Near-optimal stochastic MIMO signal detection with a mixture of t-distributions prior

Junichiro Hagiwara[†], Kazushi Matsumura, Hiroki Asumi, Yukiko Kasuga, Toshihiko Nishimura,
Takanori Sato, Yasutaka Ogawa, and Takeo Ohgane
Graduate School / Faculty of Information Science and Technology, Hokkaido University
Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido, 060-0814 Japan
Email: {jhagiwara, kazu4.matsumura, asumi, kasuga, nishim, tksato, ogawa, ohgane}@m-icl.ist.hokudai.ac.jp
[†]Current affiliation: Faculty of Social Informatics, Mukogawa Women's University

Abstract—**Multiple-input multiple-output (MIMO) systems will play a crucial role in future wireless communication, but improving their signal detection performance to increase transmission efficiency remains a challenge. To address this issue, we propose extending the discrete signal detection problem in MIMO systems to a continuous one and applying the Hamiltonian Monte Carlo method, an efficient Markov chain Monte Carlo algorithm. In our previous studies, we have used a mixture of normal distributions for the prior distribution. In this study, we propose using a mixture of t-distributions, which further improves detection performance. Based on our theoretical analysis and computer simulations, the proposed method can achieve near-optimal signal detection with polynomial computational complexity. This high-performance and practical MIMO signal detection could contribute to the development of the 6th-generation mobile network.**

Index Terms—**MIMO, signal detection, MCMC, Hamiltonian Monte Carlo, prior distribution, t-distribution**

I. INTRODUCTION

Wireless communication has become a crucial component of modern society's infrastructure, as evidenced by the widespread use and growth of 5G mobile network. To further improve convenience, research and development of 6G is underway to effectively use radio resources. Among the various technologies that play an essential role in the effective use of radio resources, multiple-input multiple-output (MIMO) technology is expected to make a significant contribution. The MIMO technology enhances wireless transmission efficiency using multiple antennas at both transmitting and receiving stations. Although MIMO with a relatively small number of antennas is already in general use, research on massive MIMO with numerous antennas just started receiving attention [1].

In MIMO systems, accurately detecting received signals is critical for improving transmission efficiency. Maximum likelihood decoding, which can achieve ideal performance, is computationally impractical for detecting received signals due to its examination of all possible combinations of transmitted signals [2]. Meanwhile, linear detection techniques, e.g., the minimum mean-square error method, are computationally less expensive but underperform [2]. The challenge lies in improving the signal detection performance while maintaining realistic computational complexity requirements. One solution

is to approach signal detection from a stochastic perspective. This allows for the theoretical consideration of uncertainty in the problem and the application of state-of-the-art probabilistic and statistical techniques.

The problem of MIMO signal detection has been approached from a stochastic perspective using several techniques, such as Bayesian methods [3], [4]. However, to the best of our knowledge, continuous prior distribution has been regarded as inappropriate for MIMO signal detection, except for our previous proposals [5], [6]. The mixed Gibbs sampling (MGS) method [7] approximates the posterior distribution using a large number of samples, where a discrete distribution is set as the prior distribution and the algorithm is based on Gibbs sampling [8], a type of Markov chain Monte Carlo (MCMC). The expectation propagation (EP) method [9] also sets a discrete distribution as the prior distribution and uses the EP algorithm [10] to estimate the parameters of the approximate posterior distribution. Moreover, the prior distribution in [11] is set as a mixture of truncated normal distributions. This method uses a variational Bayesian algorithm to determine the parameters of the approximate distribution for the posterior distribution. Previous studies on MIMO signal detection that assume stochasticity have set a discrete or discontinuous distribution as the prior distribution based on the assumption of discrete transmission symbols.

We extend the discrete signal detection problem in MIMO systems to a continuous one and apply the Hamiltonian Monte Carlo (HMC) method [12], an efficient MCMC algorithm for continuous problems. In our previous studies [5], [6], we set a mixture of normal distributions as the prior distribution. In this study, we propose a novel method using a mixture of t-distributions, which further improves signal detection performance. Through our theoretical analysis and computer simulations, the proposed method is shown to achieve near-optimal signal detection with polynomial computational complexity. Our novel MIMO signal detection method will contribute to both the practical and theoretical aspects of future wireless communications.

The remainder of this article is structured as follows: Section II presents a stochastic formulation of the problem. Section III briefly describes the MGS and EP methods as existing baseline methods. Section IV explains the details

of the proposed method. Section V discusses our theoretical analysis and computer simulation results. Finally, we present a summary in Section VI.

The following notations are used in this study. $\mathbb{C}$ denotes the field of complex numbers. $\mathrm{Re}(\mathbf{x})$ and $\mathrm{Im}(\mathbf{x})$ denote the real and imaginary parts of $\mathbf{x}$, respectively. $\hat{\mathbf{x}}$ denotes the estimate of $\mathbf{x}$. $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. $\mathbf{0}$ and $\mathbf{I}$ denote the zero vector and unit matrix, respectively. $\mathbf{X}^{\mathrm{T}}$ denotes the transpose of $\mathbf{X}$. The symbol $\sim$ indicates that the random variable on its left-hand side follows the probability distribution on its right-hand side. The symbol $\propto$ represents a proportional relationship.

II. PROBLEM FORMULATION

A. System model

Suppose a full-stream transmission in a MIMO system with N transmission antennas and M receiving antennas. The following relationship holds:

$$\mathbf{y} = \mathbf{H}\mathbf{u} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}), \quad (1)$$

where the receiving antenna's symbol vector is $\mathbf{y} = [y_1, \ldots, y_M]^{\mathrm{T}} \in \mathbb{C}^M$, the channel matrix is $\mathbf{H} \in \mathbb{C}^{M \times N}$, the transmission antenna's symbol vector is $\mathbf{u} = [u_1, \ldots, u_N]^{\mathrm{T}} \in \mathbb{C}^N$, the noise vector is $\mathbf{w} = [w_1, \ldots, w_M]^{\mathrm{T}} \in \mathbb{C}^M$, the noise variance is σ_w^2, and $\mathcal{CN}$ represents a circularly symmetric complex normal distribution. For manipulation ease, we split complex numbers into their real and imaginary components as follows: $\mathbf{y} \to \begin{bmatrix} \mathrm{Re}(\mathbf{y}) \\ \mathrm{Im}(\mathbf{y}) \end{bmatrix}$, $\mathbf{H} \to \begin{bmatrix} \mathrm{Re}(\mathbf{H}) & -\mathrm{Im}(\mathbf{H}) \\ \mathrm{Im}(\mathbf{H}) & \mathrm{Re}(\mathbf{H}) \end{bmatrix}$, $\mathbf{u} \to \begin{bmatrix} \mathrm{Re}(\mathbf{u}) \\ \mathrm{Im}(\mathbf{u}) \end{bmatrix}$, and $\mathbf{w} \to \begin{bmatrix} \mathrm{Re}(\mathbf{w}) \\ \mathrm{Im}(\mathbf{w}) \end{bmatrix}$. As a supplementary note, the symbols for vectors and matrices remain unchanged hereafter when N and M are reconsidered as $2N$ and $2M$, respectively. In this study, the channel matrix $\mathbf{H}$, noise variance σ_w, and received signal $\mathbf{y}$ are assumed to be known. The transmitted symbols are also assumed to be uniformly random across antennas due to the use of scramblers. Signal detection entails estimating the transmitted symbol vector $\mathbf{u}$, given the received signal $\mathbf{y}$.

By introducing a stochastic interpretation, "posterior distribution $\propto$ likelihood $\times$ prior distribution" holds according to Bayes' theorem:

$$p(\mathbf{u} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{u})p(\mathbf{u}). \quad (2)$$

The goal of stochastic signal detection is to obtain a point estimate of the posterior distribution $p(\mathbf{u} \mid \mathbf{y})$.

B. Likelihood

Equation (1) shows $p(\mathbf{y} \mid \mathbf{u}) = \mathcal{N}(\mathbf{y}; \mathbf{H}\mathbf{u}, \sigma_w^2 \mathbf{I})$, where $\mathcal{N}$ represents a real-value normal distribution density.

C. Prior distribution

Equation (2) implies that the prior distribution corresponds to regularization terms for correcting the likelihood. In signal detection, the prior distribution is typically a discrete multinomial distribution that reflects the possibility for discrete signal points. This improves the accuracy of the posterior
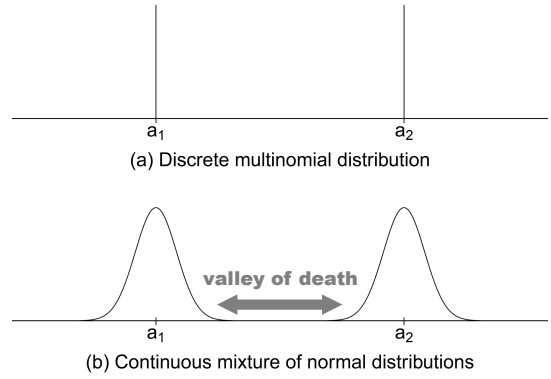

Fig. 1. Two examples of priors for BPSK.

distribution's estimation by considering the priority at the transmission signal point, as in (3) (see also Fig. 1 (a)):

$$p(\mathbf{u}) = \prod_{n=1}^{2N} \frac{1}{q} \{ \delta(u_n - a_1) + \cdots + \delta(u_n - a_q) \}, \quad (3)$$

where q denotes the square root of the modulation order, $a_1, \ldots, a_q$ denotes the real-valued coordinate of the transmission signal points, and $\delta(x)$ denotes the unit probability mass at x. For example, the MGS and EP methods assume such a prior distribution.

Assuming a continuous distribution for the prior distribution transforms the above problem into a continuous value problem. For example, the following equation demonstrates the application of a mixture of normal distributions (refer to Fig. 1 (b)):

$$p(\mathbf{u}) = \prod_{n=1}^{2N} \frac{1}{q} \{ \mathcal{N}(u_n; a_1, \sigma) + \cdots + \mathcal{N}(u_n; a_q, \sigma) \}, \quad (4)$$

where σ^2 denotes a variance of a component's normal distribution and tuning matter. In our previous studies [5], [6], we used a mixture of normal distributions as the prior distribution and set the optimal value of σ to minimize the bit error rate (BER) through a preliminary search. The optimal value of σ corresponds to the equilibrium point where the search efficiency is the best. If σ is too large, the search efficiency is low as areas other than the transmission signal points are unnecessarily explored. Meanwhile, if σ is too small, there is little overlap in the component's distribution, making it difficult to explore other possible transmission signal points (the "valley of death" in Fig. 1 (b)).

D. Posterior distribution

When the prior distribution is assumed to be a mixture of any distribution, it is impossible to obtain the posterior distribution analytically in closed form [13]. Regardless of whether the component distribution is discrete or continuous, a numerical approximation algorithm must be used to derive the posterior distribution. The point estimate $\hat{\mathbf{u}}$ of the posterior distribution can produce multiple candidates, depending on the approximation algorithm used. In addition, the $\hat{\mathbf{u}}$ may

differ from the original transmission signal point. For instance, when using a continuous distribution as the prior distribution, the search may include areas other than discrete transmission signal points, leading to deviations. To address this issue, we compute the likelihood $p(\mathbf{y} \mid \tilde{\mathbf{u}})$ after quantizing $\hat{\mathbf{u}}$ to the nearest transmission signal point $\tilde{\mathbf{u}}$ and consider $\tilde{\mathbf{u}}$ with the highest likelihood as the final point estimate.

III. PREVIOUS WORK

A. MGS method [7]

The MGS method approximates the posterior distribution by generating a large number of samples. It employs Gibbs sampling to search more intensively in regions with higher posterior probability densities. The MGS method improves the search efficiency by mixing the initialization of the search values with a probability of $1/(2N)$, which solves the problem of getting stuck in local optima. This approach can be seen as a virtual parallelization of the Markov chains. The multiple restarts technique, also proposed in [7], involves running multiple MGS methods with different initial values for the Markov chain and selecting the result with the highest likelihood. It was shown that using a sufficient number of restarts can result in near-optimal performance.

The computational complexity of the MGS method is $\mathcal{O}(MN)$ per step of the Markov chain, as the computation of the likelihood is the main factor that contributes to its complexity. With L_{MGS} as the total steps in the Markov chain, the final computational complexity is $\mathcal{O}(L_{\mathrm{MGS}}MN)$.

B. EP method [9]

The EP method approximates the posterior distribution using an uncorrelated multivariate normal distribution $q(\mathbf{u})$. In this method, the EP algorithm is used to find the parameters that minimize the Kullback–Leibler divergence $-\int p(\mathbf{u} \mid \mathbf{y}) \ln\{q(\mathbf{u})/p(\mathbf{u} \mid \mathbf{y})\}\mathrm{d}\mathbf{u}$. Specifically, the mean and variance parameters are iteratively refined, and the final mean parameter is used to estimate the transmitted symbols. According to [9], the total number of iterations, L_{EP}, can be set to at most $L_{\mathrm{EP}} = 10$ to reach the maximum detection performance.

The computational complexity of the EP method is $\mathcal{O}(N^3)$ per iteration, as the inverse matrix operation of $\mathbf{H}$ is the main factor that contributes to its complexity. With a total of $L_{\mathrm{EP}} = 10$ iterations assumed, the final computational complexity becomes $\mathcal{O}(10N^3)$.

IV. STOCHASTIC SIGNAL DETECTION WITH A MIXTURE OF t-DISTRIBUTIONS PRIOR

A. Setting the prior distribution

In this study, we propose the use of a mixture of t-distributions as the prior distribution in signal detection as in (5):

$$p(\mathbf{u}) = \prod_{n=1}^{2N} \frac{1}{q}\{\mathcal{T}(u_n; a_1, \sigma, \nu) + \cdots + \mathcal{T}(u_n; a_q, \sigma, \nu)\}, \quad (5)$$

where $\mathcal{T}$ denotes a real-valued t-distribution density [13]. The scale parameter σ and the degrees of freedom ν are
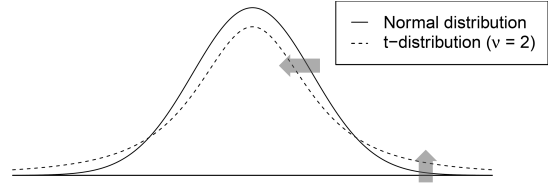

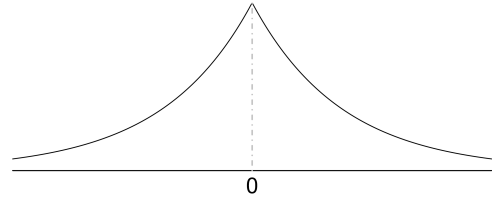Fig. 2. The normal and t-distributions with the same scale parameter σ.

Fig. 3. Laplace (double exponential) distribution.

adjustable or tunable parameters. Regarding these, we employ the optimal values obtained through preliminary searches. The t-distribution has a narrower peak and thicker tail than the normal distribution with the same scale parameter σ (Fig. 2). Therefore, compared with the normal distribution, the t-distribution can more thoroughly search around the transmission signal points. It also actively explores other potential transmission signal points by overcoming the "valley of death" between them. As a result, it is expected to have exploration properties superior to our previously proposed mixture of normal distributions [5], [6].

Furthermore, the t-distribution is suitable for discrete signal detection from the perspective of sparse estimation. The prior distribution in Bayesian estimation corresponds to the constraint (i.e., regularization term) in sparse estimation. For example, the sparse regression method known as Lasso, proposed in [14], applies the ℓ_1 norm to the regularization term in estimating regression coefficients that are mostly zero but rarely take finite values. This regularization term with the ℓ_1 norm corresponds to setting the Laplace (double exponential) distribution centered at zero as a prior distribution (Fig. 3). The kernel part of the standard Laplace distribution can be expressed as follows:

$$\exp\{-|x|\} = \exp\{-\max(-x, x)\}$$
$$\sim \exp\{-\mathrm{smooth\_max}(-x, x)\} = \exp\{-\mathrm{LogSumExp}(-x, x)\}$$
$$= \exp\left\{-\log\left(e^{-x} + e^x\right)\right\} = \left(e^{-x} + e^x\right)^{-1} \propto \{\cosh(x)\}^{-1}$$
$$\sim \left(1 + \frac{x^2}{2!}\right)^{-1}. \quad (6)$$

As the kernel part of the standard t-distribution is $(1 + \frac{x^2}{\nu})^{-(\nu+1)/2}$, the t-distribution can approximate the Laplace distribution well with $\nu \sim 2$. Therefore, setting the t-distribution centered at the signal point as a prior distribution can be interpreted as allowing sparse estimation based on the ℓ_1 norm around the signal point.

Algorithm 1 HMC method sampling

1: Initialize $\mathbf{u}$ at random
2: **for** $l = 1, \ldots, L_{\mathrm{HMC}}$ **do**
3:     Draw $\mathbf{r}$ from $\mathcal{N}(\mathbf{0}, \mathbf{I})$
4:     Numerically solve Hamilton's equations (8) to obtain $\mathbf{u}'$ and $\mathbf{r}'$
5:     Update $\mathbf{u} \leftarrow \mathbf{u}'$ with probability $\min[1, \exp\{H(\mathbf{u}, \mathbf{r}) - H(\mathbf{u}', \mathbf{r}')\}]$
6:     Regard the updated $\mathbf{u}$ as a sample from the posterior distribution $p(\mathbf{u} \mid \mathbf{y})$
7: **end for**

B. Approximation algorithm for the posterior distribution

In signal detection, setting a continuous distribution as the prior distribution yields a continuous posterior distribution, as the likelihood is a continuous normal distribution. Thus, effective approximation algorithms for continuous problems can be used in estimating the posterior distribution. In our previous study [6], we compared the results of using Newton's method, automatic differentiation variational inference method [15], and the HMC method as the approximation algorithm, under the condition of a mixture of normal distributions as the prior distribution. The results showed that the HMC method mostly achieved superior signal detection performance with the same computational complexity. This is because the HMC algorithm randomly initializes the searching value at each step of the Markov chain, which helps avoid local optima. Thus, we employ the HMC method as the approximation algorithm for the posterior distribution in this study.

The HMC method, being a type of MCMC, uses the Markov chain mechanism like the MGS method. It explores areas with higher posterior probability densities to generate samples that approximate the posterior distribution. Compared with other MCMC methods, the HMC method is efficient in terms of sampling due to its innovative use of the Hamiltonian mechanics framework. The HMC method intentionally adds a quantity, $\mathbf{r} = \mathrm{d}\mathbf{u}/\mathrm{d}\tau$ (where τ is a virtual time), that corresponds to the momentum of $\mathbf{u}$ in addition to the variable $\mathbf{u}$ being estimated. A summary of the HMC method is provided below. In this method, the system's potential energy, U, and kinetic energy, K, are defined as $U(\mathbf{u}) = -\ln(p(\mathbf{u} \mid \mathbf{y}))$ and $K(\mathbf{r}) = 1/2||\mathbf{r}||^2$, respectively. The Hamiltonian is introduced as follows:

$$H(\mathbf{u}, \mathbf{r}) = U(\mathbf{u}) + K(\mathbf{r}), \tag{7}$$

which represents the total energy of the system. Then, Hamilton's equations are expressed using these two partial differential equations:

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}\tau} &= \frac{\partial H(\mathbf{u}, \mathbf{r})}{\partial \mathbf{r}} = \mathbf{r}, \\
\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}\tau} &= -\frac{\partial H(\mathbf{u}, \mathbf{r})}{\partial \mathbf{u}} = -\frac{\partial U(\mathbf{u})}{\partial \mathbf{u}}.
\end{aligned} \tag{8}$$

Algorithm 1 shows the sampling with the HMC method. The Hamiltonian is constant based on (7). Thus, a large change in the momentum $\mathbf{r}$ significantly influences the sample value $\mathbf{u}$. In addition, as per Algorithm 1, most proposals $\mathbf{u}'$ are accepted with a probability of one, excluding cases of numerical errors.

These lead to a typically higher sampling efficiency for the HMC method than for other MCMC methods.

According to Algorithm 1, when solving (8) numerically during one step of the Markov chain, the log-posterior probability density derivative is internally evaluated L times. The value of L can vary depending on the problem and conditions, but it is typically assumed to be 10 in this study [16]. The main factor contributing to the computational complexity is the calculation of the term $(\mathbf{H}^{\mathrm{T}}\mathbf{H})\mathbf{u}$. This term is included in the derivative of the log-likelihood contained in the log-posterior probability density, and because $\mathbf{H}$ is assumed to be known, $(\mathbf{H}^{\mathrm{T}}\mathbf{H})$ only needs to be calculated once. Hence, the computational complexity of the HMC method is $\mathcal{O}(LN^2) = \mathcal{O}(10N^2)$ per step of the Markov chain. With L_{HMC} being the total steps in the Markov chain, the final computational complexity becomes $\mathcal{O}(10L_{\mathrm{HMC}}N^2)$.

V. NUMERICAL RESULTS AND DISCUSSION

We demonstrate that the proposed method can attain near-optimal signal detection in MIMO systems with polynomial computational complexity. To verify this, we perform a theoretical analysis of the computational complexity as well as computer simulations on signal detection.

A. Settings in numerical analysis

Common assumptions: We assume typical and exhaustive conditions, such as the number of antennas considering massive MIMO and the modulation order from low to high (as outlined in Table I). The BER plots are omitted if the computer simulation results are error-free.

Parameters of proposed method: The parameters of the component's normal and t-distributions are set to the values that displayed favorable performance in the preliminary rough search (Table II). Each Markov chain in the simulation is assumed to have $2N$ steps, in line with the virtual parallelization approach of the MGS method. The parallel number of Markov chains is set to $\lfloor 1000/(2N) \rfloor$, which is the minimum number required to
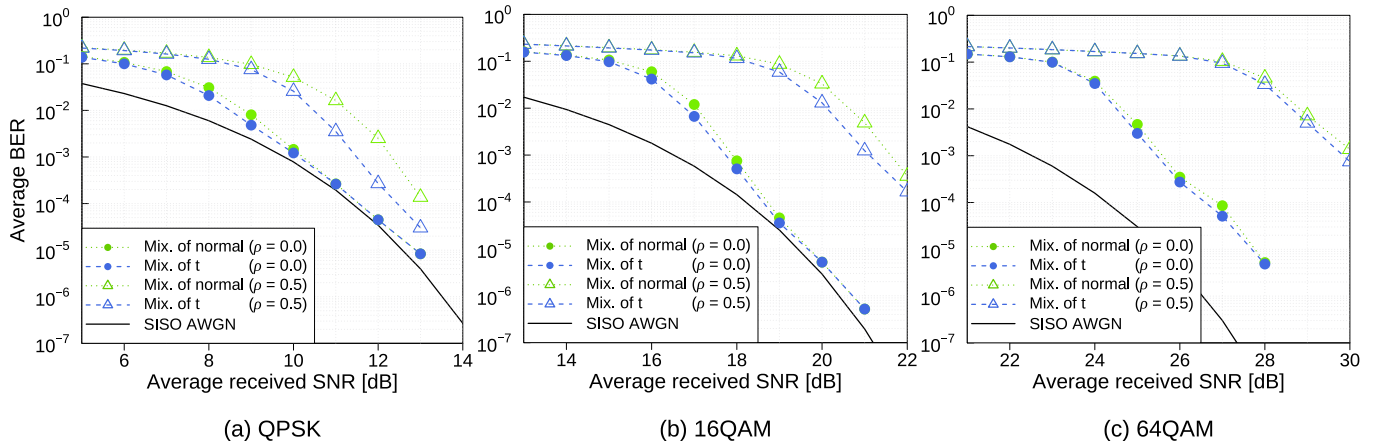
Fig. 4. Average BER vs. average received SNR for $\rho = 0$ and 0.5.

attain sufficient performance [5]. As a result, the total number of steps in the Markov chain, L_{HMC}, is equivalent to 1000.

Parameters of existing methods: The total number of steps in the Markov chain of the MGS method is set to $L_{\text{MGS}} = 1000$, equal to L_{HMC}. However, the proposed method has a computational complexity that is 10 times greater than the MGS method due to its implicit internal loop $L = 10$. To ensure a fair comparison, 10 multiple restarts of the MGS method are performed.

The total number of EP method iterations is set to $L_{\text{EP}} = 10$, which is a sufficient number to obtain adequate performance.

In the context of MIMO signal detection, the performance of a single-input single-output (SISO) transmission under additive white Gaussian noise (AWGN) serves as an ideal benchmark, as it eliminates inter-antenna interference and fading effects. The BER performance of the SISO AWGN is shown to highlight the theoretical lower bound of MIMO signal detection.

B. Computational complexity

The computational complexities of the proposed method, the MGS method, and the EP method for signal detection are $\mathcal{O}(10 \times 1000 N^2)$, $10 \times \mathcal{O}(1000 N^2)$, and $\mathcal{O}(10 N^3)$, respectively, where $N = M$. These complexities are all of polynomial orders. Particularly, the proposed method and the MGS method have the same complexity under the given conditions, whereas the EP method is the least computationally expensive due to $N = 96 \, (< 1000)$.

C. BER performance: a mixture of t-distributions prior vs. a mixture of normal distributions prior

Overall trend (Fig. 4 (a) through (c)): The higher the modulation order, the further the performance deviates from the SISO AWGN. This is because as the modulation order increases, the number of possible transmission signal points becomes larger, and the distance between these signal points becomes narrower, making the estimation more challenging.

Comparison with a mixture of normal distributions prior (Fig. 4 (a) through (c)): When $\rho = 0$, it is difficult to discern the difference in performance between the two methods as they overlap in certain areas. A clearer difference in performance can be observed when $\rho = 0.5$, which brings more challenging signal detection scenarios. For all modulation orders, the mixture of t-distributions prior is found to be superior when $\rho = 0.5$. For instance, when the BER is 10^{-3}, the mixture of t-distributions prior outperforms the mixture of normal distributions prior by 0.8, 0.5, and 0.3 dB for QPSK, 16QAM, and 64QAM, respectively. In summary, the mixture of t-distributions prior provides better signal detection performance than the mixture of normal distributions prior due to its component's narrower peak and thicker tail. The former allows for a more focused search around the transmission signal points, whereas the latter enables a more aggressive search across the "valley of death" between the transmission signal points.

The improvement obtained from using a mixture of t-distributions as the prior distribution decreases as the modulation order increases. This advantage is particularly small for 64QAM modulation. As the modulation order increases and the solution space expands, even with an aggressive search using a t-distribution component, the effectiveness of the method becomes limited with a finite number of searches.

D. BER performance: proposed method with a mixture of t-distributions prior vs. existing methods

Overall trend (Fig. 5 (a) through (c)): As aforementioned, the higher the modulation order, the further the performance is away from the SISO AWGN.

Comparison with existing methods for QPSK (Fig. 5 (a)): The performance of all methods is comparable to the SISO AWGN, with only minute differences. This is due to the signal detection ease, given the limited number of potential transmission signal points and the significant distance between them. In low signal-to-noise ratio (SNR) conditions, the proposed method performs slightly worse than the MGS and EP methods. Under noisy conditions and limited searches, exploring areas outside the transmission signal points appears less effective. In such
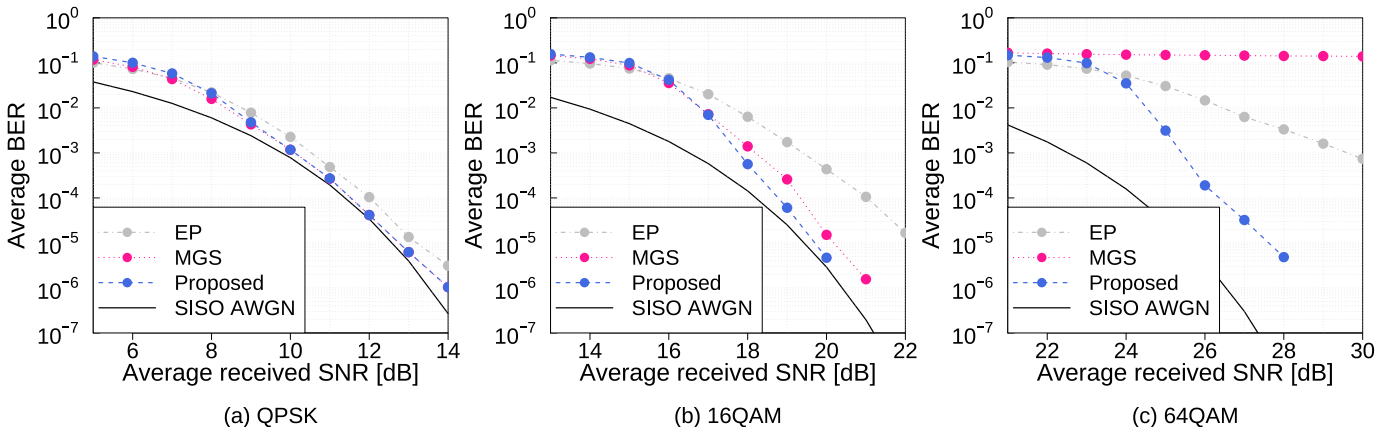
Fig. 5. Average BER vs. average received SNR for $\rho = 0$.

cases, the original impulse-like prior distribution used in the MGS and EP methods appears to be sufficient. It is noted that a low SNR region in practical communication environments usually requires the combination of other improvement techniques, such as retransmission protocol and error-correcting code. Consequently, this slight inferiority of the proposed method is expected to have only a limited impact on the overall system performance. Meanwhile, at moderate-to-high SNRs, the proposed method outperforms the EP method and is comparable to the MGS method. For instance, the SNR gain of the proposed method at a BER of 10^{-3} is close to 0 and 0.4 dB compared with the MGS and EP methods, respectively.

Comparison with existing methods for 16QAM (Fig. 5 (b)): Compared with the QPSK case, the 16QAM has an increased number of possible transmission signal points and a narrower distance between these points, making signal detection more challenging. Such a difficult condition makes the characteristics of each method more clear. The difference from the QPSK case is that the proposed method outperforms the MGS method in terms of detection performance at moderate-to-high SNRs, resulting in the best detection performance for the proposed method at moderate-to-high SNRs. In addition, compared with the QPSK case, the performance difference between the proposed method and existing methods is larger at moderate-to-high SNRs. For instance, the SNR gain of the proposed method at a BER of 10^{-3} is close to 0.4 and 1.6 dB compared with the MGS and EP methods, respectively.

Comparison with existing methods for 64QAM (Fig. 5 (c)): Compared with the QPSK and 16QAM cases, signal detection for 64QAM is the most challenging, and the distinctiveness of the proposed method is the most pronounced. This is due to the greatest number of potential transmission signal points and the smallest distance between these points. The difference between this case and the QPSK and 16QAM cases is that the proposed method outperforms the MGS method in terms of detection performance at low SNR. Although the proposed method demonstrates its best performance at moderate-to-high SNRs, similar to the 16QAM case, the difference in performance with existing methods is greater than that seen in

the 16QAM case. For instance, at a 10^{-3} BER, the proposed method shows an indeterminable large gain in SNR compared with the MGS method and a 4.2-dB gain compared with the EP method.

Additional verification is performed here because the BER performance of MGS is extremely degraded compared to QPSK and 16QAM. Fig. 6 shows excerpts from Figs. 5 (c) and 6 (c) in [5]. The "Conv." and the "iterations" in the legend indicate the MGS and the total number of iterations of the Markov chain, respectively. Fig. 6 shows that BER performance of MGS improves with fewer transmitting and receiving antennas and a higher number of iterations of the Markov chain. Compared to [5], the assumptions of this study make signal detection more difficult due to more transmitting and receiving antennas and fewer iterations of the Markov chain. Therefore, the result of MGS is considered reasonable.

E. BER performance: proposed method with a mixture of t-distributions prior vs. SISO AWGN (Fig. 5 (a) through (c))

The proposed method achieves near-optimal performance. Specifically, the SNR degradation of the proposed method at a 10^{-3} BER is within 0.3, 1.3, and 2.9 dB for QPSK, 16QAM, and 64QAM, respectively. This exceptional performance is achieved through a combination of the appropriate prior distribution setting and an effective search algorithm. In particular, the proposed prior distribution is sufficiently similar to the discrete point mass at the transmission signal point and also enables active exploration of other potential signal points. Therefore, we consider the proposed method is sufficiently reliable, as demonstrated by its outstanding performance and appropriate methodology.

VI. CONCLUSIONS

Based on a mixture of t-distributions prior and the HMC method, we proposed a signal detection method that is expected to deliver near-optimal performance with polynomial computational complexity. In terms of detection performance, the proposed method showed a substantial improvement over
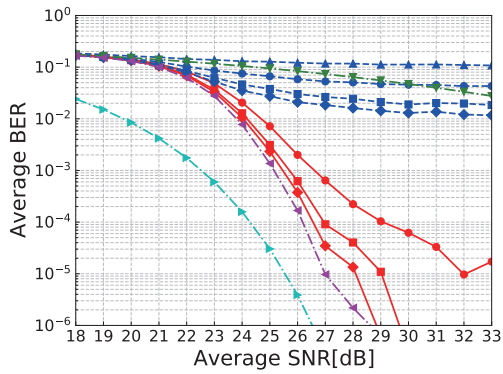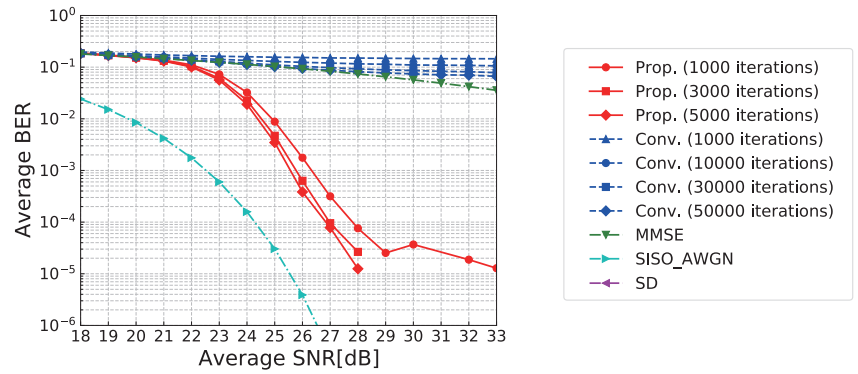
Fig. 5 (c) in [5] (N = M = 16)

Fig. 6 (c) in [5] (N = M = 32)

Fig. 6. Average BER vs. average received SNR for 64QAM and $\rho = 0$. Reprinted from [5] with permission (© 2021 IEEE).

typical existing methods for higher-order modulation. Higher-order modulation could be crucial for future wireless communications. The amount of data transferred in wireless communications is projected to grow exponentially with expectations for multi-sensory interactions beyond traditional voice and video communication. In this sense, the proposed method is deemed beneficial.

The limitations of this study are as follows. The proposed method's computational complexity, although of polynomial order, is greater than that of the EP method. In addition, the BER performance may be slightly lower than those of the MGS and EP methods at low SNR. Fine-tuning the t-distribution parameter for improved performance at low SNR may be possible, but further investigation is required. Although channel coding was not applied in this study, signal detection methods are often used in combination with error-correcting codes in practice. Therefore, we will carefully investigate its impact in the future, considering various parameters such as encoding scheme, code rate, decoding algorithm, and interleaving.

From a mathematical perspective, our study suggests that extending the discrete problem to a continuous one can lead to better solutions. This approach has the potential to revolutionize combinational optimization, including well-known problems like the traveling salesman and nurse scheduling problems.

REFERENCES

[1] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[2] A. Chockalingam and B. Rajan, *Large MIMO Systems*. Cambridge University Press, 2014.

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

[4] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019.

[5] K. Matsumura, J. Hagiwara, T. Nishimura, T. Ohgane, Y. Ogawa, and T. Sato, "A novel MIMO signal detection method using Hamiltonian Monte Carlo approach," in *2021 24th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2021, pp. 1–6.

[6] H. Asumi, Y. Kasuga, K. Matsumura, J. Hagiwara, T. Nishimura, T. Sato, Y. Ogawa, and T. Ohgane, "Comparison of performance and complexity for different search methods in stochastic MIMO signal detection," in *IEICE Technical Report RCS*, vol. 122, no. 73, June 2022, pp. 150–155, (in Japanese).

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

[8] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, 1984.

[9] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 2840–2849, 2014.

[10] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, ser. UAI'01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 362–369.

[11] P. Wang, T. Koike-Akino, P. V. Orlik, M. Pajovic, and K. J. Kim, "Variational Bayesian symbol detection for massive MIMO systems with symbol-dependent transmit impairments," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[12] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Physics Letters B*, vol. 195, no. 2, pp. 216–222, 1987.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[14] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[15] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei, "Automatic differentiation variational inference," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 430–474, Jan. 2017.

[16] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. CRC Press, 2013.