBK detection ensures exact convergence to the least squares solution

$$E[\|\mathbf{x}^L - \mathbf{x}^*\|^2] \leq \prod_{t=1}^{L}(1 - \eta^{f=0}\alpha_t)\|\mathbf{x}^0 - \mathbf{x}^*\|^2 \qquad (70)$$

for any given $\mathbf{x}^0$.

Proof. To start with, based on $\eta^{f=0}$ and $\varphi^{f=0}$ in (45) and (46), the convergence in (57) can be relaxed to a looser one, i.e.,

$$E[\|\mathbf{x}^t - \mathbf{x}^*\|^2] \leq \left[1 - (2\alpha_t - \alpha_t^2)\eta^{f=0}\right]\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \alpha_t^2 \frac{q\sigma^2}{\varphi^{f=0}}. \qquad (71)$$

Subsequently, by iteratively applying the recursive relation in (71) under the full expectation, it follows that

$$\begin{aligned}
E[\|\mathbf{x}^L - \mathbf{x}^*\|^2] &\leq \prod_{j=1}^{L}\left[1 - (2\alpha_j - \alpha_j^2)\eta^{f=0}\right]\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \\
&\qquad \sum_{j=1}^{L}\alpha_j^2 \prod_{i=j+1}^{L}\left[1 - (2\alpha_i - \alpha_i^2)\eta^{f=0}\right]\frac{q\sigma^2}{\varphi^{f=0}} \qquad (72) \\
&= g(L)\frac{q\sigma^2}{\varphi^{f=0}} \qquad (73)
\end{aligned}$$

with the following definition

$$\begin{aligned}
g(t) &\triangleq \prod_{j=1}^{t}\left[1 - (2\alpha_j - \alpha_j^2)\eta^{f=0}\right]\|\mathbf{x}^0 - \mathbf{x}^*\|^2 \frac{\varphi^{f=0}}{q\sigma^2} + \\
&\qquad \sum_{j=1}^{t}\alpha_j^2 \prod_{i=j+1}^{t}\left[1 - (2\alpha_i - \alpha_i^2)\eta^{f=0}\right]. \qquad (74)
\end{aligned}$$

To be more specific, from (74), by induction we can find the following recurrence relationship

$$g(t) = \left[1 - (2\alpha_t - \alpha_t^2)\eta^{f=0}\right]g(t-1) + \alpha_t^2 \qquad (75)$$

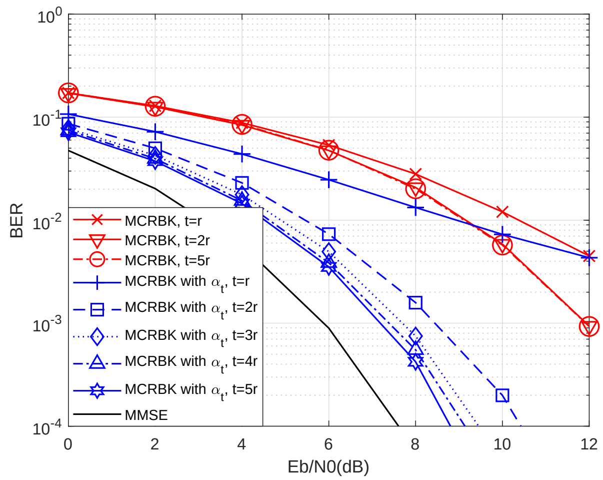

Fig. 5. BER Illustrations of MCRBK and MCRBK with dynamic α_t in 128×32 massive MIMO systems with $q = 8$ using 16-QAM.

for $t \geq 1$ with

$$g(0) = \|\mathbf{x}^0 - \mathbf{x}^*\|^2 \frac{\varphi^{f=0}}{q\sigma^2}. \tag{76}$$

Clearly, according to (75), the term $g(t-1)$ is independent of the step-size α_t. In this condition, by letting the derivative about $g(t)$ versus α_t be zero, the optimal value of α_t that makes the smallest $g(t)$ should be

$$\alpha_t = \frac{\eta^{f=0} g(t-1)}{\eta^{f=0} g(t-1) + 1} < 1. \tag{77}$$

Then, by substituting α_t in (77) back into $g(t)$ in (75), we have

$$g(t) = g(t-1)(1 - \eta^{f=0}\alpha_t). \tag{78}$$

To summarize, based on (73), (76) and (78), we can conclude that

$$E[\|\mathbf{x}^L - \mathbf{x}^*\|^2] \leq \prod_{t=1}^{L} (1 - \eta^{f=0}\alpha_t)\|\mathbf{x}^0 - \mathbf{x}^*\|^2, \tag{79}$$

completing the proof. □

According to (67) and (68) in Theorem 6, it is clear to see that $\alpha_t < 1$ decays gradually by

$$\alpha_t < \alpha_{t-1} \tag{80}$$

because of $g(t) < g(t-1)$. In this way, the proposed distributed MCRBK detection with dynamic step-size α_t will converge exactly to the linear detection solution $\mathbf{x}^*$. To give an example, in Fig. 5, we employ the BER in 128×32 uncoded massive MIMO systems with Rayleigh fading channels to illustrate the comparison between the proposed MCRBK and MCRBK with α_t. Intuitively, MCRBK stops converging at $t = 2r$ (i.e., the second iteration loop) but with a substantial performance gap compared to the MMSE detection. As a comparison, MCRBK with dynamic step α_t outperforms the basic MCRBK in terms of BER performance, which is in line with the derived results in Theorem 6. Additionally, from

---

Algorithm 1 Distributed MCRBK Detection with Dynamic step-size for Uplink Massive MIMO

Require: q, $\mathbf{H}_{\mathcal{Q}_j,:}$ and $\mathbf{y}_{\mathcal{Q}_j}$ for $1 \leq j \leq r$, $\mathbf{x}^0 = \mathbf{0}$, L
Ensure: detection solution $\widehat{\mathbf{x}}^L$
1: **for** $t = 1, \ldots, L$ **do**
2: update $\mathbf{x}^t$ by (56) with step-size α_t in (82)
3: **end for**
4: output $\widehat{\mathbf{x}}^L$ by rounding $\mathbf{x}^L$ based on constellation $\mathcal{X}^K$

Theorem 6, another merit of the proposed MCRBK algorithm with dynamic step-size is the global convergence [52], which comes from the fact that

$$\eta^{f=0}\alpha_t < 1 \tag{81}$$

for any $t \geq 1$. This means the convergence of the distributed MCRBK detection with dynamic step-size is always guaranteed, making it suitable to various massive MIMO scenarios of interest. In sharp contrast to it, the distributed detection schemes like FD-MMSE in [15], FD-CD in [14], decentralized Newton in [21] and so on only work if some specific conditions are satisfied, rendering them limited in practice. For example, the DBP-ADMM in [9] requires the condition $N \gg K$ to work while decentralized Newton in [21] converges only if $q \gg K$. More details about this point can be checked by the simulations results since some distributed schemes fail to work in some specific MIMO systems.

However, we have to point out that the derived α_t in (67) is suboptimal since it is based on the relaxed convergence in (71) rather than the original tight one in (57). Moreover, considering the computationally demanding nature of α_t in (67), here we provide a simple approximation to it by empirical methods, which is

$$\alpha_t = \frac{4K}{N}\left(1 - \frac{K}{N}\right)\frac{N/q + K}{N/q + K + t}. \tag{82}$$

However, the problem of finding easy to compute and efficient (in term of convergence) dynamic step adaptation policies is still quite open and it is left for future work.

Nevertheless, considerable potential still can be exploited at this point for achieving a better convergence, which is one of our works in future. Additionally, as shown in (70), the selection of the initial choice $\mathbf{x}^0$ is also important by enabling a faster convergence, and $\mathbf{x}^0 = \mathbf{0}$ is applied in this work. To summarize, the proposed distributed MCRBK detection with dynamic step-size for uplink massive MIMO is outlined in Algorithm 1. Additionally, to clearly illustrate the superiorities of the proposed distributed MCRBK detection with dynamic step-size, Table II is also given to show the related comparisons. Here, we claim that the index k in ADMM, CG, CD represents the number of processing iterations in each DU while k in DN denotes the number of iteration loops over the ring topology. To make a fair comparison with them, we also express the complexity of the proposed MCRBK by $\mathcal{O}(K^2 \cdot t) = \mathcal{O}(K^2 r \cdot k)$, where k indicates the number of iteration loops over the ring topology.

TABLE II
COMPARISONS AMONG DIFFERENT DISTRIBUTED DETECTION SCHEMES FOR MASSIVE MIMO.

	Computational Complexity	Data Bandwidth	Flexible number of antennas in each DU	Global Convergence	Decentralized Architecture
ADMM [9]	$\mathcal{O}(K^3r + K^2r \cdot k)$	$2K$	✓	✕	DBP
CG [10]	$\mathcal{O}(K^2N + K^2r \cdot k)$	$2K$	✓	✕	DBP
MMSE [15]	$\mathcal{O}(K^3 \cdot r)$	$2K$	✓	✕	FD
MMSE [11]	$\mathcal{O}(K^2N)$	$K^2 + K$	✓	✕	PD
CD [14]	$\mathcal{O}(KN \cdot k)$	$2K$	✓	✕	FD
DN [21]	$\mathcal{O}(K^2N + K^2r \cdot k)$	$4K$	✕	✕	Ring, Star
RLS [18]	$\mathcal{O}(K^2N)$	$K^2 + K$	✕ (limit to 1)	✓	Daisy-chain
SGD [19]	$\mathcal{O}(KN)$	K	✕ (limit to 1)	✕	Daisy-chain
ASGD [20]	$\mathcal{O}(KN)$	$2K$	✕ (limit to 1)	✕	Daisy-chain
SDK [41]	$\mathcal{O}(KN \cdot k)$	K	✕ (limit to 1)	✕	Daisy-chain
MCRBK (this work) with $f = r - 1$ and α_t	less than $\mathcal{O}(K^2 \cdot t)$ (i.e., $\mathcal{O}(K^2r \cdot k)$)	K	✓	✓	Daisy-chain Ring, Star, etc.

VII. APPLICATIONS TO OTHER MIMO NETWORKS

Based on the proposed distributed MCRBK detection with dynamic step-size, we now provide a deeper insight of its applications to other advanced MIMO networks.

A. XL-MIMO

Compared to traditional massive MIMO, much more spectral efficiency can be achieved by XL-MIMO [53]. However, as the antenna dimension goes extremely large, XL-MIMO generally works for the near-field propagation, where spatial non-stationarities occur as a result [54]–[57]. Specifically, because of the electromagnetic propagation in near-field, only a portion of the antenna array at BS can be reached by the user's transmitted signal, which leads to the visibility region (VR) for each UE [58]. It has been shown in [27] that non-stationarity in antenna arrays plays an important role upon the system performance and transceiver design. For these reasons, in the case of XL-MIMO, it is highly desired to partition the antenna array into small, separable and flexible subarrays to facilitate distributed computations within these processing units [28]–[33].

Here, we point out the proposed distributed MCRBL detection with dynamic step-size can be easily adopted to work for XL-MIMO in the way of daisy-chain. For example, as shown in Fig. 6, by ignoring the processing of subarrays out of VR, one just needs to perform the sequential processing in (56) over the active subarrays (i.e., DUs) within VR. More specifically, by Theorem 6, the more active subarrays in VR corresponds to the more DUs that can be taken into account for the distributed MCRBK detection with dynamic step-size, thus leading to a better detection performance. In other words, a flexible trade-off between performance and complexity can be achieved by the number of active subarrays in VR. Note that the merits of MCRBK like flexibility, scalability, low-complexity as well as global convergence are also maintained in the scenario of XL-MIMO, where further modifications of MCRBK for XL-MIMO will be one of our future work.

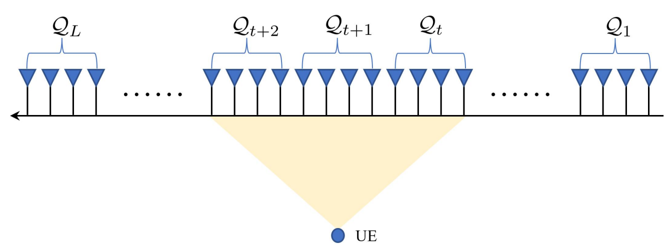

Fig. 6. Illustration of XL-MIMO with spatial non-stationary regions along the antenna array, where each user has a specific visibility region according to the channel conditions.

B. Cell-Free Massive MIMO Networks

Besides XL-MIMO, CF network is another promising extension of massive MIMO [59]–[61]. By removing the cell boundaries, it obeys the user-centric criterion for the data transmission regardless of UEs' positions to the BS or access point (AP) [62]. Typically, the same users can be cooperatively served by several APs in the geographical region, and APs are connected to a central processing unit (CPU) via the dedicated fronthaul links [63], [64]. In a word, CF massive MIMO can be viewed as an intersection of massive MIMO, distributed MIMO and cells without boundaries, where a certain level

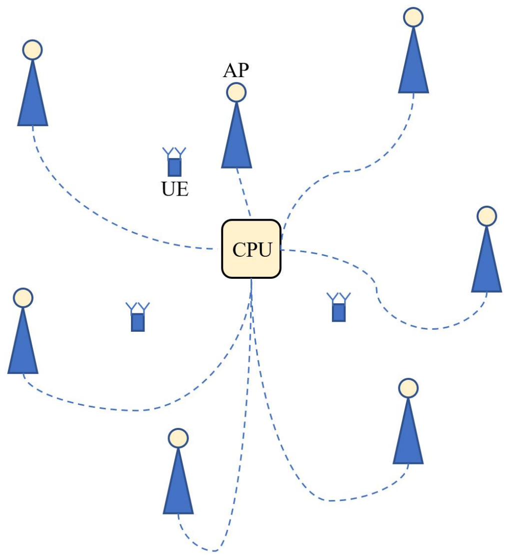

Fig. 7. Illustration of CF massive MIMO networks.

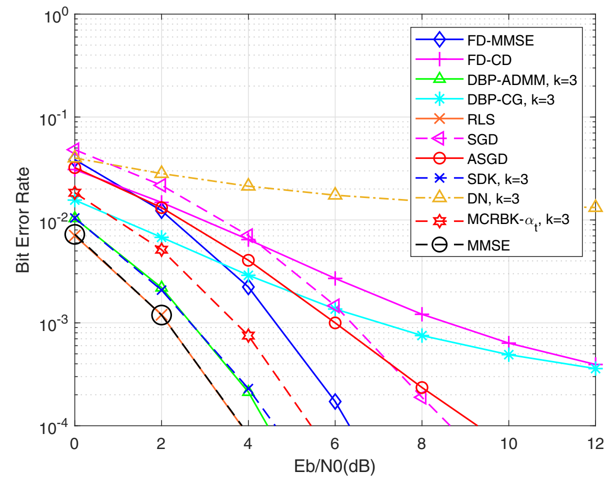

Fig. 8. BER comparison in 128×16 massive MIMO systems with $q = 4$ using 16-QAM.

of computational and transmission delays are allowed in CF massive MIMO networks.

As for the application of the proposed MCRBL scheme in CF networks, two feasible ways can be reasonably considered. On one hand, each AP can operate as a DU in the proposed distributed MCRBL detection with dynamic step-size, and sends its estimates of the transmitted signal to the CPU. After that, the CPU transfers the information to another AP for further processing and so on. On the other hand, the related APs can also send their local information (including channel estimation and received signal) to the CPU directly, where the proposed distributed MCRBL detection with dynamic step-size is carried out in a low computational complexity cost. Further investigation about this point will be given in future to achieve a better detection trade-off between performance and complexity.

VIII. Simulations

In this section, the performance and complexity of the proposed distributed MCRBK detection with dynamic step-size for uplink uncoded massive MIMO with Rayleigh fading channels are studied by simulations in full detail.

In Fig. 8, the BER performance comparison among various distributed detection schemes is shown with respect to 128×16 massive MIMO systems using 16-QAM. Specifically, besides the proposed distributed MCRBK detection with dynamic step-size (denoted by MCRBK-α_t), the FD-MMSE in [15], the FD-CD in [14], the DBP-ADMM in [9], the DBP-CG in [10], the RLS in [18], the SGD in [19], the ASGD in [20], the DN in [21] and the SDK in [41] are applied for the comparison. Here, the antenna numbers of each distributed unit in schemes of ADMM, CG, FD-MMSE, DN and MCRBK-α_t are set as $q = 4$. Different from them, RLS, SGD, ASGD and SDK schemes restrict the number of antennas in each DU as $q = 1$, which are deemed as a default setup in the following simulations. Moreover, the BER performance of centralized MMSE is also given to serve as a performance benchmark. In particular, as shown in Fig. 8, MCRBK-α_t with $k = 3$ (i.e., the third iteration loop) achieves a better detection performance than other distributed schemes except RLS. Note that as a distributed version of MMSE, RLS realizes the same detection performance as MMSE but in an iterative way. Nevertheless, from Table II, RLS requires much higher computational complexities and data bandwidth than MCRBK-α_t, which is often unaffordable in practice. Similarly, the PD-MMSE in [11] also achieves the exact MMSE detection performance at a high computational cost, which is omitted in the simulations.

In Fig. 9, the BER performance comparison among different distributed detection schemes is presented with respect to 128×32 massive MIMO systems using 16-QAM. Apart from the configurations in Fig. 8, the size $q = 8$ is applied in schemes of ADMM, CG, FD-MMSE, DN and MCRBK-α_t. In particular, compared to the case of 128×16, less received diversity gain can be exploited in the case of 128×32, so that the BER performance of all the detection schemes (including the centralized MMSE) degrade accordingly. Note that the detection schemes such as FD-MMSE, FD-CD and DN cannot work as usual in this case. For example, the operation of DN requires the conditions $q \gg K$, rendering it limited in most scenarios in practice. Meanwhile, it is clear to see that ADMM also suffers from the change of the network setup, which is inferior to the proposed MCRBK-α_t under the same iterations. On the other hand, the proposed MCRBK-α_t works normally due to its global convergence and achieves the best BER performance than all the other detection schemes except RLS. Moreover, with the increment of iterations, its detection performance gradually improves, which is in accordance with the results given in Theorem 6. As for a comparison, the BER performance of MCRBK without the aid of α_t (i.e., MCRBK with $f = r - 1$) is also added. However, similar to the results shown in Fig. 5, the BER performance of MCRBK becomes frozen after a few iterations. In addition, we point out that better BER performance can be attained by both MCRBK and

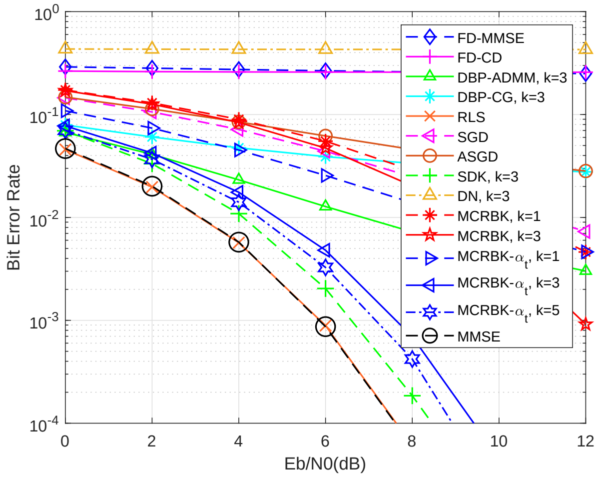

Fig. 9. BER comparison in 128×32 massive MIMO systems with $q = 8$ using 16-QAM.

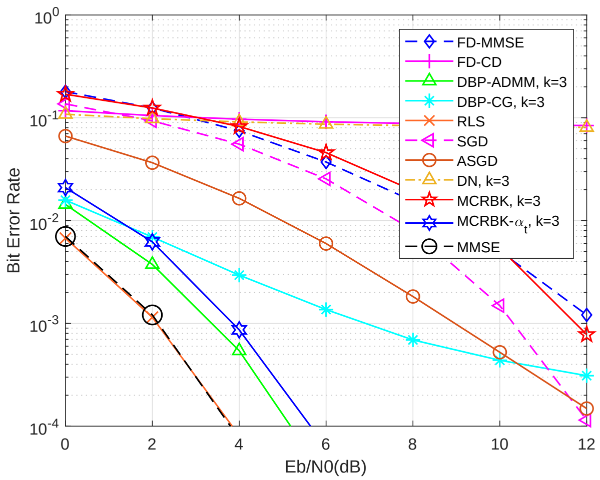

Fig. 10. BER comparison in 256×32 massive MIMO systems with $q = 8$ using 16-QAM.

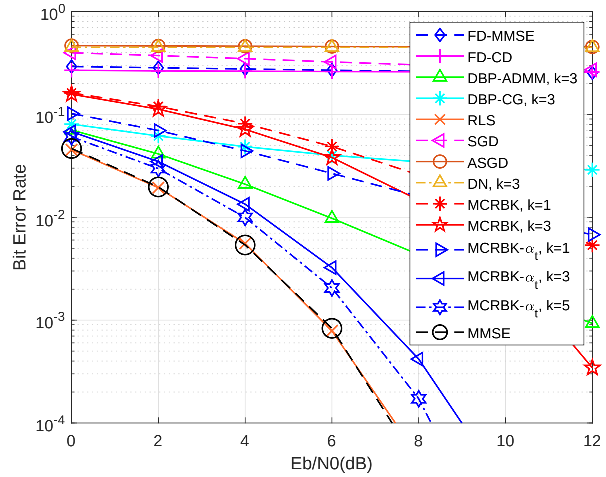

Fig. 11. BER comparison in 256×64 massive MIMO systems with $q = 8$ using 16-QAM.

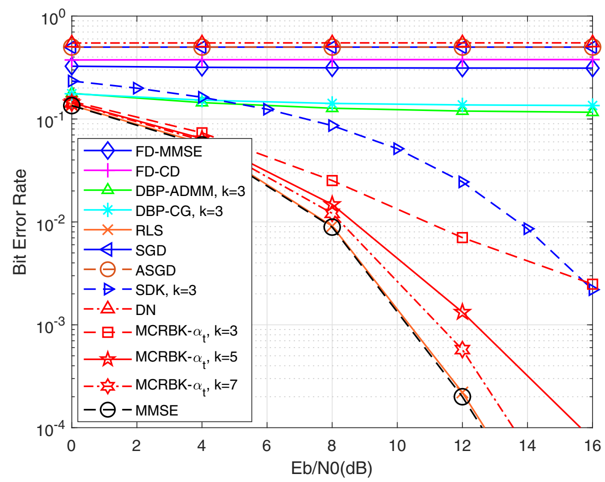

Fig. 12. BER comparison in 256×128 massive MIMO systems with $q = 8$ using 16-QAM.

MCRBK-α_t if a smaller q is employed, where more details about this point can be found in Fig. 4.

In Fig. 10, the BER performance comparison among different distributed detection schemes is illustrated with respect to 256×32 massive MIMO systems using 16-QAM, where $q = 8$ is applied in schemes of ADMM, CG, FD-MMSE, DN and MCRBK-α_t. Compared to the case of 128×32 in Fig. 9, this means more received diversity gain can be exploited, leading to an improved BER performance for all the detection schemes. However, as can be seen clearly, both FD-CD and DN schemes still fail to work in this case. On the other hand, as expected, MCRBK-α_t achieves a competitive BER performance compared to ADMM, which significantly outperforms FD-MMSE, CG, SGD, ASGD and MCRBK.

In Fig. 11, the BER performance comparison among different distributed detection schemes is shown with respect to 256×64 massive MIMO systems using 16-QAM, where $q = 8$ is applied in schemes of ADMM, CG, FD-MMSE, DN and MCRBK-α_t. Note that in this case, besides FD-MMSE, FD-CD and DN, both the SGD and ASGD detection schemes also fail to work. Clearly, because of its global convergence and rapid convergence, the proposed MCRBK-α_t is able to achieve the near-MMSE performance with a moderate choice of k (i.e., $t = rk$), which greatly outperform other distributed schemes except RLS. To further figure out the impact of the system dimension upon the detection performance, we have added the Fig. 12 to illustrate BER performance comparison regarding 256×128 massive MIMO systems using 16-QAM with $q = 8$. Interestingly, with the antenna ratio $N/K = 2$, most of the distributed detection schemes including ADMM and CG cannot work well. By contrast, MCRBK-α_t still works as usual, where further BER performance gain can be achieved with the increment of the iterations.

In Fig. 13, the computational complexity comparison among various distributed detection schemes is presented with respect to $256 \times K$ massive MIMO systems using 16-QAM. Here,

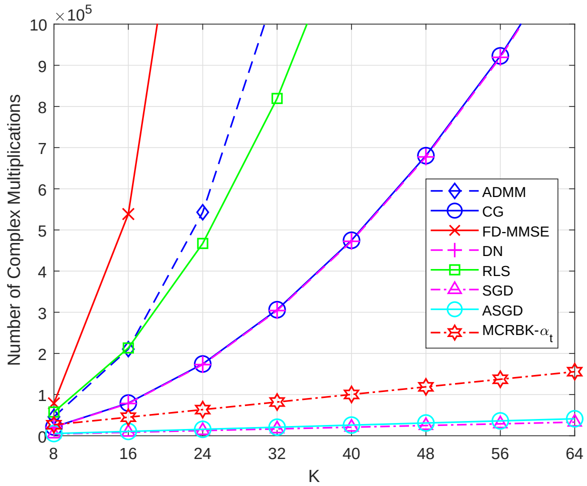

Fig. 13. Complexity comparison in terms of the number of complex multiplications for $256 \times K$ massive MIMO systems with $q = 8$ and $k = 1$.

$q = 8$ is applied in schemes of ADMM, CG, FD-MMSE, DN and MCRBK-α_t while the computational complexity is evaluated by the number of complex multiplications in a single iteration loop (i.e., $k = 1$ or $t = r$) for all distributed detection schemes. Specifically, the schemes of FD-MMSE, ADMM and RLS are quite computationally expensive, which turns out to be unaffordable in practice. CG and DN have the similar complexity cost, which is still much higher than MCRBK-α_t, SGD and ASGD. However, although SGD and ASGD incur less complexity cost than MCRBK-α_t, they only serve to work with the requirement of $q = 1$. Moreover, even under $q = 1$, their performance becomes terrible when the antenna ratio N/K goes up (see Fig. 9). In fact, as shown in Fig. 11 and Fig. 12, both of them even fail to work, rendering them limited in the practice implementation. Different from them, the proposed MCRBK-α_t has a slightly higher complexity cost (which grows in a mild way), but enjoys the fast convergence (i.e., better BER performance), the flexible number of antennas in each distributed unit (i.e., $q \geq 1$), and the global convergence (always work in different network configurations), thus making it as a perspective solution in the distributed detection in massive MIMO systems.

IX. Conclusion

In this paper, the classic Kaczmarz method was studied to facilitate the distributed detection in massive MIMO systems. Based on the derived convergence results of RBK with the general sampling probability, CRBK algorithm was proposed by incorporating the conditional sampling into iteration update, which achieves faster convergence and smaller error bound than RBK. After that, CRBK was further extended by taking more sampling results in the previous iterations into account, and this naturally leads to the proposed MCRBK algorithm. Most interestingly, we pointed out that MCRBK can be easily adapted to massive MIMO as a flexible, scalable, low-complexity distributed detection scheme. Moreover, to overcome its convergence obstacle induced by the inherent error bound, a novel dynamic step-size mechanism was further presented for the distributed MCRBK detection. By demonstration, we show that the distributed MCRBK detection with the optimized dynamic step-size converges in an exponential and global way to the results of the traditional linear detection schemes, making it promising in various scenarios of massive MIMO systems.

References

[1] I. Tomkos, D. Klonidis, E. Pikasis, and S. Theodoridis, “Toward the 6G network era: Opportunities and challenges,” *IT Professional*, vol. 22, no. 1, pp. 34–38, 2020.

[2] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, “A speculative study on 6G,” *IEEE Wireless Communications*, vol. 27, no. 4, pp. 118–125, 2020.

[3] W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[5] M. A. Albreem, A. Alhabbash, A. M. Abu-Hudrouss, and T. A. Almohamad, “Data detection in decentralized and distributed massive MIMO networks,” *Computer Communications*, vol. 189, pp. 79–99, 2022.

[6] M. A. Albreem, M. Juntti, and S. Shahabuddin, “Massive MIMO detection techniques: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019.

[7] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, “Prospective multiple antenna technologies for beyond 5G,” *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1637–1660, 2020.

[8] J. R. Sánchez, J. Vidal Alegría, and F. Rusek, “Decentralized massive MIMO systems: Is there anything to be discussed?” in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 787–791.

[9] K. Li, R. R. Sharan, Y. Chen, T. Goldstein, J. R. Cavallaro, and C. Studer, “Decentralized baseband processing for massive MU-MIMO systems,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, no. 4, pp. 491–507, 2017.

[10] K. Li, Y. Chen, R. Sharan, T. Goldstein, J. R. Cavallaro, and C. Studer, “Decentralized data detection for massive MU-MIMO on a Xeon Phi cluster,” in *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 468–472.

[11] C. Jeon, K. Li, J. R. Cavallaro, and C. Studer, “Decentralized equalization with feedforward architectures for massive MU-MIMO,” *IEEE Transactions on Signal Processing*, vol. 67, no. 17, pp. 4418–4432, 2019.

[12] ——, “On the achievable rates of decentralized equalization in massive MU-MIMO systems,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 1102–1106.

[13] K. Li, C. Jeon, J. R. Cavallaro, and C. Studer, “Decentralized equalization for massive MU-MIMO on FPGA,” in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, 2017, pp. 1532–1536.

[14] K. Li, O. Castañeda, C. Jeon, J. R. Cavallaro, and C. Studer, “Decentralized coordinate-descent data detection and precoding for massive MU-MIMO,” in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2019, pp. 1–5.

[15] K. Li, J. McNaney, C. Tarver, O. Castañeda, C. Jeon, J. R. Cavallaro, and C. Studer, “Design trade-offs for decentralized baseband processing in massive MU-MIMO systems,” in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 906–912.

[16] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, “Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 1, pp. 9–39, 2023.

[17] F. Jiande, X. Weixin, and L. Zongxiang, “A low complexity distributed multitarget detection and tracking algorithm,” *Chinese Journal of Electronics*, vol. 32, no. 3, pp. 429–437, 2023.

[18] J. R. Sanchez, F. Rusek, M. Sarajlic, O. Edfors, and L. Liu, “Fully decentralized massive MIMO detection based on recursive methods,” in *2018 IEEE International Workshop on Signal Processing Systems (SiPS)*, 2018, pp. 53–58.

[19] J. Rodrguez Sánchez, F. Rusek, O. Edfors, M. Sarajlić, and L. Liu, "Decentralized massive MIMO processing exploring daisy-chain architecture and recursive algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 687–700, 2020.

[20] Q. Liu, H. Liu, Y. Yan, and P. Wu, "A distributed detection algorithm for uplink massive MIMO systems," in *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, 2019, pp. 213–217.

[21] A. Kulkarni, M. A. Ouameur, and D. Massicotte, "Hardware topologies for decentralized large-scale MIMO detection using Newton method," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 9, pp. 3732–3745, 2021.

[22] Z. Zhang, Y. Dong, K. Long, X. Wang, and X. Dai, "Decentralized baseband processing with Gaussian message passing detection for uplink massive MU-MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2152–2157, 2022.

[23] Z. Zhang, H. Li, Y. Dong, X. Wang, and X. Dai, "Decentralized signal detection via expectation propagation algorithm for uplink massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 233–11 240, 2020.

[24] A. Chawla, R. K. Singh, A. Patel, A. K. Jagannatham, and L. Hanzo, "Distributed detection for centralized and decentralized millimeter wave massive MIMO sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7665–7680, 2021.

[25] H. Li, Y. Dong, C. Gong, X. Wang, and X. Dai, "Decentralized groupwise expectation propagation detector for uplink massive MU-MIMO systems," *IEEE Internet of Things Journal*, vol. 10, no. 6, pp. 5393–5405, 2023.

[26] X. Zhao, X. Guan, M. Li, and Q. Shi, "Decentralized linear MMSE equalizer under colored noise for massive MIMO systems," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 01–06.

[27] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 74–80, 2020.

[28] A. Amiri, S. Rezaie, C. N. Manchón, and E. de Carvalho, "Distributed receiver processing for extra-large MIMO arrays: A message passing approach," *IEEE Transactions on Wireless Communications*, vol. 21, no. 4, pp. 2654–2667, 2022.

[29] H. Wang, A. Kosasih, C.-K. Wen, S. Jin, and W. Hardjawana, "Expectation propagation detector for extra-large scale massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2036–2051, 2020.

[30] X. Yang, F. Cao, M. Matthaiou, and S. Jin, "On the uplink transmission of extra-large scale massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 229–15 243, 2020.

[31] V. Croisfelt, A. Amiri, T. Abrao, E. de Carvalho, and P. Popovski, "Accelerated randomized methods for receiver design in extra-large scale MIMO arrays," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6788–6799, 2021.

[32] A. Amiri, C. N. Manchn, and E. de Carvalho, "Uncoordinated and decentralized processing in extra-large MIMO arrays," *IEEE Wireless Communications Letters*, vol. 11, no. 1, pp. 81–85, 2022.

[33] H. He, H. Wang, X. Yu, J. Zhang, S. H. Song, and K. B. Letaief, "Distributed expectation propagation detection for cell-free massive MIMO," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 01–06.

[34] Z. H. Shaik, E. Björnson, and E. G. Larsson, "Distributed computation of a posteriori bit likelihood ratios in cell-free massive MIMO," in *2021 29th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 935–939.

[35] N. Li and P. Fan, "Distributed cell-free massive MIMO versus cellular massive MIMO under UE hardware impairments," *Chinese Journal of Electronics*, vol. 33, no. 5, pp. 1274–1285, 2024.

[36] S. M. Kaczmarz, "Angenäherte auösung von systemen linearer gleichungen," *Bulletin International de l'Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles. Série A, Sciences Mathématiques*, vol. 35, pp. 355–357, 1937.

[37] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *Journal of Theoretical Biology*, vol. 29, no. 3, pp. 471–481, 1970.

[38] G. Herman, *Fundamentals of computerized tomography: Image reconstruction from projection (2nd ed.)*. Springer, 2009.

[39] M. Sarajlić, F. Rusek, J. Rodríguez Sánchez, L. Liu, and O. Edfors, "Fully decentralized approximate Zero-Forcing precoding for massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 773–776, 2019.

[40] M. N. Boroujerdi, S. Haghighatshoar, and G. Caire, "Low-complexity statistically robust precoder/detector computation for massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6516–6530, 2018.

[41] V. Croisfelt, T. Abrão, A. Amiri, E. de Carvalho, and P. Popovski, "Decentralized design of fast iterative receivers for massive MIMO with spatial non-stationarities," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 1242–1249.

[42] T. Strohmer and R. Vershynin, "A randomized Kaczmarz algorithm with exponential convergence," *Journal of Fourier Analysis and Applications*, vol. 15, no. 2, pp. 262–278, 2009.

[43] D. Needell and J. A. Tropp, "Paved with good intentions: Analysis of a randomized block Kaczmarz method," *Linear Algebra and its Applications*, vol. 441, pp. 199–221, Jan. 2014.

[44] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

[45] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[46] E. K. P. Chong and S. H. Żak, *An Introduction to Optimization (4th ed.)*. John Wiley & Sons, Inc., 2013.

[47] M. Hanke and W. Niethammer, "On the acceleration of Kaczmarz's method for inconsistent linear systems," *Linear Algebra and its Applications*, vol. 130, pp. 83–98, 1990.

[48] D. Needell, "Randomized kaczmarz solver for noisy linear systems," *BIT Numerical Mathematics*, vol. 50, pp. 395–403, Feb. 2010.

[49] D. Wubben, R. Bohnke, V. Kuhn, and K. D. Kammeyer, "Near-maximum-likelihood detection of MIMO systems using MMSE-based lattice reduction," in *Proc. IEEE Int. Conf. Commun.(ICC'04)*, Paris, France, Jun. 2004, pp. 798–802.

[50] G. H. Golub and C. F. V. Loan, *Matrix Computations*. United States: Johns Hopkins University Press, 4th ed., 2013.

[51] A. Björck, *Numerical Methods in Matrix Computations*. Springer, 2015.

[52] Z. Wang, R. M. Gower, Y. Xia, L. He, and Y. Huang, "Randomized iterative methods for low-complexity large-scale MIMO detection," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2934–2949, 2022.

[53] Z. Wang, J. Zhang, H. Du, W. E. I. Sha, B. Ai, D. Niyato, and M. Debbah, "Extremely large-scale MIMO: Fundamentals, challenges, solutions, and future directions," *IEEE Wireless Communications*, pp. 1–9, 2023.

[54] X. Li, S. Zhou, E. Björnson, and J. Wang, "Capacity analysis for spatially non-wide sense stationary uplink massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 7044–7056, 2015.

[55] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2663–2677, 2022.

[56] X. Guo, Y. Chen, and Y. Wang, "Compressed channel estimation for near-field XL-MIMO using triple parametric decomposition," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2023.

[57] M. Cui, L. Dai, Z. Wang, S. Zhou, and N. Ge, "Near-field rainbow: Wideband beam training for XL-MIMO," *IEEE Transactions on Wireless Communications*, vol. 22, no. 6, pp. 3899–3912, 2023.

[58] A. Ali, E. D. Carvalho, and R. W. Heath, "Linear receivers in non-stationary massive MIMO channels with visibility regions," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 885–888, 2019.

[59] H. Yang and T. L. Marzetta, "Capacity performance of multicell large-scale antenna systems," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2013, pp. 668–675.

[60] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: Uniformly great service for everyone," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 201–205.

[61] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "User-centric cell-free massive MIMO networks: A survey of opportunities, challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 611–652, 2022.

[62] S. Buzzi and C. D' Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 706–709, 2017.

[63] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Communications Letters*, vol. 21, no. 8, pp. 1871–1874, 2017.

[64] M. Attarifar, A. Abbasfar, and A. Lozano, "Subset MMSE receivers for cell-free networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4183–4194, 2020.

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/TSP.2025.3558930

Zheng Wang (Senior Member, IEEE) received the B.S. degree in electronic and information engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2009, and the M.S. degree in communications from University of Manchester, Manchester, U.K., in 2010. He received the Ph.D degree in communication engineering from Imperial College London, UK, in 2015.

Since 2021, he has been an Associate Professor in the School of Information and Engineering, Southeast University, Nanjing, China. From 2015 to 2016 he served as a Research Associate at Imperial College London, UK. From 2016 to 2017 he was an senior engineer with Radio Access Network R&D division, Huawei Technologies Co.. From 2017 to 2020 he was an Associate Professor at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. At 2023, he has been recognized as a Highly Cited Chinese Researcher by Elsevier for exceptional research performance in the field of information and communications engineering. He received the Spark Award from Huawei Technologies Co. at 2023 for the research works towards Spark Challenge about Efficient Baseband Matrix Processing. He is also the Youth Editor of the Chinese Journal of Electronics (CJE). His current research interests include massive MIMO systems, machine learning and data analytic over wireless networks, and lattice theory for wireless communications.

Cunhua Pan is a full professor in Southeast University. His research interests mainly include reconfigurable intelligent surfaces (RIS), AI for Wireless, near field communications and sensing, and integrated sensing and communications. He has published over 200 IEEE journal papers. His papers got over 18500 Google Scholar citations with H-index of 69. He is Clarivate Highly Cited researcher. He is/was an Editor of IEEE Transaction on Communications, IEEE Transactions on Vehicular Technology, IEEE Wireless Communication Letters, and IEEE Communications Letters. He serves as the (leading) guest editors for IEEE Journal on Selected Areas in Communications, IEEE Journal of Selected Topics in Signal Processing, IEEE Internet of Things, IEEE Vehicular Technology Magazine, IEEE Internet of Things Magazine, IEEE Transactions on Green Communications and Networking. He received the IEEE ComSoc Leonard G. Abraham Prize in 2022, IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2022, IEEE ComSoc Fred W. Ellersick Prize in 2024, IEEE ComSoc CTTC Early Achievement Award in 2024, IEEE ComSoc SPCC Early Achievement Award in 2024, and IEEE WCSP 2022 best paper award. He supervised one Phd Student to win the IEEE Signal Processing Society Best Phd Dissertation Award.

Yongming Huang (Fellow, IEEE) received the B.S. and M.S. degrees from Nanjing University, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2007.

Since March 2007, he has been a Faculty with the School of Information Science and Engineering, Southeast University, where he is currently a Full Professor. During 2008-2009, he visited the Signal Processing Lab, Royal Institute of Technology, Stockholm, Sweden. He has authored or coauthored more than 200 peer-reviewed papers, and holds more than 80 invention patents. His research interests include intelligent 5G/6G mobile communications and millimeter wave wireless communications. He submitted around 20 technical contributions to IEEE standards, and was awarded a certificate of appreciation for outstanding contribution to the development of IEEE standard 802.11aj. He was an Associate Editor for the IEEE Transactions on Signal Processing and a Guest Editor of the IEEE Journal Selected Areas in Communications. He is currently an Editor-at-Large of the IEEE Open Journal of the Communications Society and an Associate Editor for the IEEE Wireless Communications Letters.

Shi Jin (Fellow, IEEE) received the B.S. degree in communications engineering from Guilin University of Electronic Technology, Guilin, China, in 1996, the M.S. degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2003, and the Ph.D. degree in information and communications engineering from Southeast University, Nanjing, in 2007. From June 2007 to October 2009, he was a Research Fellow with the Adastral Park Research Campus, University College London, London, U.K. He is currently with the Faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests include wireless communications, random matrix theory, and information theory. He and his co-authors received the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communication theory, the IEEE Vehicular Technology Society 2023 Jack Neubauer Memorial Award, the 2022 Best Paper Award, and the 2010 Young Author Best Paper Award by the IEEE Signal Processing Society. He is serving as an Area Editor for IEEE Transactions on Communications and IET Electronics Letters. He was an Associate Editor of IEEE Transactions on Wireless Communications, IEEE Communications Letters, and IET Communications.

G iuseppe Caire (S '92 – M '94 – SM '03 – F '05) was born in Torino in 1965. He received a B.Sc. in Electrical Engineering from Politecnico di Torino in 1990, an M.Sc. in Electrical Engineering from Princeton University in 1992, and a Ph.D. from Politecnico di Torino in 1994. He has been a post-doctoral research fellow with the European Space Agency (ESTEC, Noordwijk, The Netherlands) in 1994-1995, Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, a Professor of Electrical Engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles, and he is currently an Alexander von Humboldt Professor with the Faculty of Electrical Engineering and Computer Science at the Technical University of Berlin, Germany.

He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, the IEEE Communications Society and Information Theory Society Joint Paper Award in 2004 and in 2011, the Okawa Research Award in 2006, the Alexander von Humboldt Professorship in 2014, the Vodafone Innovation Prize in 2015, an ERC Advanced Grant in 2018, the Leonard G. Abraham Prize for best IEEE JSAC paper in 2019, the IEEE Communications Society Edwin Howard Armstrong Achievement Award in 2020, the 2021 Leibniz Prize of the German National Science Foundation (DFG), and the CTTC Technical Achievement Award of the IEEE Communications Society in 2023. Giuseppe Caire is a Fellow of IEEE since 2005. He has served in the Board of Governors of the IEEE Information Theory Society from 2004 to 2007, and as officer from 2008 to 2013. He was President of the IEEE Information Theory Society in 2011. His main research interests are in the field of communications theory, information theory, channel and source coding with particular focus on wireless communications.