

Group Information Geometry Approach for Ultra-Massive MIMO Signal Detection

Jiyuan Yang , Member, IEEE, Mingrui Fan , Yan Chen, Xiqi Gao , Fellow, IEEE, Xiang-Gen Xia , Fellow, IEEE, and Dirk Slock , Fellow, IEEE

Abstract—We propose a group information geometry approach (GIGA) for ultra-massive multiple-input multiple-output (MIMO) signal detection. The signal detection task is framed as computing the approximate marginals of the *a posteriori* distribution of the transmitted data symbols of all users. With the approximate marginals, we perform the maximization of the *a posteriori* marginals (MPM) detection to recover the symbol of each user. Based on the information geometry theory and the grouping of the components of the received signal, three types of manifolds are constructed and the approximate *a posteriori* marginals are obtained through *m*-projections. The Berry-Esseen theorem is introduced to offer an approximate calculation of the *m*-projection, while its direct calculation is exponentially complex. In most cases, increasing the number of groups tends to reduce the computational complexity of GIGA. However, when the number of groups exceeds a certain threshold, the complexity of GIGA starts to increase. Simulation results confirm that the proposed GIGA achieves better bit error rate (BER) performance within a small number of iterations, which demonstrates that it can serve as an efficient detection method in ultra-massive MIMO systems.

Index Terms—Bayesian inference, information geometry, general belief propagation, ultra-massive MIMO, signal detection.

Received 17 September 2024; revised 11 February 2025 and 3 April 2025; accepted 7 April 2025. Date of publication 30 April 2025; date of current version 19 June 2025. This work was supported in part by Jiangsu Province Major Science and Technology Project under Grant BG2024005, in part by Jiangsu Province Basic Research Project under Grant BK20192002, in part by the Fundamental Research Funds for the Central Universities under Grant 2242022k60007, in part by the Key R&D Plan of Jiangsu Province under Grant BE2022067, and in part by Huawei Cooperation Project. The associate editor coordinating the review of this article and approving it for publication was Cihan Tepedelenlioglu. (Jiyuan Yang and Mingrui Fan are co-first authors.) (Corresponding author: Xiqi Gao.)

Jiyuan Yang, Mingrui Fan, Yan Chen, and Xiqi Gao are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China, and also with Purple Mountain Laboratories, Nanjing 211111 China (e-mail: jy yang@seu.edu.cn; 220220758@seu.edu.cn; 213160372@seu.edu.cn; xqgao@seu.edu.cn).

Xiang-Gen Xia is with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA (e-mail: xxia@ee.udel.edu).

Dirk Slock is with the Department of Communication Systems, EURECOM, 06410 Biot, France (e-mail: Dirk.Slock@eurecom.fr).

I. INTRODUCTION

EMERGING as a promising solution to address the capacity demands of future communication systems, ultra-massive multiple-input multiple-output (MIMO) (also known as extremely large-scale MIMO, extra-large scale massive MIMO, etc.) has attracted a significant attention. Ultra-massive MIMO leverages a substantial number of antennas at the base station (BS), often hundreds to thousands, to serve a large number of user terminals on the same time-frequency resource, which can dramatically improve spectrum efficiency, energy efficiency, and spatial resolution [1], [2], [3], [4], [5], [6]. This technology also offers substantial beamforming gains, crucial for mitigating path losses at high-frequency bands like millimeter-wave (mmWave) and terahertz (THz) [5], [7].

To fulfill the various advantages of ultra-massive MIMO, signal detection plays an important role. Typically, signal detection is used to recover the transmitted symbols of the user terminals based on the received signal and the channel state information (CSI). In MIMO transmission, inter-symbol interference and noise pose a great challenge for signal detection. In general, the maximum *a posteriori* (MAP) or maximum-likelihood (ML) detection could provide a statistically optimal solution by means of an exhaustive search over all possible transmitted symbols. Nevertheless, the combinatorial nature of the MAP or ML detection makes conventional numerical algorithms for convex optimization unsuitable. The MAP or ML detection can be prohibitively complex in practice. On the other hand, classical linear detectors such as zero-forcing (ZF) and linear minimum mean-squared error (LMMSE) detectors, suffer from limited performance despite their polynomial-time complexity [8], [9], [10].

The massive MIMO signal detection has been a topic of great interest during the past few years. Numerous algorithms have been proposed to address this problem [11], [12], [13], [14], [15], [16], [17]. Readers interested in the development of massive MIMO signal detection techniques can refer to [18]. Among all the signal detection algorithms, those based on Bayesian inference, such as belief propagation (BP), expectation propagation (EP), and approximate message propagation (AMP), have gained a lot of interest due to their satisfactory performance-complexity profile. While possessing reduced complexity compared to MAP or ML detection, these algorithms can still achieve sub-optimal detection performance. Based on the Markov random field and message passing, a

low-complexity detection algorithm is proposed for large dimensional MIMO-ISI channels in [19]. The paper [8] pioneers the integration of EP into massive MIMO signal detection with high-order modulation schemes. In [20], an iterative detector based on AMP is proposed for large-scale multiuser MIMO-OFDM systems.

Recently, an interdisciplinary field, information geometry, has attracted great interest. It merges concepts from information theory and differential geometry to explore the geometric structures and properties of statistical models [21], [22], [23]. From the perspective of information geometry, the set of distributions can be represented as a manifold, offering a natural approach to describe the relationship between different sets of probability distributions. One common metric used to measure the distance between different probability distributions is the well-known Kullback-Leibler (K-L) divergence. Information geometry provides a mathematical foundation for analyzing and understanding the intrinsic geometric structures of various statistical models. As an active and powerful subject, it has been widely used in research related to statistics, such as radar target detection [24], [25], channel estimation [26], and quantitative fault diagnosability analysis [27].

Recently, we have proposed an information geometry approach (IGA) for ultra-massive MIMO signal detection in [28]. In [28], the signal detection problem is framed as computing the approximate marginals of the *a posteriori* probability distribution of the transmitted symbols. Based on the *a posteriori* probability distribution, we define the objective manifold (OBM) and the auxiliary manifolds (AMs), where the OBM contains the approximate marginal probability distributions and each AM is related to the received signal on a single antenna at the receiver. The calculation of the approximation is then converted to m -projections from the distributions of AMs onto the OBM. A important property of m -projection is that the solution of m -projection exists and is unique. Although it can provide high detection performance, IGA has a slow convergence rate when modulation order and signal-to-noise ratio are high.

In this work, we propose a group information geometry approach (GIGA) for signal detection of ultra-massive MIMO systems. Our goal is still to acquire the approximation of the *a posteriori* marginals, and the maximization of the *a posteriori* marginals (MPM) detection is performed to recover the transmitted data symbols from the perspective of information geometry. Different from the IGA in [28], we express the *a posteriori* distribution in a factorized manner by grouping the components of the received signal. On this basis, we define new AMs, where each AM is related to the received signal on a group of antennas at the receiver. The approximate marginals are then obtained through m -projections from the distributions of AMs onto the OBM. A direct calculation of the m -projection is first presented, whose complexity is exponential and unaffordable. To solve this problem, we propose an approximate calculation of the m -projection based on the Berry-Esseen theorem, which can reduce the complexity significantly. Simulation results show that GIGA has significant advantages over existing algorithms in both convergence rate and detection performance. In general, the complexity of GIGA decreases as the number of groups

increases. When the number of groups exceeds a certain threshold, its complexity starts to increase. Given a proper number of groups, GIGA can obtain better BER performance with lower computational complexity compared to IGA.

The remaining sections are arranged as follows. Section II introduces the system configuration and problem formulation. GIGA is developed in Section III. The calculation of the m -projection is discussed in Section IV. Simulation results are discussed in Section V. Finally, conclusions are drawn in Section VI.

Throughout this paper, upper (lower) case boldface letters denote matrices (column vectors). $\mathbb{E}_p\{\cdot\}$ denotes the expectation operation w.r.t. the distribution $p(\mathbf{h})$. $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ denote the real and imaginary parts of a complex matrix, respectively. We use y_n or $[\mathbf{y}]_n$, $a_{i,j}$ or $[\mathbf{A}]_{i,j}$ to denote the n -th component of the vector \mathbf{y} and the (i,j) -th component of the matrix \mathbf{A} , where the element indices start with 1. Given a vector $\mathbf{y} \in \mathbb{C}^P$, where $P \geq 1$ is an integer. Define a set $\mathcal{B} \triangleq \{b_1, b_2, \dots, b_Q\}$, where $1 \leq b_1 < b_2 < \dots < b_Q \leq P$ and $Q \leq P$. We denote $[\mathbf{y}]_{\mathcal{B}}$ as

$$[\mathbf{y}]_{\mathcal{B}} \triangleq \left[[\mathbf{y}]_{b_1}, [\mathbf{y}]_{b_2}, \dots, [\mathbf{y}]_{b_Q} \right]^T \in \mathbb{C}^Q.$$

Given a matrix $\mathbf{A} \in \mathbb{C}^{P \times P'}$, we use $[\mathbf{A}]_{m,:}$ to denote the m -th row of the matrix \mathbf{A} . We denote $[\mathbf{A}]_{\mathcal{B},:}$ as

$$[\mathbf{A}]_{\mathcal{B},:} \triangleq \left[[\mathbf{A}]_{b_1,:}^T, [\mathbf{A}]_{b_2,:}^T, \dots, [\mathbf{A}]_{b_Q,:}^T \right]^T \in \mathbb{C}^{Q \times P'}.$$

\odot and \otimes denote the Hadamard product and Kronecker product, respectively. We define $\mathcal{Z}_N \triangleq \{0, 1, \dots, N\}$ and $\mathcal{Z}_N^+ \triangleq \{1, 2, \dots, N\}$. $\|\cdot\|$ denotes the ℓ_2 norm. We use $p(\cdot)$ to represent the probability distribution of discrete random variables and $f(\cdot)$ to represent the probability density function (PDF) of continuous random variables, respectively. We denote the PDF of a complex Gaussian random vector $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as $f_{\text{CG}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. We denote the PDF of a real Gaussian vector $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as $f_{\text{G}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Given a positive-definite matrix $\mathbf{A} \in \mathbb{C}^{m \times m}$, then \mathbf{A} can be decomposed as $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where $\mathbf{U} \in \mathbb{C}^{m \times m}$ is unitary, $\boldsymbol{\Lambda} \triangleq \text{Diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\}$, and $\{\lambda_i\}_{i=1}^m$ are all the eigenvalues of \mathbf{A} . We denote $\mathbf{A}^{-\frac{1}{2}}$ as $\mathbf{A}^{-\frac{1}{2}} \triangleq \mathbf{U}^H \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{U}$, where $\boldsymbol{\Lambda}^{-\frac{1}{2}} \triangleq \text{Diag}\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_m}\}$. Define the delta function $\delta(x - c)$ as

$$\delta(x - c) = \begin{cases} 1, & \text{when } x = c, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where c is a constant.

II. SYSTEM CONFIGURATION AND PROBLEM FORMULATION

In this section, we give the configuration of the considered ultra-massive MIMO system. Then, we present the problem formulation of the ultra-massive MIMO signal detection.

A. System Configuration

Consider an uplink ultra-massive MIMO transmission where K single-antenna users are served by a base station (BS) with

an ultra-massive antenna array of N_r antennas. Denote the transmitted data symbol of user k as $\tilde{s}_k \in \tilde{\mathbb{S}}$, where

$$\tilde{\mathbb{S}} \triangleq \left\{ \tilde{s}^{(0)}, \tilde{s}^{(1)}, \dots, \tilde{s}^{(\tilde{L}-1)} \right\}$$

is the signal constellation, $\{\tilde{s}^{(\ell)}\}_{\ell=0}^{\tilde{L}-1}$ are the constellation points, and \tilde{L} is the modulation order (or constellation size). Throughout this work, our focus is on uncoded systems employing symmetric \tilde{L} -QAM modulation. We assume that each user selects symbols uniformly from $\tilde{\mathbb{S}}$, and all users share the same signal constellation¹. In this paper, we also assume that the average power of each transmitted symbol is normalized to unit, i.e., $\mathbb{E} \left\{ |\tilde{s}_k|^2 \right\} = 1$, $k \in \mathcal{Z}_K^+$. Denote the transmitted symbol of all users as $\tilde{\mathbf{s}} \triangleq [\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_K]^T \in \tilde{\mathbb{S}}^K$. We assume that $\tilde{\mathbf{s}}$ is transmitted through a flat-fading channel. Then, the received signal $\tilde{\mathbf{y}} \in \mathbb{C}^{N_r}$ at the BS is given by

$$\tilde{\mathbf{y}} = \tilde{\mathbf{G}}\tilde{\mathbf{s}} + \tilde{\mathbf{z}}, \quad (2)$$

where $\tilde{\mathbf{G}} \in \mathbb{C}^{N_r \times K}$ is the channel matrix, $\tilde{\mathbf{z}}$ is the circular-symmetric complex Gaussian noise, $\tilde{\mathbf{z}} \sim \mathcal{CN}(\mathbf{0}, \tilde{\sigma}_z^2 \mathbf{I})$ and $\tilde{\sigma}_z^2$ is the noise variance². In this work, we assume perfect CSI at the BS.

B. Problem Formulation

Assuming that the noise vector $\tilde{\mathbf{z}}$ and the transmitted symbol vector $\tilde{\mathbf{s}}$ are independent, as are the symbols transmitted by different users. Given the received signal model (2), the *a posteriori* probability distribution of the transmitted symbol vector $\tilde{\mathbf{s}}$ can be expressed as [28]

$$p(\tilde{\mathbf{s}}|\tilde{\mathbf{y}}) \propto \prod_{k=1}^K p_{\text{pr},k}(\tilde{s}_k) f_{\text{CG}}(\tilde{\mathbf{y}}; \tilde{\mathbf{G}}\tilde{\mathbf{s}}, \tilde{\sigma}_z^2 \mathbf{I}), \quad (3)$$

where $p_{\text{pr},k}(\tilde{s}_k)$ is the a priori probability of \tilde{s}_k , and

$$p_{\text{pr},k}(\tilde{s}_k) \Big|_{\tilde{s}_k = \tilde{s}^{(\ell)}} = \frac{1}{\tilde{L}}, k \in \mathcal{Z}_K^+, \ell \in \mathcal{Z}_{\tilde{L}-1}.$$

Given $p(\tilde{\mathbf{s}}|\tilde{\mathbf{y}})$, the MAP detector (or, equivalently, the ML detector under this case) is given by

$$\tilde{\mathbf{s}}_{\text{MAP}} = \arg \max_{\tilde{\mathbf{s}} \in \tilde{\mathbb{S}}^K} p(\tilde{\mathbf{s}}|\tilde{\mathbf{y}}). \quad (4)$$

The optimization problem above is NP-hard due to the finite-alphabet constraint $\tilde{\mathbf{s}} \in \tilde{\mathbb{S}}^K$. When the number K of users and the modulation order \tilde{L} are large, the computation of (4) will become unaffordable for practical applications.

In this paper, we process the real-valued counterpart of the received signal model in (2), which is essential for the development of GIGA. To do so, let us first define the real-valued counterpart of (2). Define real counterparts of $\tilde{\mathbf{y}}$, $\tilde{\mathbf{z}}$, $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{G}}$ as

$$\mathbf{y} \triangleq \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{y}}\} \\ \mathcal{I}\{\tilde{\mathbf{y}}\} \end{bmatrix}, \mathbf{z} \triangleq \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{z}}\} \\ \mathcal{I}\{\tilde{\mathbf{z}}\} \end{bmatrix} \in \mathbb{R}^{2N_r} \quad (5a)$$

¹The proposed GIGA can be easily extended to any modulation with varying selecting probability, provided that the symbols of different users are statistically independent, and the real and imaginary parts of each user's symbol are statistically independent as well.

²In the above notations, tildes are placed atop the mathematical symbols. This is to simplify the notation when formulating and analyzing their real counterparts later on, where the tildes are removed.

$$\mathbf{s} \triangleq \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{s}}\} \\ \mathcal{I}\{\tilde{\mathbf{s}}\} \end{bmatrix} \in \mathbb{R}^{2K}, \quad (5b)$$

$$\mathbf{G} \triangleq \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{G}}\}, & -\mathcal{I}\{\tilde{\mathbf{G}}\} \\ \mathcal{I}\{\tilde{\mathbf{G}}\}, & \mathcal{R}\{\tilde{\mathbf{G}}\} \end{bmatrix} \in \mathbb{R}^{2N_r \times 2K}, \quad (5c)$$

respectively. Then, the real-valued counterpart of (2) is given by

$$\mathbf{y} = \mathbf{G}\mathbf{s} + \mathbf{z}. \quad (6)$$

In (6), \mathbf{s} is the real-valued transmitted symbol. Denote $\mathbf{s} \triangleq [s_1, s_2, \dots, s_{2K}]^T \in \mathbb{S}^{2K}$, where $s_k \in \mathbb{S}$,

$$\mathbb{S} \triangleq \left\{ s^{(0)}, s^{(1)}, \dots, s^{(L-1)} \right\}$$

is the alphabet for the real and imaginary components of the symmetric \tilde{L} -QAM modulation, and $L = \sqrt{\tilde{L}}$. Since $\tilde{\mathbf{z}}$ is circular-symmetric complex Gaussian, it can be readily obtained that $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{I})$, and $\sigma_z^2 = \tilde{\sigma}_z^2/2$. Given (6), the *a posteriori* distribution of \mathbf{s} can be expressed as [28]

$$\begin{aligned} p(\mathbf{s}|\mathbf{y}) &\propto p_{\text{pr}}(\mathbf{s}) f(\mathbf{y}|\mathbf{s}) \\ &= \prod_{k=1}^{2K} p_{\text{pr},k}(s_k) f_G(\mathbf{y}; \mathbf{G}\mathbf{s}, \sigma_z^2 \mathbf{I}) \end{aligned} \quad (7)$$

where $p_{\text{pr}}(\mathbf{s}) = \prod_k p_{\text{pr},k}(s_k)$ is the a priori probability of \mathbf{s} , $f(\mathbf{y}|\mathbf{s}) = f_G(\mathbf{y}; \mathbf{G}\mathbf{s}, \sigma_z^2 \mathbf{I})$ is the PDF of \mathbf{y} given \mathbf{s} ,

$$p_{\text{pr},k}(s_k) \Big|_{s_k = s^{(\ell)}} = \frac{1}{L}, \ell \in \mathcal{Z}_{L-1},$$

is the a priori probability of s_k . Denote the marginals of $p(\mathbf{s}|\mathbf{y})$ as $\{p_k(s_k|\mathbf{y})\}_{k=1}^{2K}$. The goal in this work is to obtain their approximations. Given the approximate $p_k(s_k|\mathbf{y})$, $k \in \mathcal{Z}_{2K}^+$, we perform the maximization of the a posteriori marginals (MPM) detection as

$$s_{k,\text{MPM}} = \arg \max_{s_k \in \mathbb{S}} p_k(s_k|\mathbf{y}), k \in \mathcal{Z}_{2K}^+. \quad (8)$$

Consequently, the detection of the transmitted data symbol $\tilde{\mathbf{s}}$ is given by

$$\begin{aligned} \tilde{\mathbf{s}}_{\text{de}} &= [s_{1,\text{MPM}}, s_{2,\text{MPM}}, \dots, s_{K,\text{MPM}}]^T \\ &\quad + \tilde{J} [s_{K+1,\text{MPM}}, s_{K+2,\text{MPM}}, \dots, s_{2K,\text{MPM}}]^T. \end{aligned} \quad (9)$$

III. GIGA

In this section, we start by stating the signal detection problem in perspective of information geometry. We then express the a posteriori distribution in a factorization way based on the grouping of the components of the received signal. On this basis, we propose the GIGA.

A. Signal Detection in Information Geometry Perspective

In this subsection, we state the signal detection problem from the information geometry perspective. We begin with the definitions of the original manifold and the objective manifold.

Define a manifold \mathcal{S} as a set of probability distributions, which contains all possible probability distributions of \mathbf{s} , i.e.,

$$\mathcal{S} = \left\{ p(\mathbf{s}) \mid p(\mathbf{s}) > 0, \mathbf{s} \in \mathbb{S}^{2K}, \sum_{\mathbf{s} \in \mathbb{S}^{2K}} p(\mathbf{s}) = 1 \right\}. \quad (10)$$

It can be readily checked that the a priori distribution $p_{\text{pr}}(\mathbf{s})$ and the *a posteriori* distribution $p(\mathbf{s}|\mathbf{y})$ belong to \mathcal{S} . We refer to \mathcal{S} as the original manifold (OM). Then, we define a sub-manifold of \mathcal{S} . It is the set \mathcal{M}_0 of probability distributions of \mathbf{s} , where the components of \mathbf{s} are independent of each other. Define a random vector \mathbf{t} as

$$\mathbf{t} \triangleq [\mathbf{t}_1^T, \mathbf{t}_2^T, \dots, \mathbf{t}_{2K}^T]^T \in \mathbb{R}^{2K(L-1)}, \quad (11a)$$

$$\mathbf{t}_k \triangleq [t_{k,1}, t_{k,2}, \dots, t_{k,L-1}]^T \in \mathbb{R}^{(L-1)}, \quad (11b)$$

$$t_{k,\ell} \triangleq \delta(s_k - s^{(\ell)}), \ell \in \mathcal{Z}_{L-1}^+. \quad (11c)$$

We can find that the components of \mathbf{t} are determined by the value of s_k , $k \in \mathcal{Z}_{2K}^+$. Define a vector \mathbf{d} as

$$\mathbf{d} \triangleq [\mathbf{d}_1^T, \mathbf{d}_2^T, \dots, \mathbf{d}_{2K}^T]^T \in \mathbb{R}^{2K(L-1)}, \quad (12a)$$

$$\mathbf{d}_k \triangleq [d_{k,1}, d_{k,2}, \dots, d_{k,L-1}]^T \in \mathbb{R}^{(L-1)}, \quad (12b)$$

$$d_{k,\ell} = \ln \frac{p_{\text{pr},k}(s_k) \big|_{s_k=s^{(\ell)}}}{p_{\text{pr},k}(s_k) \big|_{s_k=s^{(0)}}}, \ell \in \mathcal{Z}_{L-1}^+. \quad (12c)$$

We can find that \mathbf{d} is determined by the a priori probability of \mathbf{s} . In fact, the marginal distribution of the a priori distribution $p_{\text{pr}}(\mathbf{s})$ can be expressed as [28]

$$p_{\text{pr},k}(s_k) = \exp \{ \mathbf{d}_k^T \mathbf{t}_k - \psi(\mathbf{d}_k) \},$$

where $\psi(\mathbf{d}_k) = -\ln(p_{\text{pr},k}(s_k) \big|_{s_k=s^{(0)}})$ is the normalization factor. Consequently, we can also obtain

$$p_{\text{pr}}(\mathbf{s}) = \exp \{ \mathbf{d}^T \mathbf{t} - \psi(\mathbf{d}) \}, \quad (13)$$

where $\psi(\mathbf{d}) = \sum_{k=1}^{2K} \psi(\mathbf{d}_k)$ is the normalization factor. Based on the definitions above, the sub-manifold \mathcal{M}_0 of \mathcal{S} is defined as follows

$$\mathcal{M}_0 = \left\{ p_0(\mathbf{s}; \boldsymbol{\theta}_0) \mid \boldsymbol{\theta}_0 \in \mathbb{R}^{2K(L-1)} \right\}, \quad (14a)$$

$$\begin{aligned} p_0(\mathbf{s}; \boldsymbol{\theta}_0) &= \exp \left\{ \mathbf{d}^T \mathbf{t} + \boldsymbol{\theta}_0^T \mathbf{t} - \psi_0(\boldsymbol{\theta}_0) \right\} \\ &= \prod_{k=1}^{2K} p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}), \end{aligned} \quad (14b)$$

$$\begin{aligned} p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) &= \exp \left\{ \mathbf{d}_k^T \mathbf{t}_k + \boldsymbol{\theta}_{0,k}^T \mathbf{t}_k - \psi_0(\boldsymbol{\theta}_{0,k}) \right\} \\ &= \exp \left\{ \sum_{\ell=1}^{L-1} (d_{k,\ell} + \theta_{0,k,\ell}) \delta(s_k - s^{(\ell)}) - \psi_0(\boldsymbol{\theta}_{0,k}) \right\}, \end{aligned} \quad (15)$$

where

$$\boldsymbol{\theta}_0 = [\boldsymbol{\theta}_{0,1}^T, \boldsymbol{\theta}_{0,2}^T, \dots, \boldsymbol{\theta}_{0,2K}^T]^T \in \mathbb{R}^{2K(L-1)}, \quad (16a)$$

$$\boldsymbol{\theta}_{0,k} = [\theta_{0,k,1}, \theta_{0,k,2}, \dots, \theta_{0,k,L-1}]^T \in \mathbb{R}^{(L-1)}, \quad (16b)$$

$p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$ is the marginal distribution of s_k given the joint distribution $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$, $\psi_0(\boldsymbol{\theta}_0)$ is the free energy (normalization factor) of $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$, $\psi_0(\boldsymbol{\theta}_{0,k})$ is the free energy of $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$, and

$$\begin{aligned} \psi_0(\boldsymbol{\theta}_0) &= \sum_{k=1}^{2K} \psi_0(\boldsymbol{\theta}_{0,k}) \\ &= \ln \left(\sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \left\{ \mathbf{d}^T \mathbf{t} + \boldsymbol{\theta}_0^T \mathbf{t} \right\} \right), \end{aligned} \quad (17a)$$

$$\psi_0(\boldsymbol{\theta}_{0,k}) = \ln \left(1 + \sum_{\ell=1}^{L-1} \exp \{ d_{k,\ell} + \theta_{0,k,\ell} \} \right). \quad (17b)$$

$\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_{0,k}$ are referred to as the *e*-affine coordinate system (EACS) of $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$ and $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$, respectively. From the definition, it can be checked that $p_{\text{pr},k}(s_k) = p_{0,k}(s_k; \mathbf{0})$, and $p_{\text{pr}}(\mathbf{s}) = p_0(\mathbf{s}; \mathbf{0})$. \mathcal{M}_0 is referred to as the objective manifold (OBM) since it contains all the distributions of \mathbf{s} whose components are independent of each other, and our goal in this paper is to find a distribution in it to approximate $p(\mathbf{s}|\mathbf{y})$. In information geometry theory, this process is stated as calculating the *m*-projection of $p(\mathbf{s}|\mathbf{y})$ onto \mathcal{M}_0 , i.e.,

$$p_0(\mathbf{s}; \boldsymbol{\theta}_0^*) = \arg \min_{p_0(\mathbf{s}; \boldsymbol{\theta}_0) \in \mathcal{M}_0} D_{\text{KL}} \{ p(\mathbf{s}|\mathbf{y}) : p_0(\mathbf{s}; \boldsymbol{\theta}_0) \}, \quad (18)$$

where $D_{\text{KL}}\{\cdot\}$ is the Kullback-Leibler (K-L) divergence, and

$$D_{\text{KL}} \{ p(\mathbf{s}|\mathbf{y}) : p_0(\mathbf{s}; \boldsymbol{\theta}_0) \} = \sum_{\mathbf{s} \in \mathbb{S}^{2K}} p(\mathbf{s}|\mathbf{y}) \ln \left(\frac{p(\mathbf{s}|\mathbf{y})}{p_0(\mathbf{s}; \boldsymbol{\theta}_0)} \right).$$

$p_0(\mathbf{s}; \boldsymbol{\theta}_0^*)$ can be interpreted as the distribution in \mathcal{M}_0 that is closest to $p(\mathbf{s}|\mathbf{y})$, where the distance between the two distributions is defined as the K-L divergence. Given $p_0(\mathbf{s}; \boldsymbol{\theta}_0^*)$, its marginals can be directly obtained since the components of \mathbf{s} are independent. On the other hand, the calculation of the direct *m*-projection may be unacceptable since it can be too complicated. To solve this problem, we find another distribution in the OBM \mathcal{M}_0 to approximate $p(\mathbf{s}|\mathbf{y})$ by grouping the components of the received signal \mathbf{y} and defining an extra type of manifolds.

B. Factorization of $p(\mathbf{s}|\mathbf{y})$ by Grouping the Components of the Received Signal

In this subsection, we factorize the *a posteriori* distribution $p(\mathbf{s}|\mathbf{y})$ by the grouping the components of the received signal \mathbf{y} . We first present the way of grouping.³ Define the set of indexes of all components of \mathbf{y} as $\mathcal{N}_o \triangleq \{1, 2, \dots, 2N_r\}$. We uniformly divide \mathcal{N}_o into U subsets, where U is a factor of $2N_r$. Then, the number of the elements in each subset is

$$N_u = \frac{2N_r}{U},$$

Denote each subset as \mathcal{N}_u , $u \in \mathcal{Z}_U^+$ and we have

$$\mathcal{N}_u = \{(u-1)N_u + 1, (u-1)N_u + 2, \dots, uN_u\}. \quad (19)$$

³The GIGA groups the received signal components uniformly. Utilizing correlations between the elements of channel matrix \mathbf{G} for the grouping strategy design might be beneficial, which is worth of further research.

According to the subsets $\{\mathcal{N}_u\}_{u=1}^U$, we define U sub-vectors of \mathbf{y} , where the u -th of them only contains the components of \mathbf{y} indexed in \mathcal{N}_u , i.e.,

$$\mathbf{y}_u = [\mathbf{y}]_{\mathcal{N}_u} \in \mathbb{R}^{N_u}. \quad (20)$$

Given \mathbf{s} , the PDF of \mathbf{y}_u is Gaussian, and we have

$$f(\mathbf{y}_u|\mathbf{s}) = f_G(\mathbf{y}_u; \mathbf{G}_u \mathbf{s}, \sigma_z^2 \mathbf{I}) \quad (21)$$

where

$$\mathbf{G}_u = [\mathbf{G}]_{\mathcal{N}_u, :} \in \mathbb{R}^{N_u \times 2K} \quad (22)$$

is a sub-matrix of \mathbf{G} in (6). Since given \mathbf{s} , all components of \mathbf{y} in (6) are independent with each other, thus $\{\mathbf{y}_u\}_{u=1}^U$ are also independent, and we can readily obtain that $f(\mathbf{y}|\mathbf{s}) = \prod_{u=1}^U f(\mathbf{y}_u|\mathbf{s})$. On this basis, the *a posteriori* distribution $p(\mathbf{s}|\mathbf{y})$ can be factorized as

$$p(\mathbf{s}|\mathbf{y}) \propto \prod_{k=1}^{2K} p_{\text{pr},k}(s_k) \prod_{u=1}^U f(\mathbf{y}_u|\mathbf{s}). \quad (23)$$

C. GIGA

According to (23), the *a posteriori* probability distribution $p(\mathbf{s}|\mathbf{y})$ can be further expressed as

$$p(\mathbf{s}|\mathbf{y}) = \exp \left\{ \mathbf{d}^T \mathbf{t} + \sum_{u=1}^U c_u(\mathbf{s}, \mathbf{y}_u) - \psi_q \right\}, \quad (24)$$

where

$$c_u(\mathbf{s}, \mathbf{y}_u) = -\frac{1}{2\sigma_z^2} \|\mathbf{y}_u - \mathbf{G}_u \mathbf{s}\|^2, \quad (25)$$

ψ_q is the normalization factor, and

$$\psi_q = \ln \left(\sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \left\{ \mathbf{d}^T \mathbf{t} + \sum_{u=1}^U c_u(\mathbf{s}, \mathbf{y}_u) \right\} \right). \quad (26)$$

In (24), $\mathbf{d}^T \mathbf{t}$ only contains a linear combination of single independent random variables s_k , and $\{c_u(\mathbf{s}, \mathbf{y}_u)\}_{u=1}^U$ contain all the interactions (cross-terms) between the random variables, i.e., $s_k s_{k'}$, $k \neq k'$. If we can approximate the sum $\sum_{u=1}^U c_u(\mathbf{s}, \mathbf{y}_u)$ as $\boldsymbol{\theta}_0^T \mathbf{t}$, then we have

$$p(\mathbf{s}|\mathbf{y}) \approx p_0(\mathbf{s}; \boldsymbol{\theta}_0) = \exp \left\{ (\mathbf{d} + \boldsymbol{\theta}_0)^T \mathbf{t} - \psi_0 \right\}, \quad (27)$$

where ψ_0 is the free energy, which is simple to use. Since computing the m -projection of $p(\mathbf{s}|\mathbf{y})$ onto the OBM directly is exponentially complex, which is unacceptable for the ultra-massive MIMO system, we divide the cross-terms into U groups and define U auxiliary manifolds. In this work, we obtain the approximations of $\{c_u(\mathbf{s}, \mathbf{y}_u)\}_{u=1}^U$ through iterative m -projections. Let us first define a class of sub-manifolds of \mathcal{S} . Given the number U of subsets, we define U sub-manifolds of \mathcal{S} , where the u -th of them is given by

$$\mathcal{M}_u = \left\{ p_u(\mathbf{s}; \boldsymbol{\theta}_u) \mid \boldsymbol{\theta}_u \in \mathbb{R}^{2K(L-1)} \right\}, \quad (28a)$$

$$p_u(\mathbf{s}; \boldsymbol{\theta}_u) = \exp \left\{ \mathbf{d}^T \mathbf{t} + \boldsymbol{\theta}_u^T \mathbf{t} + c_u(\mathbf{s}, \mathbf{y}_u) - \psi_u(\boldsymbol{\theta}_u) \right\}, \quad (28b)$$

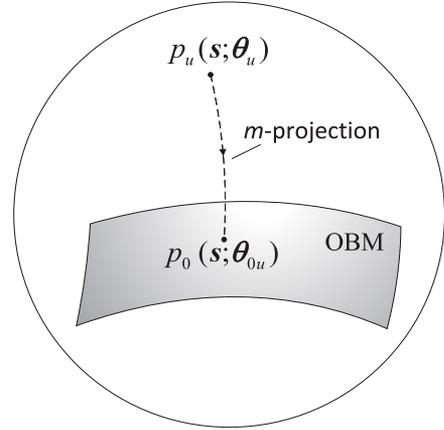


Fig. 1. Illustration of the m -Projection of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ onto the OBM.

where

$$\boldsymbol{\theta}_u = [\boldsymbol{\theta}_{u,1}^T, \boldsymbol{\theta}_{u,2}^T, \dots, \boldsymbol{\theta}_{u,2K}^T]^T \in \mathbb{R}^{2K(L-1)}, \quad (29a)$$

$$\boldsymbol{\theta}_{u,k} = [\theta_{u,k,1}, \theta_{u,k,2}, \dots, \theta_{u,k,L-1}]^T \in \mathbb{R}^{(L-1)}, \quad (29b)$$

$c_u(\mathbf{s}, \mathbf{y}_u)$ is given by (25), and the free energy ψ_u is given by

$$\psi_u(\boldsymbol{\theta}_u) = \ln \left(\sum_{\mathbf{s} \in \mathbb{S}^{2K}} \exp \left\{ \mathbf{d}^T \mathbf{t} + \boldsymbol{\theta}_u^T \mathbf{t} + c_u(\mathbf{s}, \mathbf{y}_u) \right\} \right). \quad (30)$$

$\{\mathcal{M}_u\}_{u=1}^U$ are referred to as the auxiliary manifolds (AMs), and $\boldsymbol{\theta}_u$ is referred to as the EACS of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$. Note that AMs will vary with the number U of subsets since the definitions of \mathbf{y}_u and \mathbf{G}_u will vary with U . Compared to $p(\mathbf{s}|\mathbf{y})$ in (24), $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ reserves only one interaction term $c_u(\mathbf{s}, \mathbf{y}_u)$, while all others, i.e., $\sum_{u' \neq u} c_{u'}(\mathbf{s}, \mathbf{y}_{u'})$, are replaced with $\boldsymbol{\theta}_u^T \mathbf{s}$.

$p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ of \mathcal{M}_u is the key to obtain the approximation of single interaction term $c_u(\mathbf{s}, \mathbf{y}_u)$. Given $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$, we can obtain an approximation of $c_u(\mathbf{s}, \mathbf{y}_u)$ through its m -projection on the OBM, which is illustrated in Fig. 1. Denote this m -projection as $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$, we have

$$p_0(\mathbf{s}; \boldsymbol{\theta}_{0u}) = \arg \min_{p_0(\mathbf{s}; \boldsymbol{\theta}_{0u}) \in \mathcal{M}_0} D_{\text{KL}} \{p_u(\mathbf{s}; \boldsymbol{\theta}_u) : p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})\},$$

which can be expressed in a more explicit way as

$$\boldsymbol{\theta}_{0u} = \arg \min_{\boldsymbol{\theta}_0 \in \mathbb{R}^{2K(L-1)}} D_{\text{KL}} \{p_u(\mathbf{s}; \boldsymbol{\theta}_u) : p_0(\mathbf{s}; \boldsymbol{\theta}_0)\}. \quad (31)$$

In Sec. IV, we will present a detailed calculating procedure. Here, we focus on the next steps. Suppose that $\boldsymbol{\theta}_{0u}$ is obtained, we re-express the m -projection $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$ in the following way:

$$\begin{aligned} p_0(\mathbf{s}; \boldsymbol{\theta}_{0u}) &= \exp \left\{ (\mathbf{d} + \boldsymbol{\theta}_{0u})^T \mathbf{t} - \psi_0(\boldsymbol{\theta}_{0u}) \right\} \\ &= \exp \left\{ (\mathbf{d} + \boldsymbol{\theta}_u + \boldsymbol{\xi}_u)^T \mathbf{t} - \psi_0(\boldsymbol{\theta}_{0u}) \right\}, \end{aligned} \quad (32)$$

which implies that $\boldsymbol{\theta}_{0u} = \boldsymbol{\theta}_u + \boldsymbol{\xi}_u$. In (32), we consider the EACS $\boldsymbol{\theta}_{0u}$ of $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$ as a sum of the EACS $\boldsymbol{\theta}_u$ of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ and an extra item $\boldsymbol{\xi}_u$. Comparing $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ defined in (28b) and (32), we can find that $c_u(\mathbf{s}, \mathbf{y}_u)$ in $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ is replaced by $\boldsymbol{\xi}_u^T \mathbf{t}$

in $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$. Consequently, we consider the approximation of $c_u(\mathbf{s}, \mathbf{y}_u)$ as $\boldsymbol{\xi}_u^T \mathbf{t}$, where

$$\boldsymbol{\xi}_u = \boldsymbol{\theta}_{0u} - \boldsymbol{\theta}_u, u \in \mathcal{Z}_U^+. \quad (33)$$

Then, $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$ with $\boldsymbol{\theta}_0 = \sum_{u=1}^U \boldsymbol{\xi}_u$ is considered as the approximation of the *a posteriori* distribution $p(\mathbf{s}|\mathbf{y})$ since each $c_u(\mathbf{s}, \mathbf{y}_u)$ is approximated as $\boldsymbol{\xi}_u^T \mathbf{t}$ and $\boldsymbol{\theta}_0^T \mathbf{t}$ is regarded as an approximation of $\sum_u c_u(\mathbf{s}, \mathbf{y}_u)$.

Now, let us discuss the iteration, which is similar to that for IGA in [28]. At the beginning, the EACSs are initialized as $\{\boldsymbol{\theta}_u(0)\}_{u=0}^U$. Then, given the EACSs $\{\boldsymbol{\theta}_u(t)\}_{u=0}^U$ at the t -th iteration, calculate $\{\boldsymbol{\theta}_{0u}(t)\}_{u=1}^U$ and $\{\boldsymbol{\xi}_u(t)\}_{u=1}^U$ as (31) and (33), respectively. Next, the EACS of $p_u(\mathbf{s}; \boldsymbol{\theta}_u(t))$, $u \in \mathcal{Z}_U^+$ is updated as

$$\boldsymbol{\theta}_u(t+1) = \sum_{u'=1, u' \neq u}^U \boldsymbol{\xi}_{u'}(t),$$

since $\boldsymbol{\theta}_u^T(t+1) \mathbf{t}$ replaces the sum $\sum_{u' \neq u} c_{u'}(\mathbf{s}, \mathbf{y}_{u'})$ in $p_u(\mathbf{s}; \boldsymbol{\theta}_u(t+1))$ and each $c_u(\mathbf{s}, \mathbf{y}_u)$ is approximated as $\boldsymbol{\xi}_u^T(t) \mathbf{t}$ at the t -th iteration. We update the EACS of $p_0(\mathbf{s}; \boldsymbol{\theta}_0(t))$ as $\boldsymbol{\theta}_0(t+1) = \sum_{u=1}^U \boldsymbol{\xi}_u(t)$. In practice, the following damped updating way is typically employed to ensure the convergence: given a damping factor $0 < \alpha \leq 1$, update the EACSs as

$$\boldsymbol{\theta}_u(t+1) = \alpha \sum_{u'=1, u' \neq u}^U \boldsymbol{\xi}_{u'}(t) + (1-\alpha) \boldsymbol{\theta}_u(t), u \in \mathcal{Z}_U^+, \quad (34a)$$

$$\boldsymbol{\theta}_0(t+1) = \alpha \sum_{u=1}^U \boldsymbol{\xi}_u(t) + (1-\alpha) \boldsymbol{\theta}_0(t). \quad (34b)$$

Then, repeat the above process until convergence. We refer to the above process as GIGA, and we summarize it in Algorithm 1. Note that the procedure for calculating the m -projection in Step 1 will be discussed in the next section.

IV. CALCULATION OF THE m -PROJECTION FROM $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ ONTO THE OBM

In this section, we present the calculation of the m -projection from $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ onto the OBM. We first give its direct calculation. Then, based on the Berry-Esseen theorem, we propose an approximate calculation of the m -projection. The efficient implementation of the approximate calculation is also discussed. At last, we analyze the complexities of two types of calculations of m -projections.

A. Direct Calculation

The direct calculation of the m -projection in (31) is related to the marginal distributions of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$. To express the marginals of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$, let us first define $2K$ discrete random vectors with the same dimension, where the k -th vector is denoted as $\mathbf{s}_{\setminus k}$ and $\mathbf{s}_{\setminus k}$ is of $2K-1$ dimension. $\mathbf{s}_{\setminus k}$ is obtained by removing the k -th component, i.e., s_k , in \mathbf{s} . Based on the definition, we have $\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}$, $k \in \mathcal{Z}_{2K}^+$. Given $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$, $u \in$

Algorithm 1: GIGA

Input: The a priori probability $p_{\text{pr},k}(s_k)$, $k \in \mathcal{Z}_{2K}^+$, the received signal \mathbf{y} , the channel matrix \mathbf{G} , the alphabet $\mathbb{S} = \{s^{(0)}, s^{(1)}, \dots, s^{(L-1)}\}$, the number U of subsets, the noise power σ_z^2 and the maximal iteration number t_{max} .

Initialization: Set $t = 0$. Set damping α , where $0 < \alpha \leq 1$. Calculate \mathbf{y}_u and \mathbf{G}_u , $u \in \mathcal{Z}_U^+$, as (20) and (22), respectively. Initialize the EACSs $\{\boldsymbol{\theta}_u\}_{u=0}^U$ defined in (16a) and (29). Zeros are sufficient for the initializations. Calculate the NP $d_{k,\ell}$, $k \in \mathcal{Z}_{2K}^+$, $\ell \in \mathcal{Z}_{L-1}^+$, as (12c);

repeat

1. Calculate the EACS $\boldsymbol{\theta}_{0u}$, $u \in \mathcal{Z}_U^+$, of the m -projection in (31);
2. Calculate $\boldsymbol{\xi}_u(t)$, $n \in \mathcal{Z}_{2N_r}^+$, as (33);
3. Update the EACSs as (34);
4. $t = t + 1$;

until Convergence or $t > t_{\text{max}}$;

Output: The approximate marginals, $\{p_k(s_k|\mathbf{y})\}_{k=1}^{2K}$, are given by $\{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})\}_{k=1}^{2K}$, which is defined in (38). Perform the MPM detection as (8), and $\tilde{\mathbf{s}}$ is recovered as (9).

\mathcal{Z}_U^+ , we denote its marginal probability distribution of s_k as $p_{u,k}(s_k; \boldsymbol{\theta}_u)$, and we have

$$\begin{aligned} p_{u,k}(s_k; \boldsymbol{\theta}_u) &= \sum_{s_1 \in \mathbb{S}} \cdots \sum_{s_{k-1} \in \mathbb{S}} \sum_{s_{k+1} \in \mathbb{S}} \cdots \sum_{s_{2K} \in \mathbb{S}} p_u(\mathbf{s}; \boldsymbol{\theta}_u) \\ &= \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} p_u(\mathbf{s}; \boldsymbol{\theta}_u). \end{aligned} \quad (35)$$

Further, denote the EACS of the m -projection $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$ as

$$\boldsymbol{\theta}_{0u} = [\boldsymbol{\theta}_{0u,1}^T, \boldsymbol{\theta}_{0u,2}^T, \dots, \boldsymbol{\theta}_{0u,2K}^T]^T \in \mathbb{R}^{2K(L-1)}, \quad (36a)$$

$$\boldsymbol{\theta}_{0u,k} = [\theta_{0u,k,1}, \theta_{0u,k,2}, \dots, \theta_{0u,k,L-1}]^T \in \mathbb{R}^{(L-1)}. \quad (36b)$$

Based on (14), denote the marginals of $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$ as $p_0(s_k; \boldsymbol{\theta}_{0u,k})$, $k \in \mathcal{Z}_{2K}^+$. Then, the EACS $\boldsymbol{\theta}_{0u}$ of $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$ exists and is unique since $\text{KL}\{p(\mathbf{s}; \boldsymbol{\theta}_u) : p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})\}$ is a strictly convex function of $\boldsymbol{\theta}_{0u}$. More detailed proof can be seen in [28, Theorem 1], and we have

$$p_0(s_k; \boldsymbol{\theta}_{0u,k}) = p_u(s_k; \boldsymbol{\theta}_u), s_k \in \mathbb{S}, k \in \mathcal{Z}_{2K}^+. \quad (37)$$

We can find that the marginals of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ and its m -projection are the same.

We next discuss how to obtain the value of the EACS $\boldsymbol{\theta}_{0u}$ from (37). This is related to the property of p_0 of OBM. Given any $p_0(\mathbf{s}; \boldsymbol{\theta}_0)$ in the OBM and its marginals $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$,

we have [28]

$$p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(0)}} = \frac{1}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k,\ell} + \theta_{0,k,\ell}\}}, \quad (38a)$$

$$p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(\ell)}} = \frac{\exp\{d_{k,\ell} + \theta_{0,k,\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k,\ell} + \theta_{0,k,\ell}\}}, \quad (38b)$$

where $\ell \in \mathcal{Z}_{L-1}^+$ in (38b). On the contrary, given the probability in (38), the EACS of $p_{0,k}(s_k; \boldsymbol{\theta}_{0,k})$, $k \in \mathcal{Z}_{2K}^+$ can be expressed as

$$\theta_{0,k,\ell} = \ln \frac{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(\ell)}}}{p_{0,k}(s_k; \boldsymbol{\theta}_{0,k}) \Big|_{s_k=s^{(0)}}} - d_{k,\ell}, \ell \in \mathcal{Z}_{L-1}^+. \quad (39)$$

Since the m -projection $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$ belongs to the OBM, $\theta_{0u,k,\ell}$ in (37) can be expressed as

$$\theta_{0u,k,\ell} = \ln \frac{p_{u,k}(s_k; \boldsymbol{\theta}_u) \Big|_{s_k=s^{(\ell)}}}{p_{u,k}(s_k; \boldsymbol{\theta}_u) \Big|_{s_k=s^{(0)}}} - d_{k,\ell}, \quad (40)$$

which shows that the EACS of the m -projection $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$ is determined by the marginal probability of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$. On the other hand, the closed-form solution of $p_{u,k}(s_k; \boldsymbol{\theta}_u)$ can be hard to obtain. From (35) we can find that its calculation is of exponential-complexity. When the number of users and the modulation order are large, the computational complexity of (35) is unaffordable. Inspired by the Berry-Esseen theorem, we solve this problem by computing an approximation of the marginal $p_{u,k}(s_k; \boldsymbol{\theta}_u)$, $u \in \mathcal{Z}_U^+$, $k \in \mathcal{Z}_{2K}^+$, which will be discussed in the next subsection.

B. Approximate Calculation

As mentioned above, our focus now is to calculate the approximate marginals of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$. To do so, we first express $p_{u,k}(s_k; \boldsymbol{\theta}_u)$ as follows:

$$\begin{aligned} p_{u,k}(s_k; \boldsymbol{\theta}_u) &= \sum_{\mathbf{s} \setminus k \in \mathbb{S}^{2K-1}} \exp\left\{(\mathbf{d} + \boldsymbol{\theta}_u)^T \mathbf{t} \right. \\ &\quad \left. + c_u(\mathbf{s}, \mathbf{y}_u) - \psi_u\right\} \\ &\stackrel{(a)}{\propto} \lambda_{u,k}(s_k) \kappa_{u,k}(s_k, \mathbf{y}_u), \end{aligned} \quad (41)$$

where $u \in \mathcal{Z}_U^+$, $k \in \mathcal{Z}_{2K}^+$, $s_k \in \mathbb{S}$, (a) is obtained by removing the constant independent with s_k , and

$$\begin{aligned} \lambda_{u,k}(s_k) &\triangleq \exp\left\{(\mathbf{d}_k + \boldsymbol{\theta}_{u,k})^T \mathbf{t}_k\right\} \\ &= \exp\left\{\sum_{\ell=1}^{L-1} (d_{k,\ell} + \theta_{u,k,\ell}) \delta(s_k - s^{(\ell)})\right\}, \end{aligned} \quad (42a)$$

$$\begin{aligned} \kappa_{u,k}(s_k, \mathbf{y}_u) &\triangleq \sum_{\mathbf{s} \setminus k \in \mathbb{S}^{2K-1}} \exp\left\{\sum_{k'=1, k' \neq k}^{2K} (\mathbf{d}_{k'} + \boldsymbol{\theta}_{u,k'})^T \mathbf{t}_{k'} \right. \\ &\quad \left. + c_u(\mathbf{s}, \mathbf{y}_u)\right\}. \end{aligned} \quad (42b)$$

The reason why we explicitly parameterize $\kappa_{u,k}(s_k, \mathbf{y}_u)$ by both s_k and \mathbf{y}_u is that \mathbf{y}_u will play an important role when computing an approximation of $\kappa_{u,k}(s_k, \mathbf{y}_u)$. In (41), the value of $\lambda_{u,k}(s_k)$ can be calculated directly, since we have

$$\lambda_{u,k}(s_k) = \begin{cases} 1, & s_k = s^{(0)}, \\ \exp\{d_{k,\ell} + \theta_{u,k,\ell}\}, & s_k = s^{(\ell)}, \ell \in \mathcal{Z}_{L-1}^+. \end{cases} \quad (43)$$

Under these circumstances, calculating the approximation of $\kappa_{u,k}(s_k, \mathbf{y}_u)$, $s_k \in \mathbb{S}$, becomes the critical issue. Once it is obtained, the approximate value of $p_{u,k}(s_k; \boldsymbol{\theta}_u)$, $s_k \in \mathbb{S}$, can be acquired. Also, as a note, the proportion in (41) does not influence the computation of $p_{u,k}(s_k; \boldsymbol{\theta}_u)$ because the constant corresponding to the proportion is independent with s_k . Therefore, the value of $p_{u,k}(s_k; \boldsymbol{\theta}_u)$ can be always obtained through $\sum_{s_k \in \mathbb{S}} p_{u,k}(s_k; \boldsymbol{\theta}_u) = 1$. We will not repeat this issue when similar situations arise in the next.

For now, our attention shifts to the value of $\kappa_{u,k}(s_k, \mathbf{y}_u)$, $s_k \in \mathbb{S}$. According to its definition, $\kappa_{u,k}(s_k, \mathbf{y}_u)$ can be further expressed as

$$\begin{aligned} \kappa_{u,k}(s_k, \mathbf{y}_u) &= \sum_{\mathbf{s} \setminus k \in \mathbb{S}^{2K-1}} \left(\prod_{k'=1, k' \neq k}^{2K} \exp\left\{(\mathbf{d}_{k'} + \boldsymbol{\theta}_{u,k'})^T \mathbf{t}_{k'}\right\} \right. \\ &\quad \left. \times \exp\left\{-\frac{1}{2}(\mathbf{y}_u - \mathbf{G}_u \mathbf{s})^T (\sigma_z^2 \mathbf{I})^{-1} (\mathbf{y}_u - \mathbf{G}_u \mathbf{s})\right\} \right) \\ &\stackrel{(a)}{\propto} \sum_{\mathbf{s} \setminus k \in \mathbb{S}^{2K-1}} \left(\prod_{k'=1, k' \neq k}^{2K} p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{u,k'}) f_G(\mathbf{y}_u; \mathbf{G}_u \mathbf{s}, \sigma_z^2 \mathbf{I}) \right), \end{aligned} \quad (44)$$

where $u \in \mathcal{Z}_U^+$, $k \in \mathcal{Z}_{2K}^+$, and (a) comes from adding a constant independent with s_k and \mathbf{y}_u . Next we construct $2KU$ random vectors whose PDFs are in the same form as the last line of (44). The introduction of these random vectors is the key to compute the approximations of $\kappa_{u,k}(s_k, \mathbf{y}_u)$, $u \in \mathcal{Z}_U^+$, $k \in \mathcal{Z}_{2K}^+$. Define $U \times 2K$ random vectors

$$\{\bar{\mathbf{y}}_{u,k} | u \in \mathcal{Z}_U^+, k \in \mathcal{Z}_{2K}^+\}.$$

The (u, k) -th vector is given by

$$\begin{aligned} \bar{\mathbf{y}}_{u,k} &= \mathbf{g}_{u,k} s_k + \sum_{k'=1, k' \neq k}^{2K} \mathbf{g}_{u,k'} s_{k'} + \mathbf{w} \\ &= \sum_{k'=1, k' \neq k}^{2K} \mathbf{g}_{u,k'} s_{k'} + \mathbf{w}'_{u,k}, \end{aligned} \quad (45)$$

where $u \in \mathcal{Z}_U^+$, $k \in \mathcal{Z}_{2K}^+$, $\mathbf{g}_{u,k} \in \mathbb{R}^{N_u}$ is the k -th column of \mathbf{G}_u in (22), i.e.,

$$\mathbf{G}_u = [\mathbf{g}_{u,1}, \mathbf{g}_{u,2}, \dots, \mathbf{g}_{u,2K}],$$

$\mathbf{g}_{u,k}$ is considered as a determinate and known vector, s_k is considered as a determinate and known constant,

$$\{s_{k'}\}_{k'=1, k' \neq k}^{2K}$$

are considered as $2K - 1$ independent discrete random variables, their probability distributions are $\{p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{u,k'})\}_{k' \neq k}$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{I})$ is a Gaussian random vector of N_u dimension independent with $\{s_{k'}\}_{k' \neq k}$, and $\mathbf{w}'_{u,k} = \mathbf{w} + \mathbf{g}_{u,k} s_k \sim \mathcal{N}(\mathbf{g}_{u,k} s_k, \sigma_z^2 \mathbf{I})$ is also a Gaussian random vector independent with $\{s_{k'}\}_{k' \neq k}$. The joint probability distribution of $\{s_{k'}\}_{k' \neq k}$ in (45) is given by

$$p(\mathbf{s}_{\setminus k}) = \prod_{k'=1, k' \neq k}^{2K} p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{u,k'}).$$

Then, the PDF of $\bar{\mathbf{y}}_{u,k}$ can be expressed as [29, Sec. 6.1.2]

$$\begin{aligned} f(\bar{\mathbf{y}}_{u,k}) &= \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} \left(p(\mathbf{s}_{\setminus k}) f_G \left(\bar{\mathbf{y}}_{u,k} - \sum_{k' \neq k} \mathbf{g}_{u,k'} s_{k'}; \mathbf{g}_{u,k} s_k, \sigma_z^2 \mathbf{I} \right) \right) \\ &= \sum_{\mathbf{s}_{\setminus k} \in \mathbb{S}^{2K-1}} \left(p(\mathbf{s}_{\setminus k}) f_G(\bar{\mathbf{y}}_{u,k}; \mathbf{G}_u \mathbf{s}, \sigma_z^2 \mathbf{I}) \right). \end{aligned} \quad (46)$$

It is a direct result that the PDF of $\bar{\mathbf{y}}_{u,k}$ will be equivalent to the final line of (44) when we substitute $\bar{\mathbf{y}}_{u,k}$ with \mathbf{y}_u . Consequently, we can obtain

$$\kappa_{u,k}(s_k, \mathbf{y}_u) \propto f(\bar{\mathbf{y}}_{u,k}) \Big|_{\bar{\mathbf{y}}_{u,k} = \mathbf{y}_u}, s_k \in \mathbb{S}. \quad (47)$$

$\bar{\mathbf{y}}_{u,k}$ in (45) is a hybrid random vector, which is the sum of $2K - 1$ discrete random vectors and one Gaussian random vector. The closed-form solution of its PDF is difficult to obtain, and as can be seen from (46), its computational complexity is exponential. We note that $\bar{\mathbf{y}}_{u,k}$ is obtained by summing multiple mutually independent random vectors. This is somewhat similar to the situation described by the central limit theorem, with the difference that we are dealing with the summation of multiple random vectors. In this case, the classic central limit theorem can not be applied to obtain an approximation of the probability distribution of $\bar{\mathbf{y}}_{u,k}$. Fortunately, Berry-Esseen theorem [30] can help us to obtain such an approximation. We first present the Berry-Esseen theorem.

Lemma 1 (Berry-Esseen theorem): Given N independent random vectors $\{\mathbf{x}_n\}_{n=1}^N$, where $\mathbf{x}_n \in \mathbb{R}^d$, and $d \geq 1$ is finite. Each \mathbf{x}_n has finite mean $\boldsymbol{\mu}_n \in \mathbb{R}^d$ and finite positive-definite covariance matrix $\boldsymbol{\Sigma}_n \in \mathbb{R}^{d \times d}$. Denote the summation of $\{\mathbf{x}_n\}_{n=1}^N$ as $\mathbf{x}_s \triangleq \sum_{n=1}^N \mathbf{x}_n$. Its mean and covariance matrix are denoted as $\boldsymbol{\mu}_s \triangleq \sum_{n=1}^N \boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_s \triangleq \sum_{n=1}^N \boldsymbol{\Sigma}_n$, respectively. If the following condition

$$\lim_{N \rightarrow \infty} \sum_{n=1}^N \mathbb{E} \left\{ \left\| \boldsymbol{\Sigma}_s^{-\frac{1}{2}} (\mathbf{x}_n - \boldsymbol{\mu}_n) \right\|^3 \right\} = 0$$

holds. Then, \mathbf{x}_s converges in distribution to a real Gaussian random vector $\tilde{\mathbf{x}}_s$, as N tends to infinity, i.e.,

$$\mathbf{x}_s \rightarrow \tilde{\mathbf{x}}_s \sim \mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s).$$

Inspired by the Berry-Esseen theorem, we approximate the PDF of $\bar{\mathbf{y}}_{u,k}$ as a Gaussian PDF

$$f_G(\bar{\mathbf{y}}_{u,k}; \mathbb{E}\{\bar{\mathbf{y}}_{u,k}\}, \mathbb{V}\{\bar{\mathbf{y}}_{u,k}\}), \quad (48)$$

where $\mathbb{E}\{\bar{\mathbf{y}}_{u,k}\}$ and $\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\}$ are the mean and covariance of $\bar{\mathbf{y}}_{u,k}$, respectively. In comparison to the original PDF $f(\bar{\mathbf{y}}_{u,k})$, (48) has an explicit expression and it contains only linear operations. We next calculate the mean and covariance of $\bar{\mathbf{y}}_{u,k}$. To do so, we first calculate the mean and variance of $s_{k'}, k' \neq k$, in (45). The probability distribution of $s_{k'}, k' \neq k$, is $p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{u,k'})$. According to (38), the mean and the variance of $s_{k'}, k' \neq k$, are given by

$$\begin{aligned} \mu_{u,k'} &= \sum_{s_{k'} \in \mathbb{S}} s_{k'} p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{u,k'}) \\ &= \frac{s^{(0)} + \sum_{\ell=1}^{L-1} s^{(\ell)} \exp\{d_{k',\ell} + \theta_{u,k',\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k',\ell} + \theta_{u,k',\ell}\}}, \end{aligned} \quad (49a)$$

$$\begin{aligned} v_{u,k'} &= \sum_{s_{k'} \in \mathbb{S}} s_{k'}^2 p_{0,k'}(s_{k'}; \boldsymbol{\theta}_{u,k'}) - \mu_{u,k'}^2 \\ &= \frac{(s^{(0)})^2 + \sum_{\ell=1}^{L-1} (s^{(\ell)})^2 \exp\{d_{k',\ell} + \theta_{u,k',\ell}\}}{1 + \sum_{\ell=1}^{L-1} \exp\{d_{k',\ell} + \theta_{u,k',\ell}\}} - \mu_{u,k'}^2, \end{aligned} \quad (49b)$$

respectively. Consequently, the mean and covariance of the discrete random vector $\mathbf{g}_{u,k'} s_{k'}, k' \neq k$, in (45) are

$$\mathbb{E}\{\mathbf{g}_{u,k'} s_{k'}\} = \mathbf{g}_{u,k'} \mu_{u,k'}, \quad (50a)$$

$$\mathbb{V}\{\mathbf{g}_{u,k'} s_{k'}\} = v_{u,k'} \mathbf{g}_{u,k'} \mathbf{g}_{u,k'}^T, \quad (50b)$$

respectively. Then, we can readily obtain that the mean and covariance of $\bar{\mathbf{y}}_{u,k}$, $u \in \mathcal{Z}_U^+$, $k \in \mathcal{Z}_{2K}^+$, are

$$\mathbb{E}\{\bar{\mathbf{y}}_{u,k}\} = \left(\sum_{k'=1, k' \neq k}^{2K} \mathbf{g}_{u,k'} \mu_{u,k'} + \mathbf{g}_{u,k} s_k \right) \in \mathbb{R}^{N_u}, \quad (51a)$$

$$\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\} = \left(\sum_{k'=1, k' \neq k}^{2K} v_{u,k'} \mathbf{g}_{u,k'} \mathbf{g}_{u,k'}^T + \sigma_z^2 \mathbf{I} \right) \in \mathbb{R}^{N_u \times N_u}, \quad (51b)$$

respectively. And we have the following theorem.

Theorem 1: If the condition (53) holds for $\bar{\mathbf{y}}_{u,k}$ in (45), then $\bar{\mathbf{y}}_{u,k}$ converges in distribution to a real Gaussian random vector $\tilde{\mathbf{y}}_{u,k}$, as $2K$ goes to infinity, i.e.,

$$\bar{\mathbf{y}}_{u,k} \xrightarrow{d} \tilde{\mathbf{y}}_{u,k} \sim \mathcal{N}(\mathbb{E}\{\bar{\mathbf{y}}_{u,k}\}, \mathbb{V}\{\bar{\mathbf{y}}_{u,k}\}). \quad (52)$$

From Theorem 1, we can obtain that when $2K$ is large and the condition (53) approximately holds, the PDFs of $\bar{\mathbf{y}}_{u,k}$ and $\tilde{\mathbf{y}}_{u,k}$ are approximately equivalent. In one of the simplest cases, assuming that $\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\}$ is a diagonal covariance matrix, condition (53) holds as long as the variance of each component of $\mathbf{g}_{u,k'} s_{k'}, k' \neq k$, in (45) does not tend to zero as K tends to infinity. This ensures that $\mathbf{g}_{u,k'} s_{k'}, k' \neq k$, is random

rather than deterministic, which is necessary for the application of the Berry-Esseen theorem. By replacing $f(\bar{\mathbf{y}}_{u,k})$ with $f_G(\bar{\mathbf{y}}_{u,k}; \mathbb{E}\{\bar{\mathbf{y}}_{u,k}\}, \mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})$ in (47), we can obtain

$$\kappa_{u,k}(s_k, \mathbf{y}_u) \propto f_G(\bar{\mathbf{y}}_{u,k}; \mathbb{E}\{\bar{\mathbf{y}}_{u,k}\}, \mathbb{V}\{\bar{\mathbf{y}}_{u,k}\}) \Big|_{\bar{\mathbf{y}}_{u,k}=\mathbf{y}_u}. \quad (54)$$

Combining (41) and (54), we can obtain (55), where

$$\begin{aligned} \mathbf{a}_{u,k} &\triangleq \mathbf{y}_u - \sum_{k'=1, k' \neq k}^{2K} \mathbf{g}_{u,k'} \mu_{u,k'} \\ &= \mathbf{y}_u - \mathbf{G}_u \boldsymbol{\mu}_u + \mathbf{g}_{u,k} \mu_{u,k} \in \mathbb{R}^{N_u}, \\ \boldsymbol{\mu}_u &\triangleq [\mu_{u,1}, \mu_{u,2}, \dots, \mu_{u,2K}]^T \in \mathbb{R}^{2K}, \end{aligned} \quad (56)$$

$s_k \in \mathbb{S}$, $k \in \mathcal{Z}_{2K}^+$, $u \in \mathcal{Z}_U^+$, and the derivation is given in Appendix A. Substituting $s_k = s^{(\ell)}$ into (55), we can obtain

$$p_{u,k}(s_k; \boldsymbol{\theta}_u) \Big|_{s_k=s^{(0)}} = C_{u,k} \exp \left\{ -\frac{(s^{(0)} - \tilde{\mu}_{u,k})^2}{2\tilde{v}_{u,k}} \right\}, \quad (57a)$$

$$p_{u,k}(s_k; \boldsymbol{\theta}_u) \Big|_{s_k=s^{(\ell)}} = C_{u,k} \exp \left\{ d_{k,\ell} + \theta_{u,k,\ell} - \frac{(s^{(\ell)} - \tilde{\mu}_{u,k})^2}{2\tilde{v}_{u,k}} \right\}, \quad (57b)$$

where

$$\tilde{v}_{u,k} \triangleq \frac{1}{\mathbf{g}_{u,k}^T (\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{g}_{u,k}}, \quad (58a)$$

$$\tilde{\mu}_{u,k} \triangleq \tilde{v}_{u,k} \mathbf{g}_{u,k}^T (\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{a}_{u,k}, \quad (58b)$$

$C_{u,k}$ is the normalization factor, and $\ell \in \mathcal{Z}_{L-1}^+$ in (57b). Consequently, according to the definition in (36), the relationship in (40) and (57), the EACS of the m -projection $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$ can be calculated as

$$\boldsymbol{\theta}_{0u} = [\boldsymbol{\theta}_{0u,1}^T, \boldsymbol{\theta}_{0u,2}^T, \dots, \boldsymbol{\theta}_{0u,2K}^T]^T, \quad (59a)$$

$$\boldsymbol{\theta}_{0u,k} = [\theta_{0u,k,1}, \theta_{0u,k,2}, \dots, \theta_{0u,k,L-1}]^T, \quad (59b)$$

$$\theta_{0u,k,\ell} = \frac{(s^{(0)} - s^{(\ell)}) [(s^{(0)} + s^{(\ell)}) - 2\tilde{\mu}_{u,k}]}{2\tilde{v}_{u,k}} + \theta_{u,k,\ell}, \quad (59c)$$

where $u \in \mathcal{Z}_U^+$, $k \in \mathcal{Z}_{2K}^+$, and $\ell \in \mathcal{Z}_{L-1}^+$.

We next discuss the efficient implementation of the approximate calculation. In the approximate calculation, the calculation of the inversions of $\{\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\} | u \in \mathcal{Z}_U^+, k \in \mathcal{Z}_{2K}^+\}$ in (58) is the most complex. Direct calculations of these inversions will traverse both the subscripts u and k , and thus introduce a total of $2UK$ matrix inversions of $N_u \times N_u$ dimension. Given U and K , we next reduce the total number of inversions to U by the means of Sherman-Morrison formula. Recalling the definition of $\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\}$ in (51b), we have

$$\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\} = \mathbf{G}_u \text{Diag}\{\mathbf{v}_u\} \mathbf{G}_u^T + \sigma_z^2 \mathbf{I} - v_{u,k} \mathbf{g}_{u,k} \mathbf{g}_{u,k}^T, \quad (60)$$

where $\mathbf{G}_u \in \mathbb{R}^{N_u \times 2K}$ is defined in (22), \mathbf{v}_u is defined as

$$\mathbf{v}_u \triangleq [v_{u,1}, v_{u,2}, \dots, v_{u,2K}]^T \in \mathbb{R}^{2K}, \quad (61)$$

and $v_{u,k}$ is given in (49b). According to Sherman-Morrison formula, we can obtain

$$\begin{aligned} &(\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})^{-1} \\ &= \mathbf{A}_u + \frac{v_{u,k}}{1 - v_{u,k} \mathbf{g}_{u,k}^T \mathbf{A}_u \mathbf{g}_{u,k}} \mathbf{A}_u \mathbf{g}_{u,k} (\mathbf{A}_u \mathbf{g}_{u,k})^T, \end{aligned} \quad (62)$$

where

$$\mathbf{A}_u \triangleq (\mathbf{G}_u \text{Diag}\{\mathbf{v}_u\} \mathbf{G}_u^T + \sigma_z^2 \mathbf{I})^{-1} \in \mathbb{R}^{N_u \times N_u} \quad (63)$$

is symmetric. In (62), we can find that \mathbf{A}_u only varies with u , and hence we only need to compute U inverse matrices of $N_u \times N_u$ dimension to obtain

$$\left\{ (\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})^{-1} | u \in \mathcal{Z}_U^+, k \in \mathcal{Z}_{2K}^+ \right\}.$$

Now, we discuss the calculation of \mathbf{A}_u . To obtain the inversion in (63), we can either follow (63) directly, or, based on the Woodbury identity, in the following way:

$$\mathbf{A}_u = \frac{1}{\sigma_z^2} \mathbf{I} - \frac{1}{\sigma_z^4} \mathbf{G}_u \left(\text{Diag}^{-1}\{\mathbf{v}_u\} + \frac{1}{\sigma_z^2} \mathbf{G}_u^T \mathbf{G}_u \right)^{-1} \mathbf{G}_u^T. \quad (64)$$

The computational complexities (the number of real-valued multiplications) of (63) and (64) are $\mathcal{O}(P)$ and $\mathcal{O}(Q)$, respectively, where

$$P = N_u^3 + 2KN_u^2, \quad (65a)$$

$$Q = 8K^3 + 8K^2N_u + 2KN_u^2. \quad (65b)$$

$$\lim_{K \rightarrow \infty} \left(\sum_{k' \neq k, k'=1}^{2K} \mathbb{E} \left\{ \left\| (\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})^{-\frac{1}{2}} \mathbf{g}_{u,k'} (s_{k'} - \mu_{u,k'}) \right\|^3 \right\} + \mathbb{E} \left\{ \left\| (\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})^{-\frac{1}{2}} (\mathbf{w}'_{u,k} - \mathbf{g}_{u,k} s_k) \right\|^3 \right\} \right) = 0 \quad (53)$$

$$p_{u,k}(s_k; \boldsymbol{\theta}_u) \propto \exp \left\{ \sum_{\ell=1}^{L-1} (d_{k,\ell} + \theta_{u,k,\ell}) \delta(s_k - s^{(\ell)}) - \frac{\mathbf{g}_{u,k}^T (\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{g}_{u,k}}{2} \left(s_k - \frac{\mathbf{g}_{u,k}^T (\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{a}_{u,k}}{\mathbf{g}_{u,k}^T (\mathbb{V}\{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{g}_{u,k}} \right)^2 \right\} \quad (55)$$

Algorithm 2: Approximate Calculation of the m -Projection from $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ onto the OBM

- Input:** Sub-vector \mathbf{y}_u , sub-matrix \mathbf{G}_u , number of users K , dimension N_u of \mathbf{y}_u , the alphabet $\mathbb{S} = \{s^{(0)}, s^{(1)}, \dots, s^{(L-1)}\}$, the NP \mathbf{d} , the EACS $\boldsymbol{\theta}_u$ of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$, the noise variance σ_z^2 .
1. Calculate $\{\mu_{u,k}\}_{k=1}^{2K}$ and $\{v_{u,k}\}_{k=1}^{2K}$ as (49);
 2. Calculate $\{\mathbf{a}_{u,k}\}_{k=1}^{2K}$ as (56);
 3. Calculate \mathbf{v}_u as (61);
 4. Calculate P and Q as (65);
 5. **if** $P \leq Q$ **then**
 Calculate \mathbf{A}_u as (63);
 else
 Calculate \mathbf{A}_u as (64);
 end if
 6. Calculate $(\mathbb{V}\{\bar{\mathbf{y}}\}_{u,k})^{-1}$, $k \in \mathcal{Z}_{2K}^+$, as (62);
 7. Calculate $\{\tilde{\mu}_{u,k}\}_{k=1}^{2K}$ and $\{\tilde{v}_{u,k}\}_{k=1}^{2K}$ as (58);
 8. Calculate EACS $\boldsymbol{\theta}_{0u}$ as (59);
- Output:** The EACS $\boldsymbol{\theta}_{0u}$ of the m -projection $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$.

In practice, we use the one with a lower complexity to get \mathbf{A}_u . We summarize the approximate calculation of the m -projection as Algorithm 2. It is not difficult to check that the IGA for signal detection in [28] is a special case of GIGA with $U = 2N_r$. From the following simulation results, we can find that GIGA can obtain a compromise between BER performance, convergence speed and computational complexity by adjusting the number U of groups.

C. Computational Complexity

We now give the computational complexity of the two ways of calculating the m -projection $p_0(\mathbf{s}; \boldsymbol{\theta}_{0u})$. In this work, we use the number of real-valued multiplications as the measure for computational complexity. The complexity of direct calculation is $\mathcal{O}(L^{2K})$ since we can only obtain the marginal probabilities of $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ by means of an exhaustive search. When the modulation order and the number of users are large, direct calculation will be unaffordable for practical. According to the steps of Algorithm 2, we can obtain that the computational complexity of approximately calculating the m -projection from $p_u(\mathbf{s}; \boldsymbol{\theta}_u)$ onto the OBM is

$$\mathcal{O}(\min(P, Q) + 6KN_u^2 + 4KL),$$

where P and Q are given by (65), K is the number of users, $N_u = \frac{2N_r}{U}$, N_r is the number of antennas at BS, $L = \sqrt{\tilde{L}}$, and \tilde{L} is the modulation order. Consequently, the computational complexity of GIGA with approximate calculation of m -projection is $\mathcal{O}(C_U)$ per iteration, where

$$C_U = U \min(P, Q) + 24K \frac{N_r^2}{U} + 4KUL, \quad (66)$$

U is the number of subsets. As U increases from 1, C_U decreases. However, when U is greater than a certain threshold,

TABLE I
PARAMETER SETTINGS OF THE SIMULATION

Parameter	Value
Number of BS antennas $N_{r,v} \times N_{r,h}$	16×64
UT number K	240
Center frequency f_c	4.8GHz
Simulation scenario	3GPP_38.901_UMa
Modulation Mode	QAM
Modulation Order \tilde{L}	4, 16, 64
Number of subsets U	16, 128, 512, 2048

C_U will start to increase. We will see this observation in the simulation results below. This also shows that the computational complexity of GIGA with $U = 2N_r$ (which is equivalent to IGA in [28]) is not necessarily the lowest compared to the cases when U takes other values.

V. SIMULATION RESULTS

This section provides simulation results to illustrate the performance of GIGA in ultra-massive MIMO system. The uncoded BER is used to measure the detection performance in the simulations. For generating the channel matrix, we employ the widely-used QuaDRiGa [31]. In QuaDRiGa, the BS consists of a uniform planar array (UPA) with $N_r = N_{r,v} \times N_{r,h}$ antennas, where $N_{r,v}$ and $N_{r,h}$ are the numbers of the antennas at each vertical column and horizontal row, respectively. The BS is positioned at coordinates $(0, 0, 25)$. Users are randomly distributed within a 120° sector with a radius of $r = 200\text{m}$ around $(0, 0, 1.5)$. Our results are averaged for 1000 channel matrix realizations. We summarize the main simulation parameters in Table I. Each user's average transmitted symbol power is normalized to 1. The SNR is defined as $\text{SNR} = K/\tilde{\sigma}_z^2$. Based on the received signal model (6), we compare the proposed GIGA with the following detectors.

LMMSE: The classic LMMSE detector with hard-decision,

$$\mathbf{s}_{\text{LMMSE}} = (\mathbf{G}^T \mathbf{G} + \sigma_z^2 \mathbf{I})^{-1} \mathbf{G}^T \mathbf{y}, \quad (67)$$

where the hard-decision is performed as

$$s_{k,\text{LMMSE}} = \arg \min_{s_k \in \mathbb{S}} |s_k - [\mathbf{s}_{\text{LMMSE}}]_k|^2, k \in \mathcal{Z}_{2K}^+. \quad (68)$$

The computational complexity of the LMMSE detector is $\mathcal{O}(8(2N_r K^2 + K^3))$ [8].

EP: The EP detector proposed in [8], where the hard-decision is also performed. Its computational complexity is $\mathcal{O}(8(N_r K^2 + K^3))$ per iteration [8].

AMP: The classic Bayesian inference algorithm AMP [32]. AMP could approximately calculate the *a posteriori* marginals. Then, the MPM detection (8) is performed. The computational complexity of AMP is $\mathcal{O}(8(N_r K))$ per iteration [32].

A. BER Performance

We begin with 4-QAM modulation. The BER performances of different algorithms are shown in Fig. 2. The iteration numbers of GIGA with $U = 16$, $U = 128$, $U = 512$ and $U = 2048$ are set to be 7, 10, 15 and 15, respectively. The iteration numbers of AMP and EP are set to be 30 and 40, respectively.

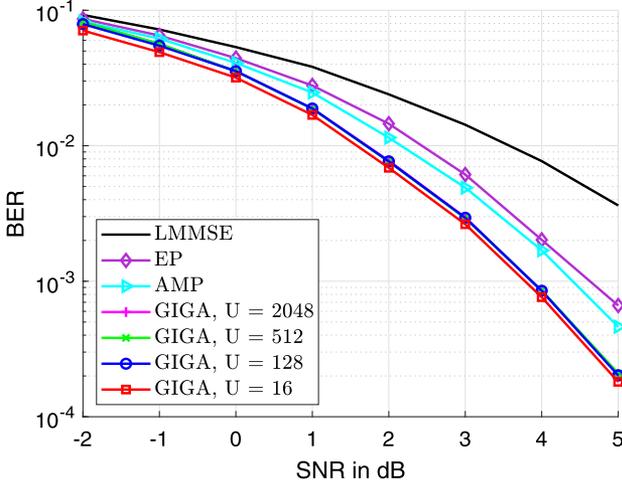


Fig. 2. BER performance of GIGA compared with AMP, EP and LMMSE under 4-QAM.

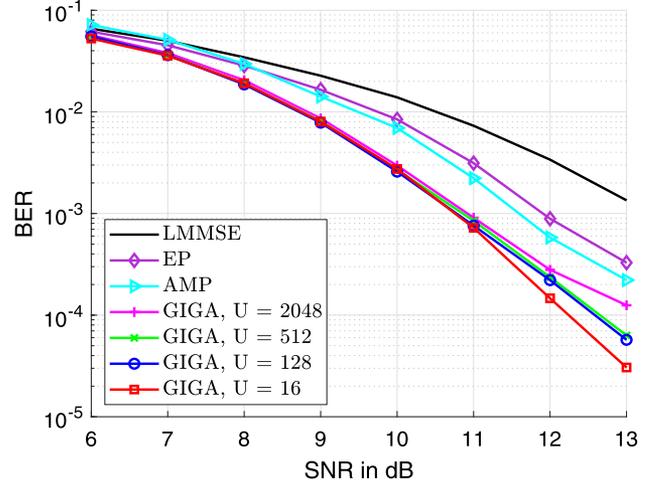


Fig. 4. BER performance of GIGA compared with AMP, EP and LMMSE under 16-QAM.

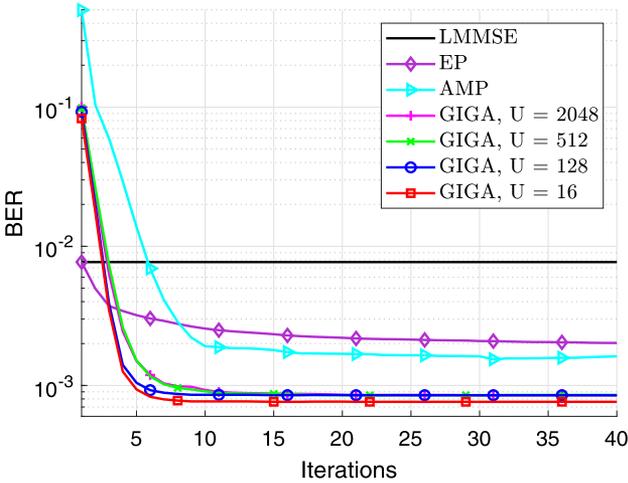


Fig. 3. Convergence performance of GIGA compared with EP and AMP at SNR = 4 dB under 4-QAM.

In Fig. 2, the BER performance of all the iterative algorithms outperform that of LMMSE detector. For GIGA, the difference in BER performance between different numbers of subsets is small. At a BER of 10^{-3} , the SNR gains of GIGA over AMP and EP are approximately 0.5dB and 0.7dB, respectively. Fig. 3 shows the convergence performance of all the iterative algorithms where the SNR is set as 4dB. In this case, with the increase of U , the number of iterations for GIGA to converge increases, while the BER performance decreases. On the other hand, overall, the performance gap and the difference in the number of iterations between different U is limited. AMP and EP converge in around 30 and 40 iterations, respectively. Additionally, it's worth noting that the BER performance of EP with a single iteration is equivalent to that of LMMSE, as the EP detector with one iteration is tantamount to the LMMSE detector [8].

Figs. 4 and 5 show the BER performances for 16-QAM and 64-QAM, respectively. In Fig. 4, the iteration numbers for GIGA with $U = 16$, $U = 128$, $U = 512$, and $U = 2048$ are set as 15, 20, 25, and 30, respectively. The iteration numbers of AMP and EP are set as 50 and 40, respectively. It can be found that GIGA still achieves the best BER performance. The performance gap between GIGA with different U becomes obvious when SNR is high. The BER performance of GIGA with $U = 16$ at SNR = 12dB is close to that of GIGA with $U = 2048$ at 13dB. At a BER of 10^{-3} , the SNR gains of GIGA over AMP and EP are about 0.7dB and 0.9dB, respectively. For 64-QAM, the iteration numbers of GIGA with $U = 16$, $U = 128$, $U = 512$, and $U = 2048$ are set as 15, 25, 30, and 35, respectively. The iteration numbers of AMP and EP are set as 70 and 40, respectively. Similar observations to those in Fig. 4 can be obtained. The BER performance of GIGA with $U = 16$ at SNR = 18dB is close to that of GIGA with $U = 2048$ at 18.5dB. At a BER of 10^{-3} , the SNR gains of GIGA over AMP and EP are around 0.7dB and 1dB, respectively. The LMMSE detector provides limited performance in ultra-massive MIMO system since the multi-dimensional Gaussian approximation in (67) is not a feasible model for massive MIMO systems with high-order constellations [8].

Figs. 6 and 7 show the convergence performances of all the iterative algorithms under 16-QAM and 64-QAM, respectively. From Fig. 6, we can find that in the scenario with 16-QAM and SNR = 12dB, the smaller the number of subsets in GIGA, the higher its convergence rate and the lower BER that can be obtained. The disparity between different U is much more pronounced than that with 4-QAM. GIGA with $U = 16$, $U = 128$, $U = 512$, and $U = 2048$ require about 10, 15, 20, and 30 iterations to converge, respectively. AMP and EP converge in around 50 and 35 iterations, respectively. From Fig. 7, we observe that in the scenario with 64-QAM and SNR of 18dB, disparity between GIGA with different U is still observable. GIGA with $U = 16$ converges within 15 iterations while GIGA with $U = 128$, 512, and 2048 converge in around 25, 35, and

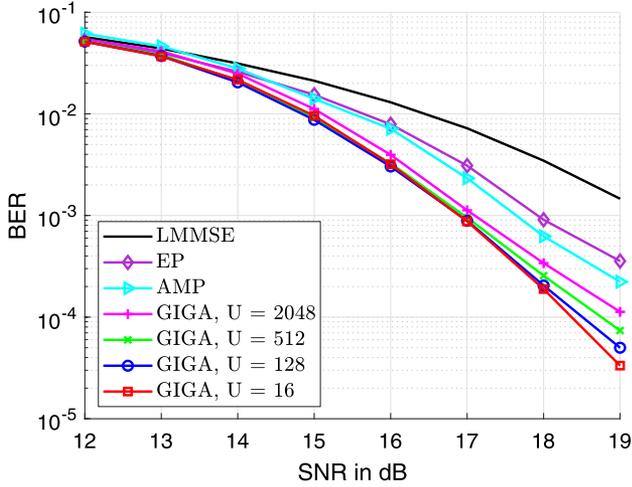


Fig. 5. BER performance of GIGA compared with AMP, EP and LMMSE under 64-QAM.

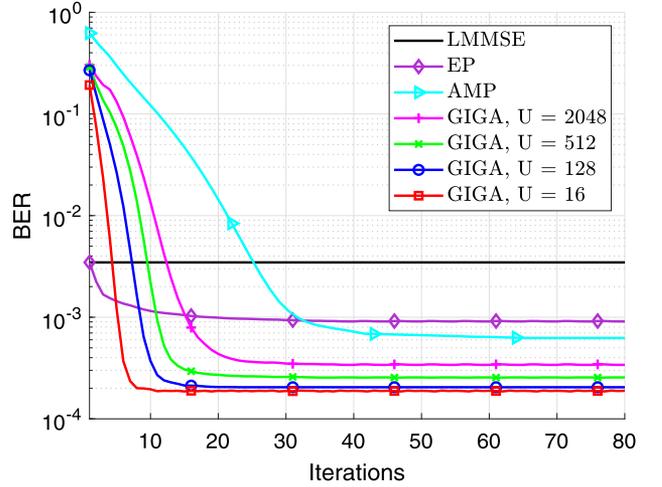


Fig. 7. Convergence performance of GIGA compared with EP and AMP at SNR = 18 dB under 64-QAM.

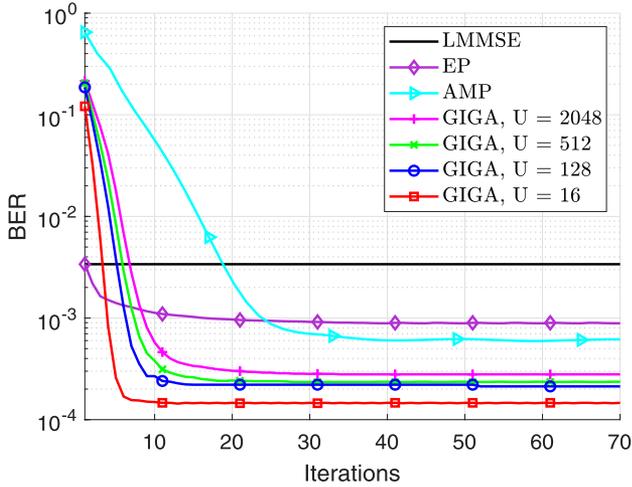


Fig. 6. Convergence performance of GIGA compared with EP and AMP at SNR = 12 dB under 16-QAM.

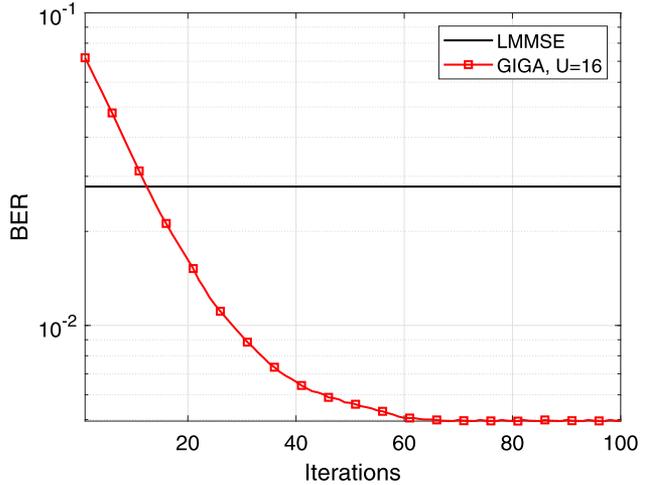


Fig. 8. Convergence performance of GIGA compared with LMMSE at SNR = 37 dB under 4-QAM.

50 iterations, respectively. AMP and EP require about 60 and 30 iterations to converge, respectively.

The simulation above is primarily conducted with $N_r = 1024$ and $K = 240$. With higher K/N_r ratios, similar convergence and BER performance can be observed for GIGA, and compared to LMMSE, AMP and EP, GIGA still provides improved performance.

Next, we show the BER performance of GIGA compared to LMMSE under high K/N_r ratios. It is known that when the number of receive antennas is significantly larger than the number of transmit antennas, even linear methods perform very well. Although GIGA is targeted for ultra-massive MIMO systems, it can still achieve improved performance under high K/N_r ratios. We consider the case $K = N_r$. In Fig. 8, it can be found that in the scenario with 4-QAM and SNR of 37 dB, the BER performance gap between GIGA with $U = 16$ and

LMMSE increases typically, and the convergence rate of GIGA becomes slower compared to the case of $K/N_r < 1$.

Fig. 9 presents the BER performance of GIGA with $U = 16$ and LMMSE under 4-QAM. We can find that although LMMSE and GIGA both have worse BER performance in the case with $K = N_r$, GIGA can still achieve improved BER performance compared to LMMSE.

B. Mutual Information Performance

It is common to use mutual information (MI) to evaluate the performance of a MIMO detector. The MI of the equivalent bit channel is used to judge on the quality of LLR soft output for a coded system. Since uncertainty quantification is important for soft input channel decoders, evaluating the MI between the transmitted bits per channel use and the detector's soft outputs can provide an understanding of the system's performance. We assume that each user selects symbols uniformly from the symbol set and no encoding method is currently used. Under

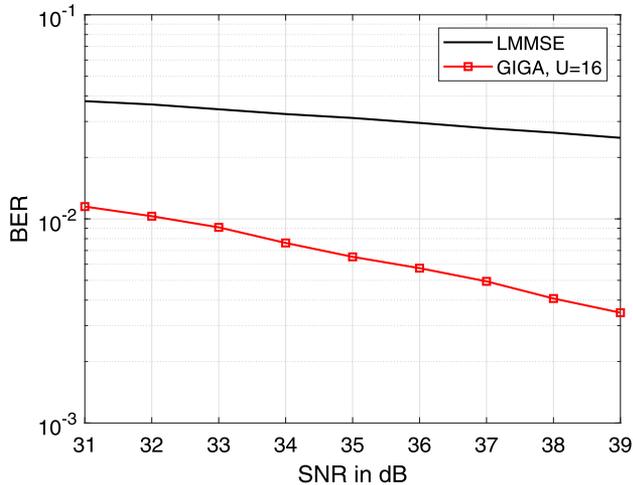


Fig. 9. BER performance of GIGA compared with LMMSE under 4-QAM.

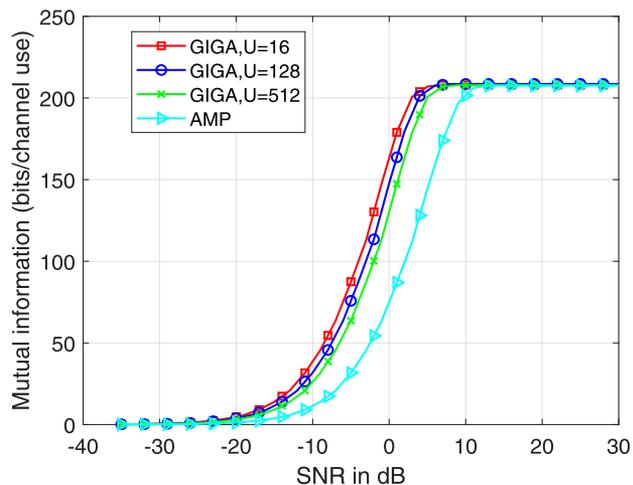


Fig. 10. Mutual information performance of GIGA compared with AMP.

finite-alphabet inputs, an approximated but efficient method for evaluating the mutual information is given by [33]:

$$I(\mathbf{y}; \mathbf{s}) = \sum_{i=1}^K I(s_i; \hat{s}_i) + C, \quad (69)$$

where C is a constant, \hat{s}_i is an estimate of s_i and $I(s_i; \hat{s}_i)$ can be calculated as

$$I(s_i | \hat{s}_i) = H(s_i) - \lim_{N_{sa} \rightarrow \infty} \frac{1}{N_{sa}} \sum_{n=1}^{N_{sa}} H(s_i | \mathbf{y}). \quad (70)$$

where N_{sa} is the number of samples. Fig. 10 presents the mutual information obtained by AMP and GIGA with $U = 16$, $U = 128$, and $U = 512$ under 4-QAM. It can be observed that GIGA has improved mutual information performance compared with AMP. Furthermore, as the number U of groups decreases, the mutual information performance of GIGA improves.

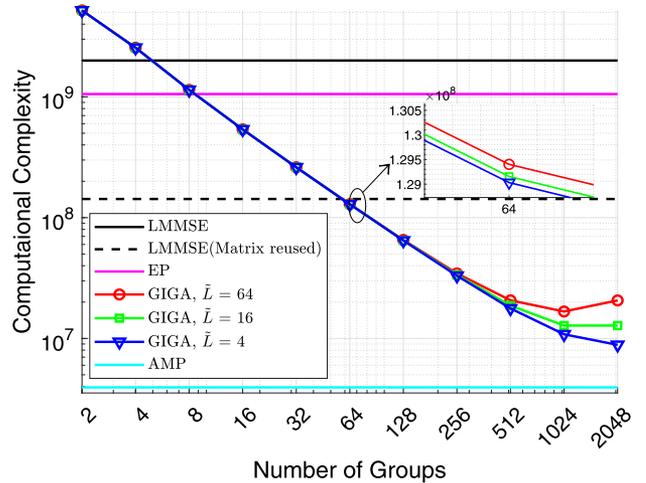


Fig. 11. Complexities of different algorithms versus the number of subsets.

C. Complexities

The computational complexities of different algorithms are plotted in Fig. 11. The x-axis is the number of subsets in GIGA. The numbers of iterations of GIGA, EP and AMP are all set to be 1. Among all the iterative algorithms, AMP has the lowest computational complexity. The complexity of GIGA decreases gradually as the number U of subsets increases. Also, we can find a special case when the modulation order is 64, its computational complexity at $U = 2048$ is slightly higher than that of $U = 1024$. When $U > 8$, the computational complexity of GIGA becomes lower than that of EP. The gap between the two increases rapidly with the number of subsets.

We now discuss the overall computational complexities of different algorithms in our simulations. Among all algorithms, the complexity of EP is the highest. Although its complexity at each iteration is lower than that of the LMMSE detection, we can see from Figs. 3, 6, and 7 that EP requires about tens of iterations to converge, which leads to its highest overall computation. In our simulations, the number of subsets for GIGA is set to be 16, 128, 512 and 2048, respectively. From Figs. 3, 6, and 7, we can find that GIGA with $U = 16$ converges within ten iterations, while under other U , it converges in tens of iterations. Under this condition, the overall computational complexity of GIGA with $U = 16$ is comparable to that of the LMMSE detection, while under other U , its overall computational complexity is lower than that of the LMMSE detection. Still, AMP has the lowest overall computational complexity. However, it is worth noting that for LMMSE detector the LMMSE matrix can be calculated once for a set of channel uses where the channels keep constant, which is computationally much less expensive. Fig. 11 also illustrates the complexity of LMMSE when the LMMSE matrix is reused, where the channel is assumed constant over 14 symbols. When the LMMSE matrix is reused, the LMMSE detector does offer an advantage in terms of computational complexity. On the other hand, in some scenarios where there are relaxed computational complexity requirements but high demands for detection performance, GIGA method may offer

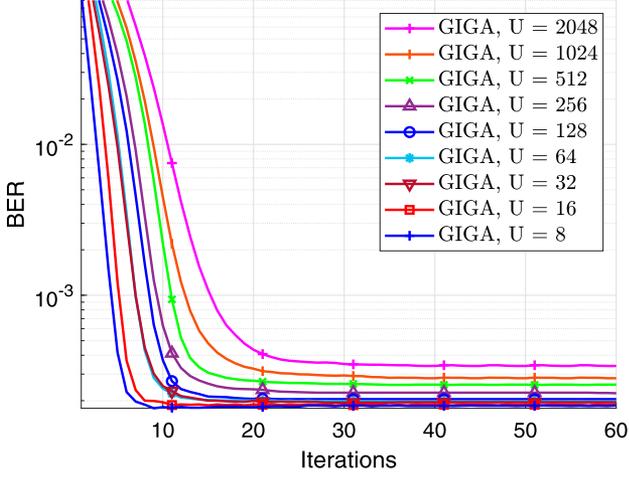


Fig. 12. Convergence performance of GIGA with multiple U at SNR = 18 dB under 64-QAM.

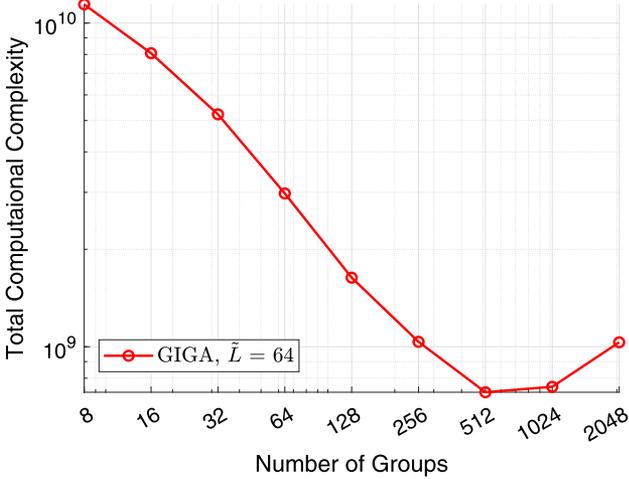


Fig. 13. Total complexities of GIGA versus the number of subsets under 64-QAM.

a more flexible choice since GIGA can adjust its grouping strategy to optimize for different cases.

To further illustrate the overall computational complexity of GIGA with different number U of subsets, we first plot the convergence performance of GIGA with more U in Fig. 12, where the modulation is 64-QAM and SNR = 18dB. Then, Fig. 13 illustrates the overall computational complexity of GIGA with different U . For each U , the overall computational complexity is defined as $T_U C_U$, where T_U is the number of iterations required for GIGA to converge when the number of subsets is U , and C_U is the complexity of its single iteration. From Fig. 13, we can find that increasing the number of subsets does not necessarily reduce the overall computational complexity. Larger U brings smaller single iteration complexity, but it also brings more iterations. In the case with 64-QAM and SNR = 18dB, the overall computational complexity is lowest when $U = 512$, and the gap between its BER performance and the best BER performance brought by $U = 8$ is relatively small.

VI. CONCLUSION

In this paper, we have proposed GIGA for ultra-massive MIMO systems. We frame the signal detection as an MPM detection problem. Leveraging information geometry theory, our objective is to compute approximations of the *a posteriori* marginals of the transmitted symbols. Through grouping the components of the received signal \mathbf{y} , we factorize the *a posteriori* probability distribution. On this basis, we define the AMs, where each AM is related to one group of components of \mathbf{y} . Then, we calculate the approximations of the *a posteriori* marginals through the m -projections from the distributions of AMs onto the OBM. We give a direct calculation of the m -projection as well as an approximate calculation based on the Berry-Esseen theorem. Efficient implementation of approximate calculation is also discussed. Simulation results demonstrate that GIGA achieves the best BER performance within a limited number of iterations compared to existing approaches. And we can flexibly make the grouping choice to serve different needs considering computational complexity and performance when GIGA is applied in a specific system configuration. This showcases the potential of GIGA as an efficient and potent detector in ultra-massive MIMO systems. As a final remark, although this paper only considers the case when all the groups have the same size, it is straightforward to generalize to the case when the groups have different sizes.

APPENDIX A CALCULATION OF (55)

From (41) and (54), we can obtain

$$p_{u,k}(s_k; \boldsymbol{\theta}_u) \stackrel{(a)}{\propto} \exp \left\{ (\mathbf{d}_k + \boldsymbol{\theta}_{u,k})^T \mathbf{t}_k - \frac{1}{2} \Theta_{u,k} \right\}, \quad (72)$$

where (a) is obtained by adding the constant independent with s_k and \mathbf{y}_u , and

$$\Theta_{u,k} = (\mathbf{y}_u - \mathbb{E} \{ \bar{\mathbf{y}}_{u,k} \})^T (\mathbb{V} \{ \bar{\mathbf{y}}_{u,k} \})^{-1} (\mathbf{y}_u - \mathbb{E} \{ \bar{\mathbf{y}}_{u,k} \}). \quad (73)$$

Combining (51a), $(\mathbf{y}_u - \mathbb{E} \{ \bar{\mathbf{y}}_{u,k} \})$ in (73) can be expressed as

$$\mathbf{y}_u - \mathbb{E} \{ \bar{\mathbf{y}}_{u,k} \} = - \left[\mathbf{g}_{u,k} s_k - \left(\mathbf{y}_u - \sum_{k'=1, k' \neq k}^{2K} \mathbf{g}_{u,k'} \mu_{u,k'} \right) \right]. \quad (74)$$

Denote $\mathbf{a}_{u,k}$ as

$$\mathbf{a}_{u,k} \triangleq \mathbf{y}_u - \sum_{k'=1, k' \neq k}^{2K} \mathbf{g}_{u,k'} \mu_{u,k'}.$$

Substituting $\mathbf{a}_{u,k}$ into (73), we can obtain (71), where (a) comes from removing a constant independent with s_k and $\mathbb{V} \{ \bar{\mathbf{y}}_{u,k} \}$ is symmetric and (b) comes from adding a constant independent with s_k . Substituting (71) into (72), we can obtain (55).

$$\Theta_{u,k} = (\mathbf{g}_{u,k} s_k - \mathbf{a}_{u,k})^T (\mathbb{V} \{\bar{\mathbf{y}}_{u,k}\})^{-1} (\mathbf{g}_{u,k} s_k - \mathbf{a}_{u,k}) \stackrel{(a)}{\propto} \mathbf{g}_{u,k}^T (\mathbb{V} \{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{g}_{u,k} s_k^2 - 2\mathbf{g}_{u,k}^T (\mathbb{V} \{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{a}_{u,k} s_k \stackrel{(b)}{\propto} \mathbf{g}_{u,k}^T (\mathbb{V} \{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{g}_{u,k} \left(s_k - \frac{\mathbf{g}_{u,k}^T (\mathbb{V} \{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{a}_{u,k}}{\mathbf{g}_{u,k}^T (\mathbb{V} \{\bar{\mathbf{y}}_{u,k}\})^{-1} \mathbf{g}_{u,k}} \right)^2 \quad (71)$$

REFERENCES

- [1] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2663–2677, Apr. 2022.
- [2] Y. Zhu, H. Guo, and V. K. N. Lau, "Bayesian channel estimation in multi-user massive MIMO with extremely large antenna array," *IEEE Trans. Signal Process.*, vol. 69, pp. 5463–5478, 2021.
- [3] H. Lu and Y. Zeng, "Communicating with extremely large-scale array/surface: Unified modeling and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4039–4053, Jun. 2022.
- [4] A. Amiri, S. Rezaie, C. N. Manchón, and E. de Carvalho, "Distributed receiver processing for extra-large MIMO arrays: A message passing approach," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2654–2667, Apr. 2022.
- [5] M. Cui, L. Dai, Z. Wang, S. Zhou, and N. Ge, "Near-field rainbow: Wideband beam training for XL-MIMO," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3899–3912, Jun. 2023.
- [6] Y. Lu and L. Dai, "Near-field channel estimation in mixed LoS/NLoS environments for extremely large-scale MIMO systems," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3694–3707, Jun. 2023.
- [7] H. Elayan, O. Amin, B. Shihada, R. M. Shubair, and M.-S. Alouini, "Terahertz band: The last piece of rf spectrum puzzle for communication systems," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1–32, 2020.
- [8] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, Aug. 2014.
- [9] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surv. Tuts.*, vol. 17, no. 4, pp. 1941–1988, 4th Quart. 2015.
- [10] J. Sun, Y. Zhang, J. Xue, and Z. Xu, "Learning to search for MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7571–7584, Nov. 2020.
- [11] Y. Wei, M.-M. Zhao, M. Hong, M.-J. Zhao, and M. Lei, "Learned conjugate gradient descent network for massive MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 6336–6349, 2020.
- [12] Q. Zhou and X. Ma, "Element-based lattice reduction algorithms for large MIMO detection," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 274–286, Feb. 2013.
- [13] T. Takahashi, A. Töllli, S. Ibi, and S. Sampei, "Low-complexity large MIMO detection via layered belief propagation in beam domain," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 234–249, Jan. 2022.
- [14] Z. Zhang, H. Li, Y. Dong, X. Wang, and X. Dai, "Decentralized signal detection via expectation propagation algorithm for uplink massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11–233–11 240, Oct. 2020.
- [15] J. Goldberger and A. Leshem, "MIMO detection for high-order QAM based on a Gaussian tree approximation," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4973–4982, Aug. 2011.
- [16] D. Gao et al., "Signal detection in MIMO systems with hardware imperfections: Message passing on neural networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 820–834, Jan. 2024.
- [17] H. He, X. Yu, J. Zhang, S. Song, and K. B. Letaief, "Message passing meets graph neural networks: A new paradigm for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4709–4723, May 2024.
- [18] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surv. Tuts.*, vol. 21, no. 4, pp. 3109–3132, 4th Quart. 2019.
- [19] P. Som, T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Low-complexity detection in large-dimension MIMO-ISI channels using graphical models," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1497–1511, Dec. 2011.
- [20] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, Oct. 2014.
- [21] Amari, S. *Information Geometry and Its Applications*. Tokyo, Japan: Springer, 2016.
- [22] S. Ikeda, T. Tanaka, and S. Amari, "Stochastic reasoning, free energy, and information geometry," *Neural Comput.*, vol. 16, no. 9, pp. 1779–1810, Sep. 2004.
- [23] S. Amari and H. Nagaoka, *Methods of Information Geometry*. Amer. Math. Soc., 2000, vol. 191.
- [24] H. Wu, Y. Cheng, X. Chen, Z. Yang, K. Liu, and H. Wang, "Power spectrum information geometry-based radar target detection in heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–16, 2024.
- [25] X. Hua, Y. Ono, L. Peng, Y. Cheng, and H. Wang, "Target detection within nonhomogeneous clutter via total Bregman divergence-based matrix information geometry detectors," *IEEE Trans. Signal Process.*, vol. 69, pp. 4326–4340, 2021.
- [26] J. Y. Yang, A.-A. Lu, Y. Chen, X. Q. Gao, X.-G. Xia, and D. T. M. Slock, "Channel estimation for massive MIMO: An information geometry approach," *IEEE Trans. Signal Process.*, vol. 70, pp. 4820–4834, 2022.
- [27] D. Wang, F. Fu, H. Yu, W. Chu, Z. Wu, and W. Li, "Artificial-intelligence-based quantitative fault diagnosability analysis of spacecraft: An information geometry perspective," *IEEE Trans. Artif. Intell.*, vol. 4, no. 4, pp. 624–635, Aug. 2023.
- [28] J. Y. Yang, Y. Chen, X. Q. Gao, D. T. M. Slock, and X.-G. Xia, "Signal detection for ultra-massive MIMO: An information geometry approach," *IEEE Trans. Signal Process.*, vol. 72, pp. 824–838, 2024.
- [29] H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes*. Kappa Res.: Galway, Ireland, 2014.
- [30] A. C. Berry, "The accuracy of the gaussian approximation to the sum of independent variates," *Trans. Amer. Math. Soc.*, vol. 49, no. 1, pp. 122–136, 1941. [Online]. Available: <http://www.jstor.org/stable/1990053>
- [31] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuADriGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.
- [32] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *Proc. IEEE Inform. Theory Workshop (ITW)*, Cairo, Jan. 2010, pp. 1–5.
- [33] P. Yang, Q. Zou, and H. Yang, "Message passing based calculation of MI and MMSE matrix for massive MIMO systems with finite-alphabet inputs," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3824–3828, Dec. 2021.



Jiyuan Yang (Member, IEEE) received the B.E. and Ph.D. degrees from the Southeast University, Nanjing, China, in 2019 and 2024, respectively. Currently, he is employed with Huawei Technologies Company Ltd. His research interests cover information geometry, Bayesian inference, and massive MIMO communications.



Mingrui Fan received the B.E. and M.E. degrees in information science and engineering from the Southeast University, Nanjing, China, in 2022 and 2025, respectively. His research interests are mainly on information geometry and massive MIMO communications.

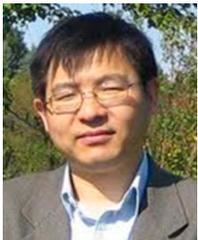


Yan Chen received the B.E. and M.E. degrees in information science and engineering from the Southeast University, Nanjing, China, in 2020 and 2023, respectively. He is currently employed with DJI. His research interests are mainly on information geometry and massive MIMO communications.



Xiqi Gao (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the Southeast University, Nanjing, China, in 1997. He joined the Department of Radio Engineering, Southeast University, in 1992. Since 2001, he has been a Professor of information systems and communications. From 1999 to 2000, he was a Visiting Scholar with Massachusetts Institute of Technology, Cambridge, and Boston University, Boston, MA. From 2007 to 2008, he visited the Darmstadt University of Technology, Darmstadt, Germany, as a Humboldt

Scholar. His current research interests include broadband multicarrier communications, massive MIMO wireless communications, satellite communications, optical wireless communications, information theory and signal processing for wireless communications. From 2007 to 2012, he served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. From 2009 to 2013, he served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. From 2015 to 2017, he served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. He received the Science and Technology Awards of the State Education Ministry of China in 1998, 2006 and 2009, the National Technological Invention Award of China in 2011, the Science and Technology Award of Jiangsu Province of China in 2014, and the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communications theory.



Xiang-Gen Xia (Fellow, IEEE) received the B.S. degree in mathematics from Nanjing Normal University, Nanjing, China, in 1983, the M.S. degree in mathematics from Nankai University, Tianjin, China, in 1986, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1992. He was a Senior/Research Staff Member with Hughes Research Laboratories, Malibu, California, from 1995 to 1996. In 1996, he joined the Department of Electrical and Computer Engineering, University of

Delaware, Newark, Delaware, where he is the Charles Black Evans Professor. His current research interests include space-time coding, MIMO and OFDM systems, digital signal processing, and SAR and ISAR imaging. He is the

author of the book “*Modulated Coding for Intersymbol Interference Channels*” (New York, Marcel Dekker, 2000). He received the National Science Foundation (NSF) Faculty Early Career Development (CAREER) Program Award in 1997, the Office of Naval Research (ONR) Young Investigator Award in 1998, and the Outstanding Overseas Young Investigator Award from the National Nature Science Foundation of China in 2001. He received the 2019 Information Theory Outstanding Overseas Chinese Scientist Award, The Information Theory Society of Chinese Institute of Electronics. He has served as an Associate Editor for numerous international journals including IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON MOBILE COMPUTING, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is a Technical Program Chair of the Signal Processing Symp., Globecom 2007 in Washington D.C. and the General Co-Chair of ICASSP 2005 in Philadelphia.



Dirk Slock (Fellow, IEEE) received the EE degree from Ghent University, Belgium, in 1982, and the M.S.EE, and M.S. degrees in statistics and the Ph.D. degree in EE from Stanford University, USA, in 1986, 1989, and 1989, respectively. In 1984, he was awarded a Fulbright Scholarship for Stanford University, USA. While at Stanford, he developed new fast recursive least-squares algorithms for adaptive filtering. From 1989 to 1991, he was a member of the Research Staff with the Philips Research Laboratory Belgium. In 1991, he joined EURECOM

where he is currently a Professor. At EURECOM, he teaches statistical signal processing (SSP) and signal processing techniques for wireless communications. He invented semi-blind channel estimation, the chip equalizer-correlator receiver used by 3G HSDPA mobile terminals, spatial multiplexing cyclic delay diversity (MIMO-CDD) now part of LTE, and his work led to the Single Antenna Interference Cancellation (SAIC) integrated in the GSM standard in 2006. Recent keywords are multicell multiuser (Massive) MIMO, imperfect CSIT, distributed resource allocation, variational and empirical Bayesian learning techniques, large random matrix analysis, audio source separation, location estimation, and exploitation. He graduated about 40 Ph.D. students, leading to an edited book and 500+ papers. In 1992, he received one Best Journal Paper Award from IEEE-SP and one from EURASIP. He is the co-author of two IEEE Globecom98, one IEEE SIU04, one IEEE SPAWC05, one IEEE WPNC'16, and one IEEE SPAWC'18 Best Student Paper Award, and a honorary mention (finalist in Best Student Paper Contest) at IEEE SSP05, IWAENC06, IEEE Asilomar06, and IEEE ICASSP'17. He was an Associate Editor for the IEEE-SP TRANSACTIONS from 1994 to 1996 and the IEEE SIGNAL PROCESSING LETTERS from 2009 to 2010. He was the General Chair of the IEEE-SP SPAWC06 and IWAENC'14 workshops, and EUSIPCO'15. He cofounded the start-ups SigTone in 2000 (music signal processing products) and Nestwave in 2014 (Ultra Low-Power Indoor and Outdoor Mobile Positioning). He is a fellow of EURASIP. In 2018, he received the URSI France Medal.