Low-Complexity Large MIMO Detection via Layered Belief Propagation in Beam Domain

Takumi Takahashi, Member, IEEE, Antti Tölli, Senior Member, IEEE, Shinsuke Ibi, Member, IEEE, and Seiichi Sampei, Fellow, IEEE

Abstract—In large multi-user multi-input multi-output systems, the computational cost and circuit scale of base stations (BSs) are effectively reduced using two-stage signal processing consisting of a slow varying outer beamformer (OBF) based on long-term channel statistics and group-specific multi-user detection for instantaneous channel variations. However, the dimensionality reduction of the group-specific beam-domain channel based on the OBF causes significant performance degradation in the subsequent spatial-filtering detection. To compensate for this drawback, this paper introduces a novel layered belief propagation (BP) detector with a concatenated structure of beam- and antenna-domain BP layers for post-stage OBF processing. The proposed detector is designed for improving the convergence of iterative detection by suppressing intra- and inter-group interference in stages. The layered structure provides the advantages of both beam and antenna domains while maintaining low signal-processing complexity. Numerical results show the validity of our proposed method in terms of the bit error rate performance in both the uncoded and coded cases and the computational complexity.

Index Terms—Large MIMO detection, belief propagation, dimensionality reduction, iterative signal detection, beam-domain signal processing, domain conversion.

I. INTRODUCTION

Large multi-user multi-input multi-output (MU-MIMO) systems, which are equipped with a large number of antennas on both the transmitter and receiver sides, have been regarded as one of the most promising technologies in the physical layer of wireless communication systems [1]–[5]. Increasing the dimensions of MIMO channels can increase data rates with the aid of spatial multiplexing gain and improve detection reliability because of spatial diversity gain, in addition to supporting a massive amount of simultaneous wireless communications. However, for large MU-MIMO systems, the precoder design in the downlink and MU detection (MUD) in the uplink are difficult and computationally expensive [2], [3]. Even if linear spatial filters, such as minimum mean square error (MMSE) filters, are used to reduce the computational cost, matrix operations based on the high-dimensional antenna-domain channel matrix are unavoidable, increasing the circuit scale of the transmitter–receiver on the base station (BS) side.

To address this issue, several approaches, including fully digital two-stage beamforming in the downlink precoder design, have been proposed [6]–[11]. A notable method using this approach is joint spatial division and multiplexing (JSDM) [6], where user equipment (UE) is grouped based on the similarity of transmit correlation matrices to design a slowly varying outer beamformer (OBF). In practice, because cellular UE devices tend to be distributed according to the terrain around the BS, they can be efficiently grouped by matching angular spread on the BS side. The OBF is constructed by selecting a set of beams oriented toward each UE group based on long-term channel statistics, and it effectively reduces the dimensions of equivalent channels in the beam (angular) domain [6], [7]. An inner beamformer (IBF) is applied for spatial multiplexing on the reduced beam-domain channel to manage both intra- and inter-group interference, significantly reducing the computational cost and transmitter circuit scale [9]–[11]. The OBF varies over long time scales compared to the IBF and, thus, requires less frequent updates.

The low-complexity strategy involving the two-stage processing with the OBF can also be considered in uplink scenarios. As shown in Fig. 1, the wireless channels between the BS and the UE device exhibit a small angular spread from the perspective of the BS, as a result of local scatterers around the UE device and the high placement of the BS antennas [7], [12]–[15]. Increasing the number of BS antennas...
leads to high spatial resolution in the beam (angular) domain. Because the received signal from each UE device is observed only in a subset of beams, large MIMO channels exhibit pseudo-sparsity (including values close to zero) in the beam domain owing to the attenuated sidelobes of the OBF. Recent studies have exploited the sparsity to propose low-complexity MUD schemes [16]–[19]. In [18], a dimensionality reduction is used to compress the channel state information (CSI) in the beam domain, with the goal of reducing memory usage. A linear detector using the compressed CSI was empirically shown to achieve improved performance, lower complexity, and reduced memory usage, in comparison with the typical antenna-domain counterparts. Beam selection algorithms and their hardware implementations were also presented in [17], [18]. As a more flexible strategy, a UE-specific OBF design was proposed in [19], where a contiguous block of beams can be selected for each UE device based on the mean square error (MSE) criterion. The resulting local linear MMSE filter achieved a detection capability comparable to that of the full antenna-domain MMSE filter, at a particularly low computational cost. However, these methods work effectively only when the number of UE devices $M$ is sufficiently small compared to the number of receive antennas $N$. As $M$ increases, the number of beams required to maintain the performance increases. An excessive dimensionality reduction aimed at computational reduction will result in information loss, and severe performance degradation is unavoidable owing to inter-group interference (overlapping angular spread), particularly when $M$ is of the same order as $N$. Additionally, $M \ll N$ must be satisfied for applying the conventional massive MIMO simplification based on the strong channel-hardening effect [2] and for the spatial filter to work well [2], [6].

To achieve reliable MUD in high-spatial-loading MU-MIMO scenarios while benefiting from dimensionality reduction, in the present study, we introduce a message passing (MP) algorithm based on belief propagation (BP) [20]–[32] in the post-stage of the OBF. The BP-based signal detector takes advantage of the law of large numbers to reduce the computational cost while maintaining the detection capability. The detection schemes can be roughly classified into matched filter (MF)-based BP and MMSE-based BP. The most common MF-based BP detector is approximate message passing (AMP) [24]–[26], which is systematically derived from a rigorous approximation of Gaussian BP (GaBP) [21]–[23] in the large-system limit, where the input and output dimensions, $M$ and $N$, respectively, are infinity for a given compression rate $\rho = N/M$. These algorithms separate the multiplexed signals using MF and consist of only scalar-wise operations; therefore, they can achieve a computational complexity of $O(MN)$ for each iteration. Furthermore, a rigorous proof of convergence to the Bayes-optimal solution in the large-system limit was presented in [25]. This proof requires the antenna-domain channel matrix entries to be independent and identically distributed (i.i.d.) random variables with zero mean. This assumption is not reasonable in practical MUD scenarios, owing to the physical limitations of the receiver and the spatial fading correlation between BS antenna elements. To address a broader class of practical applications that involve a general form of structured matrices, e.g., unitary-invariant measurement matrices, MMSE-based BP detectors based on the expectation propagation (EP) framework [33] were proposed in [28]–[30] and proven in [31], [32] to converge toward the Bayes-optimal solution in the large-system limit. The EP-based detector utilizes an Onsager correction term to decouple the self-feedback of messages across iterations, as is the case with the AMP; without the term, the functionality of the detector would be consistent with the iterative detector based on probabilistic data association (PDA) [34]. A rigorous comparative analysis of EP and PDA was presented in [32]. The potential connection between AMP and EP was investigated in [35]–[39]. However, these MMSE-based BP detectors require inverse matrix operations, which have a prohibitively high complexity, for detecting each transmit vector. Additionally, even if the inverse matrix operations are allowed on the reduced beam-domain channel, insufficient large-system conditions caused by the dimensionality reduction severely degrade the detection capability.

To resolve the lack of an appropriate MUD scheme as the post-stage of OBF for high-spatial-loading MU-MIMO systems, in this paper, we propose a novel design for a concatenated structure of beam- and antenna-domain BP detectors. First, strongly correlated intra-group interference is iteratively suppressed via local MMSE filters in the beam-domain, and reliable prior information is generated to be passed to the subsequent antenna-domain BP detector. The subsequent MF-based BP is performed in the full antenna-domain to compensate for the information loss due to dimensionality reduction and to promote convergence during iterative detection while leveraging the prior information. Each detector can be regarded as a layer with a different role, and such a multi-layer structure enables the efficient suppression of intra- and inter-group interference in stages. Furthermore, we clarify that the domain conversion from the beam-domain to the antenna-domain between the layers allows us to benefit from the advantages of beam- and antenna-domain processing simultaneously, leading to a dramatic reduction in computational cost while preserving the performance of the EP detector in the full antenna domain. The contributions of this paper are summarized as follows:

1) A novel low-complexity MUD scheme employing concatenated beam- and antenna-domain layered BP based on a given OBF is proposed for high-spatial-loading MU-MIMO systems.

2) By introducing beam-antenna domain conversion in the concatenated structure, the convergence of iterative detection is improved.

3) The proposed method is demonstrated to be feasible with a remarkably low computational cost while maintaining

---

1The present paper extends our previous conference papers [40], [41] by presenting a detailed process flow of a layered BP algorithm with domain conversion, extending the layered BP to higher-order modulation schemes, and then providing a detailed simulation analysis of computational complexity and performance in uncoded and coded cases.
the bit error rate (BER) performance of the (considerably complex) EP detector in the full antenna-domain channels.

We emphasize that, to our best knowledge, a low-complexity MUD based on the OBF for high-spatial-loading MU-MIMO scenarios, which is the key contribution of this paper, has not yet been proposed.

The remainder of this paper is organized as follows. As a preliminary step, Section II presents a system model. Section III presents the appropriate OBF design using the knowledge of channel covariance matrices to reduce the computational cost of the subsequent detector and the circuit scale on the BS side. The equivalent signal model in the beam domain is also defined. Section IV proposes the novel layered BP detector with a concatenated structure for use in general quadrature amplitude modulation (QAM) systems. The beam-antenna domain conversion method is explained throughout this paper, vectors and matrices are denoted by bold-face letters, respectively. Further-

A. Mathematical Notations

Throughout this paper, vectors and matrices are denoted by lower- and upper-case bold-face letters, respectively. Furthermore, $(\cdot)^T$, $(\cdot)^{\dagger}$ and $(\cdot)^H$ denote the conjugate, transpose, and conjugate transpose operators, respectively. $CN(a, b)$ indicates a complex Gaussian process with mean $a$ and variance $b$. $P_{ab}[a|b]$ and $P_{ab}[a|b]$ respectively represent the conditional probability mass function (PMF) and probability density function (PDF) of a realization $a$ of a random variable $a$, given the occurrence of a realization $b$ of a random variable $b$. $E_a\{\cdot\}$ is the expected value of $a$. $E_{a|b=a}\{\cdot\}$ denotes the conditional expectation of $a$, given the occurrence of $b$ of $b$. $R[\cdot]$ and $\Im[\cdot]$ are the real and imaginary parts of a complex value, respectively. $I_a$ represents an $a \times a$ square identity matrix. $\text{diag}(a)$ denotes a diagonal matrix with nonzero elements equal to $a$. $\text{det}[\cdot]$ is the determinant of a matrix. $\text{proj}(\cdot)$ denotes a projection function from an $a$-dimensional vector to a $b$-dimensional vector.

B. Antenna-Domain Signal Model

Consider an uplink MU-MIMO system in which the BS has $N$ receive (RX) antennas in a uniform linear array (ULA) pattern and $M \leq N$ UE devices are equipped with a single transmit (TX) antenna. Let $M = \{1, \ldots, m, \ldots, M\}$ be the set of UE indices and $N = \{1, \ldots, n, \ldots, N\}$ be the set of RX antenna indices. The UE devices tend to be distributed according to the terrain and large structures (e.g., buildings and roads) and therefore are not uniformly distributed around the BS, allowing UE grouping based on the angle of arrival (AoA) of the RX signal in the beam (angular) domain [6], [7]. Here, the UE devices can be grouped into $G$ groups by matching the angular spread, with the set of all groups $G = \{1, \ldots, g, \ldots, G\}$. As $M$ and $G$ increase, the angular spreads between adjacent groups inevitably overlap, leading to severe inter-group interference. Let $U_g$ be the set of all UE devices belonging to group $g \in G$, where $|U_g| = U_g$ is the number of elements in $U_g$. Thus, $M = \bigcup_{g \in G} U_g$, where $|M| = M = \sum_{g \in G} U_g$.

The $m$-th UE conveys a modulated symbol $x_m$, which represents one of $Q$ constellation points $X = \{\chi_1, \ldots, \chi_q, \ldots, \chi_Q\}$. The coefficient $w_x$, which normalizes the average TX power density to $E_s$, is given by $\sqrt{3E_s/(2(Q-1))}$ for Q-QAM [2]. Let $\mathbf{x} = [x_1, \ldots, x_m, \ldots, x_M]^T \in \mathbb{C}^{M \times 1}$ and $\mathbf{y} = [y_1, \ldots, y_m, \ldots, y_N]^T \in \mathbb{C}^{N \times 1}$ denote the TX and RX symbol vectors, respectively. The antenna-domain RX vector $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{H} \mathbf{x} + \mathbf{z},$$

where $\mathbf{H} \in \mathbb{C}^{N \times M}$ is a MIMO channel matrix in the spatial domain. The channel $\mathbf{H}$ is assumed to be static within each coherence interval of length $T$. The entries of the complex additive white Gaussian noise (AWGN) vector $\mathbf{z} = [z_1, \ldots, z_n, \ldots, z_N]^T \in \mathbb{C}^{N \times 1}$ are i.i.d. complex Gaussian variables with zero mean and $N_0$ variance, where $N_0$ is the noise power spectral density and the covariance matrix of $\mathbf{z}$ is given by $\mathbf{E}_z\{zz^H\} = N_0 \mathbf{I}_N$. The $m$-th entry of $\mathbf{y}$ is expressed as

$$y_m = \overline{h}_m \mathbf{x} + z_m,$$

where $\overline{h}_m = [h_{m,1}, \ldots, h_{m,m}, \ldots, h_{m,M}]$ denotes the $m$-th row vector of $\mathbf{H}$.

Without loss of generality, we use the geometric one-ring model [12]. Assuming a diffuse 2D field of isotropic scatterers around UE, the element in the $i$-th row and $j$-th column of the RX spatial correlation matrix for the $m$-th UE, $\Theta_m \in \mathbb{C}^{N \times N}$, is given by

$$[\Theta_m]_{i,j} = \frac{\mu_m}{\Delta \delta_m} \int_{\delta_m}^{\delta_m} \exp[j\pi(i-j)\cos(\delta)] d\delta.$$  

Here, $\delta_m = (\delta_m^{\text{max}} + \delta_m^{\text{min}})/2$ is the location of the $m$-th UE in the azimuthal direction, and waves arrive from the $m$-th UE with an angular spread $\Delta \delta_m = \delta_m^{\text{max}} - \delta_m^{\text{min}}$. The antenna element spacing is fixed to half the wavelength, and $\mu_m$ represents the path loss. The $m$-th column vector of $\mathbf{H}$ is given by $\mathbf{h}_m = \Theta_m^{1/2} \mathbf{e}_m$, where $\mathbf{e}_m \in \mathbb{C}^{N \times 1}$ represents small-scale fading and has i.i.d. complex Gaussian entries with zero mean and unit-variance [13].

III. OUTER BEAMFORMER DESIGN USING LONG-TERM CHANNEL STATISTICS

To avoid the high-dimensional matrix operations required for large-scale MUD, we adopt a two-stage receiver design consisting of the OB and BP-based MUD. The OB determines the size of the spatial filter used in the subsequent MUD, and therefore, plays a significant role in determining the overall computational cost. Let $B_g$ denote the number of beams oriented toward each UE group $g \in G$, where $\sum_{g \in G} B_g \leq N$. Here, the OBF for $g$, $\mathbf{B}_g$, contains all beams corresponding to the UE devices in $g$. Assuming that UE devices are grouped based on their long-term channel statistics (e.g., channel covariance matrices), we can design group-specific OBF matrices and use them over multiple channel realizations (over multiple coherence times).
A. Beam Selection

There are two well-known heuristic methods to find OBFs: eigen beam selection and discrete Fourier transform (DFT) beam selection [6], [7], [9]–[11]. In eigen beam selection, the eigenvectors of the channel covariance matrix are used to form the OBF matrix via eigenvalue decomposition (EVD) [6], [7]. By choosing $B_g$ eigenvectors corresponding to the $B_g$ largest eigenvalues, we obtain the OBF matrix $B_g$. In contrast, DFT beam selection uses the column vectors of the DFT matrix $D = [d_1, \ldots, d_N] \in \mathbb{C}^{N \times N}$ to form the OBF matrix [9]–[11]. Consequently, the problem reduces to finding a subset of column vectors from the unitary DFT matrix. In [11], DFT beam selection was numerically shown to be better than the eigen scheme. Without loss of generality, we use the DFT scheme to form the OBF matrix in this study.

Here, the channel covariance matrix between all the UE devices in group $g$ and the receiver on the BS side is given by

$$R_g = E_{h_i \in U^g} \left\{ \sum_{i \in U^g} h_i h_i^H \right\} = \sum_{i \in U^g} \Theta_i. \quad (4)$$

To maximize the energy collected by the DFT beams, we select $B_g$ DFT column vectors for each group $g \in G$ based on Algorithm 1. The maximization metric $d_n^H R_g d_n$ is the average received energy from the UE devices in group $g$ that can be received by the $n$-th DFT beam $d_n$. Upon finding $B_g \subseteq N$, the resulting OBF matrix that covers the strongest signal paths of each group is given by $B_g = \{d_{g(1)}, \ldots, d_{g(B_g)}\} \in \mathbb{C}^{N \times B_g}$, where $B_g B_g^H = I_{B_g}$.

B. UE Selection

Before shifting our focus to MUD, we represent a signal model in the beam domain with the OBF matrix. For each UE group $g \in G$, the beam-domain RX vector $r_g \in \mathbb{C}^{B_g \times 1}$ can be represented as

$$r_g = B_g^H y = \Xi_g x + v_g, \quad (5)$$

where $\Xi_g = B_g^H H \in \mathbb{C}^{B_g \times M}$ and $v_g = B_g z \in \mathbb{C}^{B_g \times 1}$ are the group-specific beam-domain channel and noise, respectively. Considering the computational cost for generating spatial filters, however, the number of row dimensions $M$ remains undesirably high.

For further dimensionality reduction of the group-specific beam-domain channel matrix, we ignore the negligibly small inter-group interference from UE devices that are geographically remote with respect to group $g$ owing to the attenuated sidelobes of the OBF. Similarly to DFT beam selection, in this case, we select $S_g \subseteq M \in G$ UE devices for each group $g \in G$ based on Algorithm 2. The maximization metric $\sum_{i \in S_g} d_i^H \Theta_m d_i$ is the average received energy from the $m$-th UE device that can be received by the OBF $B_g$ toward group $g$. The resulting $S_g$ includes the UE devices in group $g$ and the UE devices that strongly interfere with respect to them.

IV. Layered Belief Propagation

As described above, the OBF-based beam-domain channel can significantly reduce the dimensions of MIMO channels. However, when using only linear structures of spatial filters, insufficient spatial degrees of freedom to suppress the inter-group interference causes a high error floor in the BER performance. Above all, the information loss due to dimensionality reduction cannot be compensated on the reduced beam-domain channels after the OBF. To address these issues, we introduce a concatenated structure of beam- and antenna-domain BP detectors. As is well known, a BP detector is one of the nonlinear iterative receivers that provide better performance than linear spatial filters. The most vital point of our proposal is that the MMSE-based BP is performed in the reduced beam-domain channel; and then, using the outputs as prior information, the MF-based BP is performed in the full antenna-domain channel without dimensionality reduction. The MMSE-based BP mainly suppresses the correlated inter-group interference and generates reliable prior information for the subsequent MF-based BP, while taking advantage of the computational reduction resulting from the dimensionality reduction. The subsequent MF-based BP compensates the information loss due to dimensionality reduction and promotes the convergence of iterative detection by leveraging knowledge from the RX vector, CSI, and prior information passed from the MMSE-based BP. The name of the proposed method, layered BP, signifies its multi-stage detection process in which each BP detector has a distinct information passed from the MMSE-based BP. The name of the proposed method, layered BP, signifies its multi-stage detection process in which each BP detector has a distinct
why the extra processing for domain conversion performs the MF-based BP in antenna-domain channels instead of beam-domain channels.

A. Concatenated Beam- and Antenna-Domain Layered BP

A work flowchart of the proposed detection mechanism via layered BP is illustrated in Fig. 2(a). When both the beam- and antenna-domain CSI are available at the receiver, the domain conversion is not necessary. Fig. 2(b) schematically illustrates the belief propagation (consensus) process, which detects signals via the MMSE-based BP layer and MF-based BP layer. For each stage, we use superscripts (·)a and (·)b, respectively. The number of iterations in the MMSE-based BP is \( K^a \) and the number of iterations in the MF-based BP is \( K^b \). The corresponding iteration indices are denoted by \( k^a \in \{1, \ldots, K^a\} \) and \( k^b \in \{1, \ldots, K^b\} \), respectively. The MMSE-based BP first suppresses correlated intra-group interference among the UE devices in the subset \( S_g \) by using beam-domain local MMSE filters \( w_{g,m} \). The subsequent MF-based BP promotes the convergence of iterative detection by suppressing the residual interference in full antenna-domain channels \( H \) over all UE devices.

1) MMSE-Based BP Layer: At the first iteration \( (k^a = 1) \), no soft interference cancellation (soft IC) is conducted, because prior beliefs are unavailable. In the detection of an arbitrary TX symbol \( x_m \), \( \forall m \in S_g \) while ignoring inter-group interference from \( m \notin S_g \), the RX vector \( r_g \) of (5) can be approximated as

\[
    r_g = \tilde{r}_{g,m} + \sum_{i \in S_g \setminus \{m\}} \xi_{g,i}x_i + v_g, \tag{6}
\]

where \( \xi_{g,m} \) denotes the \( m \)-th column vector of \( \Xi_g \). In the large-system limit [2], the PDF of (6) approaches a multivariate complex Gaussian distribution because of the central limit theorem (CLT). This behavior is called vector Gaussian approximation (VGA). Accordingly, \( \tilde{r}_{g,m} \) can be approximated by an equivalent signal model as

\[
    \tilde{r}_{g,m} = \xi_{g,m}x_m + \psi_{g,m}, \tag{7}
\]

where \( \psi_{g,m} \) is the effective Gaussian noise. The covariance matrix \( \Psi_{g,m} \) is derived from the expectation of the random variables \( v_g \) and \( \{x_i, \forall i \in S_g \setminus \{m\}\} \) as

\[
    \Psi_{g,m} = \mathbb{E}_{v_g,\{x_i, \forall i \in S_g \setminus \{m\}\}} \left\{ \psi_{g,m} \psi_{g,m}^H \right\} = \Psi_g - E_s \xi_{g,m} \xi_{g,m}^H, \tag{8}
\]

with

\[
    \Psi_g = E_s \sum_{i \in S_g} \xi_{g,i} \xi_{g,i}^H + N_0 I_{B_g}. \tag{9}
\]

According to \( \xi_{g,m} \) and \( \Psi_{g,m} \), under VGA, the PDF of \( \tilde{r}_{g,m} \) based on Mahalanobis’ distance is expressed as

\[
    p_{\tilde{r}_{g,m}}(\tilde{r}_{g,m} | \chi_q) = \frac{1}{\pi^B \det \left[ \Psi_{g,m} \right]} \exp \left[ -\left( \tilde{r}_{g,m} - \xi_{g,m} \chi_q \right)^H \Psi_{g,m}^{-1} \left( \tilde{r}_{g,m} - \xi_{g,m} \chi_q \right) \right] \propto \exp \left[ 2R \left\{ \tilde{r}_{g,m}^H \Psi_{g,m}^{-1} \xi_{g,m} \chi_q - |\chi_q| 2^H \Psi_{g,m}^{-1} \xi_{g,m} \right\} \right]. \tag{10}
\]

By applying the matrix-inversion lemma to \( \Psi_{g,m}^{-1} \) in (10) to avoid matrix-inversion computations for every TX symbol detection, we obtain group-specific local MMSE filters as

\[
    w_{g,m} = \Psi_{g,m}^{-1} \xi_{g,m} = \Psi_g^{-1} \xi_{g,m} - \left[ \frac{1}{E_s} + \phi_{g,m} \right]^{-1} \phi_{g,m}, \tag{11}
\]
The conditional expectation of $\alpha$ projected to the $M$ expectation of (18).

Finally, given the prior belief vector $\alpha_{g,m}$ is likelihood information for $x_m = \chi_q$ given $\tilde{r}_{g,m}$.

In the same manner, the likelihood information $\{\alpha_{g,m}, \forall m \in S_g\}$ is computed as beliefs for each $g \in G$. To facilitate belief combining among the UE groups, $\alpha_{g,m}$ is projected to the $M$-dimensional vector $\alpha_g$ as

$$\alpha_g = [\hat{\alpha}_{g,1}, \ldots, \hat{\alpha}_{g,m}, \ldots, \hat{\alpha}_{g,M}]^T = \text{proj}_{S_{g} \to M}(\alpha_{g,m}, \forall m \in S_g),$$

where each element in $\alpha_g$ is given by

$$\hat{\alpha}_{g,m} = \begin{cases} \alpha_{g,m} & \text{if } m \in S_g \\ 0 & \text{otherwise}. \end{cases}$$

Assuming that the effective noise terms in $\alpha_1, \ldots, \alpha_G$ between the UE groups are not correlated to each other under VGA, the PDF of joint belief $p_{m,i}$ is given by

$$p_{m,i}(x_m | x_q) = \prod_{g \in G} p_{i_{g,m} | x_m}(\tilde{r}_{g,m} | x_q) \propto \exp \left[ \sum_{g \in G} \hat{\alpha}_{g,m}(x_q) \right].$$

Finally, given the prior belief vector $p_a = [p_{a_1}, \ldots, p_{a_M}]^T$, from Bayes’ rule, the soft symbol vector $\hat{x} = [\hat{x}_1, \ldots, \hat{x}_m, \ldots, \hat{x}_M]^T$ can be obtained from the conditional expectation of $\hat{x}$ as follows: [1]

$$\hat{x} = \sum_{\bar{x} \in \mathcal{X}_M \times 1} \bar{x} P_{\alpha | \bar{x}}(P_a | \bar{x}) = \sum_{\bar{x} \in \mathcal{X}_M \times 1} \bar{x} \frac{p_{\bar{x} | \alpha}(P_a | \bar{x})}{\sum_{\bar{x}' \in \mathcal{X}_M \times 1} p_{\bar{x}' | \alpha}(P_a | \bar{x}')},$$

where the occurrence of $\bar{x}$ is assumed to be equiprobable ($P(\bar{x}|\alpha) = \frac{1}{\mathcal{X}_M \times 1}$, $\forall \bar{x} \in \mathcal{X}_M \times 1$). Assuming that the effective noise terms in $p_{a_1}, \ldots, p_{a_M}$ are not correlated to each other under VGA in (7), the entries of (18) are readily given by the following symbol-wise expectation as

$$\hat{x}_m = \sum_{x_i \in \mathcal{X}} \chi_q \frac{p_{m,i} | x_m}(p_{a,m} | x_q)}{\sum_{x'_i \in \mathcal{X}} p_{i_{g,m} | x'_m}(p_{a,m} | x'_q)} = \sum_{x_i \in \mathcal{X}} \chi_q \frac{\exp \left[ \sum_{g \in G} \hat{\alpha}_{g,m}(x_q') \right]}{\sum_{x'_i \in \mathcal{X}} \chi'_q \exp \left[ \sum_{g \in G} \hat{\alpha}_{g,m}(x'_q') \right]}.$$
channels [2], the maximum ratio combining (MRC) of $s_{n,m}$ yields the following posterior belief:

$$o_m = \sum_{i \in \mathcal{N}} s_{i,m}. \quad (23)$$

Nevertheless, if $o_m$ were utilized as the prior belief, it would be subject to the ill-convergence behavior of iterative detection due to the correlation between $y_n$ and $s_{n,m}$ included in $o_m$, i.e., self-noise regression due to loopy propagation [21]. Therefore, the prior belief is typically provided by an extrinsic belief $l_{n,m}^b$, which is given by

$$l_{n,m}^b = o_m - s_{n,m} = \sum_{i \in \mathcal{N} \setminus \{n\}} s_{i,m}. \quad (24)$$

Here, a prior belief vector $l_n^b = [l_{n,1}^b, \ldots, l_{n,M}^b]^T$ is constructed by using (24).

Finally, the soft symbol vector $x_n$ can be obtained from the conditional expectation of $x$, given the prior belief vector $l_n^b$, by simply replacing $l$ with $l_n^b$ in (18). Thus, the PDF $p_{\tilde{x}|x} \left( \tilde{x}^b | x \right)$ is essential to compute the conditional expectation $\tilde{x}_n = [\tilde{x}_{n,1}, \ldots, \tilde{x}_{n,m}, \ldots, \tilde{x}_{n,M}]^T$. Assuming the strong channel-hardening effects, the symbol-wise conditional expectation is obtained by

$$\tilde{x}_{n,m} = \sum_{\chi_q \in \mathcal{X}} \chi_q \sum_{\chi_q' \in \mathcal{X}} p_{\tilde{x}|x}^{l_{n,m}}(l_{n,m} | \chi_q) \frac{\exp[\beta_{g,m}(\chi_q)]}{\sum_{\chi_q' \in \mathcal{X}} \exp[\beta_{g,m}(\chi_q')]},$$

where the likelihood information is given by

$$\beta_{n,m}(\chi_q) = \frac{1}{N_0} \left[ 2 \Re \left\{ \frac{l_{n,m}^* \chi_q}{\beta_{n,m}(\chi_q)} \right\} - |\chi_q|^2 \sum_{i \in \mathcal{N} \setminus \{n\}} |h_{i,m}|^2 \right]. \quad (26)$$

For the sake of readability, we have moved the detailed derivations of equations (25) and (26) to Appendix A.

However, the assumption of a strong channel-hardening effect is not reasonable in practical MUD scenarios, owing to the physical limitation of the receiver and the spatial fading correlation. To address this issue, an adaptively scaled belief (ASB) method was proposed in [22], which is equivalent to using the following symbol-wise expectation:

$$\tilde{x}_{n,m} = \sum_{\chi_q \in \mathcal{X}} \chi_q \sum_{\chi_q' \in \mathcal{X}} p_{\tilde{x}|x}^{l_{n,m}}(l_{n,m} | \chi_q) \frac{\exp[\sigma_{\omega_{n,m}} \cdot \beta_{n,m}(\chi_q)]}{\sum_{\chi_q' \in \mathcal{X}} \exp[\sigma_{\omega_{n,m}} \cdot \beta_{n,m}(\chi_q')]} \cdot \quad (27)$$

instead of (25), where the effective gain for $x_m$ is given by

$$\omega_{n,m} = \frac{2\sigma_0^2}{N_0} \sum_{i \in \mathcal{N} \setminus \{n\}} |h_{i,m}|^2. \quad (28)$$

According to [22], the predetermined parameter $\sigma$ at the $k$-th iteration in (27) is given by the following monotonically increasing function:

$$\sigma = f_\sigma(k^b) = a_0 + a_1 \cdot \left( \frac{k^b}{K^b} \right), \quad (29)$$

where $\sigma$ linearly increases from $a_0$ to $a_1$ in the iteration for avoiding rapid convergence to local minima. We utilize (27) for generating soft symbol replicas in the MF-based BP layer. When the number of iterations reaches the predetermined value $K^b$, $x_m$ is hard-detected as

$$\tilde{x}_m = \arg \max_{\chi_q \in \mathcal{X}} \left( 2 \Re \left\{ \frac{o_m \chi_q}{\beta_{n,m}(\chi_q)} \right\} - |\chi_q|^2 \sum_{i \in \mathcal{N} \setminus \{n\}} |h_{i,m}|^2 \right). \quad (30)$$

3) Algorithmic Structure: Algorithm 3 presents the pseudo-code of the concatenated beam- and antenna-domain layered BP detector. In the presence of spatial fading correlation among RX antennas, $s_{1,m}, \ldots, s_{N,M}$ are correlated; therefore, the belief combining in (24) cannot successfully suppress the interference and enhance the effective noise of $l_{n,m}^b$. To suppress these negative effects, the node selection (NS) method [43], [44] is introduced in the MF-based BP layer, in which the indices in $\mathcal{N}$ are divided into several subsets so that the corresponding beliefs in the same subset are as uncorrelated as possible based on the RX antenna pattern. The beliefs are sequentially updated for each set. In Algorithm 3, the RX antenna indices $\mathcal{N}$ are divided into $V$ subsets $\{\mathcal{V}_i, \ i = 1, \ldots, V\}$, where $|\mathcal{V}_i| = N/V$. Under the ULA pattern, the indices should be evenly divided as $\mathcal{V}_i = \{V \cdot (j - 1) + i \mid j = 1, \ldots, N/V\}$ to maximize the distance between two arbitrary RX antennas in each set. With the NS method, the number of iterations is $K^b \cdot V$, but the number of updated beliefs in each iteration is $|\mathcal{V}_i| = N/V$. Therefore, the NS method does not change the overall computational cost.

B. LLR Calculation for Channel-Coded Systems

In channel-coded systems, at the final iteration of the MF-based BP ($k^b = K^b$), bit LLRs for input to the subsequent channel decoder must be calculated, instead of being obtained via hard decision. Under scalar Gaussian approximation (SGA) of the interference-plus-noise term in $\tilde{y}_{n,m}$ in (21), the conditional PDF of $\tilde{y}_{n,m}$ can be expressed as

$$p_{\tilde{y}|y} \left( \tilde{y}_{n,m} | y \right) \propto \exp \left[ -\frac{|y_{n,m} - \tilde{y}_{n,m} |^2}{\psi_{n,m}} \right], \quad (31)$$

where the likelihood information is given by [22]

$$\gamma_{n,m}(\chi_q) = \frac{1}{\psi_{n,m}} \left[ 2 \Re \left\{ \tilde{y}_{n,m} \tilde{h}_{i,m} \chi_q \right\} - |\chi_q|^2 |\tilde{h}_{i,m}|^2 \right], \quad (32)$$

with

$$\psi_{n,m} = \sum_{i \in \mathcal{N} \setminus \{n\}} |h_{i,m}|^2 \left( \tilde{e}_{n,m} - \tilde{x}_{n,m} \right) + N_0, \quad (33)$$

$$\tilde{e}_{n,m} = \sum_{\chi_q \in \mathcal{X}} |\chi_q|^2 \frac{\exp[\sigma_{\omega_{n,m}} \cdot \beta_{n,m}(\chi_q)]}{\sum_{\chi_q' \in \mathcal{X}} \exp[\sigma_{\omega_{n,m}} \cdot \beta_{n,m}(\chi_q')]} \cdot \quad (34)$$

Note that $\psi_{n,m}$ is the variance of effective Gaussian noise in $\tilde{y}_{n,m}$ and $\tilde{e}_{n,m}$ is the conditional expectation of $|x_m|^2$. When $x_m$ consists of $N_o$ code bits, $c_1, \ldots, c_{N_o}$, the bit-wise
Algorithm 3 Concatenated Beam- and Antenna-Domain Layered BP

Input: \( y, H, r, \mathcal{X}, \{ \mathcal{B}_g, \mathcal{S}_y, \forall g \in \mathcal{G} \}, \{ \mathcal{V}_i \}^{V}_{i=1}, \mathcal{K}^a, \mathcal{K}^b \)
Output: \( \hat{x} = [\hat{x}_1, \ldots, \hat{x}_m, \ldots, \hat{x}_M]^T \)

/*Initialization*/
1. \( \hat{x} = [\hat{x}_1, \ldots, \hat{x}_m, \ldots, \hat{x}_M]^T = 0 \)
2. \( \forall g \in \mathcal{G}, \forall \chi_q \in \mathcal{X}: \alpha_g(\chi_q) = [\bar{\alpha}_{g,1}(\chi_q), \ldots, \bar{\alpha}_{g,m}(\chi_q), \ldots, \bar{\alpha}_{g,M}(\chi_q)]^T = 0 \)

/*Beam-domain MMSE-based BP layer*/
3. \( \forall g \in \mathcal{G}, \forall \chi_q \in \mathcal{X}: \Psi_g = E_s \sum_{i \in \mathcal{S}_g} \xi_{g,i} \xi_{g,i}^H + N_0 I_{B_g} \)
4. \( \forall g \in \mathcal{G}, \forall m \in \mathcal{S}_g: \phi_{g,m} = \bar{\xi}_{g,m}^H \Psi_g^{-1} \xi_{g,m} \)
5. for \( k^a = 1 \) to \( K^a \) do
6. \( \forall g \in \mathcal{G}, \forall m \in \mathcal{S}_g: \bar{r}_{g,m} = r_g - \Xi \hat{x} + \xi_{g,m}^H \hat{x} \)
7. \( \forall g \in \mathcal{G}, \forall m \in \mathcal{S}_g: w_{g,m} = \frac{\phi_{g,m}}{\bar{r}_{g,m}^H \phi_{g,m}} X_{g,m} \)
8. \( \forall g \in \mathcal{G}, \forall m \in \mathcal{S}_g: \eta_{g,m} = \frac{1}{E_s} \phi_{g,m} \)
9. \( \forall g \in \mathcal{G}, \forall m \in \mathcal{S}_g, \forall \chi_q \in \mathcal{X}: \hat{\alpha}_{g,m}(\chi_q) = 2 \Re \left\{ \bar{r}_{g,m} w_{g,m} \chi_q - |\chi_q|^2 \eta_{g,m} \right\} \)
10. \( \forall m \in \mathcal{M}: \hat{x}_m = \sum_{\chi_q \in \mathcal{X}} \chi_q \exp \left[ \sum_{g \in \mathcal{G}} \hat{\alpha}_{g,m}(\chi_q) \right] \frac{\exp \left[ \sum_{g \in \mathcal{G}} \hat{\alpha}_{g,m}(\chi_q) \right]}{\sum_{\chi_q \in \mathcal{X}} \exp \left[ \sum_{g \in \mathcal{G}} \hat{\alpha}_{g,m}(\chi_q) \right]} \)
11. end for

/*Antenna-domain MF-based BP layer*/
12. \( \forall n \in \mathcal{N}: \hat{x}_n = [\hat{x}_{n,1}, \ldots, \hat{x}_{n,m}, \ldots, \hat{x}_{n,M}]^T = \hat{x} \)
13. \( \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall \chi_q \in \mathcal{X}: \beta_{n,m}(\chi_q) = \frac{2}{N_0} \Re \left\{ \sum_{i \in \mathcal{N} \setminus \{n\}} \bar{y}_{i,m}^* h_i,m \chi_q - \gamma_{n,m}(\chi_q) \right\} \)
14. \( \forall n \in \mathcal{N}, \forall \chi_q \in \mathcal{X}: \omega_n = \frac{2 \Re \left\{ \bar{y}_{n,m}^* h_n,m \chi_q \right\}}{|\chi_q|^2} \gamma_{n,m}(\chi_q) \)
15. for \( k^b = 1 \) to \( K^b \) do
16. if \( k^b = 1 \) then
17. \( \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall \chi_q \in \mathcal{X}: \gamma_{n,m}(\chi_q) = \frac{1}{N_0} \Re \left\{ \sum_{i \in \mathcal{N} \setminus \{n\}} \bar{y}_{i,m}^* h_i,m \chi_q \right\} - \gamma_{n,m}(\chi_q) \)
18. \( \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall \chi_q \in \mathcal{X}: \beta_{n,m}(\chi_q) = \frac{2}{N_0} \Re \left\{ \sum_{i \in \mathcal{N} \setminus \{n\}} \bar{y}_{i,m}^* h_i,m \chi_q \right\} \)
19. \( \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall \chi_q \in \mathcal{X}: \hat{x}_m = \sum_{\chi_q \in \mathcal{X}} \chi_q \exp \left[ \sum_{n \in \mathcal{N}} \beta_{n,m}(\chi_q) \right] \frac{\exp \left[ \sum_{n \in \mathcal{N}} \beta_{n,m}(\chi_q) \right]}{\sum_{\chi_q \in \mathcal{X}} \exp \left[ \sum_{n \in \mathcal{N}} \beta_{n,m}(\chi_q) \right]} \)
20. end if
21. for \( i = 1 \) to \( V \) do
22. \( \forall m \in \mathcal{M}, \forall n \in \mathcal{V}_i: \hat{y}_{n,m} = y_n - \bar{r}_{n,m} \hat{x}_n + h_{n,m} \hat{x}_m \)
23. \( \forall m \in \mathcal{M}, \forall n \in \mathcal{V}_i, \forall \chi_q \in \mathcal{X}: \beta_{n,m}(\chi_q) = \frac{2}{N_0} \Re \left\{ \sum_{i \in \mathcal{N} \setminus \{n\}} \bar{y}_{i,m}^* h_i,m \chi_q \right\} \)
24. \( \forall m \in \mathcal{M}, \forall n \in \mathcal{V}_i, \forall \chi_q \in \mathcal{X}: \gamma_{n,m}(\chi_q) = \frac{1}{N_0} \Re \left\{ \sum_{i \in \mathcal{N} \setminus \{n\}} \bar{y}_{i,m}^* h_i,m \chi_q \right\} - \gamma_{n,m}(\chi_q) \)
25. end for
26. end if
27. end for

/*Hard decision*/
28. \( \forall m \in \mathcal{M}, \forall n \in \mathcal{N}: \hat{x}_m = \arg \max_{\chi_q \in \mathcal{X}} \sum_{n \in \mathcal{N}} \left[ 2 \Re \left\{ \bar{y}_{n,m}^* h_{n,m} \chi_q \right\} - |\chi_q|^2 |h_{n,m}|^2 \right] \)

LLR corresponding to the \( n_i \)-th bit \( c_{n_i} \) is given by

\[
\lambda_{m,n_i} = \ln \left( \frac{\sum_{\chi_q \in \mathcal{X}} \exp \left[ \sum_{n \in \mathcal{N}} \gamma_{n,m}(\chi_q) \right]}{\sum_{\chi_q \in \mathcal{X}} \exp \left[ \sum_{n \in \mathcal{N}} \gamma_{n,m}(\chi_q) \right]} \right), \quad (35)
\]

where \( X[c_{n_i} = c] \) is a set consisting of all candidate symbols such that \( c_{n_i} = c \in \{0, 1\} \). For the numerical stability, the summations as in (35) are evaluated using the Jacobian logarithm [1].

C. Beam-Antenna Domain Conversion

Finally, in this subsection, we describe the necessity of domain conversion to perform the MF-based BP in the antenna domain. To properly suppress the residual interference due to the extrinsic MRC process of (24), the effective noise included in \( s_{1,m}, \ldots, s_{N,m} \) should be as uncorrelated as possible. As a result of averaging the impact of TX symbols, their correlation strength deeply depends on the instantaneous correlation among channel coefficients in the column dimension; this can be visualized using the intensity of off-diagonal elements in a Gram matrix: \( \mathbf{G}_H = \mathbf{H}^H \mathbf{H} \). That is, when performing the MF-based BP in the beam domain without domain conversion, instead of \( \mathbf{G}_H \), we need to consider the following Gram matrix:

\[
\mathbf{G}_\Xi = \mathbf{X}^H = \mathbf{D}^H \mathbf{H}^H \mathbf{D} = \mathbf{D}^H \mathbf{G}_H \mathbf{D}. \quad (36)
\]

The noise enhancement level induced in (24) becomes more severe as the intensity of the off-diagonal elements in the Gram matrix increases, resulting in the ill-convergence behavior of iterative detection due to error propagation.

Fig. 3 shows the intensity of the off-diagonal elements in \( \mathbf{G}_\Xi \) and \( \mathbf{G}_H \) normalized by \( M \), where the diagonal elements are removed. A sector antenna with a 120° opening is considered. The angular spread for each UE device is 15°. The UE devices are randomly dropped around the BS in an \( \mathcal{N} = 64, 48 \) MU-MIMO configuration.

(a) Intensity of the off-diagonal elements in \( \mathbf{G}_\Xi \) normalized by \( M \).
(b) Intensity of the off-diagonal elements in \( \mathbf{G}_H \) normalized by \( M \)

The noise enhancement level induced in (24) becomes more severe as the intensity of the off-diagonal elements in the Gram matrix increases, resulting in the ill-convergence behavior of iterative detection due to error propagation.

Fig. 3 shows the intensity of the off-diagonal elements in \( \mathbf{G}_\Xi \) and \( \mathbf{G}_H \) normalized by \( M \), respectively, where the diagonal elements are removed. As shown in Fig. 3(a), owing to the DFT beamformer \( \mathbf{D}^H \), the intensity of the off-diagonal
elements in $G_\Xi$ is partially enhanced. In particular, the adjacent beams in the azimuthal direction increase the instantaneous correlation of the beam-domain channel coefficients, leading to the extreme off-diagonal elements. Consequently, owing to the correlation among $s_{1,m}, \ldots, s_{N,m}$, the effective noise of $b_{m}^n$ is enhanced in (24), and the soft symbols tend to contain persistent fatal errors (i.e., incorrect hard decision symbols) via error propagation. Conversely, in Fig. 3(b), the intensity of the off-diagonal elements in $G_H$ is uniformly distributed and suppressed to values smaller than those in $G_\Xi$ in the beam domain. In this case, the effective noise is not substantially enhanced, and the approximation errors in (24) can be mitigated using the soft IC in the iterative process. Thus, with the MF-based BP in the antenna domain, the convergence property of the layered BP can be improved. Note that the ill-convergence behavior in the beam-domain MF-based BP does not occur in MMSE-based BP [32] because the impact of the DFT beamformer is completely canceled out at the MMSE filtering step owing to the fact that $B_g^H B = I_{B_g}$.

In practice, the beam-domain channel $\Xi$ is estimated directly in the beam domain; thus, an extra step is needed in order to obtain the antenna-domain channel via FFT as $H = D \Xi$ when the antenna-domain channel is not available. The incremental computational cost for this operation is only $O(MN \log_2 N)$ per channel realization.

V. NUMERICAL RESULTS
A. BER Performance
Simulations were conducted to validate the performance of the proposed layered BP detector for large-scale MUD based on the OBF. The average RX powers from all TX antennas are assumed identical based on open-loop TX power control, i.e., $\mu_m = 1$, $\forall m \in \mathcal{M}$ in (3). Further, a sector antenna with a 120° opening is considered. The UE devices are naturally partitioned into $G$ segments with $U_g = M/G$, $\forall g \in G$ UE devices randomly dropped in each segment. The inter-group interference becomes severer as $M$ and $G$ increase. The angular spread for each UE device is 15°. The beam-domain channel estimation as well as the time and frequency synchronization between TX and RX are assumed to be perfect.

Fig. 4 shows BER performance with the system parameters $(G, B_g, S_g) = (8, 8, 3 \cdot U_g)$, $\forall g \in \mathcal{G}$. To properly suppress the correlated interference from the adjacent groups, $S_g$ is set to $3 \cdot U_g$. The number of UE devices is set to $M = 32$, 40, 48, 56, and 64, whereas that of RX antennas $N$ is fixed at 64; this represents a high-spatial-loading MU-MIMO configuration. The modulation scheme is Gray-coded QPSK in 4(a)–(d) and Gray-coded 16QAM in 4(e) and (f). The channel code is not utilized. The performance of the following methods is compared:

- **MMSE**: Baseline linear MMSE receiver in the antenna domain.
- **EP**: State-of-the-art MMSE-based BP detector [29], [32], where the number of iterations is $K = 8$. The MMSE filter based on the full antenna-domain channel is applied in each iteration. To improve the detection capability in practical MUD scenarios, some large-system approximations are removed from [32]. If the drawback due to dimensionality reduction is successfully suppressed, the layered BP can approach the performance of EP detector.
- **GaBP**: MF-based BP detector [21], where the number of iterations is $K = 16$. The AMP [24] and generalized AMP (GAMP) [26] are derived from GaBP using the large-system approximation; therefore, GaBP performs better than AMP and GAMP in finite-size MIMO channels. The performance results that can be achieved without whitening filters are presented.
- **MMSE w/ OBF**: Group-specific linear MMSE receiver based on the OBF.
- **Layered BP (b)**: Beam-domain layered BP detector, which is presented to denote the validity of domain conversion; the beam-antenna domain conversion is not used and all the MUD processes are performed in the beam domain, where $(K^a, K^b) = (8, 8)$.
- **Layered BP (b→a)**: Concatenated beam- and antenna-domain layered BP detector presented in Algorithm 3, where $(K^a, K^b) = (8, 8)$.
- **MFB**: Matched filter bound (MBF), which shows the ideal performance (lower bound) obtained when the perfect prior information is given. A brief explanation of MFB for MUD is presented in Appendix B.

Note that the NS method is used in “GaBP” and the MF-based BP layer of “layered BP,” where the number of sets $V$ is set to 4. Several simulations were conducted to find the sub-optimal $a_0$ and $a_1$ for minimizing the required $E_s/N_0$ at BER = $10^{-4}$ in $\rho = 2.0$. Consequently, we set $(a_0, a_1)$ to $(2, 0)$ for “layered BP (b)” and $(1, 1)$ for “layered BP (b→a).”

The results in Fig. 4 demonstrate that “MMSE w/ OBF” performs significantly worse than “MMSE” owing to dimensionality reduction based on the OBF. Even for $M = 40$ in Fig. 4(a), “MMSE w/ OBF” cannot achieve BER = $10^{-4}$. “GaBP” cannot sufficiently suppress the negative effect of severe interference due to spatial fading correlation, resulting in a high error floor in the BER performances. In contrast, “layered BP (b)” can significantly improve the detection capability compared to “GaBP” with low computational cost, owing to the local MMSE filtering in the beam domain. However, there remains the non-negligible performance gap from “EP” and it becomes larger as $M$ and $Q$ increase. The most attractive feature is that the proposed “layered BP (b→a)” can compensate for the performance gap and approach “EP” owing to domain conversion, even in the Gray-coded 16QAM cases in Fig. 4(e) and (f). The total gain w.r.t. “MMSE” at BER = $10^{-4}$ is approximately 6.0 dB, 9.5 dB, and more than 10 dB in the Gray-coded QPSK cases in Fig. 4(a)–(c), and approximately 4.0 dB and 5.0 dB in the Gray-coded 16 QAM cases in 4(e) and (f). In Figs. 4(c), (d), and (f), the error floor occurs...
even in “layered BP (b→a),” owing to the high-spatial-loading; however, such a low error floor can be suppressed using the channel decoder, as described later. In Figs. 4(a) and (e), where the spatial-loading is relatively low, the EP-based detector and the proposed method can approach “MFB.”

Let us shift our focus to other MIMO configurations. Fig. 5 shows the $E_s/N_0$ required to guarantee a BER of less than $10^{-4}$ according to $M$ for $N = 64$. As $M$ increases, the required $E_s/N_0$ increases more rapidly in “MMSE w/ OBF” than in “MMSE” owing to the severe inter-group interference caused by dimensionality reduction in the former. The required $E_s/N_0$ in “MMSE w/ OBF” approaches that in “MFB” only when $M$ is far less than $N$ and the UE devices are spatially well separated. Furthermore, “GaBP” cannot support a high number of UE devices under correlated MIMO channel conditions because of the lack of a whitening mechanism. Owing to the high error floor, it can achieve a BER of $= 10^{-4}$ only when $M \leq 16$ in Gray-coded QPSK and $M \leq 8$ in Gray-coded 16QAM. In contrast, “layered BP (b)” can drastically increase the spatial multiplexing gain in the region of low $E_s/N_0$; however, in Gray-coded 16QAM cases, the required $E_s/N_0$ rapidly increases as $M$ increases. By contrast, owing to the domain conversion, “layered BP (b→a)” can be utilized up to approximately $M = 56$ in Fig. 5(a) and $M = 40$ in Fig. 5(b). Furthermore, in the region of $M \leq 32$, the proposed method can approach “MFB” with almost the same performance as “EP.”

Another point to note is the BER performances in the coded MIMO systems, which are shown in Fig. 6. As high-spatial-loading scenarios, $(N, M) = (64, 32), (64, 40), (64, 48),$ and $(64, 64)$ MIMO configurations are considered, where the system parameters are the same as in the uncoded
Fig. 5. Required $E_s/N_0$ for achieving a BER of less than $10^{-4}$ according to the number of UE devices $M$ for $N = 64$; other settings are the same as those in Fig. 4.

Fig. 6. BER performances in the coded case with $(G, B_g, S_g) = (8, 8, 3 \cdot U_g)$, $\forall g \in G$.

In the Gram matrix of channels. As described in Section IV-A, the DFT beamformer enhances the off-diagonal elements, and the convergence property of the beam-domain MF-based BP is severely degraded. Additionally, the consistency condition [1] of the output LLRs does not hold and the channel decoder is not capable of fully performing the error correction. In contrast, the most attractive feature is that “layered BP (b→a)” can approach “EP” in terms of performance owing to the domain conversion. From the viewpoint of generating reliable LLRs that satisfy the consistency condition, the domain conversion becomes more vital in coded MIMO systems. Specifically, the performance degradation from “EP” is only approximately 0.5 dB at BER $= 10^{-4}$ in Figs. 6(a)–(c). Even in
Fig. 7 shows the BER performances of the BP-based detectors at each iteration step. The $E_s/N_0$ is fixed at $-1.0$ dB in Gray-coded QPSK and 6.0 dB in Gray-coded 16QAM cases of Fig. 7(b) and (d). The MIMO configurations are $(N, M) = (64, 48)$ and $(64, 32)$, respectively, and the other settings are the same as in Fig. 4. In Figs. 7(a) and (b), “GaBP” converges to poor local solutions owing to the spatial fading correlation, while “EP” can converge to much better solutions within a small number of iterations ($i.e.$, about 3 or 4), owing to the iterative MMSE filtering. It is worth noting here that “layered BP (b→a)” does not converge to inferior local solutions, but to solutions comparable to “EP.”

In Figs. 7(c) and (d), the BER performances of “layered BP (b→a)” at each iteration step with different $(K^a, K^b)$ settings are presented. “$(K^a, K^b) = (16, 0)$” indicates the performance of only beam-domain MMSE-based BP. Although the convergence speed is very fast owing to the fixed local MMSE filters, the local solution to which the algorithm can converge is poor (far from the optimal solution) because of the information loss caused by the dimensionality reduction. Even if the local MMSE filters are updated at each iteration step, the BER performance deteriorates via iterative processing due to error propagation. By contrast, the convergence solution of “$(K^a, K^b) = (8, 8)$” is greatly improved compared to “$(K^a, K^b) = (0, 16)$”, which confirms the effectiveness of the multi-layer structure consisting of whitening processing based on beam-domain local MMSE filters and full-antenna domain MF-based BP. Furthermore, when $K^a$ is varied while $K^b$ is fixed at 8, it can be seen that the algorithm can converge to almost the same solutions even with lower $K^a$ settings. Even for the high-spatial-loading MIMO configurations as shown in Fig. 7, setting $K^a$ to at least 3 is sufficient to converge to achievable solutions.

B. Complexity Analysis

The computational cost of each detector was evaluated in terms of the number of real multiplication operations required to detect MIMO signals and calculate LLRs passed to the channel decoder. To evaluate the approximate number of
TABLE I

| Detector | For each channel realization | For each TX vector detection |
|----------|------------------------------|-----------------------------|
| MMSE    | $12MN^2 + 8N^3 + 6MN + 3M + 2QM$ | $4MN + 3QM$ |
| EP      | Same as MMSE                 | $(K - 1) \cdot (10MN^2 + 14MN + 8N^3 + 5QMN + 3M + (\sqrt{Q}M) / 2)$ |
| GaBP    | $12MN$                       | $(28K - 5) \cdot MN + 3QM + M$ |
| MMSE w/ OBF | $\sum_{g=1}^{G} [12S_gB_g^2 + 8B_gS_g + 6S_gB_g$ | $\sum_{g=1}^{G} [4S_gB_g + 3QS_g]$ |
| Layered BP (b) | $\sum_{g=1}^{G} [12S_gB_g^2 + 8B_gS_g + 6S_gB_g$ | $\{4K^* + 12K - 5\} \cdot MN + K^* \cdot [\sum_{g=1}^{G} [4B_gS_g + 3QS_g]$ |
| Layered BP (b→a) | $\sum_{x=1}^{M} [12S_xB_x^2 + 8B_xS_x + 6S_xB_x$ | $+2QS_x] + 8MN + MN \log_2(N)$ |
|          |                              | Same as Layered BP (b)      |

Fig. 8. The number of real multiplication operations as a function of the number of UE devices $M$ (from 8 to 64) with the number of RX antennas $N$ fixed at 64. The system parameters are $(G, B_g, S_g) = (8, 8, 3 \cdot U_g), \forall g \in G$. Based on the results in Fig. 7, the number of iterations $K$ is set to 3 for “EP” and is set to 6 for “GaBP.” The number of iterations for “layered BP” is set to $(K^*, K^*) = (3, 8)$. The modulation scheme is Gray-coded 16QAM, i.e., $Q = 16$.

real multiplication operations, we adopt the following basic assumptions: [45]

- A multiplication of a complex matrix $A_1 \in \mathbb{C}^{d_1 \times d_2}$ with a complex matrix $A_2 \in \mathbb{C}^{d_2 \times d_3}$ requires $d_1d_2d_3$ multiplications (of complex numbers).
- An inversion of a complex matrix $A_3 \in \mathbb{C}^{d_1 \times d_1}$ requires approximately $2d_1^3 - d_1^2$ multiplications and $d_1^2$ divisions (of complex numbers).
- A division of complex numbers requires 1 complex multiplication (division and multiplication are assumed to have the same complexity).
- A complex multiplication requires 4 real multiplications. With these assumptions, Table I summarizes the number of real multiplication operations of different detectors; they are plotted in Fig. 8 under some system conditions. Note that we can ignore the computational cost for constructing the OBF because it can be computed offline.

First, we focus on the operations required for each channel realization (per coherence time $T$). Among all the methods, “GaBP” achieves the lowest complexity in Fig. 8(a) owing to its scalar-wise detection without any matrix operations. On the other hand, “MMSE” and “EP” require MMSE filter construction including the computation of matrix-inversion and matrix-multiplications according to the full antenna-domain channel size $N \times M$, leading to $O(MN^2 + N^3)$ complexity. Conversely, the OBF can reduce the equivalent beam-domain channel size to $B_g \times S_g$. Thus, the complexity for the construction of the local MMSE filters $\{w_g, \forall g \in G\}$ in (11) can be reduced to $O(\sum_{g=1}^{G} [S_gB_g^2 + B_g^3])$. Even when we employ the beam-antenna domain conversion in “layered BP (b→a),” its $O(MN \log_2(N))$ complexity is not dominant for determining the computational cost as shown in Fig. 8(a). Additionally, the reduction of the spatial-filter size allows us to significantly reduce the circuit scale and power consumption at the receiver, which is suitable for practical implementation.

Let us shift our focus to the operations required for the detection of each TX vector $\mathbf{x}$. In “MMSE” and “MMSE w/ OBF,” the only process required is linear filtering, and its computational cost is extremely low as shown in 8(b). Owing to the matrix operations according to the $N \times M$ antenna-domain channel matrix in each iteration, the computational cost of “EP” is prohibitively high, even if the number of iterations is relatively small. The typical “GaBP” requires $O(QMN)$ complexity in each iteration for computing the likelihood information for each constellation point. However, the low-computational implementation was proposed in [22], where the soft replica generation process is approximately functionalized so that the likelihood information need not be calculated for each constellation point, and the resulting complexity of “GaBP” can be reduced to $O(MN + QM)$ without any performance degradation. A similar methodology is applicable for the MF-based BP layer in “layered BP.” More
specifically, the lines 16–17 and 21–22 in Algorithm 3 can be replaced with the approximate soft replica generation function proposed in [22]. Consequently, the complexity of “layered BP” mainly involves the soft IC in (20) and (21) and the linear filtering process in each iteration, leading to $O(MN + QM)$.

As shown in Fig. 8(b), the computational cost of “layered BP” is suppressed below $10^6$ real multiplication operations even in the (64, 64) MIMO configuration.

As can be seen from Figs. 8(a) and (b), the dominant factor for determining the computational cost depends on the coherence time $T$. In the case of very small $T$, the proposed method can achieve lower computational cost compared to “MMSE,” but in other cases, “MMSE” achieves much lower computational cost. However, considering that the “MMSE” fails to detect signals reliably as $M$ increases and that “layered BP (b→a)” can achieve performance asymptotically approaching “EP,” we can confirm that the proposed method realizes an excellent trade-off between detection capability and computational cost. Furthermore, by heuristically optimizing the number of iterations ($K^a$, $K^b$) as shown in Fig. 7, the computational cost of “layered BP” can be further reduced according to spatial-loading, modulation level, channel model, and so on. From these results, we conclude that there are numerous scenarios in which the proposed method will be effective in terms of detection capability and computational cost.

VI. CONCLUSION

In this paper, we proposed a novel concatenated beam- and antenna-domain layered BP detector for realizing low-complexity, high-accuracy, large-scale MU-MIMO detection based on the OBF. The dimensionality reduction of beam-domain channels using the OBF, which is designed based on long-term channel statistics, can significantly reduce the computational cost for MUD and the circuit scale at the receiver. However, as the number of UE devices increases and the beam-domain channel size decreases, a high-precision MUD becomes more difficult to realize, owing to information loss and severe inter-group interference caused by dimensionality reduction. To address this problem, a concatenated structure of beam-domain MMSE-based BP and antenna-domain MF-based BP layers is introduced, which efficiently suppresses the severe interference in stages based on the pre-designed OBF. Simulation results explicitly demonstrate that the proposed method can dramatically reduce the computational cost to within a few percentage points of that of the EP-based detector while maintaining the BER performance. The most impressive feature of layered BP is the fact that the low-complexity BP detector is realized for correlated large-MUD scenarios while benefiting the dimensionality reduction.

As future research, we are considering applying our proposed layered BP detector to more flexible UE-specific OBF [19] for further computational reduction and improved convenience in beam-domain channel estimation. In such scenario, to manage severe inter-UE interference, the belief-combining operation in the MMSE-based BP layer must be improved by taking the impact of partially overlapping local MMSE filters into account.

APPENDIX A

DERIVATION OF THE SYMBOL-WISE CONDITIONAL EXPECTATION OF (25) FOR MF BELIEFS

The PDF $p_{b_n} | x \left( t_{b_n} | x \right)$ is essential to compute the exact conditional expectation $\hat{x}_n$. From (22) and (24), the prior belief vector $t_{b_n}$ can be expressed as

\[
\begin{align*}
\hat{t}_{b_n} &= [\hat{t}_{b_{n,1}}, \ldots, \hat{t}_{b_{n,m}}, \ldots, \hat{t}_{b_{n,M}}]^T = \Sigma_n \hat{x} - \Sigma_n^O \hat{x}_n + \zeta_n \\
&= \Sigma_n^D \hat{x} + \Sigma_n^O (\hat{x} - \hat{x}_n) + \zeta_n,
\end{align*}
\]

where

\[
\begin{align*}
\Sigma_n &= \sum_{i \in \mathcal{N} \setminus \{n\}} \tilde{h}_i^H h_i = H^H H - \tilde{h}_n^H h_n, \\
\zeta_n &= \sum_{i \in \mathcal{N} \setminus \{n\}} \tilde{T}_i z_i = H^H z - \tilde{h}_n z_n.
\end{align*}
\]

$\Sigma_n^D$ is a diagonal matrix with non-zero elements equal to $\Sigma_n$, and $\Sigma_n^O = \Sigma_n - \Sigma_n^D$. In the large system limit, the conditional PDF of $t_{b_n}$ is given by the following multivariate Gaussian PDF:

\[
p_{b_n} | x \left( t_{b_n} | x \right) = \frac{1}{\pi^N \text{det}(\Omega_n)} \times \exp \left[ - \left( t_{b_n} - \Sigma_n^D \hat{x} \right)^H \Omega_n^{-1} \left( t_{b_n} - \Sigma_n^D \hat{x} \right) \right],
\]

where the covariance matrix is given by

\[
\Omega_n = \Sigma_n^O \varphi_{\hat{x}_n, z} | x = \hat{x}_n \left\{ (\hat{x} - \hat{x}_n) (\hat{x} - \hat{x}_n)^H \right\} (\Sigma_n^O)^H + N_0 \Sigma_n.
\]

Under the ideal massive MIMO channel condition, owing to the strong channel-hardening effect [2], $\Sigma_n$ can be regarded as the diagonal matrix $\Sigma_n^D$. Additionally, assuming that the soft IC is successfully conducted with the highly reliable soft symbol vector ($\tilde{x}_n \approx x$) owing to the MMSE-based BP layer, the covariance matrix in (41) is approximated as

\[
\Omega_n \approx N_0 \Sigma_n^D.
\]

From (40) and (42), the following relationship is obtained as

\[
p_{b_n} | x \left( t_{b_n} | x \right) \propto \exp \left[ 2 \sqrt{R} \left\{ (t_{b_n}^H)^H \Omega_n^{-1} \Sigma_n^D \hat{x} - \hat{x} (\Sigma_n^D)^H \Omega_n^{-1} \Sigma_n^D \hat{x} \right\} - 2 \sqrt{R} \left\{ (t_{b_n}^H)^H \hat{x} - \hat{x} \right\} \Sigma_n^D \hat{x} \right] \\
\approx \exp \left[ \frac{1}{N_0} \left\{ 2 \sqrt{R} \left\{ (t_{b_n}^H)^H \hat{x} - \hat{x} \right\} \Sigma_n^D \hat{x} \right\} \right] \\
= \exp \left[ \sum_{i \in \mathcal{M}} \left\{ \frac{1}{N_0} \left\{ 2 \sqrt{R} \left\{ (t_{b_n}^H)^H \hat{x} - \hat{x} \right\} \Sigma_n^D \hat{x} \right\} \right\} \right]
\]

leading to

\[
p_{b_n} | x \left( t_{b_n} | x \right) = \prod_{i \in \mathcal{M}} p_{b_{n,i}} | x \left( t_{b_{n,i}} | x \right),
\]

\[
p_{b_{n,m}} | x \left( t_{b_{n,m}} | x \right) = \exp [\beta_{b_{n,m}} (\lambda_{x})],
\]
\[
\beta_{n,m}(\chi_q) = \frac{1}{N_0} \left( 2R \left\{ h_{n,m}^{b}(\chi_q) \right\} - |\chi_q|^2 \sum_{i \in X \setminus \{n\}} |h_{i,m}|^2 \right). \quad (46)
\]

Note that (46) is equal to (26). From (44), the conditional expectation of (xₙ,n,m) is given by
\[
\hat{x}_{n,m} = \sum_{\chi_q \in X} \chi_q p_{x_{n,m}}^{b} \left[ \chi_q | h_{n,m}^{b} \right] \left[ \chi_q | h_{n,m}^{0} \right] = \sum_{\chi_q \in X} \chi_q \sum_{\chi_q' \in X} p_{x_{n,m}}^{b} \left[ \chi_q' \right] p_{\psi_{n,m}}^{b} \left[ \chi_q' | h_{n,m}^{b} \right]. \quad (47)
\]

From (47) and Bayes’ rule, the symbol-wise expectation is given by
\[
\hat{x}_{n,m} = \sum_{\chi_q \in X} \chi_q p_{x_{n,m}}^{b} \left[ \chi_q | h_{n,m}^{b} \right] \left( \chi_q | h_{n,m}^{0} \right) = \sum_{\chi_q \in X} \chi_q \sum_{\chi_q' \in X} p_{x_{n,m}}^{b} \left[ \chi_q' \right] p_{\psi_{n,m}}^{b} \left[ \chi_q' | h_{n,m}^{b} \right]. \quad (48)
\]

From (45) and (48), (25) can be achieved.

**APPENDIX B**

**DERIVATION OF THE MATCHED FILTER BOUND (MFB) FOR MUD**

The MFB provides the ideal lower bound of BER for the BP-based detectors and also for the maximum likelihood (ML) detector. The MFB is achieved when the inter-UE interference is perfectly canceled; thus, the BP-based detectors achieve the MFB when the soft replica vector used for soft IC is equal to the transmitted vector x at the final iteration. From (21), using \( \hat{x}_n = x \), the output of perfect soft IC is given by
\[
\hat{y}_{n,m} = y_n - \sum_{i \in M \setminus \{n\}} h_{n,i}x_i = h_{n,m}x_n + z_n. \quad (49)
\]

By computing (22), (23), and (30) with \( \hat{y}_{n,m} \) of (49), the hard-decision symbol of MFB can be obtained. Similarly, by computing (32) and (35) with \( \hat{y}_{n,m} \) of (49) and \( \psi_{n,m} = N_0 \), the bit LLRs of MFB can be computed. For ease of explanation, we have used the formula of the MF-based BP layer; however, the same results can be obtained for other BP-based detectors.

**REFERENCES**

[1] L. Hanzo, T. H. Liew, B. L. Yeap, R. Y. S. Tee, and S. X. Ng, Turbo Coding, Turbo Equalisation and Space-Time Coding: EXIT-Chart-Aided Near-Capacity Designs for Wireless Channels. Hoboken, NJ, USA: Wiley, Mar. 2011.

[2] A. Chockalingam and B. S. Rajan, Large MIMO Systems. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[3] S. Yang and L. Hanzo, “Fifty years of MIMO detection: The road to large-scale MIMO systems,” *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, 4th Quart., 2015.

[4] C. Bockelmann et al., “Massive machine-type communications in 5G: Physical and MAC-layer solutions,” *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[5] I. F. Akyildiz, A. Kak, and S. Nie, “6G and beyond: The future of wireless communications systems,” *IEEE Access*, vol. 8, pp. 133985–134030, 2020.

[6] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, “Joint spatial division and multiplexing—The large-scale array regime,” *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[7] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, “Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[8] Y. Xu, G. Yue, N. Prasad, S. Rangarajan, and S. Mao, “User grouping and scheduling for large scale MIMO systems with two-stage precoding,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 5197–5202.

[9] A. Arvola, A. Tolli, and D. Gesbert, “Two-layer precoding for dimensionality reduction in massive MIMO,” in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2016, pp. 2000–2004.

[10] A. Padmanabhan and A. Tolli, “Interference management via user clustering in two-stage precoder design,” in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[11] A. Padmanabhan and A. Tolli, “An iterative approach for inter-group interference management in two-stage precoder design,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[12] W. C. Jakes, Microwave Mobile Communications, 1st ed. Hoboken, NJ, USA: Wiley, 1994.

[13] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[14] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, “A coordinated approach to channel estimation in large-scale multiple-antenna systems,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[15] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, “An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Oct. 2014.

[16] M. Mahdavi, O. Edfors, V. Owall, and L. Liu, “A low complexity massive MIMO detection scheme using angular-domain processing,” in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 181–185.

[17] M. Mahdavi, O. Edfors, V. Owall, and L. Liu, “A VLSI implementation of angular-domain massive MIMO detection,” in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2019, pp. 1–5.

[18] M. Mahdavi, O. Edfors, V. Owall, and L. Liu, “Angular-domain massive MIMO detection: Algorithm, implementation, and design tradeoffs,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 6, pp. 1948–1961, Jun. 2020.

[19] M. Abdelghany, U. Madhow, and A. Tolli, “Beamspace local LMMSE: An efficient digital backend for mmWave massive MIMO,” in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.

[20] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[21] Y. Kabashima, “A CDMA multiuser detection algorithm on the basis of belief propagation,” *J. Phys. A: Math. Gen.*, vol. 36, no. 43, 2003, Art. no. 11111.

[22] T. Takahashi, S. Ibi, and S. Sampei, “Design of adaptively scaled belief in multi-dimensional signal detection for higher-order modulation,” *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1986–2001, Mar. 2019.

[23] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and Y. Su, “Convergence analysis and assurance for Gaussian message passing iterative detector in massive MU-MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6487–6501, Sep. 2016.

[24] D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[25] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[26] L. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 2168–2172.

[27] K. Takeuchi, T. Tanaka, and T. Kawabata, “Performance improvement of iterative multiuser detection for large sparsely spread CDMA systems by spatial coupling,” *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1768–1794, Apr. 2015.

[28] M. Oppen and O. Winzen, “Expectation consistent approximate inference,” *J. Mach. Learn. Res.*, vol. 6, pp. 2177–2204, Dec. 2005. [Online]. Available: http://localhost/pubdb/p.php?3460
K. Takeuchi, “Rigorous dynamics of expectation-propagation-based inference,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 190–194.

S. Rangan, P. Schniter, and A. K. Fletcher, “Vector approximation message passing,” 2016, arXiv:1610.03082. [Online]. Available: http://arxiv.org/abs/1610.03082

K. Takeuchi, “Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements,” 2017, arXiv:1701.05284. [Online]. Available: http://arxiv.org/abs/1701.05284

T. P. Minka, “Expectation propagation for approximate Bayesian inference,” 2013, arXiv:1301.2294. [Online]. Available: http://arxiv.org/abs/1301.2294

S. Yang, T. Lv, R. G. Maudner, and L. Hanzo, “From nominal to true a posteriori probabilities: An exact Bayesian theorem based probabilistic data association approach for iterative MIMO detection and decoding,” *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2782–2793, Jul. 2013.

B. Çakmak, O. Winther, and B. H. Fleury, “S-AMP: Approximate message passing for general matrix ensembles,” 2014, arXiv:1405.2767. [Online]. Available: http://arxiv.org/abs/1405.2767

X. Meng, S. Wu, L. Kuang, and J. Lu, “An expectation propagation perspective on approximate message passing,” *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1194–1197, Aug. 2015.

X. Meng, S. Wu, and J. Zhu, “A unified Bayesian inference framework for generalized linear models,” *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 398–402, Mar. 2018.

Q. Zou, H. Zhang, C.-K. Wen, S. Jin, and R. Yu, “Concise derivation for generalized approximate message passing using expectation propagation,” *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1835–1839, Dec. 2018.

L. Liu, Y. Li, C. Huang, C. Yuen, and Y. L. Guan, “A new insight into GAMP and AMP,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8264–8269, Aug. 2019.

T. Takahashi, A. Tolli, S. Ibi, and S. Sampei, “Layered belief propagation for low-complexity large MIMO detection based on statistical beams,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

T. Takahashi, A. Tolli, S. Ibi, and S. Sampei, “Concatenated beam- and antenna-domain layered belief propagation for large MIMO detection,” in *Proc. 16th Int. Symp. Wireless Commun. Syst.* (ISWCS), Aug. 2019, pp. 592–596.

T. Takahashi, S. Ibi, and S. Sampei, “On normalization of matched filter belief in GaBP for large MIMO detection,” in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–6.

M. Hagiwara, T. Nishimura, T. Ohgane, and Y. Ogawa, “Node selection for belief propagation based channel equalization,” *IEICE Trans. Commun.*, vol. E100.B, no. 8, pp. 1285–1292, 2017.

C. Fan, X. Yuan, and Y. J. Zhang, “Source uplink signal detection in C-RANs via randomized Gaussian message passing,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5187–5200, Aug. 2017.

S. Yoon and C.-B. Chae, “Low-complexity MIMO detection based on belief propagation over pairwise graphs,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2363–2377, Jun. 2014.

Takumi Takahashi (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in communication engineering from Osaka University, Osaka, Japan, in 2016, 2017, and 2019, respectively. From 2018 to 2019, he was a Visiting Researcher with the Centre for Wireless Communications, University of Oulu, Finland. In 2019, he joined the Graduate School of Engineering, Osaka University, as an Assistant Professor. His current research interests include field of belief propagation, compressed sensing, signal processing, and wireless communications.

Antti Tolli (Senior Member, IEEE) received the Dr.Sc. (Tech.) degree in electrical engineering from the University of Oulu, Oulu, Finland, in 2008. From 1998 to 2003, he worked at Nokia Networks, as a Research Engineer and a Project Manager both in Finland and Spain. In May 2014, he was granted a five year (2014–2019) Academy Research Fellow post by the Academy of Finland. From 2015 to 2016, he visited at EURERCOM, Sophia Antipolis, France. From August 2018 to June 2019, he was visiting with the University of California at Santa Barbara, Santa Barbara, CA, USA. He is currently an Associate Professor with the Centre for Wireless Communications, University of Oulu. He has authored and co-authored numerous articles in peer-reviewed international journals and conferences and several patents all in the area of signal processing and wireless communications. His research interests include radio resource management and transceiver design for broadband wireless communications with a special emphasis on distributed interference management in heterogeneous wireless networks. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Shinsuke Ibi (Member, IEEE) received the B.E. degree in advanced engineering from the Sazuka College of Technology, Japan, in 2002, and the M.E. and Ph.D. degrees in communication engineering from Osaka University, Japan, in 2004 and 2006, respectively. From 2005 to 2006, he was a Visiting Researcher with the Centre for Wireless Communications, University of Oulu, Finland. In 2006, he joined the Graduate School of Engineering, Osaka University. From 2010 to 2011, he was a Visiting Researcher with the University of Southampton, UK. He moved to Doshisha University, in 2019, and he is currently a Professor with the Faculty of Science and Engineering. His research interests include EXIT-based coding theory, iterative detection, digital signal processing, cognitive radio, and communication theory. He received the 64th and 71st Best Paper Awards from IEICE and the 24th Telecom System Technology Award from the Telecommunication Advancement Foundation.

Seiichi Sampei (Fellow, IEEE) received the B.E., M.E., and Ph.D. degrees in electrical engineering from the Tokyo Institute of Technology, Japan, in 1980, 1982, and 1991, respectively. From 1982 to 1993, he was with the Communications Research Laboratory, Ministry of Posts and Telecommunications. From 1991 to 1992, he was a Visiting Researcher with the University of California at Davis, Davis, CA, USA. In 1993, he joined the Faculty of Engineering, Osaka University, where he is currently a Professor with the Department of Information and Communications Technology. He has been involved in the development of adaptive modulation, intelligent radio transmission/access, cognitive wireless networking, and wireless distributed network techniques. He is a member of the Institute of Image Information and Television Engineers (ITIE), and a fellow of the Institute of Electronics, Information and Communication Engineers (IEICE). He received the Shinhara Young Engineering Award, the Achievements Award, the Communications Society Best Paper Award, the Best Paper Award from IEICE, the Telecom System Technology Award, the Shida Rinzaburo Award from the Shida Rinzaburo Foundation, the DoCoMo Mobile Science Award from the Mobile Communication Foundation, the Ericsson Telecommunication Awards, and the Shida Rinzaburo Award from the President of Council for Info-Communications Promotion Month.