Collapsed VBI-DP Based Structured Sparse Channel Estimation Algorithm for Massive MIMO-OFDM

Lu, Xinhua; Manchón, Carles Navarro; Wang, Zhongyong

Published in:
IEEE Access

DOI (link to publication from Publisher):
10.1109/ACCESS.2019.2896125

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 29, 2018, accepted January 9, 2019, date of publication January 30, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896125

Collapsed VBI-DP Based Structured Sparse Channel Estimation Algorithm for Massive MIMO-OFDM

XINHUA LU[1,2], (Student Member, IEEE), CARLES NAVARRO MANCHÓN[3], AND ZHONGYONG WANG[1]

[1]School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
[2]College of Computer and Information Engineering, Nanyang Institute of Technology, Nanyang 473004, China
[3]Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark

Corresponding author: Zhongyong Wang (iezywang@sina.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61571402, in part by the Outstanding Young Talent Research Fund of Zhengzhou University under Grant 1521318003, and in part by the Major Projects in Henan Province under Grant 161100210200.

ABSTRACT Massive multiple input multiple output (MIMO) technology significantly improves the capacity of wireless communication systems by deploying hundreds of antennas at the base station. However, the large scale of the array implies higher computational complexity and pilot overhead when implementing channel estimation in the uplink. Utilizing the sparse channel structure is a promising approach to improve the channel estimation performance while circumventing such problems. In this paper, we investigate the detailed physical structure in the delay-spatial domain of uplink channels in massive MIMO-orthogonal frequency division multiplexing (MIMO-OFDM) systems and construct a hierarchical probabilistic model based on Dirichlet process (DP) prior to match the channel's structural sparse features. Based on the model, we derive a structured sparse channel estimation algorithm by implementing collapsed variational Bayesian inference (CVBI). The simulation results demonstrate that the proposed CVBI-DP algorithm can improve channel estimation performance significantly compared with the state-of-the-art methods for massive MIMO-OFDM, without increasing the computational complexity and pilot overhead.

INDEX TERMS Massive MIMO, structured sparse channel, Dirichlet process, collapsed variational Bayesian inference.

I. INTRODUCTION

Massive MIMO has the potential to significantly improve the capacity, spectral and power efficiency of cellular communication systems by deploying a massive number of antennas at the base-station [1]. The system can serve users with high data rates by exploiting the spatial degrees of freedom (DoF) of the resulting large-dimensional channel. However, few channel estimation methods take advantage of DoFs from the spatial structure of such radio channels. In addition, massive MIMO may introduce bigger bandwidth for users and higher frequency carriers, which makes the channels more complex than in traditional systems. Channel measurements and modeling [2], [3] already show there are new properties in massive MIMO channels. An important one is the fact that the channel is more sparse under higher carrier frequency [2], and this makes using sparse channel estimation methods a viable avenue to reduce the pilot overhead in massive MIMO. Another important aspect is the non-stationarity of the response over the large antenna array [3], which means the non-zero elements (support) of the channel impulse response vector may shift or disappear over the array depending on several system parameters and the propagation environment. Meanwhile, the drastic increase in the scale of antenna array complicates the channel estimation procedure, as higher computational complexity and pilot overhead is required [4]. Finding ways to exploit the spatial DoFs of the channel to improve estimation performance without increasing computational complexity and pilot overhead become, thus, an important issue to push massive MIMO technology closer to reality.

In most of the work done in the field of channel estimation for massive MIMO, the methods are mere extensions of traditional MIMO techniques, while few of the methods exploit the channel's structure. We can classify the state-of-

The associate editor coordinating the review of this manuscript and approving it for publication was Bo Li.

art methods into three main approaches. The first is an extension of the methods in traditional wireless communication systems, which estimate the channel independently for each antenna element [5] without accounting for their similarities. The second approach–and also the most popular–uses the channels spatial structure by assuming the channel impulse response (CIR) vectors have the same support, such that the support is estimated jointly for all array elements [6]–[8]. This approach, however, ignores the potential differences in the support of the channels for different array elements. The third approach consists in assuming that the channel support pattern is similar in a neighboring region [9], [10], and estimate the channel independently first. Then, the different antenna ports exchange information about their respective supports with the neighboring elements, and redo the estimation independently again by averaging the messages from neighbors to improve the accuracy.

In all the aforementioned methods, the assumption that the response of the channel at each antenna element is sparse in the delay domain is exploited. In this work, we leverage as well this assumption but from a different perspective. We first analyze the received signals over the antenna array in a massive MIMO system, and provide an uplink channel structure description in the delay-spatial domain, describing the non-stationary behavior of the response. In particular, we note that for very large arrays–as found in Massive MIMO systems– a given multi-path component may be received with significantly different propagation delay at antenna elements that are far apart in the array. The model shows that the support of the channel impulse response varies slightly from one array element to the next, resulting in channel responses that slightly shift their delay support across the elements of the array.

Inspired by the machine learning community, we propose to apply tools for estimation of clustered sparse signals in order to exploit the similarity of the support across elements of the array. Specifically we apply the Dirichlet Process (DP) Prior([11], [12]) to probabilistically model the structure of uplink channels in massive MIMO systems under Bayesian formalism, and apply variational Bayesian inference [13] in order to estimate the channel responses. Inspired by [14], we implemented a collapsed VBI (CVBI) inference method to improve estimation performance, which is based on a marginalization of one of the variables from the model. Such marginalization leads to a loss of the conjugate relationship of the variables in neighboring layers, which we circumvent by an approximation based on a Taylor expansion that leads to tractable inference. In addition–and differently to [14]– we propose to infer as well the concentration parameter of the DP-based model–which influences the number of clusters estimated from the observed data [15]– leading to more accurate results. In the experiments, we verify that the proposed structured channel estimation algorithm improves the performance compared to state-of-art methods without increasing the computational complexity or pilot overhead, and while being robust against the system parameters.

The notational convention for the rest of the paper is as follows: the superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transposition and conjugate transposition operator respectively. We use the lowercase and capital letters in bold to denote vectors and matrices respectively. The symbol $x \mid a, b \sim \mathrm{f}(a, b)$ indicates that a random variable x follows a distribution with parameters a and b. Diag$(\cdot)$ denotes a diagonal matrix with its vector argument on the main diagonal, and trace$(\cdot)$ denotes the trace operator. The symbol $\cdot \propto \cdot$ indicates proportionality. The expectation operation with respect to a function $q(x)$ is represented by $\langle\cdot\rangle_{q(x)}$. The most updated estimate of a variable, vector or matrix is denoted by $\hat{(\cdot)}$.

II. SYSTEM AND CHANNEL MODEL

We consider the uplink channel estimation problem in a single cell of a massive MIMO-OFDM system where the base-station is equipped with a uniform linear array (ULA) of R receiver elements and jointly serves many single-antenna users.

A. UP-LINK OF MASSIVE MIMO-OFDM SYSTEM

Without loss of generality, we use a generic system model inspired by that of [9]. The frequency domain signal received by r-th antenna in the array at the BS after removing the CP and multiplying by a DFT matrix can be described as

$$\begin{aligned} \boldsymbol{y}_r &= \mathrm{Diag}(\boldsymbol{x})\boldsymbol{F}\begin{pmatrix}\boldsymbol{h}_r \\ \boldsymbol{0}_{N-L}\end{pmatrix} + \boldsymbol{w}_r \\ &= \mathrm{Diag}(\boldsymbol{x})\underline{\boldsymbol{F}}\boldsymbol{h}_r + \boldsymbol{w}_r = \boldsymbol{A}\boldsymbol{h}_r + \boldsymbol{w}_r \end{aligned} \tag{1}$$

where $\boldsymbol{y}_r \in \mathbb{C}^N$ is the received signal vector, $\boldsymbol{x} \in \mathbb{C}^N$ are the transmitted signals or pilots, N is the total number of subcarriers, $\underline{\boldsymbol{F}}$ is the truncated Fourier matrix of size $N \times L$ formed by selecting the first L columns of an $N \times N$ DFT matrix $\boldsymbol{F}$, L is the length of the discrete channel impulse response vector which can include a maximum number of significant taps, $\boldsymbol{F}$ is an $N \times N$ DFT matrix whose (k, l) entry is given by $f_{k,l} = \exp\{-j\frac{2\pi}{N}kl\}$, $\boldsymbol{h}_r \in \mathbb{C}^L$ is the discrete channel impulse response vector from the user antenna to the r-th antenna in the array, and $\boldsymbol{w}_r \in \mathbb{C}^N$ is a complex, circularly-symmetric white Gaussian noise vector.

B. STRUCTURAL PHYSICAL CHANNEL MODEL

The multi-path channel impulse response from user antenna to the first antenna in the ULA at the base-station reads

$$h_1(\tau) = \sum_{i=1}^{I_P} \beta_{1,i} e^{-2\pi f_c \tau_{1,i}} \delta(\tau - \tau_{1,i}) \tag{2}$$

where $\beta_{1,i}$ and $\tau_{1,i}$ are the amplitude and delay respectively of the i-th multi-path component received at the 1st array element, I_P is the total number of multi-path components, and f_c is the carrier frequency, as illustrated in Fig. 1. As shown in the figure, when one wave impinges the antenna array,

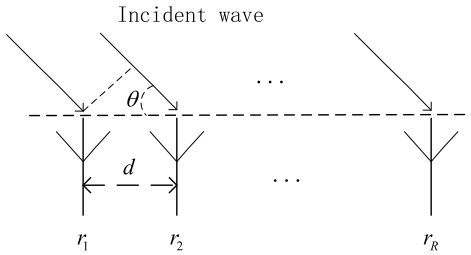


FIGURE 1. Incident wave received by the different elements in a ULA.

the gains and arrival times of the wave received are different over the array according to the different propagation distances from the user antenna to the receiving antennas. Then the multi-path channel impulse response from the user antenna to the r-th antenna in the array is

$$h_r(\tau) = \sum_{i=1}^{I_P} \beta_{r,i} e^{-2\pi f_c \frac{d_{r,i}}{C}} \delta(\tau - \frac{d_{r,i}}{C}) \tag{3}$$

where $d_{r,i}$ is the distance which the i-th wave experienced from user antenna to the r-th antenna in the array, and C is the speed of light. We assume the distance between two neighboring antennas is $\Delta d = \lambda/2$, where λ is the wavelength. Then, $d_{r,i}$ can be approximated as $d_{r,i} \approx d_{0,i} + (r-1)\Delta d \cos\theta_i$ [16], where $d_{0,i}$ is the distance of i-th path between the user antenna to the 1st antenna. With this, the propagation delay of the ith component at the r-th antenna can also be approximated as $\tau_{r,i} \approx d_{0,i}/C + (r-1)\Delta d \cos\theta_i / C$. Considering the amplitude $\beta_{r,i} \propto 1/d_{r,i}^2$ changes slightly over the array given that the distance is changed by $(r-1)\Delta d \cos\theta_i$, we ignore the differences on the i-th path gain and instead fix it as an invariant value β_i. Then the channel spatial signature induced on the antenna array by the i-th path can be described with a steering vector shape:

$$\begin{aligned} \boldsymbol{h}(\tau_i) &= [h_1(\tau_i), \ldots, h_R(\tau_i)] \\ &= \beta_i e^{-2\pi f_c \frac{d_{0,i}}{C}} [1, \cdots, e^{-j2\pi f_c \frac{(R-1)\Delta d \cos\theta_i}{C}}]. \end{aligned} \tag{4}$$

Consider now the discrete-time equivalent channel model, where we denote $h_r(l)$ as the l-th complex channel filter tap of $h_r(\tau)$. Its value is a function of mainly the gains $\beta_i e^{-2\pi f_c \tau_{r,i}}$ of the paths, whose delays $\tau_{r,i}$ are close to l/W. Then the l-th channel tap in the discrete-time equivalent channel of $\boldsymbol{h}_r(\tau)$ can be expressed as

$$h_r(l) = \sum_{i=1}^{I_P} \beta_i e^{-2\pi f_c \tau_{r,i}} \mathrm{sinc}[l - \tau_{r,i} W] \quad l = 1, \ldots, L \tag{5}$$

where the sampling function is defined as $\mathrm{sinc}(t) \triangleq \sin(\pi t)/(\pi t)$, and $\mathrm{sinc}[l] \triangleq \mathrm{sinc}(l/W)$, $W = 1/T_s$ is the system bandwidth, and T_s is the sampling period. We assume that $L \times T_s \leq \tau_{max}$, where τ_{max} is the maximum excess delay of the channel. With this model, we obtain a channel matrix

$$\begin{aligned} \boldsymbol{H} &= [\boldsymbol{h}_1 \ldots \boldsymbol{h}_R] \\ &= [(h_1(l) \ldots h_1(L))^T \ldots (h_R(l) \ldots h_R(L))^T] \end{aligned} \tag{6}$$

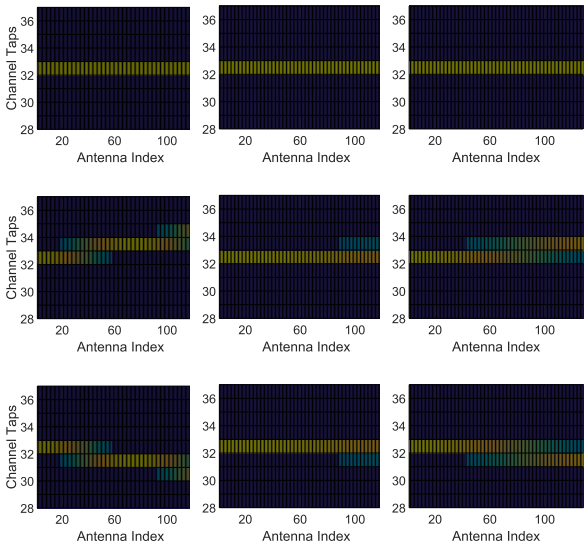


FIGURE 2. Effect of the parameters (W, f_c, θ) on the channel structure.

From the discussion above, we observe that the delay-spatial structure of the channel matrix $\boldsymbol{H}$ depends on the system parameters W and f_c, on the multi-path component parameters $d_{0,i}$ and θ_i, and on the array geometry, which is fully described by Δd and R in a ULA. To reveal the structure characteristics clearly, we illustrate an instance of the channel matrix by simulation. We set the carrier frequency as 2.6 and 6 GHz respectively, and bandwidth as 100 and 20 MHz respectively. Then generate one multi-path component and sample it by assuming 3 different DOAs. The results are shown in Fig.(2), where the 3 rows of subfigure are generated by $\theta \in \{\pi/2, \pi/4, 3\pi/4\}$ in sequence. The 3 columns are the channel patterns generated with parameters (W, f_c) with (100MHz, 2.6GHz), (20MHz, 2.6GHz), and (100MHz, 6GHz) in sequence.

From the simulation results we can see how the parameters θ, W, and f_c affect the channel support patterns. 1) From the incident signal side, when $\theta = \pi/2$ all the significant taps of the channel vectors have the same position over the array for all kinds of bandwidth and carrier frequency configurations, and therefore we say that all the channels have the same support pattern. And in the other two cases, the channel's support will shift in different directions. As a consequence, we see that 2 or 4 different types of support–i.e. number and position of the channel's significant taps– can be observed across the array elements for different configurations. 2) From the system parameter side, with a fixed carrier frequency, the support will shift more for larger bandwidths and, consequently, a larger number of different support patterns can be observed across the array. With a fixed bandwidth, increasing the carrier frequency leads to less shift on the supports across the array, as the inter-element distance in the array is inversely proportional to the carrier frequency.

In the wireless channel of massive MIMO under relatively high carrier and bandwidth scenarios, there will be few

multi-path components, which will make up structured sparse channels over the array. If two paths arrive close together, and the two waves may be summarized together to contribute to the same channel tap–some taps are strengthened while others are receded–the non-stationarity of the response over the large antenna array was observed [2], [3]. From the discussion above, we can conclude that the uplink channel of massive MIMO has a structured sparse character in the delay-spatial domain, and the structure of channel matrix is effected by the physical scenario and communication system parameters. To improve the channel estimation performance by exploiting the channel structure, a flexible structured sparse channel model and an efficient algorithm are needed.

III. COLLAPSED VBI-DP BASED ALGORITHM FOR STRUCTURED CHANNEL ESTIMATION

From the discussion in Sec.II-B, it is observed that the channel responses seen across the different array elements share many structural features, and we aim to exploit such similarities in order to improve the channel estimation accuracy. To achieve this, we first construct a probabilistic model of the received signal at the array based on DP that captures the similarities of the responses seen across the array. Then, we derive an inference method on that model that we coin "collapsed VBI".

A. PROBABILISTIC MODEL BASED ON DIRICHLET PROCESS

1) LIKELIHOOD FUNCTIONS

Assuming the noise vectors follow a complex Gaussian distribution with precision α_0, then the likelihood function of signal model in (1) can be given as

$$\boldsymbol{y}_r \mid \boldsymbol{h}_r, \alpha_0 \sim \mathcal{CN}(\boldsymbol{A}\boldsymbol{h}_r, \alpha_0^{-1}\boldsymbol{I}_N) \tag{7}$$

where $\boldsymbol{I}_N$ denotes an identity matrix of size N. The noise precision α_0 follows a Gamma distribution

$$\alpha_0 \sim \text{Gamma}(a, b). \tag{8}$$

2) SPARSITY PATTERN OF h_r

The channel vectors $\boldsymbol{h}_r$ are modeled as complex, zero-mean Gaussian vectors with covariance matrix $\boldsymbol{\Lambda}_r^{-1}$, i.e.

$$\boldsymbol{h}_r \mid \boldsymbol{\alpha}_r \sim \mathcal{CN}(0, \boldsymbol{\Lambda}_r^{-1}) \tag{9}$$

where $\boldsymbol{\Lambda}_r = \text{Diag}(\boldsymbol{\alpha}_r)$. The entries of the precision vector $\boldsymbol{\alpha}_r$ are modeled as statistically independent variables following a Gamma distribution

$$\boldsymbol{\alpha}_r \sim \prod_{l=1}^{L} \text{Gamma}(c, d) \tag{10}$$

As the entries of the precision vector are the inverse variances of the entries of the channel vectors $\boldsymbol{h}_r$, they contain information about the sparsity pattern–or support–of the channel vector. Channel taps with large value of their associated precision vector entries correspond to taps that are nearly zero, while those with low precision value correspond to taps with significant channel energy. Consequently, channel vectors having the same support pattern will have the same precision vector. As discussed in the previous section, channels corresponding to adjacent array elements will, with high probability, have the same support, and the support patterns of the channel vectors for all the array elements can therefore be described with just a few precision vectors. We call this structure a clustered-sparse structure, as the different vectors $\boldsymbol{h}_r$ vectors to be estimated can be grouped into a few clusters with the same support pattern. In order to exploit such clustered-sparse nature, we use Dirichlet Processes as a tool for modeling the precision vectors of the channel vectors.

3) DIRICHLET PROCESS WITH STICK-BREAKING

Inspired by their excellent clustering performance, we adopt the modeling by Dirichlet Process ([11], [12]). We assume $\{\boldsymbol{\alpha}_r\}_{r=1:R}$ are R random samples drawn from G, where G is a random measure drawn from $DP(\gamma, G_0)$, where γ is the concentration parameter related to the number of clusters, and G_0 is the base distribution, i.e. the mean of the process. Following the context of Bayesian non-parametric methods, we have

$$\begin{aligned} \boldsymbol{\alpha}_r &\sim G, \quad r = 1, \cdots, R \\ G &\sim DP(\gamma, G_0). \end{aligned} \tag{11}$$

Since the exact formulation of G is unavailable, at first we follow the stick-breaking construction method [13]

$$G = \sum_{k=1}^{\infty} w_k \delta_{\boldsymbol{\alpha}_k^*} \tag{12}$$

where $\{\boldsymbol{\alpha}_k^*\}_{k=1,\ldots\infty}$ are the atoms of the distribution G, each of them follows the base distribution G_0, which has the same shape as $\boldsymbol{\alpha}_r$. The variables $\{w_k\}_{k=1,\ldots\infty}$ are the weights of each atom, and can be initialized in the stick-breaking fashion

$$w_k = \pi_k \prod_{l=1}^{k-1}(1 - \pi_l). \tag{13}$$

We assume each stick π_k follows a Beta distribution

$$\pi_k \sim \text{Beta}(1, \gamma) \tag{14}$$

where γ is the concentration parameter of the DP, which controls the number of clusters. This process can be thought of as repeatedly breaking a stick of unit length into two pieces where the breakpoint is randomly sampled from the Beta distribution, and guarantees that the infinite sum of mixture weights w_k converges to 1 in probability. The purpose of choosing the stick-breaking formulation is to apply it flexibly by using the Beta distribution.

To assign a precision vector $\boldsymbol{\alpha}_r$ to each of the cluster precisions $\{\boldsymbol{\alpha}_k^*\}_{k=1,\ldots\infty}$, we introduce an indicator vector $\boldsymbol{z}_r$, which follows a Multinomial distribution

$$\boldsymbol{z}_r \mid \{w_k\}_{k=1,\ldots\infty} \sim \text{Multinomial}(\{w_k\}_{k=1,\ldots\infty}) \tag{15}$$

In the above construction, z_r is a vector with all its entries being zero except one of them, which is one, with the probabilities of each entry being one given by $\{w_k\}_{k=1,...\infty}$.

With the above, we can now pose the probabilistic model for the channel vectors as

$$\boldsymbol{h}_r \mid z_r, \{\boldsymbol{\alpha}_k^*\}_{k=1,...\infty} \sim \prod_{k=1}^{\infty} \{\mathcal{CN}(0, \boldsymbol{\Lambda}_{z_r}^{-1}\}^{1[z_r=k]} \quad (16)$$

where $1[z_r = k]$ is an indicator function equal to 1 if the k-th entry of vector z_r is one, and zero otherwise. For tractability reasons, we also truncate the infinite number of clusters to a big number K.

4) COLLAPSED DIRICHLET PROCESS

The main reason for using the stick breaking method to construct a DP is that it is conjugate to the multinomial distribution, and therefore it is easy to infer the posterior in the Bayesian framework. The original VBI assumes a factorized form for the posterior distribution which means the parameters are independent of assignment variables. However, the stick breaking representation is defined in the space of the explicit cluster labels. For a two clusters case, if we have 3 observations and assign them as ($z_1 = z_2 = 1$, $z_3 = 2$), this will result in a different solution than the assignment ($z_z = z_2 = 2, z_3 = 1$). So randomly permuting the cluster labels will change the probability of the observations [14]. This is a bad assumption when applying VBI on the model, because changes in $\{\pi_k\}_{k=1:K}$ by randomly permuting the labels will impact on $\{z_r\}_{r=1:R}$. To surmount this, we marginalize the variables $\{\pi_k\}_{k=1:K}$ out of the model and get the collapsed DP model [14]:

$$\begin{aligned} p(\boldsymbol{Z} \mid \gamma) &= \int_0^1 p(\boldsymbol{Z} \mid \boldsymbol{\pi}) p(\boldsymbol{\pi}) \mid \gamma) d\boldsymbol{\pi} \\ &= \int_0^1 \prod_{r=1}^{R} p(z_r \mid \boldsymbol{\pi}) \prod_{k=1}^{K} p(\pi_k) \mid \gamma) d\boldsymbol{\pi} \\ &= \gamma^K \prod_{k=1}^{K} \frac{\Gamma(1+N_k)\Gamma(\gamma + N_{>k})}{\Gamma(1+\gamma+N_{\geq k})} \end{aligned} \quad (17)$$

where $\boldsymbol{Z} = [z_1, \ldots, z_R]$, $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]$, $N_k \triangleq \sum_{r=1}^{R} 1[z_r = k]$, $N_{>k} \triangleq \sum_{r=1}^{R} 1[z_r > k]$, and $N_{\geq k} \triangleq N_k + N_{>k}$. We define a matrix collecting all the observations as $\boldsymbol{Y} = [\boldsymbol{y}_1, ..., \boldsymbol{y}_R]$, the set of hidden variables as $\boldsymbol{\Theta} = \{\alpha_0, \gamma, \boldsymbol{Z}, \boldsymbol{H}, \{\boldsymbol{\alpha}_k^*\}_{k=1,...,K}\}$, and the set of hyperparameters as $\boldsymbol{\Omega} = \{a, b, c, d, e, f\}$. The joint density function for all variables is given by

$$\begin{aligned} p(\boldsymbol{Y}, \boldsymbol{\Theta}) = &\prod_{r=1}^{R} p(\boldsymbol{y}_r \mid \boldsymbol{h}_r, \alpha_0) p(\boldsymbol{h}_r \mid z_r, \{\boldsymbol{\alpha}_k^*\}_{k=1:K}) \\ &\times p(\alpha_0) p(\boldsymbol{Z} \mid \gamma) \prod_{k=1}^{K} p(\boldsymbol{\alpha}_k^*). \end{aligned} \quad (18)$$

Fig.3 graphically illustrates the dependencies between the different variables based on two different kinds of DP model

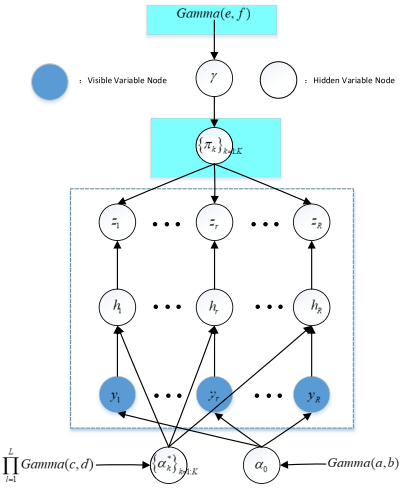


FIGURE 3. Graphical Model for Up-Link Channel in Massive MIMO Collapsed DP based.

in the same graph. The nodes under the indigo shadow exist in the original DP but don't belong to the proposed model which doesn't include the variables $\boldsymbol{\pi}$ or the prior on γ. The latter model shows a more compact dependencies relationship between the variables. We expect that inference on the collapsed model will perform better than in the original DP model because it reduces the impact of fluctuations of $\boldsymbol{\pi}$ on the latent variables.

B. VBI ALGORITHM FOR COLLAPSED DP MODEL

According to the VBI framework, we design the approximate posterior with a factorized shape

$$q(\boldsymbol{\Theta}) = q(\alpha_0) \prod_{r=1}^{R} q(z_r) q(\boldsymbol{h}_r) \prod_{k=1}^{K} q(\boldsymbol{\alpha}_k^*). \quad (19)$$

We compute $q(\boldsymbol{\Theta})$ by alternately minimizing the Kullback-Leibler divergence $D(q(\boldsymbol{\Theta})||p(\boldsymbol{\Theta}|\boldsymbol{Y}))$ with respect to each of the individual factors above. In the general rule, the optimal factor $q^*(\boldsymbol{\Theta_k})$ given all other factors can be obtained by

$$\ln q^*(\boldsymbol{\Theta}_k) \propto \langle \ln p(\boldsymbol{Y}, \boldsymbol{\Theta}) \rangle_{q(\boldsymbol{\Theta} \setminus \Theta_k)} \quad (20)$$

where $\boldsymbol{\Theta} \setminus \Theta_k$ denotes the set of all variables $\boldsymbol{\Theta}$ except Θ_k. Then the updated approximate posterior for each hidden variable can be obtained by computing (20) iteratively.

1) UPDATE OF α_0

According to (20), the update equation for α_0 reads

$$\ln q^*(\alpha_0) \propto \langle \ln p(\boldsymbol{Y}|\boldsymbol{H}, \alpha_0) p(\alpha_0) \rangle_{q(\boldsymbol{H})} \quad (21)$$

The prior of α_0 is a Gamma distribution with parameters a, b, and the posterior of it is also a Gamma distribution with the updated parameters $\hat{a}$, $\hat{b}$:

$$\begin{aligned} \hat{a} &= a - 1 + RM \\ \hat{b} &= b + \sum_{r=1}^{R} (||\boldsymbol{y}_r - \boldsymbol{A}\boldsymbol{\mu}_r||_2^2 + \text{trace}\{\boldsymbol{A}^H \boldsymbol{A} \boldsymbol{\Gamma}_r^{-1}\}) \end{aligned} \quad (22)$$

where $\boldsymbol{\mu}_r$ and $\boldsymbol{\Gamma}_r$ are the updated mean and precision of $\boldsymbol{h}_r$ respectively, whose value will be given later. The updated estimate of α_0 is then given as the mean of the approximate posterior:

$$\hat{\alpha}_0 = \frac{\hat{a}}{\hat{b}} \tag{23}$$

2) UPDATE OF h_r

According to (20), we have

$$\ln q^*(\boldsymbol{h}_r) \propto \langle \ln p(\boldsymbol{y}_r|\boldsymbol{h}_r, \alpha_0) \, p(\boldsymbol{h}_r \mid \boldsymbol{z}_r, \{\boldsymbol{\alpha}_k^*\}_{k=1:K}) \rangle_{q(\alpha_0)q(\boldsymbol{z}_r))\prod_{k=1}^K q(\boldsymbol{\alpha}_k^*)} \tag{24}$$

Since the prior of $\boldsymbol{h}_r$ is a complex Gaussian distribution with mean $\boldsymbol{\mu}_r$ and precision $\boldsymbol{\Gamma}_r$, the posterior of $\boldsymbol{h}_r$ follows the same distribution with updated mean $\hat{\boldsymbol{\mu}}_r$ and precision $\hat{\boldsymbol{\Gamma}}_r$:

$$\begin{aligned} \hat{\boldsymbol{\mu}}_r &= \hat{\alpha}_0 \hat{\boldsymbol{\Gamma}}_r^{-1} \boldsymbol{A}^H \boldsymbol{y}_r \\ \hat{\boldsymbol{\Gamma}}_r &= \sum_{k=1}^{K} q^*(z_r = k)\hat{\boldsymbol{\Lambda}}_k + \hat{\alpha}_0 \boldsymbol{A}^H \boldsymbol{A} \end{aligned} \tag{25}$$

where $\hat{\boldsymbol{\Lambda}}_k \triangleq \mathrm{Diag}\{\hat{\boldsymbol{\alpha}}_k^*\}$. We get the update of $\boldsymbol{h}_r$ as its mean vector

$$\hat{\boldsymbol{h}}_r = \hat{\boldsymbol{\mu}}_r. \tag{26}$$

3) UPDATE OF α_k^*

Following the same method as before, the update for α_k^* is obtained as

$$\ln q^*(\boldsymbol{\alpha}_k^*) \propto \langle \ln \prod_{r=1}^{R} p(\boldsymbol{h}_r|z_r = k, \{\boldsymbol{\alpha}_k^*\}_{k=1:K}) p(\boldsymbol{\alpha}_k^*) \rangle_{q(\boldsymbol{h}_r)q(z_r)} \tag{27}$$

Each element of $\boldsymbol{\alpha}_k^*$ follows a prior Gamma distribution with parameters c, d, and consequently the posterior distribution is also Gamma with updated parameters $\hat{c}, \hat{d}$

$$\begin{aligned} \hat{c} &= c + \sum_{r=1}^{R} q^*(z_r = k) \\ \hat{d} &= d + \sum_{r=1}^{R} q^*(z_r = k)(\mu_{r,l}^2 + \sigma_{r,l}^2) \end{aligned} \tag{28}$$

where $\mu_{r,l}$ and $\sigma_{r,l}^2$ are the l-th element of the mean vector $\boldsymbol{\mu}_r$ and the covariance matrix $\boldsymbol{\Gamma}_r$ respectively. The approximate posterior mean of $\boldsymbol{\alpha}_k^*$ is then

$$\hat{\boldsymbol{\alpha}}_{k,l}^* = \frac{\hat{c}}{\hat{d}} \tag{29}$$

4) UPDATE OF z_r

We update the approximate posterior of $\boldsymbol{z}_r$ according to

$$\begin{aligned} \ln q^*(z_r = k) &\propto \langle \ln p(\boldsymbol{h}_r|z_r = k, \{\boldsymbol{\alpha}_k^*\}_{k=1:K}) p(\boldsymbol{Z}|\gamma) \rangle_{q(\boldsymbol{h}_r)\prod_{k=1}^{K} q(\boldsymbol{\alpha}_k^*)\prod_{m\neq r}^{R} q(z_m)} \\ &= \underbrace{\langle \ln p(\boldsymbol{h}_r|z_r = k, \{\boldsymbol{\alpha}_k^*\}_{k=1:K}) \rangle_{q(\boldsymbol{h}_r)\prod_{k=1}^{K} q(\alpha_k^*)}}_{eq_1} \\ &\quad + \underbrace{\langle \ln p(\boldsymbol{Z}|\gamma) \rangle_{\prod_{m\neq r}^{R} q(z_m)}}_{eq_2} \end{aligned} \tag{30}$$

$$eq_1 = \sum_{l=1}^{L} \ln \hat{\alpha}_{k,l}^* - \mathrm{trace}\{\hat{\boldsymbol{\Gamma}}_r^{-1}\hat{\boldsymbol{\Lambda}}_k\} - \hat{\boldsymbol{\mu}}_r^H \hat{\boldsymbol{\Lambda}}_k \hat{\boldsymbol{\mu}}_r \tag{31}$$

$$\begin{aligned} eq_2 &= \langle \ln p(\boldsymbol{z}_r|\boldsymbol{Z}_{\setminus r}, \gamma) \rangle_{\prod_{m\neq r}^{R} q(z_m)} + \langle \ln p(\boldsymbol{Z}_{\setminus r}|\gamma) \rangle_{\prod_{m\neq r}^{R} q(z_m)} \\ &\propto \langle \ln p(\boldsymbol{z}_r|\boldsymbol{Z}_{\setminus r}, \gamma) \rangle_{\prod_{m\neq r}^{R} q(z_m)} + \text{constant} \\ &\propto \langle \ln(1+N_k^{\setminus r}) \rangle_{\prod_{m\neq r}^{R} q(z_m)} - \langle \ln(1+\gamma+N_{\geq k}^{\setminus r}) \rangle_{\prod_{m\neq r}^{R} q(z_m)} \\ &\quad + \sum_{f<k} \langle \ln(\gamma + N_{>f}^{\setminus r}) \rangle_{\prod_{m\neq r}^{R} q(z_m)} \\ &\quad - \sum_{f<k} \langle \ln(1+\gamma+N_{\geq f}^{\setminus r}) \rangle_{\prod_{m\neq r}^{R} q(z_m)} \end{aligned} \tag{32}$$

where $\boldsymbol{Z}_{\setminus r}$ denotes the matrix $\boldsymbol{Z}$ with the column $\boldsymbol{z}_r$ removed, $N_k^{\setminus r} \triangleq \sum_{m\neq r} 1[z_m = k]$, $N_{>k}^{\setminus r} \triangleq \sum_{m\neq r} 1[z_r > k]$, and $N_{\geq k}^{\setminus r} \triangleq N_k^{\setminus r} + N_{>k}^{\setminus r}$.

It's difficult to compute the expectations in (32) directly, we use a second order Taylor Expansion ([14], [15])

$$\mathrm{E}[f(x)] \approx f(\mathrm{E}[x]) + \frac{1}{2} f''(\mathrm{E}[x])\mathrm{V}[x]$$

where $\mathrm{E}[x]$ and $\mathrm{V}[x]$ are the mean and variance of a random variable x respectively. Using this, eq_2 can be approximated by

$$\begin{aligned} eq_2 \approx{}& \ln(1+\mathrm{E}[N_k^{\setminus r}]) + \frac{\mathrm{V}[N_k^{\setminus r}]}{2(1+N_k^{\setminus r})^2} \\ &- (\ln(1+\hat{\gamma}+\mathrm{E}[N_{\geq k}^{\setminus r}]) + \frac{\mathrm{V}[\hat{\gamma}+N_{\geq k}^{\setminus r}]}{2(1+\hat{\gamma}+N_{\geq k}^{\setminus r})^2}) \\ &+ \sum_{f<k} (\ln(\hat{\gamma}+\mathrm{E}[N_{>f}^{\setminus r}]) + \frac{\mathrm{V}[\hat{\gamma}+N_{>f}^{\setminus r}]}{2(\hat{\gamma}+N_{>f}^{\setminus r})^2}) \\ &- \sum_{f<k} (\ln(1+\hat{\gamma}+\mathrm{E}[N_{\geq f}^{\setminus r}]) + \frac{\mathrm{V}[\hat{\gamma}+N_{\geq f}^{\setminus r}]}{2(1+\hat{\gamma}+N_{\geq f}^{\setminus r})^2}) \end{aligned} \tag{33}$$

To compute the means and variances in (33), we use the following equations:

$$\begin{cases} E[N_k] = \sum_{r=1}^{R} q(z_r = k) \\ V[N_k] = \sum_{r=1}^{R} q(z_r = k)(1 - q(z_r = k)) \\ E[N_{>k}] = \sum_{r=1}^{R} \sum_{f>k} q(z_r = f) \\ V[N_{>k}] = \sum_{r=1}^{R} \sum_{f>k} q(z_r = f) \sum_{i\leq k} (1 - q(z_r = i)). \end{cases} \tag{34}$$

5) UPDATE OF γ

Since the role of γ in the DP model is to control the number of inferred clusters, we propose to estimate as well this variable so that the model can adapt to the different types of channel conditions observed in the received signal. Applying the VBI method, we get

$$\begin{aligned}\ln q^*(\gamma) &= \langle \ln p(\boldsymbol{Z}|\gamma)\rangle_{q(\boldsymbol{Z})} \\ &= \sum_{k=1}^{K}(\langle \ln \Gamma(1+N_k) + \ln \Gamma(\gamma + N_{>k})\rangle_{q(\boldsymbol{Z})} \\ &\quad - \langle \ln \Gamma(1+\gamma+N_{\geq k})\rangle_{q(\boldsymbol{Z})}) + K\ln\gamma \end{aligned} \tag{35}$$

where $\Gamma(\cdot)$ is the Gamma function. Since the mean of this approximate posterior is difficult to compute, we apply Minka's fixed-point iteration [17] to get the following update:

$$\hat{\gamma} = \frac{K}{\sum_{k=1}^{K}(\Psi(1+\gamma^*+\hat{N}_{\geq k}) - \Psi(1+\gamma^*+\hat{N}_{>k}))} \tag{36}$$

where $\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the Digamma function, and γ^* is the value from last iteration.

C. STRUCTURED SPARSE CHANNEL ESTIMATION ALGORITHM

The structured sparse channel estimation algorithm is obtained by iteratively calculating the updates of α_0, $\{\boldsymbol{\mu}_r, \boldsymbol{\Gamma}_r, \boldsymbol{z}_r\}_{r=1:R}$, $\{\boldsymbol{\alpha}_k^*\}_{k=1:K}$, and γ until convergence, with the estimates of the channel response being given by (26). The steps of the algorithm are summarized in Algorithm 1. As the stopping criterion, we set a threshold on the average squared distance of two successive estimates of $\boldsymbol{H}$, or maximum number of iterations It_{max}. The choice on the value of each hyperparameter in $\boldsymbol{\Omega}$ is another issue [13], [18], but we won't discuss it here. We set all the value of hyperparameters except c as 1.0×10^{-4}, and $c = 1$.

Algorithm 1 Structured Channel Estimation Algorithm

Input: Observations $\boldsymbol{Y}$, the matrix $\boldsymbol{A}$, Threshold Thr,
Maximum iteration number It_{max}.
Output: The estimated channel $\boldsymbol{H}$
Initialize: $\boldsymbol{\Omega}$, $\boldsymbol{\Theta}$; iteration counter $it = 0$;
while $||\hat{\boldsymbol{H}} - \hat{\boldsymbol{H}}^{it-1}||_2^2 < Thr$ *or* $it < It_{max}$ **do**
  Update α_0 by (23) and (22);
  Update $\boldsymbol{H}$ by (26) and (25);
  Update $\boldsymbol{\alpha}_k^*$ by (29) and (28);
  Update $\boldsymbol{Z}$ by (30), (31),(32), (33), and (34);
  Update γ by (36);
  $it = it + 1$;
end

The iteration steps in the algorithm we show here are similar to those of [12]. There are 3 different steps on updating equations of $\boldsymbol{Z}$, γ and $\boldsymbol{\pi}$ in the iteration part of the two algorithms. The most significant difference is the marginalization of $\boldsymbol{\pi}$ to improve the accuracy on cluster assignment. Compared with [14], we apply the collapsed VBI based on DPM to a channel estimation problem, which leads to a more complex model shown in Fig.3. Furthermore, we update γ to infer the number of clusters more accurately.

D. ALGORITHM ANALYSIS

To show that the proposed channel estimation algorithm works robustly and has better performance than the benchmarks without increasing the computational cost, we analyze its convergence behavior and computational complexity next.

1) CONVERGENCE

The convergence of collapsed VBI for DP models has been discussed in [14], where it was shown that the collapsed VBI algorithm reaches convergence slower than when using the non-collapsed model, but can be guaranteed by Minka's fixed-point equation [17]. To illustrate this in our system, we numerically assess the convergence for different SNR values as shown in Fig.(4). We implement the algorithm for a system with $B_W = 20$ MHz, with $N_P = 64$ evenly-spaced pilots, a channel with $I_P = 6$ multi-path components in a vector of length $L = 64$, and SNR ranging over $0 \sim 30$ dB.

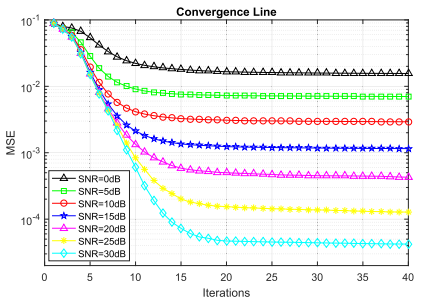


FIGURE 4. Convergence performance over iterations: $B_W = 20$MHz, $N_P = 64$, $I_P = 6$, $L = 64$, $SNR = 0 \sim 30$dB.

From the simulation results, we can see that the algorithm always reaches convergence at around 20 iterations, which will change slightly with SNR.

2) COMPUTATIONAL COMPLEXITY

From the system model Eq.(1) and the channel estimation algorithm from Eq.(23) to Eq.(36), we know that maximum cost in the iterative algorithm is the computation of the inverse of the channel covariance matrix $\hat{\boldsymbol{\Gamma}}_r$. In the update equation (25), $\hat{\boldsymbol{\Lambda}}_k$ and $\boldsymbol{A}^H\boldsymbol{A}$ are diagonal matrices, and the summation of $\hat{\boldsymbol{\Gamma}}_r$ is a diagonal matrix with the dimensions $L \times L$. So the computational complexity of this is $\mathcal{O}(RL^2)$. Similar, we can get the computational complexity for the stat-of-art algorithms, which are shown in Table 1. The Fast SBL (FSBL) estimator is proposed in [18], the T-MSBL is a efficient Block SBL estimator which makes a reasonable approximation to the T-SBL under a similar recovery performance [19], and the SABMP algorithm is a sparse distribution agnostic Bayesian matching pursuit method from [9]

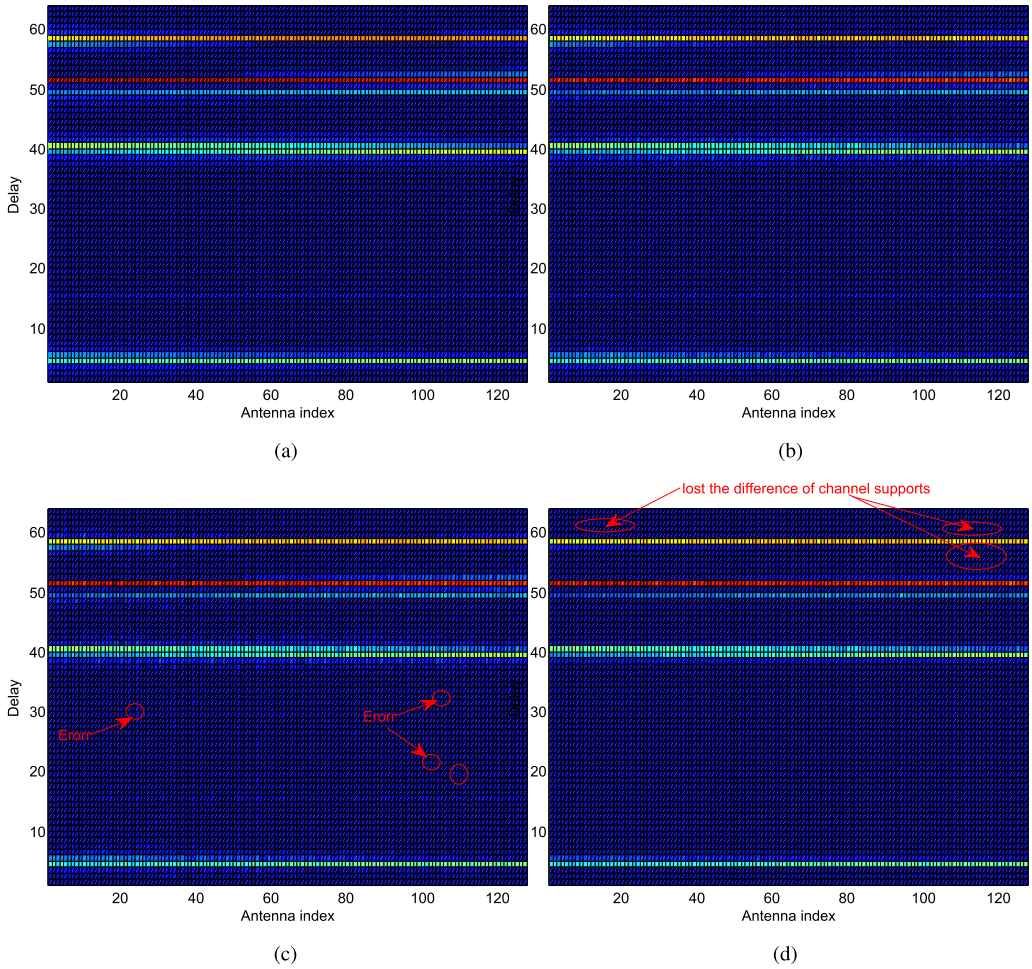


FIGURE 5. Simulated Original Channel Structure and Observed by Different Estimator: $B_W = 20$ MHz, $I_P = 6$, $SNR = 15$dB, $L = 64$, and $N_P = 64$. (a) Original Channel. (b) Channel Estimated by CVBI-DP. (c) Channel Estimated by FSBL. (d) Channel Estimated by BSBL.

TABLE 1. Complexity order of the algorithms.

Algorithm	Complexity
CVBI-DP	$\mathcal{O}(RL^2)$
FSBL	$\mathcal{O}(RL^2)$
T-MSBL	$\mathcal{O}(LN^2)$
SABMP	$\mathcal{O}(RNL^2)$

which has been proven to outperform several other compressive sensing methods, including OMP, StOMP, and Bayesian compressive sensing.

With the system parameters that we use in the next section for numerical assessment ($N = 2048, L = 64, R = 128$), we observe that the computational complexity of CVBI-DP and FSBL are the lowest of all the evaluated algorithms.

IV. NUMERICAL ASSESSMENT

A. SIMULATION SETUP

To demonstrate the performance of our proposed algorithm, we consider a MIMO-OFDM system following the description given in Section II with the parameters shown in Table 2.

The channel responses are generated following (4) and (5), with two different settings for the number of multipath components, $I_P = 4$ or $I_P = 6$.

TABLE 2. Parameters for MIMO-OFDM.

Parameter	Value	Comment
R	128	Number of antennas in ULA
f_c	2.6 G Hz	Central carrier frequency
N	2048	Number of subcarriers
B_W	$10 \sim 150$ MHz	Bandwidth
M	4	Order of QAM
L	64	Length of channel tap vector
N_P	$32 \sim 2048$	Length of pilot vector
I_P	{4,6}	Number of multi-path

B. NUMERICAL RESULTS

1) RECOVERY OF THE CHANNEL STRUCTURE

In order to illustrate the advantages of exploiting the clustered-sparse structure of the Massive MIMO channel responses in our proposed estimation algorithm, we compare its performance to that of two sparse estimators: the FSBL and T-MSBL. The former estimates independently the channels observed at each of the array elements and, therefore, doesn't exploit the similarities between them. On the other hand, the latter assumes that the sparsity pattern of the channels seen at all the array elements is the same and,

hence, estimates a single support for all responses. In Fig. 5, the resulting estimates for an instance of the algorithms are shown, along with the original channel response, for a system with $B_W = 20$ MHz, $N_P = 64$ pilot subcarriers, a channel with $I_P = 6$ multi-path components in a $L = 64$ length vector, and a signal-to-noise ratio (SNR $\triangleq$ Trace(Cov($\boldsymbol{Ah}_r$))/Trace(Cov($\boldsymbol{h}_r$))) of 15 dB. In these results, we observe that, although the sparsity pattern of the channel responses stays approximately constant across the array elements, there are slight variations of the delays of the multi-path components across the array. While our proposed CVBI-DP algorithm appears to be able to track those slight variations accurately, the T-MSBL (in the figures we rename it BSBL) estimator is not and instead estimates a constant support over all array elements. On the other hand, FSBL is able to estimate the positions of the significant channel taps well for each array element, but tends to mistakenly assign small power to portions of the response where no multi-path component is present. We attribute this to the fact that FSBL does not exploit the similarities between the channels observed at the different array elements. As we will see next, the differences between the estimators observed in Fig. 5 result in better overall accuracy for the CVBI-DP estimator.

2) ESTIMATION ACCURACY PERFORMANCE

Next, we evaluate the estimation accuracy of the proposed algorithm and the benchmarks over different SNR values. We do so by numerically evaluating the mean squared error (MSE) of the estimated responses, defined as MSE $= \mathrm{E}\{||\hat{\boldsymbol{H}} - \boldsymbol{H}||_2^2\}/(L \times R)$, over 50 Monte-Carlo runs in a system with $B_W = 20$ MHz, number of pilot $N_P = 64$, and channels with $I_P = 4$ multi-path components in a full vector length of $L = 64$ taps. In addition to the benchmarks already mentioned, we include here the SABMP algorithm from [9] and the DP-based estimators–without model collapsing–of [12] and [13]. As an ideal bound, we include an Oracle estimator which implements LMMSE estimation of the responses assuming perfect knowledge of the sparsity pattern of each channel response and of the noise variance; this can be shown to be equivalent to the Cramér-Rao lower bound (CRLB) of the channel vector assuming knowledge of its support. The results, depicted in Fig. 6, demonstrate the superior estimation accuracy of our proposed CVBI-DP algorithm over all benchmarks throughout the entire SNR range.

3) INFLUENCE OF SYSTEM PARAMETERS

We investigate next how two key system parameters, bandwidth and number of pilots, affect the performance of different channel estimation methods. To that end, we first set the SNR to 20 dB, and vary the transmission bandwidth from 10 MHz to 150 MHz. The results, shown in Fig. 7, indicate that while other estimators, such as the T-MSBL and the original DP, see a performance degradation as bandwidth increases, our proposed CVBI-DP method's performance is robust to the different bandwidth configurations.

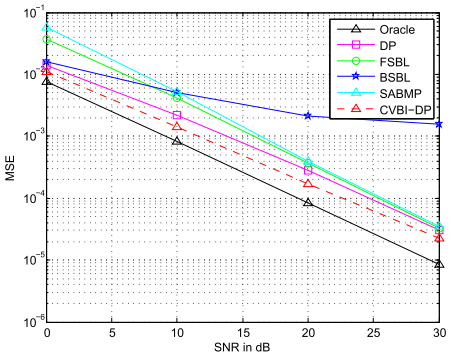


FIGURE 6. MSE performance over SNR:0 ∼ 30dB, $B_W = 20$ MHz, $I_P = 4$, $L = 64$, and $N_P = 64$.

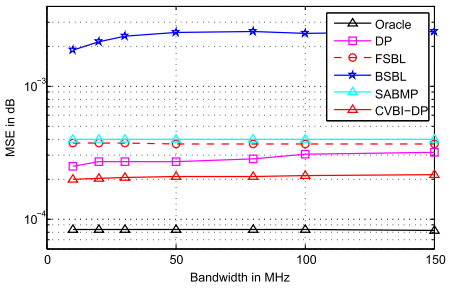


FIGURE 7. MSE performance over B_W : 10 ∼ 150MHz, SNR = 20dB, and $I_P = 4$, $L = 64$, and $N_P = 64$.

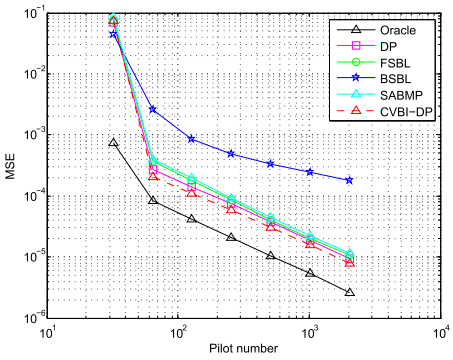


FIGURE 8. MSE performance over pilot overhead:32 ∼ 2048, $B_W = 20$MHz, $SNR = 20$dB, $I_P = 4$, $L = 64$, and $N_P = 64$.

To evaluate the effect of the amount of pilots symbols, we set the SNR to 20 dB and the bandwidth to 20 MHz, and vary the number of pilot subcarriers from 32 to 512. From the results shown in Fig. 8, we conclude that our proposed estimator performs better than all competitors regardless of the amount of pilots used.

When system bandwidth increases, there are more non-zero elements around the delay of the incident wave after sampling, and the channel structure becomes more complex. From Fig.7 we can see that the proposed method CVBI-DP is robust to bandwidth. From Fig.8 we know that regardless of the pilot overhead, the CVBI-DP always provides better estimation performance, so we can choose a low pilot overhead.

From the simulation results shown above, we draw the overall conclusion that the proposed structured channel estimation algorithm based on collapsed VBI-DP improves

the MSE performance without increasing the computational complexity and pilot overhead, while it works robustly towards other system parameters such as bandwidth.

V. CONCLUSION AND FUTURE RESEARCH

In this paper we analyze the uplink channel structure in massive MIMO-OFDM system, then derive a sparse channel estimation algorithm based on CVBI-DP. Simulation results show that the proposed channel estimation method has better performance under the MSE criterion without increasing the computational complexity and pilot overhead, and it's also robust to the system parameters.

In the future, we will continue the research on structured channel estimation for massive MIMO, but focus on two points: 1) we will improve the DP model to reinforce spatial relationship between the neighboring channels, e.g. by devising a correlation model for the neighboring taps in one channel vector; 2) we will rebuild the probabilistic channel model to adapt to different kinds of antenna arrays.

ACKNOWLEDGEMENTS

This work was supported by the China Scholarship Council for visiting the group of Prof. Bernard Fleury in the department of Electronic Systems at Aalborg University. The author would like to thank the group for the help on the research.

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] X. Gao, J. Flordelis, G. Dahman, F. Tufvesson, and O. Edfors, "Massive MIMO channel modeling—Extension of the COST 2100 model," in *Proc. Joint NEWCOM/COST Workshop Wireless Commun. (JNCW)*, vol. 1, Jun. 2015, pp. 4558–4563, doi: 10.1109/ICC.2015.7249041.

[3] S. Wu, C. X. Wang, H. Haas, E. H. M. Aggoune, M. M. Alwakeel, and B. Ai, "A non-stationary wideband channel model for massive MIMO communication systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1434–1446, Mar. 2015, doi: 10.1109/TWC.2014.2366153.

[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014, doi: 10.1109/MCOM.2014.6736761.

[5] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 2890–2895, doi: 10.1109/WCNC.2013.6555020.

[6] Y. Nan, X. Sun, and L. Zhang, "Joint channel estimation algorithm via weighted Homotopy for massive MIMO OFDM system," *Digital Signal Processing*, vol. 50, pp. 34–42, Mar. 2016, doi: 10.1016/j.dsp.2015.11.010.

[7] C. Qi, Y. Huang, S. Jin, and L. Wu, "Sparse channel estimation based on compressed sensing for massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (IEEE ICC)*, vol. 61, no. 8, Jun. 2015, pp. 4558–4563, doi: 10.1109/ICC.2015.7249041.

[8] 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. 2015, doi: 10.1109/TCOMM.2015.2508809.

[9] 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, doi: 10.1109/TSP.2014.2369005.

[10] X. Lin, S. Wu, L. Kuang, Z. Ni, X. Meng, and C. Jiang, "Estimation of sparse massive MIMO-OFDM channels with approximately common support," *IEEE Commun. Lett.*, vol. 11, no. 3, pp. 1179–1182, May 2017, doi: 10.1109/LCOMM.2017.2657620.

[11] Y. Qi, D. Liu, D. Dunson, and L. Carin, "Multi-task compressive sensing with Dirichlet process priors," in *Proc. 25th Int. Conf. Mach. Learn.*, New York, NY, USA, 2008, pp. 768–775.

[12] L. Wang, L. Zhao, G. Bi, C. Wan, L. Zhang, and H. Zhang, "Novel wideband DOA estimation based on sparse Bayesian learning with Dirichlet process priors," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 275–289, Jan. 2016, doi: 10.1109/TSP.2015.2481790.

[13] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Anal.*, vol. 1, no. 1, pp. 121–143, 2006, doi: 10.1214/06-BA104.

[14] K. Kenichi, M. Welling, and Y. W. Teh, "Collapsed variational Dirichlet process mixture models," in *Proc. 20th Int. Joint Conf. Artif. Intell.*, Jan. 2007, pp. 2796–2801.

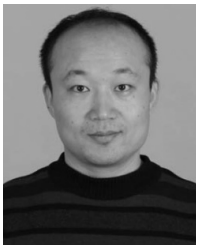
[15] A. Asuncion, M. Welling, P. Smyth, and Y. W. Teh, "On smoothing and inference for topic models," in *Proc. 25th Conf. Uncertainty Artif. Intell.*, Montreal, QC, Canada, 2009, pp. 27–34.

[16] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2004, pp. 347–352.

[17] T. P. Minka, "Estimating a Dirichlet distribution," Microsoft, Tech. Rep., 2000.

[18] N. L. Pedersen, C. N. Manchón, and B. H. Fleury, "A fast iterative Bayesian inference algorithm for sparse channel estimation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 4591–4596.

[19] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sep. 2011, doi: 10.1109/JSTSP.2011.2159773.



XINHUA LU received the B.E. and M.E. degrees from Zhengzhou University, Zhengzhou, China, in 2003 and 2007, respectively, where he is currently pursuing the Ph.D. degree. From 2015 to 2017, he was a Visiting Ph.D. Student with the Department of Electronic Systems, Aalborg University, supported by the China Scholarship Council. Since 2007, he has been a Lecturer with the College of Computer and Information Engineering, Nanyang Institute of Technology, Nanyang, China. His main research interests include massive MIMO, channel estimation, Dirichlet process, and variational Bayesian inference.

CARLES NAVARRO MANCHÓN received the degree in telecommunications engineering from the University Miguel Hernández of Elche, Spain, in 2006, and the Ph.D. degree in wireless communications from Aalborg University, Denmark, in 2011. Since 2006, he has been with Aalborg University, where he is currently an Associate Professor with the Wireless Communication Networks Section, Department of Electronic Systems. His research interest includes the area of statistical signal processing for wireless communications, including joint channel estimation and decoding algorithms, estimation and reconstruction of sparse signals, and signal processing for multiple antenna techniques.

ZHONGYONG WANG received the B.S. and M.S. degrees in automatic control from the Harbin Shipbuilding Engineering Institute, Harbin, China, in 1986 and 1988, respectively, and the Ph.D. degree in automatic control theory and application from Xi'an Jiaotong University, Xi'an, China, in 1998. Since 1988, he has been a Lecturer with the Department of Electronics, Zhengzhou University, Zhengzhou, China. From 1999 to 2002, he was an Associate Professor, and in 2002, he was promoted to Professor with the Department of Communication Engineering. His general fields of interests include numerous aspects within embedded systems, signal processing, and communication theory.