Lattice-Reduction-Aided Gibbs Algorithm for Lattice Gaussian Sampling: Convergence Enhancement and Decoding Optimization

Zheng Wang, *Member, IEEE*, Yang Huang, *Member, IEEE*, and Shanxiang Lyu

***Abstract*—Sampling from the lattice Gaussian distribution has emerged as an important problem in coding, decoding, and cryptography. In this paper, lattice reduction technique is adopted to Gibbs sampler for lattice Gaussian sampling. First, with respect to lattice Gaussian distribution, we show the convergence rate of systematic scan Gibbs sampling is characterized by the Hirschfeld-Gebelein-Rényi maximal correlation among the multivariate of being sampled. Then, Lattice-reduction-aided Gibbs algorithm is proposed to sample from an equivalent lattice Gaussian distribution but with less correlated multivariate, thus leading to a better Markov mixing. After that, we extend the proposed lattice-reduction-aided Gibbs sampling to lattice decoding, where the choice of the standard deviation for the sampling is fully investigated. A customized solution that suits for each specific decoding case by Euclidean distance is given, which results in a better tradeoff between Markov mixing and sampler decoding. Based on it, a startup mechanism is also proposed for Gibbs sampler decoding, where decoding complexity can be reduced without performance loss. Moreover, the recycling Gibbs sampling that exploits the potential of samples is also considered to improve the decoding performance in lattice decoding. Simulation results based on large-scale uncoded multiple-input multiple-output detection are presented to confirm the performance gain and complexity reduction.**

***Index Terms*—Lattice Gaussian sampling, Markov chain Monte Carlo, Gibbs sampling, lattice decoding, large-scale MIMO detection.**

I. Introduction

NOWADAYS, lattice Gaussian sampling has drawn a lot of attention in various research fields. In mathematics, Banaszczyk was the first to apply it to prove the transference theorems for lattices [1]. In coding, lattice Gaussian distribution was employed to obtain the full shaping gain for lattice coding [2]–[4], and to achieve the capacity of the Gaussian channel [5]. It was also used to achieve information-theoretic security in the Gaussian wiretap channel [6]–[8] and in the bidirectional relay channel [9], respectively. In cryptography, the lattice Gaussian distribution has become a central tool in the construction of many primitives [10]–[12]. Specifically, lattice Gaussian sampling lies at the core of signature schemes in the Gentry, Peikert and Vaikuntanathan (GPV) paradigm [13].

In decoding, lattice Gaussian sampling with a suitable variance allows to solve the closest vector problem (CVP) and the shortest vector problem (SVP) [14], [15]. In fact, the classic detection problem in multiple-input multiple-output (MIMO) systems can be viewed as a CVP problem, and the dramatically increased system size of MIMO in 5G has placed a quite pressing challenge on it [16].

However, in sharp contrast to the continuous Gaussian density, it is by no means trivial even to sample from a low-dimensional discrete Gaussian distribution. Efficient sampling schemes do exist but they only work for a few special lattices [5], [17]. Meanwhile, as the default sampling algorithm for general lattices, Klein's algorithm only works when the standard deviation is sufficiently large [13]. Therefore, in order to sample from a target lattice Gaussian distribution, Markov chain Monte Carlo (MCMC) methods were introduced [18], [19]. In principle, it randomly generates the next Markov state conditioned on the previous one; after the burn-in time, the Markov chain will step into a stationary distribution, where samples from the target distribution can be obtained thereafter [20]. As a basic MCMC method, the Gibbs algorithm, which employs univariate conditional sampling to build the Markov chain, has been introduced to lattice Gaussian sampling by showing its ergodicity [21]. In [22], the symmetric Metropolis-within-Gibbs (SMWG) algorithm was proposed for lattice Gaussian sampling to achieve the exponential convergence. Moreover, the Markov chain induced by random scan Gibbs sampling was shown to be geometric ergodicity [23], which means it converges exponentially fast. Besides Gibbs algorithm, other MCMC methods for lattice Gaussian sampling also exist, and the independent Metropolis-Hastings-Klein (IMHK) algorithm is not only uniformly ergodic but also enjoys an accessible convergence rate [18], [19].

On the other hand, thanks to the convergence theorem of MCMC, Gibbs sampling with a finite state space naturally

Manuscript received April 2, 2019; revised July 10, 2019; accepted July 12, 2019. Date of publication July 22, 2019; date of current version July 31, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. David Ramirez. This work was supported in part by the open research fund of Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space, Nanjing University of Aeronautics and Astronautics, Ministry of Industry and Information Technology, Nanjing, China, (KF20181913), in part by the open research fund of National Mobile Communications Research Laboratory, Southeast University, (2019D04), in part by the National Natural Science Foundation of China under Grant 61801216, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180420. *(Corresponding author: Zheng Wang.)*

Z. Wang and Y. Huang are with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: wznuaa@gmail.com; yang.huang.ceie@nuaa.edu.cn).

S. Lyu is with the College of Information Science and Technology and the College of Cyber Security, Jinan University, Guangzhou 510632, China (e-mail: shanxianglyu@gmail.com).

Digital Object Identifier 10.1109/TSP.2019.2929922

experiences the geometric ergodicity so that the Gibbs-based discrete Gaussian sampling decoder has already been adapted to MIMO detection to solve the CVP [24]–[28]. Moreover, the Gibbs sampling has also been introduced into soft-output decoding in MIMO systems, where the extrinsic information calculated by a priori probability (APP) detector is used to produce soft outputs [29], [30]. In [31], an investigation of Gibbs-based MCMC receivers in different communication channels is given as well. However, given those works, the choice of the standard deviation σ (also referred to as "temperature") for Gibbs sampling decoding has not been fully investigated. A common choice comes from statistics by letting σ^2 be the variance of noises, which severely suffers from the *stalling problem* in high signal-to-noise ratio (SNR) regime. Although Hassibi *et al.* suggested σ should instead be scaling at least as $\Omega(\sqrt{\mathrm{SNR}})$, it fails to exploit the decoding potential for each specific case [24]. Meanwhile, another very important point was ignored for years. Specifically, as an advanced decoder, Gibbs sampler decoding, however, is not necessary for all the decoding cases, where the optimal solution may be directly obtained by suboptimal decoding schemes especially in high SNRs. This indicates substantial computational complexity can be saved without any performance loss. In [25], [32], two stopping criterions were given for mixed-Gibbs sampler decoding schemes, but they only work for the proposed multiple restart strategies by simply terminating those trapped Markov chains.

In this paper, we advance the state of the art of the Gibbs-based lattice Gaussian sampling and lattice decoding in several fronts. First of all, in order to enhance the convergence performance of Gibbs sampling, the lattice-reduction-aided Gibbs sampling algorithm is proposed for lattice Gaussian sampling. In particular, a comprehensive analysis regarding to the convergence rate of the Markov chain induced by systematic scan Gibbs sampling is presented, and we show the convergence is essentially dominated by the Hirschfeld-Gebelein-Rényi (HGR) maximal correlation between the multiple random variables. Hence, by lattice reduction, an equivalent lattice Gaussian distribution can be established with significantly reduced HGR maximal correlation, thus leading to a boosting convergence performance.

Secondly, given the convergence gain, we then extend the lattice-reduction-aided Gibbs sampling algorithm to lattice decoding, where considerable decoding performance can be achieved compared to the conventional Gibbs sampling. After that, the investigation about optimizing the sampling probability of the target decoding point is carried out, which leads to a better trade-off between Markov mixing and sampling decoding. Specifically, we show that the choice of the standard deviation σ heavily depends on the Euclidean distance from the query point to lattice. This not only effectively avoids the stalling problem, but also provides a preferable choice of σ for each specific decoding case. Hence, for a better approximation of σ, the initial starting point of the Markov chain is strongly desired to be well chosen, which is in accordance with geometric ergodicity as the initial starting point has an indispensable impact on the convergence.

Thirdly, based on the well chosen initial starting point, we adopt the correct decoding radius from bounded distance decoding (BDD) to build a startup mechanism, which decides whether to invoke Gibbs sampler or not. Meanwhile, the demand of the high quality initial starting point can also be guaranteed through the usage of lattice reduction. In a word, our proposed Gibbs sampler decoding advances with better decoding performance and less complexity cost. Additionally, the mechanism of recycling Gibbs sampling is also applied to further improve the decoding performance, which not only works well in lattice decoding but also in hard-output detection of MIMO systems.

It should be noticed that compared to the lattice Gaussian distribution, the discrete Gaussian distribution designed for MIMO detection entails a finite state space (i.e., $\mathbf{x} \in \mathcal{X}^n$ based on the QAM constellation). After the nonlinear transformation $\mathbf{z} = \mathbf{U}^{-1}\mathbf{x}$ of lattice reduction ($\mathbf{U} \in \mathbb{Z}^{n\times n}$ is a unimodular matrix with $\det(\mathbf{U}) = \pm 1$), the state space of $\mathbf{z}$ turns out to be computationally expensive to get. For non-Gibbs sampling based detectors [33], suboptimal remedies can be carried out to restrict $\widehat{\mathbf{x}} = \mathbf{U}\mathbf{z}$ to the original set $\mathcal{X}^n$ in the end. However, for Gibbs sampler decoding, the Markov chain along with an unbounded or approximate state space of $\mathbf{z}$ tends to be unreasonably wild, which most likely results in an invalid Markov mixing. Such a problem does not exist in lattice decoding paradigm since $\mathbf{x}$ and $\mathbf{z}$ share the same state space $\mathbb{Z}^n$. To this end, lattice reduction is not recommended to be directly applied in the Markov mixing of Gibbs sampling for MIMO detection. Nevertheless, the aforementioned analysis results from lattice decoding are still applicable to MIMO detection, by simply removing lattice reduction from the Markov mixing. Additionally, besides MIMO detection, the sampler decoding strategy can also be extended to signal processing as a useful signal estimator or detector [34]–[38].

The rest of this paper is organized as follows. Section II introduces the background of lattice Gaussian distribution and briefly reviews the basics of Gibbs sampling as well as lattice reduction. In Section III, the convergence rate of systematic scan Gibbs sampling is demonstrated to be determined by the HGR maximal correlation among the multivariate. Based on it, the lattice-reduction-aided Gibbs sampling algorithm is proposed for a better Markov mixing performance. Section IV extends the lattice-reduction-aided Gibbs sampling to lattice decoding. Simulation results for large-scale uncoded MIMO detection are presented in Section V. Finally, Section VI 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 $\mathbf{B}$, $\widehat{\mathbf{b}}_i$ for the ith Gram-Schmidt vector of the matrix $\mathbf{B}$, $b_{i,j}$ for the entry in the ith row and jth column of the matrix $\mathbf{B}$. In addition, in this paper, the computational complexity is measured by the number of arithmetic operations (additions, multiplications, comparisons, etc.). Finally, $h \in L_0^2(\pi)$ and $L_0^2(\pi)$ denote the set of all mean zero and finite variance functions with respect to the target distribution π, i.e., $E_\pi[h(\mathbf{x})] = 0$ and $\mathrm{var}_\pi[h(\mathbf{x})] = v < \infty$.

II. Preliminaries

In this section, we introduce the background and mathematical tools needed to describe and analyze the following lattice-reduction-aided Gibbs sampling.

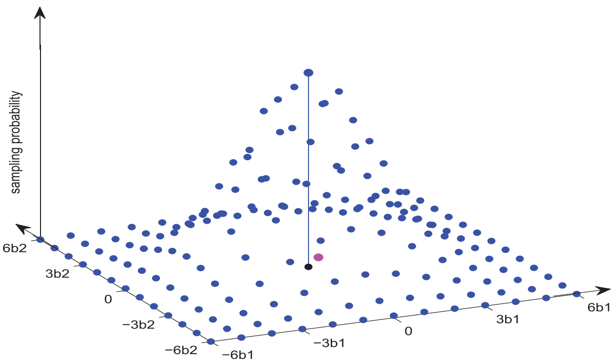

Fig. 1. Illustration of a two-dimensional lattice Gaussian distribution with $\mathbf{B} = [\mathbf{b}_1\mathbf{b}_2]$, where the red dot and blue dots respectively correspond to the query point $\mathbf{c}$ and sampling probabilities of candidate lattice points $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$. For simplicity, only the closest lattice point $\mathbf{Bx}_{\text{cvp}}$ (black dot) with the largest sampling probability is depicted.

A. Lattice Gaussian Distribution

Let $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n] \subset \mathbb{R}^n$ consist of n linearly independent vectors. The n-dimensional lattice Λ generated by $\mathbf{B}$ is defined by

$$\Lambda = \mathcal{L}(\mathbf{B}) = \{\mathbf{Bx} : \mathbf{x} \in \mathbb{Z}^n\}, \tag{1}$$

where $\mathbf{B}$ is called the lattice basis. We define the Gaussian function centered at $\mathbf{c} \in \mathbb{R}^n$ for standard deviation $\sigma > 0$ as

$$\rho_{\sigma,\mathbf{c}}(\mathbf{z}) = e^{-\frac{\|\mathbf{z}-\mathbf{c}\|^2}{2\sigma^2}}, \tag{2}$$

for all $\mathbf{z} \in \mathbb{R}^n$. When $\mathbf{c}$ or σ are not specified, we assume that they are $\mathbf{0}$ and 1 respectively. Then, the *discrete Gaussian distribution* over Λ is defined as

$$D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x}) = \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{Bx})}{\rho_{\sigma,\mathbf{c}}(\Lambda)} = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}{\sum_{\mathbf{x}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}} \tag{3}$$

for all $\mathbf{x} \in \mathbb{Z}^n$, where $\rho_{\sigma,\mathbf{c}}(\Lambda) \triangleq \sum_{\mathbf{Bx}\in\Lambda} \rho_{\sigma,\mathbf{c}}(\mathbf{Bx})$ is just a scaling to obtain a probability distribution. We remark that this definition differs slightly from the one in [10], where σ is scaled by a constant factor $\sqrt{2\pi}$ (i.e., $s = \sqrt{2\pi}\sigma$).

Fig. 1 illustrates the discrete Gaussian distribution over $\mathbb{Z}^2$. As can be seen clearly, it resembles a continuous Gaussian distribution, but is defined over a $\mathbb{Z}^2$ lattice. It has been demonstrated in [7] that only if the *flatness factor* is small enough,[1] discrete and continuous Gaussian distributions could share similar properties. Otherwise, these two distributions behave quite different in the sense of entropy rate while the difference in terms of accuracy between them is straightforward even in 1-dimensional distribution. Therefore, sampling from these two should be treated respectively in most of cases of interest especially when multi-dimensional distribution and random lattice (i.e., $\mathbf{B} \in \mathbb{R}^n$) are considered.

B. Sampler Decoding

Consider the decoding of an $n \times n$ real-valued system. The extension to the complex-valued system is straightforward [39], [40]. Let $\mathbf{x} \in \mathbb{Z}^n$ denote the transmitted signal. The corresponding received signal $\mathbf{c}$ is given by

$$\mathbf{c} = \mathbf{Bx} + \mathbf{w} \tag{4}$$

where $\mathbf{w}$ is the noise vector with zero mean and variance σ_w^2, $\mathbf{B}$ is an $n \times n$ full column-rank matrix of channel coefficients. Typically, the conventional maximum likelihood (ML) reads

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{X}^n} \|\mathbf{Bx} - \mathbf{c}\|^2 \tag{5}$$

where $\|\cdot\|$ denotes the Euclidean norm. Clearly, the ML decoding in above uncoded MIMO systems corresponds to the CVP [41]. If the received signal $\mathbf{c}$ is the origin, then ML decoding reduces to SVP.

Intuitively, the CVP given in (5) can be solved by lattice Gaussian sampling. Since the distribution is centered at the query point $\mathbf{c}$, the closest lattice point $\mathbf{Bx}_{\text{cvp}}$ to $\mathbf{c}$ is assigned the largest sampling probability. Therefore, by multiple samplings, $\mathbf{x}_{\text{cvp}}$ is most likely to be returned. It has been demonstrated that lattice Gaussian sampling is equivalent to CVP via a polynomial-time dimension-preserving reduction [42]. Compared to those existing decoding solutions by Euclidean distance, decoding by sampling has promising advantages. Firstly, sampling has the potential to be efficiently implemented, which is a charming decoding solution especially for high-dimensional systems. Secondly, the standard deviation σ can be optimized to improve the sampling probability of the target point, resulting in a better decoding performance. Thirdly, by adjusting the sample size, the sampler decoding enjoys a flexible trade-off between performance and complexity. However, the problem of sampler decoding chiefly lies on how to perform the sampling over the target lattice Gaussian distribution.

C. Lattice Reduction Technique

Lattice reduction techniques have a long tradition in the field of number theory. In 1982, the celebrated LLL algorithm was proposed as a powerful and famous lattice reduction criterion for arbitrary lattice. Specifically, a basis $\mathbf{B}$ is said to be LLL-reduced,[2] if it satisfies the following two conditions,

- $|\mu_{i,j}| \leq \frac{1}{2}$, for $1 \leq j < i \leq n$;
- $\delta\|\widehat{\mathbf{b}}_i\|^2 \leq \|\mu_{i+1,i}\widehat{\mathbf{b}}_i + \widehat{\mathbf{b}}_{i+1}\|^2$, for $1 \leq i < n$.

The first clause is called size reduction condition with $\mu_{i,j} = \langle \mathbf{b}_i, \widehat{\mathbf{b}}_j\rangle / \langle \widehat{\mathbf{b}}_j, \widehat{\mathbf{b}}_j\rangle$, while the second is known as Lovász condition. If Lovász condition is violated, the basis vectors $\mathbf{b}_i$ and $\mathbf{b}_{i+1}$ are swapped; otherwise, size reduction is carried out. If only size reduction condition is satisfied, then the basis is called size-reduced. The parameter $1/4 < \delta < 1$ controls both the convergence speed of the reduction and the degree of orthogonality of the reduced basis.

[1] This corresponds to a sufficiently large standard deviation σ, rendering many cases of interest inapplicable.

[2] Other lattice reduction schemes like Korkin-Zolotarev (KZ) reduction and Seysen reduction also exist, which are out of scope of this work. See [43], [44] for more details.

After LLL reduced, the lattice basis consists of vectors that are relatively short and orthogonal to each other. More precisely, LLL reduction is able to yield a lattice vector within $(2/\sqrt{3})^n$ of the shortest vector in lattice by average polynomial complexity $O(n^4 \log n)$ [45]. Inspired by it, the lattice-reduction-aided decoding has emerged as a powerful decoding strategy in various research fields. In MIMO detection, it has been demonstrated that the LLL reduction based minimum mean square error (MMSE) detection not only attains the full receive diversity [46], but also facilitates the diversity-multiplexing trade-off (DMT) optimal decoding [47]. Meanwhile, LLL reduction can be efficiently realized by *effective LLL reduction* with polynomial complexity $O(n^3 \log n)$ [48]. Nevertheless, the performance gap between the optimal ML decoding and lattice-reduction-aided decoding is still substantial especially in high-dimensional systems.

Algorithm 1: LLL Reduction.

```
Input: $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$
Output: $\overline{\mathbf{B}} = \mathbf{BU}$
compute Gram-Schmidt orthogonality (GSO) $\widehat{\mathbf{B}}$
$k = 2$
while $k \le n$ do
    size-reduce $\mathbf{b}_k$ against $\mathbf{b}_{k-1}$
    if $\|\widehat{\mathbf{b}}_k\|^2 < (\delta - |\mu_{k,k-1}|^2)\|\widehat{\mathbf{b}}_{k-1}\|^2$ then
        swap $\mathbf{b}_k$ and $\mathbf{b}_{k-1}$ update GSO
        $k = \max(k-1, 2)$
    else
        for $l = k-2, k-3, \ldots, 1$do
            size-reduce $\mathbf{b}_k$ against $\mathbf{b}_l$
        end for
        $k = k + 1$
    end if
end while
return $\overline{\mathbf{B}} = \mathbf{B}$
```

D. MCMC Methods

By establishing a Markov chain that randomly generates the next state, MCMC is capable of sampling from the target distribution of interest. As an important parameter which measures the time (i.e., number of Markov moves) required by a Markov chain to get close to its stationary distribution, the *mixing time* $t_{\text{mix}}(\epsilon)$ is defined as [20]

$$t_{\text{mix}}(\epsilon) = \min\{t : \max \|P^t(\mathbf{x}, \cdot) - \pi(\cdot)\|_{TV} \le \epsilon\}, \tag{6}$$

where the integer $t \ge 1$ denotes the index of Markov moves, $\|\cdot\|_{TV}$ represents the total variation distance, π is the target invariant distribution, $P^t(\mathbf{x}; \cdot)$ indicates a row of the transition matrix $\mathbf{P}$ after t Markov moves with the initial state $\mathbf{x}$.[3]

Thanks to the celebrated *coupling technique*, for any Markov chain with finite state space, exponentially fast convergence can be achieved if the underlying Markov chain is irreducible and aperiodic with an invariant distribution π.

Definition 1 ([20]): A Markov chain with stationary distribution π is geometrically ergodic if there exists $0 < \varrho < 1$ and $0 < C(\mathbf{x}) < \infty$ such that

$$\|P^t(\mathbf{x}, \cdot) - \pi(\cdot)\|_{TV} \le C(\mathbf{x})\varrho^t \tag{7}$$

for all $\mathbf{x}$ with $t \ge 1$, where function $C(\mathbf{x})$ is parameterized by the initial state $\mathbf{x}$.

Clearly, coefficient ϱ is the convergence rate of the Markov chain. In comparison, a Markov chain is said to be uniformly ergodic if it is geometrically ergodic and $C(\mathbf{x})$ is a constant independent of $\mathbf{x}$ [49]. However, in the case of lattice Gaussian sampling, the countably infinite state space that $\mathbf{x} \in \mathbb{Z}^n$ imposes a challenge.

III. Lattice-Reduction-Aided Gibbs Algorithm

In this section, the convergence analysis of systematic scan Gibbs algorithm for lattice Gaussian sampling is presented, where its convergence rate is derived by means of HGR maximal correlation. Then, based on the derived convergence rate, lattice reduction technique is adopted into Gibbs sampling for a better convergence performance.

Typically, with respect to Gibbs algorithm for lattice Gaussian sampling, each coordinate of $\mathbf{x}$ is sampled from the following 1-dimensional conditional distribution

$$P_i(x_i|\mathbf{x}_{[-i]}) = D_{\Lambda,\sigma,\mathbf{c}}(x_i|\mathbf{x}_{[-i]}) = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}{\sum_{x_i \in \mathbb{Z}} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}} \tag{8}$$

with $\sigma > 0$. Here $1 \le i \le n$ denotes the coordinate index of $\mathbf{x}, \mathbf{x}_{[-i]} \triangleq [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]^T$. During this univariate sampling, the other $n-1$ variables contained in $\mathbf{x}_{[-i]}$ are leaving unchanged. By repeating such a procedure with a certain scan scheme, a Markov chain $\{\mathbf{X}^0, \mathbf{X}^1, \ldots\}$ is established. Apart from the random scan who randomly updates the component of $\mathbf{x}$, systematic scan proceeds the update in a sequential order from x_n to x_1, thus completing a full iteration. Generally speaking, systematic scan is more preferable in lattice decoding due to its fixed update order. In fact, the mixing times of these two scan schemes do not differ by more than a polynomial factor [50].

A. Simplified Systematic Scan Gibbs Sampling

For simplicity, in order to reveal the relationship between convergence rate and correlation structure, the systematic scan scheme is considered, where its transition probability of each Markov move can be expressed as

$$P(\mathbf{X}^t = \mathbf{x}, \mathbf{X}^{t+1} = \mathbf{y}) = \prod_{i=1}^{n} P_{n-i+1}(x_{n-i+1}^{t+1}|\mathbf{x}_{[-(n-i+1)]}^t). \tag{9}$$

Clearly, for a given standard deviation $\sigma > 0$ and full rank lattice basis $\mathbf{B}$, it is easy to verify that each random variable x_i is sampled with variance

$$\text{var}[x_i|\mathbf{x}_{[-i]}] = \kappa_i > 0. \tag{10}$$

[3] The (i, j)-th entry $P(i; j)$ of transition matrix $\mathbf{P}$ represents the probability of transferring to state j from the previous state i

Therefore, all the sampling candidates of x_i are possible to be sampled theoretically, indicating an irreducible chain. In principle, the irreducible property prevents the random variables to be totally dependent, where all the components of $\mathbf{x}$ for Markov state $\mathbf{X}^t$ may be different with $\mathbf{y}$ of $\mathbf{X}^{t+1}$.

For the sake of convergence analysis, we now formulate the systematic scan Gibbs sampling to a simple version which only consists of two nominal components $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2], \mathbf{x}_1 \in \mathbb{Z}^m$ and $\mathbf{x}_2 \in \mathbb{Z}^{n-m}$. In particular, similar to (8), during a Markov move, $\mathbf{x}_1$ and $\mathbf{x}_2$ are iteratively generated by

$$\mathbf{x}_2^{t+1} \sim P_{\mathbf{x}_2}(\mathbf{x}_2|\mathbf{x}_1^t) = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}{\sum_{\mathbf{x}_2\in\mathbb{Z}^{n-m}} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}} \tag{11}$$

and

$$\mathbf{x}_1^{t+1} \sim P_{\mathbf{x}_1}(\mathbf{x}_1|\mathbf{x}_2^{t+1}) = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}{\sum_{\mathbf{x}_1\in\mathbb{Z}^{m}} e^{-\frac{1}{2\sigma^2}\|\mathbf{Bx}-\mathbf{c}\|^2}}. \tag{12}$$

Through the simplification, the above Markov chain still attains $\pi = D_{\Lambda,\sigma,\mathbf{c}}$ as the invariant distribution while its transition probability becomes

$$P(\mathbf{X}^t = \mathbf{x}, \mathbf{X}^{t+1} = \mathbf{y}) = P_{\mathbf{x}_1}(\mathbf{x}_1^{t+1}|\mathbf{x}_2^t) \cdot P_{\mathbf{x}_2}(\mathbf{x}_2^{t+1}|\mathbf{x}_1^{t+1}). \tag{13}$$

Insight into this simplified Gibbs sampler, the marginal chains $\{\mathbf{x}_1^1, \mathbf{x}_1^2, \ldots\}$ and $\{\mathbf{x}_2^1, \mathbf{x}_2^2, \ldots\}$ with respect to $\mathbf{x}_1$ and $\mathbf{x}_2$ also function as valid Markov chains. Most importantly, these marginal chains experience the same mixing performance as the original chain with convergence rate [51], [52]

$$\varrho = \varrho_1 = \varrho_2, \tag{14}$$

which implies we can obtain the convergence rate of the joint chain by only focusing on its marginal chain. Furthermore, because $\mathbf{x}_1^t$ and $\mathbf{x}_1^{t+1}$ are conditionally independent for a given $\mathbf{x}_2^{t+1}$, the *detailed balance condition* is satisfied by

$$\pi'(\mathbf{x}_1^t)P(\mathbf{x}_1^t, \mathbf{x}_1^{t+1}) = \pi'(\mathbf{x}_1^{t+1})P(\mathbf{x}_1^{t+1}, \mathbf{x}_1^t) \tag{15}$$

indicating that the marginal chain turns out to be reversible. Inspired by it, the following convergence analysis takes place in the marginal Markov chain $\{\mathbf{x}_1^1, \mathbf{x}_1^2, \ldots\}$ with target distribution π' for simplicity.[4]

B. Convergence Analysis Versus HGR Maximal Correlation

Typically, given the transition probability $P(\mathbf{X}^t, \mathbf{X}^{t+1})$, the forward operator $\mathbf{F}$ of the Markov chain is defined as [53]

$$\mathbf{F}h(\mathbf{X}^t) \triangleq \sum_{\mathbf{X}^{t+1}\in\Omega} h(\mathbf{X}^{t+1})P(\mathbf{X}^t, \mathbf{X}^{t+1}) = \mathrm{E}[h(\mathbf{X}^{t+1})|\mathbf{X}^t] \tag{16}$$

with induced operator norm

$$\|\mathbf{F}\| = \sup_{h\in L_0^2(\pi),\mathrm{var}(h)=1} \|\mathbf{F}h\|. \tag{17}$$

Here, $L^2(\pi)$ is the Hilbert space of square integrable functions with respect to π so that $L_0^2(\pi) \triangleq \{h(\mathbf{x}) : \mathrm{E}[h(\mathbf{x})] = 0, \mathrm{var}[h(\mathbf{x})] < \infty\}$ denotes the subspace of $L^2(\pi)$ consisting of functions with zero mean relative to π. More precisely, for $h(\cdot), g(\cdot) \in L_0^2(\pi)$, the inner product defined by the space is

$$\langle h(\mathbf{x}), g(\mathbf{x})\rangle = \mathrm{E}[h(\mathbf{x})g(\mathbf{x})] \tag{18}$$

with variance

$$\mathrm{var}_\pi[h(\mathbf{x})] = \langle h(\mathbf{x}), h(\mathbf{x})\rangle = \|h(\mathbf{x})\|^2. \tag{19}$$

Theorem 1: Given the invariant lattice Gaussian distribution $\pi = D_{\Lambda,\sigma,\mathbf{c}}$, the Markov chain induced by systematic scan Gibbs algorithm is geometrically ergodic

$$\|P^t(\mathbf{x}, \cdot) - D_{\Lambda,\sigma,\mathbf{c}}\|_{TV} \leq C(\mathbf{x})\varrho^t \tag{20}$$

with convergence rate

$$\varrho = \gamma^2(\mathbf{x}_1, \mathbf{x}_2) < 1. \tag{21}$$

Proof: First of all, regarding to the marginal Markov chain $\{\mathbf{x}_1^1, \mathbf{x}_1^2, \ldots\}$, the spectral radius of $\mathbf{F}_1$ is closely related with its norm as [53]

$$\mathrm{spec}(\mathbf{F}_1) = \lim_{t\to\infty} \|\mathbf{F}_1^t\|^{1/t}. \tag{22}$$

Meanwhile, the reversibility of the marginal chain corresponds to a self-adjoint operator $\mathbf{F}_1$ with [54]

$$\|\mathbf{F}_1^t\| = \|\mathbf{F}_1\|^t, \tag{23}$$

then we have

$$\mathrm{spec}(\mathbf{F}_1) = \|\mathbf{F}_1\|. \tag{24}$$

Subsequently, according to (17) and (24), the spectral radius of the forward operator $\mathbf{F}_1$ is derived as

$$\begin{aligned}
\mathrm{spec}(\mathbf{F}_1) = \|\mathbf{F}_1\| &= \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \|\mathbf{F}_1 h\| \\
&= \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \{\mathrm{var}[\mathrm{E}[h(\mathbf{x}_1^{t+1})|\mathbf{x}_1^t]]\}^{\frac{1}{2}} \\
&= \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \{\mathrm{E}[\mathrm{E}^2[h(\mathbf{x}_1^{t+1})|\mathbf{x}_1^t]] \\
&\quad - [\mathrm{E}[\mathrm{E}[h(\mathbf{x}_1^{t+1})|\mathbf{x}_1^t]]]^2\}^{\frac{1}{2}} \\
&= \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \{\mathrm{E}[\mathrm{E}^2[h(\mathbf{x}_1^{t+1})|\mathbf{x}_1^t]]\}^{\frac{1}{2}} \\
&\overset{(a)}{=} \gamma(\mathbf{x}_1^t, \mathbf{x}_1^{t+1}),
\end{aligned} \tag{25}$$

where (a) comes from the definition of HGR maximal correlation in [55] as

$$\gamma^2(\xi, \eta) = \sup_{f(\xi):\mathrm{E}(f)=0,\mathrm{var}(f)=1} \mathrm{E}[\mathrm{E}^2[f(\xi)|\eta]]. \tag{26}$$

With respect to $\gamma(\mathbf{x}_1^t, \mathbf{x}_1^{t+1})$, on one hand, it follows that

$$\begin{aligned}
\gamma(\mathbf{x}_1^t, \mathbf{x}_1^{t+1}) &= \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \mathrm{var}[\mathrm{E}[\mathrm{E}[h(\mathbf{x}_1^{t+1})|\mathbf{x}_2^{t+1}]|\mathbf{x}_1^t]] \\
&\leq \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \mathrm{var}[\mathrm{E}[h(\mathbf{x}_1^{t+1})|\mathbf{x}_2^{t+1}]] \\
&= \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \mathrm{E}[\mathrm{E}^2[h(\mathbf{x}_1) \mid \mathbf{x}_2]] \\
&= \gamma^2(\mathbf{x}_1, \mathbf{x}_2).
\end{aligned} \tag{27}$$

[4] The same result can be obtained with respect to the marginal Markov chain $\{\mathbf{x}_2^1, \mathbf{x}_2^2, \ldots\}$.

On the other hand, we have

$$\begin{aligned}\gamma(\mathbf{x}_1^t,\mathbf{x}_1^{t+1}) &\geq \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \mathrm{corr}[h(\mathbf{x}_1^t),h(\mathbf{x}_1^{t+1})]\\ &= \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \mathrm{E}[h(\mathbf{x}_1^t)h(\mathbf{x}_1^{t+1})]\\ &= \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \mathrm{E}[\mathrm{E}[h(\mathbf{x}_1^t)h(\mathbf{x}_1^{t+1})\mid \mathbf{x}_2^{t+1}]]\\ &= \sup_{h\in L_0^2(\pi'),\mathrm{var}(h)=1} \mathrm{E}[\mathrm{E}^2[h(\mathbf{x}_1)\mid \mathbf{x}_2]]\\ &= \gamma^2(\mathbf{x}_1,\mathbf{x}_2). \end{aligned} \tag{28}$$

Therefore, according to (27) and (28), we get

$$\mathrm{spec}(\mathbf{F}_1)=\gamma(\mathbf{x}_1^t,\mathbf{x}_1^{t+1})=\gamma^2(\mathbf{x}_1,\mathbf{x}_2)<1, \tag{29}$$

where the inequality holds due to the fact that $\mathbf{x}_1$ and $\mathbf{x}_2$ are random variables of each other by configuration.

Next, by invoking the following Lemma from [56], the marginal chain $\{\mathbf{x}_1^1,\mathbf{x}_1^2,\ldots\}$ turns out to be geometrically ergodic with convergence rate

$$\varrho_1=\mathrm{spec}(\mathbf{F}_1). \tag{30}$$

Lemma 1 ([56]): Given the invariant distribution π, a reversible, irreducible and aperiodic Markov chain with spectral gap $\varphi=1-\mathrm{spec}(\mathbf{F})>0$ converges exponentially as

$$\|P^t(\mathbf{x},\cdot)-\pi(\cdot)\|_{TV}\leq C(\mathbf{x})(1-\varphi)^t. \tag{31}$$

Hence, from (14) and (30), the original Markov chain $\{\mathbf{X}^1,\mathbf{X}^2,\ldots\}$ is geometric ergodicity with exponential convergence rate

$$\varrho=\gamma^2(\mathbf{x}_1,\mathbf{x}_2)<1, \tag{32}$$

completing the proof. ∎

From Theorem 1, the convergence rate ϱ is determined by the HGR maximal correlation among the multivariate of being sampled, which provides a meaningful way for the convergence enhancement. Theoretically, HGR maximal correlation is an elegant generalization of the well-known *Pearson correlation coefficient*, and serves as a normalized measure of the dependence between two random variables. Apart from Pearson correlation coefficient, $\gamma(\xi,\eta)$ is defined whenever both ξ and η are non-degenerate, which assumes values in the interval [0, 1] and vanish if and only if ξ and η are independent. Clearly, $0\leq\gamma(\mathbf{x}_1,\mathbf{x}_2)<1$ measures the dependence between $\mathbf{x}_1$ and $\mathbf{x}_2$, where $\gamma(\mathbf{x}_1,\mathbf{x}_2)=0$ if and only if $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent of each other. On the other hand, the high correlation between $\mathbf{x}_1$ and $\mathbf{x}_2$ gives rise to a larger value of $\gamma(\mathbf{x}_1,\mathbf{x}_2)$ approaching to 1. It should be noticed that such a result can be easily generalized as $\varrho=\gamma^2(x_i,\mathbf{x}_{[-i]})$, where a less correlation among x_i and $\mathbf{x}_{[-i]}$ for $1\leq i\leq n$ is also the sufficient condition for a small value of $\gamma(\mathbf{x}_1,\mathbf{x}_2)$.

Remark 1: The convergence rate of systematic scan Gibbs sampling for the lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$ is dominated by the HGR maximal correlation $\gamma(x_i,\mathbf{x}_{[-i]})$ among random variables x_i's, where the optimal convergence $\varrho=0$ happens when x_i's are independent of each other.

C. Lattice-Reduction-Aided Gibbs Sampling Algorithm

From the convergence analysis, in order to achieve an efficient Markov mixing, a smaller $\gamma(x_i,\mathbf{x}_{[-i]})$, $1\leq i\leq n$ is strongly desired. However, it is hard to explicitly calculate γ in practice. Regarding to the lattice Gaussian distribution shown in (3), it is clear that the correlation over elements of $\mathbf{x}$ is decided by matrix $\mathbf{B}$, i.e., the more orthogonal of $\mathbf{B}$, the less correlation of components in $\mathbf{x}$. For this reason, we attempt to use the *orthogonality defect* of $\mathbf{B}$ to partially characterize $\gamma(x_i,\mathbf{x}_{[-i]})$.

Specifically, the orthogonality defect of a matrix $\mathbf{B}$ is defined as [45]

$$\xi(\mathbf{B})=\frac{\prod_{i=1}^n\|\mathbf{b}_i\|}{|\det(\mathbf{B})|}, \tag{33}$$

where $\det(\cdot)$ represent the determinant of the square matrix. According to *Hadamard inequality*, the orthogonality defect is lower bounded by $\xi(\mathbf{B})\geq 1$, where the equality holds if and only if vectors in $\mathbf{B}$ are mutually orthogonal. Consequently, we can easily arrive at the following Lemma, whose proof is omitted here due to simplicity.

Lemma 2: If the full rank matrix $\mathbf{B}\in\mathbb{R}^{n\times n}$ is an orthogonal matrix with $\xi(\mathbf{B})=1$, then $\gamma(x_i,\mathbf{x}_{[-i]})=0$ for $1\leq i\leq n$, and samples from lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$ can be immediately obtained by systematic scan Gibbs sampling with convergence rate

$$\varrho=0. \tag{34}$$

Clearly, a smaller value of $\xi(\mathbf{B})$ is in high demand for the fast mixing. However, for a given lattice basis $\mathbf{B}$, any attempt to reduce $\xi(\mathbf{B})$ directly for a small $\gamma(x_i,\mathbf{x}_{[-i]})$ is impossible. Nevertheless, an alternative way can still be carried out by resorting to lattice reduction technique [45], which transfers the lattice Gaussian distribution in (3) to an equivalent one:

$$\pi(\mathbf{z})=\frac{e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}{\sum_{\mathbf{z}\in\mathbb{Z}^n}e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}, \tag{35}$$

where $\overline{\mathbf{B}}=\mathbf{B}\mathbf{U}$, $\mathbf{z}=\mathbf{U}^{-1}\mathbf{x}\in\mathbb{Z}^n$ and $\mathbf{U}\in\mathbb{Z}^{n\times n}$ is a unimodular matrix with $\det(\mathbf{U})=\pm1$.

Undoubtedly, $\overline{\mathbf{B}}\mathbf{z}$ and $\mathbf{B}\mathbf{x}$ describe the same lattice point in the space. Therefore, the target distribution $\pi=D_{\Lambda,\sigma,\mathbf{c}}$ essentially maintains unchanged during this transformation but is parameterized by $\mathbf{z}$, where there is a one-to-one correspondence between $\mathbf{x}$ and $\mathbf{z}$. Then, with respect to the Gibbs sampling, the conditional sampling probability of Gibbs sampling shown in (12) becomes

$$P_{\mathbf{z}_1}(\mathbf{z}_1|\mathbf{z}_2)=\frac{e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}{\sum_{\mathbf{z}_1\in\mathbb{Z}^m}e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}} \tag{36}$$

with $\mathbf{z}_1\in\mathbb{Z}^m$ and $\mathbf{z}_2\in\mathbb{Z}^{n-m}$, and can be further generalized to

$$P_i(z_i|\mathbf{z}_{[-i]})=\frac{e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}{\sum_{z_i\in\mathbb{Z}}e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}. \tag{37}$$

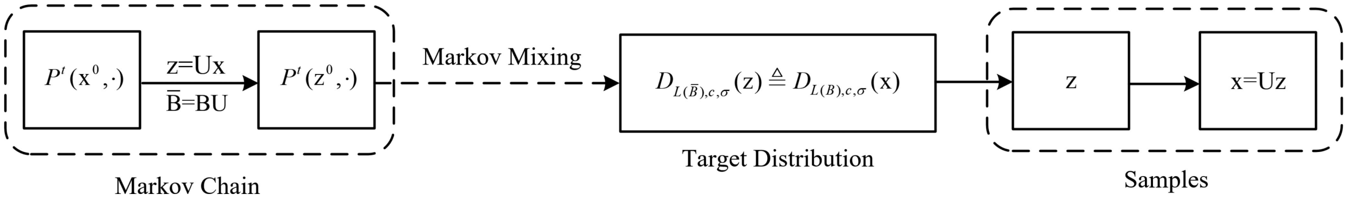

Fig. 2. Illustration of the lattice-reduction-aided Gibbs sampler for lattice Gaussian sampling.

In particular, as shown in Fig. 2, given the target distribution $\pi(\mathbf{x}) = D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$, the proposed lattice-reduction-aided Gibbs sampling consists of the following three steps:

1) Generate the equivalent lattice Gaussian distribution $D_{\mathcal{L}(\overline{\mathbf{B}}),\sigma,\mathbf{c}}(\mathbf{z})$ by LLL reduction.
2) Perform the Gibbs sampling over $D_{\mathcal{L}(\overline{\mathbf{B}}),\sigma,\mathbf{c}}(\mathbf{z})$.
3) Collect samples of $\mathbf{z}$ after the Markov mixing and output samples of $\mathbf{x}$ by $\widehat{\mathbf{x}} = \mathbf{U}\widehat{\mathbf{z}}$.

On the other hand, similarly, it is straightforward to verify the Gibbs sampling with respect to the converted lattice Gaussian distribution is also geometrically ergodic.

Theorem 2: Given $D_{\mathcal{L}(\overline{\mathbf{B}}),\sigma,\mathbf{c}}(\mathbf{z})$, the Markov chain induced by Gibbs sampling converges exponentially fast:

$$\|P^t(\mathbf{z},\cdot) - D_{\mathcal{L}(\overline{\mathbf{B}}),\sigma,\mathbf{c}}(\cdot)\|_{TV} \leq C'(\mathbf{z})(\varrho')^t, \tag{38}$$

where $D_{\mathcal{L}(\overline{\mathbf{B}}),\sigma,\mathbf{c}}(\mathbf{z}) \triangleq D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x})$.

Remarkably, such a slight change by replacing $\mathbf{x}$ with $\mathbf{z}$ introduces a significant benefit: compared to $\mathbf{B}$, the orthogonality of matrix $\overline{\mathbf{B}}$ is greatly improved by lattice reduction. More specifically, it has been demonstrated that after LLL reduction, the orthogonality defect of the reduced basis $\overline{\mathbf{B}}$ is upper bounded by [45]

$$\xi(\overline{\mathbf{B}}) \leq \beta^{\frac{n(n-1)}{4}} \tag{39}$$

with $\beta = (\delta - \frac{1}{4})^{-1}$, indicating a guaranteed reduction from $\xi(\mathbf{B})$ to $\xi(\overline{\mathbf{B}})$. Therefore, a smaller HGR maximal correlation over components within $\mathbf{z}$ is most likely to be achieved, i.e., $\gamma(\mathbf{z}_1, \mathbf{z}_2) \leq \gamma(\mathbf{x}_1, \mathbf{x}_2)$, thus leading to a better convergence rate by Theorem 2.

Remark 2: With respect to sampling from the lattice Gaussian distribution $D_{\Lambda,\sigma,\mathbf{c}}$, the usage of lattice reduction is capable of achieving less correlated random variable z_i's than x_i's, which leads to a more efficient Markov mixing.

To summarize, the proposed lattice-reduced-aided Gibbs sampling algorithm is presented in Algorithm 2. Since the operation of LLL reduction with polynomial complexity $O(n^3 \log n)$ can be performed as a preprocessing stage, the complexity of Gibbs sampling in each single Markov move is easily accepted with $O(n^2)$. Because of this, in MCMC the complexity of each Markov move is often insignificant, whereas the number of Markov moves is more critical.

Algorithm 2: Lattice-Reduction-Aided Gibbs Algorithm for Lattice Gaussian Sampling.

```
Input: $\mathbf{B}, \sigma, \mathbf{c}, \mathbf{X}^0, t_{\text{mix}}(\epsilon)$
Output: $\mathbf{x} \sim D_{\Lambda,\sigma,\mathbf{c}}$
let $\mathbf{x}^0$ denote the intial state of $\mathbf{X}^0$
obtain $\overline{\mathbf{B}} = \mathbf{B}\mathbf{U}$ and $\mathbf{z}^0 = \mathbf{U}^{-1}\mathbf{x}^0$ via LLL reduction
for $t = 1, 2, ...$ do
    for $i = n, \ldots, 1$ do
        sample $z_i^t$ from $P(z_i|\mathbf{z}_{[-i]})$ shown in (37)
    end for
    update $\mathbf{z}$ with the sampled $z_i$ and let $\mathbf{Z}^t = \mathbf{z}$
    if $t \geq t_{\text{mix}}(\epsilon)$ then
        output the state of $\mathbf{X}^t = \mathbf{U}\mathbf{Z}^t$
    end if
end for
```

IV. Lattice-Reduction-Aided Gibbs Sampling Algorithm for Lattice Decoding

In this section, we extend the proposed lattice-reduction-aided Gibbs sampling to lattice decoding. Theoretically, when MCMC method is applied for sampler decoding, its decoding performance can be evaluated by CVP decoding complexity (i.e., the number of Markov move t), which is defined by [19]

$$C_{\text{cvp}} \triangleq \frac{t_{\text{mix}}}{D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x}_{\text{cvp}})}. \tag{40}$$

Here, the mixing time t_{mix} serves as a pick-up gap to guarantee i.i.d. samples because samples from the stationary distribution tend to be correlated with each other. Besides, $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x}_{\text{cvp}})$ denotes the sampling probability of the target CVP point. Therefore, in order to strengthen the decoding performance, one can either reduce the mixing time t_{mix} (e.g., use LLL reduction to boost the convergence), or improve the sampling probability $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{x}_{\text{cvp}})$, which will be studied in the following.

A. Choice of the Sampling Deviation σ

From the point of view of simulated annealing in statistics, σ functions as "temperature" to guide the Markov mixing, which also has an impact upon t_{mix} as well. Given the lattice Gaussian distribution $\pi(\mathbf{z})$ shown in (35), although a small size σ corresponds to a relatively large decoding sampling probability $D_{\Lambda,\sigma,\mathbf{c}}(\mathbf{z}_{\text{cvp}})$, it also incurs a "cold" Markov chain which tends to be trapped by the frozen status, and vice versa [57].[5] However, since t_{mix} for Gibbs sampling is hard to get at the current stage, to balance this inherent trade-off for a better decoding performance, a feasible compromise is to ensure a reliable sampling probability given moderate σ.

[5] Actually, this is in accordance with the result of independent MHK sampling algorithm for lattice Gaussian distribution, where the exact convergence rate as well as the mixing time t_{mix} can be estimated [18], [19].

In particular, with respect to any $\mathbf{z} \in \mathbb{Z}^n$ to be sampled, we firstly extract σ from the denominator of $\pi(\mathbf{z})$ as

$$\begin{aligned}\pi(\mathbf{z}) &= \frac{e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}{\sum_{\mathbf{z}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}\\ &\overset{(a)}{\geq} \frac{e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}{\sum_{\mathbf{z}\in\mathbb{Z}^n} e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}\|^2}}\\ &\overset{(b)}{\geq} \frac{e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}{(\sqrt{2\pi}\sigma)^n\sum_{\mathbf{z}\in\mathbb{Z}^n} e^{-\pi\|\overline{\mathbf{B}}\mathbf{z}\|^2}}\\ &= f(\sigma)\cdot c \ \text{ for } \sqrt{2\pi}\sigma \geq 1 \end{aligned} \tag{41}$$

where

$$c \triangleq 1\Big/\sum_{\mathbf{z}\in\mathbb{Z}^n} e^{-\pi\|\overline{\mathbf{B}}\mathbf{z}\|^2} \tag{42}$$

is a constant and

$$f(\sigma) \triangleq \frac{e^{-\frac{1}{2\sigma^2}\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2}}{(\sqrt{2\pi}\sigma)^n} \tag{43}$$

is parameterized by σ. Here, (a) and (b) respectively obey the facts from lattice theory ([1, Lemma 1.4]) that

$$\sum_{\mathbf{v}\in\Lambda} e^{-\frac{1}{2\sigma^2}\|\mathbf{v}-\mathbf{c}\|^2} \leq \sum_{\mathbf{v}\in\Lambda} e^{-\frac{1}{2\sigma^2}\|\mathbf{v}\|^2} \tag{44}$$

and

$$\sum_{\mathbf{v}\in\Lambda} e^{-\pi s^{-1}\|\mathbf{v}\|^2} \leq s^{\frac{n}{2}} \cdot \sum_{\mathbf{v}\in\Lambda} e^{-\pi\|\mathbf{v}\|^2}, \text{ for } s \geq 1. \tag{45}$$

From (41), it is natural to see that the sampling probability for any specific $\mathbf{z}$ is lower bounded by the function $f(\sigma)$. Furthermore, the derivative of function $f(\sigma)$ with respect to $\sigma \geq 1/\sqrt{2\pi}$ is derived as follows

$$\frac{\partial f(\sigma)}{\partial \sigma} = \frac{\left(n\sigma^2 - \left\|\bar{\mathbf{B}}\mathbf{z}-\mathbf{c}\right\|^2\right)\exp\left(-\frac{\left\|\bar{\mathbf{B}}\mathbf{z}-\mathbf{c}\right\|^2}{2\sigma^2}\right)}{\sigma^{n+3}\left(\sqrt{2\pi}\right)^n}. \tag{46}$$

Subsequently, let the above derivative be zero, the optimized σ that maximizes $f(\sigma)$ is obtained as

$$\sigma = \max\left\{\frac{\|\ \overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\ \|}{\sqrt{n}}, \frac{1}{\sqrt{2\pi}}\right\}, \tag{47}$$

which implies that σ should vary with $\|\ \overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\ \|$ for a large lower bound of $\pi(\mathbf{z})$.

Clearly, the existence of the lower bound for $\pi(\mathbf{z})$ guarantees a reliable sampling probability of $\mathbf{z}$, which could be further optimized by the careful selection of σ. Meanwhile, the requirement of $\sigma \geq 1/\sqrt{2\pi}$ serves as a baseline to ensure the Markov chain evolves dynamically, even though the sampling probability below $\sigma = 1/\sqrt{2\pi}$ seems rather attractive.

Hence, as for the target point $\mathbf{z}_{\text{cvp}}$ for lattice decoding, the choice of σ due to (47) turns out to be

$$\sigma_{\text{cvp}} = \max\left\{\frac{\|\ \overline{\mathbf{B}}\mathbf{z}_{\text{cvp}}-\mathbf{c}\ \|}{\sqrt{n}}, \frac{1}{\sqrt{2\pi}}\right\}. \tag{48}$$

Generally speaking, regarding to different configurations of $\mathbf{B}$ and $\mathbf{w}$, such a flexible setting of σ_{cvp} is more beneficial to the sampler decoding by providing a specific rather than statistic choice. For small value of $\|\overline{\mathbf{B}}\mathbf{z}_{\text{cvp}}-\mathbf{c}\|$, σ_{cvp} tends to get smaller since $\mathbf{c}$ appears close to the lattice and vice versa, thus adaptively guiding the choice of σ for each $\mathbf{z}_{\text{cvp}}$.

Unfortunately, it is impossible to get $\mathbf{z}_{\text{cvp}}$ for σ_{cvp}. Therefore, in practice, the initial starting point $\mathbf{z}^0$ can be applied as an approximation. Clearly, the closer of $\mathbf{z}^0$ to $\mathbf{z}_{\text{cvp}}$, the more accurate of the selected σ. This essentially poses a stringent request for the selection of $\mathbf{z}^0$. Fortunately, thanks to the lattice reduction, the required high quality initial starting point in lattice-reduction-aided Gibbs sampling can be guaranteed. In this paper, the classic Babai's nearest plane algorithm (also known as successive interference cancelation (SIC) in MIMO detection) is utilized by

$$\mathbf{z}^0 = \mathbf{z}_{\text{lll-sic}}, \tag{49}$$

where the decoding of $\mathbf{z}_{\text{lll-sic}}$ can be executed during the transformation from $D_{\mathcal{L}(\mathbf{B}),\sigma,\mathbf{c}}(\mathbf{x})$ to $D_{\mathcal{L}(\overline{\mathbf{B}}),\sigma,\mathbf{c}}(\mathbf{z})$. To summarize, we reformat the proposed standard deviation as

$$\sigma_{\text{distance}} = \max\left\{\frac{\|\ \overline{\mathbf{B}}\mathbf{z}_{\text{lll-sic}}-\mathbf{c}\ \|}{\sqrt{n}}, \frac{1}{\sqrt{2\pi}}\right\}. \tag{50}$$

Again, we emphasize that other decoding schemes are also applicable to output $\mathbf{z}^0$ while the decoding performance improves with the accuracy of the approximation.

Besides the sampling probability, the initial starting point also plays an important role in the Markov mixing. More specifically, for the small set $C = \{\mathbf{x} : V(\mathbf{x}) = \pi(\mathbf{x})^{-c} \leq d, c > 0\}$ and $d > 2b/(1-\lambda)$, the geometric ergodicity Markov chains will converge exponentially to the stationary distribution $\pi(\mathbf{x})$ as [58]

$$\begin{aligned}\|P^n(\mathbf{x}^0,\cdot)-\pi(\cdot)\|_{TV} \leq (1-\delta)^{rn} + \left(\frac{U^r}{\alpha^{1-r}}\right)^n\\ \times\left(1+\frac{b}{1-\lambda}+V(\mathbf{x}^0)\right),\end{aligned} \tag{51}$$

where $0 < r < 1$, $0 < \lambda < 1$, $U = 1 + 2(d+b)$ and $\alpha = \frac{1+d}{1+2b+\lambda d}$. From (51), starting the Markov chain with $\mathbf{z}^0$ as close to the center of the lattice Gaussian distribution (i.e., the query point $\mathbf{c}$) as possible would be a judicious choice for the efficient t_{mix}, which is accordance with our suggestion.

On the other hand, since $\mathbf{w}$ in (5) entails the additive white Gaussian noise (AWGN) with zero mean and variance σ_w^2, it follows that

$$\|\overline{\mathbf{B}}\mathbf{z}-\mathbf{c}\|^2 = \|\mathbf{B}\mathbf{x}-\mathbf{c}\|^2 \approx n\sigma_w^2 \tag{52}$$

by the *law of large numbers*. Then, by simply substituting (52) into (47), the choice of σ can be obtained in a statistic way, that is,

$$\sigma_{\text{statistic}} = \max\left\{\sigma_w, \frac{1}{\sqrt{2\pi}}\right\}. \tag{53}$$

Interestingly, we point out that $\sigma = \sigma_w$ is just the conventional wisdom that is widely accepted by related works. However, compared to $\sigma_{\text{statistic}}$, it severely suffers from the *stalling problem* as σ_w^2 shrinks intensively with the increase of SNR. Therefore, the lower bound $\sigma \geq 1/\sqrt{2\pi}$ serves as a necessary complement to active the sampling away from the frozen status. Note that the consistency behind choices of $\sigma_{\text{statistic}}$ and σ_w suggests our analysis based on the sampling probability is tight enough, and we then advance it to more specific cases.

B. Startup Mechanism Based on Correct Decoding Radius R

The application of the initial starting point arises a natural question: whether Gibbs sampling is necessary to every decoding case? In what follows, we try to answer this question from the perspective of correct decoding radius of BDD.

Theoretically, BDD targets at solving the decoding problem when the query point is close to the lattice within a certain distance, which corresponds to a restricted variant of CVP. In BDD, the concept of correct decoding radius R was proposed to serve as a benchmark for evaluating the decoding performance [59]. More specifically, CVP is guaranteed to be solved if the distance between the query point $\mathbf{c}$ and the lattice Λ (i.e., $d(\Lambda, \mathbf{c})$) is less than R. As for Babai's nearest plane algorithm, its correct decoding radius is given by [59]

$$R_{\text{lll-sic}} = \frac{1}{2} \min_{1\leq i\leq n} \|\widehat{\overline{\mathbf{b}}}_i\|. \tag{54}$$

Here, we highlight the significance of LLL reduction again as it greatly increases $\min_i \|\widehat{\overline{\mathbf{b}}}_i\|$ compared to $\min_i \|\widehat{\mathbf{b}}_i\|$. Furthermore, it has been shown in [59] that $R_{\text{lll-sic}}$ is lower bounded as

$$R_{\text{lll-sic}} \geq \frac{1}{2\sqrt{n}\beta^{\frac{n-1}{4}}} \lambda_1(\mathbf{B}), \tag{55}$$

where $\beta = 1/(\delta - 1/4)$ and λ_1 denotes the *minimum distance* of the lattice $\mathcal{L}(\mathbf{B})$. Therefore, for the consideration of decoding efficiency, the correct decoding radius $R_{\text{lll-sic}}$ can be applied as a theoretical judgement to make the decision whether invoke Gibbs sampling or not. This means substantial decoding complexity will be saved without performance loss.

In particular, let $\mathbf{z}^0 = \mathbf{z}_{\text{lll-sic}}$, the startup mechanism works based on the threshold $R_{\text{lll-sic}}$. If $\|\overline{\mathbf{B}}\mathbf{z}^0 - \mathbf{c}\| \leq R_{\text{lll-sic}}$, then there is no need to recall Gibbs sampler as $\mathbf{z}_{\text{cvp}} = \mathbf{z}_{\text{lll-sic}}$ for sure. Otherwise, Gibbs sampler is activated for a better decoding performance. Moreover, such a judgement can be further relaxed with a constant $\alpha \geq 1$

$$\|\overline{\mathbf{B}}\mathbf{z}^0 - \mathbf{c}\| \leq \frac{\alpha}{2} \min_{1\leq i\leq n} \|\widehat{\overline{\mathbf{b}}}_i\|, \tag{56}$$

which leads to a flexible trade-off between decoding performance and efficiency.

From the perspective of efficient sampler decoding, the need for the high quality initial starting point $\mathbf{z}^0$ is also in strong demand for providing a large size of correct decoding radius R. In essence, those demands actually reveal a salient feature of geometric ergodicity: the selection of the initial starting point is an indispensable part of the Markov mixing, which is worth to be well studied. Here, we use it to work for the choice of σ, the pursuit of convergence as well as the startup mechanism, and we believe our work is just the tip of the iceberg. Meanwhile, the proposed startup mechanism based on the correct decoding radius also provides an adaptive strategy for other lattice decoding schemes especially in tackling with high-dimensional scenarios.

C. Performance Improvement by Recycling Gibbs Sampling

In standard Gibbs sampling, the state of the next Markov move is obtained when all the components of $\mathbf{x}$ are updated (i.e., systematic scan) or n times component of $\mathbf{x}$ are randomly picked up to update (i.e., random scan). However, in [60], the conception of recycling was proposed, which takes the samples with only one updated component into account, i.e.,

$$\mathbf{x}_{\text{recycle}-i} = [x_1, \ldots, x_{i-1}, x_{\text{update}-i}, x_{i+1}, \ldots, x_n]^T. \tag{57}$$

In other words, with t-times Markov moves, there are eventually $t \cdot n$ samples of $\mathbf{x}$, and the t samples outputted by standard Gibbs sampling is just a subset of it. Typically, this scheme is named as trivial recycling Gibbs (TRG) sampling.

It is shown that with the extra $t \cdot (n-1)$ samples, there are no apparent advantages of TRG over standard Gibbs in terms of the approximation of the marginal densities, and multiple recycling Gibbs (MRG) was further proposed thereafter [60]. However, in lattice decoding, those extra $t \cdot (n-1)$ samples greatly expand the candidate list without extra computational cost. This is similar to list decoding, where the decoding performance gradually improves with the valid list size [61]. Note that the large number of samples of recycling Gibbs naturally suits a more dynamic searching over the state space while a conservative choice of σ will limit its sample diversity. Therefore, a reasonably large σ is recommended in lattice decoding for a better decoding performance. Clearly, such a recycling operation can be easily adopted to the proposed lattice-reduction-aided Gibbs sampling algorithm by

$$\mathbf{z}_{\text{recycle}-i} = [z_1, \ldots, z_{i-1}, z_{\text{update}-i}, z_{i+1}, \ldots, z_n]^T. \tag{58}$$

Besides recycling Gibbs, the adaptive Gibbs sampling, which dynamically updates the transition probability of the Markov chain by learning from the collected samples, also has great potential in lattice Gaussian sampling and lattice decoding [62], [63]. However, because the adaptive algorithm has the risk to be no longer a valid Markov chain, the design as well as the related analysis is challenging. Although feasible methods for the convergence guarantee was given in [64], care must be taken in tacking with lattice Gaussian sampling. For this reason, we leave the research of adaptive Gibbs-based sampling as a future work. Note that lattice-reduction-aided recycling Gibbs sampling seems well suited for the soft-output detection in MIMO bit-interleaved coded modulation iterative detection and decoding (BICM-IDD) systems by contributing more qualified candidates, where further study with respect to coded MIMO systems by using MCMC will be one of our research works in future [65]–[68].

D. Adoption for Large-Scale MIMO Detection

In lattice decoding, both $\mathbf{x}$ and $\mathbf{z}$ have the same state space $\mathbb{Z}^n$. However, as for MIMO systems, the transmitted signal $\mathbf{x} \in \mathcal{X}^n$ normally belongs to a finite M-QAM constellation. Therefore,

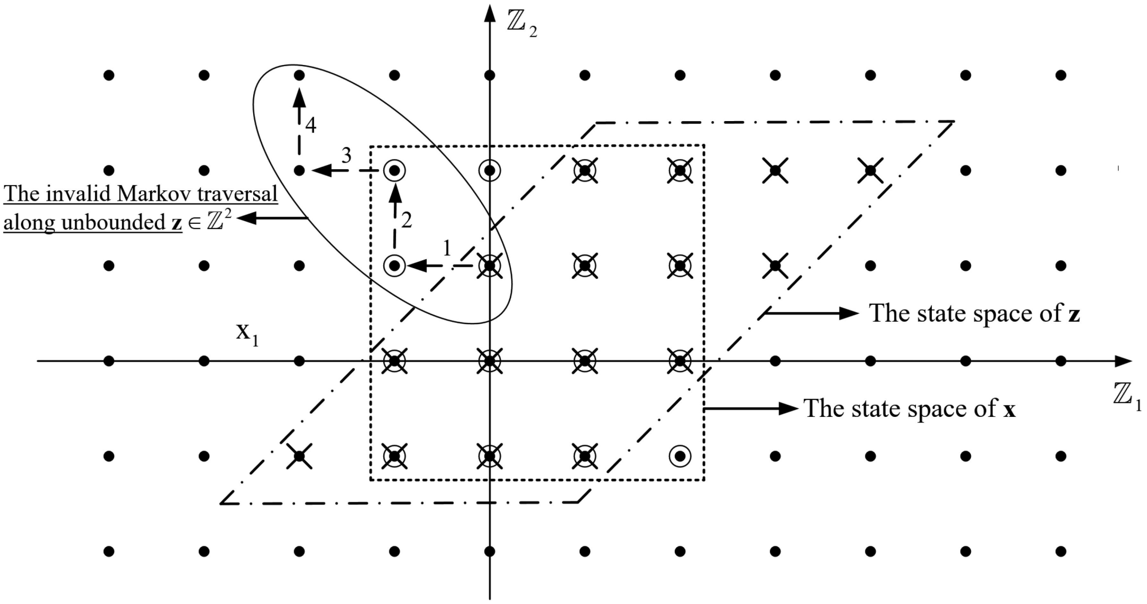
$\mathbb{Z}_2$
$\mathbb{Z}_1$
x_1
4
3
2
1
The invalid Markov traversal along unbounded $\mathbf{z} \in \mathbb{Z}^2$
The state space of $\mathbf{z}$
The state space of $\mathbf{x}$

Fig. 3. Illustration of the defective Markov moves along with the unbounded state space $\mathbf{z} = \mathbf{U}^{-1}\mathbf{x}$. The original symbols $\mathbf{x} \in \mathcal{X}^2$ with $\mathcal{X} = \{-1, 0, 1, 2\}$ and $U = [0\ 1; 1 - 1]$, and the index 1, 2, 3, 4 along the arrows stand for the possible defective Markov moves.

an inherent disadvantage associated with lattice reduction does exist since the state space of $\mathbf{z} = \mathbf{U}^{-1}\mathbf{x}$ after the nonlinear transformation is computationally expensive to get [69]. Generally, in lattice-reduction-aided detection for MIMO systems, there are two suboptimal remedies to alleviate this problem. The first one directly discards those out-of-region points, which is referred to as naive lattice decoding (NLD) [70]. Another remedy is to restrict $\widehat{\mathbf{x}} = \mathbf{U}\mathbf{z}$ to the original set $\mathcal{X}^n$, which is commonly accepted [43].

Unfortunately, both of these two remedies are not compatible with lattice-reduction-aided Gibbs sampling for MIMO detection as they only focus on the restriction of $\mathbf{x}$ at the final decision stage. In sharp contrast to them, after the transformation by lattice reduction, the Markov mixing along $\mathbf{z}$ requires a clear state space, otherwise the Markov chain is going to be invalid due to the wild mixing. Although approximation for the state space of $\mathbf{z}$ can be roughly made, it does not exactly correspond to the original state space of $\mathbf{x}$, leading to a defective Markov mixing. For a better understanding, Fig. 3 is presented as an illustration.

The essential reason behind such a problem is due to the acceptance mechanism of univariate sampling, i.e., every sampling candidate is accepted without any extra judgement or restriction. This actually raises a stringent requirement about the state space of Gibbs sampling since even a tiny disorder at the beginning would lead to a terrible error propagation along the mixing. As a comparison, the MCMC based independent Metropolis-Hastings-Klein (MHK) algorithm utilizes an acceptance ratio to decide whether to admit the sample candidate or not,[6] and LLL reduction has been well adopted to it for a better decoding performance in MIMO detection [19]. To this end, in MIMO detection, lattice reduction without clear state space of $\mathbf{z} = \mathbf{U}^{-1}\mathbf{x}$ is not recommended to participate in the Markov mixing. This is in line with the observations from [30], but it attributes the incompatibility to the increment of local minima by lattice reduction, which actually also exists in the lattice-reduction-aided detection.

Nevertheless, lattice reduction still works for Gibbs sampling in MIMO detection as a preprocessing stage to output the required initial starting point $\mathbf{x}^0 = \mathbf{x}_{\text{lll-sic}}$. Meanwhile, it is also easy to verify that our analysis about the choice of σ as well as the startup mechanism for cases of $\mathbf{z}$ suit well for cases of $\mathbf{x}$ (simple scaling and shifting with respect to $\mathbf{x}$ are necessary to make it continuous integer). Therefore, given $\mathbf{x}^0 = \mathbf{x}_{\text{lll-sic}}$, considerable performance gain and complexity reduction can be achieved.

V. Simulation

In this section, the performance of the proposed Gibbs sampling is evaluated in the large-scale uncoded MIMO detection. Specifically, the ith entry of the transmitted signal $\mathbf{x}$, denoted as x_i, is a modulation symbol taken independently from an M-QAM constellation $\mathcal{X}$ with Gray mapping. Meanwhile, we assume a flat fading environment, where the square channel matrix $\mathbf{H}$ contains uncorrelated complex Gaussian fading gains with unit variance and remains constant over each frame duration. Let E_b represents the average power per bit at the receiver, then the signal-to-noise ratio (SNR) $E_b/N_0 = n/(\log_2(M)\sigma_w^2)$ where M is the modulation level and σ_w^2 is the noise variance. Then, we can express the system model as

$$\mathbf{c} = \mathbf{H}\mathbf{x} + \mathbf{w}. \tag{59}$$

Clearly, this decoding problem of $\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{X}^n} \|\mathbf{H}\mathbf{x} - \mathbf{c}\|^2$ can be solved by sampling over the discrete Gaussian

[6] In principle, Gibbs sampling can be viewed as a special case of Metropolis-Hastings algorithm with acceptance ratio $p \equiv 1$.

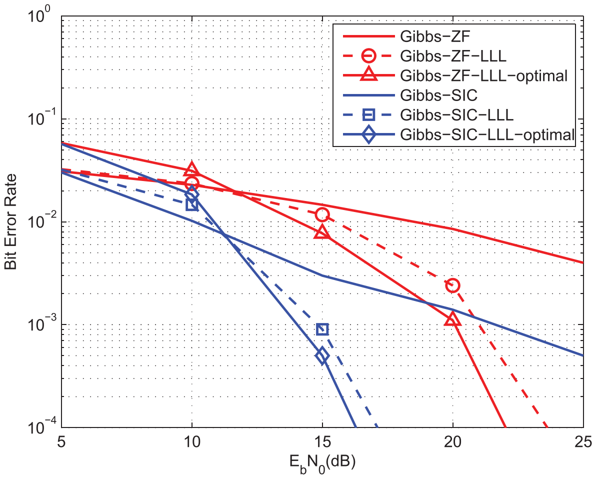

Fig. 4. Bit error rate versus average SNR per bit for the uncoded 16 × 16 MIMO system using 4-QAM.

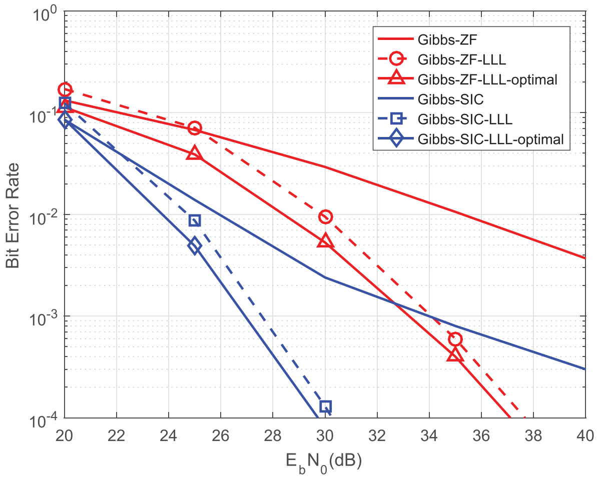

Fig. 5. Bit error rate versus average SNR per bit for the uncoded 32 × 32 MIMO system using 64-QAM.

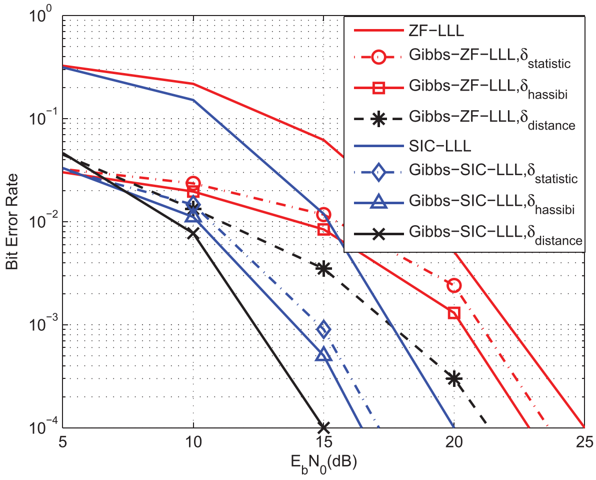

Fig. 6. Bit error rate versus average SNR per bit for the uncoded 16 × 16 MIMO system using 4-QAM.

distribution

$$P_{\mathcal{L}(\mathbf{H}),\sigma,\mathbf{c}}(\mathbf{x}) = \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{Hx}-\mathbf{c}\|^2}}{\sum_{\mathbf{x}\in\mathcal{X}^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{Hx}-\mathbf{c}\|^2}}. \tag{60}$$

Fig. 4 shows the bit error rate (BER) of lattice-reduction-aided Gibbs sampling detectors in a 16 × 16 uncoded MIMO system with 4-QAM. This corresponds to a lattice decoding scenario with restricted state space in dimension $n = 32$. The systematic scan Gibbs algorithm performs 1-dimensional conditional sampling in a backward order, thus completing a Markov move by a full iteration. As a fair comparison, the number of Markov moves is set as $t = 50$. Meanwhile, the standard deviation σ applies $\sigma_{\text{statistic}}$ in (53) for all the Gibbs samplers, which is able to avoid the stalling problem in high SNR region. As can be seen clearly, under the help of LLL reduction, significant performance gain can be achieved by the proposed lattice-reduction-aided Gibbs sampling algorithm. Moreover, if the state space of $\mathbf{z} = \mathbf{U}^{-1}\mathbf{x}$ is known, then the decoding will achieve the optimal performance. However, finding the state space of $\mathbf{z}$ is unaffordable in practice due to the exhausted search, which makes the suboptimal solution with the initial starting point outputted by the lattice-reduction-aided decoder as an alternative. Nevertheless, its performance gain compared to the standard Gibbs is still substantial. Another observation is that the decoding performance of lattice-reduction-aided Gibbs sampling based on SIC outperforms that of zero-forcing (ZF). In terms of the convergence behaviour of geometric ergodicity, this is because the initial starting point provided by SIC-LLL is closest to the center of the distribution shown in (60), thus enabling a most efficient Markov mixing. As a complement, Fig. 5 shows the BER of Gibbs sampling detectors in a 32 × 32 uncoded MIMO system with 64-QAM. As expected, considerable decoding performance gain can be obtained by the proposed lattice-reduction-aided Gibbs sampling algorithm.

Fig. 6 shows the BER performance of Gibbs sampling detectors in a 16 × 16 uncoded MIMO system with 4-QAM. This corresponds to a lattice decoding scenario with restricted state space in dimension $n = 32$. As a comparison, lattice-reduction-aided Gibbs sampler with different choices of standard deviation σ (i.e., $\delta_{\text{statistic}}$, δ_{hassibi} and δ_{distance}) are illustrated under the same Markov moves (i.e., $t = 50$). Here, the choice σ_w is contained in $\sigma_{\text{statistic}}$, and the choice δ_{hassibi} comes from [24] as

$$\sigma^2_{\text{hassibi}} = \frac{\text{SNR}}{\ln n} + \sqrt{\left(\frac{\text{SNR}}{\ln n}\right)^2 - 2\frac{\text{SNR}}{\ln n}}, \tag{61}$$

which still belongs to a statistic solution of σ. Clearly, in both lattice-reduction-aided Gibbs samplers based on ZF and SIC, the decoding performance with δ_{distance} given in (50) are the best. This confirms our analysis as it fully takes advantage of each specific decoding while the choice $\delta_{\text{statistic}}$ or δ_{hassibi} only offers a general solution by statistics.

Similarly, the same observations can also be found in Fig. 7, which shows the BER performance in a 16 × 16 uncoded MIMO

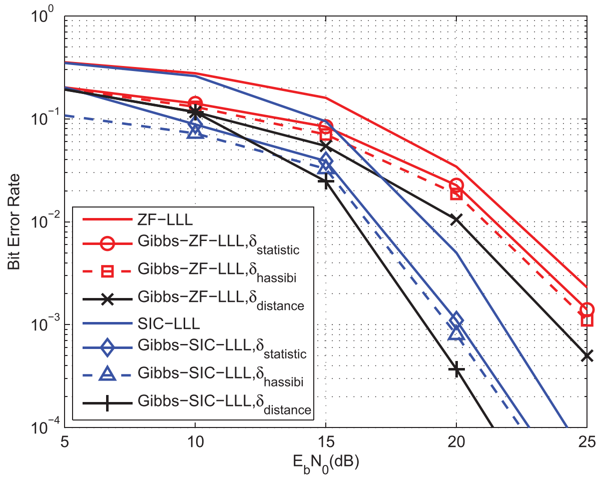

Fig. 7. Bit error rate versus average SNR per bit for the uncoded 16 × 16 MIMO system using 16-QAM.

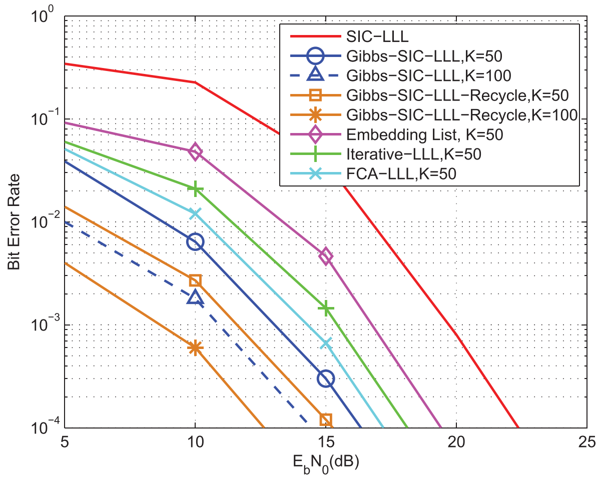

Fig. 8. Bit error rate versus average SNR per bit for the uncoded 24 × 24 MIMO system using 4-QAM.

system using 16-QAM. Intuitively, compared to the basic ZF-LLL or SIC-LLL detection, considerable performance gain is obtained by the lattice-reduction-aided Gibbs sampling detectors with $t = 50$. As expected, the choice of σ_{distance} performs the best due to its advantages of customized strategy, and the solution σ_{hassibi} is slightly better than $\sigma_{\text{statistic}}$. In addition, for the sake of stalling problem, we point out the importance of the lower bound $\sigma \geq 1/\sqrt{2\pi}$ contained in both σ_{distance} and $\sigma_{\text{statistic}}$, which prevents the Markov mixing from getting frozen.

Fig. 8 presents the BER performance comparison between the lattice-reduction-aided Gibbs sampling decoding and other decoding schemes in a 24 × 24 uncoded MIMO system with 4-QAM. Clearly, all the selected decoding schemes achieve the full receive diversity gain while the lattice-reduction-aided SIC detector serves as a basic line. In particular, compared to the embedding list algorithm in [59], fixed candidates algorithm (FCA) in [71] and iterative list decoding in [72] with 50 samples, the proposed lattice-reduction-aided Gibbs sampling detector with δ_{distance} yields a better decoding performance under the same number ($t = 50$) of Markov moves. Meanwhile, with the increase of Markov moves (i.e., t = 100), the decoding performance improves gradually. On the other hand, the BER of the recycling-based lattice-reduction-aided Gibbs sampling detector is also shown, where the decoding performance improves due to the larger size of the qualified candidate list. Accordingly, the decoding performance of recycling Gibbs improves with the increment of Markov moves as well. Note that since a large size of samples is allowed, recycling Gibbs deserves a larger σ for a more dynamic state space searching. To this end, the standard deviation we choose for it is $\sigma = 2$, thus making the collected samples more diverse. As a complement, the corresponding BER performance comparison in a 48 × 48 uncoded MIMO system with 16-QAM is also given in Fig. 9, where further performance gain can be achieved by the recycling operation without extra computational cost.

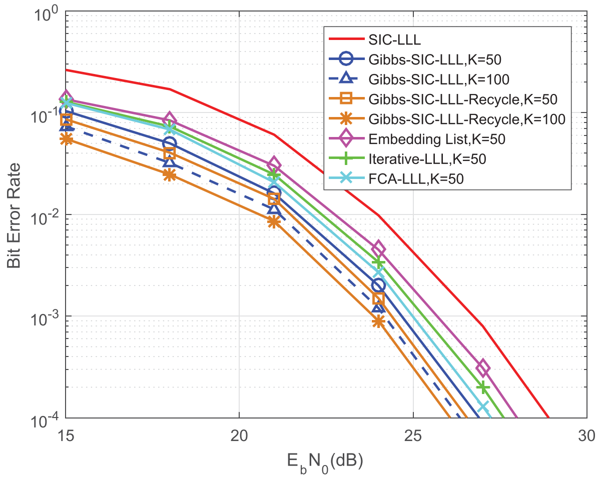

Fig. 9. Bit error rate versus average SNR per bit for the uncoded 48 × 48 MIMO system using 16-QAM.

In Fig. 10, the BER of Gibbs sampling detectors with different coefficients $\alpha \geq 1$ by means of correct decoding radius $R_{\text{lll-sic}}$ are evaluated in a 24 × 24 uncoded MIMO system with 16-QAM. The choice of σ_{distance} is applied and the number of Markov moves is set by $t = 50$. Specifically, given the initial starting point $\mathbf{x}_0 = \mathbf{x}_{\text{lll-sic}}$, $\|\mathbf{H}\mathbf{x}_{\text{sic-lll}} - \mathbf{c}\| \leq \alpha R_{\text{lll-sic}}$ serves as a judgement to decide whether to invoke Gibbs sampling detector or not. As shown in Fig. 10, the decoding performance degrades gradually with the increase of α, where $\alpha = 1$ strictly obeys $\mathbf{x}_{\text{cvp}} = \mathbf{x}_{\text{lll-sic}}$ and $\alpha > 1$ is a loose version of it. Clearly, with a moderate α (experimentally $n/8$), the decoding shows negligible performance loss, but saving considerable computational complexity from it. To be more precisely, the percentage of the direct decoding finished by the initial starting point, i.e., $\mathbf{x}_{\text{output}} = \mathbf{x}_{\text{sic-lll}}$, is depicted in Fig. 11.

Here, we highlight two salient features observed from Fig. 11. On one hand, with the increase of E_b/N_0, the noises are suppressed gradually, which significantly improves the quality of the output from SIC-LLL. In other words, more and more initial

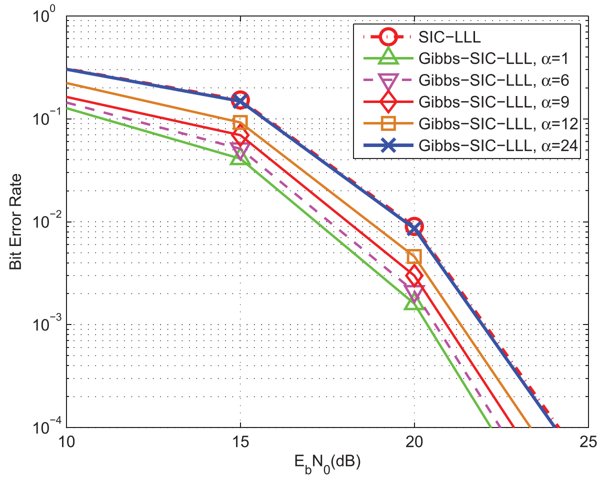

Fig. 10. Bit error rate versus average SNR per bit for the uncoded 24 × 24 MIMO system using 16-QAM.

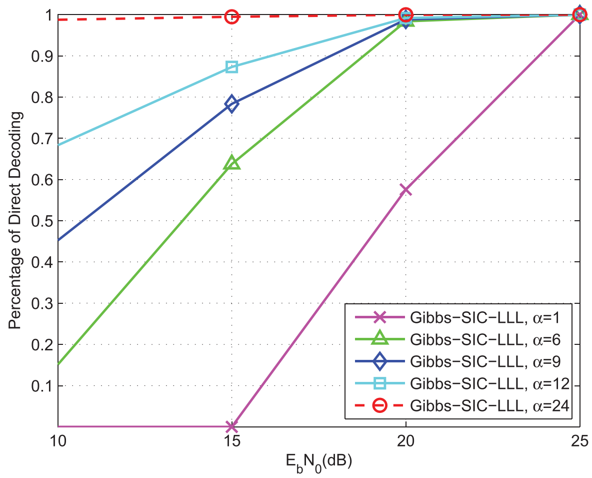

Fig. 11. Percentage of direct decoding versus average SNR per bit for the uncoded 24 × 24 MIMO system using 16-QAM.

starting points are eligible to be directly outputted along with diminishing noises. Hence, from the perspective of each specific decoding point, the demand for Gibbs sampling detector should decrease along with SNR, which emphasizes the significance of the proposed startup mechanism by removing amounts of unnecessary sampling operations. On the other hand, with the increase of α, the startup judgement becomes loose while more initial starting points are allowed to output. This naturally leads to inevitable performance degradation. However, as shown in Fig. 10, a moderate choice of α (e.g. $\alpha = n/8$) still could achieve a promising trade-off between performance loss and complexity reduction.

Following the same scenario in Fig. 10 and Fig. 11, for a better comparison to illustrate the computational cost, Fig. 12 is given to show the computational efficiency of the proposed startup mechanism. More specifically, the average elapsed running times per decoding case by lattice-reduction-aided Gibbs sampling detectors with different $\alpha \geq 1$ are drawn, which actually correspond to the invoked detection complexities respectively. The simulation is conducted by MATLAB R2016a on a single computer, with an Intel Core i7 processor at 2.7 GHz, a RAM of 8 GB and Windows 10 Enterprise Service Pack operating system. Here, the operation of LLL reduction (with polynomial complexity $O(n^3 \log n)$) is applied as a preprocessing, which is not taken into account by the running times. As for the comparison, the running times of SIC-LLL, Gibbs sampling per each Markov move (approximately 2.4×10^{-3}s) as well as standard lattice-reduction-aided Gibbs sampling detector are presented as the baselines. Clearly, the case of $\alpha = 1$ has the same decoding performance with the standard lattice-reduction-aided Gibbs sampling detector. However, under the help of the proposed startup mechanism, considerable running time can be saved with the increment of SNR, which is very important in practise. Meanwhile, with the increase of α, the judgement of the startup mechanism becomes looser and looser, resulting in performance degradation and complexity reduction.

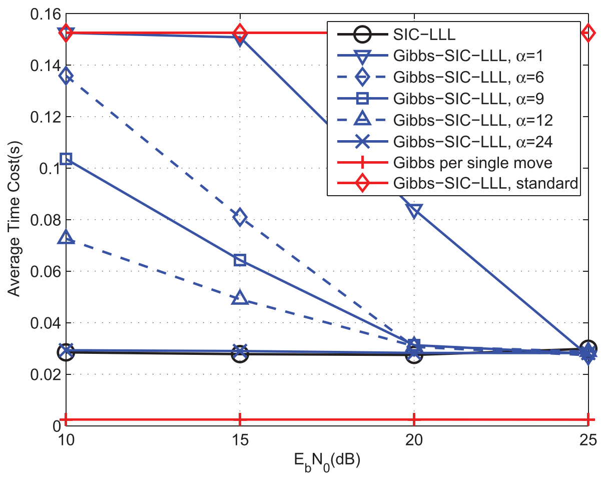

Fig. 12. Average time cost versus average SNR per bit for the uncoded 24 × 24 MIMO system using 16-QAM.

VI. Conclusion

In this paper, lattice reduction was introduced to Gibbs sampler for lattice Gaussian sampling. The convergence rate of systematic scan Gibbs sampling was investigated in full details. As demonstrated, the HGR maximal correlation over elements in lattice Gaussian sampling plays an indispensable role in the Markov mixing. Therefore, lattice reduction is applied to establish an equivalent lattice Gaussian distribution, where Gibbs sampling can be carried out with a better convergence performance. After that, we show that the proposed lattice-reduction-aided Gibbs sampling can be easily used as a sampler decoding scheme to solve the CVP. To balance the trade-off between Markov mixing and sampler decoding, the choice of the standard deviation during the sampling was studied, and a suboptimal σ based on the initial starting point was given.

Meanwhile, the recycling operation was adopted to lattice decoding in MIMO systems for the further performance improvement. Finally, to pursuit an efficient sampler decoding, a startup mechanism resorting to correct decoding radius from BDD was proposed.

Acknowledgment

The authors would like to thank Dr. C. Ling (Imperial College London, London, U.K.) for his helpful discussions and insightful suggestions.

References

[1] W. Banaszczyk, "New bounds in some transference theorems in the geometry of numbers," *Math. Ann.*, vol. 296, pp. 625–635, 1993.

[2] G. Forney and L.-F. Wei, "Multidimensional constellations–Part II: Voronoi constellations," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 941–958, Aug. 1989.

[3] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

[4] L. Liu and C. Ling, "Polar codes and polar lattices for independent fading channels," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 4923–4935, Dec. 2016.

[5] C. Ling and J.-C. Belfiore, "Achieiving the AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.

[6] L. Liu, Y. Yan, and C. Ling, "Achieving secrecy capacity of the Gaussian wiretap channel with polar lattices," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1647–1665, Mar. 2018.

[7] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, "Semantically secure lattice codes for the Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.

[8] H. Mirghasemi and J. C. Belfiore, "Lattice code design criterion for MIMO wiretap channels," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2015, pp. 277–281.

[9] S. Vatedka, N. Kashyap, and A. Thangaraj, "Secure compute-and-forward in a bidirectional relay," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2531–2556, May 2015.

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

[11] O. Regev, "On lattice, learning with errors, random linear codes, and cryptography," *J. ACM*, vol. 56, no. 6, pp. 34:1–34:40, 2009.

[12] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, , Stanford University, Stanford, CA, USA, 2009.

[13] C. Gentry, C. Peikert, and V. Vaikuntanathan, "Trapdoors for hard lattices and new cryptographic constructions," in *Proc. 40th Ann. ACM Symp. Theory Comput.*, Victoria, BC, Canada, 2008, pp. 197–206.

[14] D. Aggarwal, D. Dadush, O. Regev, and N. Stephens-Davidowitz, "Solving the shortest vector problem in 2^n time via discrete Gaussian sampling," in *Proc. Symp. Theory Comput.*, 2015, pp. 1–41.

[15] D. Aggarwal, D. Dadush, and N. Stephens-Davidowitz, "Solving the closest vector problem in 2^n time — the discrete Gaussian strike again!" in *Proc. Found. Comput. Sci.*, 2015, pp. 1–25.

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

[17] A. Campello and J.-C. Belfiore, "Sampling algorithms for lattice Gaussian codes," in *Proc. Int. Zurich Semin. Commun.*, Zurich, Switzerland, 2016, pp. 165–169.

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

[19] Z. Wang and C. Ling, "Lattice Gaussian sampling by Markov chain Monte Carlo: Bounded distance decoding and trapdoor sampling," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3630–3645, Jun. 2019.

[20] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Time*. Providence, RI, USA: Amer. Math. Soc., 2008.

[21] Z. Wang, C. Ling, and G. Hanrot, "Markov chain Monte Carlo algorithms for lattice Gaussian sampling," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 1489–1493.

[22] Z. Wang and C. Ling, "Symmetric Metropolis-within-Gibbs algorithm for lattice Gaussian sampling," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2016, pp. 394–398.

[23] Z. Wang and C. Ling, "On the geometric ergodicity of Gibbs algorithm for lattice Gaussian sampling," in *Proc. IEEE Inf. Theory Workshop*, 2017, pp. 269–273.

[24] 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 Trans. Signal Process.*, vol. 62, no. 17, pp. 4436–4450, Sep. 2014.

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

[26] P. Aggarwal and X. Wang, "Multilevel sequential Monte Carlo algorithms for MIMO demodulation," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 750–758, Feb. 2007.

[27] H. Zhu, B. Farhang-Boroujeny, and R.-R. Chen, "On performance of sphere decoding and Markov chain Monte Carlo detection methods," *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 669–672, Oct. 2005.

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

[29] B. Farhang-Boroujeny, H. Zhu, and Z. Shi, "Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1896–1909, May 2006.

[30] L. Bai, T. Li, J. Liu, Q. Yu, and J. Choi, "Large-scale MIMO detection using MCMC approach with blockwise sampling," *IEEE Trans. Commun.*, vol. 64, no. 29, pp. 3697–3707, Sep. 2016.

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

[32] J. C. Hedstrom, C. H. Yuen, R. Chen, and B. Farhang-Boroujeny, "Achieving near MAP performance with an excited Markov chain Monte Carlo MIMO detector," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7718–7732, Dec. 2017.

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

[34] K. Luo and A. Manikas, "Joint transmitter-receiver optimization in multitarget MIMO radar," *IEEE Trans. Signal Process.*, vol. 65, no. 23, pp. 6292–6302, Dec. 2017.

[35] H. Cheng, Y. Xia, Y. Huang, L. Yang, and D. P. Mandic, "A normalized complex LMS based blind I/Q imbalance compensator for GFDM receivers and its full second-order performance analysis," *IEEE Trans. Signal Process.*, vol. 66, no. 17, pp. 4701–4712, Sep. 2018.

[36] Q. Wu, G. Ding, J. Wang, and Y. Yao, "Spatial-temporal opportunity detection for spectrum-heterogeneous cognitive radio networks: Two-dimensional sensing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 516–526, Feb. 2013.

[37] J. Zhuang, H. Xiong, W. Wang, and Z. Chen, "Application of manifold separation to parametric localization for incoherently distributed sources," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2849–2860, Jun. 2018.

[38] M. Xiang, B. S. Dees, and D. P. Mandic, "Multiple-model adaptive estimation for 3-D and 4-D signals: A widely linear quaternion approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 72–84, Jan. 2019.

[39] S. Liu, C. Ling, and D. Stehlé, "Decoding by sampling: A randomized lattice algorithm for bounded distance decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5933–5945, Sep. 2011.

[40] Y. Xia and D. P. Mandic, "Augmented performance bounds on strictly linear and widely linear estimators with complex data," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 507–514, Jan. 2018.

[41] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2401, Oct. 2003.

[42] N. Stephens-Davidowitz, "Discrete Gaussian sampling reduces to CVP and SVP," in *Proc. 27th Annu. ACM-SIAM Symp. discrete algorithms (SODA)*, Arlington, Virginia, Jan. 2016, pp. 1748–1764.

[43] D. Wubben, D. Seethaler, J. Jalden, and G. Matz, "Lattice reduction," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 70–91, May 2011.

[44] S. Lyu and C. Ling, "Boosted KZ and LLL algorithms," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4784–4796, Sep. 2017.

[45] A. K. Lenstra, H. W. Lenstra, and L. Lovasz, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, pp. 515–534, 1982.

[46] M. Taherzadeh, A. Mobasher, and A. Khandani, "LLL reduction achieves the receive diversity in MIMO decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4801–4805, Dec. 2007.

[47] J. Jalden and P. Elia, "LR-aided MMSE lattice decoding is DMT optimal for all approximately universal codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jul. 2009, pp. 1263–1267.

[48] C. Ling and N. Howgrave-Graham, "Effective LLL reduction for lattice decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 2007, pp. 196–200.

[49] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[50] B. He, C. D. Sa, I. Mitliagkas, and C. Re, "Scan order in Gibbs sampling: Models in which it matters and bounds on how much," in *Proc. Neural Inf. Process. Syst.*, 2016, pp. 1–9.

[51] K. Yosida, *Functional Analysis*, 6th ed. New York, NY, USA: Springer-Verlag, 1980.

[52] J. S. Liu and W. H. Wong, "Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes," *Biometrika*, vol. 81, no. 1, pp. 27–40, 1995.

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

[54] J. S. Liu, W. H. Wong, and A. Kong, "Covariance structure and convergence rate of the Gibbs sampler with various scans," *J. Roy. Statist. Soc. Ser. B*, vol. 57, no. 1, pp. 157–169, 1995.

[55] A. Rényi, "On measures of dependence," *Acta Math. Hung.*, vol. 10, pp. 441–451, 1959.

[56] I. Kontoyannis and S. P. Meyn, "Geometric ergodicity and spectral gap of non-reversible real valued Markov chains," in *Proc. Probab. Theory Related Fields*, vol. 154, 2012, pp. 327–339.

[57] C. J. Geyer and E. A. Thompson, "Annealing Markov chain Monte Carlo with applications to ancestral inference," *J. Amer. Statist. Assoc.*, vol. 90, pp. 909–920, 1995.

[58] J. S. Rosenthal, "Minorization conditions and convergence rates for Markov chain Monte Carlo," *J. Amer. Statist. Assoc.*, vol. 90, pp. 558–566, 1995.

[59] L. Luzzi, D. Stehlé, and C. Ling, "Decoding by embedding: Correct decoding radius and DMT optimality," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2960–2973, May 2013.

[60] L. Martino, V. Elvira, and G. Camps-Valls, "The recycling Gibbs sampler for efficient learning," *Digit. Signal Process.*, vol. 74, pp. 1–13, Mar. 2018.

[61] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-best sphere decoding for MIMO detection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 491–503, Mar. 2006.

[62] K. Latuszynski, G. O. Roberts, and J. S. Rosenthal, "Adaptive Gibbs samplers and related MCMC methods," *Ann. Appl. Probab.*, vol. 23, no. 1, pp. 66–98, Feb. 2013.

[63] L. Martino, J. Read, and D. Luengo, "Independent doubly adaptive rejection Metropolis sampling within Gibbs sampling," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3123–3138, Jun. 2015.

[64] L. Martino, R. Casarin, F. Leisen, and D. Luengo, "Adaptive independent sticky MCMC algorithms," *EURASIP J. Adv. Signal Process.*, vol. 5, pp. 1–28, 2018.

[65] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[66] J. W. Choi, B. Shim, and A. C. Singer, "Efficient soft-input soft-output tree detection via an improved path metric," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1518–1533, Mar. 2012.

[67] J. Céspedes, P. M. Olmos, M. Sánchez-Fernndez, and F. Perez-Cruz, "Probabilistic MIMO symbol detection with expectation consistency approximate inference," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3481–3494, Apr. 2018.

[68] J. Céspedes, P. M. Olmos, M. Sánchez-Fernndez, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional mimo systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, Aug. 2014.

[69] C. Studer, D. Seethaler, and H. Bolcskei, "Finite lattice-size effects in MIMO detection," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, vol. 154, Oct. 2008, pp. 2032–2037.

[70] M. Taherzadeh and A. K. Khandani, "On the limitations of the naive lattice decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4820–4826, Oct. 2010.

[71] W. Zhang and X. Ma, "Low-complexity soft-output decoding with lattice-reduction-aided detectors," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2621–2629, Sep. 2010.

[72] T. Shimokawa and T. Fujino, "Iterative lattice reduction aided MMSE list detection in MIMO system," in *Proc. IEEE Int. Conf. Adv. Technol. Commun.*, Oct. 2008, pp. 50–54.

Zheng Wang (M'16) received the B.S. degree in electronic and information engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2009, the M.S. degree in communications from the Department of Electrical and Electronic Engineering, University of Manchester, Manchester, U.K., in 2010, and the Ph.D degree in communication engineering from Imperial College London, London, U.K., in 2015.

From 2015 to 2016, he served as a Research Associate with Imperial College London. From 2016 to 2017, he was an Senior Engineer with Radio Access Network R&D division, Huawei TechnologiesCompany. He is currently an Assistant Professor with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His current research interests include lattice methods for wireless communications, MIMO systems, machine learning, and physical layer security.

Yang Huang (M'17) received the B.S. and M.S. degrees from Northeastern University, Shenyang, China, in 2011 and 2013, respectively, and the Ph.D. degree from Imperial College London, London, U.K., in 2017. He is currently an Associate Professor with the Department of Information and Communication Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include wireless communications, MIMO systems, convex optimization, machine learning, signal processing for communications, 5G networks, the Internet of Things, wireless power transfer, and wireless powered communications.

Shanxiang Lyu received the B.Eng. and M.Eng. degrees in electronic and information engineering from South China University of Technology, Guangzhou, China, in 2011 and 2014, respectively, and the Ph.D. degree from the Electrical and Electronic Engineering Department, Imperial College London, London, U.K., in 2018. He is currently a Lecturer with the College of Cyber Security, Jinan University, Guangzhou, China. His research interests are in lattice theory, algebraic number theory, and their applications.