Downlink Channel Estimation for Massive MIMO Systems Relying on Vector Approximate Message Passing

Sheng Wu, Haipeng Yao, Chunxiao Jiang, Xi Chen, Linling Kuang, and Lajos Hanzo

Abstract—To reduce the pilot overhead of downlink channel estimation in massive multiple-input—multiple-output (MIMO) systems, a sparse recovery algorithm relying on the vector approximate message passing (VAMP) technique is proposed. More specifically, an a-priori channel model characterized by a multivariate Bernoulli-Gaussian distribution is invoked for exploiting the common sparsity of massive MIMO channels, and the VAMP technique is used for jointly estimating the spatially correlated channels. Moreover, the hyperparameters of the a-priori model are learned by invoking the expectation maximization (EM) algorithm. Our numerical results demonstrate that the proposed algorithm is capable of reducing the pilot overhead by 50% in massive MIMO systems.

Index Terms—Block sparsity, Channel estimation, Massive MIMO, OFDM, Vector Approximate Message Passing.

I. Introduction

To maximize the gain of massive multiple-input-multiple-output (MIMO) systems, having accurate channel state information (CSI) at the base station (BS) is essential [1]. In frequency division duplex (FDD) systems, the CSI is typically obtained by downlink (DL) channel estimation and then fed back to the BS [2], [3]. However, the conventional least squares channel estimation method requires a high pilot overhead that increases linearly with both the number of BS antennas and the length of channel impulse response (CIR), and becomes inefficient in massive MIMO systems.

In recent years, compressed sensing (CS) based channel estimation has attracted much attention due to its ability to reduce required pilots. In [4], the sparsity of the massive MIMO CIR taps was exploited by Rao and Lau. As a further advance, Rao and Lau [5] studied CS-based channel estimation in the presence of temporal correlation. Gao *et al.* [1] proposed an adaptive structured subspace pursuit based algorithm for exploiting both the common spatial and temporal sparsity in massive MIMO channels. Then, a distributed CS-based

S. Wu and H. Yao are with School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (E-mail: {thuraya, yaohaipeng}@bupt.edu.cn).

C. Jiang, X. Chen and L. Kuang are with the Tsinghua Space Center, Tsinghua University, Beijing 100084, China (E-mail: {jchx, chenxi_ee, kll}@tsinghua.edu.cn).

L. Hanzo is Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (E-mail: lh@ecs.soton.ac.uk).

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/Noo4558/1, EP/PO34284/1, COALESCE, of the Royal Society's Global Challenges Research Fund Grant as well as of the European Research Council's Advanced Fellow Grant QuantCom.

algorithm was studied by Gong *et al.* [6] for leveraging common sparsity in doubly-selective channels. Furthermore, a turbo-CS algorithm relying on the Markov chain prior was proposed by Chen *et al.* [7] for estimating MIMO channels that are structured sparse in the angular domain. Moreover, many message passing algorithms [8], [9], [10], [11], [12] have been applied to MIMO channel estimation. Mo *et al.* used the generalized AMP (GAMP) algorithm for exploiting the joint sparsity of the mmWave MIMO channel both in the angle and in the delay domain [13]. Huang *et al.* [14] proposed an iterative channel estimation algorithm based on the least square estimation and sparse message passing algorithm invoked for mmWave MIMO systems.

Against this background, we exploit the common sparsity of massive MIMO channels in the delay domain and conceive a channel estimator that intrinsically amalgamates the vector approximate message passing (VAMP) technique of Rangan et al. [12] and the EM algorithm [15] of Neal and Hinton. Specifically, an a-priori channel model characterized by a multivariate Bernoulli-Gaussian distribution is employed for exploiting the common sparsity, and the VAMP relying in a non-separable denoiser is used for jointly estimating the spatially correlated channels. At the same time, the hyperparameters of the model are learned by invoking the expectation maximization (EM) algorithm. Our numerical results demonstrate that the proposed algorithm can reduce the pilot overhead of massive MIMO systems by up to 50%, while maintaining superior performance approaching the idealized bound relying on perfectly known channel supports and hyperparameters.

Throughout the paper, the superscript T denotes the transpose operation, and H represents the conjugate transpose operation. $\|\cdot\|_2$ and $\|\cdot\|_F$ return the Euclidean norm and Frobenius norm of a vector or a matrix, respectively. $\mathcal{N}_{\mathbb{C}}(x,\hat{x},\Sigma)$ represents the Gaussian distribution function of a complex random vector x with mean \hat{x} and covariance matrix Σ . Furthermore, $\langle \cdot \rangle$ is the empirical averaging operation $\langle x \rangle = N^{-1} \sum_{n=1}^{N} x_n$, and I is an identity matrix. Finally, $\text{Tr}(\cdot)$ returns the trace of a matrix, and $\text{Diag}(\cdot)$ creates a diagonal matrix with a vector or get the diagonal elements of a matrix.

II. SYSTEM MODEL

Consider a FDD massive MIMO-OFDM system employing K OFDM subcarriers, where the BS has M antennas, while each user has only one antenna. Let $\boldsymbol{h}_{m} = [h_{m1}, \dots, h_{mL}]^{\mathsf{T}}$ denote the DL CIR taps from the mth transmit antenna (TA)

1

to a user, where h_{ml} denotes the *l*th CIR tap and *L* denotes the length of CIR taps. According to the analysis in [16], h_m . and $h_{m'}$, $\forall m' \neq m$ tend to share an identical support set in the delay domain if $\frac{d_{\text{max}}}{C} \leq \frac{1}{10\text{BW}}$, where d_{max} is the farthest distance between the TAs, C is the speed of light, and BW is the signal bandwidth. All the lth DL CIR taps from the MTAs to the user are denoted by $\boldsymbol{h}_{\cdot l} = [h_{1l}, \dots, h_{Ml}]^{\mathsf{T}}$. Due to having a limited number of CIR-tap clusters, $h_{.l}$ is spatially correlated. Therefore, the Kronecker product correlation model [17][18] is applied for characterizing $h_{\cdot l}$, which is formulated as $h_{\cdot l} = (R_l^{\mathsf{az}} \otimes R_l^{\mathsf{el}})^{\frac{1}{2}} h_l^{\mathsf{iid}}$, where $R_l^{\mathsf{az}} \otimes R_l^{\mathsf{el}}$ denote the Kronecker product of the azimuth correlation matrix R_l^{az} and the elevation correlation matrix R_l^{el} , while $h_l^{\mathsf{iid}} \in \mathbb{C}^{M \times 1}$ obeys the Gaussian distribution $\mathcal{N}_{\mathbb{C}}(h_l^{\mathsf{iid}}; \mathbf{0}, \alpha_l \mathbf{I})$, where α_l is the variance of the lth CIR tap.

The pilot sequences transmitted over different TAs are randomly and independently generated, denoted by $x_m \in$ $\mathbb{C}^{N\times 1}$, $\forall m$, but occupy the same subcarriers indexed by the set \mathcal{P} . Let $y \in \mathbb{C}^{N \times 1}$ denote the samples after discarding the cyclic prefix and the fast Fourier transformation (FFT) at the receiver, then

$$y = \sum_{m=1}^{M} \boldsymbol{\Psi}_{m} \boldsymbol{h}_{m} + \boldsymbol{\varpi} = \sum_{l=1}^{L} \boldsymbol{\Phi}_{l} \boldsymbol{h}_{l} + \boldsymbol{\varpi} = \boldsymbol{\Phi} \boldsymbol{h} + \boldsymbol{\varpi}, \quad (1)$$

where $\Psi_m = \text{diag}\{x_m\}F_{\mathcal{P}} \in \mathbb{C}^{N \times L}, F_{\mathcal{P}} \in \mathbb{C}^{N \times L}$ is comprised of the first L columns of the K-point discrete Fourier transformation (DFT) matrix and the N rows of the K-point DFT matrix specified by the indices in set \mathcal{P} , $\boldsymbol{\varpi} = [\boldsymbol{\varpi}_1 \cdots \boldsymbol{\varpi}_{ML}]^\mathsf{T} \in$ $\mathbb{C}^{ML \times 1}$ is the complex Gaussian noise obeying the distribution $p(\boldsymbol{\varpi}) \sim \mathcal{N}_{\mathbb{C}}(0, \sigma \boldsymbol{I}), \, \boldsymbol{\Phi}_{l} = [\boldsymbol{\varphi}_{1l} \cdots \boldsymbol{\varphi}_{Ml}] \in \mathbb{C}^{N \times M} \text{ stacks } \boldsymbol{\varphi}_{ml}$ being the *l*th column of matrix $\boldsymbol{\Psi}_m$, $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1 \cdots \boldsymbol{\Phi}_L] \in \mathbb{C}^{N \times ML}$, and $\boldsymbol{h} = [\boldsymbol{h}_{\cdot 1}^\mathsf{T} \cdots \boldsymbol{h}_{\cdot L}^\mathsf{T}]^\mathsf{T} \in \mathbb{C}^{ML \times 1}$.

III. Sparse Channel Estimation Relying on VAMP

To characterize the common sparsity of h, i.e., all the CIR taps in $h_{\cdot l}$, $\forall l$ are either zero simultaneously or non-zero simultaneously, a multivariate Bernoulli-Gaussian a-priori model [19] is employed:

$$p(\boldsymbol{h};\boldsymbol{\theta}_{\mathsf{p}}) = \prod_{l=1}^{L} \left[\lambda_{l} \mathcal{N}_{\mathbb{C}}(\boldsymbol{h}_{\cdot l};\boldsymbol{0}, \tau_{l} \boldsymbol{I}) + (1 - \lambda_{l}) \delta(\boldsymbol{h}_{\cdot l}) \right], \quad (2)$$

where $\theta_p \triangleq [\lambda_1, \dots, \lambda_L, \tau_1, \dots, \tau_L]^\mathsf{T}$ contains the hyperparameters of the *a-priori* channel model, $\lambda_l \in (0, 1)$ denotes the *a-priori* sparsity ratio of $h_{\cdot l}$, τ_l denotes the *a-priori* variance of every CIR tap in $\boldsymbol{h}_{\cdot l}$ when $\boldsymbol{h}_{\cdot l}$ are non-zero, and $\delta(\cdot)$ denotes the Dirac delta function. Note that as the spatial correlation matrix $\mathbf{R}_{l}^{\mathsf{az}} \otimes \mathbf{R}_{l}^{\mathsf{el}}$ is different across tap index l, modeling each $h_{\cdot l}$ with a different a-priori covariance matrix in (2) would lead to overfitting, and modeling all the $h_{\cdot l}$ by a common a-priori covariance matrix would not yield any performance gain. In view of this, the proposed prior model (2) does not capture the amplitude correlation among the taps in $h_{\cdot l}$. As the random noise ϖ in (1) is Gaussian, the likelihood function of h is written as

$$p(\mathbf{y} \mid \mathbf{h}; \sigma) = \mathcal{N}_{\mathbb{C}}(\mathbf{y}; \boldsymbol{\Phi} \mathbf{h}, \sigma \mathbf{I}). \tag{3}$$

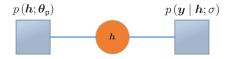


Figure 1: Factor graph for DL channel estimation.

For the channel estimation, the a-posteriori distribution of hcan be approximately calculated by the VAMP [12], and then we can get the *a-posteriori* mean of h as its estimate. However, the VAMP desires the specification of the hyperparameters in the a-priori model (2) and the likelihood function (3), i.e., $\theta \triangleq [\theta_{\rm p}, \sigma]^{\rm I}$. To this end, the EM algorithm is used to get the specification of hyperparameters,

$$Q(\theta) \triangleq \mathsf{E}\left\{\mathsf{ln}p(y \mid \boldsymbol{h}; \sigma) + \mathsf{ln}p(\boldsymbol{h}; \theta_{\mathsf{p}})\right\},\tag{4}$$

$$\theta(i) = \underset{\theta}{\operatorname{argmax}} Q(\theta), \tag{5}$$

where the expectation in (4) is w.r.t the a-posteriori distribution of h (as shown in the following (12)) that is approximately calculated by the VAMP, and the specification of hyperparameters $\theta(i)$ is updated in the maximization step as shown by (5).

Fig. 1 shows the factor graph (for more details on factor graph, please refer to [20]) corresponding to the factorization $p(y, h; \theta_p, \sigma) = p(h; \theta_p)p(y \mid h; \sigma)$, which is comprised of the only variable node h, and two function nodes, namely $p(h; \theta_D)$ and $p(y \mid h; \sigma)$. Upon invoking the VAMP and the specification of the hyperparameters, the a-posteriori probability of h at the variable node h at the ith iteration is given

$$\beta_{\boldsymbol{h}}(i,\boldsymbol{h}) \triangleq \frac{\mathcal{N}_{\mathbb{C}}(\boldsymbol{h},\boldsymbol{r}(i),\gamma_{r}^{-1}(i)\boldsymbol{I}) \prod_{l=1}^{L} p(\boldsymbol{h}_{\cdot l};\boldsymbol{\theta}_{\mathsf{p}}(i-1))}{\int_{\boldsymbol{h}} \mathcal{N}_{\mathbb{C}}(\boldsymbol{h},\boldsymbol{r}(i),\gamma_{r}^{-1}(i)\boldsymbol{I}) \prod_{l=1}^{L} p(\boldsymbol{h}_{\cdot l};\boldsymbol{\theta}_{\mathsf{p}}(i-1))}$$

$$= \prod_{l=1}^{L} (1 - \pi_l(i)) \delta(\boldsymbol{h}_{\cdot l}) + \pi_l(i) \mathcal{N}_{\mathbb{C}}(\boldsymbol{h}_{\cdot l}; \boldsymbol{\mu}_l(i), \boldsymbol{\Sigma}_l(i)), \tag{6}$$

where we have

$$\mu_l(i) = \frac{\tau_l(i)\gamma_r(i)}{1 + \tau_l(i)\gamma_r(i)} \mathbf{r}_l(i), \tag{7}$$

$$\Sigma_l(i) = \frac{\tau_l(i)}{1 + \tau_l(i)\gamma_r(i)} I, \tag{8}$$

$$\Sigma_{l}(i) = \frac{\tau_{l}(i)}{1 + \tau_{l}(i)\gamma_{r}(i)} \boldsymbol{I},$$

$$\pi_{l}(i) = \frac{\lambda_{l}(i)}{\lambda_{l}(i) + (1 - \lambda_{l}(i)) \frac{N_{\mathbb{C}}(\boldsymbol{r}_{l}(i);\boldsymbol{0},\boldsymbol{\gamma}_{r}^{-1}(i)\boldsymbol{I})}{N_{\mathbb{C}}(\boldsymbol{r}_{l}(i);\boldsymbol{0},(\boldsymbol{\tau}_{l}(i)+\boldsymbol{\gamma}_{r}^{-1}(i))\boldsymbol{I})}},$$
(9)

in terms of $r(i) \triangleq [r_1^{\mathsf{T}}(i), \dots, r_L^{\mathsf{T}}(i)]^{\mathsf{T}} \in \mathbb{C}^{ML \times 1}, \ r_l(i) \triangleq [r_{(l-1)M+1}(i), \dots, r_{lM}(i)]^{\mathsf{T}} \in \mathbb{C}^{M \times 1}, \ \text{and} \ \gamma_r(i) \ \text{defined in Tab.}$ I. Then, the *a-posteriori* mean and covariance matrix of $h_{\cdot 1}$ w.r.t (6) are given by

$$\hat{\boldsymbol{h}}_{\cdot l}(i) = \pi_l(i)\boldsymbol{\mu}_l(i),\tag{10}$$

$$C_{h,l}(i) = \pi_l(i)\Sigma_l(i) + \pi_l(i)(1 - \pi_l(i))\mu_l(i)\mu_l^{\mathsf{H}}(i).$$
 (11)

The operations of (6)-(11) show that our proposed algorithm uses an inseparable denoiser, which is different from the separable denoiser studied by most VAMP.

Then, the message emanating from the variable node \boldsymbol{h} to the function node $p(\boldsymbol{y} \mid \boldsymbol{h}; \sigma)$ is written as $\mathcal{N}_{\mathbb{C}}(h;q(i),\gamma_q^{-1}(i)I)$ by the principle of expectation propagation (EP) [12], where q(i) and $\gamma_q(i)$ are also defined in Tab. I. At the function node $p(y \mid h; \sigma)$ on the factor graph, the *a-posteriori* probability of h is given by

$$\beta_{\mathbf{y}}(i, \boldsymbol{h}) = \mathcal{N}_{\mathbb{C}}(\boldsymbol{h}; \boldsymbol{q}(i), \gamma_{q}^{-1}(i)\boldsymbol{I})p(\boldsymbol{y} \mid \boldsymbol{h}; \sigma(i-1))$$

$$\propto \mathcal{N}_{\mathbb{C}}(\boldsymbol{h}; \hat{\boldsymbol{h}}_{\mathbf{y}}(i), \boldsymbol{V}_{\mathbf{y}}(i)), \qquad (12)$$

$$\hat{\boldsymbol{h}}_{\mathbf{y}}(i) = \boldsymbol{V}_{\mathbf{y}}(i)(\sigma^{-1}(i-1)\boldsymbol{\Phi}^{\mathsf{H}}\boldsymbol{y} + \gamma_{q}(i)\boldsymbol{q}(i)), \qquad (13)$$

$$\boldsymbol{V}_{\mathbf{y}}(i) = \gamma_{q}^{-1}(i)\left(\boldsymbol{I} - \overline{\boldsymbol{V}}\mathsf{Diag}\left(\frac{|\overline{\boldsymbol{s}}|^{2}}{|\overline{\boldsymbol{s}}|^{2} + \sigma(i-1)\gamma_{q}(i)}\right)\overline{\boldsymbol{V}}^{\mathsf{H}}\right), \qquad (14)$$

where the last line of (14) is obtained by the singular value decomposition (SVD) of $\boldsymbol{\Phi}$, i.e., $\boldsymbol{\Phi} = \overline{\boldsymbol{U}} \operatorname{Diag}(\overline{\boldsymbol{s}}) \, \overline{\boldsymbol{V}}^{\mathsf{H}}$ and $\boldsymbol{\Phi}^{\mathsf{H}} \boldsymbol{\Phi} = \overline{\boldsymbol{V}} \operatorname{Diag}\left(|\overline{\boldsymbol{s}}|^2\right) \, \overline{\boldsymbol{V}}^{\mathsf{H}}$.

A. EM Learning of Hyperparameters

As mentioned above, the hyperparameters used for the *a-priori* model (2) and likelihood function (3) are learned by the EM. However, it is a challenge to perform the maximization jointly as formulated in (5). Therefore, we maximize these hyperparameters one by one [15].

Following the approach developed by Vila and Schniter [21], we have $\partial Q(\theta)/\partial \sigma = \sigma^{-2}||y - \Phi \hat{h}_f(i)||_2^2 + \sigma^{-2}\text{Tr}(\Phi V_f(i)\Phi^H) - \sigma^{-1}ML$, where $Q(\theta)$ is calculated by (4) and (12). Then, to make the derivative $\partial Q(\theta)/\partial \sigma$ zero, the *a-priori* variance of noise is updated by

$$\sigma(i) = N^{-1} \| \mathbf{y} - \mathbf{\Phi} \hat{\mathbf{h}}_f(i) \|_2^2 + N^{-1} \sigma(i-1) \operatorname{Sum} \left(\frac{|\overline{\mathbf{s}}|^2}{|\overline{\mathbf{s}}|^2 + \sigma(i-1)\gamma_q(i)} \right). \tag{15}$$

Similarly, the derivation $\partial Q(\theta)/\partial \lambda_l$ is given by $Q(\theta)/\partial \lambda_l = (\pi_l(i) - \lambda_l(i))/\lambda_l(i)/(1 - \lambda_l(i))$, and then the sparsity ratio is updated by

$$\lambda_l(i) = \pi_l(i). \tag{16}$$

Finally, the derivation of $Q(\theta)$ w.r.t the *a-priori* variance τ_l leads to $\partial Q(\theta)/\partial \tau_l = \mu_l^H(i)\mu_l(i)/\tau_l^2 + M\tau_l(i-1)/\tau_l^2/(1+\tau_l(i-1)\gamma_r(i)) + M/\tau_l$. Then, τ_l is updated by

$$\tau_l(i) = \frac{\tau_l(i-1)}{1 + \gamma_r(i)\tau_l(i-1)} + \frac{\mu_l^{\mathsf{H}}(i)\mu_l(i)}{M}.$$
 (17)

The proposed channel estimation algorithm is termed as "EM-BBG-VAMP", as shown by Tab. I. At the beginning of iterations, the *a-priori* variance of noise is guessed to be $\sigma(0) = \|\mathbf{y}\|_2^2 / N/(\mathsf{SNR}^0 + 1)$, where the signal-to-noise ratio (SNR) SNR⁰ is empirically set to 0 dB. Following the suggestion given by [21], the initial sparsity ratio could be set to

$$\lambda_{I}(0) = \frac{N}{ML} \max_{a>0} \frac{1 - 2\frac{N}{ML} \left[(1 + a^{2})B(-a) - b(a) \right]}{1 + a^{2} - 2 \left[(1 + a^{2})B(-a) - b(a) \right]}, \quad (18)$$

where $B(\cdot)$ and $b(\cdot)$ denote the cumulative distribution function and the probability density function of the standard Gaussian distribution, respectively. The *a-priori* variance τ_l is set to $\tau_l(0) = (\|\mathbf{y}\|_2^2 - N\sigma(0))/\lambda_l(0)/\|\boldsymbol{\Phi}\|_F^2, \forall l$. After I_{max}

Table I: The proposed channel estimation algorithm.

```
\begin{aligned} &\text{for } i=1,\ldots,I_{\text{max}}\,\text{do} \\ &\mu_l(i)=\frac{\tau_l(i)\gamma_r(i)}{1+\tau_l(i)\gamma_r(i)}r_l(i);\, \boldsymbol{\varSigma}_l(i)=\frac{\tau_l(i)}{1+\tau_l(i)\gamma_r(i)}\boldsymbol{I};\\ &\hat{\boldsymbol{h}}_{\cdot l}(i)=\pi_l(i)\boldsymbol{\mu}_l(i);\\ &\pi_l(i)=\frac{\lambda_l(i)}{\lambda_l(i)+(1-\lambda_l(i))}\frac{\lambda_{\mathbb{C}}(r_l(i);0,\gamma_r^{-1}(i)\boldsymbol{I})}{\lambda_{\mathbb{C}}(r_l(i);0,(\tau_l(i)+\gamma_r^{-1}(i))\boldsymbol{I})};\\ &\nu_l(i)=\left\langle \text{Diag}(\pi_l(i)\boldsymbol{\varSigma}_l(i)+\pi_l(i)(1-\pi_l(i))\boldsymbol{\mu}_l(i)\boldsymbol{\mu}_l^{\text{H}}(i))\right\rangle;\\ &\eta_1(i)=L\left\langle \sum_{l=1}^L\nu_l(i);\gamma_q(i)=\eta_1(i)-\gamma_r(i);\\ &q(i)=\left(\eta_1(i)\hat{\boldsymbol{h}}(i)-\gamma_r(i)r(i)\right)\middle/\gamma_q(i);\\ &V_y(i)=\sigma(i-1)\boldsymbol{U}^{\text{H}}\left(\boldsymbol{\Lambda}+\sigma(i-1)\gamma_q(i)\right)^{-1}\boldsymbol{U};\\ &\hat{\boldsymbol{h}}_y(i)=\boldsymbol{V}_y(i)\left(\sigma^{-1}(i-1)\boldsymbol{\Phi}^{\text{H}}\mathbf{y}+\gamma_q(i)\mathbf{q}(i)\right);\\ &\nu_2(i)=\left\langle \text{Diag}(\boldsymbol{V}_y(i))\right\rangle;\,\eta_2(i)=1\middle/\nu_2(i);\\ &\gamma_r(i)=\eta_2(i)-\gamma_q(i);\,r(i)=\left(\eta_2(i)\hat{\boldsymbol{h}}_y(i)-\gamma_q(i)\mathbf{q}\right)\middle/\gamma_r(i);\\ &\text{Update the sparse ratio }\lambda_l(i)\text{ by }(16)\text{, the noise power }\sigma(i)\text{ by }(15),\\ &\text{end for }\end{aligned}
```

Table II: Complexity comparison.

Algorithm	Number of floating point operations
SP	$O(MNL + MNN_s^2 + MNN_s^3)$
DSAMP	$O(MNL + MNi_s^2 + MNi_s^3)$
EM-BBG-EP	$O(M^3L + MNL)$
EM-BG-GAMP	O(MNL)
EM-BBG-VAMP	O(MNL)

iterations, the algorithm gives $\hat{\boldsymbol{h}}_f(I_{\text{max}})$ as the *a-posteriori* estimate of the MIMO channel. Tab. II compares the complexity defined by the number of floating point operations for the proposed EM-BBG-VAMP, for the subspace pursuit (SP) [22], for the distributed sparsity based adaptive matching pursuit (DSAMP) of Gao *et al.* [23], for the EM-BBG-EP [19], and for the EM-BG-GAMP [21], where $N_{\rm S}$ is the number of supports of \boldsymbol{h}_m , and $i_{\rm S} \leq N_{\rm S}$ is the index of the current stage. Note that since the SVD only has to be executed offline once, the complexity of SVD is not included in Tab. II.

IV. NUMERICAL SIMULATIONS

Consider a FDD massive MIMO-OFDM system having M BS TAs and K = 4096 OFDM subcarriers. The spatially correlated 3D channel model proposed in [18] is employed, and the detailed setting of the model parameters is identical to that of [19], which is omitted here owing to the pagelimit. Moreover, the idealized Oracle-LMMSE relying on perfectly known supports and perfectly known noise variance is considered as a benchmark to provide the best performance bound.

Fig. 2(a) shows the relationship between the normalized mean square error (NMSE) of channel estimation and the SNR, when M = 64 and the pilot overhead is N = 1024. It is shown that the NMSE of EM-BBG-VAMP may match the oracle-bound provided by Oracle-LMMSE, when the SNR varies between 5 dB and 35 dB. On the other hand, the EM-BBG-EP joins the oracle-bound at SNR = 10 dB, while the EM-BG-GAMP requires 32.5 dB for approaching the oracle-bound. When the target of NMSE is -15 dB, the EM-BBG-VAMP gets about 12.5 dB gain over the EM-BG-GAMP. Fig. 2(b) gives the results for the case of N = 1024 and M = 128,

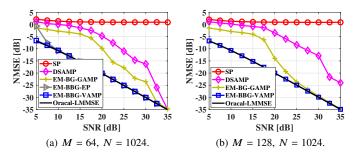


Figure 2: The NMSE versus SNR.

where the performance of EM-BBG-EP has not been evaluated due to its excessive complexity. We can find that the proposed EM-BBG-VAMP still matches the oracle-bound, considerably outperforming the SP, the DSAMP, and the EM-BG-GAMP.

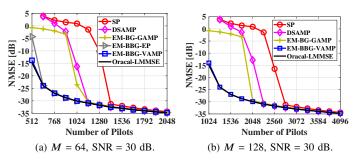


Figure 3: The NMSE versus pilot overhead.

Fig. 3(a) shows the NMSE versus the pilot overhead for M = 64 and SNR = 30 dB. The EM-BBG-VAMP matches the oracle-bound even when the pilot overhead is as low as N = 512, while the EM-BBG-EP merges with the oracle-bound at N = 640. Both the DSAMP and EM-BG-GAMP merge with the oracle-bound at N = 1152, and the SP at N = 1408. As shown in Fig. 3(b), the required pilot overhead for EM-BBG-VAMP to get the targeted NMSE = -15dB is about 50% of that of the DSAMP and EM-BG-GAMP.

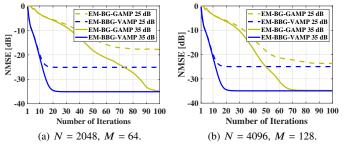


Figure 4: The NMSE versus iteration times.

Fig. 4(a) shows the performance of EM-BBG-VAMP and EM-BG-AMP with different iteration times in the case of N = 2048 and M = 64. When SNR = 25 dB, the number of iterations that the EM-BBG-VAMP requires to converge is less than 20, which is the same as that of the case of SNR = 35 dB. However, the EM-BG-AMP requires more than 60 iterations to converge at SNR = 25 dB, and more than 90 iterations at

SNR = 35 dB. In the case of N = 4096 and M = 128, as shown by Fig. 4(b), the EM-BBG-VAMP requires less than 20 iterations to converge, while other algorithms require more than 60 iterations to converge.

V. CONCLUSION

We have proposed an EM-BBG-VAMP algorithm relying on the vector message passing technique for the DL channel estimation of massive MIMO systems. The proposed algorithm is capable of reducing the number of pilots in the massive MIMO system by about 50%, compared to the SP algorithm, the EM-BG-GAMP algorithm, and the DSAMP algorithm.

REFERENCES

- Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, Feb. 2016.
- [2] W. Shen, L. Dai, Y. Shi, B. Shim, and Z. Wang, "Joint channel training and feedback for FDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8762–8767, Oct. 2016.
- [3] W. Xu, W. Xiang, Y. Jia, Y. Li, and Y. Yang, "Downlink performance of massive-MIMO systems using EVD-based channel estimation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3045–3058, Apr. 2017.
- [4] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.
- [5] —, "Compressive sensing with prior support quality information and application to massive MIMO channel estimation with temporal correlation," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4914– 4924, Sep. 2015.
- [6] B. Gong, L. Gui, Q. Qin, X. Ren, and W. Chen, "Block distributed compressive sensing-based doubly selective channel estimation and pilot design for large-scale MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9149–9161, Oct. 2017.
- [7] L. Chen, A. Liu, and X. Yuan, "Structured turbo compressed sensing for massive MIMO channel estimation using a Markov prior," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4635–4639, May 2018.
- [8] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
- [9] Q. Guo and J. Xi, "Approximate message passing with unitary transformation," arXiv preprint arXiv:1504.04799, 2015.
- [10] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.
- [11] L. Liu, C. Huang, Y. Chi, C. Yuen, Y. L. Guan, and Y. Li, "Sparse vector recovery: Bernoulli-Gaussian message passing," in *Proc. IEEE Global Telecomm. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.
- [12] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 1588–1592.
- [13] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, March 2018.
- [14] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmwave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Jan. 2019.
- [15] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in graphical models*. Springer, 1998, pp. 355–368.
- [16] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 104–118, Jan. 2015.
- [17] L. Schumacher, K. Pedersen, and P. Mogensen, "From antenna spacings to theoretical capacities - guidelines for simulating MIMO systems," in *Proc. IEEE Int. Symp. PIMRC*, Sept. 2002, pp. 587–592.
- [18] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Proc. IEEE Int. Conf. Communications (ICC)*, Sydney, NSW, Jun. 2014, pp. 5865–5870.
- [19] S. Wu, Z. Ni, X. Meng, and L. Kuang, "Block expectation propagation for downlink channel estimation in massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2225–2228, Nov. 2016.

- [20] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," IEEE Trans. Inf. Theory, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.
 W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2240, May 2000.
- 2249, May 2009. [23] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based
- adaptive channel estimation and feedback for FDD massive MIMO," IEEE Trans. Signal Process., vol. 63, no. 23, pp. 6169-6183, Dec. 2015.