Efficient Distributed Randomized Iterative Detection for Decentralized XL-MIMO Systems

Zheng Wang, *Senior Member, IEEE*, Chunguo Li, *Senior Member, IEEE*, Yongming Huang, *Fellow, IEEE*, Shi Jin, *Fellow, IEEE*, and Giuseppe Caire, *Fellow, IEEE*

***Abstract*—With the rapid scaling up of system dimensions, distributed signal detection based on decentralized baseband architectures has become an important problem in extremely large-scale MIMO (XL-MIMO). In this paper, to reduce the complexity cost of distributed detections for XL-MIMO, the efficient distributed randomized iterative detection (EDRID) algorithm is proposed. First of all, by removing the computationally expensive calculation of matrix pseudoinverse, the randomized iteration in EDRID is still able to exponentially converge into a close range around the target least squares (LS) solution but with significant complexity reduction. Then, to overcome the convergence bias of EDRID, we adopt a dynamic step-size strategy into EDRID to ensure its exact convergence to the LS solution. Theoretical analysis demonstrates that the proposed EDRID with dynamic step-size not only returns the LS solution with arbitrary accuracy for sufficient iterations, but also features the global convergence for different practical scenarios. Finally, to obtain further improvements in both convergence and efficiency, the technique of conditional sampling is also introduced to EDRID, making it a low-complexity, flexible, scalable detection choice for XL-MIMO systems.**

***Index Terms*—Distributed signal detection, decentralized baseband detection, XL-MIMO detection, low-complexity detection, distributed MIMO systems.**

I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) has established itself as a foundational technology for beyond fifth-generation (B5G) and sixth-generation (6G) wireless networks, owing to its significant potential to enhance spectral efficiency, improve energy efficiency, and increase transmission reliability [1], [2], [3], [4], [5]. However, as conventional linear detection schemes aggregate data from base station (BS) antennas to central processing unit (CPU), scaling to hundreds or even thousands of antennas imposes pressing challenges upon the uplink signal detection due to the increased complexity and bandwidth demands. This ignites the intensive interests in distributed signal detection under decentralized baseband processing architectures [6], [7], [8], [9]. In such architectures, signal processing tasks are partitioned and executed locally across multiple distributed units (DUs), thus achieving lower complexity, latency and interconnection data bandwidth. Therefore, many distributed signal detection schemes tailored for decentralized baseband architectures have been given for the uplink MIMO systems [10], [11], [12], [13], [14], [15], [16], [17], [18].

In particular, the decentralized baseband processing (DBP) architecture is given in [10] to facilitate the distributed signal detection in massive MIMO systems. By partitioning the antenna array of BS into DUs, parts of channel estimation and signal detection can be performed locally to extract the consensus information, and then a central fusion node aggregates all the consensus data to generate the final detection result. From it, methods like alternating direction method of multipliers (ADMM) and conjugate gradient (CG) have been employed for the distributed detection [11]. To get rid of the consensus information exchange between DUs, partially decentralized (PD) and fully decentralized (FD) feedforward architectures are presented respectively [12], [13], [14]. FD still utilizes the local detection in DU, and each DU releases its detection results to a feedforward fusion for the further decision. Apart from the local detection in DU, PD only provides the local preprocessing results and leaves the signal detection to the central unit. Based on them, several methods have also been proposed to alleviate the baseband obstacles in both computational complexity and interconnection data bandwidth [15], [16].

In [19], the conventional recursive algorithms are adopted into distributed signal detection based on the daisy-chain decentralized architecture. In this architecture, each DU sequentially processes the detection estimates received from its predecessor using its local observations, and then forwards the updated estimates unidirectionally to the subsequent DU [20], [21], [22]. This naturally aligns with the iterative nature of recursive methods, enabling a pipelined detection mechanism that features low interconnection bandwidth requirements and excellent scalability. Besides the daisy-chain topology, other

Received 19 July 2025; revised 17 December 2025 and 2 March 2026; accepted 10 April 2026. Date of publication 21 April 2026; date of current version 29 May 2026. This work was supported in part by the National Science and Technology Major Projects of China under Grant 2025ZD1301800, in part by the National Natural Science Foundation of China under Grant 62371124, Grant 62225107, Grant 61720106003, and Grant U25B2007, and in part by the National Key R&D Program of China under Grant 2024YFC3807900. The associate editor coordinating the review of this article and approving it for publication was Yue Wang. *(Corresponding authors: Zheng Wang; Yongming Huang.)*

Zheng Wang, Chunguo Li, Yongming Huang, and Shi Jin are with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: wznuaa@gmail.com; huangym@seu.edu.cn).

Giuseppe Caire is with the Electrical Engineering and Computer Science Department, Technische Universität Berlin, 10623 Berlin, Germany (e-mail: caire@tu-berlin.de).

Digital Object Identifier 10.1109/TSP.2026.3686191

decentralized topologies such as ring, star and tree have also been considered for the distributed detection. For instance, in [23], a decentralized Newton (DN) method is developed based on ring and star topologies, aiming for low-complexity distributed detection. However, to fully exploit the benefit of channel hardening, it only works when the number of antennas per DU is significantly larger than that of the user side, rendering its application rather limited in practice. In addition, message passing (MP) algorithms have also been explored within decentralized frameworks, where the log-likelihood ratio (LLR) information are computed as the priori information among DUs [24], [25]. Other distributed detection solutions for massive MIMO can be found in [26], [27], [28], [29], [30].

In [31], the strategy of randomized iteration is introduced into distributed detections, and a perspective distributed detection scheme named as multi-step conditional randomized block Kaczmarz (MCRBK) is proposed. It not only overcomes the convergence issue of the traditional Kaczmarz method, but also enjoys the great flexibility and scalability to suit various decentralized topologies and distributed systems. Nevertheless, although its extensions to extremely large-scale MIMO (XL-MIMO) and cell-free (CF) MIMO have been discussed in [31], further complexity reduction of it is still heavily desired to alleviate the complexity burden, especially as the system dimension of MIMO rapidly scales up in 6G [32], [33], [34], [35], [36], [37], [38], [39], [40], [41]. For this reason, in this paper, we upgrade the randomized iteration for distributed detection to a much more efficient version, which leads to the proposed efficient distributed randomized iterative detection (EDRID) algorithm. In particular, compared to the computational complexity $2q^2K + q^3 + 2qK$ (q is the number of antennas in each DU and K represents the total number of antennas at the user side) of each iteration (index by t) in MCRBK, by removing the computation of matrix pseudoinverse from randomized iterations, the computational complexity of each iteration (index by t) in EDRID is reduced to $2qK$, leading to huge complexity reduction. Meanwhile, we also demonstrate that the proposed EDRID algorithm still enjoys the exponential and global convergence to the least squares (LS) solution, making it a perspective distributed detection solution for 6G. To summarize, we advance the researches on the distributed detection for decentralized XL-MIMO systems on the following fronts.

- First of all, EDRID algorithm is presented to remove the computationally expensive matrix pseudoinverse from randomized iterations. With the derived step-size, we demonstrate that EDRID converges exponentially to a close range around the LS solution $\mathbf{x}^*$. Benefiting from it, the complexity order of the distributed randomized iterations is reduced from $\mathcal{O}(Kq^2t)$ to $\mathcal{O}(Kqt)$, which significantly alleviates the complexity hurdle of distributed MIMO systems.
- Secondly, to guarantee the exact exponential convergence to the LS solution $\mathbf{x}^*$, adaptive step-size is employed in EDRID for the further performance improvement. According to theoretical analysis, we demonstrate that the target solution $\mathbf{x}^*$ can be returned by EDRID in arbitrary accuracy with the increase of iterations. Thanks to it, remarkable complexity reduction can be realized by the proposed EDRID algorithm but still with competitive detection performance.
- Thirdly, the usage of conditional sampling is leveraged into EDRID, which is further enhanced via multi-step conditional sampling. By taking the previous multiple sampling results into account, the randomized iteration of EDRID becomes gradually deterministic. This makes EDRID applicable to different topologies of decentralized MIMO systems with enhanced iteration convergence and efficiency.

To sum up, in Table I we compare the new contributions of EDRID to the related literature, where N represents the antenna number at BS. For a fair comparison, in Table I the index k of ADMM, CG, CD indicates the number of processing iterations at each DU while k in DN, MCRBK and EDRID represents the number of iteration loops of the ring topology, namely, $t = rk$. Overall, benefiting from the low-complexity, flexility and scalability, the proposed EDRID algorithm turns out to be a perspective solution for various decentralized MIMO systems including XL-MIMO, cell-free MIMO and so on.

The organization of this paper is as follows. Section II briefly introduces the system model of the uplink MIMO detection and reviews the state of the art of the distributed detection methods. In Section III, EDRID algorithm is proposed for significant complexity reduction of the distributed randomized iterations, followed by the theoretical analysis of its convergence. To further ensure the exact convergence to the LS solution, adaptive step-size is introduced into EDRID in Section IV. In Section V, both the iteration convergence and efficiency of EDRID are improved by the usage of multi-step conditional sampling, and its applications to decentralized MIMO systems are also discussed. In Section VI, simulation results about EDRID for signal detection in XL-MIMO systems are illustrated. Finally, Section VII 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 $\mathbf{B}$. $\|\cdot\|$ 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 $\mathrm{Tr}(\cdot)$ denotes the matrix trace. $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ represent the minimum and maximum eigenvalues of a matrix while $\sigma_{\min}(\cdot)$ and $\sigma_{\max}(\cdot)$ 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 XL-MIMO systems is reviewed, followed by the background of distributed detection schemes.

A. Uplink Linear Signal Detection

We consider the signal detection in XL-MIMO systems, where all the user equipments (UEs) are located in the far field

TABLE I
COMPARISONS AMONG DIFFERENT DISTRIBUTED DETECTION SCHEMES FOR DECENTRALIZED MIMO SYSTEMS

	Computational Complexity	Flexible number of antennas in each DU	Global Convergence	Decentralized Architecture
ADMM [10]	$\mathcal{O}(K^3r + K^2r \cdot k)$	✓	×	DBP
CG [11]	$\mathcal{O}(K^2N + K^2r \cdot k)$	✓	×	DBP
FD-MMSE [16]	$\mathcal{O}(K^3 \cdot r)$	✓	×	FD
PD-MMSE [12]	$\mathcal{O}(K^2N)$	✓	×	PD
CD [15]	$\mathcal{O}(KN \cdot k)$	✓	×	FD
DN [23]	$\mathcal{O}(K^2N + K^2r \cdot k)$	×	×	Ring, Star
RLS [19]	$\mathcal{O}(K^2N)$	× (limit to 1)	✓	Daisy-chain
SGD [20]	$\mathcal{O}(KN)$	× (limit to 1)	×	Daisy-chain
ASGD [21]	$\mathcal{O}(KN)$	× (limit to 1)	×	Daisy-chain
SDK [42]	$\mathcal{O}(KN \cdot k)$	× (limit to 1)	×	Daisy-chain
MCRBK [31]	$\mathcal{O}(Kq^2 \cdot t)$ (i.e., $\mathcal{O}(KNq \cdot k)$)	✓	✓	Daisy-chain Ring, Star, etc.
EDRID (this work)	$\mathcal{O}(Kq \cdot t)$ (i.e., $\mathcal{O}(KN \cdot k)$)	✓	✓	Daisy-chain Ring, Star, etc.

of the base station (BS) antenna array. The BS is equipped with N antennas and UEs with K antennas in total are served simultaneously, $N > K$. Let $\mathbf{x} \in \mathcal{X}^K$ be the signal transmitted from UEs, where the i-th element of $\mathbf{x}$ (i.e., x_i) is a symbol in the QAM constellation $\mathcal{X}$. Given the full column rank channel matrix $\mathbf{H} \in \mathbb{C}^{N\times K}$, the received signal $\mathbf{y} \in \mathbb{C}^N$ of a generic subcarrier in an OFDMA system can be written as[1]

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

where $\mathbf{n}$ is AWGN noise vector with elements obeying $\mathcal{CN}(0, \sigma_n^2)$. 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.

To restore the transmitted signal $\mathbf{x}$ in (1), the two most common approaches for linear detection are zero forcing (ZF) and minimum mean-square error (MMSE) (i.e., $\widehat{\mathbf{x}}_{\text{zf}}$ and $\widehat{\mathbf{x}}_{\text{mmse}}$) with

$$\mathbf{x}_{\text{zf}} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y} \tag{2}$$

and

$$\mathbf{x}_{\text{mmse}} = \left(\mathbf{H}^H\mathbf{H} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y}, \tag{3}$$

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

$$\widehat{\mathbf{x}}_{\text{zf}} = \lceil \mathbf{x}_{\text{zf}} \rfloor_Q \in \mathcal{X}^K \text{ and } \widehat{\mathbf{x}}_{\text{mmse}} = \lceil \mathbf{x}_{\text{mmse}} \rfloor_Q \in \mathcal{X}^K. \tag{4}$$

Clearly, $\mathbf{x}_{\text{zf}}$ in (2) essentially accounts for the least squares (LS) solution of the problem

$$\mathbf{x} = \arg\min_{\mathbf{x}\in\mathbb{C}^K} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2. \tag{5}$$

[1]Here for simplicity we focus on the frequency-flat model, but the scheme can be easily generalized to the frequency selective case.

[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.

For notational simplicity, throughout the context, we apply $\mathbf{x}^*$ to represent $\mathbf{x}_{\text{zf}} = \mathbf{H}^\dagger\mathbf{y}$ and define the error vector $\mathbf{e} = \mathbf{y} - \mathbf{H}\mathbf{x}^* \in \mathbb{C}^N$. Note that $\mathbf{x}^*$ can be easily extended to the detection result of MMSE by replacing $\mathbf{H}$ and $\mathbf{y}$ with $\overline{\mathbf{H}} = [\mathbf{H}; \sigma_n\mathbf{I}_K]$ and $\overline{\mathbf{y}} = [\mathbf{y}; \mathbf{0}_K]$.

It has been demonstrated in [43] that if $N \gg K$ the channel matrix $\mathbf{H}$ will become near orthogonal as a benefit of *favorable propagation*, which implies that the detection performance of ZF or MMSE is close to the maximum likelihood (ML) detection performance.

B. Distributed MIMO Detection Methods

Generally, in distributed MIMO detection, the received signal $\mathbf{y}$ at BS is partitioned into r blocks of q antennas each (i.e., $N = rq$),

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

with $\mathbf{y}_i \in \mathbb{C}^q$. This corresponds to dividing the channel matrix $\mathbf{H}$ into r stacked vertically blocks, namely,

$$\mathbf{H} = [\mathbf{H}_1; \mathbf{H}_2; \ldots; \mathbf{H}_r] \text{ or } \mathbf{H} = [\mathbf{H}_1^H\mathbf{H}_2^H \ldots \mathbf{H}_r^H]^H, \tag{7}$$

where $\mathbf{H}_i \in \mathbb{C}^{q\times K}$ is the channel matrix at the i-th group of antennas. Accordingly, these groups of antennas are connected to different DUs, which obey the following setup

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{x} + \mathbf{n}_i, \ i = 1, 2, \ldots, r \tag{8}$$

with $\mathbf{n} = \left[\mathbf{n}_1^H, \mathbf{n}_2^H, \ldots, \mathbf{n}_r^H\right]^H$. Clearly, built on (8), signal detections in either cell-free MIMO systems [44] or decentralized XL-MIMO systems [45] can be well performed in a distributed fashion. Here, for a better illustration, the submatrix $\mathbf{I}_{\mathcal{Q}_i,:} \in \mathbb{C}^{q\times N}$ of the identity matrix $\mathbf{I} \in \mathbb{C}^{N\times N}$ is introduced, where $\mathcal{Q}_i = \{(i-1)q + 1, \ldots, iq\}$, $1 \leq i \leq r$ is a subset with $|\mathcal{Q}_i| = q$. According to $\mathcal{Q}_i$, it is clear to see that

$$\mathbf{H}_i = \mathbf{I}_{\mathcal{Q}_i,:}\mathbf{H}, \quad \mathbf{y}_i = \mathbf{I}_{\mathcal{Q}_i,:}\mathbf{y}, \quad \mathbf{n}_i = \mathbf{I}_{\mathcal{Q}_i,:}\mathbf{n}. \tag{9}$$

Based on (8), different decentralized baseband architectures can be designed such that the DUs exchange messages and converge to an estimate of $\widehat{\mathbf{x}}$ that somehow "pulls together" all the separate observations in (8). For example, besides DBP [10], decentralized architectures like ring [23], daisy-chain [19], and fully decentralized (FD) [16], also employ the same DU model as in (8), together with the CPU to accomplish signal detection. Based on this, distributed iterative methods such as CG, ADMM, and DN have been proposed to solve the detection problem in a decentralized manner.

In [31], based on the classic Kaczmarz method, some randomized iterative methods are introduced into the distritbuted detection of MIMO systems. Specifically, in the randomized block Kaczmarz (RBK) algorithm, the detected signal $\mathbf{x}$ is updated via random iterations

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \mathbf{H}_i^{\dagger}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}^{t-1}), \tag{10}$$

where at each iteration (indexed by t) the index $1 \le i \le r$ (i.e., the index set $\mathcal{Q}_i$) is randomly sampled from a designed discrete distribution $p(i)$, i.e., $i \sim p(i)$. According to (10), DUs can work in a cooperative way for refining the detected signal. To improve both convergence and efficiency, the technique of conditional sampling with respect to the index i is then employed. This further leads to the iterations of multi-step conditional randomized block Kaczmarz (MCRBK) algorithm

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \alpha_t\mathbf{H}_i^{\dagger}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}^{t-1}), \tag{11}$$

where $\alpha_t > 0$ denotes the step-size of each iteration.

As demonstrated in [31], with a proper choice of α_t the distributed detection based on MCRBK exponentially converges to the linear detection solution (i.e., ZF or MMSE). Meanwhile, MCRBK also enjoys the global convergence. This means its exponential convergence always holds without extra conditions or requirements, making it well fitted to different cases of interest. From (11), the computational complexity of MCRBK is $2q^2K + q^3 + 2qK$.

III. Low-Complexity Randomized Iterative Algorithm for Distributed Detection

Typically, as shown in (10) and (11), the iterations of RBK and MCRBK require the computation of the pseudoinverse of the submatrix $\mathbf{H}_i$, resulting in the considerable complexity burden to the distributed detection. To this end, we now introduce the efficient distributed randomized iterative detection (EDRID) algorithm, which effectively avoids the matrix pseudoinverse in its iteration updates.

In particular, the iteration of the proposed EDRID algorithm is given by

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \alpha\mathbf{H}_i^H(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}^{t-1}), \tag{12}$$

where $\alpha > 0$ stands for the step-size and the index $1 \le i \le r$ is randomly selected according to a uniform distribution $p(i)$, i.e.,

$$p(i) \triangleq p(\mathcal{Q}_i) = \frac{1}{r}. \tag{13}$$

Clearly, different from the iteration in (10) or (11), the matrix pseudoinverse has been removed from the iteration in (12). To be more specific, the total complexity per iteration in (12) is $2qK$, which is much smaller than those in (10) or (11).

Theorem 1: As for decentralized MIMO detection, with uniform sampling probability $p(i) = \frac{1}{r}$, the randomized iterations based on (12) converges by

$$E[\|\mathbf{x}^t - \mathbf{x}^*\|^2] \le \beta^t\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \alpha^2\eta\frac{1-\beta^t}{1-\beta} \tag{14}$$

for any given $\mathbf{x}^0$, if the step-size α satisfies

$$0 < \alpha < \frac{2}{\lambda_{\max}^*}, \tag{15}$$

where

$$\lambda_{\max}^* \triangleq \max_{1\le i\le r} \lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i), \tag{16}$$

$$\beta \triangleq 1 - \frac{1}{r}(2\alpha - \alpha^2\lambda_{\max}^*)\sigma_{\min}^2(\mathbf{H}) < 1, \tag{17}$$

$$\eta \triangleq E[\|\mathbf{H}_i^H\mathbf{I}_{\mathcal{Q}_i,:}(\mathbf{y} - \mathbf{H}\mathbf{H}^{\dagger}\mathbf{y})\|^2]. \tag{18}$$

Proof: To start with, according to (12), it follows that

$$\mathbf{x}^t - \mathbf{x}^* = \mathbf{x}^{t-1} - \mathbf{x}^* + \alpha\mathbf{H}_i^H(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}^{t-1}). \tag{19}$$

From (19), we have that

$$\begin{aligned}\|\mathbf{x}^t\mathbf{x}^*\|^2 = \|\mathbf{x}^t\mathbf{x}^*\|^2 + 2\alpha(\mathbf{x}^{t-1}\mathbf{x}^*)^H\mathbf{H}_i^H(\mathbf{y}_i\mathbf{H}_i\mathbf{x}^{t-1}) \\ + \ \alpha^2\|\mathbf{H}_i^H(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}^{t-1})\|^2.\end{aligned} \tag{20}$$

Then, by applying $\mathbf{H}\mathbf{x}^* = \mathbf{H}\mathbf{H}^{\dagger}\mathbf{y}$, the following holds

$$\begin{aligned}&\|\mathbf{H}_i^H(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}^{t1})\|^2 = \|\mathbf{H}_i^H\mathbf{I}_{\mathcal{Q}_i,:}(\mathbf{y} - \mathbf{H}\mathbf{x}^{t-1})\|^2 \\ &= \|\mathbf{H}_i^H\mathbf{I}_{\mathcal{Q}_i,:}(\mathbf{y} + \mathbf{H}\mathbf{x}^* - \mathbf{H}\mathbf{H}^{\dagger}\mathbf{y} - \mathbf{H}\mathbf{x}^{t-1})\|^2 \\ &\le \|\mathbf{H}_i^H\mathbf{H}_i(\mathbf{x}^{t-1} - \mathbf{x}^*)\|^2 + \|\mathbf{H}_i^H\mathbf{I}_{\mathcal{Q}_i,:}(\mathbf{y} - \mathbf{H}\mathbf{H}^{\dagger}\mathbf{y})\|^2.\end{aligned} \tag{21}$$

Moreover, the term $\|\mathbf{H}_i^H\mathbf{H}_i(\mathbf{x}^{t-1} - \mathbf{x}^*)\|^2$ in (21) can be upper bounded as

$$\begin{aligned}&\|\mathbf{H}_i^H\mathbf{H}_i(\mathbf{x}^{t-1} - \mathbf{x}^*)\|^2 = (\mathbf{x}^{t-1} - \mathbf{x}^*)^H(\mathbf{H}_i^H\mathbf{H}_i)^2(\mathbf{x}^{t-1} - \mathbf{x}^*) \\ &\overset{(a)}{\le} \lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)(\mathbf{x}^{t-1} - \mathbf{x}^*)^H\mathbf{H}_i^H\mathbf{H}_i(\mathbf{x}^{t-1} - \mathbf{x}^*) \\ &\le \lambda_{\max}^*(\mathbf{x}^{t-1} - \mathbf{x}^*)^H\mathbf{H}_i^H\mathbf{H}_i(\mathbf{x}^{t-1} - \mathbf{x}^*)\end{aligned} \tag{22}$$

with $\lambda_{\max}^* = \max_{1\le i\le r}\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)$. Here, the inequality (a) comes from the fact that for symmetric positive semidefinite matrix $\mathbf{H}_i^H\mathbf{H}_i$, the matrix $\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{H}_i^H\mathbf{H}_i - (\mathbf{H}_i^H\mathbf{H}_i)^2$ is also positive semidefinite (see Appendix A for the proof of this fact). Therefore, for any non-zero vector $\mathbf{z}$, it follows that

$$\mathbf{z}^H[\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{H}_i^H\mathbf{H}_i - (\mathbf{H}_i^H\mathbf{H}_i)^2]\mathbf{z} \ge 0, \tag{23}$$

which accounts for

$$\mathbf{z}^H(\mathbf{H}_i^H\mathbf{H}_i)^2\mathbf{z} \le \mathbf{z}^H\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{H}_i^H\mathbf{H}_i\mathbf{z}. \tag{24}$$

Based on (20), (21) and (22), we obtain

$$\begin{aligned}\|\mathbf{x}^t - \mathbf{x}^*\|^2 \le &\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + 2\alpha(\mathbf{x}^{t-1} - \mathbf{x}^*)^H\mathbf{H}_i^H(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}^{t-1}) \\ &+ \alpha^2\lambda_{\max}^*(\mathbf{x}^{t-1} - \mathbf{x}^*)^H\mathbf{H}_i^H\mathbf{H}_i(\mathbf{x}^{t-1} - \mathbf{x}^*) \\ &+ \alpha^2\|\mathbf{H}_i^H\mathbf{I}_{\mathcal{Q}_i,:}(\mathbf{y} - \mathbf{H}\mathbf{H}^{\dagger}\mathbf{y})\|^2.\end{aligned} \tag{25}$$

Next, based on the sampling probability $p(i)$ in (13), by taking the expectation on both sides, we have

$$\begin{aligned} E[\|\mathbf{x}^t-\mathbf{x}^*\|^2] &\le \|\mathbf{x}^{t-1}\mathbf{x}^*\|^2 \\ &+ 2\alpha(\mathbf{x}^{t-1}\mathbf{x}^*)^H E[\mathbf{H}_i^H(\mathbf{y}_i\mathbf{H}_i\mathbf{x}^{t-1})] \\ &+ \alpha^2\lambda^*_{\max}(\mathbf{x}^{t-1}-\mathbf{x}^*)^H E[\mathbf{H}_i^H\mathbf{H}_i](\mathbf{x}^{t-1}-\mathbf{x}^*)+\alpha^2\eta. \end{aligned} \tag{26}$$

Here, regarding to the term $E[\mathbf{H}_i^H(\mathbf{y}_i-\mathbf{H}_i\mathbf{x}^{t-1})]$ in (26), it follows that

$$\begin{aligned} E[\mathbf{H}_i^H(\mathbf{y}_i-\mathbf{H}_i\mathbf{x}^{t-1})] &= \mathbf{H}^H E[\mathbf{I}^H_{\mathcal{Q}_i,:}\mathbf{I}_{\mathcal{Q}_i,:}](\mathbf{y}-\mathbf{H}\mathbf{x}^{t-1}) \\ &\overset{(b)}{=} \frac{1}{r}\mathbf{H}^H(\mathbf{y}-\mathbf{H}\mathbf{x}^{t-1}) \\ &= \frac{1}{r}[\mathbf{H}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}-\mathbf{H}^H\mathbf{H}\mathbf{x}^{t-1}] \\ &= \frac{1}{r}(\mathbf{H}^H\mathbf{H}\mathbf{H}^\dagger\mathbf{y}-\mathbf{H}^H\mathbf{H}\mathbf{x}^{t-1}) \\ &= \frac{1}{r}(\mathbf{H}^H\mathbf{H}\mathbf{x}^*-\mathbf{H}^H\mathbf{H}\mathbf{x}^{t-1}) \\ &\overset{(c)}{=} -\frac{1}{r}\mathbf{H}^H\mathbf{H}(\mathbf{x}^{t-1}-\mathbf{x}^*), \end{aligned} \tag{27}$$

where (b) holds due to $E[\mathbf{I}^H_{\mathcal{Q}_i,:}\mathbf{I}_{\mathcal{Q}_i,:}]=\frac{1}{r}\mathbf{I}$ under the uniform sampling probability p_i and (c) comes from $\mathbf{H}^H\mathbf{y}=\mathbf{H}^H\mathbf{H}\mathbf{x}^*$.

Therefore, based on (27) and the fact that

$$E[\mathbf{H}_i^H\mathbf{H}_i]=\mathbf{H}^H E[\mathbf{I}^H_{\mathcal{Q}_i,:}\mathbf{I}_{\mathcal{Q}_i,:}]\mathbf{H}=\frac{1}{r}\mathbf{H}^H\mathbf{H}, \tag{28}$$

the inequality in (26) can be further expressed by

$$\begin{aligned} E[\|\mathbf{x}^t-\mathbf{x}^*\|^2] &\le \|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2 \\ &+ \frac{1}{r}(\alpha^2\lambda^*_{\max}-2\alpha)(\mathbf{x}^{t-1}-\mathbf{x}^*)^H\mathbf{H}^H\mathbf{H}(\mathbf{x}^{t-1}-\mathbf{x}^*)+\alpha^2\eta. \end{aligned} \tag{29}$$

Furthermore, as for the term $(\mathbf{x}^{t-1}-\mathbf{x}^*)^H\mathbf{H}^H\mathbf{H}(\mathbf{x}^{t-1}-\mathbf{x}^*)$ in (29), because of

$$\mathbf{u}^H\mathbf{A}^H\mathbf{A}\mathbf{u}\ge\sigma^2_{\min}(\mathbf{A})\|\mathbf{u}\|^2 \tag{30}$$

for matrix $\mathbf{A}$ and vector $\mathbf{u}$, it follows that

$$(\mathbf{x}^{t-1}-\mathbf{x}^*)^H\mathbf{H}^H\mathbf{H}(\mathbf{x}^{t-1}-\mathbf{x}^*)\ge\sigma^2_{\min}(\mathbf{H})\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2. \tag{31}$$

Then, the inequality in (29) can be expressed by

$$\begin{aligned} E[\|\mathbf{x}^t-\mathbf{x}^*\|^2] &\le \left[1-\frac{1}{r}(2\alpha-\alpha^2\lambda^*_{\max})\sigma^2_{\min}(\mathbf{H})\right]\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2 \\ &+\alpha^2\eta=\beta\|\mathbf{x}^{t-1}-\mathbf{x}^*\|^2+\alpha^2\eta, \end{aligned} \tag{32}$$

where $\beta=1-\frac{1}{r}(2\alpha-\alpha^2\lambda^*_{\max})\sigma^2{}_{\min}(\mathbf{H})$. Then, by unrolling the recurrence in (32), we can obtain

$$\begin{aligned} E[\|\mathbf{x}^t-\mathbf{x}^*\|^2] &\le \beta^t\|\mathbf{x}^0-\mathbf{x}^*\|^2+\alpha^2\eta\sum_{j=0}^{t-1}\beta^j \\ &\overset{(d)}{\approx} \beta^t\|\mathbf{x}^0-\mathbf{x}^*\|^2+\alpha^2\eta\frac{1-\beta^t}{1-\beta}, \end{aligned} \tag{33}$$

where the approximation in (d) follows the *geometric series* $\sum_{i=k}^n z^i=\frac{z^k-z^{n+1}}{1-z}$.

On the other hand, to ensure the convergence in (33), the condition $1>\beta>0$ should be made, namely,

$$1>\frac{1}{r}(2\alpha-\alpha^2\lambda^*_{\max})\sigma^2_{\min}(\mathbf{H})>0, \tag{34}$$

which corresponds to

$$0<\alpha<\frac{2}{\lambda^*_{\max}}. \tag{35}$$

□

In order to ensure convergence of the algorithm, i.e., condition (35), one needs a bound on the maximum eigenvalue $\lambda_{\max}$. This depends on the specific channel statistics and can be learned for the specific propagation scenario. For example, when the matrix $\mathbf{H}$ contains Gaussian i.i.d. elements, from classical random matrix theory we have that [46]

$$\lambda_{\max}(\mathbf{H}^H\mathbf{H})\to N\left(1+\sqrt{\frac{K}{N}}\right)^2, \tag{36}$$

$$\lambda_{\min}(\mathbf{H}^H\mathbf{H})\to N\left(1-\sqrt{\frac{K}{N}}\right)^2, \tag{37}$$

when $N\to\infty$ and $K\to\infty$. Similar to (36) and (37), the following results can also be got

$$\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)=\lambda_{\max}(\mathbf{H}_i\mathbf{H}_i^H)\to K\left(1+\sqrt{\frac{q}{K}}\right)^2, \tag{38}$$

$$\lambda_{\min}(\mathbf{H}_i^H\mathbf{H}_i)=\lambda_{\min}(\mathbf{H}_i\mathbf{H}_i^H)\to K\left(1-\sqrt{\frac{q}{K}}\right)^2, \tag{39}$$

when $K\to\infty$ and $q\to\infty$. From (38), the range of α can be roughly estimated by

$$0<\alpha<\frac{2}{K\left(1+\sqrt{\frac{q}{K}}\right)^2}, \tag{40}$$

where the choice of $\alpha=1/K$ is recommended for the sake of efficiency. More generally, in practical scenarios, the channel statistics can be learned at the time of the system setup (e.g., installation of the XL-MIMO base station) via measurements (i.e., in a data-driven way), and the parameter α can be adapted as a consequence.

According to Theorem 1, with proper choice of α, the proposed EDRID algorithm converges exponentially to the least squares solution $\mathbf{x}^*$ but within a specified error bound. Put it in another way, due to this estimation inaccuracy, the distributed detection based on (12) only converges to within a radius of $\mathbf{x}^*$, which may degrade the detection performance of decentralized MIMO systems.

IV. Exact Convergence to LS Solution by Dynamic Step-Size

In order to guarantee the exact convergence of EDRID to the LS solution $\mathbf{x}^*$, we now introduce the dynamic step-size $\alpha^t>0$ into the iterations in (12). By dynamic choice 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, by the appropriate choice

of $\alpha^t > 0$, the proposed EDRID algorithm is able to return the approximation of $\mathbf{x}^*$ with arbitrary accuracy as t increases.

In particular, under the help of dynamic step-size α_t, the randomized iteration in (12) becomes

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \alpha_t \mathbf{H}_i^H (\mathbf{y}_i - \mathbf{H}_i \mathbf{x}^{t-1}) \tag{41}$$

with

$$0 < \alpha_t < \frac{2}{\lambda^*_{\max}}. \tag{42}$$

Theorem 2: As for decentralized MIMO detection, the proposed EDRID algorithm following $p(i) = \frac{1}{r}$ ensures exact convergence to the LS solution by

$$E[\|\mathbf{x}^L - \mathbf{x}^*\|^2] \le \prod_{t=1}^{L} (1 - \omega\alpha_t) \|\mathbf{x}^0 - \mathbf{x}^*\|^2 \tag{43}$$

with dynamic step-size

$$\alpha_t = \frac{\omega g(t-1)}{\lambda^*_{\max} \omega g(t-1) + 1}, \tag{44}$$

where

$$g(t) = g(t-1)(1 - \omega\alpha_t), \tag{45}$$

$$g(0) = \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{\eta}, \tag{46}$$

$$\omega \triangleq \sigma^2_{\min}(\mathbf{H})/r. \tag{47}$$

Proof: Under dynamic step-size α_t, the convergence in (32) can be expressed as

$$E[\|\mathbf{x}^t - \mathbf{x}^*\|^2] \le \left[1 - (2\alpha_t - \alpha_t^2 \lambda^*_{\max})\omega\right] \|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \alpha_t^2 \eta \tag{48}$$

with $\omega = \sigma^2_{\min}(\mathbf{H})/r$.

By iteratively applying the recursive relation in (48), it follows that

$$\begin{aligned} E[\|\mathbf{x}^L - \mathbf{x}^*\|^2] &\le \prod_{j=1}^{L} \left[1 - (2\alpha_j - \alpha_j^2 \lambda^*_{\max})\omega\right] \|\mathbf{x}^0 - \mathbf{x}^*\|^2 \\ &\quad + \sum_{j=1}^{L} \alpha_j^2 \prod_{i=j+1}^{L} \left[1 - (2\alpha_i - \alpha_i^2 \lambda^*_{\max})\omega\right] \eta \\ &= g(L)\eta \end{aligned} \tag{49}$$

with the following definition

$$\begin{aligned} g(t) \triangleq \prod_{j=1}^{t} \left[1 - (2\alpha_j - \alpha_j^2 \lambda^*_{\max})\omega\right] \|\mathbf{x}^0 - \mathbf{x}^*\|^2 \frac{1}{\eta} \\ + \sum_{j=1}^{t} \alpha_j^2 \prod_{i=j+1}^{t} \left[1 - (2\alpha_i - \alpha_i^2 \lambda^*_{\max})\omega\right]. \end{aligned} \tag{50}$$

To be more specific, from (50), by induction we can find the following recurrence relationship

$$g(t) = \left[1 - (2\alpha_t - \alpha_t^2 \lambda^*_{\max})\omega\right] g(t-1) + \alpha_t^2 \tag{51}$$

for $t \ge 1$ with

$$g(0) = \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{\eta}. \tag{52}$$

According to (51), the term $g(t-1)$ is independent of the step-size α_t while $g(t)$ is a function of α_t. Therefore, by letting the derivative about $g(t)$ versus α_t be zero, the optimal value of α_t that minimizes $g(t)$ is given by

$$\alpha_t = \frac{\omega g(t-1)}{\lambda^*_{\max} \omega g(t-1) + 1}. \tag{53}$$

Then, by substituting α_t in (53) back into $g(t)$ in (51), we have

$$g(t) = g(t-1)(1 - \omega\alpha_t). \tag{54}$$

To summarize, based on (49), (52) and (54), we can conclude that

$$E[\|\mathbf{x}^L - \mathbf{x}^*\|^2] \le \prod_{t=1}^{L} (1 - \omega\alpha_t) \|\mathbf{x}^0 - \mathbf{x}^*\|^2, \tag{55}$$

completing the proof. □

According to Theorem 2, by making use of the dynamic step-size α_t, the error bound during the convergence of EDRID (i.e., the term $\alpha^2 \eta \frac{1-\beta^t}{1-\beta}$ in (14)) has been well addressed, which facilitates the convergence to the exact linear detection solution $\mathbf{x}^*$. Note that the choice of α_t in (44) always satisfies the requirement in (42) due to

$$\alpha_t = \frac{1}{\lambda^*_{\max} + \frac{1}{\omega g(t-1)}} < \frac{2}{\lambda^*_{\max}}. \tag{56}$$

Corollary 1: The proposed EDRID algorithm with dynamic step-size α_t in (44) achieves the global convergence as its convergence in (43) always holds due to

$$0 < \omega\alpha_t < 1. \tag{57}$$

Proof: On one hand, according to the definitions of ω and α_t, it is intuitive to confirm $\omega\alpha_t > 0$. On the other hand, based on (47) and (53), we have

$$\begin{aligned} \omega\alpha_t &= \frac{\sigma^2_{\min}(\mathbf{H})}{r} \frac{\omega g(t-1)}{\lambda^*_{\max} \omega g(t-1) + 1} \\ &\le \frac{\min\limits_{1 \le i \le r} \sigma^2_{\min}(\mathbf{H}_i)}{r} \frac{\omega g(t-1)}{\lambda^*_{\max} \omega g(t-1) + 1} \\ &= \frac{\min\limits_{1 \le i \le r} \lambda_{\min}(\mathbf{H}_i^H \mathbf{H}_i)}{\lambda^*_{\max}} \frac{1}{r} \frac{\omega g(t-1)}{\omega g(t-1) + \frac{1}{\lambda^*_{\max}}} \\ &= \frac{\lambda^*_{\min}}{\lambda^*_{\max}} \frac{1}{r} \frac{\omega g(t-1)}{\omega g(t-1) + \frac{1}{\lambda^*_{\max}}} \\ &< 1. \end{aligned} \tag{58}$$

Here, $\lambda^*_{\min} \triangleq \min_{1 \le i \le r} \lambda_{\min}(\mathbf{H}_i^H \mathbf{H}_i)$ and the inequality shown above comes from the fact that [47]

$$\sigma_{i+k+l}(\mathbf{A}) \le \sigma_i(\mathbf{B}) \le \sigma_i(\mathbf{A}), \tag{59}$$

where $\mathbf{B}$ denotes an $(m-k) \times (n-l)$ submatrix of $\mathbf{A} \in \mathbb{C}^{m \times n}$. □

From Corollary 1 and Theorem 2, the convergence of EDRID with dynamic step-size is always guaranteed due to the convergence rate $0 < 1 - \omega\alpha_t < 1$ (i.e., global convergence), making it suitable to various decentralized MIMO scenarios of interest. It is interesting to notice that several other decentralized

detection algorithms such as FD-MMSE in [16], FD-CD in [15], decentralized Newton in [23], do not satisfy this property and convergence is ensured only under specific conditions. Furthermore, because of (83), we notice that $g(t)$ is decreasing with the increment of t, i.e., $g(t) < g(t-1)$. This implies that the dynamic step-size α_t in (53) also forms a monotonically decreasing sequence, i.e.,

$$\alpha_t < \alpha_{t-1}. \tag{60}$$

To summarize, in the framework of randomized iterations, the problem of distributed MIMO detection is transformed into an equivalent formation, namely,

$$\begin{aligned} \mathbf{x} &= \arg\min_{\mathbf{x}\in\mathbb{C}^K} \|\mathbf{I}_{\mathcal{Q}_i,:}\mathbf{H}\mathbf{x} - \mathbf{I}_{\mathcal{Q}_i,:}\mathbf{y}\|^2 \\ s.t. \quad & p(\mathcal{Q}_i) = \tfrac{1}{r}. \end{aligned} \tag{61}$$

By executing the randomized iterations in (41), we show that the common solution $\mathbf{x}^*$ that fits every subproblem $\mathbf{H}_i\mathbf{x} - \mathbf{y}_i$ (i.e., $\mathbf{I}_{\mathcal{Q}_i,:}(\mathbf{H}\mathbf{x} - \mathbf{y})$) would be gradually refined. Clearly, such a problem reformulation combines the factors of both Euclidean distance and statistics, which paves an effective way for the following problem solving. Besides the original Euclidean domain, by benefiting from the extra system freedom introduced by statistics, remarkable performance improvement or complexity reduction can be achieved in a reasonable way, where more related randomized schemes based on this methodology can be found in our previous works [31], [48], [49], [50], [51]. Here, to be more specific, we refer such a methodology to as hybrid-domain signal processing, which firstly establishes a novel but equivalent problem formulation based on different domains and then solve it by fully utilizing the introduced system freedoms.

V. Application to Distributed Detection via Conditional Sampling

The proposed EDRID algorithm relies on the random sampling about the index i. To facilitate its application for decentralized MIMO systems and enhance its convergence performance at the same time, we now introduce the technique of conditional sampling.

A. Convergence Improvement by Conditional Sampling

Specifically, considering the index j is sampled at iteration $t-1$ (i.e. the index set $\mathcal{Q}_j$ is selected), then by conditional sampling, the index i at iteration t is sampled according to the following conditional distribution:

$$\overline{p}(i) \triangleq \overline{p}(\mathcal{Q}_i) = \frac{p(i)}{1-p(j)},\ i \neq j, \tag{62}$$

where the index j sampled at the previous iteration is removed from the sampling candidate list of i. As uniform sampling is applied, this corresponds to

$$\overline{p}(i) = \frac{1}{r-1},\ i \neq j. \tag{63}$$

Intuitively, thanks to the conditional sampling in (63), the sampling repetitions over two consecutive iterations can be effectively avoided, which leads to an improved sampling diversity[3].

Theorem 3: As for decentralized MIMO detection, the proposed EDRID algorithm following $\overline{p}(i)$ in (63) ensures exact convergence to LS solution by

$$E[\|\mathbf{x}^L - \mathbf{x}^*\|^2] \leq \prod_{t=1}^{L}(1 - \overline{\omega}\overline{\alpha}_t)\|\mathbf{x}^0 - \mathbf{x}^*\|^2 \tag{64}$$

with dynamic step-size

$$\overline{\alpha}_t = \frac{\overline{\omega}\overline{g}(t-1)}{\lambda^*_{\max}\overline{\omega}\overline{g}(t-1) + 1}, \tag{65}$$

where

$$\overline{g}(t) = \overline{g}(t-1)(1 - \overline{\omega}\overline{\alpha}_t), \tag{66}$$

$$\overline{g}(0) = \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{\overline{\eta}}, \tag{67}$$

$$\overline{\omega} \triangleq \frac{\sigma^2_{\min}(\mathbf{H})}{r-1}, \tag{68}$$

$$\overline{\eta} \triangleq E_{\overline{p}(i)}[\|\mathbf{H}_i^H\mathbf{I}_{\mathcal{Q}_i,:}(\mathbf{y} - \mathbf{H}\mathbf{H}^\dagger\mathbf{y})\|^2]. \tag{69}$$

Proof: To begin with, when the conditional sampling $\overline{p}(i)$ is applied, the expectation term $E_{\overline{p}(i)}[\mathbf{H}_i^H\mathbf{H}_i]$ becomes

$$\begin{aligned} E_{\overline{p}(i)}[\mathbf{H}_i^H\mathbf{H}_i] &= \mathbf{H}^H E_{\overline{p}(i)}[\mathbf{I}^H_{\mathcal{Q}_i,:}\mathbf{I}_{\mathcal{Q}_i,:}]\mathbf{H} \\ &= \frac{1}{r-1}\mathbf{H}^H\mathbf{I}_{[-\mathcal{Q}_j]}\mathbf{H} \\ &= \frac{1}{r-1}\mathbf{H}^H_{[-\mathcal{Q}_j]}\mathbf{H}_{[-\mathcal{Q}_j]} \end{aligned} \tag{70}$$

where $\mathbf{H}_{[-\mathcal{Q}_j]} = \mathbf{I}_{[-\mathcal{Q}_j]}\mathbf{H}$ and $\mathbf{I}_{[-\mathcal{Q}_j]}$ denotes the identity matrix but the corresponding diagonal elements related to the index set $\mathcal{Q}_j$ are 0. For example, given $\mathcal{Q}_j = \{3,4\}$, then it follows

$$\mathbf{I}_{[-\{3,4\}]} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{71}$$

Based on (70), the inequality in (26) can be written as

$$\begin{aligned} &E\left[\left\|\mathbf{x}^t - \mathbf{x}^*\right\|^2\right] \leq \left\|\mathbf{x}^{t-1} - \mathbf{x}^*\right\|^2 \\ &+ \frac{1}{r-1}\left(\alpha^2\lambda^*_{\max} - 2\alpha\right)\left(\mathbf{x}^{t-1} - \mathbf{x}^*\right)^H\mathbf{H}^H_{[-\mathcal{Q}_j]}\mathbf{H}_{[-\mathcal{Q}_j]}\left(\mathbf{x}^{t-1} - \mathbf{x}^*\right) \\ &+ \alpha^2\bar{\eta} \\ &\leq \left[1 - \frac{1}{r-1}\left(2\alpha - \alpha^2\lambda^*_{\max}\right)\sigma^2_{\min}\left(\mathbf{H}_{[-\mathcal{Q}_j]}\right)\right]\left\|\mathbf{x}^{t-1} - \mathbf{x}^*\right\|^2 \\ &+ \alpha^2\bar{\eta} \\ &\overset{(e)}{\leq} \left[1 - \frac{1}{r-1}\left(2\alpha - \alpha^2\lambda^*_{\max}\right)\sigma^2_{\min}(\mathbf{H})\right]\left\|\mathbf{x}^{t-1} - \mathbf{x}^*\right\|^2 \\ &+ \alpha^2\bar{\eta}. \end{aligned} \tag{72}$$

[3]Here, we use sampling diversity to describe the diversity of samples during a certain iterations.

with $\overline{\eta} = E_{\overline{p}(i)}[\|\mathbf{H}_i^H \mathbf{I}_{\mathcal{Q}_i,:}(\mathbf{y} - \mathbf{H}\mathbf{H}^\dagger \mathbf{y})\|^2]$ and inequality (e) comes from (59). Then, by applying the dynamic step-size $\overline{\alpha}_t$, the inequality in (72) can be expressed by

$$E[\|\mathbf{x}^t - \mathbf{x}^*\|^2] \leq \left[1 - (2\overline{\alpha}_t - \overline{\alpha}_t^2 \lambda_{\max}^*)\overline{\omega}\right] \|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 + \overline{\alpha}_t^2 \overline{\eta} \tag{73}$$

with $\overline{\omega} = \sigma_{\min}^2(\mathbf{H})/(r-1)$.

Subsequently, following the derivation in Theorem 2, we obtain

$$E[\|\mathbf{x}^L - \mathbf{x}^*\|^2] \leq \prod_{t=1}^{L} (1 - \overline{\omega}\overline{\alpha}_t) \|\mathbf{x}^0 - \mathbf{x}^*\|^2, \tag{74}$$

which completes the proof. □

By Theorem 3, EDRID with conditional sampling probability $\overline{p}(i)$ still converges to LS solution in an exponential manner. In fact, it is readily to confirm that the global convergence also holds as well. Meanwhile, due to $\overline{\omega} > \omega$, the convergence improvement can also be verified in the comparison between (48) and (73). Motivated by these, in what follows we extend the multi-step conditional sampling into EDRID for further convergence gains.

B. Enhancement by Multi-Step Conditional Sampling

In particular, replaces the conditional sampling probability $\overline{p}(i)$ in (62) with the multi-step conditional sampling $\overline{p}^f(i)$ that, at t-th iteration, excludes the choice of the previous $1 < f \leq r-1$ sampling results, i.e.,

$$\overline{p}^f(i) \triangleq \overline{p}^f(\mathcal{Q}_i) = \frac{p(i)}{1 - p(j) - \ldots - p(l)}, \tag{75}$$

where the indexes $j, \ldots, l$ indicate the multi-step conditional sampling results of iterations $t-1, \ldots, t-f$ respectively. Accordingly, this corresponds to $\mathcal{Q}_i \nsubseteq \{\mathcal{Q}_j, \ldots, \mathcal{Q}_l\}$, implying that there must be no sampling repetitions from the iteration $t-f$ to t. Meanwhile, the maximum value of f is $r-1$ otherwise there would be no sampling candidate for $\mathcal{Q}^t$ when $t > r$.

According to (75), the sampling probabilities $p(i)$ in (13) and $\overline{p}(i)$ in (62) can be viewed as the special cases of $\overline{p}^f(i)$ with $f = 0$ and $f = 1$ respectively, which leads to a generalized multi-step condition sampling with $0 \leq f \leq r-1$. Subsequently, the iteration performance of EDRID with $\overline{p}^f(i)$ can be obtained as follows, where the proof is omitted due to the similarity with Theorem 3.

Theorem 4: As for decentralized MIMO detection, the proposed EDRID algorithm following $\overline{p}^f(i)$ in (75) ensures exact convergence to LS solution by

$$E[\|\mathbf{x}^L - \mathbf{x}^*\|^2] \leq \prod_{t=1}^{L} (1 - \overline{\omega}^f \overline{\alpha}_t^f) \|\mathbf{x}^0 - \mathbf{x}^*\|^2 \tag{76}$$

with dynamic step-size

$$\overline{\alpha}_t^f = \frac{\overline{\omega}^f \overline{g}^f(t-1)}{\lambda_{\max}^* \overline{\omega}^f \overline{g}^f(t-1) + 1}, \tag{77}$$

where

$$\overline{g}^f(t) = \overline{g}^f(t-1)(1 - \overline{\omega}^f \overline{\alpha}_t^f), \tag{78}$$

$$\overline{g}^f(0) = \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{\overline{\eta}^f}, \tag{79}$$

$$\overline{\omega}^f \triangleq \frac{\sigma_{\min}^2(\mathbf{H})}{r-f}, \tag{80}$$

$$\overline{\eta}^f \triangleq E_{\overline{p}^f(i)}[\|\mathbf{H}_i^H \mathbf{I}_{\mathcal{Q}_i,:}(\mathbf{y} - \mathbf{H}\mathbf{H}^\dagger \mathbf{y})\|^2]. \tag{81}$$

From Theorem 4, the exact and exponential convergence of EDRID with $\overline{p}^f(i)$ still can be achieved. Meanwhile, according to the fact that

$$\overline{\omega}^f \overline{\alpha}_t^f = \frac{1}{r-f} \frac{\sigma_{\min}^2(\mathbf{H})}{\lambda_{\max}^*} \frac{\overline{\omega}^f \overline{g}^f(t-1)}{\overline{\omega}^f \overline{g}^f(t-1) + \frac{1}{\lambda_{\max}^*}}, \tag{82}$$

we can readily validate the global and enhanced convergence of EDRID with $\overline{p}^f(i)$.

Corollary 2: Following $\overline{p}^f(i)$ in (75), the proposed EDRID algorithm with dynamic step-size $\overline{\alpha}_t^f$ in (77) achieves the global convergence as its convergence in (76) always holds due to

$$0 < \overline{\omega}^f \overline{\alpha}_t^f < 1. \tag{83}$$

Theorem 5: The exponential convergence rate of EDRID with $\overline{\alpha}_t^f$ improves with the increment of $0 \leq f \leq r-1$ due to

$$0 < \omega\alpha_t < \overline{\omega}^1 \overline{\alpha}_t^1 < \cdots < \overline{\omega}^{r-1} \overline{\alpha}_t^{r-1} < 1. \tag{84}$$

According to Theorem 5, the choice of $f = r-1$ is strongly preferred, which effectively boosts the convergence performance by the largest value of $\overline{\omega}^f$. Typically, when $f = r-1$ is applied, as t increases the sampling of the index i in EDRID would become deterministic gradually due to a shrinking sampling candidate list. More interestingly, once $t > r-1$, there is only one sampling candidate left for each sampling of i (i.e., $\mathcal{Q}_i$), thus completely removing the randomness from EDRID. Undoubtedly, this greatly facilitates the implementation of EDRID in different kinds of decentralized MIMO systems without any performance loss, where both convergence and efficiency gains can be obtained.

Due to the usage of multi-step conditional sampling, no matter what kind of the original sampling probability $p(i)$ is applied, $\overline{p}^{r-1}(i)$ would be become deterministic when $t > r$. Because of it, the uniform distribution becomes an easy choice of $p(i)$, and the detection order (i.e., $\mathbf{H}_i, \mathbf{H}_j, \ldots, \mathbf{H}_l$) is determined only by the sampling in the first round-robin. Typically, under ring architecture, EDRID works by

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \alpha_t \mathbf{H}_i^H (\mathbf{y}_i - \mathbf{H}_i \mathbf{x}^{t-1}) \tag{85}$$

with $i = (t-1) \mod r + 1$. To make it more specific, Fig. 1 illustrates the operations of the proposed EDRID algorithm for XL-MIMO systems. Under ring architecture, each iteration in (85) corresponds to the processing in a single DU, where a full processing by all the DUs requires r iterations, i.e.,

$$\text{DU } 1 \to \text{DU } 2 \to \text{DU } 3 \to \cdots \to \text{DU } r. \tag{86}$$

This is the reason why index $t = rk$ is applied for the fair compassion between EDRID and other distributed schemes, where more details can be found in Table I and simulations. From (85), EDRID can also be readily extended to other distributed topologies by flexibly adjusting the processing order.

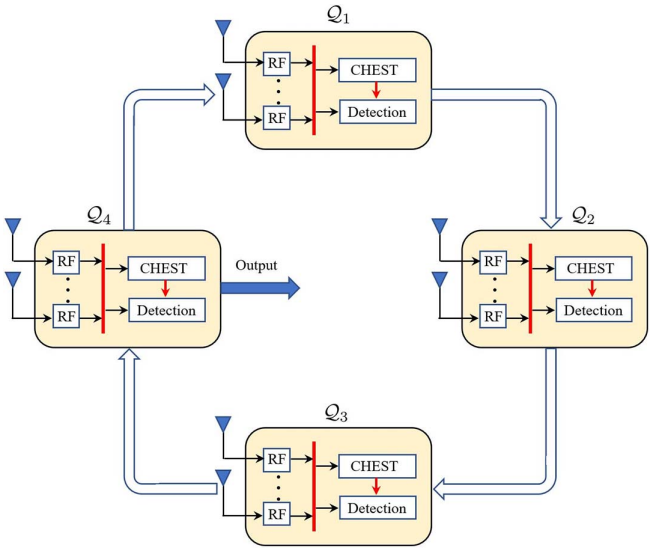

Fig. 1. Illustration of the proposed EDRID algorithm for uplink XL-MIMO with a ring topology consisting of $r = 4$ DUs. Each DU performs local radio-frequency (RF) processing and channel estimation (CHEST).

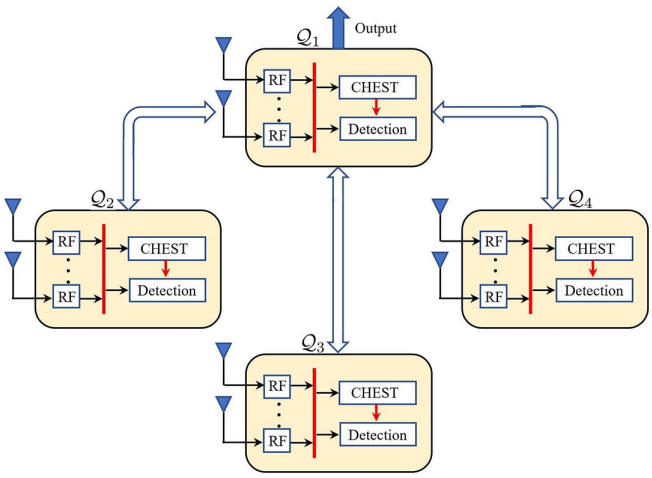

Fig. 2. Illustration of the proposed EDRID algorithm for uplink XL-MIMO, using a star topology with $r = 4$ DUs.

For example, the star topology shown in Fig. 2 can be emulated by the sequential order

$$\text{DU } 1 \rightarrow \text{DU } 2 \rightarrow \text{DU } 1 \rightarrow \text{DU } 3 \rightarrow \cdots \rightarrow \text{DU } 1. \tag{87}$$

However, compared with the ring order in (86), this processing order provides less processing diversity, which in turn leads to additional uplink bandwidth requirements in XL-MIMO systems.

Since each DU can be flexibly added or removed as a system module, the implementation of EDRID enjoys great scalability. Therefore, similar to MCRBK in [31], the proposed EDRID also suits well decentralized MIMO systems like cell-free MIMO and so on but with much lower complexity costs. Note that the data bandwidth required by EDRID is summarized in Table II, and is determined by the total number of complex values exchanged over all links per detection. In contrast to centralized detection, where the full channel matrix $\mathbf{H}$ and received signal $\mathbf{y}$ must be uploaded to CPU, the bandwidth requirement of EDRID is independent of the number of BS antennas, since different DUs communicate with each other only through the estimated vector $\mathbf{x}^t$.

TABLE II
COMPARISONS OF DATA BANDWIDTH

Methods	Data Bandwidth
Centralized Detection	$NK + N$
ADMM [10]	$2Kr \cdot k$
CG [11]	$2Kr \cdot k$
FD-MMSE [16]	$2Kr$
PD-MMSE [12]	$(K^2 + K)r$
CD [15]	$2Kr \cdot k$
DN [23]	$4Kr \cdot k$
RLS [19]	$(K^2 + K)r$
SGD [20]	Kr
ASGD [21]	$2Kr$
SDK [42]	Kr
MCRBK [31]	Krk
EDRID Ring Topology	$Kr \cdot k$
EDRID Star Topology	$2K(r-1) \cdot k$

Algorithm 1 EDRID Algorithm with Dynamic step-size for Uplink XL-MIMO

Require: q, $\mathbf{H}_{\mathcal{Q}_j,:}$ and $\mathbf{y}_{\mathcal{Q}_j}$ for $1 \le j \le r$, $\mathbf{x}^0 = \mathbf{0}$, L
Ensure: detection output $\widehat{\mathbf{x}}^L$
1: **for** $t = 1, \ldots, L$ **do**
2: update $\mathbf{x}^t$ by (85) with step-size α_t in (88)
3: **end for**
4: output $\widehat{\mathbf{x}}^L$ by rounding $\mathbf{x}^L$ based on constellation $\mathcal{X}^K$

As shown in Table I, the proposed EDRID algorithm has much lower complexity cost than the MCRBK algorithm in [31], namely, $\mathcal{O}(Kq \cdot t)$ versus $\mathcal{O}(Kq^2 \cdot t)$, thus leading to significant complexity reduction. Moreover, in terms of index k, we can observe that the complexity of EDRID (i.e., $\mathcal{O}(KN \cdot k)$) is actually insensitive to the antenna number q in each DU. Put it in other way, given N and K, the complexity of EDRID remains unchanged with different q, which enables it more flexible to practical deployment of decentralized MIMO systems. On the other hand, to ease the computation of α_t in (44), following the work in [31], we still take the following choice as an efficient alternative

$$\alpha_t = \frac{4}{N}\left(1 - \frac{K}{N}\right)\frac{N/q + K}{N/q + K + t}. \tag{88}$$

In addition, as illustrated in (43), a better choice of $\mathbf{x}^0$ is also beneficial to the convergence performance while we apply $\mathbf{x}^0 = \mathbf{0}$ in this work. To sum up, the proposed EDRID algorithm with dynamic step-size is shown in Algorithm 1.

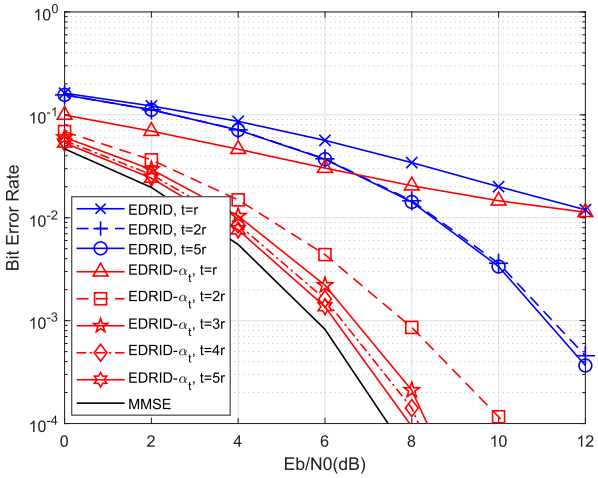

Fig. 3. BER Illustrations of EDRID and EDRID with dynamic α_t in 256×64 XL-MIMO systems with $q = 8$ using 16-QAM.

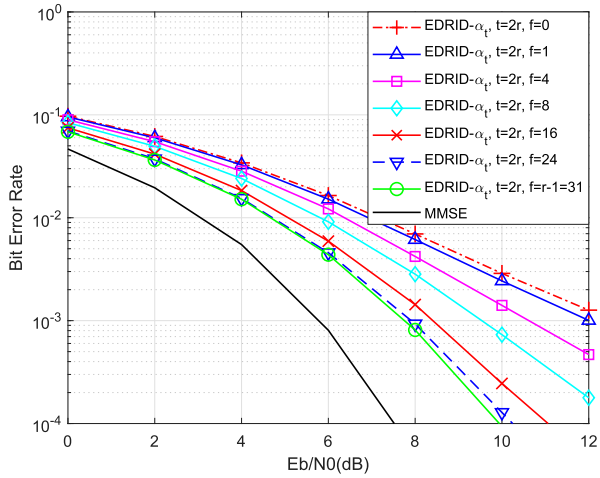

Fig. 4. BER Illustrations of EDRID with dynamic α_t under different f in 256×64 XL-MIMO systems with $q = 8$ using 16-QAM.

VI. Simulations

In this section, the performance and complexity of the proposed EDRID algorithm are thoroughly investigated through simulations for uplink XL-MIMO systems over Rayleigh fading channels.

Fig. 3 presents a comparison of the bit error rate (BER) performance between the proposed EDRID and EDRID with dynamic step-size α_t in 256×64 XL-MIMO systems using 16-QAM. The centralized MMSE is presented as a performance benchmark in the following simulations. As illustrated, EDRID rapidly approaches its performance upper bound, yet a substantial performance gap does exist compared to the MMSE detection. As shown in Theorem 1, such a performance obstacle is due to the fact that the standard EDRID fails to converge to the LS solution. In comparison, EDRID with dynamic step-size α_t (denoted by EDRID-α_t) demonstrates superior BER performance over the standard EDRID. Note that if not particularly indicated, both the standard EDRID and EDRID-α_t in simulations are enhanced by the multi-step conditional sampling with $f = r - 1$. With the increment of iterations, the detection performance of EDRID-α_t gradually converges to that of MMSE, which is consistent with the convergence results in Theorem 4. Therefore, compared to standard EDRID, EDRID with dynamic step-size α_t turns out to be a much better solution for decentralized MIMO systems.

On the other hand, to illustrate the impact of f in the multi-step conditional sampling, the same XL-MIMO configuration is still considered in Fig. 4 with different sizes of $f = 0, 1, 4, 8, 16, 24, 31$. With the increment of f, the detection performance of EDRID-α_t gradually improves. This is accordance with Theorem 5 since more knowledge of previous iterations have been taken into account to serve for the current random iteration. Benefiting from it, both the iteration efficiency and convergence are significantly enhanced as a result, making $f = r - 1$ a preferred choice to EDRID-α_t.

Fig. 5 compares the BER performance of different distributed detection schemes in 256×32 XL-MIMO systems with 16-QAM. Specifically, besides the proposed EDRID with dynamic step-size α_t, the FD-MMSE in [16], the FD-CD in [15], the DBP-ADMM in [10], the DBP-CG in [11], the RLS in [19], the SGD in [20], the ASGD in [21], the DN in [23], the SDK in [42], and the MCRBK-α_t in [31] are applied for the fair comparison. For EDRID-α_t, MCRBK-α_t, ADMM, CG, FD-MMSE, and DN, each DU has 8 antennas, i.e., $q = 8$, while RLS, SGD, ASGD, and SDK adopt the $q = 1$ configuration as their default setup. As demonstrated in Fig. 5, under the same iterations $k = 3$, EDRID-α_t outperforms all distributed detection schemes except RLS. Although RLS achieves equivalent detection performance with MMSE in a distributed manner, it has much higher computational complexity and data bandwidth requirements than EDRID-α_t, where the details can be found in Tables I and II. Consistent with RLS, PD-MMSE [12] also achieves MMSE-equivalent detection performance with high computational complexity, which is omitted in simulations. Although SDK achieves comparable detection performance with EDRID-α_t, it requires each DU only with one antenna, which is inapplicable in most cases of interest. Note that the detection schemes such as FD-CD and DN fail to work in this scenario as they require extra conditions for the distributed implementation.

Fig. 6 extends the BER performance comparison among different distributed detection schemes to 256×64 XL-MIMO systems with 16-QAM, where $q = 8$ is applied. Notably, 256×64 MIMO system provides less receive diversity gain than a 256×32 MIMO, which results in degraded BER performance for all detection schemes including MMSE. As can be seen clearly, ADMM is sensitive to this configuration change, demonstrating inferior results compared to EDRID-α_t with the same number of iterations. On the other hand, besides the complexity reduction, the proposed EDRID-α_t scheme also achieves better detection performance than MCRBK-α_t, which realizes an advanced detection trade-off between performance and complexity. Meanwhile, the improved detection performance of EDRID-α_t can be observed with the increasing iterations, which is in line with the convergence results in Theorem 4. Additionally, the standard EDRID is also added to the comparison as well. However, as expected, its detection

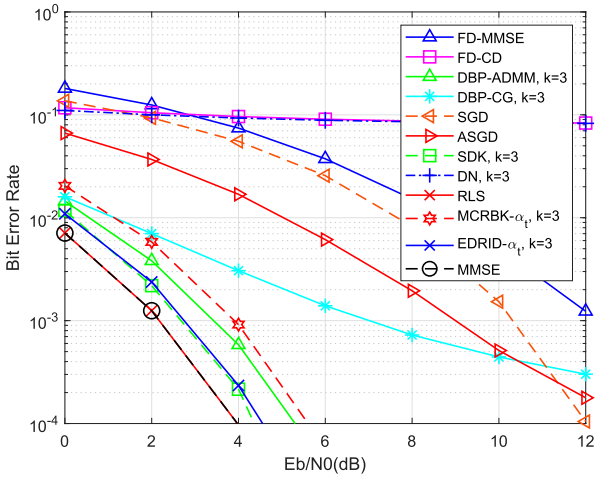

Fig. 5. BER comparison in 256×32 XL-MIMO systems with $q = 8$ using 16-QAM.

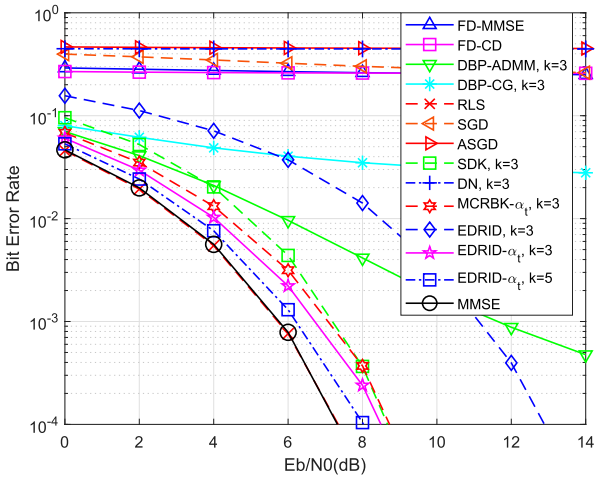

Fig. 6. BER comparison in 256×64 XL-MIMO systems with $q = 8$ using 16-QAM.

performance is not that competitive because it suffers from the convergence bottleneck shown in Theorem 1.

Apart from the independent, identically distributed (i.i.d.) Rayleigh channels, the correlated channels of XL-MIMO systems are also studied to illustrate the convergence performance of EDRID. Typically, following the configurations in [52], [53], the correlated channel matrix $\mathbf{R}_{\text{cor}}^{\frac{1}{2}}\mathbf{H}\mathbf{T}_{\text{cor}}^{\frac{1}{2}}$ is applied with the normalized correlation coefficient $1 \geq \psi \geq 0$, where $\mathbf{R}_{\text{cor}} \in \mathbb{C}^{N\times N}$ and $\mathbf{T}_{\text{cor}} \in \mathbb{C}^{K\times K}$ represent the receive and transmit correlation matrices respectively, i.e.,

$$\mathbf{R}_{\text{cor}} = \begin{bmatrix} 1 & \psi & \psi^4 & \cdots & \psi^{(N-1)^2} \\ \psi & 1 & \psi & \cdots & \vdots \\ \psi^4 & \psi & 1 & \cdots & \psi^4 \\ \vdots & \vdots & \vdots & \ddots & \psi \\ \psi^{(N-1)^2} & \cdots & \psi^4 & \psi & 1 \end{bmatrix},$$

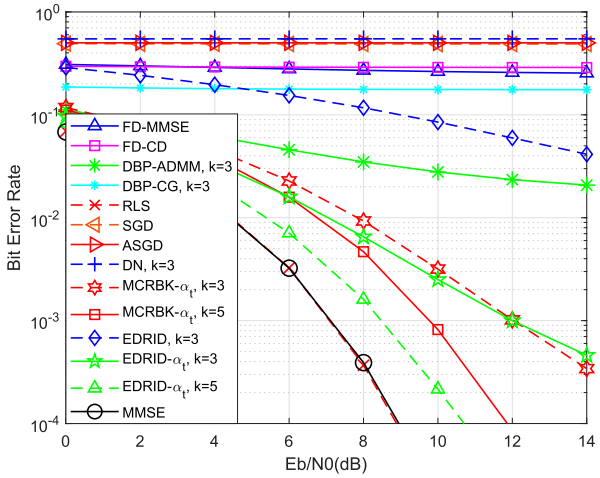

Fig. 7. BER comparison in 256×64 XL-MIMO systems with $q = 8$ using 16-QAM with $\psi = 0.05$.

$$\mathbf{T}_{\text{cor}} = \begin{bmatrix} 1 & \psi & \psi^4 & \cdots & \psi^{(K-1)^2} \\ \psi & 1 & \psi & \cdots & \vdots \\ \psi^4 & \psi & 1 & \cdots & \psi^4 \\ \vdots & \vdots & \vdots & \ddots & \psi \\ \psi^{(K-1)^2} & \cdots & \psi^4 & \psi & 1 \end{bmatrix}.$$

Specifically, an uncorrelated scenario entails $\psi = 0$ and a completely correlated case corresponds to $\psi = 1$. Intuitively, with $\psi = 0.05$ in Fig. 7, the BER performance of all the detection degrade accordingly compared to the results given in Fig. 6. Clearly, detection schemes based on random iteration enjoy the global convergence so that their convergence are still guaranteed.

In Fig. 8, the BER performance comparison among different distributed detection schemes is presented for 512×64 XL-MIMO systems using 16-QAM. In this case, the size $q = 16$ is applied in schemes of ADMM, CG, FD-MMSE, DN, MCRBK-α_t, and EDRID-α_t. Intuitively, 512×64 MIMO provides a higher receive diversity gain than the 256×64 case in Fig. 6, resulting in the enhanced BER performance for all detection schemes. Unfortunately, the schemes DN, ASGD, SGD, FD-CD, FD-MMSE that fail to work in Fig. 6 still remain ineffective. Meanwhile, competitive BER performance is achieved by EDRID-α_t with significant improvements observed over ADMM, standard EDRID, and MCRBK-α_t.

Fig. 9 evaluates the BER performance for 512×128 XL-MIMO systems using 16-QAM, where $q = 16$ is applied in schemes of ADMM, CG, MCRBK-α_t, and EDRID-α_t. We notice that in this case, both ADMM and CG detection schemes suffer from the severe performance degradation, where the similar observations can also be found in the case of 256×64 shown in Fig. 6. On the contrary, thanks to global and exponential convergence, the proposed EDRID-α_t not only outperforms all other distributed schemes under the same iterations $k = 3$, but also attains the near-MMSE performance with the increment of iterations. We point out that the performance superiority of EDRID-α_t over MCRBK-α_t can still be found, making it a better choice for decentralized MIMO detection. Fig. 10 extends

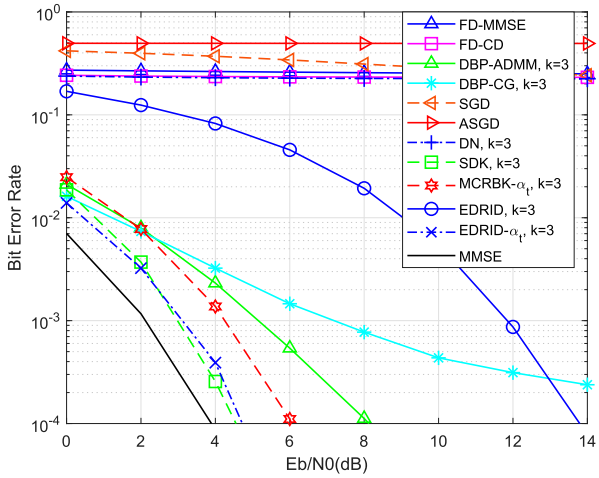

Fig. 8. BER comparison in 512×64 XL-MIMO systems with $q = 16$ using 16-QAM.

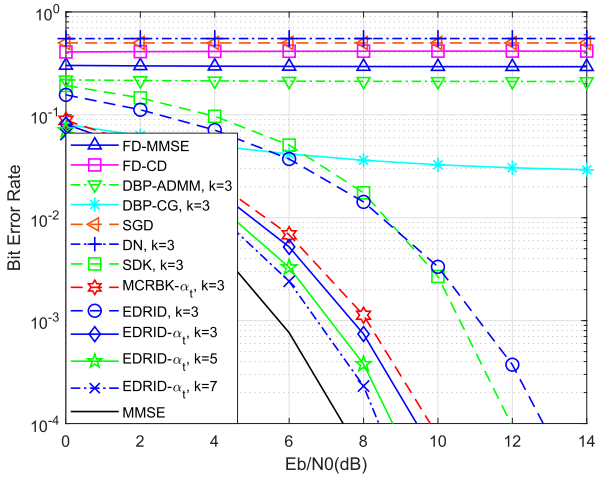

Fig. 10. BER comparison in 1024×256 XL-MIMO systems with $q = 32$ using 16-QAM.

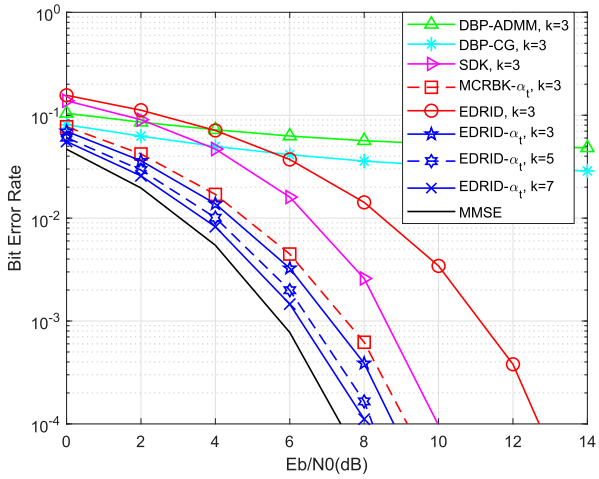

Fig. 9. BER comparison in 512×128 XL-MIMO systems with $q = 16$ using 16-QAM.

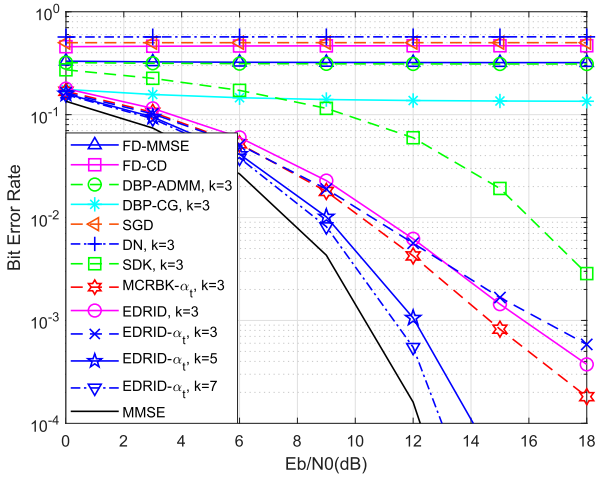

Fig. 11. BER comparison in 512×256 XL-MIMO systems with $q = 32$ using 16-QAM.

the BER performance comparison to 1024×256 XL-MIMO systems using 16-QAM, where $q = 32$ is applied in schemes of ADMM, CG, FD-MMSE, DN, MCRBK-α_t, and EDRID-α_t. Note that in this case, the schemes that fail to work in Fig. 9 still turn out to be ineffective as well. On the other hand, as expected, the proposed EDRID-α_t is able to achieve the near-MMSE performance with a moderate size k.

To further investigate the impact of the system dimension upon the detection performance, in Fig. 11 the BER performance comparison for 512×256 XL-MIMO systems using 16-QAM with $q = 32$ is presented. Notably, with the antenna ratio $N/K = 2$, most distributed detection schemes including SDK experience severely degraded performance. In contrast, EDRID-α_t still works as usual, with its BER performance showing progressive improvement as the iteration increases. Interestingly, in this case, we find that MCRBK-α_t achieves a better detection performance than EDRID-α_t under the same iterations $k = 3$, which is quite different from the simulation results given in other cases. Nevertheless, as shown in Table I, we emphasize that under the same sizes of $k = 3$ and $q = 32$, the complexity cost of MCRBK-α_t (i.e., $\mathcal{O}(KNq \cdot k)$) is 32 times as much as that of EDRID-α_t (i.e., $\mathcal{O}(KN \cdot k)$).

Fig. 12 studies the impact of the parameter q on the detection performance of the proposed EDRID-α_t algorithm, where a 512×128 XL-MIMO system with 16-QAM is applied here. Specifically, given the fixed total number of received antennas $N = 512$, by $N = qr$, different setups of the number of antennas in each DU q and the number of DUs r are employed, where a smaller q accounts for a larger r. As can be seen from Fig. 12, under the same iterations $k = 3$, with $r = 8, 16, 32, 64, 128$, the increased number of DUs leads to the improved detection performance, which corresponds to $q = 64, 32, 16, 8, 4$ respectively. From it, in terms of BER performance, we can see that the large r or small q (more number of DUs) is preferred as it means more iterations (i.e., $t = rk$ in total) are invoked in EDRID even though each iteration involves small sizes of $\mathbf{H}_i$ and $\mathbf{y}_i$. Put it in another way, for large q (i.e., the number of antennas at each DU is large), the number of recalling the randomized iterations is limited due to the small r, such that the performance potential cannot be fully exploited. Nevertheless, more latency over large

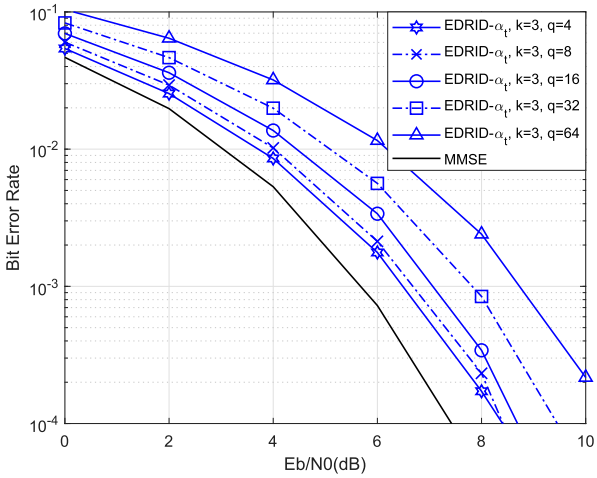

Fig. 12. BER comparison in 512×128 XL-MIMO systems with $q = 4, 8, 16, 32, 64$ using 16-QAM.

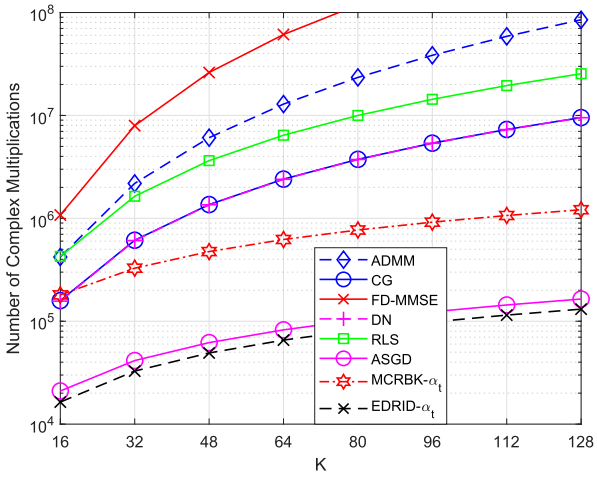

Fig. 13. Complexity comparison in terms of the number of complex multiplications for $512 \times K$ XL-MIMO systems with $q = 8$ and $k = 1$.

number of DUs also will be incurred for large r and small q, thus resulting in a trade-off therein. Most importantly, the deployment of DUs in practice has been determined physically, which can not be easily altered.

Fig. 13 presents the computational complexity comparison of various distributed detection schemes for $512 \times K$ massive MIMO systems employing 16-QAM, where $q = 8$ is implemented in schemes of ADMM, CG, FD-MMSE, DN, MCRBK-α_t, and EDRID-α_t. Specifically, complexity evaluation is based on the number of complex multiplications required per single iteration (i.e., $k = 1$ or $t = r$). As illustrated, FD-MMSE, ADMM, and RLS demand prohibitively high computational costs for practical implementation. Meanwhile, CG and DN have similar complexity levels, which are substantially higher than MCRBK-α_t, ASGD, and EDRID-α_t schemes. Although ASGD demonstrates computational complexity comparable to EDRID-α_t, its application are fundamentally constrained to decentralized MIMO systems with $q = 1$. Compared to MCRBK-α_t, EDRID-α_t achieves both lower computational complexity and better BER performance (except the case of 512×256 MIMO system with $N/K = 2$), making it a better solution for decentralized XL-MIMO.

VII. Conclusion

In this paper, we proposed an efficient distributed randomized iterative detection (EDRID) algorithm to enable scalable and low-complexity signal detection in decentralized XL-MIMO systems. The EDRID algorithm eliminates the need for matrix pseudoinverse computations within its randomized iterations, thereby significantly reducing computational complexity. Then, for achieving the exact convergence to the LS detection solution, the strategy of dynamic step-size has been adopted into EDRID, such that the linear detection performance can be obtained with exponential and global convergence. Moreover, by introducing the conditional sampling technique, both the convergence performance and efficiency of EDRID can be further improved. Overall, the proposed EDRID algorithm provides a low-complexity, flexible, and scalable solution for different cases of distributed detection, making it perspective in decentralized MIMO systems.

Appendix A
Proof of $[\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{H}_i^H\mathbf{H}_i - (\mathbf{H}_i^H\mathbf{H}_i)^2] \succeq \mathbf{0}$

Proof: First of all, the symmetric matrix $\mathbf{H}_i^H\mathbf{H}_i$ is positive semidefinite because of

$$\mathbf{w}^H\mathbf{H}_i^H\mathbf{H}_i\mathbf{w} = (\mathbf{H}_i\mathbf{w})^H\mathbf{H}_i\mathbf{w} \geq 0 \tag{89}$$

for any vector $\mathbf{w} \in \mathbb{C}^K$.

Next, let λ_i be an eigenvalue of $\mathbf{H}_i^H\mathbf{H}_i$ with corresponding eigenvector $\mathbf{v}_i$, we have

$$\mathbf{H}_i^H\mathbf{H}_i\mathbf{v}_i = \lambda_i\mathbf{v}_i. \tag{90}$$

Then, given the symmetric matrix $\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{I} - \mathbf{H}_i^H\mathbf{H}_i$, we can obtain

$$[\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{I} - \mathbf{H}_i^H\mathbf{H}_i]\mathbf{v}_i = [\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i) - \lambda_i]\mathbf{v}_i \tag{91}$$

which implies that $\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i) - \lambda_i$ is an eigenvalue of matrix $\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{I} - \mathbf{H}_i^H\mathbf{H}_i$. Furthermore, since $\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i) - \lambda_i \geq 0$ always holds, all the eigenvalues of the symmetric matrix $\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{I} - \mathbf{H}_i^H\mathbf{H}_i$ are non-negative. From it, we can conclude that the matrix $\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{I} - \mathbf{H}_i^H\mathbf{H}_i$ is positive semidefinite.

In the following, considering the two symmetric positive semidefinite matrices $\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{I} - \mathbf{H}_i^H\mathbf{H}_i$ and $\mathbf{H}_i^H\mathbf{H}_i$, it is straightforward to verify that the product of them is still symmetric positive semidefinite, i.e.,

$$\mathbf{z}^H[(\lambda_{\max}(\mathbf{H}_i^H\mathbf{H}_i)\mathbf{I} - \mathbf{H}_i^H\mathbf{H}_i)\mathbf{H}_i^H\mathbf{H}_i]\mathbf{z} \geq 0 \tag{92}$$

for any vector $\mathbf{z}$, which completes the proof. □

REFERENCES

[1] I. Tomkos, D. Klonidis, E. Pikasis, and S. Theodoridis, "Toward the 6G network era: Opportunities and challenges," *IT Prof.*, 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 Commun.*, vol. 27, no. 4, pp. 118–125, Aug. 2020.

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

[4] C.-X. Wang et al., "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, 2nd Quart. 2023.

[5] Y. Gao et al., "Artificial intelligence enabled joint channel estimation and signal detection for massive MIMO systems," *Chin. J. Electron.*, vol. 35, no. 1, pp. 178–195, 2026.

[6] M. A. Albreem, A. Alhabbash, A. M. Abu-Hudrouss, and T. A. Almohamad, "Data detection in decentralized and distributed massive MIMO networks," *Comput. Commun.*, vol. 189, pp. 79–99, 2022.

[7] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3109–3132, 4th Quart. 2019.

[8] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.

[9] J. R. Sánchez, J. Vidal Alegría, and F. Rusek, "Decentralized massive MIMO systems: Is there anything to be discussed?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2019, pp. 787–791.

[10] K. Li, R. R. Sharan, Y. Chen, T. Goldstein, J. R. Cavallaro, and C. Studer, "Decentralized baseband processing for massive MU-MIMO systems," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 7, no. 4, pp. 491–507, Dec. 2017.

[11] 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 *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, 2016, pp. 468–472.

[12] C. Jeon, K. Li, J. R. Cavallaro, and C. Studer, "Decentralized equalization with feedforward architectures for massive MU-MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4418–4432, Sep. 2019.

[13] C. Jeon, K. Li, J. R. Cavallaro, and C. Studer, "On the achievable rates of decentralized equalization in massive MU-MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2017, pp. 1102–1106.

[14] K. Li, C. Jeon, J. R. Cavallaro, and C. Studer, "Decentralized equalization for massive MU-MIMO on FPGA," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, 2017, pp. 1532–1536.

[15] K. Li, O. Castañeda, C. Jeon, J. R. Cavallaro, and C. Studer, "Decentralized coordinate-descent data detection and precoding for massive MU-MIMO," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2019, pp. 1–5.

[16] K. Li et al., "Design trade-offs for decentralized baseband processing in massive MU-MIMO systems," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.,* 2019, pp. 906–912.

[17] 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 J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 9–39, Jan. 2023.

[18] F. Jiande, X. Weixin, and L. Zongxiang, "A low complexity distributed multitarget detection and tracking algorithm," *Chin. J. Electron.*, vol. 32, no. 3, pp. 429–437, 2023.

[19] J. R. Sanchez, F. Rusek, M. Sarajlic, O. Edfors, and L. Liu, "Fully decentralized massive MIMO detection based on recursive methods," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, 2018, pp. 53–58.

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

[21] Q. Liu, H. Liu, Y. Yan, and P. Wu, "A distributed detection algorithm for uplink massive MIMO systems," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, 2019, pp. 213–217.

[22] Q. Chen, Z. Wang, C. Qi, Z. Gao, Y. Huang, and D. Niyato, "Decentralized likelihood ascent search-aided detection for distributed large-scale MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 24, no. 5, pp. 4160–4173, May 2025.

[23] A. Kulkarni, M. A. Ouameur, and D. Massicotte, "Hardware topologies for decentralized large-scale MIMO detection using Newton method," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 9, pp. 3732–3745, Sep. 2021.

[24] 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 Trans. Veh. Technol.*, vol. 71, no. 2, pp. 2152–2157, Feb. 2022.

[25] Z. Zhang, H. Li, Y. Dong, X. Wang, and X. Dai, "Decentralized signal detection via expectation propagation algorithm for uplink massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11233–11240, Oct. 2020.

[26] Q. Chen, Z. Wang, C. Ma, X. Dai, and D. W. K. Ng, "General recursive least square algorithm for distributed detection in massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 73, no. 8, pp. 12137–12142, Aug. 2024.

[27] 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 Trans. Veh. Technol.*, vol. 70, no. 8, pp. 7665–7680, Aug. 2021.

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

[29] X. Zhao, X. Guan, M. Li, and Q. Shi, "Decentralized linear MMSE equalizer under colored noise for massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 1–6.

[30] X. Pan, Z. Zheng, X. Huang, and Z. Fei, "On the uplink distributed detection in UAV-enabled aerial cell-free mMIMO systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 13812–13825, Oct. 2024.

[31] Z. Wang, C. Pan, Y. Huang, S. Jin, and G. Caire, "Randomized iterative algorithms for distributed massive MIMO detection," *IEEE Trans. Signal Process.*, vol. 73, pp. 2304–2319, 2025.

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

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

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

[35] X. Yang, F. Cao, M. Matthaiou, and S. Jin, "On the uplink transmission of extra-large scale massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15229–15243, Dec. 2020.

[36] 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 Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6788–6799, Jul. 2021.

[37] A. Amiri, C. N. Manchón, and E. de Carvalho, "Uncoordinated and decentralized processing in extra-large MIMO arrays," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 81–85, Jan. 2022.

[38] 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 *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 1–6.

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

[40] N. Li and P. Fan, "Distributed cell-free massive MIMO versus cellular massive MIMO under UE hardware impairments," *Chin. J. Electron.*, vol. 33, no. 5, pp. 1274–1285, 2024.

[41] W. Yu, Y. Ma, H. He, S. Song, J. Zhang, and K. B. Letaief, "Deep learning for near-field XL-MIMO transceiver design: Principles and techniques," *IEEE Commun. Mag.*, vol. 63, no. 1, pp. 52–58, 2025.

[42] 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 *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, 2021, pp. 1242–1249.

[43] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[44] 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 Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 611–652, 1st Quart. 2022.

[45] Z. Wang et al., "Extremely large-scale MIMO: Fundamentals, challenges, solutions, and future directions," *IEEE Wireless Commun.*, vol. 31, no. 3, pp. 117–124, Jun. 2024.
[46] A. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends*, pp. 1–190, 2004. Available: https://ieeexplore.ieee.org/book/8187216
[47] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2013.
[48] Z. Wang and C. Ling, "Lattice Gaussian sampling by Markov chain Monte Carlo: Bounded distance decoding and trapdoor sampling," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3630–3645, Jun. 2019.
[49] Z. Wang, R. M. Gower, Y. Xia, L. He, and Y. Huang, "Randomized iterative methods for low-complexity large-scale MIMO detection," *IEEE Trans. Signal Process.*, vol. 70, pp. 2934–2949, 2022.
[50] Z. Wang, W. Xu, Y. Xia, Q. Shi, and Y. Huang, "A new randomized iterative detection algorithm for uplink large-scale MIMO systems," *IEEE Trans. Commun.*, vol. 71, no. 9, pp. 5093–5107, Sep. 2023.
[51] Z. Wang, C. Ling, S. Jin, Y. Huang, and F. Gao, "Probabilistic searching for MIMO detection based on lattice Gaussian distribution," *IEEE Trans. Commun.*, vol. 72, no. 1, pp. 85–100, Jan. 2024.
[52] B. Costa, A. Mussi, and T. Abrao, "MIMO detectors under correlated channels," *Semina: Ciencias Exatas e Tecnologicas*, vol. 37, no. 1, pp. 3–12, 2016.
[53] R. De Lamare and R. Sampaio-Neto, "Detection estimation algorithms in massive MIMO systems, " Aug. 2014, *arXiv:1408.4853.*

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 the University of Manchester, Manchester, U.K., in 2010. He received the Ph.D. degree in communication engineering from Imperial College London, U.K., in 2015. Since 2021, he has been an Associate Professor with the School of Information and Engineering, Southeast University, Nanjing, China. From 2015 to 2016, he served as a Research Associate with Imperial College London, U.K. From 2016 to 2017, he was a Senior Engineer with the Radio Access Network R&D division, Huawei Technologies Company. From 2017 to 2020, he was an Associate Professor with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. In 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 Company in 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 research interests include massive MIMO systems, machine learning and data analytic over wireless networks, and lattice theory for wireless communications.

Chunguo Li (Senior Member, IEEE) received the bachelor's degree in wireless communications from Shandong University in 2005, and the Ph.D. degree in wireless communications from the Southeast University in 2010. In 2010, he joined as a Faculty Member with the Southeast University, where he was an Associate Professor from 2012 to 2016 and a Full Professor since 2017. From 2012 to 2013, he was the Postdoctoral Researcher with Concordia University, Montreal, Canada. From 2013 to 2014, he was with DSL laboratory of Stanford University as a Visiting Associate Professor. From 2017 to 2019, he was the Adjunct Professor with Xizang Minzu University under the supporting Tibet program organized by China National Human Resources Ministry. He is the fellow of IET, fellow of China Institute of Communications (CIC), Chair of IEEE Computational Intelligence Society Nanjing Chapter, and Chair of Advisory Committee for Instruments industry in Jiangsu province. He is currently the Subject Editor for *Journal of The Franklin Institute* since 2025, and is serving as the Editor for IEEE TRANSACTIONS ON COMMUNICATIONS since 2025. He has served as an Associate Editor for a couple of international journals and as session chair for many international conferences. His research interests include cell-free massive MIMO networks, AI-based wireless communications, information theories, and AI-based audio signal processing.

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 the Southeast University, Nanjing, China, in 2007. Since 2007, he has been a Faculty with the School of Information Science and Engineering, Southeast University, where he is currently a Full Professor. From 2008 to 2009, he visited the Signal Processing Laboratory, 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 IEEE TRANSACTIONS ON SIGNAL PROCESSING and a Guest Editor of IEEE JOURNAL SELECTED AREAS IN COMMUNICATIONS. He is currently an Editor-at-Large of IEEE Open Journal of the Communications Society and an Associate Editor for 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 the Southeast University, Nanjing, in 2007. From 2007 to 2009, he was a Research Fellow with 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 coauthors 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*.

Giuseppe Caire (Fellow, IEEE) was born in Torino, in 1965. He received the B.Sc. degree in electrical engineering from the Politecnico di Torino in 1990, the M.Sc. degree in electrical engineering from Princeton University in 1992, and the Ph.D. degree from the Politecnico di Torino in 1994. He has been a Postdoctoral Research Fellow with the European Space Agency (ESTEC, Noordwijk, The Netherlands) in 1994–1995, an Assistant Professor in telecommunications with the Politecnico di Torino, an Associate Professor with the University of Parma, Italy, Professor with the Department of Mobile Communications, Eurecom Institute, Sophia-Antipolis, France, a Professor of electrical engineering with 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, 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 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. He has served in the Board of Governors of the IEEE Information Theory Society from 2004 to 2007, and as an Officer from 2008 to 2013. He was the President of the IEEE Information Theory Society in 2011. His research interests include field of communications theory, information theory, channel and source coding with particular focus on wireless communications.