

Iterative Demodulation and Decoding Algorithm for 3GPP/LTE-A MIMO-OFDM Using Distribution Approximation

Tao Cui, FeiFei Gao, *Senior Member, IEEE*, Arumugam Nallanathan, *Fellow, IEEE*, Hai Lin, *Senior Member, IEEE*, and Chintha Tellambura, *Fellow, IEEE*

Abstract—Soft iterative detection/decoding algorithms are fundamentally necessary for multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) adopted in the 3rd Generation Long Term Evolution - Advanced (LTE-A) in order to increase the capacity and achieve high data rates. However, their high performance critically requires log likelihood ratio (LLR) computations with prohibitive complexity. This challenge will be addressed in this paper. We first use the assumption of Gaussian transmit symbols to show the equivalence among several existing algorithms. We next develop a non-Gaussian approximation for high-order constellations, which paves the way for interference cancellation based detectors. Based on both Gaussian and non-Gaussian approximations, we thus develop several capacity-achieving iterative MIMO-OFDM demodulation and decoding algorithms. To this end, we adopt K -best algorithms to take advantage of both types of approximations and the list decoder. Unlike existing algorithms, our proposed K -best algorithms make use of the *a priori* probabilities to generate the list. Simulations of standard-compliant LTE systems demonstrate that the proposed algorithms outperform existing ones.

Index Terms—Iterative decoding, MIMO-OFDM, K -best, Gaussian approximation, LTE-A, turbo principle, 3GPP.

I. INTRODUCTION

THE 3rd Generation Partnership Project (3GPP) has standardized the Long Term Evolution - Advanced (LTE-A) in Release 10 [1] with ambitious data rates at more than 1 Gb/s in local areas and 100 Mb/s in wide areas. To meet these challenging requirements, multiple-input multiple-output (MIMO) techniques as well as high spectrum allocation of 100 MHz promise a linear increase of the wireless link capacity. To counteract the delay spread and to facilitate flexible resource allocation, in the downlink, orthogonal frequency-division multiplexing (OFDM) is combined with MIMO. However, to

attain the capacity gains of MIMO-OFDM, one of the fundamental challenges is the complexity of the detection/decoding algorithms measured by decoding power consumption/silicon area [2]–[5].

Even for uncoded MIMO systems, the complexity of optimal maximum likelihood (ML) decoding is exponential in the number of antennas. Low complexity but sub-optimal receiver structures such as the vertical Bell Laboratories Layered Space-Time (V-BLAST) nulling/cancelling algorithm and many others are thus developed [6]–[10]. On the other hand, the complexity of ML decoding can be significantly reduced, especially in the high signal-to-noise ratio (SNR) region, by sphere decoding (SD) [11]. Nevertheless, if channel coding is added to MIMO, then the size of the search space is exponential in the product of the code blocklength and the number of antennas. The optimal joint detector/decoder is prohibitively complex notwithstanding the use of SD. For this reason, the main idea of near-capacity iterative detection/decoding method [12] is the iterative exchange between the MIMO detector and the channel decoder — the MIMO detector uses log likelihood ratio (LLR) information from the channel decoder, which then incorporates soft information from the MIMO detector. These iterations continue until desired performance is achieved. The decoder computing exact LLR is called maximum a posteriori (MAP) decoder. However, repeated LLR computations are essential.

The computation burden of LLR for each bit must be reduced. To achieve this goal, [12] modifies the standard SD to generate a list of candidates at the detector and only the contributions from the list are included in the LLR values needed for the iterative process. Moreover, the LLR computation is simplified via the max-log approximation. Although these two steps allow [12] to reach near-optimal performance with lower complexity, the complexity depending on SNR can nevertheless be high. Another idea is to avoid the rapid expansion of the search tree by retaining only a fixed number of nodes at each step. Following this idea, a K -best SD is proposed in [13] for both uncoded and coded MIMO systems, which has a constant complexity across the entire range of SNR. Another approach [14]–[17] to adopt soft information in iterative detection and decoding is to use nonlinear interference cancellation followed by a single-input single-output (SISO) iterative detector and decoder in [18], which is shown to have low complexity. Yet another idea is for the computation of the LLR of each bit, the other symbols are ranked according to

Tao Cui is a visiting scholar with Institute of Information Processing, Department of Automation, Tsinghua University, Beijing 100084, China (Email: cuitao42@yahoo.com).

Feifei Gao is with Institute of Information Processing, Department of Automation, Tsinghua University, Beijing 100084, China (Email: feifeigao@tsinghua.edu.cn).

Arumugam Nallanathan is with Faculty of Natural & Mathematical Sciences, King's College London, London WC2R 2LS, U.K. (Email: arumugam.nallanathan@kcl.ac.uk).

Hai Lin is with Graduate School of Engineering, Osaka Prefecture University, Osaka, Japan (Email: hai.lin@ieee.org).

Chintha Tellambura is with Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (Email: chintha@ece.ualberta.ca).

Manuscript received February 13, 2017; revised July 18, 2017, October 23, 2017; accepted November 16, 2017.

their contributions to the LLR of the said bit, and a subset of those symbols with less contributions are approximated as Gaussian variables [19].

In this paper, we consider efficient iterative detection/decoding for MIMO-OFDM based LTE-A systems. As in the aforementioned works, the burden of LLR computations is the critical factor. The method of LLR computation is what differentiates the various algorithms in [12], [14]–[17]. The algorithms in [14]–[16] are equivalent to using a Gaussian approximation. However, it is ineffective for high order modulations such as 64 quadrature amplitude modulation (QAM), which is needed in LTE-A. We thus propose non-Gaussian approximation by exploiting the probability mass function (PMF) of the constellation points into a unified form followed by relaxing the discrete variable to take continuous values. By integrating over the resultant non-Gaussian probability density function (PDF), we approximate the LLR value. Next, we propose combining the K -best algorithm and the Gaussian or non-Gaussian approximation to take advantage of both. In the proposed modified K -best algorithms, we generate a list of K lattice points. There are two major differences between our proposed algorithm and that in [12]. First, we use sum-log rather than max-log to compute the LLR value by summing over all lattice points in the list. Second, the list generation method is different. To generate the list, we adopt the *a priori* probability, which is approximated using Gaussian and non-Gaussian distributions. We also discuss several variations of the basic algorithms and efficient implementation in commercial MIMO-OFDM receivers. In addition to non-Gaussian approximation for high order constellations, we propose several list-decoding algorithms while all lattice points are used in [19] for the symbols not using Gaussian approximation. Different from [20] where a zero-forcing detector is used to approximate max-log LLR and Gaussian approximation is used to study the distribution of LLR, we use Gaussian and non-Gaussian distributions for sum-log listing decoding. Simulation results using an LTE simulator show that the proposed algorithms outperform the existing ones, especially for higher-order modulations.

The rest of the paper is organized as follows. In Section II, the MIMO-OFDM system model is presented. In Section III, the iterative receiver structure is outlined and several existing algorithms are reviewed. A non-Gaussian approximation based algorithm and several modified K -best algorithms are given in Section IV. Simulation results are provided in Section V and Section VI draws the conclusions.

II. SYSTEM MODEL

We consider a MIMO-OFDM system with M transmit and N receive antennas. The system has N_s subcarriers in an OFDM block. There are M data streams to be transmitted¹. The constellation \mathcal{Q}_m is applied on stream m , where C_m is the number of bits per constellation symbol. The incoming bits of

¹Note that in LTE standard, when there are 4 antennas, there are still two data streams rather than four data streams but the coded symbols of each stream are sent on different antennas. The discussion in this paper can be readily extended to this case. To simplify discussion, we use the current notation.

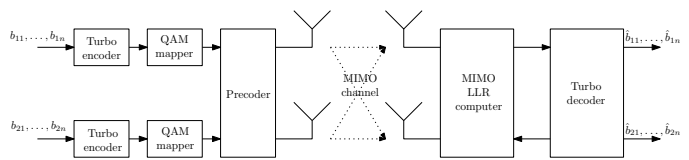


Fig. 1. The diagram of a MIMO-OFDM system.

each stream m of length $N_s C_m R_m$, $m=1, \dots, M$, is encoded using a channel code (typically a convolutional or turbo code) of rate R_m , resulting in a bit vector \mathbf{b}_m . The encoded bits are converted into symbols using a mapping function $x_{i,m} = \mathcal{M}_m(\mathbf{b}_m((i-1)C_m+1:iC_m))$ (e.g., Gray mapping and set partitioning mapping), $i=0, \dots, N_s-1$, where $x_{i,m}$ is the symbol to be transmitted over subcarrier i and antenna m . The inverse discrete Fourier transform (IDFT) of the data block $x_{0,m}, \dots, x_{N_s-1,m}$ yields the time domain sequence, i.e.,

$$X_{j,m} = \frac{1}{\sqrt{N_s}} \sum_{i=0}^{N_s-1} x_{i,m} e^{j2\pi ij/N_s}, j=0, \dots, N_s-1. \quad (1)$$

After adding cyclic prefix (CP), the received signal on carrier i after DFT can be written in vector form as

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i + \mathbf{w}_i, i=0, \dots, N_s-1, \quad (2)$$

where $\mathbf{x}_i = [x_{i,1}, \dots, x_{i,M}]^T$ is the transmitted signal on M transmit antennas, $\mathbf{y}_i = [y_{i,1}, \dots, y_{i,N}]^T$ and $\mathbf{w}_i = [w_{i,1}, \dots, w_{i,N}]^T$ are the received signal and the additive noise on N receive antennas, respectively, and \mathbf{H}_i denotes the $N \times M$ channel matrix. Each entry in \mathbf{w}_i has mean zero and variance σ^2 . The MIMO-OFDM system is depicted in Fig. 1. Each subcarrier (2) is a distinct MIMO system. In this paper except in Section IV-B, we will neglect the subscript i in (2) for brevity. We assume channel matrix $(\mathbf{H}_i, i=0, 1, \dots, N_s-1)$ is known at the receiver which can be estimated using algorithms in [21]–[26].

III. EXISTING ITERATIVE DETECTION AND DECODING ALGORITHMS

A. Basic Concepts

The optimal joint detector and decoder must compute the likelihood of each bit given the received signals $\mathbf{y}_0, \dots, \mathbf{y}_{N_s-1}$ on all subcarriers. However, this is computationally infeasible, the algorithms in [12], [14]–[16], [18] thus use the “turbo principle”, where information is exchanged between the detector and decoder iteratively. In this paper, we focus on how to generate extrinsic information at each subcarrier using the received signals on this subcarrier and the *a priori* information on each bit from the channel decoder. The generated extrinsic information on all subcarriers is then input to the soft-in-soft-out channel decoder for the next iteration of decoding and detection. We briefly review several such existing algorithms and unravel their relationships, which will motivate our algorithms in Section IV.

B. Iterative Detection and Decoding

The *a priori* probability (APP) is usually expressed as a LLR value (L -value). The sign of L -value indicates if a bit is zero or one. The magnitude of L -value indicates the reliability of a bit decision; e.g., magnitude close to zero shows an unreliable bit. We represent the logical zero for a bit by $b_i = -1$ and logical one by $b_i = +1$, respectively. Given the APP from the channel decoder, the *a posteriori* LLR value of the bit b_i conditioned on the received vector \mathbf{y} is

$$L(b_i|\mathbf{y}) = \log \frac{\Pr(b_i = +1|\mathbf{y})}{\Pr(b_i = -1|\mathbf{y})}. \quad (3)$$

We can rewrite (3) using Bayes' theorem as

$$L(b_i|\mathbf{y}) = \log \frac{\sum_{\mathbf{x} \in \mathcal{X}_{i,+1}} \Pr(\mathbf{y}|\mathbf{x}) \prod_{j=0, j \neq i}^{M-1} \Pr(b_j = \mathcal{B}_j(\mathbf{x}))}{\sum_{\mathbf{x} \in \mathcal{X}_{i,-1}} \Pr(\mathbf{y}|\mathbf{x}) \prod_{j=0, j \neq i}^{M-1} \Pr(b_j = \mathcal{B}_j(\mathbf{x}))} \quad (4)$$

$$= \underbrace{\log \frac{\Pr(b_i = +1)}{\Pr(b_i = -1)}}_{L_A(b_i)} + \underbrace{\log \frac{\sum_{\mathbf{x} \in \mathcal{X}_{i,+1}} \Pr(\mathbf{y}|\mathbf{x}) \prod_{j=0, j \neq i}^{M-1} \Pr(b_j = \mathcal{B}_j(\mathbf{x}))}{\sum_{\mathbf{x} \in \mathcal{X}_{i,-1}} \Pr(\mathbf{y}|\mathbf{x}) \prod_{j=0, j \neq i}^{M-1} \Pr(b_j = \mathcal{B}_j(\mathbf{x}))}}_{L_E(b_i|\mathbf{y})},$$

where $\mathcal{X}_{i,+1}$ and $\mathcal{X}_{i,-1}$ are the set of $2^{\sum_{m=1}^{M-1} C_m - 1}$ vectors such that the i -th bit is $+1$ or -1 , respectively; Namely $\mathcal{X}_{i,\pm 1} = \{\mathbf{x} | \mathcal{M}(\mathbf{b}) = \mathbf{x}, b_i = \pm 1\}$, $\mathbf{b} = \mathcal{B}(\mathbf{x})$ is the inverse mapping of $\mathbf{x} = \mathcal{M}(\mathbf{b})$ and $\mathcal{B}_j(\mathbf{x})$ is the j -th bit of $\mathcal{B}(\mathbf{x})$.

In the case of Gaussian channel (2), we can further write $L(b_i|\mathbf{y})$ as

$$L(b_i|\mathbf{y}) = \log \frac{\sum_{\mathbf{x} \in \mathcal{X}_{i,+1}} \exp \left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2}{\sigma^2} + \sum_{m=1}^{M-1} \sum_{j=0, j \neq i, b_j=1}^{C_m-1} L_A(b_j) \right)}{\sum_{\mathbf{x} \in \mathcal{X}_{i,-1}} \exp \left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2}{\sigma^2} + \sum_{m=1}^{M-1} \sum_{j=0, j \neq i, b_j=1}^{C_m-1} L_A(b_j) \right)} + L_A(b_i). \quad (5)$$

To find the L -value for each b_i using (5), a search is needed over $2^{\sum_{m=1}^{M-1} C_m - 1}$ terms, which is exponential in the total number of bits $\sum_{m=1}^{M-1} C_m$. Since this complexity is prohibitive, a list sphere decoder (LSD) is adopted in [12].

We consider a commercial iterative decoding and demodulation architecture for MIMO-OFDM in Fig. 2 adapted from [27]. $L_E(b_i|\mathbf{y})$ computation in (4) corresponds to the demodulator in Fig. 2. Besides iterative decoding in turbo decoder, there is an outerloop iterative procedure to exchange $L_A(b_i)$ between demodulator and turbo decoder. In LTE-A, rate matching is used to match the number of bits in transport block to the number of bits that can be transmitted in the given allocation [1]. Rate matching involves many things including sub-block interleaving, bit collection and pruning.

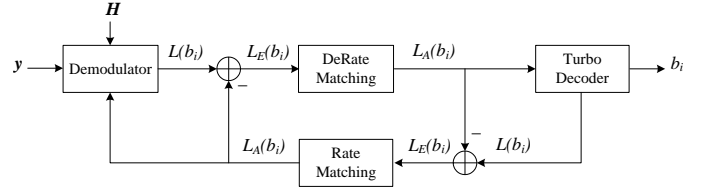


Fig. 2. Iterative decoding and demodulation architecture for MIMO-OFDM.

Derate matching is the reverse operation of rating matching. We can find the corresponding symbol \mathbf{x} for b_i in (4) through rate matching.

C. Iterative Detection and Decoding with Gaussian Approximation

Here we review a Gaussian approximation for iterative detection and decoding and relate different existing algorithms, where we can gain insights to develop new algorithms. To this end, note that we can rewrite (4) as

$$L(b_i|\mathbf{y}) = \log \frac{\sum_{x_m \in \mathcal{X}_{i,+1}^m} \Pr(x_m) \sum_{\mathbf{x}_{-m}} \Pr(\mathbf{y}|\mathbf{x}_{-m}, x_m) \Pr(\mathbf{x}_{-m})}{\sum_{x_m \in \mathcal{X}_{i,-1}^m} \Pr(x_m) \sum_{\mathbf{x}_{-m}} \Pr(\mathbf{y}|\mathbf{x}_{-m}, x_m) \Pr(\mathbf{x}_{-m})}, \quad (6)$$

where x_m denotes the symbol that b_i belongs to, \mathbf{x}_{-m} denotes the vector that contains all entries of \mathbf{x} except for the m -th entry, and $\mathcal{X}_{i,\pm 1}^m$ ($\mathcal{X}_{i,\pm 1}^m$) is the set of 2^{C_m-1} vectors such that b_i is $+1$ (-1). For any given x_m , we need to compute $\sum_{\mathbf{x}_{-m}} \Pr(\mathbf{y}|\mathbf{x}_{-m}, x_m) \Pr(\mathbf{x}_{-m})$. A suboptimal approach is to replace the summation over \mathbf{x}_{-m} with an integration over a continuous distribution. One common choice is the Gaussian distribution. If we assume the entries of \mathbf{x}_{-m} are independent Gaussian random variables with mean and variance

$$\mu_{m'} = E\{x_{m'}\}, \nu_{m'}^2 = E\{|x_{m'}|^2\} - E^2\{x_{m'}\}, \quad (7)$$

$m' = 1, \dots, M$, $m' \neq m$ and Gaussian channel model (2) is used, then we have

$$\Pr(\mathbf{y}|x_m) = \sum_{\mathbf{x}_{-m}} \Pr(\mathbf{y}|\mathbf{x}_{-m}, x_m) \Pr(\mathbf{x}_{-m})$$

$$\approx \int_{-\infty}^{+\infty} \Pr(\mathbf{y}|\mathbf{x}_{-m}, x_m) f(\mathbf{x}_{-m}) d\mathbf{x}_{-m} \quad (8)$$

$$\propto \exp \left(-(\mathbf{y} - \mathbf{H}_{-m} \boldsymbol{\mu}_{-m} - \mathbf{h}_m x_m)^H \mathbf{R}_m^{-1} \right. \\ \left. \times (\mathbf{y} - \mathbf{H}_{-m} \boldsymbol{\mu}_{-m} - \mathbf{h}_m x_m) \right),$$

where the integral is from $-\infty$ to $+\infty$ in each dimension, \mathbf{H}_{-m} contains the columns of \mathbf{H} except for the m -th column, \mathbf{h}_m is the m -th column of \mathbf{H} , $\boldsymbol{\mu}_{-m} = [\mu_1, \dots, \mu_{m-1}, \mu_{m+1}, \dots, \mu_M]^T$,

$$\mathbf{R}_m = \mathbf{H}_{-m} \text{diag} \{ \nu_1^2, \dots, \nu_{m-1}^2, \nu_{m+1}^2, \dots, \nu_M^2 \} \mathbf{H}_{-m}^H + \sigma^2 \mathbf{I}_N, \quad (9)$$

and \mathbf{I}_N is an $N \times N$ identity matrix. To maximize $\Pr(\mathbf{y}|x_m)$ in (8) we only need to search through 2^{C_m} possible x_m 's. The complexity of computing LLR reduces from $2^{\sum_{m=1}^M C_m}$ by searching through all possible constellation combinations to 2^{C_m} .

It can be shown that different algorithms [14]–[16] are equivalent to the Gaussian approximation (8).

IV. PROPOSED ITERATIVE DETECTION AND DECODING ALGORITHM

In this section, we assume all transmit symbols are from a square-QAM constellation, as in LTE-A [1], but our work can also readily extend to arbitrary constellations. To achieve a good performance-complexity tradeoff especially for high order constellations such as 64-QAM, we can try two different avenues. First, by following [12], we may compute L-values with the LSD. However, the number of summands in (4) is exponential. Since LSD depends on the maximally possible searching time, its detection throughput is variable, which would demand I/O buffers adding an extra overhead in a practical system [13]. Second, we may use the Gaussian approximation to get the closed-form (8) but its performance is not good for higher order modulations [16]. Due to these reasons, our goal is to combine LSD and continuous integral approximation of discrete summation by taking the advantage of both. Before doing that, for square-QAM, we write (2) as a real system, i.e.,

$$\underbrace{\begin{bmatrix} \Re(\mathbf{y}_i) \\ \Im(\mathbf{y}_i) \end{bmatrix}}_{\tilde{\mathbf{y}}_i} = \underbrace{\begin{bmatrix} \Re(\mathbf{H}_i) & -\Im(\mathbf{H}_i) \\ \Im(\mathbf{H}_i) & \Re(\mathbf{H}_i) \end{bmatrix}}_{\tilde{\mathbf{H}}_i} \underbrace{\begin{bmatrix} \Re(\mathbf{x}_i) \\ \Im(\mathbf{x}_i) \end{bmatrix}}_{\tilde{\mathbf{x}}_i} + \underbrace{\begin{bmatrix} \Re(\mathbf{w}_i) \\ \Im(\mathbf{w}_i) \end{bmatrix}}_{\tilde{\mathbf{w}}_i}, \quad (10)$$

where $i=0, \dots, N_s-1$ and $\Re(x)$ and $\Im(x)$ denote the real part and imaginary part of x , respectively and the entries of $\tilde{\mathbf{x}}_i$ are from pulse-amplitude modulation (PAM) constellations. With a slight abuse of notations, we still use (2) to represent the real system with the entries of \mathbf{x}_i from PAM.

A. Iterative Detection and Decoding with Non-Gaussian Approximation

To motivate our non-Gaussian approximation, we start with the Binary Phase Shift Keying (BPSK), i.e., $X \in \{+1, -1\}$. Let $\Pr(X=+1)=p$ and $\Pr(X=-1)=1-p$, where p is from $L_A(b_i)$ in (4). We can write this PMF into a single equation as

$$\Pr(X=x) = p \left(\frac{x+1}{2}\right)^2 (1-p) \left(\frac{x-1}{2}\right)^2, \quad x = \pm 1. \quad (11)$$

A natural continuous approximation of this PMF is obtained by relaxing x to be a real number with a scaling factor to keep $\int \Pr(X=x)dx=1$. Note that there are several choices of the PMF (11). For example, we may choose $\Pr(X=x) = p \frac{x+1}{2} (1-p) \frac{1-x}{2}$. But this function will diverge when x goes to ∞ , which is undesirable. We may also choose $\Pr(X=x) = p \frac{|x+1|}{2} (1-p) \frac{|x-1|}{2}$. But this function is not amenable to closed-form integrations.

The PMF in (11) can be extended to higher modulations. For a given modulation \mathcal{Q} with $\Pr(X=x_i)=p_i$ and $\sum p_i=1$, we can write the PMF into a single equation as

$$\Pr(X=x) = \prod_{x_i \in \mathcal{Q}} p_i \frac{\prod_{x_j \in \mathcal{Q}, x_j \neq x_i} (x-x_j)^2}{\prod_{x_j \in \mathcal{Q}, x_j \neq x_i} (x_i-x_j)^2} = \exp \left(\sum_{l=0}^{2(|\mathcal{Q}|-1)} a_l x^l \right), \quad (12)$$

where $x \in \mathcal{Q}$ and a_l is a constant depending on the constellation, e.g.,

$$a_l = \frac{d^l \left(\sum_{x_i \in \mathcal{Q}} \log p_i \frac{\prod_{x_j \in \mathcal{Q}, x_j \neq x_i} (x-x_j)^2}{\prod_{x_j \in \mathcal{Q}, x_j \neq x_i} (x_i-x_j)^2} \right)}{dx^l} \Big|_{x=0}.$$

The PDF can be obtained by relaxing x to be a real number. When $|\mathcal{Q}| > 2$, if we use (12) directly in (8), then the integral involves a polynomial greater than the second order in the exponent, whose closed-form may be hard to obtain. Therefore, we approximate the PMF (12) with a second order polynomial in the exponent for any \mathcal{Q} , i.e.,

$$\Pr(X=x) = \exp(- (c+2rx+ax^2)), \quad (13)$$

where c, r, a are constants. Note that the Gaussian distribution is a special case of (13), which contains a nonnegative a . The coefficients a, r, c are found by solving

$$\min_{a,r,c} \sum_i \omega_i (\exp(- (c+2rx_i+ax_i^2)) - p_i)^2, \quad (14)$$

where $\omega_i \geq 0$ is a weight for symbol x_i . We may choose uniform weights or assign higher weights to the symbols with large probability. At each iteration of decoding, we can get p from $L_A(b_i)$ in (4) and solve (14) using convex optimization.

The integration in (8) is from $-\infty$ to $+\infty$, which may distort the LLR value. Practical constellations typically contain finite alphabets, e.g., 2D-PAM is $\{-2D+1, -2D+3, \dots, 2D-3, 2D-1\}$. We can integrate from $-U$ to U instead. Some possible choices of U are $2D$ or $2D-1+\sigma$. With (13) and $U=2D$, we can write (8) as

$$\begin{aligned} & \Pr(\mathbf{y}|x_m) \\ & \stackrel{(a)}{\approx} \sum_{\mathbf{x}_{-m}} \Pr(\mathbf{y}|\mathbf{x}_{-m}, x_m) \prod_{k \neq m, 2d_k-1=x_k} \int_{2d_k-2}^{2d_k} f(x_k) dx_k \\ & \stackrel{(b)}{\approx} \int_{-U}^{+U} \Pr(\mathbf{y}|\mathbf{x}_{-m}, x_m) f(\mathbf{x}_{-m}) d\mathbf{x}_{-m} \\ & \propto \exp \left(- \frac{\|\mathbf{y} - \mathbf{h}_m x_m\|^2}{\sigma^2} \right) \\ & \quad \times \int_{-U}^{+U} \exp \left(- 2 \underbrace{\left(\mathbf{r}_{-m}^T - \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{h}_m x_m)^T \mathbf{H}_{-m} \right)}_{\mathbf{b}_{-m}^T} \mathbf{x}_{-m} \right. \\ & \quad \left. - \underbrace{\mathbf{x}_{-m}^T \left(\mathbf{A}_{-m} + \frac{\mathbf{H}_{-m}^T \mathbf{H}_{-m}}{\sigma^2} \right) \mathbf{x}_{-m}}_{\mathbf{R}_m} \right) d\mathbf{x}_{-m}, \end{aligned} \quad (15)$$

where $\mathbf{r}_{-m} = [r_1, \dots, r_{m-1}, r_{m+1}, \dots, r_M]^T$ and $\mathbf{A}_{-m} = \text{diag}\{a_1, \dots, a_{m-1}, a_{m+1}, \dots, a_M\}$, $r_{m'}$ and $a_{m'}$ are obtained from (14). In (a), we have approximate the PMF using a finite integration over PDF as

$$Pr(x_k = 2d_k - 1) \approx \int_{2d_k-2}^{2d_k} f(x_k) dx_k.$$

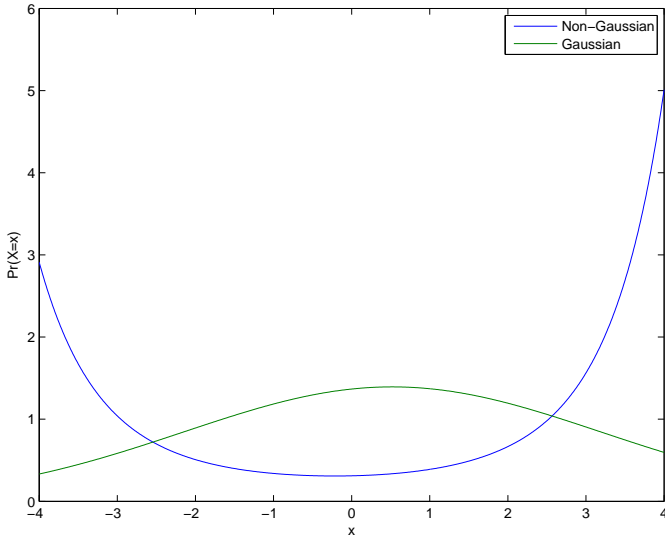


Fig. 3. PDF comparison between non-Gaussian and Gaussian approximations for 4-PAM with $\mathcal{Q}=\{-3,-1,1,3\}$ when $\Pr(X=-3)=0.32$, $\Pr(X=-1)=0.08$, $\Pr(X=1)=0.12$, and $\Pr(X=3)=0.48$.

In (b), we have applied the mean value theorem for definite integrals where

$$\begin{aligned} & \int_{2d_k-2}^{2d_k} \Pr(\mathbf{y}|\mathbf{x}_{-(m,k)}, x_m, x_k) f(x_k) dx_k \\ &= \Pr(\mathbf{y}|\mathbf{x}_{-(m,k)}, x_m, \tilde{x}_k) \int_{2d_k-2}^{2d_k} f(x_k) dx_k \\ &\approx \Pr(\mathbf{y}|\mathbf{x}_{-m}, x_m) \int_{2d_k-2}^{2d_k} f(x_k) dx_k, \end{aligned}$$

where $\mathbf{x}_{-(m,k)}$ denotes the vector that contains all entries of \mathbf{x} except for the m -th and k -th entries.

Comparing (15) with (8), we see that there are two main differences. First, \mathbf{r}_{-m} and \mathbf{A}_{-m} are not from the matched mean and variance but from matching the PMF directly. Second, the integral is from $-U$ to U .

Let the singular value decomposition of \mathbf{R}_m be $\mathbf{V}^T \mathbf{\Lambda} \mathbf{V}$ and $\mathbf{g}(x_m) = \mathbf{V} \mathbf{b}_{-m}$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_{M-1}\}$. We change variables by defining $\mathbf{z} = \mathbf{V} \mathbf{x}_{-m}$. However, the integration region of \mathbf{z} is an $M-1$ dimensional polytope, making the integral hard to compute. For simplicity, we enlarge the integration region by setting a bound $Z_i = U \sum_{j=1}^{M-1} |V_{i,j}|$ for dimension i . We can then approximate (15) as

$$\begin{aligned} \Pr(\mathbf{y}|x_m) &\propto \exp\left(-\frac{\|\mathbf{y} - \mathbf{h}_m x_m\|^2}{\sigma^2}\right) \\ &\times \prod_{i=1}^{M-1} \int_{-Z_i}^{+Z_i} \exp(-2g_i(x_m)z_i - \lambda_i z_i^2) dz_i, \end{aligned} \quad (16)$$

where $g_i(x_m)$ is the i -th entry of $\mathbf{g}(x_m) = \mathbf{V} \mathbf{b}_{-m}$. When $\lambda_i > 0$, we can rewrite the integral in (16) into a Q-function. When $\lambda_i < 0$, we can compute the integral using numerical methods.

To compare Gaussian and non-Gaussian approximations, we consider 4-PAM with $\mathcal{Q}=\{-3,-1,1,3\}$. Two bits (b_1, b_2) are mapped to \mathcal{Q} via gray mapping $(0,1) \rightarrow -3, (0,0) \rightarrow$

$-1, (1,0) \rightarrow 1, (1,1) \rightarrow 3$. After several decoding iterations, we may have $\Pr(b_1=1)=0.6$ and $\Pr(b_2=1)=0.8$, which gives $\Pr(X=-3)=0.32$, $\Pr(X=-1)=0.08$, $\Pr(X=1)=0.12$, and $\Pr(X=3)=0.48$. We compare the PDF's of non-Gaussian (14) and Gaussian approximations in Fig. 3. The areas between $2i-2$ and $2i$ (as an approximation of the symbol probability $P(X=2i-1)$), $i=-1,0,1,2$ are 0.31, 0.09, 0.10, 0.49 for the non-Gaussian approximation and 0.15, 0.29, 0.33, 0.23 for the Gaussian approximation. Clearly the Gaussian approximation does not match the discrete distribution when some bits are unreliable. This problem is a lot more severe when each symbol contains more than 2 bits, which may be the reason that the probabilistic data association (PDA) performs poorly for high order modulations as observed in [16].

For general bit mappings, due to the constraint on the second order polynomial, (13) may not fit the PMF for all symbols in large constellations. Moreover, it is observed in [18] that Gray mapping does not perform well in iterative joint detection and decoding. Other mappings such as set partitioning mapping may perform better. For set partitioning mapping, we can resolve the polynomial order constraint via constellation decomposition. Let $b_i = +1$ for logical one and $b_i = -1$ for logical zero. Let C be a positive integer. We can write the set partitioning mapping for 2^C -PAM ($\mathcal{Q} = \{-2^C+1, -2^C+3, \dots, 2^C-3, 2^C-1\}$) as

$$x = \sum_{i=0}^{C-1} 2^i b_i = \mathbf{d}^T \mathbf{b}, \quad (17)$$

where $\mathbf{d} = [1, 2, \dots, 2^{C-1}]^T$ and $\mathbf{b} = [b_0, \dots, b_{C-1}]^T$. Note that we have ignore the scaling factor in the modulation to keep unit average power. As each entry of b takes BPSK, the continuous approximation to the PMF is given in (11), where (13) is exact with

$$a = c = -\frac{1}{4}(\log p + \log(1-p)), \quad b = -\frac{1}{4}(\log p - \log(1-p)). \quad (18)$$

Define $\tilde{\mathbf{H}}_{-m} = \mathbf{H}_{-m} \text{diag}\{\mathbf{d}_1^T, \dots, \mathbf{d}_{m-1}^T, \mathbf{d}_{m+1}^T, \dots, \mathbf{d}_M^T\}$ and $\tilde{\mathbf{b}}_{-m} = [\mathbf{b}_1^T, \dots, \mathbf{b}_{m-1}^T, \mathbf{b}_{m+1}^T, \dots, \mathbf{b}_M^T]^T$. By replacing \mathbf{H}_{-m} with $\tilde{\mathbf{H}}_{-m}$ and \mathbf{x}_{-m} with $\tilde{\mathbf{b}}_{-m}$ in (15), we can obtain a similar form as (16). The only difference is that the new eigenvalue $\tilde{\lambda}_i$ is nonnegative as a in (18) is nonnegative. We can thus rewrite (16) as

$$\begin{aligned} \Pr(\mathbf{y}|x_m) &\propto \exp\left(-\frac{\|\mathbf{y} - \mathbf{h}_m x_m\|^2}{\sigma^2} + \sum_{i=1}^{M-1} \frac{\tilde{g}_i^2(x_m)}{\lambda_i}\right) \\ &\times \prod_{i=1}^{M-1} \left(Q\left(\sqrt{2\tilde{\lambda}_i} Z_i + \sqrt{\frac{2}{\tilde{\lambda}_i}} \tilde{g}_i(x_m)\right) \right. \\ &\quad \left. - Q\left(-\sqrt{2\tilde{\lambda}_i} Z_i + \sqrt{\frac{2}{\tilde{\lambda}_i}} \tilde{g}_i(x_m)\right) \right), \end{aligned} \quad (19)$$

where $\tilde{\lambda}_i$ and $\tilde{g}_i(x_m)$ are defined similarly as in (16) by replacing \mathbf{H}_{-m} with $\tilde{\mathbf{H}}_{-m}$. This approach may be extended to other similar bit mappings resulting in constellation partitioning.

B. K -Best algorithm with Discrete Distribution Approximation

In [13], a K -best iterative MIMO receiver is proposed. The K -best algorithm offers the advantages of parallel implementation and fixed throughput regardless of the operating SNR. However, the original K -best [13] does not consider the impact of APP on the K best paths. Here, we consider a K -best algorithm which will generate different sets of K -best paths at each iteration by using a distributional approximation.

The LSD only considers the maximum term among all the $2^{\sum_{m=1}^M C_m - 1}$ terms in (4), and the list is generated by using $\Pr(\mathbf{y}|x_1, \dots, x_M)$ only without the *a priori* information $\Pr(x_{m'})$, $m'=1, \dots, M$. Moreover, when the LSD comes to the i -th data stream, it only checks the symbols satisfying

$$\left(\tilde{y}_i - R_{i,i}x_i - \sum_{j=i+1}^M R_{i,j}\tilde{x}_j\right)^2 + \sum_{j=i+1}^M \left(\tilde{y}_j - \sum_{l=j}^M R_{j,l}\tilde{x}_l\right)^2 \leq r^2, \quad (20)$$

where \tilde{y}_i is the i -th entry of $\mathbf{Q}^H \mathbf{y}$, $R_{i,j}$ is the (i,j) -th entry of \mathbf{R} with QR decomposition $\mathbf{H} = \mathbf{Q}\mathbf{R}$, and \tilde{x}_j is the trial value of x_j . Using (20) is myopic and does not consider the effect of choosing x_i on the data streams $1, \dots, i-1$. On the other hand, the Gaussian approximation algorithm (Section III-C) considers the summation in (4), but is not good for high-order constellations. We next develop an algorithm that unifies both approaches. The key idea is to use the Gaussian approximation or non-Gaussian approximation as a metric to guide the search, by taking into account the impact of stream i on streams $1, \dots, i-1$.

As with LSD, we also want to find a list of K lattice points. But there are three key differences from LSD. First, we try to find a list $\mathcal{L}_{i,\pm 1}$ containing K points for each $b_i = \pm 1$ rather than a common list for both b_i 's. The LLR value of b_i in (4) is then approximated as

$$L(b_i|\mathbf{y}) \approx \log \frac{\sum_{\mathbf{x} \in \mathcal{L}_{i,+1}} \Pr(\mathbf{x}|\mathbf{y})}{\sum_{\mathbf{x} \in \mathcal{L}_{i,-1}} \Pr(\mathbf{x}|\mathbf{y})}. \quad (21)$$

Second, we use sum-log rather than max-log used in LSD. The third difference lies in the way we generate the list. We want to find K lattice points $\mathbf{x} \in \mathcal{X}_{i,\pm 1}$ such that $\Pr(\mathbf{x}|\mathbf{y})$ is maximized rather than maximizing $\Pr(\mathbf{y}|\mathbf{x})$, where the *a priori* information is exploited in the former case.

In the initial step of the proposed method, we assume that b_i belongs to data stream m and check each $\tilde{x}_m \in \mathcal{X}_{i,\pm 1}^m$ to find the K candidates such that $\Pr(x_m|\mathbf{y})$ is maximized. Then we add m into a set \mathcal{V} containing the streams that have been checked. We can write $\Pr(\tilde{x}_m|\mathbf{y})$ as

$$\Pr(\tilde{x}_m|\mathbf{y}) \propto \sum_{\mathbf{x}_{-m}} \Pr(\mathbf{y}|\mathbf{x}_{-m}, \tilde{x}_m) \Pr(\mathbf{x}_{-m}). \quad (22)$$

Direct computation of (22) requires $2^{\sum_{m'=1, m' \neq m}^M C_{m'}}$ summations, which may be computationally prohibitive. We replace the summation in (22) as an integral

$$\Pr(\tilde{x}_m|\mathbf{y}) \propto \int \Pr(\mathbf{y}|\mathbf{x}_{-m}, \tilde{x}_m) f(\mathbf{x}_{-m}) d\mathbf{x}_{-m}, \quad (23)$$

where $f(\mathbf{x}_{-m})$ is the matched PDF of \mathbf{x}_{-m} , which could be either Gaussian or non-Gaussian. With Gaussian approximation (non-Gaussian approximation can be derived similarly), we have

$$\Pr(\tilde{x}_m|\mathbf{y}) \propto \exp\left(-(\mathbf{y} - \mathbf{H}_{-m}\boldsymbol{\mu}_{-m} - \mathbf{h}_m\tilde{x}_m)^H \times \mathbf{R}_{-m}^{-1}(\mathbf{y} - \mathbf{H}_{-m}\boldsymbol{\mu}_{-m} - \mathbf{h}_m\tilde{x}_m)\right), \quad (24)$$

where $\boldsymbol{\mu}_{-m}$ and \mathbf{R}_{-m} are defined in (8). The K \tilde{x}_m 's with the largest $\Pr(\tilde{x}_m|\mathbf{y})$ are added into a list \mathcal{L} , which is initialized to be \emptyset .

The process then goes for x_1, x_2, \dots, x_M . Before it reaches x_j , $j \neq m$, we have $\mathcal{V} = \{m, 1, \dots, j-1\}$ and the list \mathcal{L} contains K candidates, each of which has the form $\mathbf{z}_{\mathcal{V}} = [x_m, x_1, \dots, x_{j-1}]^T$. For each $\mathbf{z}_{\mathcal{V}} \in \mathcal{L}$ and $\tilde{x}_j \in \mathcal{Q}_j$, we compute $\Pr(\mathbf{z}_{\mathcal{V}}, \tilde{x}_j|\mathbf{y})$. Among the resultant $K|\mathcal{Q}_j|$ candidates, we only choose K of them such that $\Pr(\mathbf{z}_{\mathcal{V}}, \tilde{x}_j|\mathbf{y})$ is maximized, update the list \mathcal{L} with the K chosen vectors, and add j into \mathcal{V} . We can approximate $\Pr(\mathbf{z}_{\mathcal{V}}, \tilde{x}_j|\mathbf{y})$ in the same way as in (23). In case of Gaussian approximation, we have

$$\Pr(\mathbf{z}_{\mathcal{V}}, \tilde{x}_j|\mathbf{y}) \propto \exp\left(-(\mathbf{y} - \mathbf{H}_{-\{\mathcal{V},j\}}\boldsymbol{\mu}_{-\{\mathcal{V},j\}} - \mathbf{H}_{\mathcal{V}}\mathbf{z}_{\mathcal{V}} - \mathbf{h}_j\tilde{x}_j)^H \times \mathbf{R}_{\{\mathcal{V},j\}}^{-1}(\mathbf{y} - \mathbf{H}_{-\{\mathcal{V},j\}}\boldsymbol{\mu}_{-\{\mathcal{V},j\}} - \mathbf{H}_{\mathcal{V}}\mathbf{z}_{\mathcal{V}} - \mathbf{h}_j\tilde{x}_j)\right), \quad (25)$$

where $\boldsymbol{\mu}_{-\mathcal{A}}$ constitutes the entries of $\boldsymbol{\mu}$ that are not in \mathcal{A} , $\mathbf{H}_{-\mathcal{A}}$ consists of the columns of \mathbf{H} that are not in \mathcal{A} , and

$$\mathbf{R}_{\{\mathcal{V},j\}} = \mathbf{H}_{-\{\mathcal{V},j\}} \text{diag}\left\{\boldsymbol{\nu}_{-\{\mathcal{V},j\}}^2\right\} \mathbf{H}_{-\{\mathcal{V},j\}}^H + \sigma^2 \mathbf{I}_N. \quad (26)$$

The process ends when $j=M$. After obtaining a list for $b_i = \pm 1$ respectively, we use (21) to compute the LLR value.

C. Extensions

The basic K -best algorithm (Section IV-B), named as sum-algorithm, can be extended in several ways.

Max-Algorithm: Different from the sum-algorithm where $\Pr(\mathbf{x}_{\mathcal{V}}, \tilde{x}_j|\mathbf{y})$ is maximized consecutively, in the max-algorithm, we maximize $\Pr(\mathbf{x}|\mathbf{y})$ directly. At the first step, for each $\tilde{x}_m \in \mathcal{X}_{i,\pm 1}^m$, we find the corresponding $\tilde{\mathbf{x}}_{-m}$ such that

$$\begin{aligned} \tilde{\mathbf{x}}_{-m} &= \arg \max_{\mathbf{x}_{-m} \in \mathcal{X}^{-m}} \Pr(\tilde{x}_m, \mathbf{x}_{-m}|\mathbf{y}) \\ &= \arg \max_{\mathbf{x}_{-m} \in \mathcal{X}^{-m}} \Pr(\mathbf{y}|\tilde{x}_m, \mathbf{x}_{-m}) \Pr(\tilde{x}_m, \mathbf{x}_{-m}), \end{aligned} \quad (27)$$

where \mathcal{X}^{-m} includes all possible lattice points. We put K \tilde{x}_m into the list \mathcal{L} such that $\Pr(\tilde{x}_m, \tilde{\mathbf{x}}_{-m}|\mathbf{y})$ is the largest and add m into a set \mathcal{V} . Similarly we replace $\Pr(\tilde{x}_m, \mathbf{x}_{-m})$ with its continuous Gaussian or non-Gaussian approximations and relax the discrete set \mathcal{X}^{-m} into a continuous set \mathcal{C}^{-m} .

When \mathcal{C}^{-m} is bounded, the boundary on x_j is defined by the largest and the smallest elements in \mathcal{Q}_j . For example, when $\mathcal{Q}_j = \{-3, -1, 1, 3\}$, we choose $-3 \leq x_j \leq 3$. When the non-Gaussian approximation (13) is used, we need to solve

$$\begin{aligned} \hat{\mathbf{x}}_{-m} &= \arg \min_{\mathbf{x}_{-m} \in \mathcal{C}^{-m}} \|\mathbf{y} - \mathbf{H}_{-m}\mathbf{x}_{-m} - \mathbf{h}_m\tilde{x}_m\|^2 \\ &\quad + 2\sigma^2 \mathbf{r}_{-m}^T \mathbf{x}_{-m} + \sigma^2 \mathbf{x}_{-m}^T \mathbf{A}_{-m} \mathbf{x}_{-m}. \end{aligned} \quad (28)$$

As (28) is quadratic in \mathbf{x}_{-m} , when the objective function of (28) is convex, $\hat{\mathbf{x}}_{-m}$ can be found using convex optimization tools. If not, the problem is shown to be NP-hard [28]. We thus find a local minimum using numerical methods such as Newton's method around

$$\operatorname{argmin}_{\mathbf{x}_{-m} \in \mathcal{C}^{-m}} \|\mathbf{y} - \mathbf{H}_{-m}\mathbf{x}_{-m} - \mathbf{h}_m \tilde{x}_m\|^2.$$

We can set $\tilde{\mathbf{x}}_{-m} = \hat{\mathbf{x}}_{-m}$ or map $\hat{\mathbf{x}}_{-m}$ to the closest lattice point in \mathcal{X}^{-m} . Comparing with (20), (28) uses the *a priori* information through \mathbf{r}_{-m} and \mathbf{A}_{-m} , and it counts the impact of symbol \tilde{x}_m on $\Pr(\tilde{x}_m, \tilde{\mathbf{x}}_{-m} | \mathbf{y})$.

The process then goes to x_1, x_2, \dots, x_M . Before it reaches x_j , $j \neq m$, $V = \{m, 1, \dots, j-1\}$, the list \mathcal{L} contains K candidates, each of which has the form $\mathbf{z}_V = [x_m, x_1, \dots, x_{j-1}]^T$. For each $\mathbf{z}_V \in \mathcal{L}$ and each $\tilde{x}_j \in \mathcal{Q}_j$, we find the corresponding $\tilde{\mathbf{x}}_{-\{V,j\}}$ such that

$$\tilde{\mathbf{x}}_{-\{V,j\}} = \operatorname{argmax}_{\mathbf{x}_{-\{V,j\}} \in \mathcal{X}^{-\{V,j\}}} \Pr(\mathbf{z}_V, \tilde{x}_j, \mathbf{x}_{-\{V,j\}} | \mathbf{y}). \quad (29)$$

Among the resulting $K|\mathcal{Q}_j|$ $[\tilde{\mathbf{x}}_V^T, \tilde{x}_j]^T$, we only choose K of them such that $\Pr(\mathbf{z}_V, \tilde{x}_j, \tilde{\mathbf{x}}_{-\{V,j\}} | \mathbf{y})$ is maximized, update the list \mathcal{L} with the K chosen vectors, and add j into V . As in (28), we can approximate $\tilde{\mathbf{x}}_{-m}$ by solving

$$\begin{aligned} \hat{\mathbf{x}}_{-\{V,j\}} = \operatorname{argmin}_{\mathbf{x}_{-\{V,j\}}} & \|\mathbf{y} - \mathbf{H}_{-\{V,j\}}\mathbf{x}_{-\{V,j\}} - \mathbf{H}_V \mathbf{z}_V - \mathbf{h}_j \tilde{x}_j\|^2 \\ & + 2\sigma^2 \mathbf{r}_{-\{V,j\}}^T \mathbf{x}_{-\{V,j\}} + \sigma^2 \mathbf{x}_{-\{V,j\}}^T \mathbf{A}_{-\{V,j\}} \mathbf{x}_{-\{V,j\}}, \end{aligned} \quad (30)$$

where the notations are similar to those in (25) and (28).

Interestingly, when \mathcal{C}^{-m} is unbounded and the Gaussian approximation is used, we can show that the sum algorithm is equivalent to the max algorithm (see Appendix).

Bit-wise Algorithm: The sum-algorithm proceeds from symbol to symbol, which can also be applied on bits. For example, when set partitioning mapping is used, 2^C -PAM ($\mathcal{Q} = \{-2^C + 1, -2^C + 3, \dots, 2^C - 3, 2^C - 1\}$) can be written as a weighted sum of bits using (17). Taking the sum-algorithm as an example, to compute $L(b_i | \mathbf{y})$, we start with b_i and compute $\Pr(b_i = \pm 1 | \mathbf{y}) = \sum_{\mathbf{x} \in \mathcal{X}_{i,\pm 1}} \Pr(\mathbf{x} | \mathbf{y})$. In (23), we replace every x_j except x_m with a Gaussian or non-Gaussian continuous variable and $\Pr(b_i = \pm 1 | \mathbf{y})$ is computed by summing over all possible x_m in $\mathcal{X}_{i,\pm 1}^m$. We can even approximate x_m as a continuous variable. For example, when x_m is assumed to be Gaussian, we can compute the matched mean and variance as

$$\mu_{m,i,\pm 1} = \sum_{x_m \in \mathcal{X}_{i,\pm 1}^m} \Pr(x_m) x_m \quad (31)$$

and

$$\nu_{m,i,\pm 1}^2 = \sum_{x_m \in \mathcal{X}_{i,\pm 1}^m} \Pr(x_m) |x_m|^2 - |\mu_{m,i,\pm 1}|^2. \quad (32)$$

When the non-Gaussian distribution is used, we can get the distribution by fitting the distribution over the symbols in $\mathcal{X}_{i,\pm 1}^m$ only. We can obtain $\Pr(b_i | \mathbf{y})$ as (24). When the algorithm reaches bit b_j and its corresponding symbol is $x_{m'}$, let $\mathbf{b}_j = [b_1, \dots, b_j]^T$ where symbols $x_{m'+1}, \dots, x_{m-1}, x_{m+1}, \dots, x_M$ have not been visited. For any

$\tilde{\mathbf{b}}_j$ from the list \mathcal{L} , we can compute the matched mean and variance for $x_{m'}$ as

$$\mu_{m',\mathbf{b}_j,\tilde{\mathbf{b}}_j} = \sum_{x_{m'} \in \mathcal{X}_{\mathbf{b}_j,\tilde{\mathbf{b}}_j}^{m'}} \Pr(x_{m'}) x_{m'} \quad (33)$$

and

$$\nu_{m',\mathbf{b}_j,\tilde{\mathbf{b}}_j}^2 = \sum_{x_{m'} \in \mathcal{X}_{\mathbf{b}_j,\tilde{\mathbf{b}}_j}^{m'}} \Pr(x_{m'}) |x_{m'}|^2 - |\mu_{m',\mathbf{b}_j,\tilde{\mathbf{b}}_j}|^2, \quad (34)$$

where $\mathcal{X}_{\mathbf{b}_j,\tilde{\mathbf{b}}_j}^{m'}$ is the set of constellation points for $x_{m'}$ such that the corresponding bits in \mathbf{b}_j is equal to $\tilde{\mathbf{b}}_j$. The rest of the algorithm is identical to that of the symbol based algorithm.

The advantage of the bit-wise algorithm is that some symbols can be pruned early when the first few bits of the corresponding symbols are not chosen in the list with K elements.

Early Stopping and Varied K: The sum-algorithm stops after reaching x_M . We can stop the algorithm at any x_j . In this case, we can compute the LLR value as (35).

We can then approximate

$$\sum_{x_{j+1}, \dots, x_{m-1}, x_{m+1}, \dots, x_M} \Pr(x_{j+1}, \dots, x_{m-1}, x_{m+1}, \dots, x_M) \Pr(\mathbf{y} | \mathbf{x})$$

using Gaussian or non-Gaussian approximation. When $j = m$, $K = |\mathcal{Q}_m|$ and Gaussian approximation are used, the K -best algorithm with early stopping reduces to that in Section III-C or the PDA method in [16]. The stopping level gives a tradeoff between performance and complexity. This early stopping can also be used when some symbols are not reliable, e.g., every symbol in the constellation has roughly the same probability or LLR value is less than a threshold. In this case, different candidates may have roughly the same metric and choosing the best K candidates may not be good. We can reorder the symbols such that the unreliable symbols correspond to the last few symbols and use early stopping when the algorithm reaches the unreliable symbols.

We can also vary the list size K for different symbols. The list size K_j can be chosen as K_j after symbol x_j is visited. For example, K_j can be chosen to be a large value for the first few visited symbols as the choice of these symbols is important to the overall performance, and K_j is chosen to be a small value when the algorithm is close to the end to save complexity.

Remarks:

- Note that the proposed algorithms are different from the K best algorithm in [13] in the way how the K best candidates are generated and updated and how the LLR value is computed. In [13], LLR value is computed using only the candidate from the list with the largest $\Pr(\mathbf{x} | \mathbf{y})$ while we consider all the candidates in the list.
- There are several interesting special cases of the proposed algorithms. When $K = 1$, the proposed algorithms become an improved soft version of V-BLAST algorithm. When $K = +\infty$, they reduce to the optimal MAP detection in Section III-B.
- The proposed K -best sum-algorithm can be considered as a combination of soft successive interference cancellation (SIC) and hard SIC, where a stream of data is canceled

$$L(b_i|\mathbf{y}) \approx \log \frac{\sum_{[x_1, \dots, x_j, x_m]^T \in \mathcal{L}_{i,+1}} \Pr(x_1, \dots, x_j, x_m)}{\sum_{[x_1, \dots, x_j, x_m]^T \in \mathcal{L}_{i,-1}} \Pr(x_1, \dots, x_j, x_m)} \frac{\sum_{x_{j+1}, \dots, x_{m-1}, x_{m+1}, \dots, x_M} \Pr(x_{j+1}, \dots, x_{m-1}, x_{m+1}, \dots, x_M) \Pr(\mathbf{y}|\mathbf{x})}{\sum_{x_{j+1}, \dots, x_{m-1}, x_{m+1}, \dots, x_M} \Pr(x_{j+1}, \dots, x_{m-1}, x_{m+1}, \dots, x_M) \Pr(\mathbf{y}|\mathbf{x})}. \quad (35)$$

by hard decision on the bits of a symbol followed by re-modulation of the hard decoded bits. The sum-algorithm uses soft SIC when choosing candidates from the list and uses hard SIC before it proceeds to the next symbol.

- Instead of using sum-log to compute the LLR value, we can also use max-log.
- In practical protocols there always exists some CRC check bits. When a particular data stream passes the CRC check, we do not need to include this data stream in the future iterative demodulation and decoding. We can cancel this data stream directly or hard SIC, which will reduce the complexity of the proposed K -best algorithms.

D. Complexity Reduction

Direct computation of (25) requires matrix inversion and matrix multiplication for every $\tilde{x}_j \in \mathcal{Q}_j$. From the expression of $\mathbf{R}_{\{\nu, j\}}$ in (26) and the matrix inversion lemma [29], we have

$$\mathbf{R}_{\{\nu, j\}}^{-1} = (\mathbf{R}_\nu + \nu_j^2 \mathbf{h}_j \mathbf{h}_j^H)^{-1} = \mathbf{R}_\nu^{-1} - \mathbf{g}_j (\nu_j^{-2} + \mathbf{h}_j^H \mathbf{g}_j)^{-1} \mathbf{g}_j^H, \quad (36)$$

where $\mathbf{g}_j = \mathbf{R}_\nu^{-1} \mathbf{h}_j$. Initially, we need to compute $(\mathbf{H} \text{diag}\{\nu^2\} \mathbf{H}^H + \sigma^2 \mathbf{I}_N)^{-1}$, which has a complexity $O(N^{2.376} + NM^2)$. Substituting (36) into (25), we obtain

$$\begin{aligned} & (\mathbf{y} - \mathbf{H}_{-\{\nu, j\}} \boldsymbol{\mu}_{-\{\nu, j\}} - \mathbf{H}_\nu \mathbf{x}_\nu - \mathbf{h}_j \tilde{x}_j)^H \mathbf{R}_{\{\nu, j\}}^{-1} \\ & \quad \times (\mathbf{y} - \mathbf{H}_{-\{\nu, j\}} \boldsymbol{\mu}_{-\{\nu, j\}} - \mathbf{H}_\nu \mathbf{x}_\nu - \mathbf{h}_j \tilde{x}_j) \quad (37) \\ & = C - 2\tilde{x}_j B + \tilde{x}_j^2 A \end{aligned}$$

where

$$\begin{aligned} A &= \mathbf{h}_j^H \mathbf{g}_j - \mathbf{h}_j^H \mathbf{g}_j (\nu_j^{-2} + \mathbf{h}_j^H \mathbf{g}_j)^{-1} \mathbf{g}_j^H \mathbf{h}_j \\ B &= \left(1 - \mathbf{h}_j^H \mathbf{g}_j (\nu_j^{-2} + \mathbf{h}_j^H \mathbf{g}_j)^{-1}\right) \mathbf{g}_j^H \\ & \quad \times (\mathbf{y} - \mathbf{H}_{-\nu} \boldsymbol{\mu}_{-\nu} + \mathbf{h}_j \mu_j - \mathbf{H}_\nu \mathbf{x}_\nu) \\ C &= \mathbf{y}^H \mathbf{R}_\nu^{-1} \mathbf{y} - \mathbf{y}^H \mathbf{g}_j (\nu_j^{-2} + \mathbf{h}_j^H \mathbf{g}_j)^{-1} \mathbf{g}_j^H \mathbf{y} \end{aligned} \quad (38)$$

Computing $\mathbf{h}_j^H \mathbf{g}_j$ and $\mathbf{g}_j^H \mathbf{y}$ needs $2(N-1)$ additions and $2N$ multiplications. Moreover, $\mathbf{y}^H \mathbf{R}_\nu^{-1} \mathbf{y}$ and $\mathbf{y} - \mathbf{H}_{-\nu} \boldsymbol{\mu}_{-\nu}$ are inherited from the previous step. $\mathbf{H}_\nu \mathbf{x}_\nu$ is updated and stored in the list and the update needs KN multiplications and KN additions. Computing $\mathbf{y} - \mathbf{H}_{-\nu} \boldsymbol{\mu}_{-\nu} + \mathbf{h}_j \mu_j - \mathbf{H}_\nu \mathbf{x}_\nu$ needs N multiplications and $2N$ additions. The total number of additions to compute the coefficients A, B, C for all the elements in the list is $3(K+1)N + K - 2$ and the total number of multiplications is $(2K+3)N + 5$. As (37) is a scalar function in \tilde{x}_j , we can search over \mathcal{Q}_j for each \mathbf{x}_ν to find the K candidates with the maximum (25). This simple algorithm requires $2K|\mathcal{Q}_j|$ multiplications and $2K|\mathcal{Q}_j|$ additions.

Assuming exponential function can be computed by lookup table, Table I compares the complexity of the proposed K -Best

algorithm, Gaussian approximation and solving (5) optimally denoted as MAP (maximum a posteriori) in a $N \times M$ MIMO system with constellation size Q , where we have assumed matrix inversion uses Gauss-Jordan elimination [29]. From the big-O notation, we can see that Gaussian approximation reduces the complexity from $O(Q^M NM)$ exponential in M to $O(QNM^2)$. K -best algorithm can further reduce the complexity to $O(KNM)$.

V. SIMULATION RESULTS

In this section, we present simulation results to verify the effectiveness of the proposed algorithms. We consider a MIMO-OFDM system with 1024 subcarriers and 960 subcarriers are used for data transmission. Perfect knowledge of channel state information is assumed. Each transmit antenna is assigned power P . The SNR is defined as P/N_0 , where N_0 is the noise power. The simulations are conducted using an LTE simulator with the encoder, interleaver, rate matching procedure following LTE standard specified in [30]. We consider Extended Vehicular A model (EVA) [31] in this section with delay profile [0 30 150 310 370 710 1090 1730 2510] ns and power profile [0 -1.5 -1.4 -3.6 -0.6 -9.1 -7 -12 -16.9] dB. The channel power profile is normalized to unity. The scheme of turbo encoder is a Parallel Concatenated Convolutional Code (PCCC) with two 8-state constituent encoders and one turbo code internal interleaver. The coding rate of turbo encoder is 1/3. The transfer function of the PCCC is: $G(D) = [1, \frac{g_0(D)}{g_1(D)}]$ [30], where $g_0(D) = 1 + D^2 + D^3$, $g_1(D) = 1 + D + D^3$. Eight iterations are performed within the turbo decoder. Finally, 64 QAM and Gray mapping are considered in this section.

The algorithm using (5) is denoted as MAP. LSD in [12] is denoted as LSD (List sphere decoder). The Gaussian approximation using (8) is denoted as Gaussian. Estimating \hat{x}_i from (2) using a minimum mean square error (MMSE) estimator followed by a turbo decoder without iterative demodulation and decoding is denoted as MMSE. In LSD, we choose the list size $L=512$ to be consistent with [12]. The K -best algorithm using common list is denoted as K -best, while the K -Best algorithm using (25) for each bit is denoted as K -Best-Bit. The non-Gaussian approximation algorithm in Section IV-A is denoted as Non-Gaussian.

A. Bit Error Rate (BER) Comparison of Different Algorithms

We first consider fixed scheduling, where both data streams transmit using transport block size (TBS) 1916. The bit error rates of different algorithms after simulating 20000 subframes in a 2×2 MIMO-OFDM system are shown in Fig. 4. The channel varies independently from subframe to subframe. All algorithms use 6 iterations. It is clear that all iterative algorithms benefit from the information exchange between the demapper and decoder as compared with MMSE. We can see

TABLE I
COMPLEXITY COMPARISON OF DIFFERENT ALGORITHMS TO COMPUTE $L(b_i|\mathbf{y})$.

	additions	multiplications	Big-O
MAP	$Q^M N(M+1) - 2$	$Q^M (N(M+1) + Q - 1)$	$O(Q^M NM)$
Gaussian	$Q(N(M-1)^2 + \frac{1}{3}N^3 + \frac{3}{2}N^2 + \frac{1}{6}N + MN) - 2$	$Q(N(M-1)^2 + \frac{1}{3}N^3 + \frac{3}{2}N^2 + \frac{4}{3}N + MN + 1)$	$O(QNM^2)$
K -Best	$M(3(K+1)N + 2KQ + K - 2) + 2KN(M+1) - 2$	$M((2K+3)N + 2KQ + 5) + 2K(N(M+1) + Q + 1)$	$O(KNM)$

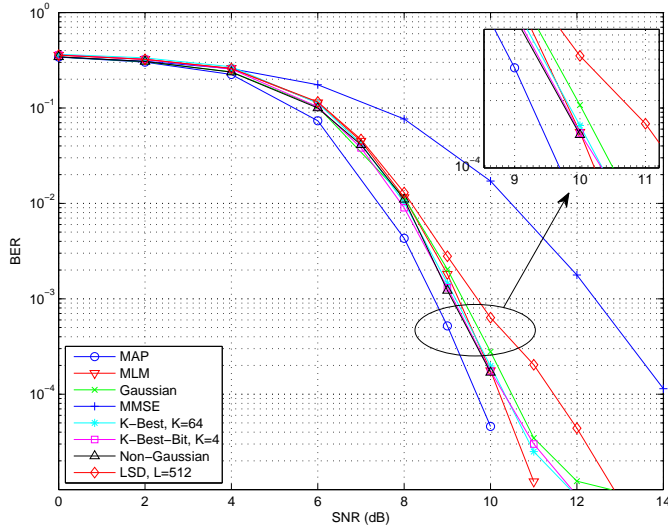


Fig. 4. BER comparison of different algorithms in a MIMO-OFDM system with 2 transmit and 2 receive antennas over the EVA channel.

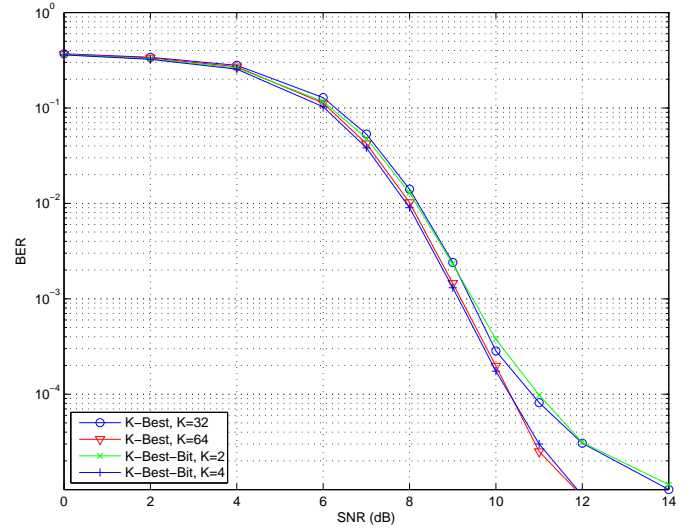


Fig. 5. K -best algorithm BER comparison with list size K in a MIMO-OFDM system with 2 transmit and 2 receive antennas over the EVA channel.

that Gaussian approximation incurs a 0.5 dB loss over MAP at $\text{BER}=10^{-3}$. The Gaussian approximation only needs to sum over 64 terms while the MAP needs to compute $32 \times 64 = 2048$ terms in the numerator and denominator in (5), respectively. With the proposed K -best algorithms, K -Best, $K=64$ has a 0.08 dB gain over Gaussian approximation, and K -Best-Bit, $K=4$ has a 0.15 dB gain over Gaussian approximation at $\text{BER}=10^{-3}$. K -Best-Bit, $K=4$ only needs to sum over $K=4$ terms in the LLR computation but with improved performance over Gaussian approximation. LSD with $L=512$ incurs a 1 dB loss over MAP at $\text{BER}=10^{-3}$ but with a higher complexity than Gaussian approximation and the proposed K -best algorithms. The non-Gaussian approximation achieves a 0.3 dB gain over Gaussian approximation at $\text{BER}=10^{-3}$. The proposed K -Best algorithms achieve good performance with reduced complexity.

B. Effects of the List Size K

In Fig. 5, the performance of the K -best algorithms are compared with different list size K in a 2×2 MIMO-OFDM system. We can see that for both K -Best and K -Best-Bit, doubling K gives a 0.5 dB gain at $\text{BER}=10^{-3}$. In high SNR, the error floor is also reduced by increasing K . But doubling K also means that the complexity is roughly doubled. Moreover, more exponentials are needed to compute in the LLR computation, which is very expensive in hardware implementation.

C. Effects of the Number of Iterations

In practice, it is crucial that the decoding time satisfies the timeline constraint in LTE-A [1]. It is therefore important to understand the performance of the iterative decoders in terms of the number of iterations. Fig. 6 shows the BER performance of the proposed K -best algorithm with $K=64$ and different number of iterations in a 2×2 MIMO-OFDM system (Fig. 6(a)) and in a 4×4 MIMO-OFDM system (Fig. 6(b)). We can see that marginal performance gain diminishes as the number of iterations increases. When comparing $\text{SNR}=8$ dB in Fig. 6(a) with $\text{SNR}=6$ dB in Fig. 6(b), we find that convergence speed of the proposed K -best algorithm decreases as the number of antennas increases. In general, three to four iterations is good enough in a 2×2 MIMO-OFDM system to achieve close to optimal performance using the proposed K -best algorithms, which means the proposed algorithms can be implemented with fixed complexity and fixed detection throughput.

D. Throughput Comparison of Different Algorithms

BER does not translate to throughput performance directly. A higher TBS may have higher BER even though data rate is higher. In Fig. 7, we compare the throughput gain of different algorithms over Gaussian approximation in a 2×2 MIMO-OFDM system. The throughput gain is obtained after averaging 50 channel realizations. For each channel realization, we find the maximum TBS pair on the two streams that can achieve average 10% block error rate (BLER) on both data

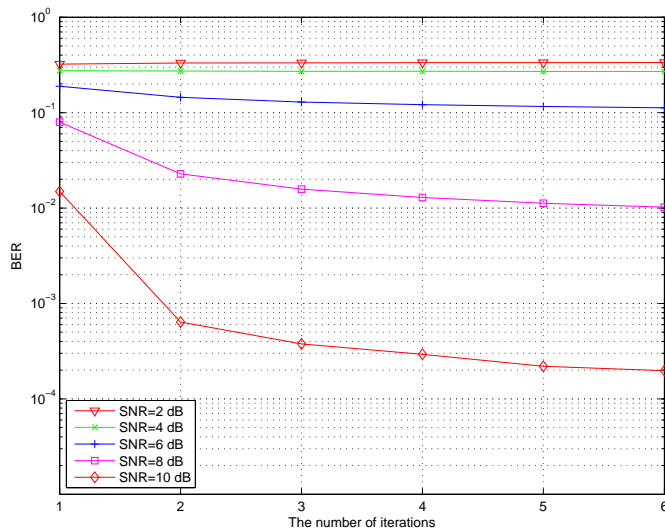
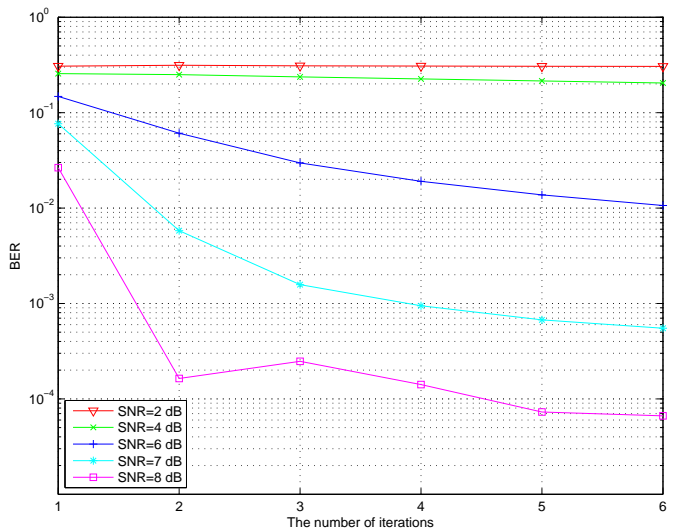

 (a) A 2×2 MIMO-OFDM system

 (b) A 4×4 MIMO-OFDM system

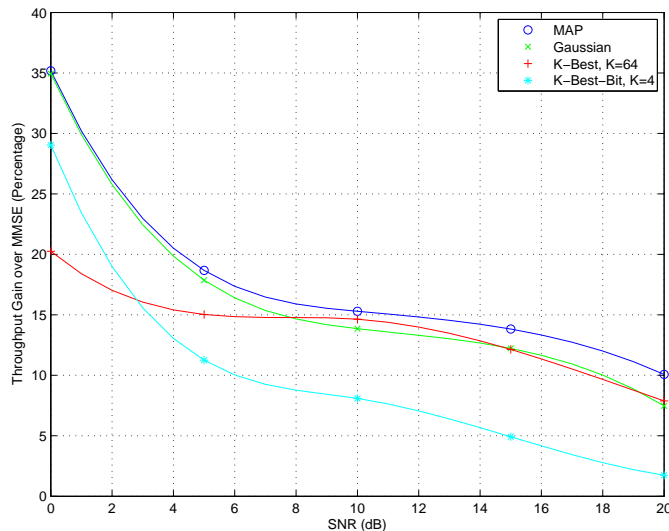
 Fig. 6. BER comparison of K -best algorithm with $K=64$ and different number of iterations


Fig. 7. Throughput gain comparison of different algorithms over MMSE in a MIMO-OFDM system with 2 transmit and 2 receive antennas over the EVA channel.

streams. We choose 10% BLER because optimal throughput can be achieved under most channel conditions by targeting 10% BLER [27], [32]. The modulation corresponding to each TBS is chosen according to [30] and Hybrid ARQ is used. We can see that the throughput gain of all algorithms decreases as SNR increases. This is because in high SNR the MIMO channel capacity with finite constellation is saturated to 6 bits/s/Hz as the largest constellation is 64QAM. In high SNR, all algorithms can approach this limit. K -Best-Bit, $K=4$ achieves almost the same throughput as MAP in all SNRs. Both K -Best, $K=64$ and K -Best-Bit, $K=4$ achieve more than 5% gain over Gaussian approximation when SNR is greater than 5 dB but the former two are less complex than the latter one.

VI. CONCLUSION

The design of low complexity MIMO-OFDM LTE-A receivers to meet the data rate requirement while achieving power consumption and silicon area savings is critically important. In this paper, we developed several such low-complexity iterative detection and decoding algorithms. Non-Gaussian approximation was proposed to enhance the performance of interference cancellation based detectors with large constellations. Several modified K -best algorithms were also developed to take advantages of both Gaussian or non-Gaussian approximation and the list decoder, which provide a flexible performance and complexity tradeoff. Simulation results demonstrated that the proposed low complexity algorithms can achieve a performance gain over existing ones for practical systems that use high-order constellations.

APPENDIX

THE EQUIVALENCE BETWEEN THE SUM AND MAX ALGORITHMS

In this appendix, we consider a general system

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \quad (39)$$

where \mathbf{x} is of mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{R}_x and \mathbf{w} is of mean zero and covariance matrix \mathbf{R}_w . We show that by using Gaussian approximation the sum algorithm (25) is equivalent to the max algorithm. To do so, we solve

$$\min_{\mathbf{x}_{-m} \in \mathcal{C}^{-m}} \|\mathbf{y} - \mathbf{H}_{-m}\mathbf{x}_{-m} - \mathbf{h}_m \tilde{x}_m\|^2 + (\mathbf{x}_{-m} - \boldsymbol{\mu}_{-m})^H \boldsymbol{\Lambda}_{-m} (\mathbf{x}_{-m} - \boldsymbol{\mu}_{-m}), \quad (40)$$

where $\boldsymbol{\Lambda}_{-m} = \text{diag}\{\nu_1^2, \dots, \nu_{m-1}^2, \nu_{m+1}^2, \dots, \nu_M^2\}$. Define

$$f(\mathbf{y}) = \exp\left(-(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})^H (\mathbf{H}\mathbf{R}_x\mathbf{H}^H + \mathbf{R}_w)^{-1} (\mathbf{y} - \mathbf{H}\boldsymbol{\mu})\right) \quad (41)$$

which corresponds to (25). Define

$$\begin{aligned} g(\mathbf{y}) &= \max_{\mathbf{x}} \exp \left(-\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 - (\mathbf{x} - \boldsymbol{\mu})^H \mathbf{R}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \\ &\propto \exp \left(-\mathbf{y}^H \mathbf{R}_w \mathbf{y} + (\mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{y} + \mathbf{R}_x^{-1} \boldsymbol{\mu})^H \right. \\ &\quad \left. \times (\mathbf{R}_x^{-1} + \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{H})^{-1} (\mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{y} + \mathbf{R}_x^{-1} \boldsymbol{\mu}) \right), \end{aligned} \quad (42)$$

which corresponds to (40). From the definition of LLR value, we want to show that $f(\mathbf{y})$ is proportional to $g(\mathbf{y})$, i.e., $f(\mathbf{y}) \propto g(\mathbf{y})$.

By using the matrix inversion lemma [29], we have

$$\begin{aligned} \mathbf{R}_w - \mathbf{R}_w^{-1} \mathbf{H} (\mathbf{R}_x^{-1} + \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^H \mathbf{R}_w^{-1} \\ = (\mathbf{H} \mathbf{R}_x \mathbf{H}^H + \mathbf{R}_w)^{-1}, \end{aligned} \quad (43)$$

which means that

$$\begin{aligned} \mathbf{y}^H \mathbf{R}_w \mathbf{y} - \mathbf{y}^H \mathbf{R}_w^{-1} \mathbf{H} (\mathbf{R}_x^{-1} + \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{y} \\ = \mathbf{y}^H (\mathbf{H} \mathbf{R}_x \mathbf{H}^H + \mathbf{R}_w)^{-1} \mathbf{y}. \end{aligned} \quad (44)$$

We also have

$$\begin{aligned} &\boldsymbol{\mu}^H \mathbf{R}_x^{-1} (\mathbf{R}_x^{-1} + \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{y} \\ &= \boldsymbol{\mu}^H \mathbf{R}_x^{-1} \left(\mathbf{R}_x - \mathbf{R}_x \mathbf{H}^H (\mathbf{H} \mathbf{R}_x \mathbf{H}^H + \mathbf{R}_w)^{-1} \mathbf{H} \mathbf{R}_x \right) \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{y} \\ &= \boldsymbol{\mu}^H \left(\mathbf{I}_N - \mathbf{H}^H (\mathbf{H} \mathbf{R}_x \mathbf{H}^H + \mathbf{R}_w)^{-1} \mathbf{H} \mathbf{R}_x \right) \mathbf{H}^H \mathbf{R}_w^{-1} \mathbf{y} \\ &= \boldsymbol{\mu}^H \mathbf{H}^H \left(\mathbf{I}_N - (\mathbf{H} \mathbf{R}_x \mathbf{H}^H + \mathbf{R}_w)^{-1} \mathbf{H} \mathbf{R}_x \mathbf{H}^H \right) \mathbf{R}_w^{-1} \mathbf{y} \\ &= \boldsymbol{\mu}^H \mathbf{H}^H (\mathbf{H} \mathbf{R}_x \mathbf{H}^H + \mathbf{R}_w)^{-1} \mathbf{y}. \end{aligned} \quad (45)$$

Substituting (44) and (45) into (42) and comparing with (41), it can be readily seen that $f(\mathbf{y}) \propto g(\mathbf{y})$. Therefore, the sum algorithm (25) is equivalent to the max algorithm (40) under the Gaussian approximation.

REFERENCES

- [1] "Evolved universal terrestrial radio access (E-UTRA); long term evolution (LTE) physical layer; general description," the 3rd Generation Partnership Project, TS 36.201, Mar.13, 2009.
- [2] M. Rumney, *LTE and the Evolution to 4G Wireless: Design and Measurement Challenges*, 2009.
- [3] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*, 2011.
- [4] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE-advanced," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 127–135, 2013.
- [5] F. Ghavimi and H.-H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: Architectures, service requirements, challenges, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 525–549, Secondquarter 2015.
- [6] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–15, Jan. 1999.
- [7] T. Cui and C. Tellambura, "An efficient generalized sphere decoder for rank-deficient MIMO systems," *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 423–425, 2005.
- [8] T. Cui, C. Tellambura, and Y. Wu, "Constrained detection for spatial-multiplexing multiple-input-multiple-output systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1537–1547, 2008.
- [9] T. Cui, S. Han, and C. Tellambura, "Probability-distribution-based node pruning for sphere decoding," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1586–1596, 2013.
- [10] H. Zhu, W. Chen, B. Li, and F. Gao, "An improved square-root algorithm for V-BLAST based on efficient inverse cholesky factorization," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 43–48, January 2011.
- [11] E. Viterbo and J. Bours, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1642, Jul. 1999.
- [12] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, March 2003.
- [13] 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, March 2006.
- [14] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.
- [15] M. Tüchler, A. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.
- [16] D. Pham, K. Pattipati, P. Willett, and J. Luo, "A generalized probabilistic data association detector for multiple antenna systems," *IEEE Commun. Lett.*, vol. 8, no. 4, pp. 205–207, April 2004.
- [17] S. Liu and Z. Tian, "Near-optimum soft decision equalization for frequency selective MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 721–733, Mar. 2004.
- [18] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. of IEEE Globecom*, Nov. 1998, pp. 579–584.
- [19] Q. Guo, D. D. Huang, S. Nordholm, J. Xi, and L. Ping, "Soft-in soft-out detection using partial Gaussian approximation," *IEEE Access*, vol. 2, pp. 427–436, May 2014.
- [20] M. Cirkic, D. Persson, E. G. Larsson, and J.-k. Larsson, "Gaussian approximation of the LLR distribution for the ML and partial marginalization MIMO detectors," in *ICASSP*, IEEE, 2011, pp. 3232–3235.
- [21] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay network," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May 2008.
- [22] T. Cui, F. Gao, T. Ho, and A. Nallanathan, "Distributed space-time coding for two-way wireless relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 658–671, Feb. 2009.
- [23] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024–3033, Oct. 2009.
- [24] —, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4443–4455, Nov. 2009.
- [25] S. Silva, G. Amarapura, C. Tellambura, and M. Ardakani, "Relay selection strategies for MIMO two-way relay networks with spatial multiplexing," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4694–4710, Dec. 2015.
- [26] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, November 2016.
- [27] P. Subrahmanya, A. Sendonaris, J. Tang, A. Salvekar, S. Khare, J. Park, B. Banister, and T. Cui, "Unified iterative decoding architecture using joint LLR extraction and a priori probability," 2015, US Patent 8,976,903.
- [28] P. M. Pardalos and S. A. Vavasis, "Quadratic programming with one negative eigenvalue is NP-hard," *Journal of Global Optimization*, vol. 1, no. 1, pp. 15–22, 1991.
- [29] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.
- [30] "Evolved universal terrestrial radio access (E-UTRA); multiplexing and channel coding," the 3rd Generation Partnership Project, TS 36.212, Jun. 8, 2009.
- [31] "Evolved universal terrestrial radio access (E-UTRA); user equipment (UE) radio transmission and reception," the 3rd Generation Partnership Project, TS 36.104, Jun.18, 2009.
- [32] T. Cui, F. Lu, V. Sethuraman, A. Goteti, S. P. Rao, and P. Subrahmanya, "Throughput optimization in high speed downlink packet access (HS-DPA)," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 474–483, 2011.

Tao Cui received the M.Sc. degree in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, in 2005, and the Ph.D. degree from the Department of Electrical Engineering, California Institute of Technology, Pasadena, USA, in 2009. His research interests are in networking theory, communication theory, information theory, machine learning and financial engineering.

Dr. Cui received the Best Paper Award at the IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS) in 2007 and the Second Place in the ACM Student Research Competition at the 2007 Richard Tapia Celebration of Diversity in Computing Conference. Dr. Cui has authored or coauthored over 80 journal and conference papers with total citations more than 2900 and an h-index of 29 from Google Scholar.



Feifei Gao (M'09-SM'14) received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China in 2002, the M.Sc. degree from McMaster University, Hamilton, ON, Canada in 2004, and the Ph.D. degree from National University of Singapore, Singapore in 2007. He was a Research Fellow with the Institute for Infocomm Research (I2R), A*STAR, Singapore in 2008 and an Assistant Professor with the School of Engineering and Science, Jacobs University, Bremen, Germany from 2009 to 2010. In 2011, he joined the Department of Automation, Tsinghua University, Beijing, China, where he is currently an Associate Professor.

Prof. Gao's research areas include communication theory, signal processing for communications, array signal processing, and convex optimizations, with particular interests in MIMO techniques, multi-carrier communications, cooperative communication, and cognitive radio networks. He has authored/coauthored more than 120 refereed IEEE journal papers and more than 140 IEEE conference proceeding papers, which have been cited more than 4200 times from Google Scholar.

Prof. Gao has served as an Editor of IEEE Transactions on Wireless Communications, IEEE Communications Letters, IEEE Signal Processing Letters, IEEE Wireless Communications Letters, International Journal on Antennas and Propagations, and China Communications. He has also served as the symposium co-chair for 2015 IEEE Conference on Communications (ICC), 2014 IEEE Global Communications Conference (GLOBECOM), 2014 IEEE Vehicular Technology Conference Fall (VTC), as well as Technical Committee Members for many other IEEE conferences.



Arumugam Nallanathan (S'97-M'00-SM'05-F'17) is Professor of Wireless Communications and Head of the Communication Systems Research (CSR) group in the School of Electronic Engineering and Computer Science at Queen Mary University of London since September 2017. He was with the Department of Informatics at King's College London from December 2007 to August 2017, where he was Professor of Wireless Communications from April 2013 to August 2017 and a Visiting Professor from September 2017. He was an Assistant Professor in

the Department of Electrical and Computer Engineering, National University of Singapore from August 2000 to December 2007. His research interests include 5G Wireless Networks, Internet of Things (IoT) and Molecular Communications. He published more than 350 technical papers in scientific journals and international conferences. He is a co-recipient of the Best Paper Award presented at the IEEE International Conference on Communications 2016 (ICC2016) and IEEE International Conference on Ultra-Wideband 2007 (ICUWB 2007). He is an IEEE Distinguished Lecturer. He has been selected as a Web of Science (ISI) Highly Cited Researcher in 2016.

He is an Editor for IEEE Transactions on Communications. He was an Editor for IEEE Transactions on Wireless Communications (2006-2011), IEEE Transactions on Vehicular Technology (2006-2017), IEEE Wireless Communications Letters and IEEE Signal Processing Letters. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of IEEE Communications Society and Technical Program Chair and member of Technical Program Committees in numerous IEEE conferences. He received the IEEE Communications Society SPCE outstanding service award 2012 and IEEE Communications Society RCC outstanding service award 2014.



Hai Lin (M'05-SM'12) received the B.E. degree from Shanghai JiaoTong University, China, in 1993, the M.E. degree from University of the Ryukyus, Japan, in 2000, and the Dr. Eng. degree from Osaka Prefecture University, Japan, in 2005. Since 2000, he was a research associate in the Graduate School of Engineering, Osaka Prefecture University, where now he is an associate professor.

Dr. Lin has acted many times as technical program committee member for IEEE ICC, GLOBECOM, WCNC, VTC, etc. He served as a technical program co-chair for the Wireless Communications Symposium of IEEE ICC 2011, the Signal Processing for Communications Symposium of IEEE ICC 2013, and the Wireless Communications Symposium of IEEE GLOBECOM 2013. Dr. Lin was an Editor of the IEEE Transactions on Wireless Communications, and now is an Editor of the IEEE Transactions on Vehicular Technology. His research interests are in signal processing for communications, wireless communications, and statistical signal processing. He is a member of the IEICE.



Chintha Tellambura (F'11) received the B.Sc. degree (with first-class honor) from the University of Moratuwa, Sri Lanka in 1986, the MSc degree in Electronics from King's College, University of London, United Kingdom in 1988, and the PhD degree in Electrical Engineering from the University of Victoria, Canada, in 1993.

He was with Monash University, Australia, from 1997 to 2002. Presently, he is a Professor with the Department of Electrical and Computer Engineering, University of Alberta. His current research interests

include the design, modelling and analysis of cognitive radio, heterogeneous cellular networks and 5G wireless networks.

Prof. Tellambura served as an editor for both IEEE Transactions on Communications (1999-2011) and IEEE Transactions on Wireless Communications (2001-2007) and for the latter he was the Area Editor for Wireless Communications Systems and Theory during 2007-2012. He has received best paper awards in the Communication Theory Symposium in 2012 IEEE International Conference on Communications (ICC) in Canada and 2017 ICC in France. He is the winner of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. In 2011, he was elected as an IEEE Fellow for his contributions to physical layer wireless communication theory. In 2017, he was elected as a Fellow of Canadian Academy of Engineering. Prof. Tellambura has authored or coauthored over 500 journal and conference papers with total citations more than 14,000 and an h-index of 62 (Google Scholar).