Randomized Iterative Sampling Decoding Algorithm For Large-Scale MIMO Detection

Zheng Wang, *Senior Member, IEEE*, Shengheng Liu, *Senior Member, IEEE*, Yili Xia, *Member, IEEE*, Cong Ling, *Member, IEEE* and Yongming Huang, *Senior Member, IEEE*

***Abstract*—In this paper, the paradigm of the traditional iterative decoding schemes for the uplink large-scale MIMO detection is extended by sampling in an Markov chain Monte Carlo (MCMC) way. Different from iterative decoding whose performance is upper bounded by the suboptimal linear decoding scheme like ZF or MMSE, the proposed iterative random sampling decoding (IRSD) algorithm is capable of achieving the optimal ML decoding performance with the increment of Markov moves, thus establishing a flexible trade-off between suboptimal and optimal decoding performance. According to convergence analysis, we show that the Markov chain induced by IRSD algorithm experiences the exponential convergence, and its related convergence rate is also derived in detail. Based on it, the Markov mixing becomes tractable, followed by the decoding optimization with respect to the standard deviation of the target distribution. Meanwhile, further decoding performance enhancement and parallel implementation are also studied so that the proposed IRSD algorithm is well suited for various cases of large-scale MIMO systems.**

***Index Terms*—Large-scale MIMO detection, massive MIMO detection, iterative methods, sampling decoding, MCMC.**

I. Introduction

The large-scale multiple-input multiple-output (MIMO) system has become a promising extension of MIMO in 5G and beyond 5G, which boosts the network capacity on a much greater scale without extra bandwidth [1]–[4]. However, the dramatically increased system size also places a pressing challenge on the signal detection in the uplink [5], [6]. To this end, a number of low-complexity iterative decoding schemes have been proposed, which approach the suboptimal linear decoding performance via an iterative way [7]–[12]. Specifically, the iterative decoding schemes for large-scale MIMO detection are mainly based on series expansion and matrix splitting methods [13]–[16]. Compared to series expansion, iterative decoding schemes based on matrix splitting like Jacobi, Richardson and Gauss-Seidel (GS) iterations turn out to be more efficient [17]. In [18], a randomized iterative decoding named as the modified randomized iterative detection algorithm (MRIDA) is given, which not only achieves a faster iteration convergence but also enjoys the global convergence.

However, all those iteration-based decoding schemes belong to the suboptimal linear decoding. Theoretically, there is a substantial performance gap between the suboptimal linear decoding and the optimal ML decoding, and the optimal ML decoding performance can be approximated by linear decoding like zero forcing (ZF) and minimum mean-square error (MMSE) only when the number of received antennas at base station (BS) (denoted by n) is sufficiently larger than the number of transmitted antennas of user equipments (UE) (denoted by m), i.e., $n \gg m$ [19]. In fact, compared to the increased number of the received antennas at BS, the number of the transmitted antennas at UE side has also improved accordingly. Moreover, with the rapid increment of UE, the total number of antennas at UE side has increased significantly. Therefore, the environment of wireless communications has become much more complicated than before [20], [21], so that a flexible decoding scheme that well suits various scenarios of large-scale MIMO systems is highly desired. Unfortunately, since the performance of iterative decoding is upper bounded by the linear decoding performance, those iteration-based decoding schemes fail to achieve a better decoding performance even though numerous number of iterations are carried out, rendering them rather limited in the various cases of large-scale MIMO [22], [23].

Recently, sampling turns out to be a powerful strategy for solving decoding problems [24], [25]. In particular, the traditional decoding problem can be cast as an equivalent sampling problem [26]–[29]. By doing this, the optimal decoding solution can be encountered by sampling from a multi-dimensional discrete Gaussian distribution since it naturally entails the largest sampling probability. However, decoding by sampling heavily relies on how to successfully sample from the target discrete Gaussian distribution [30], [31], and it is rather difficult in contrast to the case of sampling from the continuous Gaussian density. Therefore, to effectively exploit the potential behind the randomness, the methods of Markov chain Monte Carlo (MCMC) were introduced into sampling decoding [32]–[34]. Specifically, MCMC attempts to sample from the target distribution via a Markov chain, which randomly generates the next sample conditioned on previous samples. After a burn-in period, which is also known as the mixing time, the Markov chain will reach into a stationary distribution, and successful sampling from the complex target distribution can be carried out thereafter [35].

This work was supported in part by the National Natural Science Foundation of China under Grant 61801216, 62225107, 61771124 and 61720106003, in part by supported by the National Key R&D Program of China under Grant 2018YFB1800801, in part by ZTE Corporation under Research Program 2023ZTE01-04, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180420, in part by the Zhi Shan Young Scholar Program of Southeast University, in part by the Fundamental Research Funds for the Central Universities 2242022k60002.

Z. Wang, S. Liu Y. Xia and Y. Huang are with School of Information Science and Engineering, and with Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 210096, China (e-mail: z.wang@ieee.org). Cong Ling is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: cling@ieee.org).

In this paper, in order to achieve a flexible iterative decoding scheme with tractable performance between suboptimal linear and optimal ML decoding, a general sampling decoding framework with respect to the traditional iterative methods is proposed, which adopts the random sampling to the iterations in an MCMC way. First of all, the iterative random sampling decoding (IRSD) algorithm is proposed. Without loss of generalization, it is described based on the traditional GS iteration and we show it is capable of achieving the optimal decoding performance with the increase of the number of Markov moves. By doing this, a flexible decoding performance can be achieved with the adjustable number of Markov moves. Secondly, the convergence analysis of the Markov mixing in the proposed IRSD algorithm is performed, where the explicit convergence rate as well as the mixing time of the Markov chain is derived. Then, based on the accessible mixing time, the sampling decoding by IRSD algorithm is optimized for a better decoding efficiency, which leads to a better choice of the standard deviation $\sigma > 0$ in the target distribution Π. Thirdly, the extensions of the proposed IRSD algorithm to successive overrelaxation (SOR) and Jacobi iterations are studied in detail, where further decoding enhancement and parallel implementation can be achieved respectively.

The rest of this paper is organized as follows. Section II introduces the system model and the traditional iterative decoding schemes, and briefly reviews the basics of MCMC methods. In Section III, the proposed IRSD algorithm is described based on the classic GS iteration, which results in a flexible decoding performance controlled by the number of Markov moves. Meanwhile, convergence analysis is also given to show the accessible convergence rate and mixing time of the Markov chain in the proposed IRSD algorithm. In Section IV, the optimization about the standard deviation is presented for a better decoding efficiency. In Section V, extensions of the proposed IRSD algorithm to other traditional iterations like SOR and Jacobi iterations are presented to achieve further decoding enhancement and parallel implementation respectively. After that, simulations of the proposed IRSD algorithm over different iteration schemes for uplink large-scale MIMO detection are presented in Section VI. Finally, Section VII concludes the paper.

Notation: Matrices and column vectors are denoted by upper and lowercase boldface letters, and the transpose, inverse, pseudoinverse of a matrix $\mathbf{B}$ by $\mathbf{B}^T, \mathbf{B}^{-1}$, 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 $\mathbf{B}$. $\lceil x \rfloor$ denotes rounding to the integer closest to x. If x is a complex number, $\lceil x \rfloor$ rounds the real and imaginary parts separately. Finally, in this paper, the computational complexity is measured by the number of arithmetic operations (additions, multiplications, comparisons, etc.).

II. Preliminary

In this section, we introduce the system model, the traditional iterative decoding schemes as well as the classic MCMC methods needed to describe and analyze the proposed IRSD algorithm.

A. System Model

Consider the decoding of an $n \times m$ real-valued system with $n \geq m$. The extension to the complex-valued system is straightforward [24]. Let $\mathbf{s} \in \mathcal{X}^m \subseteq \mathbb{Z}^m$ denote the transmitted signal, where The corresponding received signal $\mathbf{c}$ is given by

$$\mathbf{c} = \mathbf{H}\mathbf{s} + \mathbf{w} \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^n$ is the noise vector with zero mean and variance σ_w^2, $\mathbf{H}$ is an $n \times m$ matrix of channel coefficients. Typically, the conventional maximum likelihood (ML) reads

$$\widehat{\mathbf{s}}_{\text{ml}} = \arg\min_{\mathbf{s}\in\mathcal{X}^m} \|\mathbf{c} - \mathbf{H}\mathbf{s}\|^2 \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm. In essence, this problem belongs to the integer least squares (ILS) problem, and is also known as the closest vector problem (CVP) in lattice decoding, which is NP-hard in high dimensional systems [36], [37].

In [19], by fully taking advantages of receive diversity when $n \gg m$, it has been shown that the near optimal decoding performance can be achieved by the traditional linear decoding schemes like ZF or MMSE. Specifically, the linear ZF decoding returns the following solution

$$\widehat{\mathbf{s}}_{\text{zf}} = \lceil \widetilde{\mathbf{s}}_{\text{zf}} \rfloor_Q \in \mathcal{X}^m \tag{3}$$

with

$$\widetilde{\mathbf{s}}_{\text{zf}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{c}, \tag{4}$$

where $\lceil\cdot\rfloor_Q$ denotes the direct rounding according to the discrete constellation $\mathcal{X}^m$. The similar operation works in MMSE decoding with $\widetilde{\mathbf{s}}_{\text{mmse}} = \left(\mathbf{H}^T\mathbf{H} + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{H}^T\mathbf{c}$. More precisely, $\widetilde{\mathbf{s}}_{\text{zf}}$ is the decoding solution of the following least squares (LS) problem

$$\widetilde{\mathbf{s}}_{\text{zf}} = \arg\min_{\mathbf{s}\in\mathbb{R}^m} \|\mathbf{c} - \mathbf{H}\mathbf{s}\|^2, \tag{5}$$

which offers a good approximation to the solution of ILS problem in (2) when $n \gg m$. Nevertheless, the implementation of ZF or MMSE decoding is still challenging due to the matrix inversion with computational complexity $O(m^3)$.

B. Suboptimal Linear Decoding based on Iterative Methods

In order to bypass the matrix inversion, a number of low-complexity decoding schemes based on iterative methods are proposed. Specifically, the ZF decoding in (4) (or MMSE decoding) can be interpreted by an equivalent system model, which decodes a linear system

$$\mathbf{A}\mathbf{s} = \mathbf{b}. \tag{6}$$

Here, $\mathbf{b} = \mathbf{H}^T\mathbf{c} \in \mathbb{R}^m$, the ZF filtering matrix $\mathbf{A} = \mathbf{H}^T\mathbf{H} \in \mathbb{R}^{m\times m}$ is a symmetric positive matrix (so is the MMSE filtering matrix $\mathbf{A} = \mathbf{H}^T\mathbf{H} + \sigma_w^2\mathbf{I}$).

To solve the linear system in (6), the matrix splitting about $\mathbf{A} = \mathbf{M} + \mathbf{N}$ is applied with $\mathbf{M} \in \mathbb{R}^{m\times m}$ and $\mathbf{N} \in \mathbb{R}^{m\times m}$ so that

$$\mathbf{s} = \mathbf{G}\mathbf{s} + \mathbf{g} \tag{7}$$

with $\mathbf{G} = -\mathbf{M}^{-1}\mathbf{N} = \mathbf{I} - \mathbf{M}^{-1}\mathbf{A} \in \mathbb{R}^{m\times m}$ and $\mathbf{g} = \mathbf{M}^{-1}\mathbf{b} \in \mathbb{R}^m$. From (7), iterative methods compute the successive approximations to the solution by repeatedly applying the following iterations [38]

$$\widetilde{\mathbf{s}}^{t+1} = \mathbf{G}\widetilde{\mathbf{s}}^{t} + \mathbf{g}, \tag{8}$$

where $\mathbf{G}$ is known as the *iteration matrix*. Moreover, the convergence of iterative methods is guaranteed if [39]

$$\lim_{t\to\infty} \mathbf{G}^t = \mathbf{0}, \tag{9}$$

and $\widetilde{\mathbf{s}}^{t+1}$ will gradually approach to $\widetilde{\mathbf{s}}_{\text{zf}}$ along the increment of iterations.

Intuitively, the choices of $\mathbf{M}$ and $\mathbf{N}$ for the matrix splitting $\mathbf{A} = \mathbf{M} + \mathbf{N}$ are key to iterative methods. In Jacobi iteration, the matrix splitting is set with $\mathbf{M} = \mathbf{D}$ and $\mathbf{N} = \mathbf{L} + \mathbf{U}$, which leads to the following iterations [17]

$$\mathbf{D}\widetilde{\mathbf{s}}^{t+1} = -(\mathbf{U} + \mathbf{L})\widetilde{\mathbf{s}}^{t} + \mathbf{b}. \tag{10}$$

Here, $\mathbf{D} \in \mathbb{R}^{m\times m}$, $\mathbf{L} \in \mathbb{R}^{m\times m}$ and $\mathbf{U} \in \mathbb{R}^{m\times m}$ respectively stand for the diagonal components, the strictly lower triangular components and the strictly upper triangular components of matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{D} + \mathbf{L} + \mathbf{U}$ and $\mathbf{L} = \mathbf{U}^T$. More precisely, the update of $\widetilde{s}_i^{t+1}$, $1 \le i \le m$ in Jacobi iteration is calculated by

$$\widetilde{s}_i^{t+1} = -\frac{1}{d_{i,i}}\left(\sum_{j=1}^{i-1} l_{i,j}\widetilde{s}_j^{t} + \sum_{j=i+1}^{m} u_{i,j}\widetilde{s}_j^{t} - b_i\right), \tag{11}$$

where $u_{i,j} \in \mathbb{R}$, $l_{i,j} \in \mathbb{R}$ and $d_{i,j} \in \mathbb{R}$ standard for the element of matrix $\mathbf{U}$, $\mathbf{L}$ and $\mathbf{D}$ respectively. As for Richardson iteration, $\mathbf{M}$ and $\mathbf{N}$ are set as $\mathbf{M} = \frac{1}{\omega}\mathbf{I}$ and $\mathbf{N} = \mathbf{A} - \frac{1}{\omega}\mathbf{I}$ respectively, where the coefficient $\omega > 0$ is known as the relaxation factor.

In order to achieve a faster convergence performance, Gauss-Seidel (GS) iteration with $\mathbf{M} = \mathbf{D} + \mathbf{U}$ and $\mathbf{N} = \mathbf{L}$ is introduced as [40]

$$(\mathbf{D} + \mathbf{U})\widetilde{\mathbf{s}}^{t+1} = -\mathbf{L}\widetilde{\mathbf{s}}^{t} + \mathbf{b}. \tag{12}$$

Different from Jacobi and Richardson iterations, the iteration of $\widetilde{\mathbf{s}}$ in GS is carried out element by element in a sequential order, where the updated elements of $\widetilde{\mathbf{s}}$ at iteration $t+1$ (i.e., $\widetilde{s}_i^{t+1}$) are also taken into account to update the rest of elements in $\widetilde{\mathbf{s}}$. Typically, the update of $\widetilde{s}_i^{t+1}$ from $i = m$ to $i = 1$ in a backwards order is calculated as

$$\widetilde{s}_i^{t+1} = -\frac{1}{d_{i,i}}\left(\sum_{j=1}^{i-1} l_{i,j}\widetilde{s}_j^{t} + \sum_{j=i+1}^{m} u_{i,j}\widetilde{s}_j^{t+1} - b_i\right). \tag{13}$$

Consequently, given the iteration results $\widetilde{\mathbf{s}}^L$, $L \ge 1$, the final decoding solution of iterative methods for the ILS problem in (2) is outputted by the direct quantization as

$$\widehat{\mathbf{s}} = \lceil \widetilde{\mathbf{s}}^L \rfloor_Q \in \mathcal{X}^m. \tag{14}$$

For a better understanding, the illustration of the decoding process based on GS iteration is shown in Fig. 1. In [18], another iteration-based decoding named as MRIDA is proposed, which achieves a faster convergence performance than GS iteration.

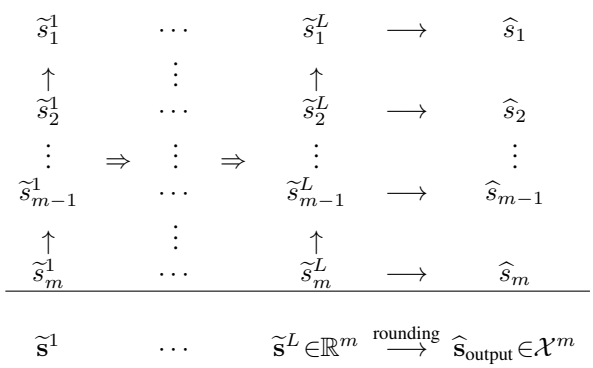

Fig. 1. Illustration of the detection based on Gauss-Seidel iteration. Arrows about $\widetilde{s}_i^t \to \widetilde{s}_{i-1}^t$ comes from the calculation in (13) while arrows about $\widetilde{s}_i^L \to \widehat{s}_i$ denote the direct rounding by $\widehat{s}_i = \lceil \widetilde{s}_i^L \rfloor_Q \in \mathcal{X}$.

Nevertheless, MRIDA still belongs to the suboptimal linear decoding so that its performance is also upper bounded by ZF or MMSE no matter how many iterations are carried out.

C. Markov Chain Monte Carlo Methods

As a foremost sampling scheme in MCMC, the Metropolis-Hastings (MH) algorithm tries to sample from the target invariant distribution $\Pi(\mathbf{s})$ via a proposal distribution [35] $q(\mathbf{x}, \mathbf{y})$[1]. To be more specific, given the current state $\mathbf{x}$ for Markov chain $\mathbf{S}^t$, a state candidate $\mathbf{y}$ for the next Markov move $\mathbf{S}^{t+1}$ is generated from the proposal distribution $q(\mathbf{x}, \mathbf{y})$. Then the acceptance ratio α is computed by [41]

$$\alpha = \min\left\{1, \frac{\Pi(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\Pi(\mathbf{x})q(\mathbf{x}, \mathbf{y})}\right\}, \tag{15}$$

and $\mathbf{y}$ will be accepted as the new state (i.e., $\mathbf{S}^{t+1} = \mathbf{y}$) with probability α. Otherwise, $\mathbf{x}$ will be retained as $\mathbf{S}^{t+1} = \mathbf{x}$. In this way, a Markov chain $\{\mathbf{S}^0, \mathbf{S}^1, \ldots\}$ is established with the transition probability $P(\mathbf{x}, \mathbf{y})$ as follows:

$$P(\mathbf{x}, \mathbf{y}) = \begin{cases} q(\mathbf{x}, \mathbf{y})\alpha & \text{if } \mathbf{y} \ne \mathbf{x}, \\ 1 - \sum_{\mathbf{z}\ne\mathbf{x}} q(\mathbf{x}, \mathbf{z})\alpha & \text{if } \mathbf{y} = \mathbf{x}. \end{cases} \tag{16}$$

Interestingly, the proposal distribution $q(\mathbf{x}, \mathbf{y})$ in MH algorithm can be any fixed distribution, but it is quite challenging to find a suitable one with satisfactory convergence. In principle, Gibbs sampling is a special case of MH algorithm by letting the proposal distribution be the univariate conditional distribution, i.e., [42]

$$q(\mathbf{x}, \mathbf{y}) = \Pi(s_i | \mathbf{s}_{[-i]}), \tag{17}$$

where $\mathbf{s}_{[-i]} \triangleq [s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_m]^T$ is leaving unchanged during the univariate sampling about s_i. Clearly, with (17), it is easy to verify that the acceptance ratio α of Gibbs algorithm is always 1. By repeating such a procedure with a certain scan scheme, a Markov chain $\{\mathbf{S}^0, \mathbf{S}^1, \ldots\}$ is established. Apart from the random scan who randomly updates the component of $\mathbf{s}$, systematic scan proceeds the update of s_i in a sequential order from s_m to s_1 (i.e., $\mathbf{s}_{[-i]} = [s_1^t, \ldots, s_{i-1}^t, s_{i+1}^{t+1}, \ldots, s_m^{t+1}]^T$), which is more preferable in practice.

[1] $q(\mathbf{x}, \mathbf{y})$ can also be expressed as $q(\mathbf{y} \mid \mathbf{x})$, namely, $q(\mathbf{x}, \mathbf{y}) \triangleq q(\mathbf{y} \mid \mathbf{x})$.

Additionally, according to the *convergence theorem* of MCMC, the exponential convergence of MH algorithm and Gibbs sampling can be easily verified.

Theorem 1 ([35]). *For finite state space $\mathbf{s} \in \Omega$, suppose that the transition matrix of the Markov chain is irreducible and aperiodic with the invariant stationary distribution Π, then there exist constants $0 < \varrho < 1$ and $C > 0$ such that*

$$\|P^t(\mathbf{s}, \cdot) - \Pi(\cdot)\|_{TV} \leq C\varrho^t. \tag{18}$$

Here, $t \geq 1$ denotes the index of Markov moves, $\|\cdot\|_{TV}$ represents the total variation distance, Π is the target invariant distribution, $P^t(\mathbf{s}; \cdot)$ indicates a row of the transition matrix $\mathbf{P}$ after t Markov moves with the initial state $\mathbf{s}$. Clearly, coefficient ϱ is the convergence rate of the Markov chain, where a smaller ϱ corresponds to a faster convergence performance in approaching the target distribution Π.

III. Iterative Random Sampling Decoding

In order to achieve low decoding complexity, iterative decoding schemes are proposed to approach the performance of linear ZF or MMSE iteration by iteration. Nevertheless, the performance of these iterative decoding schemes are strictly upper bounded by ZF or MMSE decoding no matter how many iterations are carried out, namely,

$$\lim_{t\to\infty} \lceil \widetilde{\mathbf{s}}^t \rfloor_Q = \widehat{\mathbf{s}}_{\text{zf}}. \tag{19}$$

In a word, these traditional iterative decoding schemes are inapplicable when the flexible performance between suboptimal linear and optimal ML decoding is needed. In fact, such a performance requirement does widely exist in various cases of large-scale MIMO systems except the case $n \gg m$. Theoretically, this is due to the fact that traditional iterative schemes are designed to solve the LS problem in (5) rather than the ILS problem in (2), and the solution of LS problem only offers an approximation to that of the ILS problem [18]. To overcome the performance limitation of the traditional iterative decoding schemes, we adopt the random sampling into iterations, which is capable of solving the ILS problem in (2) as

$$\lim_{t\to\infty} \widehat{\mathbf{s}}^t = \widehat{\mathbf{s}}_{\text{ml}}. \tag{20}$$

Different from iterative decoding, the proposed IRSD algorithm tries to return the optimal decoding solution $\widehat{\mathbf{s}}_{\text{ml}}$ by sampling from an invariant target distribution Π

$$\Pi(\mathbf{s}) = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{s}-\mathbf{b}\|^2}}{\sum_{\mathbf{s}\in\mathcal{X}^m} e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{s}-\mathbf{b}\|^2}}. \tag{21}$$

Typically, it is clear to see that the optimal decoding solution $\widehat{\mathbf{s}}_{\text{ml}}$ of the ILS problem in (2) also has the largest sampling probability in distribution Π, i.e.,

$$\begin{aligned}\widehat{\mathbf{s}}_{\text{ml}} &= \arg\min_{\mathbf{s}\in\mathcal{X}^m} \|\mathbf{H}\mathbf{s}-\mathbf{c}\|^2 = \arg\min_{\mathbf{s}\in\mathcal{X}^m} \|\mathbf{A}\mathbf{s}-\mathbf{b}\|^2 \\ &= \arg\max_{\mathbf{s}\in\mathcal{X}^m} \Pi(\mathbf{s}).\end{aligned} \tag{22}$$

Therefore, if one can sample from Π, the optimal decoding solution $\widehat{\mathbf{s}}_{\text{ml}}$ can be obtained by sampling. Most importantly, due to its largest sampling probability, $\widehat{\mathbf{s}}_{\text{ml}}$ is most likely to be sampled during the multiple independent samplings, thus providing an efficient way to decode the ILS problem in (2). Unfortunately, it is rather difficult to perform the sampling even from a low-dimensional discrete Gaussian distribution. For this reason, MCMC methods are applied in the proposed IRSD algorithm, and we show that the distribution induced by the Markov moves converges to the target distribution Π in an exponential way. By doing this, the ILS problem in (2) can be effectively solved by sampling from Π.

A. Algorithm Description

Here, for a better presentation, the proposed iterative random sampling decoding (IRSD) algorithm is described based on GS iteration, where extensions to other iterative schemes are straightforward. In particular, different from GS iteration that sequentially computes $\widetilde{s}_i^{t+1} \in \mathbb{R}$ in (13), we update $\widehat{s}_i^{t+1} \in \mathcal{X}$ by random sampling from a discrete Gaussian distribution centered at $\widetilde{s}_i^{t+1}$, namely,

$$\widehat{s}_i^{t+1} \sim p_{\text{gs}}(s_i) = \frac{e^{-\frac{1}{2\sigma_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}}{\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\sigma_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \tag{23}$$

with

$$\widetilde{s}_i^{t+1} = -\frac{1}{d_{i,i}}\left(\sum_{j=1}^{i-1} l_{i,j}\widehat{s}_j^t + \sum_{j=i+1}^{m} u_{i,j}\widehat{s}_j^{t+1} - b_i\right), \tag{24}$$

where $\sigma_i = \sigma/|d_{i,i}|$. More specifically, as shown in (24), $\widehat{s}_1^t, \ldots, \widehat{s}_{i-1}^t$ and the updated $\widehat{s}_{i+1}^{t+1}, \ldots, \widehat{s}_m^{t+1}$ are applied to calculate $\widetilde{s}_i^{t+1}$, then the random sampling in (23) is performed to yield the updated $\widehat{s}_i^{t+1}$. In this way, the update of $\widehat{s}_i^{t+1}$ is carried out sequentially from $i = m$ to $i = 1$, which leads to $\mathbf{s}^{t+1}$. To make it clear, this decoding process is depicted in Fig. 2.

Typically, let us focus on $p_{\text{gs}}(s_i)$ in (23), and it follows that

$$\begin{aligned}
p_{\text{gs}}(s_i) &= \frac{e^{-\frac{1}{2\sigma_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}}{\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\sigma_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \\
&= \frac{e^{-\frac{1}{2\sigma^2}\|d_{i,i}s_i+\sum_{j=i+1}^m u_{i,j}\widehat{s}_j^{t+1}+\sum_{j=1}^{i-1} l_{i,j}\widehat{s}_j^t-b_i\|^2}}{\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\sigma^2}\|d_{i,i}s_i+\sum_{j=i+1}^m u_{i,j}\widehat{s}_j^{t+1}+\sum_{j=1}^{i-1} l_{i,j}\widehat{s}_j^t-b_i\|^2}} \\
&= \frac{e^{-\frac{1}{2\sigma^2}\|a_{i,i}s_i+\sum_{j=i+1}^m a_{i,j}\widehat{s}_j^{t+1}+\sum_{j=1}^{i-1} a_{i,j}\widehat{s}_j^t-b_i\|^2}}{\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\sigma^2}\|a_{i,i}s_i+\sum_{j=i+1}^m a_{i,j}\widehat{s}_j^{t+1}+\sum_{j=1}^{i-1} a_{i,j}\widehat{s}_j^t-b_i\|^2}}
\end{aligned} \tag{25}$$

$$\begin{aligned}
&= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{h}_i^T\mathbf{h}_i s_i+\sum_{j=i+1}^m \mathbf{h}_i^T\mathbf{h}_j\widehat{s}_j^{t+1}+\sum_{j=1}^{i-1}\mathbf{h}_i^T\mathbf{h}_j\widehat{s}_j^t-\mathbf{h}_i^T\mathbf{c}\|^2}}{\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\sigma^2}\|\mathbf{h}_i^T\mathbf{h}_i s_i+\sum_{j=i+1}^m \mathbf{h}_i^T\mathbf{h}_j\widehat{s}_j^{t+1}+\sum_{j=1}^{i-1}\mathbf{h}_i^T\mathbf{h}_j\widehat{s}_j^t-\mathbf{h}_i^T\mathbf{c}\|^2}} \\
&= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{h}_i^T(\mathbf{H}\mathbf{s}-\mathbf{c})\|^2}}{\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\sigma^2}\|\mathbf{h}_i^T(\mathbf{H}\mathbf{s}-\mathbf{c})\|^2}}
\end{aligned} \tag{26}$$

with $\mathbf{s} = [\widehat{s}_1^t, \ldots, \widehat{s}_{i-1}^t, s_i, \widehat{s}_{i+1}^{t+1}, \ldots, \widehat{s}_m^{t+1}]^T$. Here, $a_{i,j}$ standards for the element of matrix $\mathbf{A}$ snd $\mathbf{h}_i$ denotes the i-th column of matrix $\mathbf{H}$. Intuitively, the sampling mechanism of $p_{\text{gs}}(s_i)$ is similar to Gibbs sampling as the other $m-1$ components of $\mathbf{s}$ (i.e., $\mathbf{s}_{[-i]} = [\widehat{s}_1^t, \ldots, \widehat{s}_{i-1}^t, \widehat{s}_{i+1}^{t+1}, \ldots, \widehat{s}_m^{t+1}]$) are also taken into account to update $\widehat{s}_i^{t+1}$. However, due to

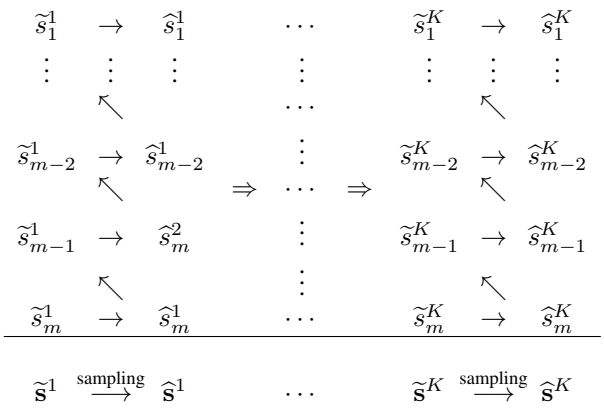

Fig. 2. Illustration of the proposed IRSD algorithm. Arrows about $\widetilde{s}_i^t \to \widehat{s}_i^t$ denote the sampling by (23) while arrows about $\widehat{s}_i^t \to \widetilde{s}_{i-1}^t$ represent the calculations in (24).

the existence of the term $\mathbf{h}_i^T$ in (26), the univariate sampling of $p_{\text{gs}}(s_i)$ can not be attributed to Gibbs sampling, which means the convergence of the above random sampling to the target distribution Π is not guaranteed. For this reason, the mechanism of Metropolis-Hastings (MH) algorithm in MCMC is introduced to ensure the convergence.

Specifically, given $\widehat{\mathbf{s}}^t$, the sampling probability $p(\widehat{\mathbf{s}}^{t+1}|\widehat{\mathbf{s}}^t)$ can be expressed as

$$p(\widehat{\mathbf{s}}^{t+1}|\widehat{\mathbf{s}}^t) = p_{\text{gs}}(\widehat{s}_m^{t+1}) \cdots p_{\text{gs}}(\widehat{s}_1^{t+1}). \tag{27}$$

Then, with respect to the invariant target distribution Π in (21), the sampling probability $p(\widehat{\mathbf{s}}^{t+1}|\widehat{\mathbf{s}}^t)$ in (27) can be used as the proposal distribution $q_{\text{gs}}(\mathbf{x}, \mathbf{y})$, namely,

$$q_{\text{gs}}(\mathbf{x} = \widehat{\mathbf{s}}^t, \mathbf{y} = \widehat{\mathbf{s}}^{t+1}) = p(\widehat{\mathbf{s}}^{t+1}|\widehat{\mathbf{s}}^t). \tag{28}$$

Therefore, at each iteration, a sampling candidate $\widehat{\mathbf{s}}^{t+1}$ is obtained by sampling from the proposal distribution $q_{\text{gs}}(\mathbf{x}, \mathbf{y})$. After that, the acceptance ratio α in (15) is employed to make the decision about whether accept this sampling candidate $\mathbf{y} = \widehat{\mathbf{s}}^{t+1}$ as the Markov state of $\mathbf{S}^{t+1}$ or not. To make it clear, we summarize the related operations of generating a Markov state as the following three basic steps:

1) *Sample from the proposal distribution* $q_{gs}(\mathbf{x} = \widehat{\mathbf{s}}^t, \mathbf{y} = \widehat{\mathbf{s}}^{t+1})$ *in (28) to obtain a candidate state* $\widehat{\mathbf{s}}^{t+1}$ *for* $\mathbf{S}^{t+1}$.

2) *Calculate the acceptance ratio* α in (15).

3) *With probability* α *accept* $\mathbf{S}^{t+1} = \widehat{\mathbf{s}}^{t+1}$*; otherwise, reject* $\widehat{\mathbf{s}}^{t+1}$ *and let* $\mathbf{S}^{t+1} = \widehat{\mathbf{s}}^t$.

By doing this, in the proposed IRSD algorithm, a valid Markov chain $\{\mathbf{S}^0, \mathbf{S}^1, \ldots\}$ is established in the way of MH algorithm. From Theorem 1, we can easily arrive at the following result about the exponential convergence to the target distribution Π.

Corollary 1. *For finite state space* $\mathbf{s} \in \mathcal{X}^m$*, the Markov chain induced by the proposed IRSD algorithm is irreducible and aperiodic with the invariant stationary distribution* Π*, so that there exist constant* $0 < \varrho < 1$ *and* $C > 0$ *such that*

$$\|P^t(\mathbf{s}, \cdot) - \Pi(\cdot)\|_{TV} \leq C\varrho^t. \tag{29}$$

Clearly, Corollary 1 verifies the sampling decoding by the proposed IRSD algorithm. Among all the samples during the Markov chain, the one with the smallest Euclidean distance $\|\mathbf{As} - \mathbf{b}\|$ is selected as the decoding solution. From (29), the choice of the initial starting state of $\mathbf{S}^0$ also plays an important role in Markov mixing. The initial state $\mathbf{s}^0$ can be chosen from $\mathcal{X}^m$ arbitrarily but a closer choice of $\mathbf{s}^0$ to the center of the target distribution Π would be helpful to Markov mixing [43]. To this end, the output of GS iteration can be applied as a good initial setting for the underlying Markov mixing. To summarize, the proposed IRSD algorithm over GS iteration is outlined in Algorithm 1. Note that the rejected samples by the judgement based on the acceptance ratio α can also be taken into account for the consideration of decoding diversity. Here, we claim that the proposed IRSD algorithm is essentially different from MRIDA in [18] as the latter aims to achieve the suboptimal linear decoding performance with low complexity while the former enjoys the flexible performance between suboptimal linear decoding and optimal ML decoding.

B. Convergence Analysis

Compared to Gibbs sampling over the target distribution Π in (21), the proposed IRSD algorithm stems directly from the traditional iterative methods, making it easier to be analyzed in a theoretic way. In what follows, convergence analysis with respect to the proposed IRSD algorithm is carried out and we show that the convergence rate of the Markov mixing in IRSD is accessible. Most importantly, this leads to a tractable Markov chain while further analysis and optimization can be performed for better decoding performance and efficiency. In sharp contrast to the proposed IRSD algorithm, convergence analysis about Gibbs sampling is hard to perform, rendering it a heuristic sampling scheme for a long time [43], [44]. Moreover, besides GS iteration, further extensions to other traditional iteration schemes can be easily made by the proposed IRSD algorithm, which offers a general sampling decoding framework for the traditional iterations.

First of all, the proposal distribution $q(\mathbf{x} = \widehat{\mathbf{s}}^t, \mathbf{y} = \widehat{\mathbf{s}}^{t+1})$ in IRSD algorithm can be expressed as

$$\begin{aligned} q_{\text{gs}}(\mathbf{x} = \widehat{\mathbf{s}}^t, \mathbf{y} = \widehat{\mathbf{s}}^{t+1}) &= p_{\text{gs}}(s_m) \cdots p_{\text{gs}}(s_1) \\ &= \frac{e^{-\frac{1}{2\sigma^2}\|(\mathbf{D}+\mathbf{U})\widehat{\mathbf{s}}^{t+1} + \mathbf{L}\widehat{\mathbf{s}}^t - \mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i \in \mathcal{X}} e^{-\frac{1}{2\sigma_i^2}\|s_i - \widetilde{s}_i^{t+1}\|^2}} \\ &= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{As} - \mathbf{L}\Delta\mathbf{s} - \mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i \in \mathcal{X}} e^{-\frac{1}{2\sigma_i^2}\|s_i - \widetilde{s}_i^{t+1}\|^2}} \end{aligned} \tag{30}$$

with $\Delta\mathbf{s} = \widehat{\mathbf{s}}^{t+1} - \widehat{\mathbf{s}}^t \in \mathbb{Z}^n$.

Lemma 1. *In the Markov mixing of the proposed IRSD algorithm, there exists constant* $\beta > 0$ *such that*

$$\frac{q_{gs}(\mathbf{x} = \widehat{\mathbf{s}}^t, \mathbf{y} = \widehat{\mathbf{s}}^{t+1})}{\Pi(\mathbf{y} = \widehat{\mathbf{s}}^{t+1})} \geq \beta \cdot f(\Delta\mathbf{s}) \tag{31}$$

for all $\mathbf{x} = \widehat{\mathbf{s}}^t \in \mathcal{X}^m$ *with function* $f(\Delta\mathbf{s}) = e^{-\frac{1}{2\sigma^2}\|\mathbf{L}\Delta\mathbf{s}\|^2}$ *and*

$$\beta = \frac{\sum_{\mathbf{s} \in \mathcal{X}^m} e^{-\frac{1}{2\sigma^2}\|\mathbf{As} - \mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i \in \mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|s_i\|^2}} \tag{32}$$

Algorithm 1 IRSD algorithm over GS iteration

Require: $\mathbf{H}, \mathbf{c}, \sigma, L, K$;
Ensure: $\widehat{\mathbf{s}}_{\text{output}}$;
use traditional GS iteration in (13) to get $\widetilde{\mathbf{s}}^L$
let initial state $\widehat{\mathbf{s}}^0 = \lceil \widetilde{\mathbf{s}}^L \rfloor_Q \in \mathcal{X}^m$
for $t = 1, \ldots, K$ **do**
  sample $\widehat{\mathbf{s}}^t$ from $q_{\text{gs}}(\mathbf{x} = \widehat{\mathbf{s}}^{t-1}, \mathbf{y} = \widehat{\mathbf{s}}^t)$ in (28)
  calculate the acceptance ratio α in (15)
  generate a sample u from the uniform density $U[0, 1]$
  if $u \leq \alpha$ **then**
    $\widehat{\mathbf{s}}^t$ is accepted
  else
    $\widehat{\mathbf{s}}^t$ is rejected and let $\widehat{\mathbf{s}}^t = \widehat{\mathbf{s}}^{t-1}$
  end if
end for
Output the sample with the smallest $\|\mathbf{H}\widehat{\mathbf{s}}^t - \mathbf{c}\|$ as $\widehat{\mathbf{s}}_{\text{output}}$

The proof of Lemma 1 is provided in Appendix A.

By (15) and (30), the transition probability $P(\mathbf{x}, \mathbf{y})$ of the Markov chain in the proposed IRSD algorithm can be derived as

$$P(\mathbf{x},\mathbf{y})=\begin{cases}\min\left\{q(\mathbf{x},\mathbf{y}), \frac{\Pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\Pi(\mathbf{x})}\right\} & \text{if } \mathbf{y}\neq\mathbf{x},\\ q(\mathbf{x},\mathbf{x})+\sum\limits_{\mathbf{z}\neq\mathbf{x}}\max\left\{0, q(\mathbf{x},\mathbf{z})-\frac{\Pi(\mathbf{z})q(\mathbf{z},\mathbf{x})}{\Pi(\mathbf{x})}\right\} & \text{if } \mathbf{y}=\mathbf{x}.\end{cases} \tag{33}$$

According to Lemma 1, it is straightforward to check that the following relationship holds

$$P(\mathbf{x}, \mathbf{y}) \geq \beta f(\Delta\mathbf{x})\Pi(\mathbf{y}) \geq \delta\Pi(\mathbf{y}) \tag{34}$$

for all $\mathbf{x} = \widehat{\mathbf{s}}^t, \mathbf{y} = \widehat{\mathbf{s}}^{t+1} \in \mathcal{X}^m$, and $\Delta\mathbf{x} = \mathbf{x} - \mathbf{y} = \Delta\mathbf{s} \in \mathbb{Z}^n$ with

$$\delta = \beta \cdot \min[f(\Delta\mathbf{s})] > 0, \tag{35}$$

where $\min[\cdot]$ denotes the minimum value of the function. Clearly, δ is a constant, which is mainly determined by the given $\mathbf{A}$ and $\mathbf{b}$.

Based on (34), the *coupling technique* in MCMC is applied to arrive at the following Theorem, which specifies the exponential convergence rate of the Markov chain induced by the proposed IRSD algorithm. The related proof can be found in Theorem 1 in [26], which is omitted here.

Theorem 2. *Given the invariant target distribution* Π*, the Markov chain established by the proposed IRSD algorithm converges exponentially as*

$$\|P^t(\mathbf{s}, \cdot) - \Pi(\cdot)\|_{TV} \leq \varrho^t = (1 - \delta)^t \tag{36}$$

for all $\mathbf{s} \in \mathcal{X}^m$ *with* $\delta = \beta \cdot min[f(\Delta\mathbf{s})]$.

Obviously, $\varrho = (1-\delta)$ is the convergence rate of the Markov chain[2]. Moreover, given the value of δ in (35), the mixing time of the Markov chain $t_{\text{mix}}(\epsilon)$, which measures the time required by a Markov chain to get close to its target distribution, can be calculated as

$$t_{\text{mix}}(\epsilon) = \frac{\ln\epsilon}{\ln(1-\delta)} \leq (-\ln\epsilon) \cdot \left(\frac{1}{\delta}\right), \quad \epsilon < 1 \tag{37}$$

where the bound $\ln(1-\delta) < -\delta$ for $0 < \delta < 1$ is used here. Therefore, the mixing time is proportional to $1/\delta$, and becomes smaller as $\delta \to 1$. Here, we point out that the system setup also has an impact upon the convergence rate ϱ. Specifically, when $n \gg m$ the columns in channel matrix $\mathbf{H}$ are nearly orthogonal to each other, which results in a diagonal dominant matrix $\mathbf{A}$. In this condition, both the coefficient β and the function $f(\Delta\mathbf{s})$ approach 1, which leads to a faster Markov mixing accordingly.

IV. Decoding Optimization

We now investigate the choice of the standard deviation $\sigma > 0$ in the target distribution Π. On one hand, with the increment of σ, the distribution Π turns out to be uniform gradually, which results in a better convergence performance. This is easy to understand as a uniform distribution is straightforward to approximate [33]. However, on the other hand, a larger σ also implies a smaller probability of $\widehat{\mathbf{s}}_{\text{ml}}$ in Π. For the consideration of reliability, this means more samplings are required to obtain the optimal decoding solution $\mathbf{s}_{\text{ml}}$. To this end, how to choose σ in a reasonable way is a key problem in the proposed IRSD algorithm. An alternative choice coming from statistics can be applied by letting σ^2 be the variance of noises. Specifically, from (22), the noise $\overline{\mathbf{w}} = \mathbf{H}^T\mathbf{w}$ in system $\mathbf{As} + \overline{\mathbf{w}} = \mathbf{b}$ follows $\mathcal{N}(\mathbf{0}, \sigma_w^2\mathbf{H}^T\mathbf{H})$, which leads to

$$\sigma_{\text{noise}}^i = \sigma_w\|\mathbf{h}_i\|, \ 1 \leq i \leq m. \tag{38}$$

Unfortunately, such a choice severely suffers from the *stalling problem* as σ_w shrinks intensively with the increase of SNR [28].

Generally, in MCMC samples from the target distribution tend to be correlated with each other so that the Markov mixing time t_{mix} can be used as a sampling gap to pick up the desired independent and identically distributed (i.i.d.) random samples. In this way, the average complexity cost (in terms of the number of Markov moves) of obtaining a specific sample $\mathbf{s}$ via MCMC can be estimated by

$$C_{\text{average}}(\mathbf{s}) \triangleq \frac{t_{\text{mix}}}{\Pi(\mathbf{s})}, \tag{39}$$

which offers a straightforward way to evaluate the decoding efficiency. More specifically, the average complexity $C_{\text{average}}(\mathbf{s})$ is upper bounded as follows

$$\begin{aligned} C_{\text{average}}(\mathbf{s}) &< \log\left(\frac{1}{\epsilon}\right) \cdot \frac{1}{\delta} \cdot \frac{\sum_{\mathbf{s}\in\mathcal{X}^m} e^{-\frac{1}{2\sigma^2}\|\mathbf{As}-\mathbf{b}\|^2}}{e^{-\frac{1}{2\sigma^2}\|\mathbf{As}-\mathbf{b}\|^2}} \\ &\leq \log\left(\frac{1}{\epsilon}\right) \cdot \frac{1}{\min[f(\Delta\mathbf{s})]} \cdot \frac{\prod_{i=1}^m \sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|s_i\|^2}}{e^{-\frac{1}{2\sigma^2}\|\mathbf{As}-\mathbf{b}\|^2}} \\ &= \log\left(\frac{1}{\epsilon}\right) \cdot \frac{1}{\min[f(\Delta\mathbf{s})]} \cdot C, \end{aligned} \tag{40}$$

where

$$C = \frac{\prod_{i=1}^m \sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|s_i\|^2}}{e^{-\frac{1}{2\sigma^2}\|\mathbf{As}-\mathbf{b}\|^2}}. \tag{41}$$

[2]In theory, δ offers a lower bound for the spectral gap of the transition probability of the underlying Markov chain [26].

From (40), it is clear to see that the upper bound of $C_{\text{average}}(\mathbf{s})$ chiefly depends on C in (41). In order to achieve a lower upper bound of C, we arrive at the following result.

Theorem 3. *With $\sigma \leq \min_{1\leq i\leq m}|d_{i,i}|/(2\sqrt{\pi})$, C is upper bounded by*

$$C \leq e^{\frac{1}{2\sigma^2}\|\mathbf{As}-\mathbf{b}\|^2}. \tag{42}$$

Proof: First of all, by recalling the *Jacobi theta function* $\vartheta_3(\tau)=\sum_{i=-\infty}^{+\infty} e^{-\pi\tau i^2}$ with $\tau>0$ [45], [46], we have

$$\prod_{i=1}^{m}\sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|s_i\|^2}=\prod_{i=1}^{m}\vartheta_3(|d_{i,i}|^2/2\pi\sigma^2). \tag{43}$$

Then, by simple calculation, we can get that [47]

$$\vartheta_3(2)=\sum_{i=-\infty}^{+\infty} e^{-2\pi i^2}=\frac{\sqrt[4]{6\pi+4\sqrt{2}\pi}}{2\Gamma(\frac{3}{4})}=1.0039, \tag{44}$$

where $\Gamma(\cdot)$ stands for the *Gamma function*. Since $\vartheta_3(\tau)$ is monotonically decreasing with τ (i.e., $\lim_{\tau\to\infty}\inf\vartheta_3(\tau)=1$), one can set

$$\frac{\min_{1\leq i\leq m}|d_{i,i}|^2}{2\pi\sigma^2}\geq 2, \tag{45}$$

so that the following bound holds

$$\prod_{i=1}^{m}\vartheta_3(|d_{i,i}|^2/2\pi\sigma^2)\leq\vartheta_3^m(2)=1.0039^m. \tag{46}$$

In this way, the product of exponential terms $\prod_{i=1}^{m}\vartheta_3(|d_{i,i}|^2/2\pi\sigma^2)$ in C is close to 1 even for values of m up to hundreds (e.g., $1.0039^{100}=1.4467$). Therefore, if σ satisfies the condition in (45), namely

$$\sigma\leq\min_{1\leq i\leq m}|d_{i,i}|/(2\sqrt{\pi}), \tag{47}$$

then we have

$$C\leq 1.0039^m\cdot e^{\frac{1}{2\sigma^2}\|\mathbf{As}-\mathbf{b}\|^2}\approx e^{\frac{1}{2\sigma^2}\|\mathbf{As}-\mathbf{b}\|^2}, \tag{48}$$

completing the proof. ∎

Given $\sigma\leq\min_{1\leq i\leq m}|d_{i,i}|/(2\sqrt{\pi})$ in Theorem 3, a large σ is preferred for a small value of $e^{\frac{1}{2\sigma^2}\|\mathbf{As}-\mathbf{b}\|^2}$. Hence, for a lower upper bound of C, we set

$$\sigma_{\text{optimized}}=\min_{1\leq i\leq m}|d_{i,i}|/(2\sqrt{\pi}) \tag{49}$$

in the proposed IRSD algorithm, thus leading to an efficient sampling decoding. Then, it is clear to see that the upper bound of the average complexity $C_{\text{average}}(\mathbf{s})$ is proportion to the Euclidean distance $\|\mathbf{As}-\mathbf{b}\|$ while the optimal decoding solution $\widehat{\mathbf{s}}_{\text{ml}}$ has the smallest complexity cost to be obtained. To be more specific, we can arrive at the following result, where the proof is omitted due to the simplicity.

Corollary 2. *The average number of Markov moves required to solve the ILS problem in (2) by the proposed IRSD algorithm over GS iteration is $O(e^{\frac{2\pi}{\min_i^2|d_{i,i}|}\|\mathbf{A}\widehat{\mathbf{s}}_{ml}-\mathbf{b}\|^2})$.*

Given Corollary 2, a flexible decoding performance trade-off between GS iteration and ML decoding can be achieved by adjusting the number of Markov move t. However, as shown in Corollary 2, sufficiently large number of Markov moves is required to achieve the ML decoding performance. Therefore, a mild number of Markov moves is recommended in practice to obtain considerable performance gain with acceptable number of Markov moves. More specifically, as for the computational complexity of each Markov move, given (25), the complexity of perform sampling about $p_{\text{gs}}(s_i)$ with $s_i\in\mathcal{X}$ is $O(m\cdot|\mathcal{X}|)$, so that the complexity of generating the sample from the proposed distribution is $O(m^2\cdot|\mathcal{X}|)$. Meanwhile, complexity $O(m^2\cdot|\mathcal{X}|)$ is required to calculate the acceptance ratio α while the complexity of each iteration in GS is $O(m^2)$. Therefore, the total computational complexity of each Markov move in IRSD algorithm over GS iteration is $O(m^2\cdot|\mathcal{X}|)$. When small constellation $\mathcal{X}$ (e.g., 4-QAM or 16-QAM) is used, the computational complexity per Markov move or iteration the proposed IRSD algorithm can be attributed to $O(m^2)$, which is competitive compared to traditional iterative decoding schemes.

V. Further Enhancement and Parallel Implementation

In this section, the proposed IRSD algorithm is adopted to successive overrelaxation (SOR) iteration and Jacobi iteration respectively for further decoding enhancement and parallel implementation.

A. Decoding Enhancement by SOR

Based on GS iteration, SOR iteration is proposed for a better convergence performance, which introduces the relaxation factor $1<\omega<2$ into the iterations as [40]

$$(\mathbf{D}+\omega\mathbf{U})\widetilde{\mathbf{s}}^{t+1}=[(1-\omega)\mathbf{D}-\omega\mathbf{L}]\widetilde{\mathbf{s}}^{t}+\omega\mathbf{b}. \tag{50}$$

From (50), each element of $\widetilde{\mathbf{s}}^{t+1}$ in SOR iteration is calculated by

$$\widetilde{s}_i^{t+1}=(1-\omega)\widetilde{s}_i^{t}-\frac{\omega}{d_{i,i}}\left(\sum_{j=1}^{i-1}l_{i,j}\widetilde{s}_j^{t}+\sum_{j=i+1}^{m}u_{i,j}\widetilde{s}_j^{t+1}-b_i\right). \tag{51}$$

Clearly, with $\omega=1$, the expression of $\widetilde{s}_i^{t+1}$ in (51) is the same with that in (13), so that GS iteration can be viewed as a special case of SOR iteration. Compared to GS iteration which applies $\mathbf{s}_{[-i]}=[\widetilde{s}_1^t,\ldots,\widetilde{s}_{i-1}^t,\widetilde{s}_{i+1}^{t+1},\ldots,\widetilde{s}_m^{t+1}]$ to update s_i^{t+1}, besides $\mathbf{s}_{[-i]}$, s_i^t is also taken into account for updating s_i^{t+1} in SOR iteration.

Therefore, based on (51), random sampling can be well adopted as

$$\widehat{s}_i^{t+1}\sim p_{\text{sor}}(s_i)=\frac{e^{-\frac{1}{2\overline{\sigma}_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}}{\sum_{s_i\in\mathcal{X}}e^{-\frac{1}{2\overline{\sigma}_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \tag{52}$$

with

$$\widetilde{s}_i^{t+1}=(1-\omega)\widehat{s}_i^{t}-\frac{\omega}{d_{i,i}}\left(\sum_{j=1}^{i-1}l_{i,j}\widehat{s}_j^{t}+\sum_{j=i+1}^{m}u_{i,j}\widehat{s}_j^{t+1}-b_i\right) \tag{53}$$

and $\overline{\sigma}_i=\frac{\omega\sigma}{|d_{i,i}|}$, where the rest of operations keep the same with the aforementioned IRSD algorithm over GS iteration.

For a better understanding, we denote the $\widetilde{s}_i^{t+1}$ in (53) and (24) as $\widetilde{s}_{i-\text{sor}}^{t+1}$ and $\widetilde{s}_{i-\text{gs}}^{t+1}$ respectively. Then, the relationship between them can be revealed as

$$\widetilde{s}_{i-\text{sor}}^{t+1} = \widehat{s}_{i-\text{gs}}^{t+1} + (\omega-1)(\widehat{s}_{i-\text{gs}}^{t+1} - \widehat{s}_i^t). \tag{54}$$

From (52), $\widetilde{s}_i^{t+1}$ stands for the center of the 1-dimensional discrete Gaussian distribution. According to (54), when $\widetilde{s}_{i-\text{sor}}^{t+1}$ is applied there is a center shift introduced by the difference $\widetilde{s}_{i-\text{gs}}^{t+1} - \widehat{s}_i^t$, which makes the distribution center get far away from $\widehat{s}_i^t$. Theoretically, a reasonable center shift is helpful to the Markov mixing due to a larger sampling probability $p(\widehat{s}_i^{t+1} \neq \widehat{s}_i^t)$ (i.e., a smaller sampling probability $p(\widehat{s}_i^{t+1} = \widehat{s}_i^t)$), so that the underlying Markov moves become more dynamic in state space exploration.

To be more specific, the proposed distribution in IRSD over SOR can be obtained as

$$\begin{aligned} q_{\text{sor}}(\mathbf{x}=\widehat{\mathbf{s}}^t, \mathbf{y}=\widehat{\mathbf{s}}^{t+1}) &= p_{\text{sor}}(s_m)\cdots p_{\text{sor}}(s_1) \\ &= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\widehat{\mathbf{s}}^{t+1} - ((1-1/\omega)\mathbf{D}+\mathbf{L})\Delta\mathbf{s}-\mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \end{aligned} \tag{55}$$

with $\Delta\mathbf{s} = \widehat{\mathbf{s}}^{t+1} - \widehat{\mathbf{s}}^t \in \mathbb{Z}^n$. Intuitively, with $\Delta\mathbf{s} = \mathbf{0}$, it follows that

$$\begin{aligned} q_{\text{sor}}(\mathbf{x}=\widehat{\mathbf{s}}^t, \mathbf{y}=\widehat{\mathbf{s}}^t) &= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\widehat{\mathbf{s}}^t-\mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \\ &> \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\widehat{\mathbf{s}}^t-\mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \\ &\geq \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\widehat{\mathbf{s}}^t-\mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i\|^2}}, \end{aligned} \tag{56}$$

and $q_{\text{sor}}(\mathbf{x}=\widehat{\mathbf{s}}^t, \mathbf{y}=\widehat{\mathbf{s}}^t)$ achieves a smaller lower bound than $q_{\text{gs}}(\mathbf{x}=\widehat{\mathbf{s}}^t, \mathbf{y}=\widehat{\mathbf{s}}^t)$, i.e.,

$$\frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\widehat{\mathbf{s}}^t-\mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i\|^2}} < \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\widehat{\mathbf{s}}^t-\mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|s_i\|^2}}, \tag{57}$$

where the above inequality holds due to a larger σ naturally leads to an increased $\sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i\|^2}$. Then, according to (33), it is clear to see that the transition probability $P_{\text{sor}}(\mathbf{x},\mathbf{y})$ with $\mathbf{x}=\mathbf{y}$ has a smaller lower bound than $P_{\text{gs}}(\mathbf{x},\mathbf{y})$ with $\mathbf{x}=\mathbf{y}$, so as to a larger upper bound of the transition probability $P_{\text{sor}}(\mathbf{x},\mathbf{y})$ with $\mathbf{x}\neq\mathbf{y}$ than $P_{\text{gs}}(\mathbf{x},\mathbf{y})$ with $\mathbf{x}\neq\mathbf{y}$.

Therefore, by carefully adopting the random sampling into SOR iteration in the same way, the induced Markov chain becomes more dynamic in state space exploration, which is beneficial to the Markov mixing. Moreover, this can also be interpreted by Hirschfeld-Gebelein-Rényi (HGR) maximal correlation, where a less correlation between two consecutive samples corresponds to a better convergence performance [42]. Nevertheless, we point out that such a center shift introduced by $\widetilde{s}_{i-\text{gs}}^{t+1} - \widehat{s}_i^t$ should be controlled carefully with a mild choice of ω, otherwise the convergence may be impeded even with an increased sampling probability $p(\widehat{s}_i^{t+1} \neq \widehat{s}_i^t)$. For this reason, here we apply the choice $\omega = \frac{2}{1+\sqrt{1-[\rho(\mathbf{I}-\mathbf{D}^{-1}\mathbf{A})]^2}}$ (i.e., $\rho(\mathbf{A})$ is the spectral radius of matrix $\mathbf{A}$) in [38], where further investigation about this point will be carried out as a future work.

Algorithm 2 IRSD algorithm over SOR or Jacobi iteration

Input: $\mathbf{H}, \mathbf{c}, \sigma, L, K$;
Output: $\widehat{\mathbf{s}}_{\text{output}}$;

use SOR in (51) or Jacobi in (11) to get $\widetilde{\mathbf{s}}^L$
let $\widehat{\mathbf{s}}^0 = \lceil \widetilde{\mathbf{s}}^L \rfloor_Q \in \mathcal{X}^m$
for $t = 1, \ldots, K$ **do**
  sample $\widehat{\mathbf{s}}^t$ from $q_{\text{sor}}(\mathbf{x}=\widehat{\mathbf{s}}^{t-1}, \mathbf{y}=\widehat{\mathbf{s}}^t)$ in (55) or sample $\widehat{\mathbf{s}}^t$ from $q_{\text{jacobi}}(\mathbf{x}=\widehat{\mathbf{s}}^{t-1}, \mathbf{y}=\widehat{\mathbf{s}}^t)$ in (60)
  calculate the acceptance ratio α in (15)
  generate a sample u from the uniform density $U[0,1]$
  if $u \leq \alpha$ **then**
    $\widehat{\mathbf{s}}^t$ is accepted
  else
    $\widehat{\mathbf{s}}^t$ is rejected and let $\widehat{\mathbf{s}}^t = \widehat{\mathbf{s}}^{t-1}$
  end if
end for
Output the sample with the smallest $\|\mathbf{H}\widehat{\mathbf{s}}^t - \mathbf{c}\|$ as $\widehat{\mathbf{s}}_{\text{output}}$

B. Parallel Implementation by Jacobi

The proposed IRSD algorithm can be easily extended to Jacobi iteration. In particular, based on (11), random sampling is performed as

$$\widehat{s}_i^{t+1} \sim p_{\text{jacobi}}(s_i) = \frac{e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}}{\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \tag{58}$$

with

$$\widetilde{s}_i^{t+1} = -\frac{1}{d_{i,i}}\left(\sum_{j=1}^{i-1} l_{i,j}\widehat{s}_j^t + \sum_{j=i+1}^{m} u_{i,j}\widehat{s}_j^t - b_i\right), \tag{59}$$

where the rest of operations keep the same with the aforementioned IRSD algorithm. Interestingly, it is clear to see that the update of $\widehat{s}_i^{t+1}$ only depends on $\widehat{\mathbf{s}}^t$, which allows a parallel structure in sampling decoding [48], [49].

Accordingly, the related proposed distribution in IRSD can be derived as

$$\begin{aligned} q_{\text{jacobi}}(\mathbf{x}=\widehat{\mathbf{s}}^t, \mathbf{y}=\widehat{\mathbf{s}}^{t+1}) &= p_{\text{jacobi}}(s_m)\cdots p_{\text{jacobi}}(s_1) \\ &= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\widehat{\mathbf{s}}^{t+1} - (\mathbf{U}+\mathbf{L})\Delta\mathbf{s}-\mathbf{b}\|^2}}{\prod_{i=1}^m \sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \end{aligned} \tag{60}$$

with $\Delta\mathbf{s} = \widehat{\mathbf{s}}^{t+1} - \widehat{\mathbf{s}}^t \in \mathbb{Z}^n$. Then, based on Lemma 1 and Theorem 2, we can arrive at the following results, where the proof is omitted due to the simplicity.

Corollary 3. *The proposed IRSD algorithm over Jacobi iteration has a slower convergence performance than the IRSD algorithm over GS iteration due to a smaller convergence coefficient* $\delta = \beta \cdot min[f_{jacobi}(\Delta\mathbf{s})]$ *with* $f_{jacobi}(\Delta\mathbf{s}) = e^{-\frac{1}{2\sigma^2}\|(\mathbf{U}+\mathbf{L})\Delta\mathbf{s}\|^2}$*, where*

$$min[f_{jacobi}(\Delta\mathbf{s})] < min[f_{gs}(\Delta\mathbf{s})]. \tag{61}$$

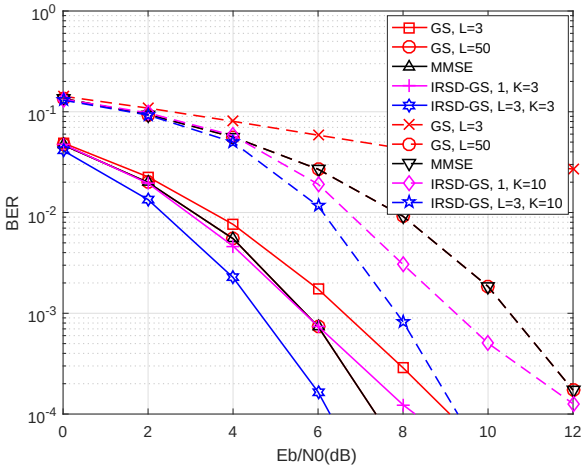

Fig. 3. Bit error rate versus average SNR per bit for large-scale MIMO system with 16-QAM, where solid and dashed curves denote 32×128 and 64×128 respectively.

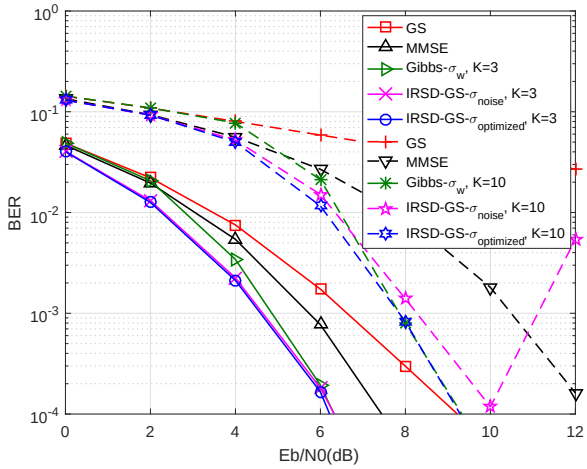

Fig. 4. Bit error rate versus average SNR per bit for large-scale MIMO system with 16-QAM, where solid and dashed curves denote 32×128 and 64×128 respectively.

Although the Jacobi-based IRSD algorithm has a slower convergence performance than the GS-based IRSD algorithm, it allows parallel implementation, which greatly facilitates its application in practice. Moreover, for a better decoding performance, a damped parameter $\omega = n/(n+m)$ is adopted into Jacobi iteration to update its iteration matrix as [22]

$$\widetilde{\mathbf{s}}^{t+1} = (1-\omega)\widetilde{\mathbf{s}}^{t} + \omega \mathbf{D}^{-1}(\mathbf{b} - \mathbf{R}\widetilde{\mathbf{s}}^{t}). \tag{62}$$

Therefore, based on damped Jacobi, we can simply update it by the proposed IRSD as

$$\widehat{s}_i^{t+1} \sim p_{\text{damped jacobi}}(s_i) = \frac{e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i - \widetilde{s}_i^{t+1}\|^2}}{\sum_{s_i \in \mathcal{X}} e^{-\frac{1}{2\widetilde{\sigma}_i^2}\|s_i - \widetilde{s}_i^{t+1}\|^2}} \tag{63}$$

with

$$\widetilde{s}_i^{t+1} = (1-\omega)\widehat{s}_i^{t} - \frac{\omega}{d_{i,i}}\left(\sum_{j=1}^{i-1} l_{i,j}\widehat{s}_j^{t} + \sum_{j=i+1}^{m} u_{i,j}\widehat{s}_j^{t} - b_i\right). \tag{64}$$

To summarize, the proposed IRSD algorithm over SOR or Jacobi iteration is outlined in Algorithm 2. We point out that IRSD algorithm can also be extended to other traditional iterative decoding schemes like Newton iteration, Richardson iteration and so on to achieve the flexible decoding performance, which is omitted here due to the simplicity. In addition, the multiple-try mechanism of MCMC may also be applied in the proposed IRSD to improve the decoding performance, which will be considered as a future work [31], [50], [51].

VI. Simulation Results

In this section, the performance of the proposed IRSD algorithm for uplink large-scale MIMO systems is investigated by simulations in full details.

In Fig. 3, the choice of the initial state $\mathbf{s}^0$ in the proposed IRSD over GS iteration is studied in both 32×128 and 64×128 large-scale MIMO systems with 16-QAM. Specifically, two choices $\widehat{\mathbf{s}}^0 = \mathbf{1}$ and $\widehat{\mathbf{s}}^0 = \lceil \widetilde{\mathbf{s}}^L \rfloor_Q$, $L = 3$ in IRSD over GS iteration are evaluated respectively in terms of the bit error rates (BERs), where the standard deviation $\sigma_{\text{optimized}}$ in (49) is applied. Besides, the BER performance of GS iteration with $L = 3$, $L = 50$ and MMSE are also shown for a better comparison. As expected, although GS iterative detection with $L = 50$ achieves a better BER performance than that with $L = 3$, its detection performance is exactly upper bounded by MMSE detector no matter how many iterations are carried out. For this reason, IRSD algorithm is proposed to bridge the detection performance between iterative decoding schemes and ML decoding. As can be seen clearly, in both 32×128 and 64×128 systems, IRSD over GS iteration are able to significantly improve the BER performance of GS iterative detection. Meanwhile, in 32×128 system under the same number of Markov moves $K = 3$, IRSD over GS iteration with $\widehat{\mathbf{s}}^0 = \lceil \widetilde{\mathbf{s}}^L \rfloor_Q$, $L = 3$ achieves a better BER performance than that with $\widehat{\mathbf{s}}^0 = \mathbf{1}$. This is accordance with the fact that a better choice of $\widehat{\mathbf{s}}^0$ is helpful to the Markov mixing. Similarly, the same observation can also be found in 64×128 system with $K = 10$, implying that the output yielded by the traditional GS iterative detector provides a good initial setup for the sampling detection in IRSD. Therefore, it is highly recommended in practice and we will apply $\widehat{\mathbf{s}}^0 = \lceil \widetilde{\mathbf{s}}^L \rfloor_Q$, $L = 3$ as a default setup for IRSD algorithm in the following simulations. Also, the GS iterative detection in the following simulations are all applied with iteration number $L = 3$ by default.

In Fig. 4, the choice of the standard deviation $\sigma > 0$ in the proposed IRSD over GS iteration is studied in both 32×128 and 64×128 large-scale MIMO systems with 16-QAM. In particular, two choices σ_{noise} in (38) and $\sigma_{\text{optimized}}$ in (49) are evaluated respectively. Clearly, in 32×128 system with the same number of Markov moves $K = 3$, nearly the same BER performance can be achieved by the proposed IRSD over GS iteration with both σ_{noise} and $\sigma_{\text{optimized}}$. However, in 64×128 system with $K = 10$, the performance difference between these two choices can be found, where IRSD with $\sigma_{\text{optimized}}$ achieves a better BER performance than that with σ_{noise}. Moreover, the BER performance of IRSD with σ_{noise} becomes fluctuated with the increase of SNR, which is due to

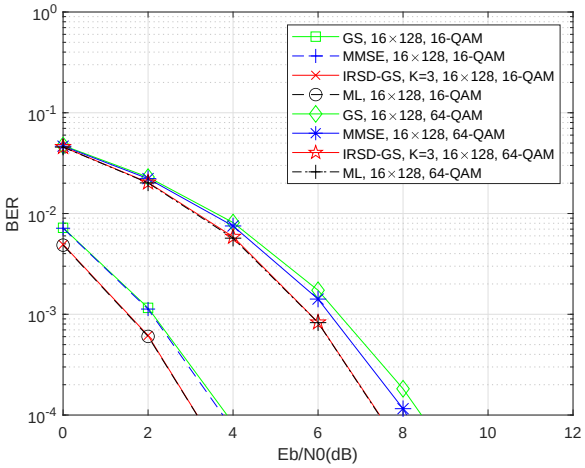

Fig. 5. Bit error rate versus average SNR per bit for 16×128 large-scale MIMO system.

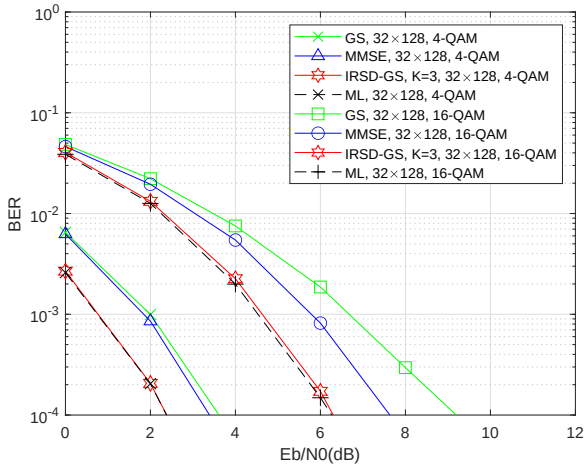

Fig. 6. Bit error rate versus average SNR per bit for 32×128 large-scale MIMO system.

the stalling effect since the value of σ_{noise} is getting smaller and smaller. Meanwhile, for a better comparison, the BER performance of Gibbs sampling is also presented with standard deviation σ_w. It is clear to see that under the same number of Markov moves the proposed IRSD over GS outperforms Gibbs sampling in both 32×128 and 64×128 systems. Besides the performance gain, compared to Gibbs sampling, the proposed IRSD algorithm has an accessible convergence rate, making it promising in the related analysis and optimization. For notational simplicity, in following simulations, we apply $\sigma_{\text{optimized}}$ as the standard deviation in IRSD-based detection schemes by default.

In Fig. 5, the performance comparison with respect to the proposed IRSD over GS iteration with Markov moves $K = 3$ is illustrated in 16×128 large-scale MIMO systems with 16-QAM and 64-QAM respectively. For a better understanding, the traditional GS iteration with $K = 3$, MMSE and ML detection (implemented by the classic sphere decoding [52]) are also shown. In particular, in both cases of 16 and 64-QAM, small performance gaps between MMSE and ML detection can be observed. Due to the favorable propagation of the channel matrix $\mathbf{H}$ when the dimensions satisfy $n \gg m$, near ML performance can be approximated by the traditional linear detection schemes like MMSE, where GS iteration provides a low complexity implementation of MMSE. Nevertheless, the proposed IRSD over GS iteration outperforms MMSE and GS iteration with a better detection performance, where ML performance can be exactly achieved by it with $K = 3$. For a better understanding, in Fig. 6, the performance comparison with respect to the proposed IRSD over GS iteration with Markov moves $K = 3$ is presented with respect to 32×128 large-scale MIMO systems with 4-QAM and 16-QAM respectively. Clearly, the condition $n \gg m$ is unfulfilled in a 32×128 large-scale MIMO system so that the performance gaps between MMSE and ML detection become substantial. Moreover, such a performance gap will get larger when the number of transmitted antennas is increased subsequently. On the other hand, compared to the traditional GS iteration who

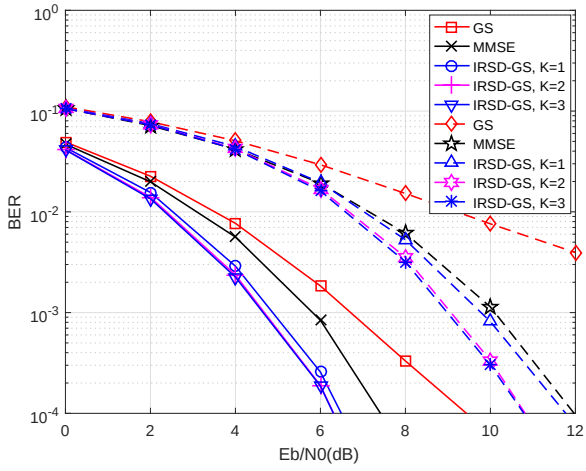

Fig. 7. Bit error rate versus average SNR per bit for 32×128 MIMO system, where solid and dashed curves denote 16-QAM and 64-QAM respectively.

is limited by the linear MMSE performance, we can also find that near ML detection performance can be achieved by the proposed IRSD over GS iteration with $K = 3$.

In Fig. 7, the performance comparison with respect to the proposed IRSD over GS iteration with different numbers of Markov moves $K = 1, 2, 3$ is illustrated in a 32×128 uncoded large-scale MIMO system with 16-QAM and 64-QAM respectively. For a better comparison, the BER performance of GS iterative detection with $L = 3$ and MMSE detection are also shown. Clearly, by sampling detection, the BER performance of IRSD over GS iteration significantly outperforms GS iterative detection and MMSE. As expected, with the increase of the number of Markov moves K, the BER performance of IRSD over GS iteration improves gradually in both cases of 16-QAM and 64-QAM. This is in line with the convergence result given in Corollary 1. More precisely, as shown in Corollary 2, a better sampling approximation can be achieved with the increment of K, which leads to a better detection result accordingly. On the other hand, we point out that the performance gain of IRSD becomes marginal with

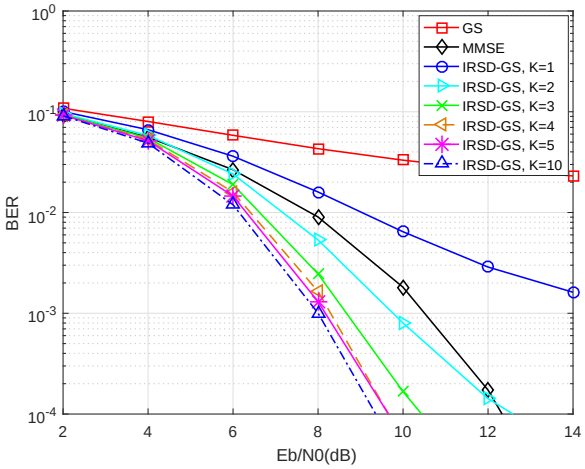

Fig. 8. Bit error rate versus average SNR per bit for 64×128 large-scale MIMO system with 16-QAM.

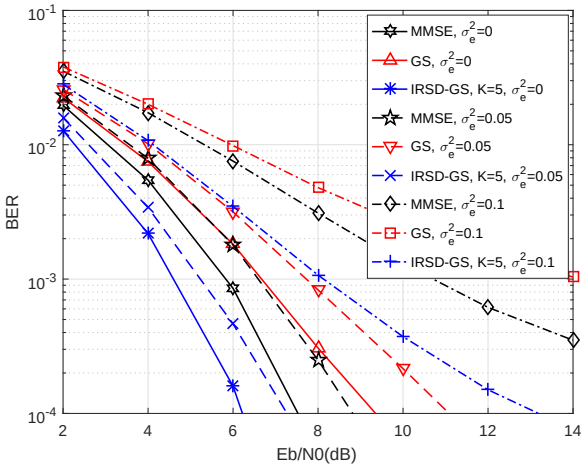

Fig. 9. Bit error rate versus average SNR per bit for 32×128 large-scale MIMO system with imperfect CSI using 16-QAM.

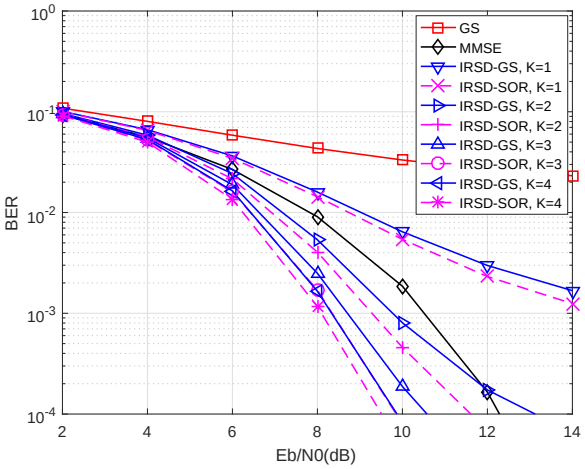

Fig. 10. Bit error rate versus average SNR per bit for 64×128 large-scale MIMO system with 16-QAM.

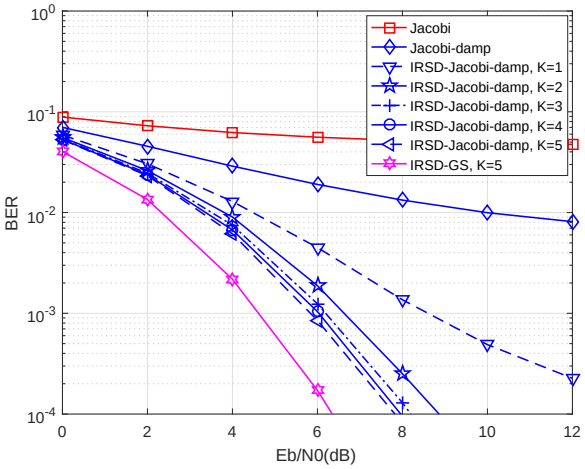

Fig. 11. Bit error rate versus average SNR per bit for 32×128 large-scale MIMO system with 16-QAM.

the increase of K. For this reason, a mild choice of K is recommended for the sake of complexity.

In Fig. 8 another performance comparison about IRSD over GS iteration with different numbers of Markov moves K is presented regarding to a 32×128 uncoded large-scale MIMO system with 16-QAM. Similarly, the BER performance of IRSD over GS iteration improves with the increment of K. Theoretically, compared to 32×128 system, less receive diversity gain can be exploited by 64×128 system due to the larger number of transmitted antennas, making the related uplink signal detection more difficult. To this end, more number of Markov moves is needed to achieve a better BER performance than the traditional MMSE detection. Nevertheless, considerable detection gain still can be obtained with a moderate size of K.

Fig. 9 is given to evaluate the detection performance of the proposed IRSD algorithm over GS iteration under different cases of channel state information (CSI), where a 16×128 uncoded large-scale MIMO system using 64-QAM is applied. Specifically, $\widehat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}$ stands for the imperfect CSI at the receiver side, where $\Delta\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2\mathbf{I})$ denotes the channel estimation errors with $\sigma_e^2 = \frac{m}{n_p \cdot E_p}$ [53]. Here, n_p and E_p indicate the number and the power of pilot symbols respectively, and we set $\sigma_e^2 = 0, 0.05, 0.1$ respectively for the comparison. Undoubtedly, $\sigma_e^2 = 0$ corresponds to a perfect CSI as before. Compared to the results of perfect CSI, the BER performance of all the detection schemes under imperfect CSI degrade gradually with the increase of σ_e^2. Nevertheless, it is clear to see that the proposed IRSD algorithm still works under imperfect CSI while significant performance gain can be achieved with $K = 5$.

As we claimed, the proposed IRSD algorithm is well suited to various traditional iteration methods for a better decoding performance. In Fig. 10, the performance comparison with respect to the proposed IRSD over SOR iteration is presented in a 64×128 uncoded large-scale MIMO system with 16-QAM. Besides GS iterative detection and MMSE detection, IRSD over GS iteration is also added for a better comparison. As can be seen clearly, with the increment of the number of Markov moves K, both the BER performance of IRSD over

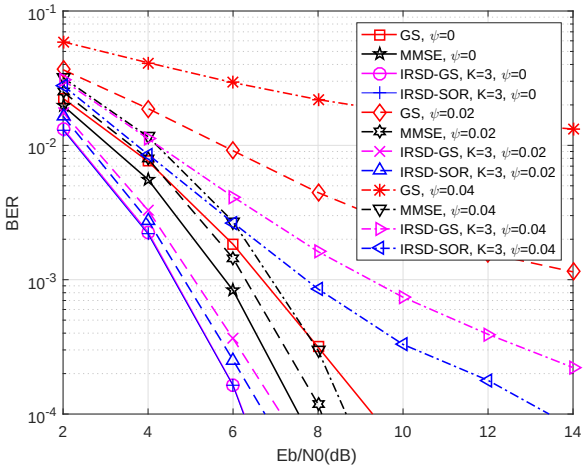

Fig. 12. Bit error rate versus average SNR per bit for 32×128 large-scale MIMO system with 16-QAM and normalized correlation index ψ.

GS and SOR iterations improve gradually. As expected, under the same K, the BER performance of IRSD over SOR achieves a better BER performance than that of IRSD over GS. This is accordance with our analysis in Section V since IRSD over SOR iteration can be viewed as an enhancement version of IRSD over GS iteration.

In Fig. 11, the performance comparison with respect to the proposed IRSD over damped Jacobi iteration is shown in a 32×128 uncoded large-scale MIMO system with 16-QAM. As shown in (58), the traditional Jacobi iteration can be easily adopted to the IRSD over Jacobi iteration. Since damped Jacobi offers a faster iteration convergence than the traditional Jacobi, it is better to adopt the damped Jacobi rather than the traditional Jacobi into IRSD. Specifically, the BER performance of IRSD over damped Jacobi iteration improves gradually with the increment of K. It is clear that the detection performance of IRSD over damped Jacobi iteration is not as good as that of IRSD over GS iteration under the same number of Markov moves. Nevertheless, as we pointed out, IRSD over damped Jacobi iteration allows a parallel decoding structure, which is beneficial to the implementation in practice.

In Fig. 12, the performance comparison with respect to the proposed IRSD algorithm under correlated channels of large-scale MIMO systems is investigated. Specifically, following the setups of correlation channels in [54], the correlated channel matrix is set by $\mathbf{R}_{\text{cor}}^{\frac{1}{2}}\mathbf{H}\mathbf{T}_{\text{cor}}^{\frac{1}{2}}$, where $\mathbf{R}_{\text{cor}} \in \mathbb{C}^{n\times n}$ and $\mathbf{T}_{\text{cor}} \in \mathbb{C}^{m\times m}$ denotes the receive correlation matrix and the transmit correlation matrix respectively. Note that the normalized correlation coefficient $1 \geq \psi \geq 0$ is employed to adjust the correlation degree within them. More precisely, a totally uncorrelated scenario corresponds to $\psi = 0$ while a fully correlated scenario implies $\psi = 1$. As can be seen clearly, with $\psi = 0.02$, the detection performance of GS iteration, MMSE, IRSD over GS iteration and IRSD over SOR iteration slightly degrade compared to the i.i.d. case with $\psi = 0$. Accordingly, with the increase of ψ, these detection performance are getting worse gradually. However, the decoding gain brought by the proposed IRSD still can be clearly confirmed. Meanwhile, as expected, under the same K, IRSD over SOR iteration achieves a better BER performance than IRSD over GS iteration.

VII. Conclusion

In this paper, based on MCMC methods, a general sampling decoding framework over the traditional iteration methods is proposed for the uplink detection of large-scale MIMO systems. Compared to these iterative decoding schemes, the proposed IRSD algorithm is able to achieve a flexible performance between suboptimal and optimal decoding, which is adjusted by the number of Markov moves. Compared to Gibbs sampling, the proposed IRSD algorithm entails an accessible convergence rate with tractable mixing time. Based on it, further optimization, enhancement and extension with respect to IRSD algorithm can be carried out for better decoding performance and efficiency, making the proposed IRSD algorithm well suited to various cases of large-scale MIMO systems.

Appendix A
Proof of Lemma 1

Proof: Using (30) and (21), we have

$$\begin{aligned}\frac{q_{\text{gs}}(\mathbf{x},\mathbf{y})}{\Pi(\mathbf{y})} &= \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{s}-\mathbf{L}\Delta\mathbf{s}-\mathbf{b}\|^2}}{\prod_{i=1}^{m}\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\sigma_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \cdot \frac{\sum_{\mathbf{s}\in\mathcal{X}^m} e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{s}-\mathbf{b}\|^2}}{e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{s}-\mathbf{b}\|^2}} \\ &\overset{(a)}{\geq} \frac{\sum_{\mathbf{s}\in\mathcal{X}^m} e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{s}-\mathbf{b}\|^2}}{\prod_{i=1}^{m}\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\sigma_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \cdot e^{-\frac{1}{2\sigma^2}\|\mathbf{L}\Delta\mathbf{s}\|^2} \\ &\overset{(b)}{\geq} \frac{\sum_{\mathbf{s}\in\mathcal{X}^m} e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{s}-\mathbf{b}\|^2}}{\prod_{i=1}^{m}\sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|s_i-\widetilde{s}_i^{t+1}\|^2}} \cdot e^{-\frac{1}{2\sigma^2}\|\mathbf{L}\Delta\mathbf{s}\|^2} \\ &\overset{(c)}{\geq} \frac{\sum_{\mathbf{s}\in\mathcal{X}^m} e^{-\frac{1}{2\sigma^2}\|\mathbf{A}\mathbf{s}-\mathbf{b}\|^2}}{\prod_{i=1}^{m}\sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma_i^2}\|s_i\|^2}} \cdot e^{-\frac{1}{2\sigma^2}\|\mathbf{L}\Delta\mathbf{s}\|^2} \\ &= \beta \cdot e^{-\frac{1}{2\sigma^2}\|\mathbf{L}\Delta\mathbf{s}\|^2}, \end{aligned} \tag{65}$$

where (a) follows the triangle inequality

$$\|\mathbf{A}\mathbf{s}-\mathbf{L}\Delta\mathbf{s}-\mathbf{b}\| \leq \|\mathbf{A}\mathbf{s}-\mathbf{b}\| + \|\mathbf{L}\Delta\mathbf{s}\|, \tag{66}$$

(b) and (c) hold due to the fact that [55]

$$\begin{aligned}\sum_{s_i\in\mathcal{X}} e^{-\frac{1}{2\sigma^2}\|s_i-\widetilde{s}_i^{t+1}\|^2} &\leq \sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}\|s_i-\widetilde{s}_i^{t+1}\|^2} \\ &\leq \sum_{s_i\in\mathbb{Z}} e^{-\frac{1}{2\sigma^2}\|s_i\|^2}. \end{aligned} \tag{67}$$

■

References

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

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

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

[4] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 1941–1988, Sep. 2015.

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

[6] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge University Press, 2005.

[7] F. Jin, Q. Liu, H. Liu, and P. Wu, "A low complexity signal detection scheme based on improved Newton iteration for massive MIMO systems," *IEEE Communications Letters*, vol. 23, no. 4, pp. 748–751, 2019.

[8] C. Tang, C. Liu, L. Yuan, and Z. Xing, "High precision low complexity matrix inversion based on Newton iteration for data detection in the massive MIMO," *IEEE Communications Letters*, vol. 20, no. 3, pp. 490–493, 2016.

[9] X. Qin, Z. Yan, and G. He, "A near-optimal detection scheme based on joint steepest descent and Jacobi method for uplink massive MIMO systems," *IEEE Communications Letters*, vol. 20, no. 2, pp. 276–279, 2016.

[10] J. Tu, M. Lou, J. Jiang, D. Shu, and G. He, "An efficient massive MIMO detector based on second-order Richardson iteration: From algorithm to flexible architecture," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 11, pp. 4015–4028, 2020.

[11] L. Dai, X. Gao, X. Su, S. Han, C.-L. I, and Z. Wang, "Low-complexity soft-output signal detection based on Gauss-Seidel method for uplink multiuser large-scale MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4839–4845, 2015.

[12] D. Zhu, B. Li, and P. Liang, "On the matrix inversion approximation based on Neumann series in massive MIMO systems," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 1763–1769.

[13] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[14] M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-scale MIMO detection for 3GPP LTE: Algorithms and FPGA implementations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 916–929, 2014.

[15] Y. Zhang, A. Yu, X. Tan, Z. Zhang, X. You, and C. Zhang, "Adaptive damped Jacobi detector and architecture for massive MIMO uplink," in *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2018, pp. 203–206.

[16] X. Gao, L. Dai, C. Yuen, and Y. Zhang, "Low-complexity MMSE signal detection based on Richardson method for large-scale MIMO systems," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, 2014, pp. 1–5.

[17] A. Björck, Numerical Methods for Least Squares Problems. Philadelphia, PA, USA: SIAM, 1996.

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

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

[20] 3GPP, "User equipment (UE) conformance specification," in *TS 38.521-4 v.15.0.0, Tech. Spec.*, May 2019, pp. 1–5.

[21] ——, "Study on scenarios and requirements for next generation access technologies," in *TR 38.913 v.14.3.0, Tech. Rep.*, June 2017, pp. 1–41.

[22] J. Minango, C. de Almeida, and C. Daniel Altamirano, "Low-complexity MMSE detector for massive MIMO systems based on damped Jacobi method," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–5.

[23] P. Zhang, L. Liu, G. Peng, and S. Wei, "Large-scale MIMO detection design and FPGA implementations using SOR method," in *2016 8th IEEE International Conference on Communication Software and Networks (ICCSN)*, 2016, pp. 206–210.

[24] S. Liu, C. Ling, and D. Stehlé, "Decoding by sampling: A randomized lattice algorithm for bounded distance decoding," *IEEE Trans. Inform. Theory*, vol. 57, pp. 5933–5945, Sep. 2011.

[25] P. Klein, "Finding the closest lattice vector when it is unusually close," in *ACM-SIAM Symp. Discr. Algorithms*, 2000, pp. 937–941.

[26] Z. Wang and C. Ling, "Lattice Gaussian sampling by Markov chain Monte Carlo: Bounded distance decoding and trapdoor sampling," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3630–3645, June 2019.

[27] J. Choi, "An MCMC-MIMO detector as a stochastic linear system solver using successive overrelexation," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1445–1455, 2016.

[28] B. Hassibi, M. Hansen, A. Dimakis, H. Alshamary, and W. Xu, "Optimized Markov Chain Monte Carlo for signal detection in MIMO systems: An analysis of the stationary distribution and mixing time," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4436–4450, Sep. 2014.

[29] T. Datta, N. Kumar, A. Chockalingam, and B. Rajan, "A novel Monte Carlo sampling based receiver for large-scale uplink multiuser MIMO systems," *IEEE Transactions on Vehicular Technology,*, vol. 62, no. 7, pp. 3019–3038, Sep. 2013.

[30] Z. Wang, "Markov chain Monte Carlo methods for lattice Gaussian sampling: Convergence analysis and enhancement," *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 6711–6724, 2019.

[31] Z. Wang, Y. Huang, and S. Lyu, "Lattice-reduction-aided Gibbs algorithm for lattice Gaussian sampling: Convergence enhancement and decoding optimization," *IEEE Transactions on Signal Processing*, vol. 67, no. 16, pp. 4342–4356, 2019.

[32] Z. Wang, C. Ling, and G. Hanrot, "Markov chain Monte Carlo algorithms for lattice Gaussian sampling," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Honolulu, USA, Jun. 2014, pp. 1489–1493.

[33] Z. Wang and C. Ling, "On the geometric ergodicity of Metropolis-Hastings algorithms for lattice Gaussian sampling," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 738–751, Feb. 2018.

[34] Z. Wang, S. Lyu, Y. Xia, and Q. Wu, "Expectation propagation-based sampling decoding: Enhancement and optimization," *IEEE Transactions on Signal Processing*, vol. 69, pp. 195–209, 2021.

[35] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Time*, American Mathematical Society, 2008.

[36] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, 1986.

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

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

[39] Y. Saad, Iterative Methods for Sparse Linear Systems. Philadelphia, PA, USA: SIAM, 1997.

[40] A. Greenbaum, Iterative Methods for Solving Linear Systems. Philadelphia, PA, USA: SIAM, 1997.

[41] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[42] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, New York: Springer-Verlag, 2001.

[43] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*. UK, Cambridge University Press, 2009.

[44] R. Chen, J. Liu, and X. Wang, "Convergence analysis and comparisons of Markov chain Monte Carlo algorithms in digital communications," *IEEE Trans. on Signal Process.*, vol. 50, no. 2, pp. 255–270, 2002.

[45] J. H. Conway and N. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1998.

[46] J. Yi, "Theta-function identities and the explicit formulas for Theta-function and their applications," in *Proc J. Math. Anal. Appl.*, vol. 292, 2004, pp. 381–400.

[47] B. Berndt, *Ramanujan's Notebooks, Part V*. Springer-Verlag, New York, 1998.

[48] A. Margaris, S. Souravlas, and M. Roumeliotis, "Parallel implementations of the Jacobi linear algebraic systems solve," *ArXiv*, vol. abs/1403.5805, 2014.

[49] A. Ahmadi, F. Manganiello, A. Khademi, and M. C. Smith, "A parallel Jacobi-embedded Gauss-Seidel method," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 6, pp. 1452–1464, 2021.

[50] L. Martino, "A review of multiple try MCMC algorithms for signal processing," *Digital Signal Processing*, vol. 75, pp. 134–152, 2018.

[51] B. Calderhead, "A general construction for parallelizing Metropolis-Hastings algorithms," *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, vol. 111, no. 49, pp. 17 408–17 413, 2014.

[52] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2389–2401, Oct. 2003.

[53] Q. Chen, S. Zhang, S. Xu, and S. Cao, "Efficient MIMO detection with imperfect channel knowledge - A deep learning approach," in *2019 IEEE*

Wireless Communications and Networking Conference (WCNC), 2019, pp. 1–6.

[54] 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.

[55] D. Micciancio and O. Regev, "Worst-case to average-case reductions based on Gaussian measures," in *Proc. Ann. Symp. Found. Computer Science*, Rome, Italy, Oct. 2004, pp. 372–381.

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

Since 2021, he has been an Associate Professor in the School of Information and Engineering, Southeast University (SEU), Nanjing, China. From 2015 to 2016 he served as a Research Associate at Imperial College London, UK. From 2016 to 2017 he was an senior engineer with Radio Access Network R&D division, Huawei Technologies Co.. From 2017 to 2020 he was an Associate Professor at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. His current research interests include massive MIMO systems, machine learning and data analytics over wireless networks, and lattice theory for wireless communications.

Shengheng Liu (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in electronics engineering from the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, in 2010 and 2017, respectively. Dr. Liu is currently an Associate Professor with the School of Information Science and Engineering, Southeast University (SEU), Nanjing, China. He has also been with the Purple Mountain Laboratories, Nanjing, since 2020. Prior to joining SEU, he held a postdoctoral position at the Institute for Digital Communications, The University of Edinburgh, Edinburgh, U.K., from 2017 to 2018. He also worked as a Visiting Research Associate from 2015 to 2016 at the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA, USA, under the support of the China Scholarship Council. Dr. Liu is a recipient of the 2017 National Excellent Doctoral Dissertation Award from the China Institute of Communications. His research interests mainly focus on intelligent sensing and wireless communications.

Yili Xia (Member, IEEE) received the B.Eng. degree in Information Engineering from Southeast University, Nanjing, China, in 2006, and the M.Sc. (Distinction) degree in Communications & Signal Processing from Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., in 2007. He obtained his Ph.D. degree in adaptive signal processing from Imperial College London in 2011.

Since 2013, he has been an Associate Professor in Signal Processing with the School of Information Science and Engineering, Southeast University, Nanjing, China, where he is currently the Deputy Head of the Department of Information and Signal Processing Engineering. His research interests include complex and hypercomplex statistical analysis, detection and estimation, linear and nonlinear adaptive filters, as well as their applications on communications and power systems. Dr. Xia was a recipient of the Best Student Paper Award at the International Symposium on Neural Networks (ISNN) in 2010 (coauthor), and the Education Innovation Award at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in 2019. He is currently an Associate Editor for the IEEE Transactions on Signal Processing.

Cong Ling received the bachelors and masters degrees from the Nanjing Institute of Communications Engineering, China, in 1995 and 1997 respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2005. He is currently a Reader (equivalent to Professor/Associate Professor) with the Electrical and Electronic Engineering Department, Imperial College London. He is also a member of the Academic Centre of Excellence in Cyber Security Research, Imperial College, and an affiliated member of the Institute of Security Science and Technology, Imperial College. Before joining Imperial College, he had been on the faculties of the Nanjing Institute of Communications Engineering and Kings College. He visited The Hong Kong University of Science and Technology as a Hong Kong Telecom Institute of Information Technology (HKTIIT) fellow in 2009.

He has been an Associate Editor (in multiterminal communications and lattice coding) of IEEE Transactions on Communications, and an Associate Editor of IEEE Transactions on Vehicular Technology and on the program committees of several international conferences, including IEEE Information Theory Workshop, Globecom, and ICC.

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

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