Low-complexity Beam-domain Channel Estimation with Long-term Statistics for Large MIMO Detection

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

[*]Department of Information and Communications Technology, Osaka University, Yamada-oka 2-1, Suita 565-0871, Japan
[†] Centre for Wireless Communications (CWC), FI-90014 University of Oulu, Finland
[‡] Faculty of Science and Engineering, Doshisha University, 1-3 Tataramiyakodani, Kyotanabe-shi, 610-0394, Japan
Email: [*]{takahashi, sampei}@comm.eng.osaka-u.ac.jp, [†]antti.tolli@oulu.fi, [‡]sibi@mail.doshisha.ac.jp

Abstract—This paper proposes low-complexity beam-domain channel estimation using long-term channel statistics in belief propagation (BP) based large multi-input multi-output (MIMO) detection. When the channel correlation matrix between the base station (BS) and each user equipment (UE) is available and used as prior information, maximum a-posteriori probability (MAP) estimation provides the optimal estimation performance. However, it requires undesirably complex large-scale matrix operations at any time the channel statistics is changed. By appropriately selecting beam-domain angular bins for each UE, the proposed method allows us to significantly reduce the computational cost while maintaining the near-optimal performance in terms of the mean square error (MSE) of estimated channel. The selection threshold is adaptively determined based on the prior information such as the channel correlation matrix, statistical beam, and receive SNR. For the subsequent BP-based signal detection, an appropriate covariance matrix is designed while considering the detrimental impact of channel estimation errors. Numerical results show that the proposed method can reduce the computational cost to less than 4% as compared to the MAP estimation, while providing similar MSE performance.

I. INTRODUCTION

Large multi-input multi-output (MIMO) systems are getting to a mature state with a significant deployment in the physical layer of future wireless communication systems [1], [2]. In particular, uplink multi-user MIMO (MU-MIMO) aimed at collecting data from a massive number of wireless terminals has attracted more and more attention [2], [3]. Under such scenarios, sophisticated channel estimation and signal detection play crucial roles for realizing low-complexity and high-accuracy MU detection (MUD) at a base station (BS).

The signal detection is performed based on the estimated channel with pilot sequences. Therefore, highly accurate channel estimation is indispensable for realizing large-scale MUD [4]. When the long-term channel statistics between the receiver and each wireless terminals are available based on observation at the BS, maximum a-posteriori probability (MAP) estimation using the knowledge as prior information achieves the maximum estimation performance in probability theory. However, the optimal estimation requires high-dimensional inverse matrix operations whenever the statistical information is updated, resulting in prohibitively high computational burden. On the other hand, a spatial filter derived from the least square (LS) criterion without the prior information allows us to estimate channel with lower computational cost, but the estimation accuracy is severely degraded.

To solve the problem, we focus on beam-domain channel estimation [4]–[6]. Since the user equipments (UEs) tend to be distributed according to the terrain around the BS, the angular spread of receive waves is naturally limited based on the local scatters around the UEs. Thus, their energy is concentrated in a subspace of the beam-domain channel. By exploiting the approximate sparse property, the beam-domain signal processing allows us to alleviate various challenging issues typified by pilot contamination [4]–[7].

The approximate sparsity of beam-domain channel indicates the fact that the estimation quality for each UE is different among the statistical beams. From the above perspective, the present study focus on angular–UE (beam–UE) bin selection. The statistical beam selection is a low-complexity strategy to improve the channel estimation quality while utilizing the prior information, where we can basically select a set of statistical beams oriented towards each UE to improve the pilot signal-to-noise ratio (SNR). That is, by estimating beam-domain channel based only on the bins whose SNR is more than 1 in linear scale while the other bins are discarded for removing noise (denoising), the estimation accuracy can be improved to near-MAP performance in large MU-MIMO scenarios. The selection threshold is given by the solution of the sum mean square error (MSE) minimization problem of the estimated channel, and it can be updated with remarkably low computational cost compared to the MAP estimation. The resultant criterion provides the same optimal solution as the exhaustive search. Additionally, the beneficial effect becomes larger as the number of transmit and receive antennas increase, thus, the proposed method has high scalability in large MU-MIMO scenarios.

Regarding the subsequent signal detection, message passing (MP) algorithm based on belief propagation (BP) [8]–[10] is introduced as large-scale MUD. The most noticed BP-based detector is approximate message passing (AMP) [8], but it is vulnerable to spatial fading correlations. Alternatively, we employ an iterative detector based on expectation propagation (EP) framework [11], which is proposed by [9], [10]. It can be naturally regarded as a vectorized AMP, and is proven in [10] to converge toward Bayes-optimal performance under some statistical assumptions in the large system limit. To maintain the attractive convergence property by propagating the correct belief in iterative detection, we also design appropriate covariance matrices such that the detrimental impact of channel estimation errors is properly taken into consideration.

Contributions of this paper are summarized as follows:
- Low-complexity beam-domain channel estimation is con-

sidered utilizing the long-term channel statistics.
- Appropriate covariance matrices are designed for the subsequent EP-based iterative detection, while taking into account the estimation errors.
- The proposed method is demonstrated to be feasible with remarkably low complexity while maintaining near-MAP performance in large-scale MUD.

The remainder of this paper is organized as follows. Sect. II presents a system model, and then the traditional channel estimation and covariance matrix required for belief generation are derived, respectively. A novel low-complexity beam-domain channel estimation based on the long-term channel statistics is proposed in Sect. III. Sect. IV briefly presents the EP-based detector using the modified covariance matrices in the presence of channel estimation errors. Sect. V validates the proposed method through computer simulations. Finally, Sect. VI concludes the paper with a summary.

Throughout this paper, $P_{\mathsf{a|b}}[a|b]$ and $p_{\mathsf{a|b}}(a|b)$ respectively represent the conditional probability mass function (PMF) and the probability density function (PDF) of a realization a of random variable a given the occurrence of a realization b of random variable b. $\mathbb{E}_{\mathsf{a}}\{\cdot\}$ is the expected value of random variable a. $\mathbb{E}_{\mathsf{a|b}=b}\{\cdot\}$ denotes the conditional expectation of random variable a given the occurrence of a realization b of random variable b. $\mathrm{diag}\,[a]$ denotes a diagonal matrix with the elements of a. The special operators $\otimes$ and $\odot$ respectively represent the Kronecker product and Hadamard product.

II. PRELIMINARY STEP

A. System model

Consider an uplink MUD system, where the BS has N receive (RX) antennas in uniform linear array (ULA) pattern and $M(\leq N)$ UEs are equipped with a single transmit (TX) antenna. At the k-th discrete time instance, the m-th UE transmits a TX symbol $x_m[k]$, which represents one among Q constellation points. A vector of the TX symbols is denoted by $x[k] \in \mathbb{C}^{M \times 1}$. The averaged energy of each symbol $x_m[k]$ is normalized to E_s. At the k-th discrete time instance, the n-th RX antenna receives a RX symbol $y_n[k]$ at the receiver. A vector of the RX symbols is denoted by $y[k] \in \mathbb{C}^{N \times 1}$. Assuming frequency flat $N \times M$ antenna-domain MIMO channel matrix $H \in \mathbb{C}^{N \times M}$, $y[k]$ is represented as

$$y[k] = Hx[k] + z[k], \tag{1}$$

where the vector $z[k] \in \mathbb{C}^{N \times 1}$ is a complex additive white Gaussian noise (AWGN) vector with $z[k] \sim \mathcal{CN}(0, N_0 I_N)$, and where N_0 is the noise spectral density.

To represent the spatial fading correlation among RX antennas based on UE locations in the azimuthal direction, the geometric one-ring model [12] is utilized. Assuming diffuse 2–D field of isotropic scatters around the UEs, the (i, j) element of the RX spatial correlation matrix for the m-th UE, $\Theta_m \in \mathbb{C}^{N \times N}$, is given by

$$[\Theta_m]_{i,j} = \frac{1}{\Delta\psi_m} \int_{\psi_m^{\min}}^{\psi_m^{\max}} \exp\left[\mathrm{j}\pi(i-j)\cos(\psi)\right] d\psi, \tag{2}$$

which represents the correlation coefficient between the i-th and j-th RX antenna elements. Here, waves arrive from the m-th UE with an angular spread $\Delta\psi_m = \psi_m^{\max} - \psi_m^{\min}$. The antenna element spacing is fixed to half the wavelength. From (2), the RX spatial correlation becomes more severe with a decrease in the angular spread of UEs. With the assistance of (2), the m-th column vector of H is computed by

$$h_m = \Theta_m^{1/2} v_m, \tag{3}$$

where $v_m \sim \mathcal{CN}(0, I_N)$.

Then, the spatio-temporal symbol matrices X, Y, and Z can be expressed as

$$X = [x[1], \ldots, x[k], \ldots, x[K]] \in \mathbb{C}^{M \times K}, \tag{4}$$
$$Y = [y[1], \ldots, y[k], \ldots, y[K]] \in \mathbb{C}^{N \times K}, \tag{5}$$
$$Z = [z[1], \ldots, z[k], \ldots, z[K]] \in \mathbb{C}^{N \times K}. \tag{6}$$

Under the above-mentioned assumptions, (1) is extended as

$$Y = HX + Z, \tag{7}$$

where we assume that the channel fluctuations of H are static during $1 \leq k \leq K$. In the TX symbol matrix X, each UE forms a frame with a length of K symbols, which include K_p pilot ($1 \leq k_\mathrm{p} = k \leq K_\mathrm{p}$) and K_d data ($1 \leq k_\mathrm{d} = k - K_\mathrm{p} \leq K_\mathrm{d}$) symbols, i.e. $K = K_\mathrm{p} + K_\mathrm{d}$. Thus, the spatio-temporal matrices of (4)–(6) are sub-divided into pilot and data parts as

$$X = [X_\mathrm{p}, X_\mathrm{d}], \quad \left(X_\mathrm{p} \in \mathbb{C}^{M \times K_\mathrm{p}}, X_\mathrm{d} \in \mathbb{C}^{M \times K_\mathrm{d}}\right), \tag{8}$$
$$Y = [Y_\mathrm{p}, Y_\mathrm{d}], \quad \left(Y_\mathrm{p} \in \mathbb{C}^{N \times K_\mathrm{p}}, Y_\mathrm{d} \in \mathbb{C}^{N \times K_\mathrm{d}}\right), \tag{9}$$
$$Z = [Z_\mathrm{p}, Z_\mathrm{d}], \quad \left(Z_\mathrm{p} \in \mathbb{C}^{N \times K_\mathrm{p}}, Z_\mathrm{d} \in \mathbb{C}^{N \times K_\mathrm{d}}\right), \tag{10}$$

where the pilot sequence for the m-th UE is defined as $s_m = [s_{m,1}, \ldots, s_{m,K_\mathrm{p}}]^\mathrm{T}$, thus, $X_\mathrm{p} = [s_1, \ldots, s_M]^\mathrm{T}$.

B. Channel estimation and error covariance matrix

Based on the RX symbol matrix of pilot part: $Y_\mathrm{p} = HX_\mathrm{p} + Z_\mathrm{p}$, channel estimation is performed before signal detection. Here, we assume the knowledge of the RX correlation matrices $\Theta_1, \ldots, \Theta_M$ in addition to Y_p and X_p are given at the BS.

1) MAP estimation: First, Y_p can be vectorized as

$$y_\mathrm{p} = \left[y[1]^\mathrm{T}, \ldots, y[k_\mathrm{p}]^\mathrm{T}, \ldots, y[K_\mathrm{p}]^\mathrm{T}\right]^\mathrm{T} = Sh + z_\mathrm{p}, \tag{11}$$

where we define

$$S = [s_1 \otimes I_N, \ldots, s_m \otimes I_N, \ldots, s_M \otimes I_N] \in \mathbb{C}^{NK_\mathrm{p} \times NM}, \tag{12}$$
$$h = \left[h_1^\mathrm{T}, \ldots, h_m^\mathrm{T}, \ldots, h_M^\mathrm{T}\right]^\mathrm{T} \in \mathbb{C}^{MN \times 1}, \tag{13}$$
$$z_\mathrm{p} = \left[z[1]^\mathrm{T}, \ldots, z[k_\mathrm{p}]^\mathrm{T}, \ldots, z[K_\mathrm{p}]^\mathrm{T}\right]^\mathrm{T} \in \mathbb{C}^{NK_\mathrm{p} \times 1}. \tag{14}$$

The MAP estimation that maximizes the posterior probability using all the given knowledge is formulated with the following maximization criterion of

$$\hat{h} = \arg\max_{\underline{h}} p_{\mathsf{h}|\mathsf{y_p}, \mathsf{S}, \Theta_1, \ldots, \Theta_M}\left(\underline{h}|y_\mathrm{p}, S, \Theta_1, \ldots, \Theta_M\right). \tag{15}$$

Hereafter, the prior variables $S, \Theta_1, \ldots, \Theta_M$ are appropriately omitted to simplify the mathematical notations. Using Bayes' theorem, the posterior probability can be rewritten as

$$p_{\mathsf{h}|\mathsf{y_p}}(h|y_\mathrm{p}) = \frac{p_{\mathsf{y_p}|\mathsf{h}}(y_\mathrm{p}|h)\,p_{\mathsf{h}}(h)}{}$$

From (3), the stochastic behavior of $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M$ is independent of each other. Therefore, a prior PDF is given by

$$p_{\mathsf{h}}(\boldsymbol{h}) = \prod_{m=1}^{M} p_{\mathsf{h}_m}(\boldsymbol{h}_m) = A \exp\left[-\sum_{m=1}^{M} \boldsymbol{h}_m^{\mathrm{H}} \boldsymbol{\Theta}_m^{-1} \boldsymbol{h}_m\right], \quad (17)$$

where we have $A = 1/(\pi^{MN} \cdot (\det[\boldsymbol{\Theta}_1] \cdots \det[\boldsymbol{\Theta}_M])^N)$ and $\mathbb{E}_{\mathsf{h}_m}\{\boldsymbol{h}_m \boldsymbol{h}_m^{\mathrm{H}}\} = \boldsymbol{\Theta}_m$. Substitution of (17) and the following likelihood function:

$$p_{\mathsf{y}_{\mathrm{p}}|\mathsf{h}}(\boldsymbol{y}_{\mathrm{p}}|\boldsymbol{h}) = \frac{1}{(\pi N_0)^{NK_{\mathrm{p}}}} \exp\left[-\frac{(\boldsymbol{y}_{\mathrm{p}} - \boldsymbol{S}\boldsymbol{h})^{\mathrm{H}}(\boldsymbol{y}_{\mathrm{p}} - \boldsymbol{S}\boldsymbol{h})}{N_0}\right], \quad (18)$$

into (16) yields

$$p_{\mathsf{h}|\mathsf{y}_{\mathrm{p}}}(\boldsymbol{h}|\boldsymbol{y}_{\mathrm{p}}) \;\propto\; \exp\left[-l(\boldsymbol{h})\right], \quad (19)$$

$$\because l(\boldsymbol{h}) \;=\; \boldsymbol{h}^{\mathrm{H}}\boldsymbol{\Theta}^{-1}\boldsymbol{h} + \frac{(\boldsymbol{y}_{\mathrm{p}} - \boldsymbol{S}\boldsymbol{h})^{\mathrm{H}}(\boldsymbol{y}_{\mathrm{p}} - \boldsymbol{S}\boldsymbol{h})}{N_0}, \quad (20)$$

where $\boldsymbol{\Theta} = \mathrm{diag}[\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_M] \in \mathbb{C}^{MN \times MN}$ is a block diagonal matrix. Thus, the MAP criterion of (15) can be rewritten as

$$\hat{\boldsymbol{h}} = \underset{\underline{\boldsymbol{h}} \in \mathbb{C}^{MN \times 1}}{\arg\min} \, l(\underline{\boldsymbol{h}}). \quad (21)$$

Finally, the MAP estimator can be found from the solution of Wirtinger derivative: $\frac{\partial l(\boldsymbol{h})}{\partial \boldsymbol{h}^{\mathrm{H}}} = 0$ as

$$\hat{\boldsymbol{h}} = \left[\boldsymbol{\Theta}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S} + N_0 \boldsymbol{I}_{MN}\right]^{-1} \boldsymbol{\Theta}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{y}_{\mathrm{p}}. \quad (22)$$

Interestingly, thanks to the matrix inversion identity: $(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B})^{-1}\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A})^{-1}$, the MAP estimator in (22) coincides with the minimum mean square error (MMSE) estimator: [5]

$$\hat{\boldsymbol{h}} = \boldsymbol{\Theta}\boldsymbol{S}^{\mathrm{H}} \left[\boldsymbol{S}\boldsymbol{\Theta}\boldsymbol{S}^{\mathrm{H}} + N_0 \boldsymbol{I}_{NK_{\mathrm{p}}}\right]^{-1} \boldsymbol{y}_{\mathrm{p}}. \quad (23)$$

The estimated channel matrix $\hat{\boldsymbol{H}}_{\mathrm{MAP}}$ is given by reshaping $\hat{\boldsymbol{h}}$ to $N \times M$ matrix. The estimation error matrix is defined as

$$\hat{\boldsymbol{\Delta}}_{\mathrm{MAP}} = \boldsymbol{H} - \hat{\boldsymbol{H}}_{\mathrm{MAP}}. \quad (24)$$

The sum MSE is computed by the mean square Frobenius norm of error matrix in (24) as

$$J_{\mathrm{MAP}} = \mathbb{E}_{\mathsf{H}, \mathsf{Z}_{\mathrm{p}}} \left\{\left|\hat{\boldsymbol{\Delta}}_{\mathrm{MAP}}\right|_{\mathrm{F}}^2\right\} = \mathbb{E}_{\mathsf{H}, \mathsf{Z}_{\mathrm{p}}} \left\{\mathrm{tr}\left[\left(\boldsymbol{h} - \hat{\boldsymbol{h}}\right)\left(\boldsymbol{h} - \hat{\boldsymbol{h}}\right)^{\mathrm{H}}\right]\right\}$$

$$= \mathrm{tr}\left[\boldsymbol{\Theta}\left(\boldsymbol{I}_{MN} + \frac{1}{N_0}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\boldsymbol{\Theta}\right)^{-1}\right]. \quad (25)$$

Here, with the use of $\hat{\boldsymbol{H}}_{\mathrm{MAP}}$, the data part of (7): $\boldsymbol{Y}_{\mathrm{d}} = \boldsymbol{H}\boldsymbol{X}_{\mathrm{d}} + \boldsymbol{Z}_{\mathrm{d}}$ can be rewritten as

$$\boldsymbol{Y}_{\mathrm{d}} \;=\; \hat{\boldsymbol{H}}_{\mathrm{MAP}}\boldsymbol{X}_{\mathrm{d}} + \hat{\boldsymbol{Z}}_{\mathrm{d}}, \quad \because \hat{\boldsymbol{Z}}_{\mathrm{d}} = \hat{\boldsymbol{\Delta}}_{\mathrm{MAP}}\boldsymbol{X}_{\mathrm{d}} + \boldsymbol{Z}_{\mathrm{d}}, \quad (26)$$

where $\hat{\boldsymbol{Z}}_{\mathrm{d}}$ is the effective noise including channel estimation errors. Let $\hat{\boldsymbol{z}}[k_{\mathrm{d}}]$ denote the k_{d}-th column vector of $\hat{\boldsymbol{Z}}_{\mathrm{d}}$. In this case, the covariance matrix of $\hat{\boldsymbol{z}}[k_{\mathrm{d}}]$ is given by

$$\hat{\boldsymbol{\Psi}}_{\mathrm{MAP}} = \mathbb{E}_{\mathsf{H}, \mathsf{Z}}\left\{\hat{\boldsymbol{z}}[k_{\mathrm{d}}]\hat{\boldsymbol{z}}^{\mathrm{H}}[k_{\mathrm{d}}]\right\} = E_{\mathrm{s}} \cdot \mathbb{E}_{\mathsf{H}, \mathsf{Z}_{\mathrm{p}}}\left\{\hat{\boldsymbol{\Delta}}_{\mathrm{MAP}}\hat{\boldsymbol{\Delta}}_{\mathrm{MAP}}^{\mathrm{H}}\right\} + N_0 \boldsymbol{I}_N$$

$$= E_{\mathrm{s}} \cdot \mathbb{E}_{\mathsf{H}, \mathsf{Z}_{\mathrm{p}}}\left\{\sum_{m=1}^{M}\left(\boldsymbol{h}_m - \hat{\boldsymbol{h}}_m\right)\left(\boldsymbol{h}_m - \hat{\boldsymbol{h}}_m\right)^{\mathrm{H}}\right\} + N_0 \boldsymbol{I}_N. \quad (27)$$

When the pilot matrix $\boldsymbol{X}_{\mathrm{p}}$ consists of orthogonal sequences (e.g., Zadoff-Chu (ZC) sequences [13]) for minimizing the Frobenius norm of error matrix, $\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S} = K_{\mathrm{p}}\boldsymbol{I}_{MN}$ holds. Conse-

quently, (27) can be rewritten as

$$\hat{\boldsymbol{\Psi}}_{\mathrm{MAP}} = E_{\mathrm{s}} \cdot \sum_{m=1}^{M} \boldsymbol{\Theta}_m \left[\boldsymbol{I}_N + \frac{K_{\mathrm{p}}}{N_0}\boldsymbol{\Theta}_m\right]^{-1} + N_0 \boldsymbol{I}_N. \quad (28)$$

By generating the belief based on $\hat{\boldsymbol{H}}_{\mathrm{MAP}}$ and $\hat{\boldsymbol{\Psi}}_{\mathrm{MAP}}$ in the subsequent BP-based detector, highly accurate signal detection can be achieved. However, the update of (22) for tracking changes in the channel statistics $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_M$ becomes computationally expensive and infeasible, due to high-dimensional matrix operations, as both M and N increase. The detailed complexity analysis is presented in Sect. V-A.

2) LS estimation: The channel estimator on the basis of LS criterion is given by

$$\hat{\boldsymbol{H}}_{\mathrm{LS}} \;=\; \underset{\underline{\boldsymbol{H}}}{\arg\min} \, p_{\mathsf{Y}_{\mathrm{p}}|\mathsf{H}, \mathsf{X}_{\mathrm{p}}}(\boldsymbol{Y}_{\mathrm{p}}|\underline{\boldsymbol{H}}, \boldsymbol{X}_{\mathrm{p}})$$

$$= \;\underset{\underline{\boldsymbol{H}}}{\arg\min} \, \left|\boldsymbol{Y}_{\mathrm{p}} - \underline{\boldsymbol{H}}\boldsymbol{X}_{\mathrm{p}}\right|_{\mathrm{F}}^2 = \boldsymbol{Y}_{\mathrm{p}}\boldsymbol{X}_{\mathrm{p}}^{\dagger}, \quad (29)$$

where $\boldsymbol{X}_{\mathrm{p}}^{\dagger} = \boldsymbol{X}_{\mathrm{p}}^{\mathrm{H}}\left[\boldsymbol{X}_{\mathrm{p}}\boldsymbol{X}_{\mathrm{p}}^{\mathrm{H}}\right]^{-1}$. The estimation error matrix of LS estimator is defined as

$$\hat{\boldsymbol{\Delta}}_{\mathrm{LS}} = \boldsymbol{H} - \hat{\boldsymbol{H}}_{\mathrm{LS}} = -\boldsymbol{Z}_{\mathrm{p}}\boldsymbol{X}_{\mathrm{p}}^{\dagger}, \quad (30)$$

and the sum MSE is given by

$$J_{\mathrm{LS}} \;=\; \mathbb{E}_{\mathsf{Z}_{\mathrm{p}}}\left\{\left|\hat{\boldsymbol{\Delta}}_{\mathrm{LS}}\right|_{\mathrm{F}}^2\right\} = \mathrm{tr}\left[\mathbb{E}_{\mathsf{Z}_{\mathrm{p}}}\left\{\hat{\boldsymbol{\Delta}}_{\mathrm{LS}}^{\mathrm{H}}\hat{\boldsymbol{\Delta}}_{\mathrm{LS}}\right\}\right]$$

$$= \;NN_0\mathrm{tr}\left[\left[\boldsymbol{X}_{\mathrm{p}}\boldsymbol{X}_{\mathrm{p}}^{\mathrm{H}}\right]^{-1}\right]. \quad (31)$$

With the use of $\hat{\boldsymbol{H}}_{\mathrm{LS}}$, $\boldsymbol{Y}_{\mathrm{d}}$ in (7) can be rewritten as

$$\boldsymbol{Y}_{\mathrm{d}} = \hat{\boldsymbol{H}}_{\mathrm{LS}}\boldsymbol{X}_{\mathrm{d}} + \hat{\boldsymbol{Z}}_{\mathrm{d}}, \quad \because \hat{\boldsymbol{Z}}_{\mathrm{d}} = \hat{\boldsymbol{\Delta}}_{\mathrm{LS}}\boldsymbol{X}_{\mathrm{d}} + \boldsymbol{Z}_{\mathrm{d}}. \quad (32)$$

Assuming the use of orthogonal pilots, the covariance matrix of $\hat{\boldsymbol{z}}[k_{\mathrm{d}}]$ is given by

$$\hat{\boldsymbol{\Psi}}_{\mathrm{LS}} \;=\; \mathbb{E}_{\mathsf{Z}}\left\{\hat{\boldsymbol{z}}[k_{\mathrm{d}}]\hat{\boldsymbol{z}}^{\mathrm{H}}[k_{\mathrm{d}}]\right\} = N_0 \left(E_{\mathrm{s}}\frac{M}{K_{\mathrm{p}}} + 1\right)\boldsymbol{I}_N. \quad (33)$$

From (29) and (33), the LS estimator does not require the large-scale matrix operations found in MAP estimation, leading to significant reduction of the computational cost. However, the knowledge of RX correlation matrices $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_M$ can not be utilized as the prior information, and the estimation accuracy is severely degraded compared to MAP estimation. Therefore, a novel method for using the long-term channel statistics with lower computational cost is worth considering.

III. Low-complexity Beam-domain Channel Estimation Using Long-term Statistics

From (29), the essence of the LS estimator is the maximization of the likelihood function: $p_{\mathsf{Y}_{\mathrm{p}}|\mathsf{H}, \mathsf{X}_{\mathrm{p}}}(\boldsymbol{Y}_{\mathrm{p}}|\boldsymbol{H}, \boldsymbol{X}_{\mathrm{p}})$, and the estimated channel is given by the solution of optimization problem using only the knowledge of $\boldsymbol{Y}_{\mathrm{p}}$ and $\boldsymbol{X}_{\mathrm{p}}$. In the beam-domain channel estimation, the estimation quality can be enhanced using the long-term channel statistics $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_M$ and N_0 as the prior information.

A. Beam-domain channel estimation

Assuming RX digital beamforming using discrete Fourier transform (DFT) at the receiver, a pilot part of the spatio-

temporal RX matrix in the beam domain is given by

$$R_\mathrm{p} = D^\mathrm{H} Y_\mathrm{p} = \Xi X_\mathrm{p} + D^\mathrm{H} Z_\mathrm{p}, \qquad (34)$$

where $D = [d_1, \ldots, d_N] \in \mathbb{C}^{N \times N}$ denotes a DFT matrix, and $\Xi = D^\mathrm{H} H$ denotes the equivalent beam-domain channel matrix. As in the cases of (29) and (30), the LS estimator and error matrix in the beam domain are expressed as

$$\hat{\Xi} = R_\mathrm{p} X_\mathrm{p}^\dagger, \quad \hat{\varLambda}_\Xi = D^\mathrm{H} \hat{\varLambda}_\mathrm{LS} = \Xi - \hat{\Xi}, \qquad (35)$$

respectively. Too many selected bins just increase the noise power, and too few do not sufficiently collect the desired signal energy. Based on the above equations in (35), an appropriate set of angular bins should be selected for the each UE in the beam domain to maximize the estimation quality. Denoting the selection matrix by $G \in \{0,1\}^{N \times M}$, the optimal selection (denoising) threshold can be given by the solution of the following sum MSE minimization problem as

$$G = \arg \min_{\underline{G} \in \{0,1\}^{N \times M}} \underbrace{\mathbb{E}_{\mathsf{H}, \mathsf{Z}_\mathrm{p}} \left\{ \left| \Xi - \underline{G} \odot \hat{\Xi} \right|_\mathrm{F}^2 \right\}}_{J_\Xi}. \qquad (36)$$

Using (35), the sum MSE of (36) can be rewritten as

$$\begin{aligned} J_\Xi &= \mathbb{E}_{\mathsf{H}, \mathsf{Z}_\mathrm{p}} \left\{ \left| (1 - \underline{G}) \odot \Xi + \underline{G} \odot \hat{\varLambda}_\Xi \right|_\mathrm{F}^2 \right\} \\ &= \mathbb{E}_{\mathsf{H}, \mathsf{Z}_\mathrm{p}} \left\{ \sum_{n=1}^N \sum_{m=1}^M \left| (1 - \underline{g}_{n,m}) \cdot \xi_{n,m} + \underline{g}_{n,m} \cdot \hat{\delta}_{n,m} \right|^2 \right\}, \end{aligned} \qquad (37)$$

where $\underline{g}_{i,j}$, $\xi_{i,j}$, and $\hat{\delta}_{i,j}$ denote the element in the i-th row and the j-th column of $\underline{G}$, Ξ, and $\hat{\varLambda}_\Xi$, respectively. Using the property: $(1 - \underline{g}_{i,j}) \cdot \underline{g}_{i,j} = 0$, (37) can be expressed as

$$\begin{aligned} J_\Xi &= \sum_{n=1}^N \sum_{m=1}^M \left[(1 - \underline{g}_{n,m})^2 \cdot \mathbb{E}_\mathsf{H} \left\{ |\xi_{n,m}|^2 \right\} + \underline{g}_{n,m}^2 \cdot \mathbb{E}_{\mathsf{Z}_\mathrm{p}} \left\{ |\hat{\delta}_{n,m}|^2 \right\} \right] \\ &= \sum_{n=1}^N \sum_{m=1}^M \left[(1 - \underline{g}_{n,m}) \cdot d_n^\mathrm{H} \Theta_m d_n + \underline{g}_{n,m} \cdot \frac{J_\mathrm{LS}}{MN} \right], \end{aligned} \qquad (38)$$

where we use

$$\mathbb{E}_\mathsf{H} \left\{ |\xi_{n,m}|^2 \right\} = d_n^\mathrm{H} \mathbb{E}_{\mathsf{v}_m} \left\{ h_m h_m^\mathrm{H} \right\} d_n = d_n^\mathrm{H} \Theta_m d_n, \qquad (39)$$

$$\mathbb{E}_{\mathsf{Z}_\mathrm{p}} \left\{ |\hat{\delta}_{n,m}|^2 \right\} = \frac{1}{MN} \left| \hat{\varLambda}_\mathrm{LS} \right|_\mathrm{F}^2 = \frac{J_\mathrm{LS}}{MN} \quad \because D^\mathrm{H} D = I_N. \qquad (40)$$

Note that the LS estimation error is uniformly superimposed on the estimation value. From (38), the optimal selection (denoising) threshold for each element to minimize the sum MSE is readily given by

$$g_{n,m} = \begin{cases} 1 & \text{if } d_n^\mathrm{H} \Theta_m d_n > \frac{J_\mathrm{LS}}{MN} \\ 0 & \text{otherwise} \end{cases}. \qquad (41)$$

Consequently, the resultant denoised beam-domain channel matrix can be expressed as

$$\hat{\Xi}_\mathrm{D} = G \odot \hat{\Xi}. \qquad (42)$$

The above operation is essentially equivalent to estimating the each UE channel with the set of optimal statistical beams for minimizing the sum MSE of the estimated channel.

As N increases, the beam-domain resolution becomes higher, and the estimation quality is significantly improved

compared to the typical antenna-doamin LS estimation owing to a large number of zero elements in $\hat{\Xi}_\mathrm{D}$.

B. Beam-domain covariance matrix

When the iterative detection and channel decoder are utilized in the subsequent signal detection, an appropriate design of covariance matrices for channel estimation errors is crucial to prevent the degradation of convergence property.

From (37), the beam-domain estimation error matrix of (42) can be expressed as

$$\hat{\varLambda}_\mathrm{D} = \overline{G} \odot \Xi + G \odot \hat{\varLambda}_\Xi, \qquad (43)$$

where we define $\overline{G} = 1 - G$. With the use of Ξ_D, the beam-domain RX symbol matrix of data part can be expressed as

$$R_\mathrm{d} = D^\mathrm{H} Y_\mathrm{d} = \hat{\Xi}_\mathrm{D} X_\mathrm{d} + \hat{\Phi}_\mathrm{d}, \qquad (44)$$

where we define

$$\hat{\Phi}_\mathrm{d} = \hat{\varLambda}_\mathrm{D} X_\mathrm{d} + D^\mathrm{H} Z_\mathrm{d}. \qquad (45)$$

Denoting the k_d-th column vector of $\hat{\Phi}_\mathrm{d}$ by $\hat{\phi}[k_\mathrm{d}]$, the covariance matrix is given by

$$\begin{aligned} \hat{\Sigma} &= \mathbb{E}_{\mathsf{H}, \mathsf{Z}} \left\{ \hat{\phi}[k_\mathrm{d}] \hat{\phi}^\mathrm{H}[k_\mathrm{d}] \right\} = E_\mathrm{s} \cdot \mathbb{E}_{\mathsf{H}, \mathsf{Z}_\mathrm{p}} \left\{ \hat{\varLambda}_\mathrm{D} \hat{\varLambda}_\mathrm{D}^\mathrm{H} \right\} + N_0 I_N \\ &= E_\mathrm{s} \cdot \left[\mathbb{E}_\mathsf{H} \left\{ \left(\overline{G} \odot \Xi \right) \left(\overline{G} \odot \Xi \right)^\mathrm{H} \right\} \right. \\ &\quad + \left. \mathbb{E}_{\mathsf{Z}_\mathrm{p}} \left\{ \left(G \odot \hat{\varLambda}_\Xi \right) \left(G \odot \hat{\varLambda}_\Xi \right)^\mathrm{H} \right\} \right] + N_0 I_N, \end{aligned} \qquad (46)$$

where the cross terms are 0 since $\mathbb{E}_{\mathsf{Z}_\mathrm{p}} \{Z_\mathrm{p}\} = 0$. To avoid the non-linear operation $\odot$, we should focus on the each element of (46). From (35), (39), and (40), the element in the i-th row and the j-th column of $\hat{\Sigma}$ can be expressed as

$$\hat{\sigma}_{i,j} = \begin{cases} \sum_{m=1}^M \overline{g}_{i,m}^2 \cdot d_i^\mathrm{H} \Theta_m d_i \\ \qquad + \frac{J_\mathrm{LS}}{MN} \sum_{m=1}^M g_{i,m}^2 \cdot d_i^\mathrm{H} d_i + N_0 & (i = j) \\ \sum_{m=1}^M \overline{g}_{i,m} \overline{g}_{j,m} \cdot d_i^\mathrm{H} \Theta_m d_j \\ \qquad + \frac{J_\mathrm{LS}}{MN} \sum_{m=1}^M g_{i,m} g_{j,m} \cdot d_i^\mathrm{H} d_j & (i \neq j) \end{cases}$$

$$= \begin{cases} \sum_{m=1}^M \left[\overline{g}_{i,m} \cdot d_i^\mathrm{H} \Theta_m d_i + g_{i,m} \cdot \frac{J_\mathrm{LS}}{MN} \right] + N_0 & (i = j) \\ \sum_{m=1}^M \overline{g}_{i,m} \overline{g}_{j,m} \cdot d_i^\mathrm{H} \Theta_m d_j & (i \neq j) \end{cases}, \qquad (47)$$

where $\overline{g}_{i,j} = 1 - g_{i,j}$. Owing to the orthogonality of the statistical beams, the off-diagonal elements of (47) are negligible small values and much smaller than the diagonal elements. Therefore, the covariance matrix $\hat{\Sigma}$ can be approximated with high-accuracy to the diagonal matrix as

$$\hat{\Sigma} \approx \mathrm{diag} \left[\hat{\sigma}_{1,1}, \ldots, \hat{\sigma}_{n,n}, \ldots, \hat{\sigma}_{N,N} \right]. \qquad (48)$$

From (41), (47) and (48), all the operations required to track changes in the long-term channel statistics is computation of the metric $\{d_n \Theta_m d_n, \forall m, \forall n\}$, whose computational cost is much smaller than that of the matrix operations found in the MAP estimator of (22).

IV. EP-BASED ITERATIVE DETECTION

In this section, we briefly provide an iterative detector in the presence of channel estimation errors based on the EP framework [9], [10]. For convenience, the estimated antenna-domain channel matrices in (22) and (29) are represented by $\hat{H}$. To simplify the notations, $[k_\mathrm{d}]$ is omitted because the same

Algorithm 1 EP-based iterative detector in the antenna domain

---

Input: $\boldsymbol{y}$, $\hat{\boldsymbol{H}} = \left[\hat{\boldsymbol{h}}_1, \ldots, \hat{\boldsymbol{h}}_M \right]$, $\hat{\boldsymbol{\Psi}}$, U

Output: $\beta_1^{(U)}, \ldots, \beta_m^{(U)}, \ldots, \beta_M^{(U)}$ ▷ Input to channel decoder

  Initialization

1: $\check{\boldsymbol{x}}^{(0)} = \left[\check{x}_1^{(0)}, \ldots, \check{x}_m^{(0)}, \ldots, \check{x}_M^{(0)} \right]^{\mathrm{T}} = \boldsymbol{0}$

2: $\check{\boldsymbol{\Phi}}^{(0)} = \mathrm{diag}\left[\check{\phi}_1^{(0)}, \ldots, \check{\phi}_m^{(0)}, \ldots, \check{\phi}_M^{(0)} \right] = E_{\mathrm{s}} \boldsymbol{I}_M$

  Iteration

3: **for** $u = 1$ to U **do**

4: $\quad \boldsymbol{\Psi}^{(u)} = \hat{\boldsymbol{H}} \check{\boldsymbol{\Phi}}^{(u-1)} \hat{\boldsymbol{H}}^{\mathrm{H}} + \hat{\boldsymbol{\Psi}}$

5: $\quad$ **for** $m = 1$ to M **do**

  $\quad\quad$ *Belief Generation (BG)*

6: $\quad\quad \tilde{\boldsymbol{y}}_m^{(u)} = \boldsymbol{y} - \hat{\boldsymbol{H}} \check{\boldsymbol{x}}^{(u-1)} + \hat{\boldsymbol{h}}_m \check{x}_m^{(u-1)}$

7: $\quad\quad \gamma_m^{(u)} = \hat{\boldsymbol{h}}_m^{\mathrm{H}} \left[\boldsymbol{\Psi}^{(u)} \right]^{-1} \hat{\boldsymbol{h}}_m$, $\quad w_m^{(u)} = \frac{4 w_x}{1 - \gamma_m^{(u)} \check{\phi}_m^{(u-1)}}$

8: $\quad\quad \beta_m^{(u)} = w_m^{(u)} \cdot \hat{\boldsymbol{h}}_m^{\mathrm{H}} \left[\boldsymbol{\Psi}^{(u)} \right]^{-1} \tilde{\boldsymbol{y}}_m^{(u)}$, $\quad \omega_m^{(u)} = w_m^{(u)} \cdot \gamma_m^{(u)}$

  $\quad\quad$ *Soft Symbol Generation (soft SG)*

9: $\quad\quad \overline{x}_m^{(u)} = w_x \cdot \left(\tanh\left[\Re\left\{ \frac{\beta_m^{(u)}}{2} \right\} \right] + \mathrm{j} \tanh\left[\Im\left\{ \frac{\beta_m^{(u)}}{2} \right\} \right] \right)$

10: $\quad\quad \overline{\phi}_m^{(u)} = E_{\mathrm{s}} - |\overline{x}_m^{(u)}|^2$

11: $\quad\quad \frac{1}{\check{\phi}_m^{(u)}} = \frac{1}{\overline{\phi}_m^{(u)}} - \frac{1}{4 w_x} \omega_m^{(u)}$

12: $\quad\quad \check{x}_m^{(u)} = \check{\phi}_m^{(u)} \left(\frac{\overline{x}_m^{(u)}}{\overline{\phi}_m^{(u)}} - \frac{1}{4 w_x} \beta_m^{(u)} \right)$

13: $\quad$ **end for**

14: **end for**

---

process is performed at all discrete times ($1 \le k_{\mathrm{d}} \le K_{\mathrm{d}}$). Furthermore, we assume quadrature phase shift keying (QPSK) signaling ($\mathcal{X} = \{\pm w_x \pm \mathrm{j} w_x\}$, $w_x = \sqrt{E_{\mathrm{s}}/2}$).

The pseudo-code of EP-based detector in the antenna domain is provided in Algorithm 1. It consists of two modules: belief generator (BG) and soft symbol generator (soft SG). The BG performs MUD using the linear MMSE filter and computes a log likelihood ratio (LLR) belief. On the other hand, the soft SG computes a soft replica of the TX symbol from the belief using a non-linear activation function. By propagating the beliefs between each iteration process via the modules, the EP-based detector can gradually improve the detection capability. The detailed derivations of Algorithm 1 are omitted due to limitations of space, but it is naturally derived as an extension of [10] with the assistance of the modified covariance matrices and some mathematical manipulations[1].

When using the denoised beam-domain channel presented in Sect. III, the subsequent signal processing is also performed in the beam domain. In this case, $\boldsymbol{y}$, $\hat{\boldsymbol{H}}$, and $\hat{\boldsymbol{\Psi}}$ are respectively replaced with $\boldsymbol{r} = \boldsymbol{D}^{\mathrm{H}} \boldsymbol{y}$, $\hat{\boldsymbol{\Xi}}_{\mathrm{D}}$, and $\hat{\boldsymbol{\Sigma}}$ in Algorithm 1[2].

V. NUMERICAL RESULTS

A. Complexity analysis

To validate the proposed method, we focus on the required computational cost at any time the long-term channel statistics

---

[1] Some simplifications on the premise of the large system limit are removed to improve the convergence property in practical scenarios.

[2] By exploiting the sparse property of beam-domain channel, the computational cost of signal detection can be also reduced [14]. The comprehensive design and evaluation is left for the future work.
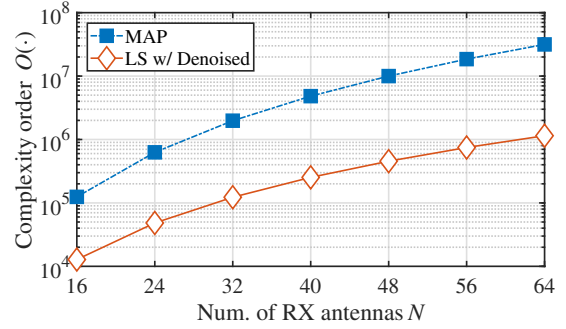

Fig. 1. The complexity order for updating the estimator according to the channel statistics as a function of N (from 16 to 64), whereas the number of UEs M is varied as $5N/8$ (from 10 to 40).

---

is changed. The computational cost becomes dominant at the update timing. Assuming the use of orthogonal pilot sequences, the MAP estimator of (22) can be rewritten as

$$\hat{\boldsymbol{h}} = \mathrm{diag}\left[\left(K_{\mathrm{p}} \boldsymbol{\Theta}_1 + N_0 \boldsymbol{I}_N \right)^{-1} \boldsymbol{\Theta}_1, \ldots \right.$$
$$\left. \ldots, \left(K_{\mathrm{p}} \boldsymbol{\Theta}_M + N_0 \boldsymbol{I}_N \right)^{-1} \boldsymbol{\Theta}_M \right] \boldsymbol{S}^{\mathrm{H}} \boldsymbol{y}_{\mathrm{p}}. \quad (49)$$

When computed in the proper order, the computation of the block diagonal matrix included in (49) is dominant factor for determining the computational complexity. This operation needs to compute $N \times N$ inverse matrix M times, leading to about $O(3MN^3 + M^2 N)$ complexity, in terms of the number of complex number multiplications, which is unacceptable complexity, especially in large-scale MUD.

On the other hand, only the metric computation for determining the selection threshold of (41) is sufficient in the proposed method, and therefore large-scale matrix operations are not necessary. Consequently, the required computational cost can be reduced to $O(MN^2 \log_2 N + MN^2)$, thus, is roughly $\log_2 N / (3N)$ fraction of the cases of optimal estimation. The order of complexities are plotted under some system conditions in Fig. 1. As both M and N increase, the gap between "MAP" and "LS w/ Denoised" gradually spread. Remarkably, at $(N, M) = (64, 40)$, the proposed method performs with about 3% of computational cost of the MAP estimator.

B. MSE performances

Computer simulations were conducted to validate the performance of the proposed beam-domain channel estimation in the EP-based large-scale MUD. The average RX power from each TX antenna is assumed to be identical on the basis of slow TX power control. A sector antenna of 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 realizations. The angular spread for each UE is 15 degrees. The modulation scheme is Gray-coded QPSK. The length of pilot sequence K_{p} is M for maintaining the orthogonality.

Fig. 2 shows MSE performances in $(N, M) = (64, 40)$ MU-MIMO configuration. Compared to the MAP estimator "MAP," the MSE performance of "LS" is significantly degraded, especially in the low SNR region. By contrast, the proposed beam-domain channel estimation "LS w/ Denoised" can approach
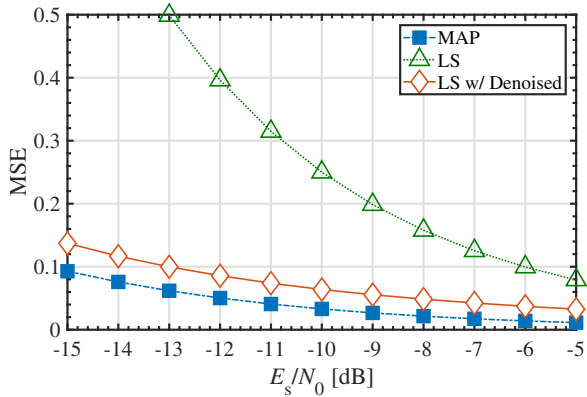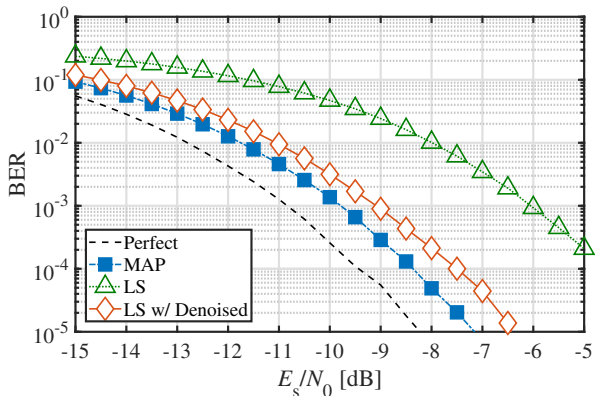
Fig. 2. MSE performances: $(N, M) = (64, 40)$

Fig. 3. BER performances: $(N, M) = (64, 40)$

ideal "MAP" performance with less than 4% complexity of "MAP" using the long term channel statistics for appropriately selecting the angular–UE (beam–UE) bins.

C. BER performance

Let shift our focus to the BER performances in Fig. 3 to evaluate the beam-domain channel estimation including the covariance matrix design used in the subsequent EP-based signal detection. The low-density parity-check (LDPC) code of rate 0.5 and length 1944 bits 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 $U = 8$ iterative detection. "Perfect" indicates the EP-based detection performance with perfect channel estimation as a limit property.

The other three curves show the BER performances in the presence of channel estimation errors. Owing to the appropriate covariance matrix design, the extreme deterioration of convergence property caused by the model error of beliefs does not appear in any case. When using the LS estimator, the performance degradation becomes higher than 4.5 dB at BER $= 10^{-4}$ as compared to the limit property "Perfect". On the other hand, in the MAP estimator, the degradation can be suppressed to about 1.0 dB using the long-term channel statistics as the prior information in exchange for high computational burden. The most attractive feature is that the proposed beam-domain channel estimation "LS w/ Denoised" can approach

"MAP" property with remarkably low complexity, and the performance gap is less than 1.0 dB at BER $= 10^{-4}$.

VI. CONCLUSION

In this paper, we proposed low-complexity beam-domain channel estimation using the long-term channel statistics in large-scale iterative MUD. The optimal MAP estimation requires high-dimensional matrix operations to track changes in the channel statistics, leading to the extremely high computational cost in large MU-MIMO scenarios. To reduce the computational burden, beam-domain channel estimation is exploited. By utilizing the approximate sparse property of beam-domain channel, we select the angular-UE bins so that minimizes the sum MSE of the estimated channel based on the prior information. As both M and N increase, the beneficial effect becomes larger, thus, the proposal has high scalability. To maintain the convergence property of iterative detection, the appropriate beam-domain covariance matrix is also designed. The proposed method can reduce the computational cost to less than 4% with the BER performance degradation of less than 1.0 dB at BER $= 10^{-4}$ compared to the MAP estimation.

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] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

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

[6] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. on Signal Processing*, vol. 62, no. 12, pp. 3261–3271, June 2014.

[7] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexingThe large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct 2013.

[8] 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.

[9] 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.

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

[11] T. P. Minka, "Expectation propagation for approximate Bayesian inference," *Proc. 17th Conf. UAI*, pp. 362–369, Aug. 2001.

[12] W. C. Jakes, *Microwave Mobile Communications, 1st ed.* Wiley-IEEE Press, 1994.

[13] D. Chu, "Polyphase codes with good periodic correlation properties (coresp.)," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, July 1972.

[14] 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.