

Subspace Marginalized Belief Propagation for mmWave Overloaded MIMO Signal Detection

Takumi Takahashi^{*}, Shinsuke Ibi[†], Antti Tölli[‡], and Seiichi Sampei^{*}

^{*}Department of Information and Communications Technology, Osaka University, Yamada-oka 2-1, Suita 565-0871, Japan

[†] Faculty of Science and Engineering, Doshisha University, 1-3 Tataramiyakodani, Kyotanabe-shi, 610-0394, Japan

[‡] Centre for Wireless Communications (CWC), FI-90014 University of Oulu, Finland

Email: ^{*}{takahashi, sampei}@comm.eng.osaka-u.ac.jp, [†]sibi@mail.doshisha.ac.jp, [‡]antti.tolli@oulu.fi

Abstract—This paper deals with mmWave overloaded multi-user multi-input multi-output (MU-MIMO) detection, where the number of receive antennas is less than that of transmitted streams. Belief propagation (BP) is well known strategy for achieving large-scale MU detection (MUD) with low-complexity and high-accuracy. However, in mmWave massive MUD, the BP-based signal detector is subject to ill convergence behavior of iterative detection due to under-determined problem induced by spatial overloading and strong correlation among user channels induced by narrow angular spread of receive signal and line-of-sight (LOS) environments. To alleviate these impairments, we propose a novel iterative MUD approach based on beam-domain subspace marginalized BP (SMBP). Exploiting the approximate sparsity of beam-domain channels, the maximum likelihood (ML) principle is used to combine the strongly correlated signal subspace with reduced dimension while the BP-based detection is used for the remaining complementary subspace. The space partitioning criterion is adaptively determined based on channel state information (CSI) so that the two subspaces are as orthogonal as possible. Numerical results show that the proposed method is able to serve a massive number of wireless connections with low computational complexity even in the LOS environment, while providing excellent BER performance.

I. INTRODUCTION

Large multi-input multi-output (MIMO), which can simultaneously provide a high degree of spatial multiplexing gain and high detection reliability, has been regarded as one of the most promising technologies in the physical layer of wireless communication systems [1], [2]. In uplink scenarios, overloaded massive multi-user MIMO (MU-MIMO) aimed at simultaneously supporting more wireless connections than the number of receive antennas at a base station (BS) plays a crucial role for dealing with the explosive increase in the number of wireless terminals [3]–[6]. However, large-scale MU detection (MUD) in strongly correlated fading channels caused by the narrow angular spread of receive signal and line-of-sight (LOS) environments assumed for mmWave MIMO communications, has not been sufficiently investigated.

As low-complexity MUD solutions, linear spatial filters such as minimum mean square error (MMSE) filters are often utilized. However, the matrix operations based on the high dimensional channel matrix are unavoidable, resulting in high computational burden and large circuit scale of the receiver. Additionally, the performance is considerably worse than that of optimal maximum likelihood (ML) detection. To improve the detection capability while reducing the computational

cost, stochastic iterative detection schemes based on belief propagation (BP) have been proposed [7]–[13]. The BP-based signal detector can take advantage of the law of large numbers to simplify the computation of conditional symbol expectation. The most noticed BP-based detector is approximate message passing (AMP) [10], [11], which is systematically derived from an exact approximation of Gaussian BP (GaBP) [8], [9] in the large-system limit occurring when the input and output dimensions, M and N respectively, increase to infinity for a given compression rate $\rho = N/M$. These algorithms consist of only scalar-wise operations, and therefore can achieve the computational complexity $O(MN)$ for each iteration process. Furthermore, a rigorous proof of the convergence of BP-based detectors to the Bayes-optimal solution in the large-system limit was presented in [11], [13]. However, this proof requires that the antenna-domain channel matrix entries follow an independent identically distributed (i.i.d.) with zero mean. In other words, the convergence property of iterative detection is severely degraded under the strong spatial fading correlation found in mmWave MIMO channels. Even when we employ iterative detectors based on the expectation propagation (EP) framework [14], which was proposed in [12], [13] to overcome AMP's vulnerability against channel correlation, the performance degradation is inevitable due to noise enhancement induced by the spatial filtering step, especially in overloaded configurations. The matrix inversion operation required for each iteration is also unsuitable for large-scale MUD.

To solve the lack of an appropriate signal detector, the present study particularly focuses on the beam-domain signal processing [15]–[17] on the premise of fully digital beamforming (FDBF) at the receiver. In practice, the user equipments (UEs) tend to be distributed according to the terrain around the BS, and the angular spread of receive waves is tightly limited in mmWave communications. Consequently, the received signal power is concentrated in a subspace of the beam-domain channel, and the interference between corresponding UEs at close angular positions is strongly correlated. In this case, it is hard to separate multiplexed signals using the typical BP-based detectors. On the other hand, the interference due to the sidelobe leakage in beam domain is relatively small and with low correlation among users with larger spatial separation, and hence, it is more suitable for the BP-based detection.

Subspace marginalization (SM) uses the ML principle to

jointly detect only a part of signal space and it is one of the most effective ways to deal with MUD scenarios where the severe interference is highly concentrated. The origin of SM method dates back to sphere detection (SD) and tree-search detection (TSD) [2], [7]. Although these SM-based detection schemes can achieve the near-ML performance even in the presence of severe interference among UEs, setting a fixed criterion of search range is infeasible for large-scale MUD with many search candidates. To tackle this problem, subspace marginalization with interference suppression (SUMIS) [18] was proposed, where the SM method is introduced for mitigating the negative impact of the instantaneous correlation caused by insufficient channel hardening effect [7] in small-scale MUD. The multiple matrix inversion operations are still unavoidable, but the near-SD performance can be achieved with much less computational cost.

Inspired by the idea of [18], in this paper, we propose a novel beam-domain subspace marginalized belief propagation (SMBP) which allows us to use BP-based iterative detection even in mmWave overloaded massive MUD. By taking advantage of the above statistical properties of beam-domain channels, the proposed SMBP method marginalizes strongly correlated signal subspace based on the ML principle while the remaining complementary subspace is combined using the Gaussian BP approach with low computational cost.

Throughout this paper, $P_{a|b}[a|b]$ and $p_{a|b}(a|b)$ respectively represent the conditional probability mass function (PMF) and the probability density function (PDF) of a realization a of random variable \mathbf{a} given the occurrence of a realization b of random variable \mathbf{b} . $\mathbb{E}_{\mathbf{a}}\{\cdot\}$ is the expected value of random variable \mathbf{a} . $\mathbb{E}_{\mathbf{a}|b}\{\cdot\}$ denotes the conditional expectation of random variable \mathbf{a} given the occurrence of a realization b of random variable \mathbf{b} . $CN(a, b)$ indicates a complex-valued Gaussian process with mean a and variance b . \mathbf{I}_a represents an $a \times a$ square identity matrix. $\text{diag}[\mathbf{a}]$ denotes a diagonal matrix with the elements of \mathbf{a} . \odot represents Hadamard product.

II. SYSTEM MODEL

A. Antenna-domain signal model

Consider an uplink overloaded MUD system, where the BS has N' receive (RX) antennas in uniform linear array (ULA) pattern and M' ($> N'$) UEs are equipped with a single transmit (TX) antenna. The antenna-domain RX vector is given by the following linear regression model:

$$\mathbf{y}^c = [y_1^c, \dots, y_{N'}^c, \dots, y_{N'}^c]^T = \mathbf{H}^c \mathbf{x}^c + \mathbf{z}^c, \quad (1)$$

where $\mathbf{x}^c = [x_1^c, \dots, x_{M'}^c, \dots, x_{M'}^c]^T \in \mathcal{X}^c$ and $\mathbf{z}^c = [z_1^c, \dots, z_{N'}^c, \dots, z_{N'}^c]^T \in \mathbb{C}^{N' \times 1}$ respectively denote the TX and noise vectors. $\mathbf{H}^c \in \mathbb{C}^{N' \times M'}$ is $N' \times M'$ antenna-domain MIMO channel matrix. The m' -th UE conveys a modulated symbol $x_{m'}$, which represents one of Q' constellation points: $\mathcal{X}^c = \{\chi_1^c, \dots, \chi_{Q'}^c, \dots, \chi_{Q'}^c\}$. The average power density of the constellations in the set \mathcal{X}^c is denoted by E_s . As an example, $\mathcal{X}^c = \{\pm w_x \pm jw_x\}$ ($w_x = \sqrt{E_s/2}$) for quadrature phase shift keying (QPSK) signaling. The vector $\mathbf{z}^c \in \mathbb{C}^{N' \times 1}$ is a complex

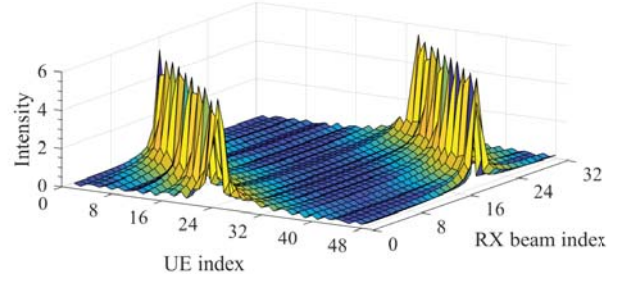


Fig. 1. Intensity of the elements in beam-domain channel \mathcal{E}^c , where we assume a $(N', M') = (32, 48)$ LOS-MIMO configuration. To clarify the characteristics, all the UEs are equally spread in the angular direction of the receiver and their indices are ordered based on the angular position.

additive white Gaussian noise (AWGN) vector whose entries z_n^c obey $CN(0, N_0)$, where N_0 is the noise spectral density. The covariance matrix is given by $\mathbb{E}_{\mathbf{z}^c} \{\mathbf{z}^c \mathbf{z}^{cH}\} = N_0 \mathbf{I}_{N'}$.

Without loss of generality, we here assume a LOS channel from each UE to the receiver at the BS. In this case, the channel vector for the m' -th UE (the m' -th column vector of \mathbf{H}^c) can be written as follows: [19]

$$\mathbf{h}_{m'} = a_{m'} [1 \quad \exp[j\Omega_{m'}] \quad \dots \quad \exp[j(N' - 1)\Omega_{m'}]]^T, \quad (2)$$

where $\Omega_{m'} = \pi \sin \theta_{m'}$, and the antenna element space is fixed to half the wavelength. $\theta_{m'}$ is the m' -th UE location in the azimuthal direction, and $a_{m'}$ depends on the radial location of the m' -th UE and wavelength for representing the path loss.

B. Beam-domain signal model

When the discrete Fourier transform (DFT) beamformer is used as the FDBF, the beam-domain RX vector is given by

$$\mathbf{r}^c = [r_1^c, \dots, r_{N'}^c, \dots, r_{N'}^c]^T = \mathbf{D}^H \mathbf{y}^c = \mathcal{E}^c \mathbf{x}^c + \mathbf{v}^c, \quad (3)$$

where $\mathbf{D} \in \mathbb{C}^{N' \times N'}$ denotes the DFT matrix. $\mathcal{E}^c = \mathbf{D}^H \mathbf{H}^c \in \mathbb{C}^{N' \times M'}$ and $\mathbf{v}^c = \mathbf{D}^H \mathbf{z}^c \in \mathbb{C}^{N' \times 1}$ are the beam-domain channel matrix and noise, respectively. Fig. 1 shows the intensity of the elements in \mathcal{E}^c . According to this characteristic (regularly biased and partial enhanced) beam-domain channel structure, SMBP enables to suppress severe interference.

For ease of algebraic manipulations, the complex-valued signal model of (3) can be interpreted as a double-size real-valued signal model on the basis of pulse amplitude modulation (PAM) symbols: [7]

$$\mathbf{r} = \mathcal{E} \mathbf{x} + \mathbf{v} \quad (4)$$

where

$$\mathbf{r} = \begin{bmatrix} \Re \{\mathbf{r}^c\} \\ \Im \{\mathbf{r}^c\} \end{bmatrix}, \quad \mathcal{E} = \begin{bmatrix} \Re \{\mathcal{E}^c\} & -\Im \{\mathcal{E}^c\} \\ \Im \{\mathcal{E}^c\} & \Re \{\mathcal{E}^c\} \end{bmatrix}, \\ \mathbf{x} = \begin{bmatrix} \Re \{\mathbf{x}^c\} \\ \Im \{\mathbf{x}^c\} \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \Re \{\mathbf{v}^c\} \\ \Im \{\mathbf{v}^c\} \end{bmatrix}. \quad (5)$$

Here, we define $\mathbf{r} \in \mathbb{R}^{N \times 1}$, $\mathcal{E} \in \mathbb{R}^{N \times M}$, $\mathbf{x} \in \mathbb{R}^{M \times 1}$, and $\mathbf{z} \in \mathbb{R}^{N \times 1}$, where $M = 2M'$ and $N = 2N'$. The m -th PAM symbol x_m in \mathbf{x} represents one of $Q (= \sqrt{Q'})$ PAM constellation points $\mathcal{X} = \{\chi_1, \dots, \chi_q, \dots, \chi_Q\}$ whose entries are amplitudes of the real and imaginary components of \mathcal{X}^c . The average

power density of the constellations in the set \mathcal{X} is denoted by $E_s/2$. As an example, $\mathcal{X} = \{\pm w_x\}$ for QPSK signaling. The n -th entry of \mathbf{r} can be represented by

$$r_n = \xi_n \mathbf{x} + v_n, \quad (6)$$

where $\xi_n = [\xi_{n,1}, \dots, \xi_{n,m}, \dots, \xi_{n,M}]$ is the n -th row vector of \mathbf{H} . The noise vector \mathbf{v} is a real-valued AWGN vector whose entries v_n obey $\mathcal{N}(0, N_0/2)$. With the above-mentioned signal model, the conditional PDF of RX vector \mathbf{r} is given by

$$p_{\mathbf{r}|\mathbf{x}}(\mathbf{r}|\mathbf{x}) \propto \exp\left[-\frac{(\mathbf{r} - \xi \mathbf{x})^2}{N_0}\right] = \prod_{n=1}^N \exp\left[-\frac{(r_n - \xi_n \mathbf{x})^2}{N_0}\right]. \quad (7)$$

III. SUBSPACE MARGINALIZED BELIEF PROPAGATION

Based on the beam-domain real-valued signal model of (4), the detailed algorithm of SMBP is presented. SMBP consists of three modules: soft interference canceller (soft IC), belief generator, and soft symbol generator (activation function). By propagating the belief (information reflecting the detection reliability) between each iteration process via the modules, SMBP can gradually improve the detection capability. To simplify the mathematical notations, we assume QPSK signaling ($\mathcal{X} = \{\pm w_x\}$, $w_x = \sqrt{E_s/2}$)¹.

A. Initial detection

At the first iteration ($k = 1$), no soft IC is conducted as there are no available prior information, thus, signal detection is performed using the knowledge of \mathbf{r} and ξ . First, the signal space is divided into the subspace that are combined based on the ML principle and the remaining complementary subspace that are combined using the GaBP approach based on the law of large numbers. For the detection of an arbitrary TX symbol x_m , $\exists m \in \{1, \dots, M\}$, the RX symbol r_n , $\exists n \in \{1, \dots, N\}$ can be expressed as the following partitioned model of (6):

$$\begin{aligned} r_n &= \xi_{n,m} x_m + \underbrace{\begin{bmatrix} \bar{\xi}_{n,m} & \tilde{\xi}_{n,m} \end{bmatrix}}_{\text{permut. of } \xi_{n,i} (i \neq m)} \underbrace{\begin{bmatrix} \bar{\mathbf{x}}_{n,m} \\ \tilde{\mathbf{x}}_{n,m} \end{bmatrix}}_{\text{permut. of } x_i (i \neq m)} + v_n \\ &= \xi_{n,m} x_m + \bar{\xi}_{n,m} \bar{\mathbf{x}}_{n,m} + \tilde{\xi}_{n,m} \tilde{\mathbf{x}}_{n,m} + v_n, \end{aligned} \quad (8)$$

where $\bar{\mathbf{x}}_{n,m} \in \mathbb{R}^{s \times 1}$ consists of s TX symbols that are expected to cause serious interference with x_m , and $\bar{\xi}_{n,m} \in \mathbb{R}^{1 \times s}$ denotes the corresponding channel vector. The subspace division of (8) is described in Section III.D. $\tilde{\xi}_{n,m} \tilde{\mathbf{x}}_{n,m}$ is the interference consisting of the other $M - (s + 1)$ TX symbols, and this term can be regarded as the summation of random variables.

In the large system, therefore, the PDF of (8) approaches an i.i.d. real Gaussian distribution as a result of the central limit theorem; this behavior is referred to as scalar Gaussian approximation (SGA) [7]. Accordingly, by approximating $\tilde{\xi}_{n,m} \tilde{\mathbf{x}}_{n,m} + v_n$ in (8) as a Gaussian random variable, r_n can be approximated by a Gaussian signal model:

$$r_n \approx r_{n,m} = \xi_{n,m} x_m + \bar{\xi}_{n,m} \bar{\mathbf{x}}_m + \phi_{n,m}, \quad \phi_{n,m} \sim \mathcal{N}(0, \psi_{n,m}), \quad (9)$$

¹The proposed SMBP detector can be extended to QAM signaling after some appropriate mathematical manipulations proposed in [9].

where $\phi_{n,m}$ is the effective noise, and the variance is derived from the expectation of random variables \mathbf{z}_n and $\bar{\mathbf{x}}_m$ as

$$\psi_{n,m} = \mathbb{E}_{\mathbf{z}_n, \bar{\mathbf{x}}_m} \{\phi_{n,m}^2\} = (E_s/2) \cdot \tilde{\xi}_{n,m} \tilde{\xi}_{n,m}^T + (N_0/2). \quad (10)$$

From (9) and (10) under the SGA condition, the conditional PDF of $r_{n,m}$ can be expressed as

$$\begin{aligned} p_{r_{n,m}|x_m, \bar{\mathbf{x}}_m}(r_{n,m}|x_m, \bar{\mathbf{x}}_m) \\ \propto \exp\left[-\frac{1}{2\psi_{n,m}} \left(r_{n,m} - \left[\xi_{n,m} x_m + \bar{\xi}_{n,m} \bar{\mathbf{x}}_m\right]\right)^2\right]. \end{aligned} \quad (11)$$

To perform signal detection based on Bayesian statistics, we need to accurately approximate a posterior log likelihood ratio (LLR) [1] using the likelihood function of (11).

Applying Bayes' theorem and sum rule, the posterior LLR l_m for detecting x_m is given by

$$\begin{aligned} l_m &= \ln \frac{P_{x_m|\mathbf{r}}[x_m = +w_x|\mathbf{r}]}{P_{x_m|\mathbf{r}}[x_m = -w_x|\mathbf{r}]} = \ln \frac{p_{\mathbf{r}|x_m}(\mathbf{r} + w_x) P_{x_m}[+w_x]}{p_{\mathbf{r}|x_m}(\mathbf{r} - w_x) P_{x_m}[-w_x]} \\ &= \ln \frac{\sum_{\bar{\mathbf{x}}_m} p_{\mathbf{r}|x_m, \bar{\mathbf{x}}_m}(\mathbf{r} + w_x, \bar{\mathbf{x}}_m) P_{x_m, \bar{\mathbf{x}}_m}[+w_x, \bar{\mathbf{x}}_m]}{\sum_{\bar{\mathbf{x}}'_m} p_{\mathbf{r}|x_m, \bar{\mathbf{x}}'_m}(\mathbf{r} - w_x, \bar{\mathbf{x}}'_m) P_{x_m, \bar{\mathbf{x}}'_m}[-w_x, \bar{\mathbf{x}}'_m]}, \end{aligned} \quad (12)$$

where $\bar{\mathbf{x}}_m \in \mathcal{X}^s$ represents a realization candidate of random vector $\bar{\mathbf{x}}_m$. Note that the exact computation of l_m requires the ML complexity. When the occurrence of TX symbol from \mathcal{X} is equiprobable ($P_{x_m}[\chi_q] = \frac{1}{Q}, \forall m, \forall q$), then

$$l_m = \ln \frac{\sum_{\bar{\mathbf{x}}_m} p_{\mathbf{r}|x_m, \bar{\mathbf{x}}_m}(\mathbf{r} + w_x, \bar{\mathbf{x}}_m)}{\sum_{\bar{\mathbf{x}}'_m} p_{\mathbf{r}|x_m, \bar{\mathbf{x}}'_m}(\mathbf{r} - w_x, \bar{\mathbf{x}}'_m)}. \quad (13)$$

Assuming that $r_{1,m}, \dots, r_{N,m}$ are (conditional²) independent under the SGA condition in (9), l_m can be approximated by

$$l_m \approx \ln \frac{\sum_{\bar{\mathbf{x}}_m} \prod_{i=1}^N p_{r_{i,m}|x_m, \bar{\mathbf{x}}_m}(r_{i,m} + w_x, \bar{\mathbf{x}}_m)}{\sum_{\bar{\mathbf{x}}'_m} \prod_{j=1}^N p_{r_{j,m}|x_m, \bar{\mathbf{x}}'_m}(r_{j,m} - w_x, \bar{\mathbf{x}}'_m)}. \quad (14)$$

According to (14), the posterior LLR can be approximately computed using the likelihood function of (11). Obviously, the approximation accuracy depends on the SGA accuracy, and it can be enhanced when the strongly correlated signal subspace is properly marginalized by the SM method due to the sidelobe attenuation in the beam-domain channel.

Nevertheless, if the approximated l_m in (14) is utilized as the prior belief, SMBP is subject to ill-convergence behavior of iterative detection, due to the noise correlation between r_n and $r_{n,m}$ included in l_m at the next iteration process. To mitigate the harmful impact of self-noise regression, in BP regime [8], the prior belief for the n -th RX symbol r_n is provided by subtracting the self-likelihood information as

$$\begin{aligned} l_{n,m} &= \ln \frac{\sum_{\bar{\mathbf{x}}_m} \prod_{i=1, i \neq n}^N p_{r_{i,m}|x_m, \bar{\mathbf{x}}_m}(r_{i,m} + w_x, \bar{\mathbf{x}}_m)}{\sum_{\bar{\mathbf{x}}'_m} \prod_{j=1, j \neq n}^N p_{r_{j,m}|x_m, \bar{\mathbf{x}}'_m}(r_{j,m} - w_x, \bar{\mathbf{x}}'_m)} \\ &= \ln \frac{\sum_{\bar{\mathbf{x}}_m} \exp\left[-\sum_{i=1, i \neq n}^N \frac{(r_{i,m} - \xi_{i,m} w_x - \bar{\xi}_{i,m} \bar{\mathbf{x}}_m)^2}{2\psi_{i,m}}\right]}{\sum_{\bar{\mathbf{x}}'_m} \exp\left[-\sum_{j=1, j \neq n}^N \frac{(r_{j,m} + \xi_{j,m} w_x - \bar{\xi}_{j,m} \bar{\mathbf{x}}'_m)^2}{2\psi_{j,m}}\right]}. \end{aligned} \quad (15)$$

²At the second and later iterations, $r_{1,m}, \dots, r_{N,m}$ are assumed to be conditional independent given the prior belief from the previous iteration.

Finally, the conditional expectation of \mathbf{x} given the prior belief vector $\mathbf{l}_n = [l_{n,1}, \dots, l_{n,m}, \dots, l_{n,M}]$ is given by

$$\begin{aligned} \check{\mathbf{x}}_n &= [\check{x}_{n,1}, \dots, \check{x}_{n,m}, \dots, \check{x}_{n,M}]^T = \mathbb{E}_{\mathbf{x}|\mathbf{l}_n} \{\mathbf{x}\} \\ &= \sum_{\chi} \chi P_{\mathbf{x}|\mathbf{l}_n}[\chi|\mathbf{l}_n] = \sum_{\chi} \chi \frac{p_{\mathbf{l}_n|\mathbf{x}}(\mathbf{l}_n|\chi)}{\sum_{\chi'} p_{\mathbf{l}_n|\mathbf{x}'}(\mathbf{l}_n|\chi')}, \end{aligned} \quad (16)$$

where $\chi \in \mathcal{X}^M$ represents a realization candidate of random vector \mathbf{x} . Assuming that $l_{n,m}$ satisfies the consistency condition [20], and that its statistical behavior is uncorrelated among the elements of \mathbf{l}_n under the SGA condition, the entries of (16) are readily given by the following activation function: [1]

$$\check{x}_{n,m} = \eta(l_{n,m}) = w_x \cdot \tanh\left[\frac{l_{n,m}}{2}\right]. \quad (17)$$

B. Iterative detection

At the second and later iterations ($k \neq 1$), signal detection is performed with the aid of $\check{x}_{n,m}, \forall n$, and $\forall m$ generated in the previous iteration process. In the detection of x_m , the soft IC for the n -th RX symbol r_n is conducted by subtracting the interference $\check{\xi}_{n,m}\check{\mathbf{x}}_m$ in (8) as

$$r_n - \check{\xi}_{n,m}\check{\mathbf{x}}_m = \xi_{n,m}x_m + \bar{\xi}_{n,m}\bar{\mathbf{x}}_m + \check{\xi}_{n,m}(\bar{\mathbf{x}}_m - \check{\mathbf{x}}_m) + \nu_n, \quad (18)$$

where $\check{\mathbf{x}}_m$ is the soft replica corresponding to $\bar{\mathbf{x}}_m$. By approximating the residual interference-plus-noise term in (18) as Gaussian random variable, the output belief can be characterized by a Gaussian signal model as

$$r_{n,m} = \xi_{n,m}x_m + \bar{\xi}_{n,m}\bar{\mathbf{x}}_m + \phi_{n,m}, \quad \phi_{n,m} \sim \mathcal{N}(0, \psi_{n,m}), \quad (19)$$

where the variance of $\phi_{n,m}$ is derived from the conditional expectation of random variables Z_n and $\bar{\mathbf{x}}_m$ given $\mathbf{l}_n = \mathbf{l}_n$ as

$$\begin{aligned} \psi_{n,m} &= \mathbb{E}_{Z_n, \bar{\mathbf{x}}_m|\mathbf{l}_n} \{\phi_{n,m}^2\} \\ &= \check{\xi}_{n,m} \mathbf{A}_{n,m} \check{\xi}_{n,m}^T + (N_0/2), \end{aligned} \quad (20)$$

$$\begin{aligned} \mathbf{A}_{n,m} &= \mathbb{E}_{\bar{\mathbf{x}}_m|\mathbf{l}_n} \left\{ (\bar{\mathbf{x}}_m - \check{\mathbf{x}}_m) (\bar{\mathbf{x}}_m - \check{\mathbf{x}}_m)^T \right\} \\ &= (E_s/2) \cdot \mathbf{I}_{M-(s+1)} - \text{diag}[\check{\mathbf{x}}_{n,m} \odot \check{\mathbf{x}}_{n,m}]. \end{aligned} \quad (21)$$

After updating the prior belief in (15) with (19) and (20), (17) is computed again in each iteration. When the number of iterations reaches the predetermined value K , x_m is detected by hard-decision of l_m as

$$\hat{x}_m = w_x \cdot \text{sgn}(l_m), \quad (22)$$

where $\text{sgn}(\cdot)$ denotes the operation for extracting the sign of a number.

C. Computational reduction with max-log approximation

For the numerical stability, the summations as in (15) are typically evaluated using the Jacobian logarithm: [1]

$$\ln\left(\sum_{u=1}^U \exp[a_u]\right) = f(\dots f(f(a_1, a_2), a_3), \dots), a_U), \quad (23)$$

where $f(a, b) = \max(a, b) + \ln(1 + \exp[-|a - b|])$. To avoid undesirable sequential processing including MNQ^{s+1} look-up-table (LUT) operations for evaluating $\exp[\cdot]$, max-log

approximation [21] is effective, where the second term of $f(a, b)$ is neglected. In the case of (15), there should be a clear magnitude relationship among the metrics based on the Euclidean distance due to the decaying sidelobes of beam-domain channels found in Fig. 1, therefore, the accurate approximation can be expected. Applying the max-log approximation to (15), the extrinsic belief can be computed as

$$l_{n,m} \approx \sum_{i=1, i \neq n}^N \frac{2w_x \xi_{i,m} (r_{i,m} - \bar{\xi}_{i,m} \bar{\mathbf{x}}_m^{(\max)})}{\psi_{i,m}}, \quad (24)$$

where $\bar{\mathbf{x}}_m^{(\max)}$ is a realization of $\bar{\mathbf{x}}_m$ to maximize the metric. Although the complexity order does not change due to the exhaustive search for max operation, (24) can significantly reduce the processing delay in practical implementation.

D. Criterion for partitioning signal space

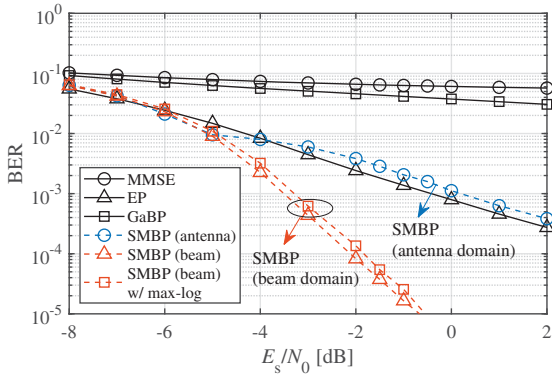
The remaining issue is how to partition the signal space in (8), i.e., how to select s TX symbols for SM, where s is a predetermined parameter for trade-off between the performance and computational complexity. Intuitively, when considering the detection of x_m , the detection capability of SMBP is improved as the impact of interference vector $\bar{\mathbf{x}}_m$ on $\bar{\mathbf{x}}_m$ is decreased. This is essentially equivalent to aligning the interference into the null space of $\bar{\mathbf{x}}_m$, that is, the signal space should be partitioned so that the inner product between the column vectors of $\bar{\mathbf{E}}_m = [\bar{\xi}_{1,m}^T, \dots, \bar{\xi}_{N,m}^T]^T$ and those of $\check{\mathbf{E}}_m = [\check{\xi}_{1,m}^T, \dots, \check{\xi}_{N,m}^T]^T$ becomes as small as possible. Note that SMBP is exactly the same as the optimal ML detection if the column spaces of $\bar{\mathbf{E}}_m$ and $\check{\mathbf{E}}_m$ are orthogonal. In order to select s symbols based on the above criterion, we should focus on the intensity of the off-diagonal elements in the following Gram matrix: $\mathbf{G} = \bar{\mathbf{E}}_m^T \check{\mathbf{E}}_m$. Denoting the m -th column vector by $\mathbf{g}_m = [g_{1,m}, \dots, g_{M,m}]^T$, we can select s indices corresponding to the largest values of $|g_{i,m}|$ ($i \neq m$) as a significant interference source. As a result, an appropriate subspace $\bar{\mathbf{E}}_m$ consists of s corresponding column vectors of $\bar{\mathbf{E}}_m$, and the rest of column vectors are placed in $\check{\mathbf{E}}_m$.

In this paper, the signal space partitioning is executed for each channel realization. However, if the long-term channel statistics, e.g., channel covariance matrix, is available at the BS, we can continue to use the same partitioned model, which is designed based on the average characteristic for each statistical beam, over multiple channel realization (multiple coherence times). Particularly in mmWave MU-MIMO channels, the long-term channel statistics can be clearly described with spatial correlation matrices, therefore, the signal space partition requires less frequent updates according to small fluctuation caused by slow UEs movement. It results in significant computational reduction in practical large-scale systems.

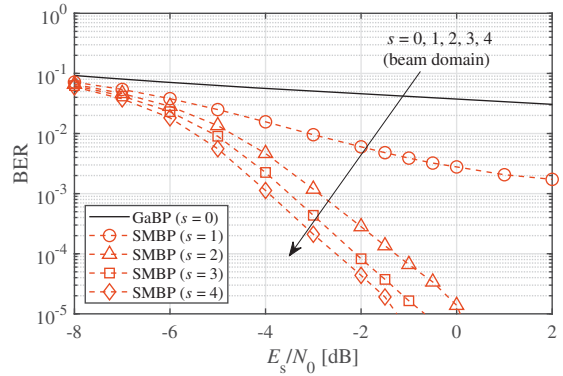
IV. NUMERICAL RESULTS

A. BER performance

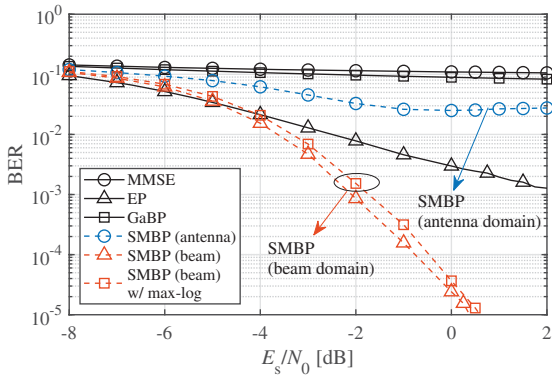
Computer simulations were conducted to validate the performance of the proposed SMBP detector for overloaded massive



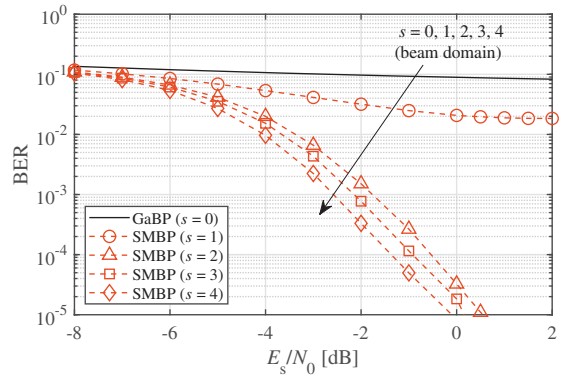
(a) $(N, M) = (64, 80)$: 125% overloaded MIMO, $(L_c, R_c) = (648, 5/6)$



(a) $(N, M) = (64, 80)$: 125% overloaded MIMO, $(L_c, R_c) = (648, 5/6)$



(b) $(N, M) = (64, 96)$: 150% overloaded MIMO, $(L_c, R_c) = (648, 2/3)$



(b) $(N, M) = (64, 96)$: 150% overloaded MIMO, $(L_c, R_c) = (648, 2/3)$

Fig. 2. BER performances of different detectors in overloaded massive LOS-MIMO detection, where the predetermined parameter $s = 3$ in SMBP.

Fig. 3. BER performances of the proposed SMBP-based iterative detection with various parameters $s = 0, 1, 2, 3, 4$, respectively.

MUD in mmWave communications. To represent the severe fading correlation, we use the LOS-MIMO channel model in (2). A sector antenna with 120 degrees opening is considered, and the UEs are randomly dropped in the above angular region around the BS, where the performance metrics are averaged over 1000 independent UE drops and channel realizations³. The average RX power from each TX antenna is assumed to be identical on the basis of slow TX power control ($a_{m'} = 1, \forall m'$), and the time and frequency synchronization are assumed to be perfect. The modulation scheme is Gray-coded QPSK.

Fig. 2 shows BER performances of different detectors in $(N, M) = (64, 80)$ and $(64, 96)$ overloaded MU-MIMO configurations, respectively. The low-density parity-check (LDPC) code of length $L_c = 648$ bits and rate $R_c = 5/6, 2/3$ used in the IEEE 802.11n standard is applied as the channel code. The error correction by the channel decoder is conducted only once after $K = 12$ iterative detection. To improve the convergence property, damping method [7] is utilized in all iterative detectors, where the damping factor ζ is set to $\zeta = 0.2$ for EP and $\zeta = 0.5$ for the other detectors⁴. As a baseline

³We enforce a minimum separation in spatial frequency between any two UEs for avoiding the occurrence of excessive interference caused by LOS channels, choosing it as half the 3dB beamwidth: $\Delta\Omega_{min} = \frac{2.783}{N^2}$ [19].

⁴The damping factor ζ is determined to minimize the BER at $E_s/N_0 = -2.0$ dB in computer simulations.

performance, the curve of typical linear MMSE filter “linear MMSE” is drawn. As a comparison with the state-of-the-art BP-based detectors, “GaBP” [9] and “EP” [12] performances are also presented. Note that the performances of linear MMSE and EP are equal in both antenna- and beam-domain because the impact of \mathbf{D}^H in (3) is completely canceled out at the linear filtering step due to the characteristic of $\mathbf{D}^H \mathbf{D} = \mathbf{I}_N$. By contrast, GaBP and SMBP should be performed in the beam domain because the contribution of \mathbf{D}^H remains through the non-linear operations in (15), and hence, allows to enhance the conditional independence among beliefs.

Under the strong spatial fading correlation, GaBP cannot sufficiently separate MIMO signals based on matched filters (MFs) [8], resulting in high-level error floors. EP can suppress the harmful impact of correlation owing to iterative MMSE filtering in exchange for the high computational burden, but the ill-convergence due to the noise enhancement at filtering step caused by under-determined MUD is inevitable. The most attractive feature is that our proposed “SMBP (beam)” can significantly improve the detection capability compared to the other methods. Remarkably, “SMBP (beam)” obtains about 6.5 dB gain from “EP” at $\text{BER} = 10^{-4}$ in Fig. 2(a) and achieves $\text{BER} = 10^{-5}$ by suppressing the error floor in 2(b). Note that such performance improvements are not obtained in “SMBP (antenna)”, where SMBP is performed in the antenna domain.

TABLE I
COMPUTATIONAL COMPARISON OF DIFFERENT DETECTORS

Detector	linear MMSE	EP [12]	SMBP
Complexity order	$O(MN^2 + N^3)$	$O(K(MN^2 + N^3))$	$O(KMNQ^{s+1})$

These results imply the effectiveness of SM method in the BP-based detector when utilising the statistical filtering property of beam-domain processing. Furthermore, even when the max-log approximation is applied for reducing the processing delay, the degradation is just 0.5 dB or less.

Let us now focus on the parameter s that determines the computational cost. Fig. 3 shows BER performance with various parameter settings $s = 0, 1, 2, 3, 4$. The other system parameters are the same as in Fig. 2. Note that $s = 0$ without the SM method is equivalent to GaBP. Interestingly, even when only $s = 2$ real-valued symbols (equivalent to 1 QPSK symbol) are marginalized, the convergence property of iterative detection is significantly improved compared to ‘‘GaBP’’. It is because that the high reliability of beliefs generated based on the SM method becomes the starting point for iterative convergence and enhances the reliability of the entire belief via soft IC. For the given system configuration, setting the parameter $s = 2, 3$ provides a reasonable trade-off between the performance and computational cost.

B. Complexity analysis

Table I summarizes the order of complexities. For evaluating the computational cost of iterative detection scheme, let us focus on the complexity order required for each TX vector \mathbf{x} detection. The complexity bottleneck of EP is the matrix operations according to $N \times M$ channel matrix in each iteration, resulting in $O(K(MN^2 + N^3))$ complexity. It rapidly grows as M and N increase. On the other hand, SMBP does not require any matrix operations, and the dominant factor for determining the computational cost is the computation of extrinsic prior beliefs in (15), leading to $O(KMNQ^{s+1})$ complexity. That is, SMBP can perform with lower computational cost compared to EP in the cases of $Q^{s+1} < N$. As an example, when the number of RX antennas is fixed to $N = 64$, the complexity of SMBP is roughly 7% and 14% fraction of the cases of EP for $s = 2$ and 3, respectively. Applying the max-log approximation, the practical computational burden (e.g., the required number of multiplications) becomes further smaller percentage while providing similar BER performance.

V. CONCLUSION

In this paper, we proposed a novel iterative detector based on beam-domain SMBP for overloaded massive MUD in mmWave MU-MIMO communications. Under such scenarios, the convergence property of BP-based signal detector is severely degraded due to under-determined problem induced by spatial overloading and strong correlation among user channels induced by narrow angular spread of receive signal and LOS environments. To deal with these impairments, the SM method on the premise of DFT beamforming at the receiver is introduced into the BP framework, according to

the approximate sparsity of beam-domain channels due to the sidelobe attenuation. Through computer simulations, we can confirm that the proposed method is able to significantly improve the detection capability as compared to the EP-based detector with much lower computational cost.

ACKNOWLEDGEMENT

This work was financially supported by JSPS KAKENHI Grant Number JP18H03765 and JP19K23516, Japan.

REFERENCES

- [1] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems Sphere-Packing, Iterative Detection and Cooperation*. Wiley-IEEE Press, May 2009.
- [2] S. Yang and L. Hanzo, ‘‘Fifty years of MIMO detection: The road to large-scale MIMOs,’’ *IEEE Commun. Surveys Tutorials*, vol. 17, no. 4, pp. 1941–1988, Fourthquarter 2015.
- [3] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, ‘‘Massive machine-type communications in 5G: physical and MAC-layer solutions,’’ *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sept. 2016.
- [4] K. Wong, A. Paulraj, and R. D. Murch, ‘‘Efficient high-performance decoding for overloaded MIMO antenna systems,’’ *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1833–1843, May 2007.
- [5] S. Yoon and C. Chae, ‘‘Low-complexity MIMO detection based on belief propagation over pairwise graphs,’’ *IEEE Trans. on Vehicular Technol.*, vol. 63, no. 5, pp. 2363–2377, June 2014.
- [6] R. Hayakawa and K. Hayashi, ‘‘Convex optimization-based signal detection for massive overloaded MIMO systems,’’ *IEEE Trans. on Wireless Commun.*, vol. 16, no. 11, pp. 7080–7091, Nov. 2017.
- [7] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge University Press, 2014.
- [8] Y. Kabashima, ‘‘A CDMA multiuser detection algorithm on the basis of belief propagation,’’ *J.Phys. A: Math. Gen.*, vol. 36, no. 43, p. 11111, Oct. 2003.
- [9] 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.
- [10] D. L. Donoho, A. Maleki, and A. Montanari, ‘‘Message-passing algorithms for compressed sensing,’’ *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.
- [11] 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.
- [12] 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.
- [13] K. Takeuchi, ‘‘Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements,’’ Jan. 2017. [Online]. Available: <http://arxiv.org/abs/1701.05284>
- [14] T. P. Minka, ‘‘Expectation propagation for approximate Bayesian inference,’’ *Proc. 17th Conf. UAI*, pp. 362–369, Aug. 2001.
- [15] A. Adhikary, J. Nam, J. 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.
- [16] M. Abdelghany, U. Madhow, and A. Tolli, ‘‘Beamspace local LMMSE: An efficient digital backend for mmWave massive MIMO,’’ in *2019 IEEE 20th Int. Workshop on SPAWC*, July 2019, pp. 1–5.
- [17] T. Takahashi, A. Tölli, S. Ibi, and S. Sampei, ‘‘Layered belief propagation for low-complexity large MIMO detection based on statistical beams,’’ in *Proc. IEEE ICC 2019, Shanghai, China*, May. 2019.
- [18] M. Ćirkić and E. G. Larsson, ‘‘SUMIS: Near-optimal soft-in soft-out MIMO detection with low and fixed complexity,’’ *IEEE Trans. Signal Processing*, vol. 62, no. 12, pp. 3084–3097, June 2014.
- [19] C. A. Balanis, *Antenna Theory: Analysis and Design*. New York, NY, USA: Wiley-Interscience, 2005.
- [20] J. Hagenauer, ‘‘The EXIT chart - introduction to extrinsic information transfer,’’ in *Proc. EUSIPCO 2004*, Vienna, Austria, Sept. 2004, pp. 1541–1548.
- [21] J. Vogt and A. Finger, ‘‘Improving the max-log-map turbo decoder,’’ *Electronics Letters*, vol. 36, no. 23, pp. 1937–1939, Nov 2000.