SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2019, Vol. 62 042302:1–042302:13
https://doi.org/10.1007/s11432-018-9637-5

Lattice reduction aided belief propagation for massive MIMO detection

Senjie ZHANG[1*], Zhiqiang HE[1], Kai NIU[1], Shi JIN[2] & Hong CHENG[3]

[1]*Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing* 100876*, China;*
[2]*National Mobile Communications Research Laboratory, Southeast University, Nanjing* 210096*, China;*
[3]*Intel Labs China, Beijing* 100190*, China*

Received 4 September 2018/Revised 6 October 2018/Accepted 12 November 2018/Published online 29 December 2018

Abstract Efficient massive MIMO detection for practical deployment, which is with spatially correlated channel and high-order modulation, is a challenging topic for the fifth generation mobile communication (5G). In this paper, lattice reduction aided belief propagation (LRA-BP) is proposed for massive MIMO detection. LRA-BP applies the message updating rules of Markov random field based belief propagation (MRF-BP) in lattice reduced MIMO system. With the lattice reduced, well-conditioned MIMO channel, LRA-BP obtains better message updating and detection performance in spatially correlated channel than MRF-BP. Log-domain arithmetic is used in LRA-BP for computational complexity reduction. Simulation result shows that LRA-BP outperforms MRF-BP with 3–10 dB in terms of required SNR for 1% packet error rate in spatially correlated channel for 256-QAM. We also show that LRA-BP requires much lower complexity compared with MRF-BP.

Keywords massive MIMO, MIMO detection, belief propagation, graph-based detection, lattice reduction, Markov random field

Citation Zhang S J, He Z Q, Niu K, et al. Lattice reduction aided belief propagation for massive MIMO detection. Sci China Inf Sci, 2019, 62(4): 042302, https://doi.org/10.1007/s11432-018-9637-5

1 Introduction

Massive multiple-input and multiple-output (massive MIMO or large-scale MIMO) is a key technology for the fifth generation mobile communication (5G) [1–3]. The efficient detections in massive MIMO draw lots of attention [4]. Classical algorithms for MIMO detection include minimum mean-square error (MMSE) and sphere decoding (SD) [5, 6]. Another category of MIMO detection is graph based detection [7–11]. Graph based MIMO detection is with high parallelism. It benefits hardware implementation for massive MIMO detection with low latency which is critical to 5G. Also it provides the possibility for a uniform processing architecture for MIMO detection and channel decoding.

In graph based detection, MIMO system is modeled as a fully connected graph. Two types of graphical MIMO model are proposed: bipartite graph and Markov random field (MRF). In bipartite graph, observation nodes, variable nodes and the edges represent the received signals, hidden data symbols to be detected and MIMO channel. In MRF, observation nodes are embodied in the edges which describe local dependencies among variable nodes (data symbols). Graph based detection relies on the belief propagation (BP) algorithm [12] or the sum-product algorithm [13] since they are efficient tools in solving inference problems in probabilistic graphical models. These algorithms are used widely in channel

* Corresponding author (email: senjie.zhang@bupt.edu.cn)

decoding, such as the turbo codes and low density parity check codes. They has also been extensively studied for the MIMO detection with bipartite and pairwise graphical MIMO model.

As a direct migation from channel decoding, bipartite graph based BP is used for MIMO detection with complexity reduction based on edge pruning [7] and Gaussian assumption [8, 9, 14]. In Gaussian BP, the input data and messages are all assumed to be Gaussian so that the message and posterior probability can be represented by probabilistic mean and variance, resulting in a very simple message update rule. But as Ref. [10] shown, bipartite graph based Gaussian BP converges to MMSE solution only. Also it does not work well for non-Gaussian input, particularly for high-order modulation. Belief scaling is considered in [15, 16] to improve performance for high-order modulation but only 16-QAM is verified. MRF based belief propagation (MRF-BP) can be used to detect high-order modulation [10, 11, 17]. In MRF-BP, the conditional a posteriori probability under Gaussian input assumption is used to approximate the marginal probability density function (PDF) of non-Gaussian data. The messages exchanged between variable nodes are not treated as Gaussian. In MRF-BP, the message is obtained with conditional MMSE estimator [11] which works well for independent identically distributed (i.i.d.) Rayleigh channel. In spatially correlated channel, the performance of conditional MMSE degrades due to the noise amplification effect. Consequently the performance of MRF-BP degrades. Unfortunately many practical massive MIMO channel is spatially correlated due to practical limitations like form factor. As simulation result shown later, the performance gain of MRF-BP over MMSE reduces to be negligible in 3GPP 3D channel model [18, 19].

In this paper, we propose lattice reduction aided belief propagation (LRA-BP) for efficient massive MIMO detection for practical deployment, which is with spatially correlated channel and high-order modulation. LRA-BP utilizes lattice reduction to improve the performance in spatially correlated channel. Lattice reduction [20] is a powerful concept for solving diverse problems involving point lattices. With lattice reduction, spatially correlated channel is transformed to a more orthogonal channel which assists MIMO detection. Better detection performance by combining lattice reduction with MMSE and SD are reported in [21, 22]. Its application in massive MIMO has been proposed [23]. With the transformed channel, LRA-BP has better message updating and consequent MIMO detection performance in spatially correlated channel. To support high-order modulation, LRA-BP leverages the message exchanging and updating of MRF-BP instead of bipartite graph based BP. We also applies log-domain arithmetic in LRA-BP for computational complexity reduction. Simulation shows that LRA-BP outperforms MRF-BP with 3–10 dB in terms of required SNR for 1% packet error rate (PER) in spatially correlated channel for 256-QAM. Also LRA-BP outperforms MRF-BP in independent channel.

This paper is organized as follows. In Section 2, we briefly review the system model and MRF-BP. In Section 3, LRA-BP is proposed. The performance is evaluated and compared via link-level simulations in Section 4, and the computational complexity is analyzed in Section 5. Finally, in Section 6, the concluding remarks are given.

2 System model and MRF-BP

MIMO system model. A MIMO system with N_r receiving antennas and N_t transmitting antennas is modeled as

$$\boldsymbol{r} = \boldsymbol{G} \cdot \boldsymbol{s} + \boldsymbol{\omega}, \tag{1}$$

where $\boldsymbol{r}$ is an $N_r \times 1$ complex vector for received signals, $\boldsymbol{G}$ is an $N_r \times N_t$ complex matrix for channel coefficients, $\boldsymbol{s}$ is an $N_t \times 1$ complex vector for transmitted data symbols, $\boldsymbol{\omega}$ is an $N_r \times 1$ complex vector for noise. The noise vector $\boldsymbol{\omega}$ is complex Gaussian with mean 0 and covariance $\mathrm{E}[\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{H}}] = \sigma^2 \boldsymbol{I}$, where $\mathrm{E}[\cdot]$ denotes expectation. In practice, each element of $\boldsymbol{s}$ is a constellation point drawn from a finite constellation $\boldsymbol{\Omega}_{\boldsymbol{s}}$ of size 2^{Q_m} such as quadrature phase-shift keying (QPSK) and 256-QAM, for which $Q_m = 2$ and 8, respectively.

To facilitate computation and reduce complexity, we use an equivalent real-domain system model

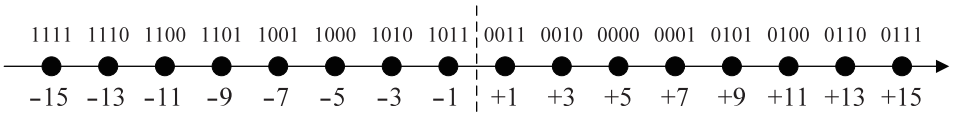

Figure 1 Real-domain constellation for 256-QAM.

derived from real value decomposition (RVD) [24]. The real-domain system model can be expressed as

$$\boldsymbol{y} = \boldsymbol{H} \cdot \boldsymbol{x} + \boldsymbol{n} = \sum_{k=1}^{2N_t} \boldsymbol{h}_k x_k + \boldsymbol{n}, \tag{2}$$

where

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \cdots & \boldsymbol{h}_{2N_t} \end{bmatrix} = \begin{bmatrix} \mathcal{R}(\boldsymbol{G}) & -\mathcal{I}(\boldsymbol{G}) \\ \mathcal{I}(\boldsymbol{G}) & \mathcal{R}(\boldsymbol{G}) \end{bmatrix},$$

$$\boldsymbol{y} = \begin{bmatrix} \mathcal{R}(\boldsymbol{r}) \\ \mathcal{I}(\boldsymbol{r}) \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} \mathcal{R}(\boldsymbol{s}) \\ \mathcal{I}(\boldsymbol{s}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{n} = \begin{bmatrix} \mathcal{R}(\boldsymbol{\omega}) \\ \mathcal{I}(\boldsymbol{\omega}) \end{bmatrix}$$

is the Gaussian noise vector with mean 0 and covariance $0.5\sigma^2\boldsymbol{I}$. Here $\mathcal{R}(\cdot)$ takes the real part, and $\mathcal{I}(\cdot)$ takes the imaginary part. Although the system's dimension doubles, the constellation size reduces by a factor of $2^{-Q_m/2}$ in real-domain system model. Since belief propagation takes each constellation point into account, it benefits a lot in terms of complexity.

The element of $\boldsymbol{x}$, noted as x_k, is drawn from a finite real-domain constellation $\boldsymbol{\Omega}_{\boldsymbol{x}}$ of size $2^{Q_m/2}$. For 256-QAM, $\boldsymbol{\Omega}_{\boldsymbol{x}} = \{\pm1, \pm3, \pm5, \pm7, \pm9, \pm11, \pm13, \pm15\}/\sqrt{170}$. With the definition of constellation and inter-stream independency, the Gaussian assumption on $\boldsymbol{x}$ can be formulated as $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{I})$, where $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ representing a Gaussian PDF of mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ defined as

$$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

The Gray mapping function between constellation point and its bits, noted as $x_k(b_1, b_2, \ldots, b_j, \ldots, b_{Q_m/2})$, can be found in [25]. Figure 1 shows the mapping function for 256-QAM.

Considering jth bit (b_j), the finite constellation $\boldsymbol{\Omega}_{\boldsymbol{x}}$ can be divided into two subsets, $\boldsymbol{\Omega}_{\boldsymbol{x}}^{j0}$ and $\boldsymbol{\Omega}_{\boldsymbol{x}}^{j1}$, where $\boldsymbol{\Omega}_{\boldsymbol{x}}^{jb} = \{x_k(b_1, b_2, \ldots, b_j = b, \ldots, b_{Q_m/2})\}, b = 0$ or 1, represents all constellation points corresponding to $b_j = b$ respectively. For channel decoding, log-likelihood ratio (LLR) for informatoin bits are desired. The LLR of $b_{k,j}$, the jth bit carried by kth X-axis transmitted symbol x_k, can be computed as

$$\mathrm{LLR}(b_{k,j}) = \log\frac{P(b_{k,j}=0)}{P(b_{k,j}=1)} = \log\frac{\sum_{x_k \in \boldsymbol{\Omega}_{\boldsymbol{x}}^{j0}} P(x_k)}{\sum_{x_k \in \boldsymbol{\Omega}_{\boldsymbol{x}}^{j1}} P(x_k)}, \tag{3}$$

where $P(x_k)$ is the the belief (probability) of x_k. With lattice reduction, the system model can be reformulated as

$$\boldsymbol{y} = \boldsymbol{H} \cdot \boldsymbol{x} + \boldsymbol{n} = \tilde{\boldsymbol{H}} \cdot \boldsymbol{z} + \boldsymbol{n} = \sum_{k=1}^{2N_t} \tilde{\boldsymbol{h}}_k z_k + \boldsymbol{n}, \tag{4}$$

where $\tilde{\boldsymbol{H}} = \boldsymbol{H} \cdot \boldsymbol{T}$ and $\boldsymbol{z} = \boldsymbol{T}^{-1}\boldsymbol{x}$. Here $\boldsymbol{T}$ and $\boldsymbol{T}^{-1}$ are both $2N_t \times 2N_t$ integer unimodular matrix. The most commonly used algorithm to obtain $\tilde{\boldsymbol{H}}$ and $\boldsymbol{T}$ is the Lenstra-Lenstra-Lovász (LLL) reduction algorithm [26]. With lattice reduction, the new channel matrix $\tilde{\boldsymbol{H}}$ is more orthogonal than $\boldsymbol{H}$. Since $\boldsymbol{z} = \boldsymbol{T}^{-1}\boldsymbol{x}$, we have $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{C})$, where $\boldsymbol{C} = [c_{ij}] = 0.5 \cdot \boldsymbol{T}^{-1} \cdot \boldsymbol{T}^{-\mathrm{T}}$. So the Gaussian assumption on $\boldsymbol{z}$ remains and $p(z_k) = \mathcal{N}(z_k; 0, c_{kk}) = \mathcal{N}(z_k; 0, \sigma_k^2)$. However the inter-stream independency is no longer valid. The conditional PDF of z_j given z_i can be obtained as [27]

$$p(z_j|z_i) = \mathcal{N}\left(z_j; \rho\frac{\sigma_j}{\sigma_i} z_i, (1-\rho^2)\sigma_j^2\right), \tag{5}$$

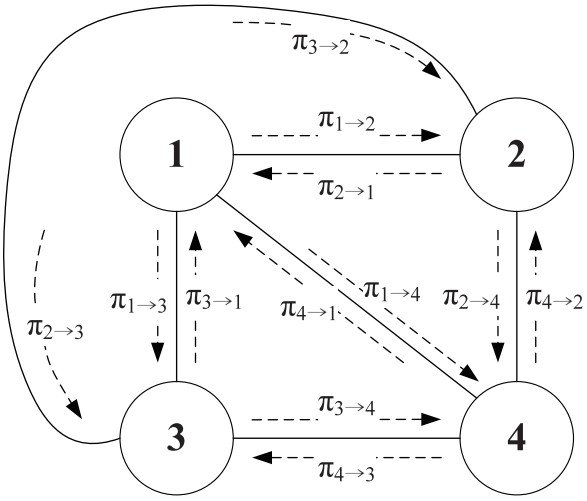

Figure 2 Pairwise graph for MIMO detection.

where the correlation coefficient $\rho = c_{ij}/(\sigma_j\sigma_i)$.

With real-domain system model, the pairwise MRF graphical MIMO model with $2N_t$ observation nodes can be obtained. The observation node stands for a transmitted data symbol and connects each other in pair. For a compact expression, we denote the edge connecting nodes i and j as $e(i,j)$ and the set of neighbors of the jth node as $V(j)$, where both i and j belong to the integer set $[1,\ldots,2N_t]$ and $i \neq j$. Figure 2 shows the MRF for a MIMO system with four observation nodes ($N_t = 2$).

In MRF-BP, the ith observation node stands for x_i in (2) and the message from ith node to jth node (noted as $\pi_{i\to j}$) contains the conditional probability of x_j (for each constellation points) given x_i's information. LRA-BP is based on the same pairwise MRF graphical MIMO model. In LRA-BP, the ith observation node stands for z_i in (4) and $\pi_{i\to j}$ contains the conditional probability of z_j (for each Z-axis constellation points) given z_i's information.

Overview of MRF-BP. MRF-BP exchanges and updates messages in each edge bidirectionally. The messgae updating rules is [11]

$$\pi_{i\to j}^{(t)}(x_j) = \alpha \sum_{x_i\in\boldsymbol{\Omega}_{\boldsymbol{x}}} \left\{ p\big(x_j\big|x_i,\boldsymbol{y}\big) \prod_{k\in V(i)\setminus j} \pi_{k\to i}^{(t-1)}(x_i) \right\}, \quad \forall x_j \in \boldsymbol{\Omega}_{\boldsymbol{x}}, \tag{6}$$

where t is the iteration number, $V(i)\setminus j$ denotes all elements in $V(i)$ except j, $\beta = \boldsymbol{h}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\boldsymbol{h}_j + 0.5^{-2}$, $\boldsymbol{K}_{\{j,i\}} = 0.5\sigma^2\boldsymbol{I} + \sum_{k\neq j,i}\boldsymbol{h}_k\boldsymbol{t}_k\boldsymbol{t}_k^{\mathrm{T}}\boldsymbol{h}_k^{\mathrm{T}}$ and

$$p\big(x_j\big|x_i,\boldsymbol{y}\big) = \sqrt{\frac{\beta}{2\pi}}\exp\left\{\frac{-(\beta x_j - \boldsymbol{h}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\boldsymbol{y} + \boldsymbol{h}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\boldsymbol{h}_i x_i)^2}{2\beta}\right\}. \tag{7}$$

As (6) and (7) shown, MRF-BP's message updadting is based on conditional MMSE estimator [11]. It is contaminated by inter-stream interference and residual noise, which is more severe in spatially correlated channel than independent channel. So MRF-BP's performance degrades in spatially correlated channel.

3 Lattice reduction aided belief propagation

In this section, LRA-BP is proposed for efficient MIMO detection in spatially correlated channel by conducting message updating rules like (7) in lattice reduced system model (4). With lattice reduction, orignal MIMO channel is transformed to a more orthogonal channel and the impact of inter-stream interference and residual noise is weakened. Consequently the message's quality and detection performance are improved, especially for spatially correlated channel. In order to use lattice reduced system model (4),

LRA-BP should determines Z-axis constellation at first. Correspondingly, LRA-BP should revise the message updating rules in (7) according to the lattice reduced channel $\tilde{\boldsymbol{H}}$ and Z-axis transmitted symbol $\boldsymbol{z}$. The LLR generation from $\boldsymbol{z}$ is also desirable. We also apply log-domain arithmetic in LRA-BP for computational complexity reduction.

3.1 Determination of Z-axis constellation

Since $x_k \in \boldsymbol{\Omega}_{\boldsymbol{x}}$ is finite, z_k is also finite. The original (X-axis) real-domain constellation $\boldsymbol{\Omega}_{\boldsymbol{x}}$ is transformed to the Z-axis real-domain constellation $\boldsymbol{\Omega}_{\boldsymbol{z}}^{(k)}$. It should be noted that unlike regular constellation $\boldsymbol{\Omega}_{\boldsymbol{x}}$ which is common for all streams, $\boldsymbol{\Omega}_{\boldsymbol{z}}^{(k)}$ differs for different stream (i.e., different k). To determine the exact constellation points of $\boldsymbol{\Omega}_{\boldsymbol{z}}^{(k)}$, all possible $\boldsymbol{x}$ should be visited, which leads to a prohibitively high complexity since there are $(Q_m/2)^{2N_t}$ possible combinations. Instead of the full Z-axis constellation, we use a sub-optimal subset in Z-axis constellation as $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(k)}$ in LRA-BP. In order to make the constellation point in $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(k)}$ with high probability, $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(k)}$ is centered at the MMSE estimate of z_k. The slicing of z_k in integer set [28] is used for computational complexity reduction.

Considering $\boldsymbol{z}' = c \cdot \boldsymbol{z} + \boldsymbol{v}$, where $\boldsymbol{v} = (1/2) \cdot \boldsymbol{T}^{-1} \cdot [1, 1, \ldots, 1]^{\mathrm{T}}_{2N_t \times 1}$ is the displacement vector and c is the QAM power normalization constant (e.g., $\sqrt{2}/2$ for QPSK and $\sqrt{170}/2$ for 256-QAM), it is clear that $\boldsymbol{z}' \in \mathbb{Z}$ (the set of integers) [28]. Commonly the possible values of z'_k (the kth element of $\boldsymbol{z}'$) spans an interval of continuous integers, noted as $\boldsymbol{\Theta}^{(k)}$. Like $\boldsymbol{\Omega}_{\boldsymbol{z}}^{(k)}$, $\boldsymbol{\Theta}^{(k)}$ differs for different k.

Given $\boldsymbol{z}$'s MMSE estimation

$$\bar{\boldsymbol{z}} = \left(\tilde{\boldsymbol{H}}^{\mathrm{H}} \cdot \tilde{\boldsymbol{H}} + \frac{1}{2}\sigma^2 \cdot \boldsymbol{T}^{\mathrm{H}} \cdot \boldsymbol{T} \right)^{-1} \cdot \tilde{\boldsymbol{H}}^{\mathrm{H}} \cdot \boldsymbol{y}, \tag{8}$$

and $\lfloor \bar{\boldsymbol{z}} \rceil = \mathrm{round}(c \cdot \bar{\boldsymbol{z}} + \boldsymbol{v})$ where $\mathrm{round}(\cdot)$ denotes the operation to take nearest integer, the integer interval $\boldsymbol{\Theta}^{(k)}$ can be determined as

$$\boldsymbol{\Theta}^{(k)} = \Big[-R + \lfloor \bar{z} \rceil_k, \lfloor \bar{z} \rceil_k + R \Big], \tag{9}$$

where $\lfloor \bar{z} \rceil_k$ is the kth element of $\lfloor \bar{\boldsymbol{z}} \rceil$ and R is the configurable radius of the interval. The optimal value of R depends on the post MMSE processing SNR. From simulation we found that a empirical value with same order of magnitude as the size of X-axis constellation is enough.

The subset of Z-axis constellation used in LRA-BP containing the constellat points with high probability in $\boldsymbol{\Omega}_{\boldsymbol{z}}^{(k)}$, can be derived from $\boldsymbol{\Theta}^{(k)}$ as

$$\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(k)} \left\{ z_k \Big| z_k = \frac{1}{c}(\theta - v_k),\ \forall \theta \in \boldsymbol{\Theta}^{(k)} \right\}, \tag{10}$$

where v_k is the kth element of the displacement vector $\boldsymbol{v}$.

3.2 Message updating

Like MRF-BP, LRA-BP exchanges messages of the probability for Z-axis constellation points in each edge bidirectionally. Denoting the message from the ith node to the jth node as $\pi_{i \to j}(z_j)$, its initial value is

$$\pi_{i \to j}^{(0)}(z_j) = \left| \tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(j)} \right|^{-1}. \tag{11}$$

The message updating rule of LRA-BP for any possible pair of (i, j) can be described as

$$\pi_{i \to j}^{(t)}(z_j) = \alpha \sum_{z_i \in \tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(i)}} \left\{ p\big(z_j \big| z_i, \boldsymbol{y}\big) \prod_{k \in V(i) \setminus j} \pi_{k \to i}^{(t-1)}(z_i) \right\}, \quad \forall z_j \in \tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(j)}, \tag{12}$$

where $p(z_j | z_i, \boldsymbol{y})$ is the translation function determined by transformed channel $\tilde{\boldsymbol{H}}$, α is the normalization coefficient and $V(i) \setminus j$ denotes all elements in $V(i)$ except j.

The translation function $p(z_j|z_i,\boldsymbol{y})$ is constant during iterations. It can be precomputed before iteration starts with probability relations as

$$p(z_j|z_i,\boldsymbol{y})=\frac{p(\boldsymbol{y}|z_i,z_j)p(z_i,z_j)}{p(z_i,\boldsymbol{y})}=\frac{p(\boldsymbol{y}|z_i,z_j)p(z_j|z_i)}{p(\boldsymbol{y}|z_i)}. \tag{13}$$

The Gaussian assumption leads to a conditional MMSE estimator for z_j given z_i as [11]

$$y'_{j|i}=\tilde{\boldsymbol{h}}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\boldsymbol{y}=a_{j|i,i}z_i+\sigma_{j|i}^2z_j+n'_{j|i}, \tag{14}$$

with

$$\boldsymbol{K}_{\{j,i\}}=\frac{1}{2}\sigma^2\boldsymbol{I}+\sum_{k\neq j,i}\tilde{\boldsymbol{h}}_k\boldsymbol{t}_k\boldsymbol{t}_k^{\mathrm{T}}\tilde{\boldsymbol{h}}_k^{\mathrm{T}}, \tag{15}$$

$$a_{j|i,i}=\tilde{\boldsymbol{h}}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\tilde{\boldsymbol{h}}_i, \tag{16}$$

$$n'_{j|i}=\tilde{\boldsymbol{h}}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\left(\sum_{k\neq j,i}\tilde{\boldsymbol{h}}_kz_k+\boldsymbol{n}\right), \tag{17}$$

$$\sigma_{j|i}^2=\mathrm{E}\left|n'_{j|i}\right|^2=\tilde{\boldsymbol{h}}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\tilde{\boldsymbol{h}}_j, \tag{18}$$

where $\boldsymbol{c}_{j|i}$ is MMSE filtering vector, $\boldsymbol{K}_{\{j,i\}}$ is the correlation matrix, $\boldsymbol{t}_k$ is the kth row of $\boldsymbol{T}^{-1}$, $n'_{j|i}$ is inter-stream interference and residual noise and $\sigma_{j|i}^2$ is the power of $n'_{j|i}$.

With (14), Eq. (13) can be rewritten as

$$p(z_j|z_i,\boldsymbol{y})=p(z_j|z_i,y'_{j|i}), \tag{19}$$

then the translation function $p(z_j|z_i,\boldsymbol{y})$ can be obtained with following theorem.

Theorem 1.

$$p(z_j|z_i,\boldsymbol{y})=\sqrt{\frac{\beta}{2\pi}}\mathrm{e}^{d_{i\to j}(z_j)}, \tag{20}$$

where

$$\beta=\sigma_{j|i}^2+\sigma_j^{-2}, \tag{21}$$

$$d_{i\to j}(z_j)=-(\beta z_j-y'_{j|i}+a_{j|i,i}z_i)^2/(2\beta). \tag{22}$$

Proof. With Gaussian assumption, and considering ρ is relative small, Eq. (13) can be rewritten as

$$p(z_j|z_i,\boldsymbol{y})=p(z_j|z_i,y'_{j|i})=\frac{p(y'_{j|i}|z_i,z_j)p(z_j|z_i)}{p(y'_{j|i}|z_i)}, \tag{23}$$

where

$$p(y'_{j|i}|z_i,z_j)=\mathcal{N}(y'_{j|i};a_{j|i,i}z_i+\sigma_{j|i}^2z_j,\sigma_{j|i}^2), \tag{24}$$

$$p(z_j|z_i)\approx p(z_j)=\mathcal{N}(z_j;0,\sigma_j^2), \tag{25}$$

$$p(y'_{j|i}|z_i)=\mathcal{N}(y'_{j|i};a_{j|i,i}z_i,\sigma_{j|i}^2+\sigma_{j|i}^4\sigma_j^2). \tag{26}$$

With the properties of Gaussian PDF [11,12,29] as follows:

$$\mathcal{N}(x;\mu,\sigma^2)=\mathcal{N}(\mu;x,\sigma^2)=\mathcal{N}(x-\mu;0,\sigma^2), \tag{27}$$

$$\mathcal{N}(ax+b;\mu,\sigma^2)=\mathcal{N}\left(x;\frac{\mu-b}{a},\frac{\sigma^2}{|a|^2}\right), \tag{28}$$

$$\mathcal{N}(x;\mu_1,\sigma_1^2)\cdot\mathcal{N}(x;\mu_2,\sigma_2^2)=\mathcal{N}\left(x;\frac{\sigma_1^{-2}\mu_1+\sigma_2^{-2}\mu_2}{\sigma_1^{-2}+\sigma_2^{-2}},\frac{1}{\sigma_1^{-2}+\sigma_2^{-2}}\right)\cdot\mathcal{N}(\mu_1;\mu_2,\sigma_1^2+\sigma_2^2), \tag{29}$$

we have

$$
\begin{aligned}
p\big(z_j|z_i, y'_{j|i}\big) &= \frac{\mathcal{N}\big(y'_{j|i}; a_{j|i,i}z_i + \sigma^2_{j|i}z_j, \sigma^2_{j|i}\big) \cdot \mathcal{N}\big(z_j; 0, \sigma^2_j\big)}{\mathcal{N}\big(y'_{j|i}; a_{j|i,i}z_i, \sigma^2_{j|i} + \sigma^4_{j|i}\sigma^2_j\big)} \\
&\overset{(27)(28)}{=\!=} \frac{\mathcal{N}\big(z_j; \sigma^{-2}_{j|i} \cdot (y'_{j|i} - a_{j|i,i}z_i), \sigma^{-2}_{j|i}\big) \cdot \mathcal{N}\big(z_j; 0, \sigma^2_j\big)}{\mathcal{N}\big(y'_{j|i}; a_{j|i,i}z_i, \sigma^2_{j|i} + \sigma^4_{j|i}\sigma^2_j\big)} \\
&\overset{(29)}{=\!=} \frac{\mathcal{N}\big(z_j; \frac{y'_{j|i} - a_{j|i,i}z_i}{\sigma^2_{j|i} + \sigma^{-2}_j}, \frac{1}{\sigma^2_{j|i} + \sigma^{-2}_j}\big) \cdot \mathcal{N}\big(\sigma^{-2}_{j|i} \cdot (y'_{j|i} - a_{j|i,i}z_i); 0, \sigma^{-2}_{j|i} + \sigma^2_j\big)}{\mathcal{N}\big(y'_{j|i}; a_{j|i,i}z_i, \sigma^2_{j|i} + \sigma^4_{j|i}\sigma^2_j\big)} \\
&\overset{(28)}{=\!=} \mathcal{N}\big(z_j; (y'_{j|i} - a_{j|i,i}z_i)/(\sigma^2_{j|i} + \sigma^{-2}_j), 1/(\sigma^2_{j|i} + \sigma^{-2}_j)\big) \\
&\overset{(21)}{=\!=} \mathcal{N}\big(z_j; (y'_{j|i} - a_{j|i,i}z_i)/\beta, 1/\beta\big) \\
&\overset{(28)}{=\!=} \mathcal{N}\big(\beta z_j; (y'_{j|i} - a_{j|i,i}z_i), \beta\big).
\end{aligned} \tag{30}
$$

Remark 1. If $\boldsymbol{T} = \boldsymbol{I}$ and $\sigma^2_j = 0.5$, LRA-BP's message updating rule (12) and (19) is identical to MRF-BP's message updating rule (6) and (7). So MRF-BP is a special case of LRA-BP if lattice reduction is not conducted.

3.3 Log-likelihood ratio generation

After T iterations of message updating rule (12), the probability of the jth Z-axis symbol z_j is given by

$$
P(z_j) = \prod_{k \in V(j)} \pi^{(T)}_{k \to j}(z_j). \tag{31}
$$

The probability of Z-axis constellation points should be converted back to X-axis for LLR generation. Similar to the determination of $\boldsymbol{\Omega}^{(k)}_{\boldsymbol{z}}$, it is with prohibitively high complexity to exactly convert $P(z_j)$ to $P(x_j)$ since all possible $\boldsymbol{x}$ should be visited.

An approximation solution can be obtained based on Gaussian assumption on z_j and x_j for $j \in [1, 2N_t]$. Denoting $\mu^{(z)}_j$ as the mean of z_j and $\mu^{(x)}_j$ as the mean of x_j, we have

$$
\begin{bmatrix} \mu^{(x)}_1 \\ \vdots \\ \mu^{(x)}_{2N_t} \end{bmatrix} = \boldsymbol{T} \times \begin{bmatrix} \mu^{(z)}_1 \\ \vdots \\ \mu^{(z)}_{2N_t} \end{bmatrix}, \tag{32}
$$

where

$$
\mu^{(z)}_j = \mathrm{E}\big(z_j\big) = \sum_{z_j \in \tilde{\boldsymbol{\Omega}}^{(j)}_{\boldsymbol{z}}} P(z_j) z_j.
$$

Similarly $\nu^{(z)}_j$ and $\nu^{(x)}_j$, the covariance of z_j and x_j, is connected with

$$
\begin{bmatrix} \nu^{(x)}_1 \\ \vdots \\ \nu^{(x)}_{2N_t} \end{bmatrix} = \big(\boldsymbol{T} \circ \boldsymbol{T}\big) \times \begin{bmatrix} \nu^{(z)}_1 \\ \vdots \\ \nu^{(z)}_{2N_t} \end{bmatrix}, \tag{33}
$$

where $\nu^{(z)}_j = \mathrm{E}(|z_j - \mu^{(z)}_j|^2)$ and $\boldsymbol{T} \circ \boldsymbol{T}$ stands for Hadamard product of $\boldsymbol{T}$.

With $\mu^{(x)}_j$ and $\nu^{(x)}_j$, the probability of the jth X-axis symbol x_j can be obtained as

$$
P(x_j) = \lambda \cdot \mathcal{N}\left(x_j; \mu^{(x)}_j, \nu^{(x)}_j\right), \tag{34}
$$

where λ is normalization coefficient for $\sum_{x_j \in \boldsymbol{\Omega}_{\boldsymbol{x}}} P(x_j) = 1$. Then Eq. (3) can be used to generate LLR.

3.4 Application of log-domain arithmetic

The translation function $p(z_j|z_i, \boldsymbol{y})$ in (20) requires to compute lots of exponential function. The computation can be avoided using Jacobian logarithm [30].

By transforming the messages between nodes from the probability of constellation points (noted as $\pi_{i\to j}(z_j)$) to the log probability (noted as $m_{i\to j}(z_j)$), where

$$m_{i\to j}(z_j) = \log\big(\pi_{i\to j}(z_j)\big), \tag{35}$$

and applying (20) to (12), the message updating rule can be rewritten as

$$\begin{aligned} \mathrm{e}^{m_{i\to j}^{(t)}(z_j)} &= \alpha \sum_{z_i\in\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(i)}} \left\{ p\big(z_j\big|z_i, \boldsymbol{y}\big) \prod_{k\in V(i)\backslash j} \pi_{k\to i}^{(t-1)}(z_i) \right\} \\ &= \alpha \sum_{z_i\in\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(i)}} \left\{ \sqrt{\frac{\beta}{2\pi}} \mathrm{e}^{d_{i\to j}(z_j)} \prod_{k\in V(i)\backslash j} \mathrm{e}^{m_{k\to i}^{(t-1)}(z_i)} \right\} \\ &= \alpha\sqrt{\frac{\beta}{2\pi}} \sum_{z_i\in\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(i)}} \exp\left\{ d_{i\to j}(z_j) + \sum_{k\in V(i)\backslash j} m_{k\to i}^{(t-1)}(z_i) \right\}. \end{aligned} \tag{36}$$

Jacobian logarithm leads to the following approximation:

$$\log(\mathrm{e}^x + \mathrm{e}^y) = \max\big(x, y\big) + f_c(|x-y|) \approx \max\big(x, y\big). \tag{37}$$

Applying (37) to (36) and discarding common constant, we have the message updating rule in log-domain as

$$m_{i\to j}^{(t)}(z_j) = \underset{z_i\in\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(i)}}{\operatorname{argmax}} \left\{ d_{i\to j}(z_j) + \sum_{k\in V(i)\backslash j} m_{k\to i}^{(t-1)}(z_i) \right\}. \tag{38}$$

Accordingly the message should be initialized as

$$m_{i\to j}^{(0)}(z_j) = \log\left(\left| \boldsymbol{\Omega}_{\boldsymbol{z}}^{(j)} \right|^{-1} \right). \tag{39}$$

At output stage, the log probability of the jth transmitted symbol z_j is given by

$$M(z_j) = \sum_{k\in V(j)} m_{k\to j}^{(T)}(z_j), \tag{40}$$

and the probability is obtained with $P(z_j) = \alpha \mathrm{e}^{M(z_j)}$ (α is the normalization coefficient).

With log-domain arithmetic, the complexity for computing exponential function is reduced from $2N_t(2N_t-1)|\boldsymbol{\Omega}_{\boldsymbol{z}}^{(j)}|^2$ times to $2N_t|\boldsymbol{\Omega}_{\boldsymbol{z}}^{(j)}|$ times. Also the multiplications in (12) are converted into additions in (38). The overall complexity of LRA-BP reduces significantly.

3.5 Summary of LRA-BP

LRA-BP is summarized as Algorithm 1 shown. If lattice reduction is not conducted, i.e., $\boldsymbol{T} = \boldsymbol{I}$ and $\sigma_j^2 = 0.5$, LRA-BP rollbacks to a log-domain implementation of MRF-BP. Log-domain MRF-BP is with much lower complexity than origianl MRF-BP since it converts the multiplications into additions and requries less exponential function.

Algorithm 1 LRA-BP

Initialization:

1: $m_{i\to j}^{(0)}(z_j)=\log(|\boldsymbol{\Omega}_{\boldsymbol{z}}^{(j)}|^{-1})$ for $j=1,2,\ldots,2N_t$.

Pre-processing:

2: $\boldsymbol{K}_{\{j,i\}}=\frac{1}{2}\sigma^2\boldsymbol{I}+\sum_{k\neq j,i}\tilde{\boldsymbol{h}}_k\boldsymbol{t}_k\boldsymbol{t}_k^{\mathrm{T}}\tilde{\boldsymbol{h}}_k^{\mathrm{T}}$; $y'_{j|i}=\tilde{\boldsymbol{h}}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\boldsymbol{y}$; $a_{j|i,i}=\tilde{\boldsymbol{h}}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\tilde{\boldsymbol{h}}_i$; $\sigma_{j|i}^2=\tilde{\boldsymbol{h}}_j^{\mathrm{T}}\boldsymbol{K}_{\{j,i\}}^{-1}\tilde{\boldsymbol{h}}_j$.

3: $p(z_j|z_i,\boldsymbol{y})=\sqrt{\frac{\beta}{2\pi}}\mathrm{e}^{d_{i\to j}(z_j)}$, $\beta=\sigma_{j|i}^2+\sigma_j^{-2}$, $d_{i\to j}(z_j)=-(\beta z_j-y'_{j|i}+a_{j|i,i}z_i)^2/(2\beta)$.

Iteration: For all possible (i,j) pairs repeat message updating as

4: **for** $t=1,2,\ldots,T$ and $\forall z_j\in\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(j)}$ **do**

5: $\quad m_{i\to j}^{(t)}(z_j)=\mathrm{argmax}_{z_i\in\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(i)}}\left\{d_{i\to j}(z_j)+\sum_{k\in V(i)\backslash j}m_{k\to i}^{(t-1)}(z_i)\right\}$.

6: **end for**

Output:

7: $M(z_j)=\sum_{k\in V(j)}m_{k\to j}^{(T)}(z_j),\forall z_j\in\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(j)}$.

8: $P(z_j)=\alpha\mathrm{e}^{M(z_j)}$.

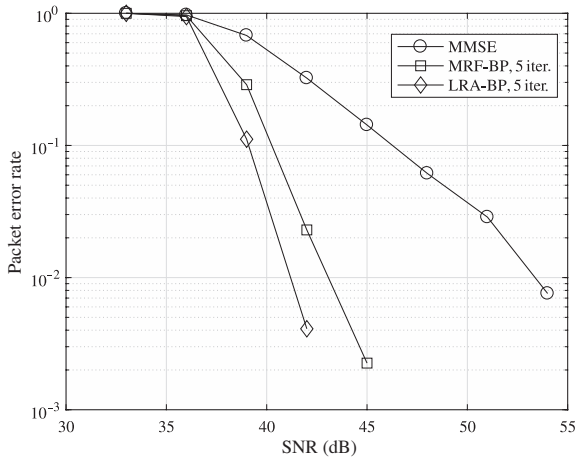

Figure 3 Comparison of receiver performance: 12×12, Rayleigh channel.

4 Simulation results

In this section, the performance of LRA-BP is compared with classical MMSE and MRF-BP.

The performance is measured in terms of packer error rate (PER). To verify the convergence of iteration, high-order modulation (256-QAM according to [25]) is used. The size of regular real-domain constellation for 256-QAM ($|\boldsymbol{\Omega}_{\boldsymbol{x}}|$) is 16. As described in Section 3, the size of Z-axis constellation ($|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(k)}|$) is configurable. In simulation, we use an interval with radius $R=4$ then $|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{z}}^{(k)}|=9$. For channel coding, we use 3GPP LTE Turbo code of rate 3/4 and length 3392 along with 3GPP LTE rate matching (interleaver) [31].

Simulation is conducted by means of Monte Carlo simulations with independent identically distributed (i.i.d.) Rayleigh channel and the 3D channel model proposed by 3GPP [18].

4.1 Results for Rayleigh channel

Rayleigh channel is widely used in performance evaluation for massive MIMO when N_r and N_t are both large. Considered the number of supportable spatial streams in massive MIMO is limited by channel estimation, and the number of 5G pilot sequences is 12, we choose $N_r=N_t=12$. According to [11]'s suggestion, the iteration number of MRF-BP is 5. LRA-BP adopts same iteration number as MRF-BP in this simulation. Figure 3 shows that LRA-BP (5 iterations) outperforms classical MMSE and classical MRF-BP (5 iterations) with 10 dB and 2 dB in terms of required signal-to-noise ratio (SNR) for 1% PER in 12×12 Rayleigh channel. The performance gain over MRF-BP reveals that LRA-BP effectively combines the advantage of lattice reduction and belief propagation.

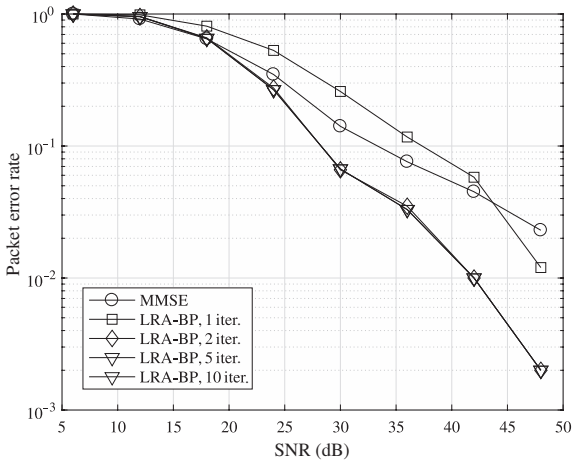

Figure 4 Comparison of iteration number: 32×4, 3GPP 3D channel, SU.

4.2 Results for 3GPP 3D channel

For more practical massive MIMO scenario, where $N_r \gg N_t$, 3GPP 3D channel model is used to evaluate LRA-BP performance in uplink transmission. 3GPP 3D channel model defines a series of deployment scenario. We use urban macro non-light-of-sight scenario (3D-UMa NLOS) in simulation. In 3GPP 3D channel model, a 2D planar antenna array is defined. There are $N_r = M \times N \times P$ antenna elements in the 2D planar array, where N is the number of columns, M is the number of antenna elements with the same polarization in each column and P is the number of polarization. Antenna elements are uniformly placed (0.5λ) in both vertical and horizontal direction.

We use cross-polarized antenna element ($P = 2$). Considered practical issues like form factor limitation and easy to install, Ref. [18] suggests $M \gg N$ and $N = 1$ (single-column array) or $N = 2$ (dual-column array). Accordingly, we choose the value of M to be 16 or 8 to make $N_r = 32$. With these antenna configurations, the generated channel is spatially correlated. Since most of scatterers is close to the ground, single-column 2D planar antenna array leads to higher channel correlation than dual-column array.

In 3GPP 3D channels, linear receivers like MMSE cannot obtain optimal performance as shown in [32] and non-linear receivers like LRA-BP show their performance advantage.

Both single-user MIMO (SU-MIMO) and multiuser MIMO (MU-MIMO) are evaluated.

Iteration number of LRA-BP. As Figure 4 shown, LRA-BP converges with 2 iterations. The first iteration of LRA-BP obtains higher slope in PER curve than MMSE, but suffers SNR loss. This is due to the impact of a known SNR loss phenomenon associated with the symbol slicing consequent to lattice reduction [33]. The second iteration of LRA-BP compensates the SNR loss effectively. No more iteration is necessary. So the following simulations adopt 2 iterations for LRA-BP.

Single-user MIMO. In SU-MIMO, a user equipment (UE) with $N_t = 4$ antennas transmits data to base station with 2D planar antenna array described above. Then a 32×4 MIMO system is established. Figure 5 shows that LRA-BP outperforms MMSE and MRF-BP with 10 dB (single-column array) and 7 dB (dual-column array) in terms of required SNR for 1% PER in SU-MIMO.

We also found that MRF-BP loses its performance advantage to MMSE in 3GPP 3D channel model. It is due to the degradation of MRF-BP's message updating in spatially correlated channel. Lattice reduction restores the orthogonality between columns of channel coefficients matrix $\boldsymbol{H}$ and compensates the degradation of MRF-BP's message updating. It boosts LRA-BP's performance in spatially correlated channel.

Since MMSE gets better performance when channel correlation decreases, LRA-BP's performance gain over MMSE reduces when switching from single-column antenna array to dual-column array. As Figure 5 shown, LRA-BP's gain over MMSE differs 3 dB for different antenna arrays.

Multiuser MIMO. In MU-MIMO, two UEs (each with $N_t = 2$ antennas) transmit data to base

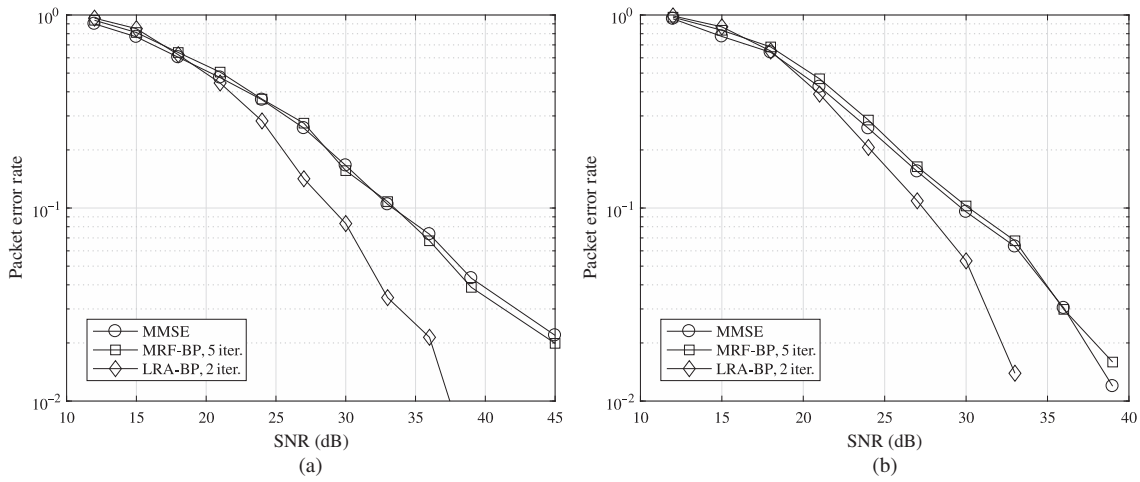

Figure 5 Comparison of receiver performance: 32×4, 3GPP 3D channel, SU-MIMO. (a) $N_r = 32$ $(M = 16, N = 1, P = 2), N_t = 4$; (b) $N_r = 32$ $(M = 8, N = 2, P = 2), N_t = 4$.

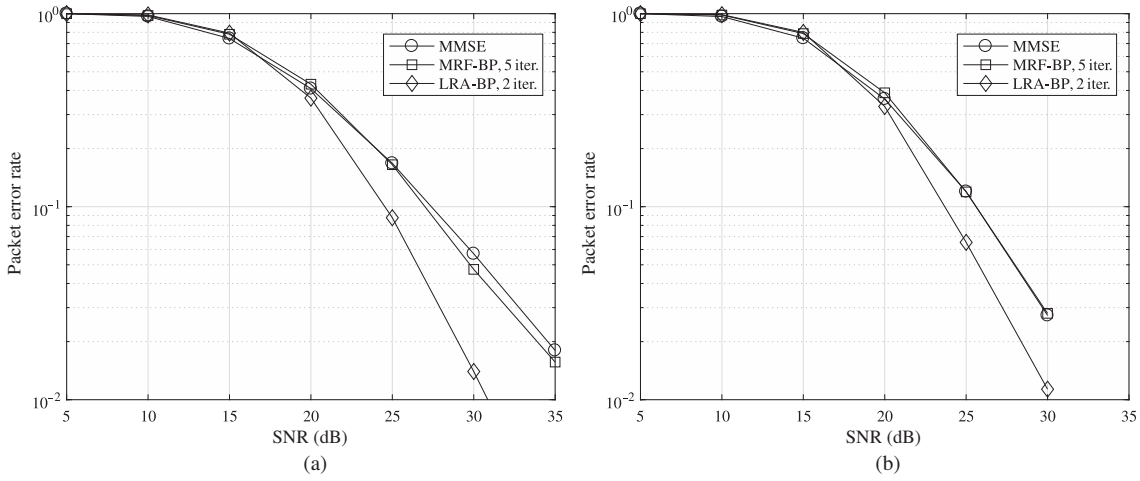

Figure 6 Comparison of receiver performance: 32×4, 3GPP 3D channel, MU-MIMO (two uplink 2-antenna UEs). (a) $N_r = 32$ $(M = 16, N = 1, P = 2), N_t = 4$; (b) $N_r = 32$ $(M = 8, N = 2, P = 2), N_t = 4$.

station and a 32×4 virtual-MIMO system is established. Figure 6 shows that LRA-BP outperforms MMSE and MRF-BP with 6 dB (single-column array) and 3 dB (dual-column array) in terms of required SNR for 1% PER in MU-MIMO.

In MU-MIMO, it is seldom for two UEs to locate closely. The channel correlation for MU-MIMO is lower than SU-MIMO. Consequently MMSE has better performance in MU-MIMO than SU-MIMO. Like comparing single-column antenna array to dual-column array, LRA-BP's performance gain over MMSE reduces 4 dB when switching from SU-MIMO to MU-MIMO, as Figures 5 and 6 shown. But even in a good channel (dual-column antenna array and MU-MIMO), LRA-BP still outperforms MMSE and MRF-BP with 3 dB.

5 Complexity analysis

In this section we compare the computational complexity of LRA-BP and MRF-BP. Since log-domain MRF-BP is with much lower complexity than origianl MRF-BP, we compare LRA-BP with log-domain MRF-BP at first.

The major computational complexity in common to both LRA-BP and MRF-BP includes the iteration in (38) and the translation function $p(z_j|z_i, \boldsymbol{y})$ in (20). For the translation function, its complexity

Table 1 General computational complexity (in terms of equivalent additions)

	Log-domain MRF-BP	LRA-BP
Iteration (38)	$2TN_t(2N_t-1)^2\left\|\boldsymbol{\Omega}\right\|^2$	
Translation function (20)	$80N_r^2N_t$	
LLL reduction	0	$24N_r^2N_t$
In total	$2TN_t(2N_t-1)^2\left\|\boldsymbol{\Omega}\right\|^2+80N_r^2N_t$	$2TN_t(2N_t-1)^2\left\|\boldsymbol{\Omega}\right\|^2+104N_r^2N_t$

Table 2 Computational complexity with specific parameters (in terms of equivalent additions)

	Rayleigh channel		3GPP 3D channel	
	Log-domain MRF-BP	LRA-BP	Log-domain MRF-BP	LRA-BP
N_r	12		32	
N_t	12		4	
T	5		5	2
$\left\|\boldsymbol{\Omega}\right\|$	16	9	16	9
Complexity	16389120	5321592	829440	489488

is mainly for matrix inversion to get $\boldsymbol{K}_{\{j,i\}}^{-1}$. With Sherman-Morrison formula, its complexity can be estimated as $16N_r^2N_t$ multiplications and $16N_r^2N_t$ additions. For ease of comparison, we assumes a multiplication has same complexity as four additions. Then the complexity for translation function, which is common for MRF-BP and LRA-BP, is estimated to be $80N_r^2N_t$ equivalent additions.

For LRA-BP, extra complexity is required to condcut the LLL reduction algorithm and related processing. The LLL reduction algorithm has dynamic complexity. With our observation, it requires $12N_r^2N_t$ equivalent additions to cover most of channel instances. And we doubles this estimation to cover other extra lattice reduction required operations in LRA-BP, including the determination of Z-axis constellation in (8) and LLR generation in (32) and (33).

Table 1 summarizes the computational complexity, where T is the number of iterations and $|\boldsymbol{\Omega}|$ is the size of constellation.

In simulation, we use two sets of parameters, e.g., T and $\left|\boldsymbol{\Omega}\right|$, for Rayleigh channel and 3GPP 3D channel. The detail parameters and corresponding complexity are summarized in Table 2.

Table 2 shows that LRA-BP requires about 33%–60% computational complexity compared with log-domain MRF-BP for different MIMO configurations and algorithm parameters. Since log-domain MRF-BP is with much lower complexity than original MRF-BP, LRA-BP is also with much lower complexity than MRF-BP.

6 Conclusion

In this paper, lattice reduction aided belief propagation (LRA-BP) is proposed for massive MIMO detection. LRA-BP improves MIMO detection performance in spatially correlated channel by applying message updating rules from MRF-BP in lattice reduced MIMO system model, and it also benefits in independent channel. Log-domain arithmetic is used in LRA-BP for computational complexity reduction. For different massive MIMO configurations, LRA-BP outperforms MRF-BP with 3–10 dB in terms of required SNR for 1% packet error rate. LRA-BP requires 33%–60% computational complexity compared with log-domain MRF-BP and consequently much lower complexity compared with MRF-BP.

References

1 Andrews J G, Buzzi S, Choi W, et al. What will 5G be? IEEE J Sel Areas Commun, 2014, 32: 1065–1082
2 Boccardi F, Heath R W, Lozano A, et al. Five disruptive technology directions for 5G. IEEE Commun Mag, 2014, 52: 74–80
3 Ji H, Kim Y, Lee J, et al. Overview of full-dimension mimo in lte-advanced pro. IEEE Commun Mag, 2017, 55: 176–184
4 Yang S, Hanzo L. Fifty years of MIMO detection: the road to large-scale MIMOs. IEEE Commun Surv Tut, 2015, 17: 1941–1988

5 Tang C, Tao Y, Chen Y, et al. Approximate iteration detection and precoding in massive MIMO. China Commun, 2018, 15: 183–196
6 Chen Y, Gao X Q, Xia X G, et al. Robust MMSE precoding for massive MIMO transmission with hardware mismatch. Sci China Inf Sci, 2018, 61: 042303
7 Hu J, Duman T M. Graph-based detector for blast architecture. In: Proceedings of IEEE International Conference on Communications, Glasgow, 2007. 1018–1023
8 Yang J, Zhang C, Liang X, et al. Improved symbol-based belief propagation detection for large-scale mimo. In: Proceedings of IEEE Workshop on Signal Processing Systems (SiPS), Hangzhou, 2015. 1–6
9 Long F, Lv T, Cao R, et al. Single edge based belief propagation algorithms for mimo detection. In: Proceedings of the 34th IEEE Sarnoff Symposium, Princeton, 2011. 1–5
10 Bickson D, Dolev D. Linear detection via belief propagation. In: Proceedings of the 45th Annual Allerton Conference on Communication, Control, and Computing, Allerton, 2007. 7
11 Yoon S, Chae C B. Low-complexity MIMO detection based on belief propagation over pairwise graphs. IEEE Trans Veh Technol, 2014, 63: 2363–2377
12 Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco: Morgan Kaufmann Publishers Inc., 1988
13 Kschischang F R, Frey B J, Loeliger H A. Factor graphs and the sum-product algorithm. IEEE Trans Inform Theor, 2001, 47: 498–519
14 Montanari A, Prabhakar B, Tse D. Belief propagation based multiuser detection. In: Proceedings of the 43th Allerton Conference on Communications, Control and Computing, Monticello, 2005. 86
15 Takahashi T, Ibi S, Sampei S. Design of adaptively scaled belief in large mimo detection for high-order modulation. In: Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, 2017. 1800–1505
16 Som P, Datta T, Chockalingam A, et al. Improved large-mimo detection based on damped belief propagation. In: Proceedings of IEEE Information Theory Workshop on Information Theory, Cairo, 2010. 1–5
17 Bickson D, Dolev D, Shental O, et al. Gaussian belief propagation based multiuser detection. In: Proceedings of IEEE International Symposium on Information Theory, Toronto, 2008. 1878–1882
18 3rd Generation Partnership Project (3GPP). Study on 3d channel model for LTE. TR-36.873. http://www.3gpp.org/DynaReport/36873.htm
19 Xu W L. Capacity improvement analysis of 3D-beamforming in small cell systems. Sci China Inf Sci, 2018, 61: 022305
20 Wubben D, Seethaler D, Jalden J, et al. Lattice reduction. IEEE Signal Process Mag, 2011, 28: 70–91
21 Wubben D, Bohnke R, Kuhn V, et al. Mmse-based lattice-reduction for near-ml detection of mimo systems. In: Proceedings of ITG Workshop on Smart Antennas, Munich, 2004. 106–113
22 Shabany M, Gulak P G. The application of lattice-reduction to the k-best algorithm for near-optimal mimo detection. In: Proceedings of IEEE International Symposium on Circuits and Systems, Seattle, 2008. 316–319
23 Peng G, Liu L, Zhou S, et al. Algorithm and architecture of a low-complexity and high-parallelism preprocessing-based k-best detector for large-scale MIMO systems. IEEE Trans Signal Process, 2018, 66: 1860–1875
24 Liu T H. Comparisons of two real-valued MIMO signal models and their associated ZF-SIC detectors over the Rayleigh fading channel. IEEE Trans Wirel Commun, 2013, 12: 6054–6066
25 3rd Generation Partnership Project (3GPP). Physical channels and modulation. TS-36.211. http://www.3gpp.org/DynaReport/36211.htm
26 Lenstra A K, Lenstra H W, Lovász L. Factoring polynomials with rational coefficients. Math Ann, 1982, 261: 515–534
27 Jensen J. Statistics for Petroleum Engineers and Geoscientists. Amsterdam: Elsevier, 2000. 207
28 Windpassinger C. Detection and precoding for multiple input multiple output channels. Dissertation for Ph.D. Degree. Nurnberg: University Erlangen, 2004. 33–36
29 Rasmussen C E, Williams C. Gaussian Processes for Machine Learning. Cambridge: The MIT Press, 2006
30 Robertson P, Hoeher P, Villebrun E. Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding. Eur Trans Telecomm, 1997, 8: 119–125
31 3rd Generation Partnership Project (3GPP). Multiplexing and channel coding. TS-36.212. http://www.3gpp.org/DynaReport/36212.htm
32 Rusek F, Persson D, Lau B K, et al. Scaling up mimo: opportunities and challenges with very large arrays. IEEE Signal Process Mag, 2013, 30: 40–60
33 Studer C, Seethaler D, Bolcskei H. Finite lattice-size effects in mimo detection. In: Proceedings of the 42nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2008. 2032–2037