



Residual-Based Detections and Unified Architecture for Massive MIMO Uplink

Chuan Zhang¹ · Yufeng Yang² · Shunqing Zhang³ · Zaichen Zhang¹ · Xiaohu You¹

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Massive multiple-input multiple-output (M-MIMO) technique brings better energy efficiency and coverage but higher computational complexity than small-scale MIMO. For linear detections such as minimum mean square error (MMSE), prohibitive complexity lies in solving large-scale linear equations. For a better trade-off between bit-error-rate (BER) performance and computational complexity, iterative linear algorithms like conjugate gradient (CG) have been applied and have shown their feasibility in recent years. In this paper, residual-based detection (RBD) algorithms are proposed for M-MIMO detection, including minimal residual (MINRES) algorithm, generalized minimal residual (GMRES) algorithm, and conjugate residual (CR) algorithm. RBD algorithms focus on the minimization of residual norm per iteration, whereas most existing algorithms focus on the approximation of exact signal. Numerical results have shown that, for 64-QAM 128×8 MIMO, RBD algorithms are only 0.13 dB away from the exact matrix inversion method when $\text{BER} = 10^{-4}$. Stability of RBD algorithms has also been verified in various correlation conditions. Complexity comparison has shown that, CR algorithm require 87% less complexity than the traditional method for 128×60 MIMO. The unified hardware architecture is proposed with flexibility, which guarantees a low-complexity implementation for a family of RBD M-MIMO detectors.

Keywords Massive MIMO · Residual-based detection · Minimal residual · Conjugate residual · Unified hardware

1 Introduction

Multiple-input multiple-output (MIMO) is a key technique for wireless communications [1] and has been incorporated into standards such as the 3rd generation partnership project (3GPP) long term evolution (LTE) and IEEE 802.11n [2]. By equipping hundreds of antennas at transmitters and serving relatively a small number of users [3], its advanced version massive MIMO (M-MIMO) provides significant improvement in spectral efficiency, interference reduction, transmit-power efficiency, and link reliability [4].

Because of the large antenna number at base station (BS) or user side, computational complexity becomes unaffordable in M-MIMO detection. Among existing detections, zero forcing (ZF) is a basic way, which neglects the effect of noise [5]. However, its performance is not satisfactory.

Though linear schemes like minimum mean square error (MMSE) [6] improve the performance compared with ZF, its computation complexity still increases drastically as the antenna number grows. For a M-MIMO channel \mathbf{H} , computational complexity of MMSE inversion is $\mathcal{O}(M^3)$, which makes it costly in applications [7]. To avoid matrix inversion, *Neumann* series expansion (NSE) [8–10] has been employed for approximation. However, complexity remains unaffordable when NSE terms become more than 2. Thus, iterative linear solvers are proposed for further reduction, such as Gauss-Seidel [11, 12] and conjugate gradient (CG) [13, 14]. Methods like successive over-relaxation (SOR) [15, 16] and its variation [17] are also considered. Meanwhile, efficient optimizations of algorithms are also proposed like precondition [18, 19]. Most iterative linear detectors can reduce MMSE's complexity to $\mathcal{O}(M^2)$ with tolerable performance loss.

It is worth noting that existing algorithms mainly focus on approximating exact solution [20], whereas this paper proposes residual-based detection (RBD) algorithms which focus on the minimization of residual norm. Firstly, minimal residual (MINRES) algorithm, which is a basic RBD, is considered. Its extended version, generalized minimal residual (GMRES) by [21] is also considered, with

Chuan Zhang and Yufeng Yang contributed equally to this work.

✉ Chuan Zhang
chzhang@seu.edu.cn

Extended author information available on the last page of the article.

unavoidable drawbacks, which will be detailed below. In M-MIMO scenario, GMRES can be derived into another version: conjugate residual (CR). Computation process and convergence proof of each RBD algorithm are elaborated. Numerical results under different antenna configurations and correlations are given as well. For lower complexity, the iteration number is chosen as 2, 3, and 4, respectively. Complexity comparison among proposed RBD algorithms and the traditional one is also shown, to demonstrate RBD algorithms' advantages in performance and complexity.

For application considerations, efficient hardware architectures of RBD algorithms are required. In this paper, hardware architectures of MINRES and CR algorithms are proposed. However, as a family of algorithm, computation similarity can be referred and accordingly a unified design method is proposed. The hardware architecture can be given by two common modules: iterative module and coefficient module. Unified hardware architecture for both MINRES and CR algorithms is further proposed, which can also take care of GMRES. Moreover, the proposed design method can be also applied in some RBD and other iterative detectors.

The remainder of the paper is organized as follows. Section 2 gives the system models of non- and correlated M-MIMO detectors. Section 3 employs RBD algorithms and shows the convergence proof for each algorithm. Numerical results are given in Section 4. Section 5 elaborates the computational complexity of RBD algorithms. Section 6 proposes the hardware architectures of RBD algorithms and the unified design method. Finally, Section 7 concludes the entire paper.

Notation The lowercase and upper bold face letters stand for column vector and matrix, respectively. The operations $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate transpose, respectively. The entry in the i -th row and j -th column of \mathbf{A} is $\mathbf{A}(i, j)$. The vector α in the k -th iteration is α_k . Complexity is denoted in terms of complex-valued multiplication number.

2 System Model for Massive MIMO Uplink

2.1 Linear Detection Model

Consider an uplink of a massive MIMO system with N antennas at the base station (BS), which simultaneously serves M single antenna users. Here, N is always much bigger than M ($N \gg M$). The transmitted signal and received vectors are denoted by $\mathbf{s} = [s_1, s_2, \dots, s_M]^T$ and $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$, respectively, where $\mathbf{s} \in \mathbb{C}^M$, $\mathbf{y} \in \mathbb{C}^N$. Then the system model is

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (1)$$

where \mathbf{H} is an $N \times M$ uplink channel matrix, \mathbf{n} is the vector representing Additive White Gaussian Noise (AWGN) with zero-mean and variance σ^2 .

According to MMSE equalization scheme, at the BS side, the estimate of the transmitted symbol vector $\hat{\mathbf{s}}$ is

$$\hat{\mathbf{s}} = (\mathbf{H}^H \mathbf{H} + \sigma^2 \mathbf{I}_M)^{-1} \mathbf{H}^H \mathbf{y} = \mathbf{A}^{-1} \tilde{\mathbf{y}}, \quad (2)$$

where the matrix \mathbf{I} means identity matrix with dimension M , and the MMSE filtering matrix \mathbf{A} is defined based on Gram matrix \mathbf{G} :

$$\mathbf{A} = \mathbf{G} + \sigma^2 \mathbf{I}_M, \quad (3)$$

where $\mathbf{G} = \mathbf{H}^H \mathbf{H}$.

Correspondingly, output of matched filter $\tilde{\mathbf{y}}$ is

$$\tilde{\mathbf{y}} = \mathbf{H}^H \mathbf{y}. \quad (4)$$

Nevertheless, computational complexity of exact matrix inversion \mathbf{A}^{-1} is $\mathcal{O}(M^3)$. Methods such as *Cholesky* decomposition based method are not suitable for M-MIMO detection when its scale increases.

2.2 Correlated Channel Model

Consider correlation of antennas for M-MIMO, this paper applies *Kronecker* model in [22] and \mathbf{H} can be denoted by $\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{W} \mathbf{R}_t^{1/2}$, where $\mathbf{W} \in \mathbb{C}^{N \times M}$ is an $N \times M$ i.i.d. channel matrix with zero mean and unit variance. Meanwhile $\mathbf{R}_r \in \mathbb{C}^{N \times N}$ and $\mathbf{R}_t \in \mathbb{C}^{M \times M}$ are spatial correlation matrices at BS and user side:

$$\mathbf{R}_r(i, k) = \begin{cases} (\zeta_r e^{j\theta})^{k-i}, & i \leq k, \\ \mathbf{R}_r^*(k, i), & i > k; \end{cases} \quad (5)$$

$$\mathbf{R}_t(i, k) = \begin{cases} (\zeta_t e^{j\theta})^{k-i}, & i \leq k, \\ \mathbf{R}_t^*(k, i), & i > k. \end{cases} \quad (6)$$

The i -th row and k -th column is denoted by $\mathbf{R}(i, k)$. \mathbf{R}_t^* and \mathbf{R}_r^* are conjugate matrices of \mathbf{R}_t and \mathbf{R}_r , respectively. This paper contains four scenarios of correlation condition to elaborate common M-MIMO detectors.

- *Uncorrelated*: In this condition, correlations of BS and users are ignored, which means correlation factor $\zeta_t = \zeta_r = 0$. Under this circumstance, $\mathbf{R}_t^{1/2}$ and $\mathbf{R}_r^{1/2}$ are actually \mathbf{I}_N and \mathbf{I}_M , respectively. Then \mathbf{H} is the ideal i.i.d. Rayleigh fading channel matrix.
- *User Correlated*: For multi-antenna users, if the distance between two BS antennas is larger than half-wavelength, correlation between BS antennas can be neglected. In this condition $\mathbf{R}_r^{1/2}$ becomes diagonal matrix \mathbf{D}_r thus $\mathbf{H} = \mathbf{D}_r \mathbf{W} \mathbf{R}_t^{1/2}$.
- *BS Correlated*: For single-antenna users, correlation among users is omitted. Nevertheless, as M-MIMO

contains large-scale antenna array, pathloss between BS and users cannot be ignored. Thus the channel is $\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{W} \mathbf{D}_t$, where \mathbf{D}_t is a diagonal matrix where pathloss attenuation factor is represented.

- *Fully Correlated*: When fully correlated, both user and BS should be considered. Thus matrix remains $\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{W} \mathbf{R}_t^{1/2}$, where $\mathbf{R}_r^{1/2}$ and $\mathbf{R}_t^{1/2}$ are shown in Eqs. 5 and 6, respectively.

3 Residual-Based Detection Algorithms

In this section residual-based detection (RBD) is proposed as a series. For a linear detection problem

$$\mathbf{A}\mathbf{s} = \mathbf{y}, \quad (7)$$

suppose that \mathbf{s}_* denotes the exact estimation of detection signal, existing detection methods mainly focus on the approximation of \mathbf{s} to \mathbf{s}_* , which is denoted by the absolute error $\|\mathbf{s} - \mathbf{s}_*\|$. Whereas vector $\mathbf{r} = \|\mathbf{y} - \mathbf{A}\mathbf{s}\|$ denotes the residual norm of the signal, RBD algorithms mainly focus on the minimization of vector \mathbf{r} in the computation process. This section will give detailed description of RBD algorithms and the relationship between these algorithms will be given too. Also, although this section focuses on hard-output detection, it can be extended into soft-output detection and the implementation method can be seen in [23].

3.1 Minimal Residual Algorithm

As a kind of projection algorithm for massive MIMO detection, proposed minimal residual (MINRES) [24] is the simplest algorithm for its short calculation process, which is shown in *Algorithm 1*.

Algorithm 1 Minimal Residual Algorithm

Input: \mathbf{A} and $\tilde{\mathbf{y}}$

- 1: **for** $k = 0, \dots, K$ **do**
- 2: $\mathbf{r}_k = \tilde{\mathbf{y}} - \mathbf{A}\mathbf{s}_k$
- 3: $\alpha_k = \frac{\mathbf{r}_k^H \mathbf{A} \mathbf{r}_k}{\|\mathbf{A} \mathbf{r}_k\|^2}$
- 4: $\mathbf{s}_{k+1} = \mathbf{s}_k + \alpha_k \mathbf{r}_k$
- 5: **end for**

Output: $\hat{\mathbf{s}} = \mathbf{s}_{K+1}$

It is easily shown that MINRES minimizes the function $f(\mathbf{s}) = \|\mathbf{y} - \mathbf{A}\mathbf{s}\|_2^2$ in the direction of \mathbf{r} . Since MINRES is the simplest RBD algorithm, it requires the filtering matrix \mathbf{A} only to be positive definite. Since the MMSE filtering matrix \mathbf{A} is symmetric positive definite (SPD), the requirement can be met easily. So

$$\begin{aligned} \|\mathbf{r}_{k+1}\|_2^2 &= (\mathbf{r}_k - \alpha_k \mathbf{A} \mathbf{r}_k, \mathbf{r}_k - \alpha_k \mathbf{A} \mathbf{r}_k) \\ &= (\mathbf{r}_k - \alpha_k \mathbf{A} \mathbf{r}_k, \mathbf{r}_k) - \alpha_k (\mathbf{r}_k - \alpha_k \mathbf{A} \mathbf{r}_k, \mathbf{A} \mathbf{r}_k). \end{aligned} \quad (8)$$

For the vector $\mathbf{r}_k - \alpha_k \mathbf{A} \mathbf{r}_k$ is orthogonal to search direction $\mathbf{A} \mathbf{r}_k$, thus the right side of Eq. 8 vanishes and therefore

$$\begin{aligned} \|\mathbf{r}_{k+1}\|_2^2 &= (\mathbf{r}_k - \alpha_k \mathbf{A} \mathbf{r}_k, \mathbf{r}_k) \\ &= (\mathbf{r}_k, \mathbf{r}_k) - \alpha_k (\mathbf{A} \mathbf{r}_k, \mathbf{r}_k) \\ &= \|\mathbf{r}_k\|^2 \left(1 - \frac{(\mathbf{A} \mathbf{r}_k, \mathbf{r}_k)}{(\mathbf{r}_k, \mathbf{r}_k)} \frac{(\mathbf{A} \mathbf{r}_k, \mathbf{r}_k)}{(\mathbf{A} \mathbf{r}_k, \mathbf{A} \mathbf{r}_k)}\right) \\ &= \|\mathbf{r}_k\|^2 \left(1 - \frac{(\mathbf{A} \mathbf{r}_k, \mathbf{r}_k)^2}{(\mathbf{r}_k, \mathbf{r}_k)^2} \frac{\|\mathbf{r}_k\|_2^2}{\|\mathbf{A} \mathbf{r}_k\|_2^2}\right). \end{aligned} \quad (9)$$

For the positive definite matrix \mathbf{A} ,

$$\frac{(\mathbf{A} \mathbf{x}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})} \geq \lambda_{\min}(\mathbf{A} + \mathbf{A}^T)/2 > 0. \quad (10)$$

Since matrix \mathbf{A} is positive definite, its inversion \mathbf{A}^{-1} is positive definite, too. Similarly, let $\mathbf{t} = \mathbf{A} \mathbf{x}$ then

$$\frac{(\mathbf{A} \mathbf{x}, \mathbf{x})}{(\mathbf{A} \mathbf{x}, \mathbf{A} \mathbf{x})} = \frac{(\mathbf{t}, \mathbf{A}^{-1} \mathbf{t})}{(\mathbf{t}, \mathbf{t})} \geq \lambda_{\min}(\mathbf{A}^{-1} + \mathbf{A}^{-T})/2 > 0. \quad (11)$$

Finally, let $\mu(\mathbf{A})$ denotes $\lambda_{\min}(\mathbf{A} + \mathbf{A}^T)/2$, then

$$\|\mathbf{r}_{k+1}\|_2^2 \leq (1 - \mu(\mathbf{A})\mu(\mathbf{A}^{-1}))\|\mathbf{r}_k\|_2^2. \quad (12)$$

From the derivation given, residual norm in MINRES algorithm decreases after each iteration, thus the convergence of MINRES can be proven.

3.2 Generalized Minimal Residual Algorithm

The Generalized Minimal Residual (GMRES) Algorithm is an iterative method to calculate the solution of nonsymmetric system of linear systems [24]. It is the generalized version of MINRES, GMRES inference canceller was proposed in [21, 25] first and in this paper, the essence of GMRES will be introduced. Some computation processes to elaborate the computation process of GMRES are supplemented in this paper. As a projection method based on $\kappa = \kappa_V$ in which κ_V is V -th Krylov subspace, GMRES can minimize the residual norm to approximate the exact solution of $\mathbf{A}\mathbf{s} = \mathbf{y}$ by the vector $\mathbf{s}_k \in \kappa_k$, where

$$\kappa_V = \text{span}\{\mathbf{y}, \mathbf{A}\mathbf{y}, \mathbf{A}^2\mathbf{y}, \dots, \mathbf{A}^{V-1}\mathbf{y}\}. \quad (13)$$

To avoid the linear independence of vectors $\mathbf{y}, \mathbf{A}\mathbf{y}, \dots, \mathbf{A}^{V-1}\mathbf{y}$, Arnoldi iteration [26] is used to form orthogonal

basis $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_V$ for κ_V . Thus vector $\mathbf{s}_V \in \kappa_V$ can be rewritten as $\mathbf{s} = \mathbf{s}_0 + \mathbf{Q}_V \mathbf{p}_V$, where \mathbf{Q}_V is an m -by- V matrix formed by basis $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_V$.

Meanwhile, a $(V+1)$ -by- V upper Hessenberg matrix $\tilde{\mathbf{H}}_V$ is produced in the Arnoldi iteration process, where

$$\mathbf{A}\mathbf{Q}_V = \mathbf{Q}_{V+1}\tilde{\mathbf{H}}_V. \quad (14)$$

Thus, the whole GMRES process can be deduced. Define

$$\begin{aligned} J(\mathbf{p}) &= \|\mathbf{y} - \mathbf{A}\mathbf{s}\|_2 = \|\mathbf{y} - \mathbf{A}(\mathbf{s}_0 + \mathbf{Q}_V \mathbf{p})\|_2 \\ &= \|\mathbf{r}_0 - \mathbf{A}\mathbf{Q}_V \mathbf{p}\|_2 \\ &= \|\beta \mathbf{q}_1 - \mathbf{Q}_{V+1} \tilde{\mathbf{H}}_V \mathbf{p}\|_2 \\ &= \|\mathbf{Q}_{V+1}(\beta \mathbf{e}_1 - \tilde{\mathbf{H}}_V \mathbf{p})\|_2. \end{aligned} \quad (15)$$

Since the column-vectors of \mathbf{Q}_{V+1} are orthogonal, it is easy to understand that

$$J(\mathbf{p}) = \|\beta \mathbf{e}_1 - \tilde{\mathbf{H}}_V \mathbf{p}\|_2. \quad (16)$$

With the definition of $J(\mathbf{p})$, GMRES algorithm minimizes it and make the signal approximating $\mathbf{s}_0 + \kappa_V$. After knowing this, GMRES approximation can be denoted by simple equation

$$\mathbf{s}_V = \mathbf{s}_0 + \mathbf{Q}_V \mathbf{p}_V, \quad (17)$$

where

$$\mathbf{p}_V = \arg \min \|\beta \mathbf{e}_1 - \tilde{\mathbf{H}}_V \mathbf{p}_V\|_2. \quad (18)$$

Accordingly, the computation process of GMRES algorithm is shown in *Algorithm 2*.

Algorithm 2 Generalized Minimal Residual Algorithm

Input: \mathbf{A} and $\tilde{\mathbf{y}}$

```

1:  $\mathbf{r}_0 = \tilde{\mathbf{y}} - \mathbf{A}\mathbf{s}_0$ ,  $\beta = \|\mathbf{r}_0\|_2$  and  $\mathbf{q}_1 = \mathbf{r}_0/\beta$ 
2: Define the  $(V+1)$ -by- $V$  matrix  $\tilde{\mathbf{H}}_V$ .  $\tilde{\mathbf{H}}_V = \mathbf{0}$ 
3: for  $j = 1, \dots, V$  do
4:    $\mathbf{w}_j = \mathbf{A}\mathbf{q}_j$ 
5:   for  $i = 1, \dots, j$  do
6:      $\tilde{\mathbf{H}}(i, j) = \mathbf{q}_i^T \mathbf{w}_j$ 
7:      $\mathbf{w}_j = \mathbf{w}_j - \tilde{\mathbf{H}}(i, j)\mathbf{q}_i$ 
8:   end for
9:    $\tilde{\mathbf{H}}(j+1, j) = \|\mathbf{w}_j\|_2$ 
10:  if  $\tilde{\mathbf{H}}(j+1, j) = 0$  then
11:     $V = j$  and go to 13
12:  end if
13:   $\mathbf{q}_{j+1} = \mathbf{w}_j / \tilde{\mathbf{H}}(j+1, j)$ 
14:   $\mathbf{p}_V = \arg \min \|\beta \mathbf{e}_1 - \tilde{\mathbf{H}}_V \mathbf{p}\|_2$ 
15:   $\mathbf{s}_k = \mathbf{s}_{k-1} + \mathbf{Q}_V \mathbf{p}_V$ 
16: end for

```

Output: $\hat{\mathbf{s}} = \mathbf{s}_V$

With the information given, M-MIMO detection problems can be solved easily. However, key step of GMRES is step-12 in *Algorithm 2*, which is not mentioned in [21, 25]. To supplement the process of GMRES and make it easier to be understood, Givens rotation to solve this optimization problem is introduced in this paper and can be seen in [Appendix](#).

For matrix \mathbf{A} , $(\mathbf{A}^T + \mathbf{A})/2$ is positive definite, then in the k -th iteration,

$$\|\mathbf{r}_k\| \leq (1 - \frac{\lambda_{\min}^2(1/2(\mathbf{A}^T + \mathbf{A}))}{\lambda_{\max}(\mathbf{A}^T \mathbf{A})})^{n/2} \|\mathbf{r}_0\|, \quad (19)$$

where $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$ denote the minimum and maximum eigenvalue of matrix \mathbf{M} , respectively.

While in M-MIMO detection scheme, matrix \mathbf{A} is SPD, then Eq. 19 can be deformed to

$$\|\mathbf{r}_k\| \leq (\frac{\tau_2(\mathbf{A})^2 - 1}{\tau_2(\mathbf{A})^2})^{n/2} \|\mathbf{r}_0\|, \quad (20)$$

where $\tau_2(\mathbf{A})$ is the condition number of \mathbf{A} .

From Eqs. 19 and 20, it can be seen that residual norm of GMRES strictly decreases after iterations, which shows the convergence of it. Synthesizing Arnoldi GMRES algorithm and Givens rotation, the complete GMRES algorithm is a kind of advanced algorithm as a M-MIMO detection method by minimizing the norm of the residual vector.

3.3 Conjugate Residual Algorithm

As can be seen in Section 3.2, complete GMRES algorithm needs too many operations and some of them are square root, and even matrix inversion from Givens rotation, which should be avoided in M-MIMO detection scheme. To remedy this and keep the performance of the algorithm for M-MIMO detection, GMRES can be updated to an advanced version.

Consider GMRES is an algorithm for nonsymmetric problem, while M-MIMO detection is solving a SPD problem, some restrictions can be added to GMRES, which makes the GMRES algorithm involving into the proposed conjugate residual (CR) algorithm. Switching nonsymmetric problems to Hermitian problems, CR can lower the computational complexity of GMRES. Being another Krylov subspace iterative method, CR also minimizes the residual vector in each iteration and is feasible in M-MIMO detection. Computation process of CR algorithm is shown in *Algorithm 3*.

Algorithm 3 CR for MMSE detection**Input:** \mathbf{A} and $\tilde{\mathbf{y}}$

```

1:  $\mathbf{s}_0 = \mathbf{0}, \mathbf{r}_0 = \tilde{\mathbf{y}} - \mathbf{A}\mathbf{s}_0, \mathbf{p}_0 = \mathbf{r}_0$ 
2:  $\mathbf{e}_0 = \mathbf{A}\mathbf{p}_0, \mathbf{m}_0 = \mathbf{A}\mathbf{r}_0$ 
3: for  $k = 1, \dots, K$  do
4:    $\alpha_k = \mathbf{r}_{k-1}^H \mathbf{m}_{k-1} / \|\mathbf{e}_{k-1}\|^2$ 
5:    $\mathbf{s}_k = \mathbf{s}_{k-1} + \alpha_k \mathbf{p}_{k-1}$ 
6:    $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_k \mathbf{e}_{k-1}$ 
7:    $\mathbf{m}_k = \mathbf{A}\mathbf{r}_k$ 
8:    $\beta_k = \mathbf{r}_k^H \mathbf{m}_k / \mathbf{r}_{k-1}^H \mathbf{m}_{k-1}$ 
9:    $\mathbf{p}_k = \mathbf{r}_k + \beta_k \mathbf{p}_{k-1}$ 
10:   $\mathbf{e}_k = \mathbf{m}_k + \beta_k \mathbf{e}_{k-1}$ 
11: end for

```

Output: $\hat{\mathbf{s}} = \mathbf{s}_K$

The output $\hat{\mathbf{s}}$ can be proved to support the convergence of the algorithm [27]. For CR on an SPD system,

$$\|\mathbf{s}_k\|^2 - \|\mathbf{s}_{k-1}\|^2 = 2\alpha_k \mathbf{s}_{k-1}^T \mathbf{p}_{k-1} + \mathbf{p}_{k-1}^T \mathbf{p}_{k-1} \geq 0. \quad (21)$$

Therefore,

$$\|\mathbf{s}_k\| \geq \|\mathbf{s}_{k-1}\|. \quad (22)$$

Then, final solution can be expressed as $\mathbf{s}_l = \mathbf{s}^*$,

$$\begin{aligned}
\mathbf{s}_l &= \mathbf{s}_{l-1} + \alpha_{l-1} \mathbf{p}_{l-1} \\
&= \dots \\
&= \mathbf{s}_k + \alpha_{k+1} \mathbf{p}_k + \dots + \alpha_{l-1} \mathbf{p}_{l-1} \\
&= \mathbf{s}_{k-1} + \alpha_k \mathbf{p}_{k-1} + \alpha_{k+1} \mathbf{p}_k + \dots + \alpha_{l-1} \mathbf{p}_{l-1}. \quad (23)
\end{aligned}$$

From the conclusion above, it can be deduced that

$$\begin{aligned}
&\|\mathbf{s}_l - \mathbf{s}_{k-1}\|^2 - \|\mathbf{s}_l - \mathbf{s}_k\|^2 \\
&= 2\alpha_k \mathbf{p}_{k-1}^T (\alpha_{k+1} \mathbf{p}_k + \dots + \alpha_{l-1} \mathbf{p}_{l-1}) \\
&\quad + \alpha_k^2 \mathbf{p}_{k-1}^T \mathbf{p}_{k-1} \geq 0. \quad (24)
\end{aligned}$$

While for the MMSE linear detection problem, linear equation $\mathbf{A}\mathbf{s} = \mathbf{y}$ is to be solved. Thus

$$\begin{aligned}
&\|\mathbf{s}_l - \mathbf{s}_{k-1}\|_{\mathbf{A}}^2 - \|\mathbf{s}_l - \mathbf{s}_k\|_{\mathbf{A}}^2 \\
&= 2\alpha_k \mathbf{p}_{k-1}^T \mathbf{A} (\alpha_{k+1} \mathbf{p}_k + \dots + \alpha_{l-1} \mathbf{p}_{l-1}) + \alpha_k^2 \mathbf{p}_{k-1}^T \mathbf{A} \mathbf{p}_{k-1} \\
&= 2\alpha_k \mathbf{q}_{k-1}^T (\alpha_{k+1} \mathbf{p}_k + \dots + \alpha_{l-1} \mathbf{p}_{l-1}) + \alpha_k^2 \mathbf{q}_{k-1}^T \mathbf{p}_{k-1} > 0. \quad (25)
\end{aligned}$$

The derivation above indicates that the residual norm is strictly decreasing. Thus CR is feasible for massive MIMO detection.

4 Numerical Results and Comparison

4.1 Results with Different Antenna Configurations

With 64-QAM and i.i.d. channel model, the bit-error-rate (BER) comparison between each RBD algorithm and two antenna configurations are considered. Here iteration time k is set as 2, 3 and 4, respectively.

It is worth noting that because CR algorithm is a derivation of GMRES algorithm in M-MIMO scheme, they have the same BER performance as mentioned in Section 3.3. For better elaboration, it is also shown in Fig. 1. It can be seen that when $k = 4$, both of them approximate traditional matrix inversion. To be specific, when $\text{BER} = 9 \times 10^{-3}$, CR has only 0.25 dB gap between *Cholesky* decomposition.

To show the relationship between RBD algorithms and other algorithms, here we take coordinate descent (CD) [28] into consideration. In [29], optimized version of CD is proposed as optimized coordinate descent (OCD). In this paper, it is fair to compare the BER performance of original RBD algorithms with original CD algorithm. Here Figs. 2 and 3 compare BER performance of each RBD algorithm and CD algorithm when the antenna configuration is $N \times M = 128 \times 16$ and 128×8 , respectively. In Fig. 2, when $\text{BER} = 10^{-3}$ and iteration time $k = 4$, MINRES has 1.63 dB drawback compared with *Cholesky* decomposition while GMRES and CR have 10.25 dB gap with *Cholesky* decomposition. Performance of RBD algorithms improve a lot along with the increment of iteration time k . Meanwhile, GMRES and CR outperform MINRES a lot. For example, when $\text{BER} = 10^{-2}$ and iteration time $k = 3$, CR and

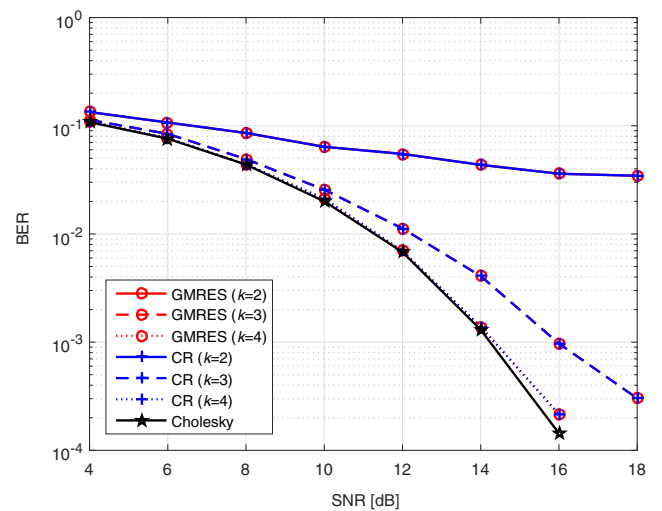


Figure 1 Performance comparison with $N \times M = 128 \times 16$.

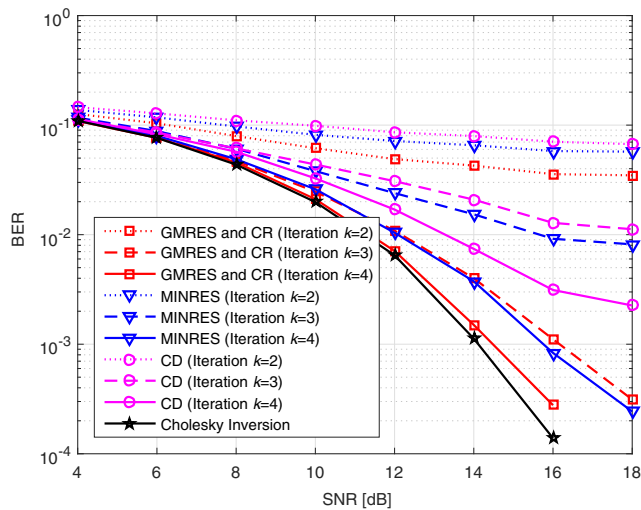


Figure 2 Performance comparison with $N \times M = 128 \times 16$.

GMRES outperform MINRES by nearly 4 dB SNR gap. Also, it can be seen that CD cannot outperform RBD algorithms. When $\text{BER} = 8 \times 10^{-2}$ after 4 times of iterations, GMRES and CR outperform CD by 2.75 dB and MINRES outperforms CD by 1.5 dB.

For another antenna configuration $N \times M = 128 \times 8$, as shown in Fig. 3, RBD algorithms perform well in approximating *Cholesky* decomposition scheme. MINRES has huge performance improvement as iteration time increases. Take $\text{BER} = 7 \times 10^{-2}$ for instance, MINRES has 6.1 dB gain when iteration time increases from $k = 2$ to $k = 3$. CR and GMRES have almost the same performance with exact matrix inversion when iteration time $k \geq 3$, in which condition SNR gap between them is less than 0.2 dB. In this case, CD can approximate the performance of MINRES,

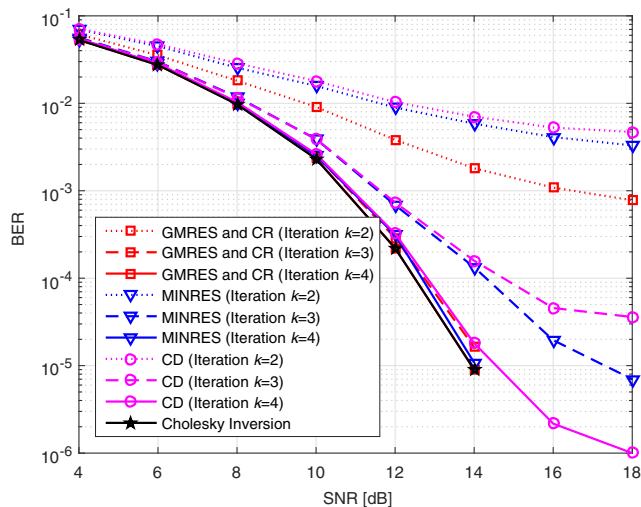


Figure 3 Performance comparison with $N \times M = 128 \times 8$.

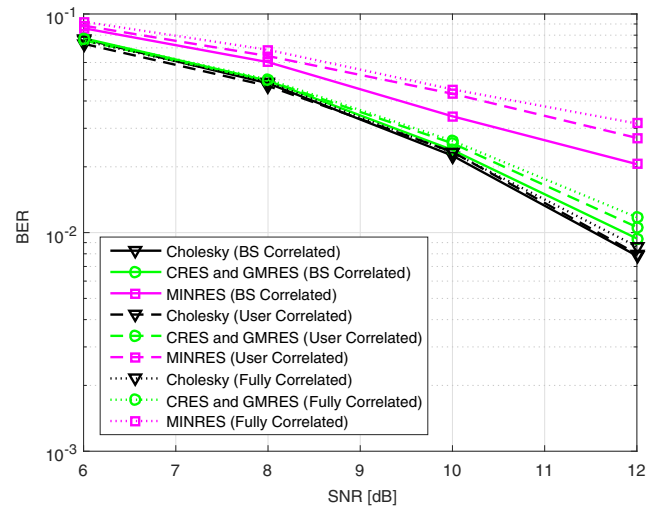


Figure 4 Performance comparison with correlations.

while has no advantage over RBD algorithms. When $\text{BER} = 9 \times 10^{-4}$, CD has 0.13 dB drawback than MINRES and 0.23 dB drawback than GMRES and CR after 4 iterations. From the results above, RBD algorithms perform better than CD in M-MIMO detection problems.

4.2 Results with Different Correlation Conditions

Consider $N \times M = 128 \times 16$ M-MIMO system and iteration time is 4, BER performances of each RBD algorithm and *Cholesky* decomposition are given in Fig. 4. Here three conditions are considered: i) *User Correlated* case ($\zeta_t = 0.4, \zeta_r = 0$), ii) *BS correlated* case ($\zeta_t = 0, \zeta_r = 0.4$), iii) *Fully Correlated* case: ($\zeta_t = 0.4, \zeta_r = 0.4$).

In Fig. 4, as the correlation factor ζ varies, performance of GMRES and CR remain stable and the performance loss is less than 0.5 dB. MINRES algorithm will suffer from the change of the correlation condition. However, MINRES loses up to 1.8 dB when $\text{BER} = 8 \times 10^{-1}$. Thus RBD algorithms are not very sensitive to correlation conditions for M-MIMO.

5 Computational Complexity Analysis

Computational complexity of each RBD algorithm is compared to describe the complexity issue of RBD algorithms. In this section computational complexity is analyzed for better understanding of RBD algorithms. As mentioned in Section 3.1, MINRES algorithm is the basic algorithm in RBD algorithms. GMRES algorithm is the generalized version of MINRES and is complex in computation process. To meet the requirement of M-MIMO system, GMRES can be derived into CR algorithm, which is suitable for M-MIMO detection. Table 1 concludes the

Table 1 Complexity comparison of different algorithms.

| Operation | MINRES | GMRES | CR |
|----------------|---------------|--|--------------------|
| Addition | $2kM$ | $(\frac{1}{2}k^2 + \frac{3}{2}k + 1)M$ | $(4K + 1)M$ |
| Multiplication | $4kM^2 + 2kM$ | $2kM^2 + (\frac{1}{2}k^2 + \frac{1}{2}k + 2)M$ | $(k + 2)M^2 + 4kM$ |

comparison of different algorithms in terms of complex-valued additions and complex-valued multiplications.

The detection complexity is mainly contributed by complex-valued multiplication. The complexity of each algorithm is compared with *Cholesky* decomposition. Suppose the antenna number at BS is 128 and SNR is 20 dB. Complexity comparison is shown in Figs. 5 and 6. Here Fig. 6 is another showcase of Fig. 5 because it transform the axes of Fig. 5 into logarithmic and the relationship between different algorithms can be seen easily.

5.1 Complexity of MINRES

Being the basic RBD, MINRES has the simplest computation process, though in Fig. 5 its complexity is not the least. However, when user antenna number is 60, MINRES can achieve 76% complexity reduction compared with traditional matrix inversion after 3 iterations and 84% after 2 iterations. Thus the complexity reduction outweighs the performance loss in terms of trade-off.

5.2 Complexity of GMRES

As a generalized version of MINRES, GMRES has more application scenarios. However, its complexity rises also. As shown in Fig. 5, GMRES has higher complexity than other

RBD algorithms. Similarly, when user antenna number is 60, GMRES reduces the complexity of traditional method by 75% after 3 iterations and 86% after 2 iterations.

5.3 Complexity of CR

It is clear that CR has the lowest complexity of RBD algorithms: in the same condition, CR reduces the complexity of traditional method by 89% when user antenna number is 60 after 3 iterations and 91% after 2 iterations. Having the BER performance in Section 4, CR is the best algorithm in RBD algorithms and can substitute GMRES in M-MIMO.

6 Hardware Architecture for RBD Algorithms

Computational process of RBD algorithms is introduced in Section 3. To further elaborate RBD algorithms, corresponding hardware architectures are shown in Section 6.1 and 6.2. Since GMRES algorithm maintains the same BER performance as CR algorithm with unaffordable computational complexity, GMRES algorithm is not hardware friendly. Thus the implementation of GMRES is replaced by CR algorithm. A method to unify the hardware design is also proposed in Section 6.3. Using the new design method, RBD algorithms can be designed by only two basic

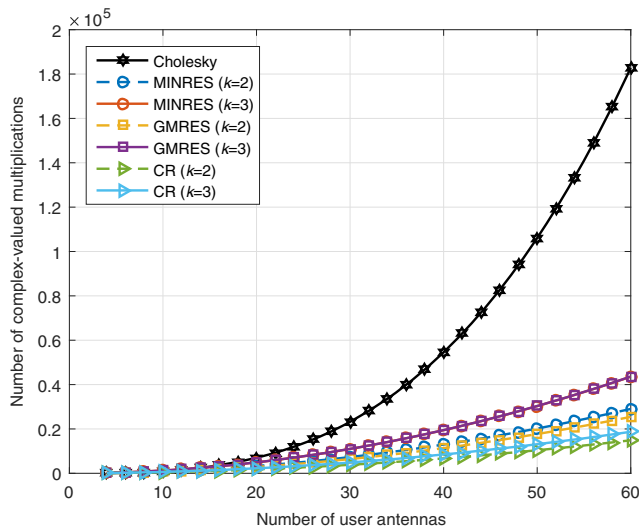
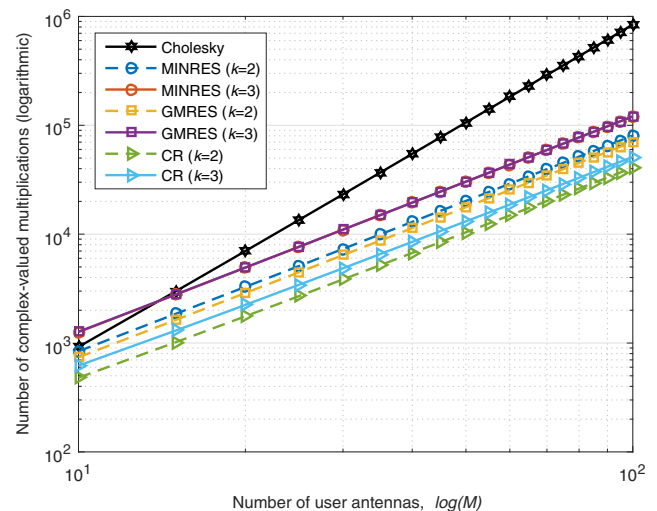
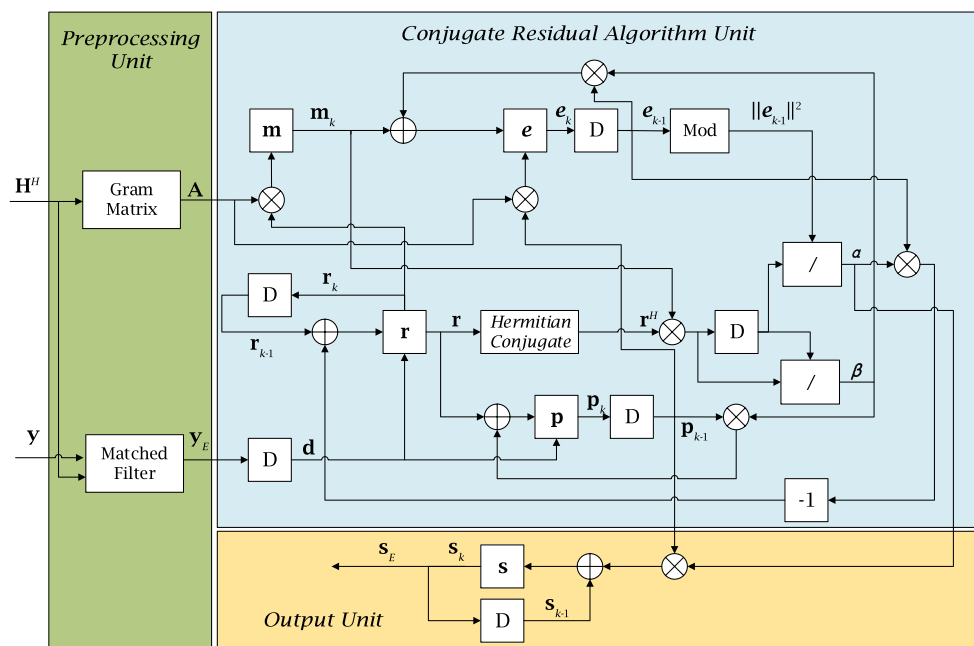
**Figure 5** Computational complexity comparison.**Figure 6** Logarithmic computational complexity comparison.

Figure 7 Hardware architecture of CR method.



modules. Unified architectures of MINRES and CR are proposed in Section 6.3 to validate the design method (Fig. 7).

6.1 Hardware Architecture of MINRES Algorithm

As mentioned in Section 3.1, MINRES is the most basic algorithm of RBD algorithm, thus the hardware architecture of it is not very complex. Being divided into two units, Fig. 8 shows the architecture of MINRES algorithm. Preprocessing Unit computes the output of matched filter \tilde{y} and matrix A . Minimal Residual Algorithm Unit is the main unit of the architecture and it minimizes the residual norm in each iteration. In Fig. 8, \tilde{y} is denoted by y_E and the symbol output is denoted by s_{k+1} , where the index k is iteration time.

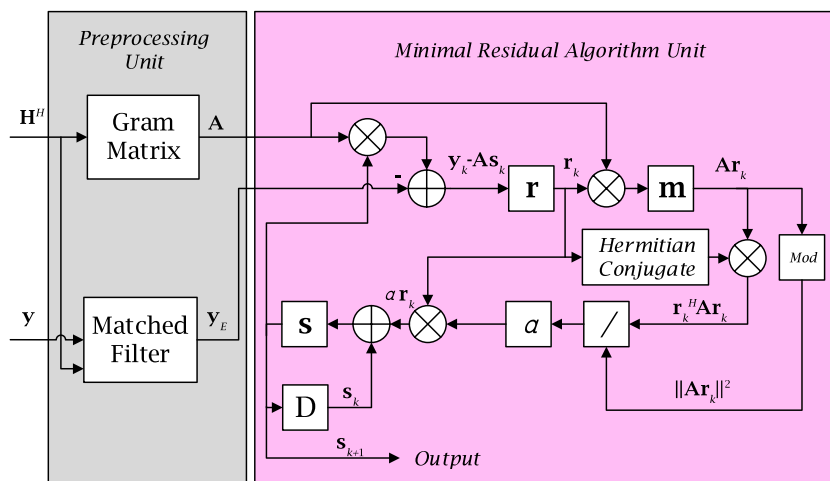
6.1.1 Preprocessing Unit

In this unit, matched filter module computes \tilde{y} by $\tilde{y} = H^H y$ while MMSE filtering matrix A is computed by Gram matrix module. Since matrix A is Hermitian, $M \times M$ lower triangular systolic array is adopted to compute it. Each processing element (PE) performs a multiply-accumulate (MAC) operation with same inputs.

6.1.2 Minimal Residual Algorithm Unit

In this unit, symbol signals are stored and computed iteratively. Square module with r , m , s , α stores the corresponding signals of each iteration. The hermitian of symbol is given after hermitian conjugate module and module with “/” means division operation, in which the

Figure 8 Hardware architecture of MINRES method.



input from downside is the divisor. Module with “D” is the delayer which can provide the signal of last iteration for the algorithm. “Mod” module computes the modulus of the input signal. At the end of the iteration, output s_k is the final symbol output of MINRES.

6.2 Hardware Architecture of CR Algorithm

Being another RBD algorithm, CR has much lower computational complexity compared with traditional exact matrix inversion. As is shown in Section 6.1.1, CR has better performance than MINRES. Thus CR performs well in terms of performance within RBD algorithms. Hardware architecture of CR contains three parts: preprocessing unit, conjugate residual algorithm unit, output unit. Same with that of Section 6.1.1, preprocessing unit computes $\tilde{\mathbf{y}}$ and \mathbf{A} . Conjugate residual algorithm unit is the main unit and it computes symbol signals iteratively to minimize residual norm as well.

6.2.1 Preprocessing Unit

Functioning as a preliminary unit, this unit has the same architecture with that in Section 6.1.1.

6.2.2 Conjugate Residual Algorithm Unit

As the main computing unit of CR algorithm, this unit adopts similar functional modules with that in Section 6.1.2. Differently, with the output of preprocessing unit, CR algorithm needs initialization, which is denoted by the input from the downside of the storage module of vector \mathbf{r} , \mathbf{p} , \mathbf{e} , \mathbf{m} . Meanwhile, in this architecture the left input of division module is dividend and the upper input is the divisor of division operation.

6.2.3 Output Unit

With the output of CR algorithm unit, we provide the estimation of transmitted signal stored in \mathbf{s} . When the iteration ends, final symbol output is denoted by \mathbf{s}_E .

6.3 Unified Hardware Architecture

Being RBD algorithm, MINRES and CR have different architectures. Thus in terms of implementation they are uncorrelated. Thanks to the special characteristic of RBD algorithm that the minimization of residual norm, RBD algorithms can be designed by a unified design method. In this part a design method to normalize the hardware architecture of RBD algorithm is firstly introduced and then unified architecture of MINRES and CR are given.

6.3.1 Normalizing Design Method of RBD Algorithms

The purpose of the normalizing design method is to make the hardware design of RBD algorithms flexible and reusable, users can switch detectors with existing hardware resources as long as they want to. To meet this purpose, synthesizing the characteristic of RBD algorithms, which is minimizing the residual norm, the normalizing design method is then applied.

Consider the computation process and hardware module of RBD algorithms, this method takes two modules as basic modules: iterative module and coefficient module. Iterative module can iteratively update the signal or compute the residual norm in each iteration. Another module is coefficient module which computes the coefficient of each vector in computation process. Hardware architectures of both basic modules are shown in Fig. 9.

Iterative module consists of two operation unit, a multiplier and an accumulator, which performs a MAC operation. Coefficient module consists of two Hermitian conjugate module, a division module and a multiplier, which provides the coefficient for each iteration module. Upper input of this module is the dividend and the input from downside is the divisor. Operation of each module can be denoted by

$$\begin{cases} \mathbf{y} = \mathbf{x} + \mathbf{a}\mathbf{b}, \\ \mathbf{c} = \frac{\mathbf{m}^H \mathbf{n}}{\mathbf{p}^H \mathbf{q}}. \end{cases} \quad (26)$$

Having these two basic modules, hardware architectures of RBD algorithms can be unified. Besides basic modules,

Figure 9 Basic modules of unified hardware architecture.

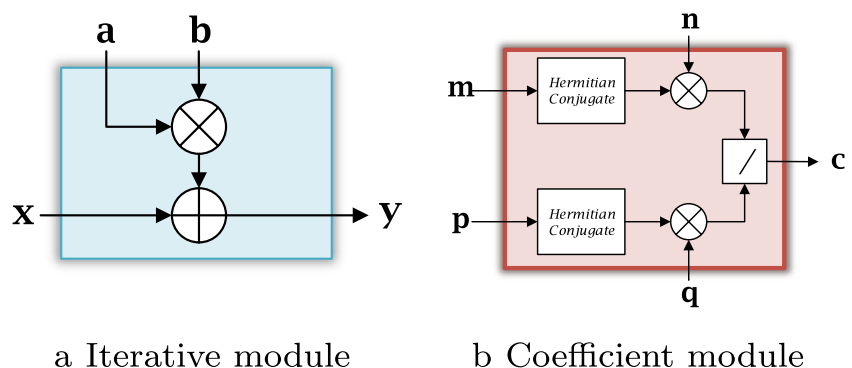
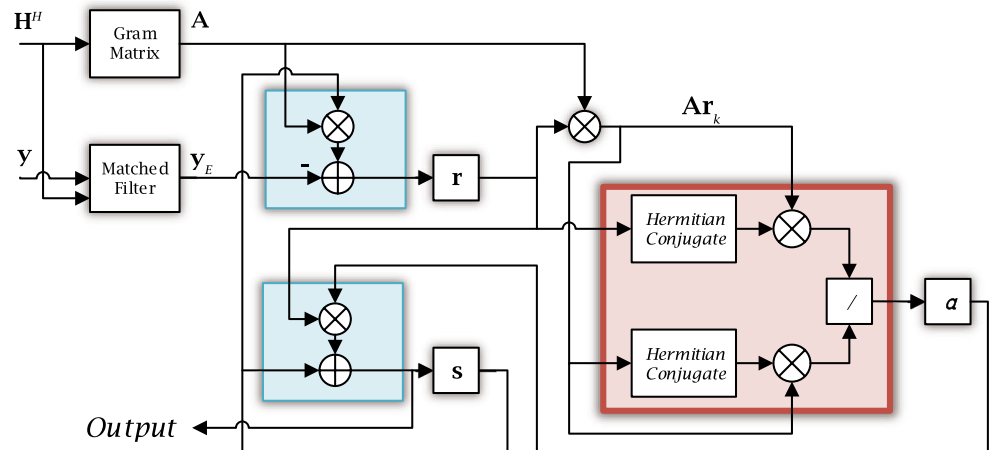


Figure 10 Unified architecture of MINRES algorithm.



only some multipliers and delays are needed. Thus the flexibility of hardware can be improved and the architecture can be reused for further usage. Those two basic modules can also be used in some other iterative detection algorithms. To validate the reasonability of this design method, unified hardware architectures are given in Sections 6.3.2 and 6.3.3.

6.3.2 Unified Architecture of MINRES Algorithm

As the basic algorithm of RBD algorithms, MINRES does not need many basic modules, the unified architecture of it contains two iterative modules and a coefficient module.

Input signal of this unified architecture is also computed from Gram matrix module and matched filter. By

normalizing design method, MINRES algorithm adopts two iterative modules to store the residual r and signal s . Coefficient module serves for the coefficient α . Aside from basic modules, unified hardware architecture of MINRES only has an additional multiplier. After the computation of iteration, symbol output is given as the output of an iterative module (Fig. 10).

6.3.3 Unified Hardware Architecture of CR Algorithm

Traditional hardware architecture of CR algorithm is kind of complex as shown in Fig. 7. After normalization, the architecture is shown in Fig. 11.

Unified hardware architecture of CR algorithm contains four iterative modules and two coefficient modules. Iterative

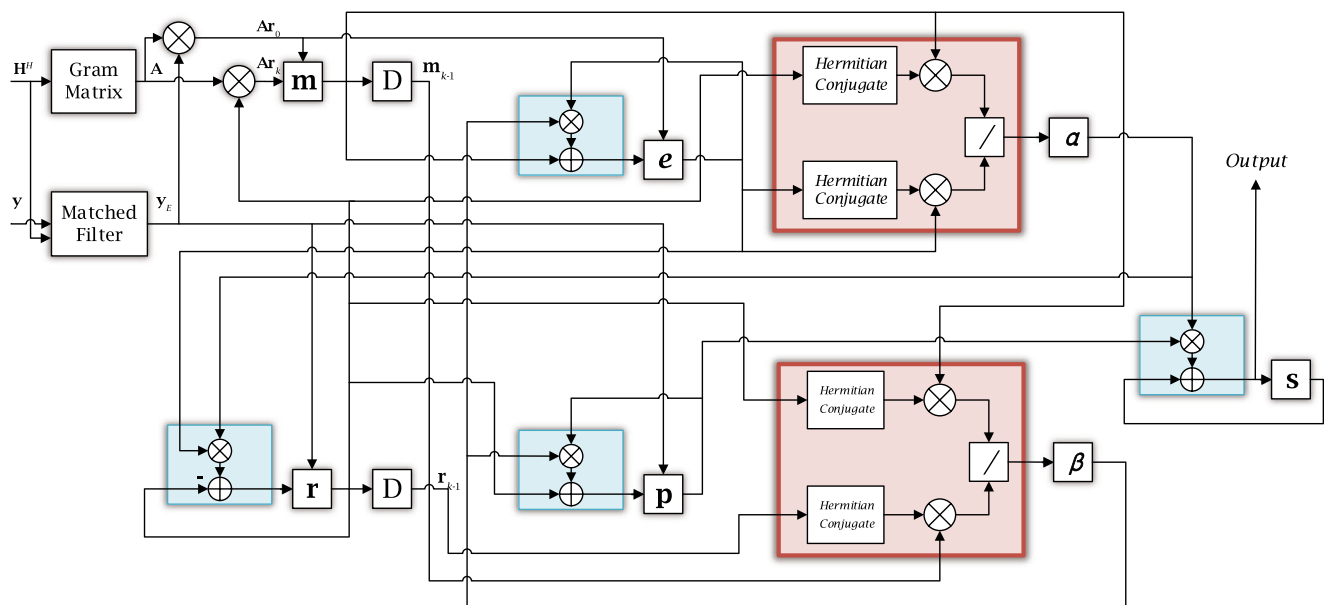


Figure 11 Unified architecture of CR algorithm.

modules are placed for the storage of signal \mathbf{r} , \mathbf{e} , \mathbf{p} and \mathbf{s} . Coefficient modules store the value of coefficient α and β . Within each iteration, signal \mathbf{m} is updated by a multiplier and two delayers in the architecture store corresponding signal of last iteration, as mentioned in *Algorithm 3*. Initialization of each signal is the upper input of each module.

With the proposed method in Section 6.3.1, hardware architectures of RBD algorithms can be unified. Furthermore, the design method can also be applied to other linear iterative detectors like CG.

7 Conclusion

In this paper, RBD algorithms are first proposed, including MINRES algorithm, GMRES algorithm and CR algorithm. Distinguished from most of other iterative linear detection algorithms, proposed RBD algorithms focus on the minimization of residual norm. Numerical results of different antenna configurations and correlation conditions have demonstrated the approximation to the performance of traditional matrix inversion and the stability of algorithms, respectively. In addition, computational complexity of RBD algorithms are compared and the comparison with matrix inversion shows the complexity reduction advantage of RBD algorithms. Finally hardware architectures of RBD algorithms are first given and the following proposed normalizing design method is adopted, then the unified hardware architectures of RBD algorithms are proposed. Therefore, the proposed RBD algorithms are of good performance, low complexity, and correlation robustness, which are favorable for M-MIMO systems. Future work will be directed towards FPGA implementation of RBD algorithms and further optimization of RBD algorithms.

Acknowledgements This work is supported in part by NSFC under grants 61871115 and 61501116, Jiangsu Provincial NSF for Excellent Young Scholars, Huawei HIRP Flagship under grant YB201504, the Fundamental Research Funds for the Central Universities, the SRTP of Southeast University, State Key Laboratory of ASIC & System under grant 2016KF007, ICRI for MNC, and the Project Sponsored by the SRF for the Returned Overseas Chinese Scholars of MoE.

Appendix : Derivation of Givens Rotation

Given the problem $\mathbf{p} = \arg \min \|\beta \mathbf{e}_1 - \tilde{\mathbf{H}}_V \mathbf{p}\|_2$, knowing that $\tilde{\mathbf{H}}_V$ is a $(V+1)$ -by- V matrix. It is shown that an over-constrained linear system of $V+1$ equations for V unknowns is given and the minimum can be computed by QR decomposition [30]. An $(V+1)$ -by- $(V+1)$ orthogonal matrix Ω_V and an $(V+1)$ -by- V upper triangular matrix $\tilde{\mathbf{R}}_V$ such that $\Omega_V \tilde{\mathbf{H}}_V = \tilde{\mathbf{R}}_V$.

Because of the characteristic of matrix $\tilde{\mathbf{H}}_V$ and $\tilde{\mathbf{R}}_V$, they can be denoted as

$$\tilde{\mathbf{H}}_{V+1} = \begin{bmatrix} \tilde{\mathbf{H}}_V & \mathbf{h}_{V+1} \\ 0 & h_{V+2,V+1} \end{bmatrix}, \tilde{\mathbf{R}}_V = \begin{bmatrix} \mathbf{R}_V \\ 0 \end{bmatrix}, \quad (27)$$

where $\mathbf{h}_{V+1} = (h_{1,V+1}, \dots, h_{V+1,V+1})^T$. Premultiplying the Hessenberg matrix with Ω_V , a nearly triangular matrix can be yielded with zeros and a row with multiplicative identity as

$$\begin{bmatrix} \Omega_V & 0 \\ 0 & 1 \end{bmatrix} \tilde{\mathbf{H}}_{V+1} = \begin{bmatrix} \mathbf{R}_V & \mathbf{r}_{V+1} \\ 0 & \rho \\ 0 & \sigma \end{bmatrix}. \quad (28)$$

If $\sigma = 0$, this matrix would be triangular. Givens rotation [31] will remedy this as

$$\mathbf{G}_{V+1} = \begin{bmatrix} \mathbf{I}_V & 0 & 0 \\ 0 & c_V & b_V \\ 0 & -b_V & c_V \end{bmatrix}, \quad (29)$$

where

$$c_V = \frac{\rho}{\sqrt{\rho^2 + \sigma^2}} \text{ and } b_V = \frac{\sigma}{\sqrt{\rho^2 + \sigma^2}}. \quad (30)$$

After the processing of Givens rotation, matrix Ω_V can be formed as

$$\Omega_{V+1} = \mathbf{G}_V \begin{bmatrix} \Omega_V & 0 \\ 0 & 1 \end{bmatrix}. \quad (31)$$

Meanwhile, a triangular matrix is yielded as

$$\Omega_{V+1} \tilde{\mathbf{H}}_{V+1} = \begin{bmatrix} \mathbf{R}_V & r_{V+1} \\ 0 & r_{V+1,V+1} \\ 0 & 0 \end{bmatrix}, \quad (32)$$

where $r_{V+1,V+1} = \sqrt{\rho^2 + \sigma^2}$.

Then given the QR decomposition, the minimization problem can be solved by the transform that

$$\begin{aligned} \|\tilde{\mathbf{H}}_V \mathbf{p}_V - \beta \mathbf{e}_1\| &= \|\Omega_V (\tilde{\mathbf{H}}_V \mathbf{p}_V - \beta \mathbf{e}_1)\| \\ &= \|\tilde{\mathbf{R}}_V \mathbf{p}_V - \beta \Omega_V \mathbf{e}_1\|. \end{aligned} \quad (33)$$

Afterwards, using vector $\tilde{\mathbf{g}}_V$ to denote $\beta \Omega_V \mathbf{e}_1$ as

$$\tilde{\mathbf{g}}_V = \begin{bmatrix} \mathbf{g}_V \\ \gamma_V \end{bmatrix}, \quad (34)$$

where $\mathbf{g}_V \in \mathbb{R}_V$ and $\gamma_V \in \mathbb{R}$.

Finally, norm $\|\tilde{\mathbf{H}}_V \mathbf{p}_V - \beta \mathbf{e}_1\|$ can be denoted by

$$\begin{aligned} \|\tilde{\mathbf{H}}_V \mathbf{p}_V - \beta \mathbf{e}_1\| &= \|\tilde{\mathbf{R}}_V \mathbf{p}_V - \beta \Omega_V \mathbf{e}_1\| \\ &= \left\| \begin{bmatrix} \mathbf{R}_V \\ 0 \end{bmatrix} \mathbf{p}_V - \begin{bmatrix} \mathbf{g}_V \\ \gamma_V \end{bmatrix} \right\|. \end{aligned} \quad (35)$$

So vector \mathbf{p} that minimizes the norm is

$$\mathbf{p}_V = \mathbf{R}_V^{-1} \mathbf{g}_V, \quad (36)$$

where vector \mathbf{g}_V can be updated easily and the minimization problem can be solved.

References

- Larsson, E., et al. (2014). Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2), 186–195.
- Lee, J., Han, J.-K., Zhang, J.C. (2009). MIMO technologies in 3GPP LTE and LTE-advanced. *EURASIP Journal on Wireless Communications and Networking*, 2009(1), 1–10.
- Huh, H., Caire, G., Papadopoulos, H.C., Ramprasad, S.A. (2012). Achieving massive MIMO spectral efficiency with a not-so-large number of antennas. *IEEE Transactions on Wireless Communications*, 11(9), 3226–3239.
- Lu, L., Li, G.Y., Swindlehurst, L.A., Ashikhmin, A., Zhang, R. (2014). An overview of massive MIMO Benefits and challenges. *IEEE Journal of Selected Topics in Signal Processing*, 8(5), 742–758.
- Spencer, Q.H., Swindlehurst, L.A., Haardt, M. (2004). Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels. *IEEE Transactions on Signal Processing*, 52(2), 461–471.
- Larsson, E.G. (2009). MIMO detection methods: How they work [lecture notes]. *IEEE Signal Processing Magazine*, 26(3), 91–95.
- Krishnamoorthy, A., & Menon, D. (2013). Matrix inversion Cholesky decomposition. In *Proceedings of the IEEE Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pp. 70–72.
- Wang, F., Zhang, C., Yang, J., Liang, X., You, X., Xu, S. (2015). Efficient matrix inversion architecture for linear detection in massive MIMO systems. In *Proceedings of the IEEE, Digital Signal Processing (DSP)*, pp. 248–252.
- Liang, X., Zhang, C., Xu, S., You, X. (2015). Coefficient adjustment matrix inversion approach and architecture for massive mimo systems. In *Proc. Inter. Conf. on ASIC (ASICON)*, pp. 1–4.
- Wu, M., Bei, Y., Guohui, W., Dick, C., Cavallaro, J.R., Studer, C. (2014). Large-scale MIMO detection for 3GPP LTE: Algorithms and FPGA implementations. *IEEE Journal of Selected Topics in Signal Processing*, 8(5), 916–929.
- Dai, L., Linglong, Gao, X., Xinyu, Xin, S. u., Han, S., Huangfeng, Chih-Lin, I., Wang, Z. (2015). Low-complexity soft-output signal detection based on Gauss–Seidel method for uplink multiuser large-scale mimo systems. *IEEE Transactions on Vehicular Technology*, 64(10), 4839–4845.
- Wu, Z., Chuan, Z., Xue, Y., Xu, S., You, X. (2016). Efficient architecture for soft-output massive MIMO detection with Gauss–Seidel method. In *Proceedings of the IEEE, Circuits and Systems (ISCAS)*, pp. 1886–1889.
- Yin, B., Wu, M., Cavallaro, J.R., Studer, C. (2015). VLSI Design of large-scale soft-output mimo detection using conjugate gradients. In *Proceedings of the IEEE Circuits and Systems (ISCAS)*, pp. 1498–1501.
- Yin, B., Wu, M., Cavallaro, J.R., Studer, C. (2014). Conjugate gradient-based soft-output detection and precoding in massive MIMO systems. In *Proceedings of the IEEE, International Workshop on Green Communications, parallel with IEEE GLOBECOM*, pp. 3696–3701.
- Zhang, P., Liu, L., Peng, G., Wei, S. (2016). Large-scale MIMO detection design and FPGA implementations using SOR method. In *Proceedings of the IEEE Inter. Conf. on Communication Software and Networks (ICCSN)*, pp. 206–210.
- Gao, X., Dai, L., Hu, Y., Wang, Z., Wang, Z. (2014). Matrix inversion-less signal detection SOR method for uplink large-scale MIMO systems. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pp. 3291–3295.
- Yu, A., Zhang, C., Zhang, S., You, X. (2016). Efficient SOR-based detection and architecture for large-scale MIMO uplink. In *Proceedings of the IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, pp. 402–405.
- Xue, Y., Zhang, C., Zhang, S., You, X. (2016). A fast-convergent pre-conditioned conjugate gradient detection for massive MIMO uplink. In *Proceedings of the IEEE, Digital Signal Processing (DSP)*, pp. 331–335.
- Jin, J., Xue, Y., Ueng, Y.-L., You, X., Zhang, C. (2017). A split pre-conditioned conjugate gradient method for massive MIMO detection. In *Proceedings of the IEEE, International Workshop on Signal Processing Systems (SiPS)*, pp. 1–6.
- Yang, S., & Hanzo, L. (2015). Fifty years of MIMO detection The road to large-scale MIMO. *IEEE Communication Surveys and Tutorials*, 17(4), 1941–1988.
- Abdaoui, A., Berbineau, M., Snoussi, H. (2007). GMRES interference canceler for doubly iterative MIMO system with a large number of antennas. In *Proceedings of the IEEE, International Symposium on Signal Processing and Information Technology*, pp. 449–453.
- Kermoal, J.-P., Schumacher, L., Pedersen, K.I., Mogensen, P.E., Frederiksen, F. (2002). A stochastic MIMO Radio channel model with experimental validation. *IEEE Journal on Selected Areas in Communications*, 20(6), 1211–1226.
- Studer, C., Fateh, S., Seethaler, D. (2011). ASIC implementation of soft-input soft-output MIMO detection using MMSE parallel interference cancellation. *IEEE Journal of Solid-State Circuits*, 46(7), 1754–1765.
- Saad, Y. (2003). Iterative methods for sparse linear systems SIAM.
- Abdaoui, A., Berbineau, M., Snoussi, H. (2008). GMRES interference canceller for MIMO relay network. In *Proceedings of the IEEE, GLOBECOM*, pp. 1–5.
- Voß, H. (2004). An Arnoldi method for nonlinear eigenvalue problems. *BIT Numerical Mathematics*, 44(2), 387–401.
- Fong, D.C.-L., & Saunders, M.A. (2012). CG versus MINRES An empirical comparison. *SQU Journal for Science*, 17(1), 44–62.
- Wright, S.J. (2015). Coordinate descent algorithms. *Mathematical Programming*, 151(1), 3–34.
- Wu, M., Dick, C., Cavallaro, J.R., Studer, C. (2016). High-throughput data detection for massive MU-MIMO-OFDM using coordinate descent. arXiv:1611.08779.
- Wubben, D., Bohnke, R., Kuhn, V., Kammeyer, K.-D. (2003). MMSE, extension of V-BLAST based on sorted QR decomposition. In *Proceedings of the IEEE Vehicular technology conference (VTC)-Fall 1*, pp. 508–512.
- Ling, F. (1991). Givens rotation based least squares lattice and related algorithms. *IEEE Transactions on Signal Processing*, 39(7), 1541–1551.



Chuan Zhang received the B.E. degree (summa cum laude) in microelectronics and the M.E. degree in very-large scale integration (VLSI) design from Nanjing University, Nanjing, China, in 2006 and 2009, respectively, and the M.S.E.E. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Minnesota, Twin Cities (UMN), USA, in 2012. He is currently an Associate Professor with the National Mobile Com-

munications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing. His current research interests include low-power high-speed VLSI design for digital signal processing and digital communication, bio-chemical computation and neuromorphic engineering, and quantum communication. He is a member of the Seasonal School of Signal Processing and Design and Implementation of Signal Processing Systems TC of the IEEE Signal Processing Society, and Circuits and Systems for Communications TC, VLSI Systems and Applications TC, and Digital Signal Processing TC of the IEEE Circuits and Systems Society. He was a co-recipient of the Best Paper Award of the IEEE Asia Pacific Conference on Circuits and Systems (APCCAS) in 2016, the Best (Student) Paper Award of the IEEE International Conference on DSP in 2016, three excellent paper awards and the Excellent Poster Presentation Award of the International Collaboration Symposium on Information Production and Systems in 2016 and 2017, two best (student) paper awards of the IEEE International Conference on ASIC in 2015 and 2017, the Best Paper Award Nomination of the IEEE Workshop on Signal Processing Systems in 2015, the Merit (Student) Paper Award of the IEEE APCCAS in 2008. He received the Three-Year University-Wide Graduate School Fellowship of UMN, and the Doctoral Dissertation Fellowship of UMN.



Yufeng Yang received the Bachelor of Engineering degree in Information Engineering from Southeast University, China, in 2018. He is working towards the Master of Science degree in Electrical and Computer Engineering from Georgia Institute of Technology, USA. His research interests include efficient algorithms and VLSI architectures for massive MIMO signal detection, automatic speaker recognition.



Shunqing Zhang received the B.S. degree from the Department of Microelectronics, Fudan University, Shanghai, China, in 2005, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, in 2009.

He was with the Communication Technologies Laboratory, Huawei Technologies, as a Research Engineer and then a Senior Research Engineer from 2009 to 2014, and a

Senior Research Scientist of Intel Collaborative Research Institute on Mobile Networking and Computing, Intel Labs from 2015 to 2017. Since 2017, he has been with the School of Communication and Information Engineering, Shanghai University, Shanghai, China, as a Full Professor. His current research interests include energy efficient 5G/5G+ communication networks, hybrid computing platform, and joint radio frequency and baseband design. He has published over 60 peer-reviewed journal and conference papers, as well as over 50 granted patents. He is a senior member of IEEE, received the National Young 1000-Talents Program, and won the paper award for Advances in Communications from IEEE Communications Society in 2017.



Zaichen Zhang was born in Nanjing, China in 1975. He received BS and MS degrees in Electrical and Information Engineering from Southeast University, Nanjing, China in 1996 and 1999, respectively, and PhD degree in Electrical and Electronic Engineering from The University of Hong Kong, Hong Kong, China in 2002. From 2002 to 2004, he was a Postdoctoral Fellow at the National Mobile Communications Research Lab., Southeast University, China.

He joined School of Information Science and Engineering, Southeast University, China in 2004 and is currently a Professor there. He has published over 180 papers and issued 26 patents. His current research interests include beyond 5G intelligent wireless systems, optical wireless communications, and quantum information technologies.



Xiaohu You received B.S., M.S. and Ph.D. degrees in electrical engineering from Nanjing Institute of Technology, Nanjing, China, in 1982, 1985, and 1989, respectively. From 1987 to 1989, he was with Nanjing Institute of Technology as a Lecturer. From 1990 to the present time, he has been with Southeast University, first as an Associate Professor and later as a Professor. His research interests include mobile communications, adaptive signal

processing, and artificial neural networks with applications to communications and biomedical engineering. He has contributed over 40 IEEE journal papers and 2 books in the areas of adaptive signal processing, neural networks and their applications to communication systems. He was the Premier Foundation Investigator of the China National Science Foundation. From 1999 to 2002, he was the Principal Expert of the C3G Project, responsible for organizing China's 3G Mobile Communications R&D Activities. From 2001-2006, he was the Principal Expert of the national 863 FuTURE Project. He received the excellent paper award from the China Institute of Communications in 1987 and Elite Outstanding Young Teacher Awards from Southeast University in 1990, 1991, and 1993. He was a recipient of the 1989 Young Teacher Award of Fok Ying Tung Education Foundation, State Education Commission of China. Dr. You now is the Chairman of IEEE Nanjing Section. He was selected as IEEE Fellow for his contributions to development of mobile communications in China in 2012.

Affiliations

Chuan Zhang¹  · Yufeng Yang² · Shunqing Zhang³ · Zaichen Zhang¹ · Xiaohu You¹

Yufeng Yang
yfyang@gatech.edu

Shunqing Zhang
shunqing@shu.edu.cn

Zaichen Zhang
zczhang@seu.edu.cn

Xiaohu You
xhyu@seu.edu.cn

¹ Lab of Efficient Architectures for Digital-communication and Signal-processing (LEADS), National Mobile Communications Research Laboratory, Quantum Information Center, Southeast University, Nanjing, China

² School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA USA

³ Shanghai Institute for Advanced Communications and Data Science, Shanghai University, Shanghai, China