2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) | 979-8-3503-6224-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/PIMRC59610.2024.10817353

Joint Activity Detection and Channel Estimation for MIMO Grant-Free Random Access through Bayesian Learning

Boran Yang*, Xiaoxu Zhang*, Li Hao*, George K. Karagiannidis†, and Pingzhi Fan*
*Southwest Jiaotong University, Chengdu, China
†Aristotle University of Thessaloniki, Thessaloniki, Greece and Lebanese American University, Beirut, Lebanon
Email: bryang@my.swjtu.edu.cn

***Abstract*—Massive machine-type communications (mMTC) are anticipated to be supported by grant-free random access (GF-RA) and multiple-input multiple-output (MIMO) techniques. Compressed sensing (CS) is extensively advocated to accommodate massive connectivity due to bursty data transmission. In this paper, we formulate the joint activity detection and channel estimation for the MIMO-enabled GF-RA system as a block single measurement vector (SMV) problem. First, the hierarchical block sparse Bayesian learning (BSBL) framework is developed to solve this problem, where the potential block sparse properties induced by multiantenna reception are exploited by assigning the structured hyperprior. Then, by integrating the generalized approximate message passing (GAMP) approach into the BSBL formulation to effectively approximate the posterior distribution, we propose a computationally efficient Bayesian learning algorithm named GAMP-BSBL. Fortunately, the proposed Bayesian algorithms enable automatic learning of block sparse solutions without requiring noise level and user sparsity ratio as explicit conditions. Simulation results show that the proposed algorithms provide improved performance gains as compared with the standard CS methods.**

***Index Terms*—Grant-free random access, compressed sensing, block single measurement vector, block sparse Bayesian learning, generalized approximate message passing.**

I. Introduction

Massive machine-type communications (mMTC) are predicted to accommodate widespread and unparalleled connectivity for cellular Internet of Things (IoT) applications [1]. Unlike human-type communications, the main characteristics of mMTC include large-scale users, short-time transmissions, and sporadic data traffic. Therefore, how to design effective and reliable wireless access techniques to support mMTC is an open-ended topic [2]. As a branch of wireless access techniques, the standard request-grant random access agreements were commonly used in 5G new radio (NR) [3]. However, for massive connectivity, the tedious handshake interaction between users and base station (BS) is highly susceptible to signaling cost and end-to-end delay. To remove this issue, the grant-free random access (GF-RA) solutions have gained great attention, since it authorizes the active users to access the BS directly without the need for cumbersome scheduling procedures [4]. Intuitively, the key interests of GF-RA are which users are active and how their channel state information is available at any given time. However, due to sporadic traffic transmission, active user identification and channel estimation are typically transformed into the sparse signal reconstruction paradigm and appropriately tackled via compressed sensing (CS) algorithms [5]–[7].

Recently, well-established greedy CS methods, such as the matching pursuit (MP) and its variants [8]–[10], have been introduced to achieve massive connectivity in uplink GF-RA systems. On the other hand, Bayesian-type algorithms, e.g., approximate message passing (AMP) [11] and sparse Bayesian learning (SBL) [12], have been developed to perform the parallel tasks of user identification and channel estimation. Furthermore, our previous work [13] proposed a low-complexity SBL method to circumvent high-dimensional matrix inversion by using the generalized approximate message passing technique. However, the aforementioned studies [8]–[12] failed to utilize multiple-input multiple-output (MIMO) techniques to further enhance the massive detection performance. Inspired by the above observations, the researchers in [14] developed an AMP-based multiple measurement vector (AMP-MMV) algorithm for GF-RA systems with multiantenna reception. In addition, block greedy CS approaches, such as the block MP and its extensions [15]–[17], have been proposed to enable massive random access. However, the established works [14]–[17] necessitated precise insights into user activity probability or noise level, which is impractical for mMTC with occasional traffic characteristics.

In this paper, the joint activity detection and channel estimation (JADCE) in MIMO-enabled GF-RA system is transformed into the form of block sparse signal reconstruction. For this purpose, two sparse detection algorithms based on Bayesian learning are developed to support massive connectivity. The main contributions of this paper are summarised as follows:

- First, we introduce the block sparse Bayesian learning (BSBL) [18] framework to capture the underlying structured sparsity properties, where the sparse pattern of each block is dominated by a unique hyperparameter. Different from

This work was partially funded by the Sichuan Science and Technology Program (No. 2022NSFSC0910), and in part by the National Science Foundation of China Project (No. 62020106001).

979-8-3503-6224-4/24/$31.00 ©2024 IEEE

the related BSBL algorithm, we employ a hierarchical Gaussian-Gamma prior model to promote the sparsity of the channel vector.
- Then, to overcome the complexity problem, we develop a cost-efficient Bayesian learning algorithm by integrating the generalized approximate message passing (GAMP) method [19] into the BSBL formulation. Since high-dimensional matrix inversion is bypassed, the proposed GAMP-BSBL algorithm significantly reduces the computational cost of BSBL.
- In contrast to existing studies [14]–[17], the proposed Bayesian algorithms enable automatic learning of block sparse solutions without imposing any restrictions on noise magnitude and user activity level.
- The superiority of the proposed algorithms in terms of reconstruction accuracy and cost efficiency is validated by extensive simulation experiments.

The remainder is arranged as follows. Section II presents the system model for MIMO-enabled GF-RA. Sections III and IV describe the proposed BSBL and GAMP-BSBL algorithms for JADCE. Section V shows the performance analysis. Finally, Section VI concludes the paper.

Notation: The transpose, conjugate transpose, expectation, column-wise vectorization, and Kronecker product are marked as $(\cdot)^T$, $(\cdot)^H$, $\mathrm{E}[\cdot]$, $\mathrm{vec}(\cdot)$, and $\otimes$. $\|\cdot\|_2$ indicates the l_2-norm. $\mathrm{tr}[\,\cdot\,]$ represents the matrix's trace. $\Gamma(\cdot)$ denotes the Gamma function. $\mathbf{0}_r$ denotes a vector with all zero elements and size r. $\boldsymbol{I}_r$ denotes a identity matrix of size $r \times r$. $\mathrm{diag}(\mathbf{r})$ stands for a diagonal matrix comprising the elements $\mathbf{r}$.

II. System Model

This paper considers a uplink MIMO-enabled GF-RA system containing a BS with M antennas and K one-antenna users, as illustrated in Fig. 1. As the access time slot arrives, only K_a users among K are active, i.e., $K_a \ll K$, corresponding to sporadic service traffic of mMTC. Since the total users for massive connectivity are generally greater than the finite resource blocks, it is assumed that each user is endowed with a non-orthogonal pilot $\boldsymbol{s}_k = (s_{k,1}, s_{k,2}, \cdots, s_{k,N})^T$ of length N. For simplicity, the case of synchronized reception is considered within a coherent block. Thus, the received signal of multiple antennas can be formulated as

$$\mathbf{Y} = \sum_{k=1}^{K} \boldsymbol{s}_k \boldsymbol{h}_k^T + \mathbf{W} = \mathbf{S}\mathbf{H} + \mathbf{W}, \tag{1}$$

where $\mathbf{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_M)$ with $\boldsymbol{y}_m \in \mathbb{C}^{N\times 1}$ is the received signal of the m-th antenna, $\boldsymbol{h}_k = (h_{k,1}, h_{k,2}, \cdots, h_{k,M})^T$ is the channel gain between user k and the BS, $\mathbf{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_M)$ with $\boldsymbol{w}_m \in \mathbb{C}^{N\times 1}$ is additive Gaussian noise satisfying mean $\mathbf{0}_N$ and covariance $\sigma^2 \boldsymbol{I}_N$, $\mathbf{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_K)$, and $\mathbf{H} = (\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_K)^T$. The row sparse matrix $\mathbf{H}$ exhibits the same sparse pattern in each column, as shown in Fig. 1.

To extract the underlying structured sparse properties of user activity induced by multiantenna reception, we further recombine $\mathbf{H}$ into a new sparse vector $\mathbf{h}$ by sequentially extracting each row vector of $\mathbf{H}$. Similarly, the matrices $\mathbf{Y}$, $\mathbf{S}$, and $\mathbf{W}$ need to be transformed into a suitable form to fit the reshaped profile of $\mathbf{H}$. Considering this, Equation (1) is converted to the equivalent block single measurement vector (SMV) model,

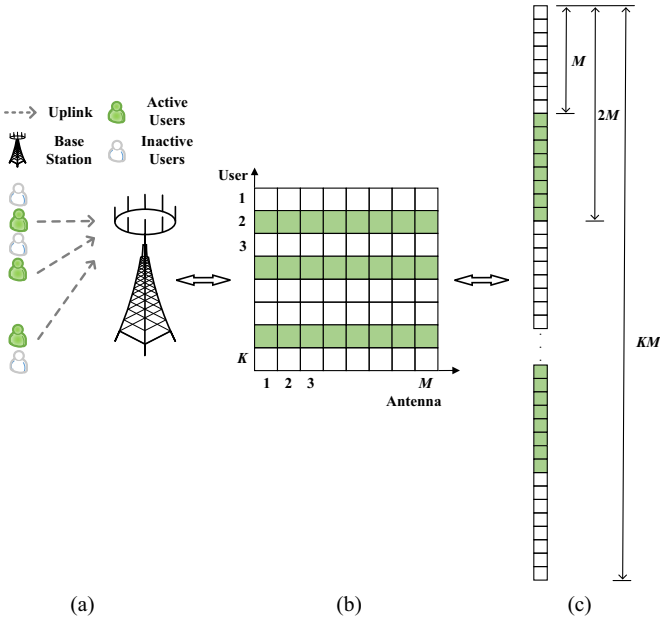


Fig. 1. A typical massive connectivity scenario and the equivalent CS models. (a) Uplink GF-RA system with multiantenna reception. (b) MMV model. (c) Block sparse SMV model.

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{w} = (\mathbf{S} \otimes \boldsymbol{I}_M)\mathbf{h} + \mathbf{w}, \tag{2}$$

where $\mathbf{y} = \mathrm{vec}(\mathbf{Y}^T)$ is the linear observation signal, $\mathbf{w} = \mathrm{vec}(\mathbf{W}^T)$ is the linear Gaussian noise, $\mathbf{h} = \mathrm{vec}(\mathbf{H}^T)$ is the linear channel vector, and $\mathbf{A} = \mathbf{S} \otimes \boldsymbol{I}_M$ is the equivalent measurement matrix. Note that $\mathbf{h}$ is a block sparse vector, i.e., each block of $\mathbf{h}$ shares the same sparse mode, as presented in Fig. 1(c). This potential structured sparsity properties can be exploited to improve the accuracy of massive user detection.

Throughout this paper, the main objective is joint active user identification and channel estimation. This involves extracting the block sparse vector $\mathbf{h}$ from the linear observation vector $\mathbf{y}$ based on the equivalent measurement matrix $\mathbf{A}$.

III. BSBL for JADCE

In the uplink GF-RA system with multiantenna reception, the sparse patterns of all coefficients within a block of $\mathbf{h}$ are statistically dependent. To extract the underlying block sparsity properties, we adopt a two-layered hierarchical prior model with a unique hyperparameter assigned to each block. The hierarchical prior structure provides improved sparsity-promoting property as compared with the standard BSBL framework [18]. More precisely, in the first layer, we assign

a multivariate Gaussian prior to $\mathbf{h}$,

$$\begin{aligned} p(\mathbf{h}\,;\boldsymbol{\alpha}) &= \prod_{k=1}^{K}\prod_{m=1}^{M}\mathcal{CN}(h_{k,m}\mid 0\,,\alpha_k^{-1}) \\ &= (2\pi)^{-KM}|\boldsymbol{\Lambda}|\exp(-\mathbf{h}^H\boldsymbol{\Lambda}\mathbf{h}), \end{aligned} \tag{3}$$

where $h_{k,m}$ is the km-th coefficient of $\mathbf{h}$, α_k is the hyperparameter of k-th block, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_K)^T$ is the hyperparameter vector used to tune the structured sparsity of $\mathbf{h}$, and $\boldsymbol{\Lambda} = \mathrm{diag}(\mathrm{vec}[(\boldsymbol{\alpha}, \boldsymbol{\alpha}, \cdots, \boldsymbol{\alpha})^T])$. It is clear that the k-th block of $\mathbf{h}$ becomes zero as the hyperparameter α_k tends to positive infinity.

In the second layer, we assign Gamma prior distributions to the hyperparameters $\boldsymbol{\alpha}$,

$$\begin{aligned} p(\boldsymbol{\alpha}) &= \prod_{k=1}^{K}\mathrm{Gamma}(\alpha_k\mid a\,,b) \\ &= \prod_{k=1}^{K}\Gamma(a)^{-1}b^a\alpha_k^{a-1}e^{-b\alpha_k}, \end{aligned} \tag{4}$$

where a and b are empirically defined parameters employed to encourage sparsity of $\mathbf{h}$.

For clarity, we denote θ as $1/\sigma^2$, which indicates the inverse precision of noise level. Furthermore, it is assumed that the noise vector $\mathbf{w}$ satisfies a common Gaussian prior,

$$p(\mathbf{w}) = \mathcal{CN}(\mathbf{w}\mid \mathbf{0}_{NM}\,,\theta^{-1}\boldsymbol{I}_{NM}). \tag{5}$$

Then, the Gamma distribution is assigned to the hyperparameter θ,

$$p(\theta) = \mathrm{Gamma}(\theta\mid c\,,d) = \Gamma(c)^{-1}d^c\theta^{c-1}e^{-d\theta}, \tag{6}$$

where c and d are empirically determined parameters. As a result, the linear observation signal $\mathbf{y}$ satisfies a Gaussian likelihood distribution,

$$p(\mathbf{y}\mid\mathbf{h}\,;\theta) = \mathcal{CN}(\mathbf{y}\mid\mathbf{A}\mathbf{h}\,,\theta^{-1}\boldsymbol{I}_{NM}). \tag{7}$$

Based on Bayes' theory, the posterior probability for $\mathbf{h}$ is given by

$$p(\mathbf{h}\mid\mathbf{y}\,;\boldsymbol{\alpha}\,,\theta) = \frac{p(\mathbf{y}\mid\mathbf{h}\,;\theta)p(\mathbf{h}\,;\boldsymbol{\alpha})}{\int p(\mathbf{y}\mid\mathbf{h}\,;\theta)p(\mathbf{h}\,;\boldsymbol{\alpha})d\mathbf{h}}. \tag{8}$$

Substituting (3) and (7) into (8), and employing some mathematical simplifications [5], it is proved that $\mathbf{h}$ approximately satisfies a Gaussian distribution,

$$p(\mathbf{h}\mid\mathbf{y}\,;\boldsymbol{\alpha}\,,\theta) = \mathcal{CN}(\mathbf{h}\mid\boldsymbol{\mu}\,,\boldsymbol{\Sigma}), \tag{9}$$

where the mean and variance of the posterior probability are given by

$$\begin{aligned} \mathbf{h}^{\mathrm{BSBL}} &= \boldsymbol{\mu} = \theta\boldsymbol{\Sigma}\mathbf{A}^H\mathbf{y}, \\ \boldsymbol{\Sigma} &= (\theta\mathbf{A}^H\mathbf{A} + \boldsymbol{\Lambda})^{-1}. \end{aligned} \tag{10}$$

The mean of posterior distribution is treated as the minimum mean square error (MMSE) estimate of $\mathbf{h}$, which depends on the estimated hyperparameters. Then, we focus on estimating the hyperparameters $\{\boldsymbol{\alpha}\,,\theta\}$. Specifically, the expectation maximization (EM) strategy [20] is performed to renew the hyperparameters.

Algorithm 1 BSBL

Initialization: $\forall k$: $\boldsymbol{\mu}^{(0)} = \mathbf{0}$, $\alpha_k^{(0)} = 1$, $\theta^{(0)} = 1$, and $t = 1$.
Step 1. Updating the MMSE estimation of the linear channel vector $\mathbf{h}$ refer to (10).
Step 2. Estimating the hyperparameters refer to (12) and (13).
Iterating the full process until $\|\boldsymbol{\mu}^{(t)} - \boldsymbol{\mu}^{(t-1)}\|_2 < \eta$, where η is the maximum permissible error of 10^{-8}.

1) E-step: In the first phase, the expectation value of the full log-likelihood distribution, also known as the Q-function, is computed,

$$\begin{aligned} &Q^{(t-1)}(\mathbf{y}\,,\mathbf{h}\mid\boldsymbol{\alpha}\,,\theta) \\ &= \sum_{k=1}^{K}\sum_{m=1}^{M}(\frac{1}{2}\ln\alpha_k - \frac{1}{2}\alpha_k\mathrm{E}[h_{k,m}^2]) \\ &\quad + \sum_{k=1}^{K}((a-1)\ln\alpha_k - b\alpha_k) + \frac{NM}{2}\ln\theta \\ &\quad - \frac{\theta}{2}\mathrm{E}[\|\mathbf{y}-\mathbf{A}\mathbf{h}\|_2^2] + (c-1)\ln\theta - d\theta. \end{aligned} \tag{11}$$

2) M-step: In the second phase, the hyperparameters are renewed by maximizing the Q-function. Extrapolating the first-order partial derivatives of (11) versus $\boldsymbol{\alpha}$ and θ, respectively, and taking zero values, we can access the update rules for α_k and θ,

$$\alpha_k^{(t)} = \frac{M + 2a - 2}{2b + \sum_{m=1}^{M}(\mu_{k,m}^2 + \Sigma_{km,km})}, \tag{12}$$

$$\theta^{(t)} = \frac{NM + 2c - 2}{2d + \|\mathbf{y}-\mathbf{A}\boldsymbol{\mu}\|_2^2 + \mathrm{tr}[\boldsymbol{\Sigma}\mathbf{A}^H\mathbf{A}]}. \tag{13}$$

The BSBL algorithm is summarized in Algorithm 1.

IV. GAMP-BSBL FOR JADCE

The BSBL algorithm enables automatic learning of block sparse solutions without the fundamental information of noise level and user sparsity ratio. However, the computation of the posterior distribution of $\mathbf{h}$ incorporates the inverse operation of $KM \times KM$ matrix. The BSBL algorithm has a complexity order of $\mathcal{O}(K^3M^3)$, which limits its applicability especially in massive connectivity scenarios. To overcome the expensive cost, we embed the GAMP approach into the BSBL formulation to achieve the effective approximation of posterior distribution, and propose a computationally efficient alternative named GAMP-BSBL. GAMP is a refined version of the traditional belief propagation algorithm by simplifying the message passing rules at the factor and variable nodes [19].

In this case, we define $\boldsymbol{\Theta} = \{\boldsymbol{\alpha}\,,\theta\}$ as the set of hyperparameters and $\mathbf{z} = \mathbf{A}\mathbf{h}$ as the noiseless output signal. GAMP-BSBL algorithm provides a series of scalar estimates for the

unknown vectors $\mathbf{h}$ and $\mathbf{z}$. Firstly, GAMP-BSBL approximates the true posterior probability $p(h_{k,m} \mid \mathbf{y}, \hat{r}_{k,m}, \tau^r_{k,m}, \mathbf{\Theta})$ as

$$\begin{aligned} &p(h_{k,m} \mid \mathbf{y}, \hat{r}_{k,m}, \tau^r_{k,m}, \mathbf{\Theta}) \\ &= \frac{p(h_{k,m} \mid \mathbf{\Theta})\mathcal{CN}(h_{k,m} \mid \hat{r}_{k,m}, \tau^r_{k,m})}{\int p(h_{k,m} \mid \mathbf{\Theta})\mathcal{CN}(h_{k,m} \mid \hat{r}_{k,m}, \tau^r_{k,m}) dh_{k,m}}, \end{aligned} \tag{14}$$

where $\hat{r}_{k,m}$ denotes the variable of $h_{k,m}$ plus noise and $\tau^r_{k,m}$ denotes the variance of noise. Substituting (3) into (14) and making some simplifications, it can be evidenced that $p(h_{k,m} \mid \mathbf{y}, \hat{r}_{k,m}, \tau^r_{k,m}, \mathbf{\Theta})$ approximately satisfies a Gaussian distribution $\mathcal{CN}(h_{k,m} \mid \mu^h_{k,m}, \phi^h_{k,m})$ with mean and variance as follows

$$\begin{aligned} \mu^h_{k,m} &= \frac{\hat{r}_{k,m}}{\alpha_k \tau^r_{k,m} + 1}, \\ \phi^h_{k,m} &= \frac{\tau^r_{k,m}}{\alpha_k \tau^r_{k,m} + 1}. \end{aligned} \tag{15}$$

Moreover, GAMP-BSBL approximates the true posterior probability $p(z_{n,m} \mid \mathbf{y}, \hat{p}_{n,m}, \tau^p_{n,m}, \mathbf{\Theta})$ as

$$\begin{aligned} &p(z_{n,m} \mid \mathbf{y}, \hat{p}_{n,m}, \tau^p_{n,m}, \mathbf{\Theta}) \\ &= \frac{p(y_{n,m} \mid z_{n,m}, \mathbf{\Theta})\mathcal{CN}(z_{n,m} \mid \hat{p}_{n,m}, \tau^p_{n,m})}{\int p(y_{n,m} \mid z_{n,m}, \mathbf{\Theta})\mathcal{CN}(z_{n,m} \mid \hat{p}_{n,m}, \tau^p_{n,m}) dz_{n,m}}, \end{aligned} \tag{16}$$

where $z_{n,m}$ denotes the nm-th coefficient of $\mathbf{z}$, $\hat{p}_{n,m}$ denotes the variable of $z_{n,m}$ plus noise, and $\tau^p_{n,m}$ denotes the variance of noise. The Gaussian noise setting produces $p(y_{n,m} \mid z_{n,m}, \mathbf{\Theta}) = p(y_{n,m} \mid z_{n,m}, \theta^{-1})$. Hence, $p(z_{n,m} \mid \mathbf{y}, \hat{p}_{n,m}, \tau^p_{n,m}, \mathbf{\Theta})$ also approximately satisfies a Gaussian distribution $\mathcal{CN}(z_{n,m} \mid \mu^z_{n,m}, \phi^z_{n,m})$ with mean and variance as follows

$$\begin{aligned} \mu^z_{n,m} &= \frac{\tau^p_{n,m}\theta y_{n,m} + \hat{p}_{n,m}}{\theta \tau^p_{n,m} + 1}, \\ \phi^z_{n,m} &= \frac{\tau^p_{n,m}}{\theta \tau^p_{n,m} + 1}. \end{aligned} \tag{17}$$

Since the main cost of computing the posterior distribution stems from matrix-vector multiplications, the GAMP-BSBL algorithm has a complexity order of $\mathcal{O}(NKM^2)$, such that it is more suitable for massive connectivity cases. Then, the EM strategy [20] is implemented to estimate the hyperparameters by taking $\mathbf{h}$ as the hidden variables and optimizing the Q-function, i.e.,

$$\begin{aligned} &Q^{(t-1)}(\mathbf{y}, \mathbf{h} \mid \boldsymbol{\alpha}, \theta) \\ &= \mathrm{E}[\ln p(\mathbf{h}; \boldsymbol{\alpha}) p(\boldsymbol{\alpha})] + \mathrm{E}[\ln p(\mathbf{y} \mid \mathbf{h}; \theta) p(\theta)] \\ &\triangleq Q(\boldsymbol{\alpha} \mid \boldsymbol{\alpha}^{(t-1)}, \theta^{(t-1)}) + Q(\theta \mid \boldsymbol{\alpha}^{(t-1)}, \theta^{(t-1)}). \end{aligned} \tag{18}$$

For the hyperparameters $\boldsymbol{\alpha}$, we have

$$\begin{aligned} Q(\boldsymbol{\alpha} \mid \boldsymbol{\alpha}^{(t-1)}, \theta^{(t-1)}) = &\sum_{k=1}^{K}\sum_{m=1}^{M}\left(\frac{1}{2}\ln \alpha_k - \frac{1}{2}\alpha_k \mathrm{E}[h^2_{k,m}]\right) \\ &+ \sum_{k=1}^{K}((a-1)\ln \alpha_k - b\alpha_k). \end{aligned} \tag{19}$$

Algorithm 2 GAMP-BSBL

Initialization: $\forall k, m, n$: $\alpha_k^{(0)} = 1$, $\theta^{(0)} = 1$, $\mu^{h\,(0)}_{k,m} = 1$, $\phi^{h\,(0)}_{k,m} = 1$, $\hat{s}^{(0)}_{n,m} = 0$, and $t = 1$.

Step 1. $\forall n \in \{1, 2, \cdots, N\}$ and $\forall m \in \{1, 2, \cdots, M\}$:

$$z^{(t)}_{n,m} = \sum_{k=1}^{K}\sum_{m=1}^{M} A_{nm,km}\mu^{h\,(t-1)}_{k,m}$$

$$\tau^{p\,(t)}_{n,m} = \sum_{k=1}^{K}\sum_{m=1}^{M} A^2_{nm,km}\phi^{h\,(t-1)}_{k,m}$$

$$\hat{p}^{(t)}_{n,m} = z^{(t)}_{n,m} - \tau^{p\,(t)}_{n,m}\hat{s}^{(t-1)}_{n,m}$$

Step 2. $\forall n \in \{1, 2, \cdots, N\}$ and $\forall m \in \{1, 2, \cdots, M\}$:

$$\hat{s}^{(t)}_{n,m} = \frac{1}{\tau^{p\,(t)}_{n,m}}\left(\frac{\tau^{p\,(t)}_{n,m}\theta^{(t-1)}y_{n,m} + \hat{p}^{(t)}_{n,m}}{\theta^{(t-1)}\tau^{p\,(t)}_{n,m} + 1} - \hat{p}^{(t)}_{n,m}\right)$$

$$\tau^{s\,(t)}_{n,m} = \frac{\theta^{(t-1)}}{\theta^{(t-1)}\tau^{p\,(t)}_{n,m} + 1}$$

Step 3. $\forall k \in \{1, 2, \cdots, K\}$ and $\forall m \in \{1, 2, \cdots, M\}$:

$$\tau^{r\,(t)}_{k,m} = \left(\sum_{n=1}^{N}\sum_{m=1}^{M} A^2_{nm,km}\tau^{s\,(t)}_{n,m}\right)^{-1}$$

$$\hat{r}^{(t)}_{k,m} = \mu^{h\,(t-1)}_{k,m} + \tau^{r\,(t)}_{k,m}\sum_{n=1}^{N}\sum_{m=1}^{M} A_{nm,km}\hat{s}^{(t)}_{n,m}$$

Step 4. $\forall k \in \{1, 2, \cdots, K\}$ and $\forall m \in \{1, 2, \cdots, M\}$:

$$\mu^{h\,(t)}_{k,m} = \frac{\hat{r}^{(t)}_{k,m}}{\alpha_k^{(t-1)}\tau^{r\,(t)}_{k,m} + 1}$$

$$\phi^{h\,(t)}_{k,m} = \frac{\tau^{r\,(t)}_{k,m}}{\alpha_k^{(t-1)}\tau^{r\,(t)}_{k,m} + 1}$$

Step 5. Updating of α_k:

$$\alpha_k^{(t)} = \frac{M + 2a - 2}{2b + \sum_{m=1}^{M}\mathrm{E}[h^{2\,(t)}_{k,m}]}$$

Step 6. Updating of θ:

$$\theta^{(t)} = \frac{NM + 2c - 2}{2d + \sum_{n=1}^{N}\sum_{m=1}^{M}\mathrm{E}[(y_{n,m} - z^{(t)}_{n,m})^2]}$$

Iterating the full process until $\sqrt{\sum_{k=1}^{K}\sum_{m=1}^{M}\left|\mu^{h\,(t)}_{k,m} - \mu^{h\,(t-1)}_{k,m}\right|^2} < \eta$, where η is the maximum permissible error of 10^{-8}.

Deriving the first-order derivative of (19) versus α_k, and letting it be zero, we get

$$\alpha_k^{(t)} = \frac{M + 2a - 2}{2b + \sum_{m=1}^{M}\mathrm{E}[h^2_{k,m}]}. \tag{20}$$

Since $\mathcal{CN}(h_{k,m} \mid \mu^h_{k,m}, \phi^h_{k,m})$ obeys a Gaussian probability with mean $\mu^h_{k,m}$ and variance $\phi^h_{k,m}$, $\mathrm{E}[h^2_{k,m}]$ can be formulated as

$$\begin{aligned} \mathrm{E}[h^2_{k,m}] &= (\mu^h_{k,m})^2 + \phi^h_{k,m} \\ &= \left(\frac{\hat{r}_{k,m}}{\alpha_k\tau^r_{k,m} + 1}\right)^2 + \frac{\tau^r_{k,m}}{\alpha_k\tau^r_{k,m} + 1}. \end{aligned} \tag{21}$$

For the hyperparameter θ, we have

$$Q(\theta \mid \boldsymbol{\alpha}^{(t-1)}, \theta^{(t-1)}) = \frac{NM}{2} \ln \theta + (c-1) \ln \theta - d\theta - \frac{\theta}{2} \sum_{n=1}^{N} \sum_{m=1}^{M} \mathrm{E}[(y_{n,m} - z_{n,m})^2]. \tag{22}$$

Deriving the first-order derivative of (22) versus θ, and letting it be zero, we get

$$\theta^{(t)} = \frac{NM + 2c - 2}{2d + \sum_{n=1}^{N} \sum_{m=1}^{M} \mathrm{E}[(y_{n,m} - z_{n,m})^2]}. \tag{23}$$

Since $\mathcal{CN}(z_{n,m} \mid \mu^z_{n,m}, \phi^z_{n,m})$ obeys a Gaussian probability with mean $\mu^z_{n,m}$ and variance $\phi^z_{n,m}$, $\mathrm{E}[(y_{n,m} - z_{n,m})^2]$ can be given by

$$\begin{aligned} &\mathrm{E}[(y_{n,m} - z_{n,m})^2] \\ &= (y_{n,m} - \mu^z_{n,m})^2 + \phi^z_{n,m} \\ &= (y_{n,m} - \frac{\tau^p_{n,m} \theta y_{n,m} + \hat{p}_{n,m}}{\theta \tau^p_{n,m} + 1})^2 + \frac{\tau^p_{n,m}}{\theta \tau^p_{n,m} + 1}. \end{aligned} \tag{24}$$

The GAMP-BSBL algorithm is summarized in Algorithm 2.

V. Simulation Results

The performance of the developed BSBL and GAMP-BSBL algorithms are evaluated and compared with the classical SBL [12], block orthogonal MP (B-OMP) [15], block compressive sampling MP (B-CoSaMP) [16], and AMP-MMV [14] methods. The primary system parameters are listed below. The total users and pilot sequence length are set to $K = 200$ and $N = 100$. The antennas deployed by the BS are $M = 4$, which is a typical parameter for realistic scenarios. In all simulations, a flat Rayleigh fading channel model is considered, i.e., $\boldsymbol{h}_k \sim \mathcal{CN}(\mathbf{0}_M, \boldsymbol{I}_M)$. The informative parameters are $a = 1.5$ and $b = 10^{-8}$. The non-informative parameters are $c = 10^{-8}$ and $d = 10^{-8}$. Moreover, the pilot sequences are generated with a random Gaussian distribution and the simulation results are averaged over 1000 Monte Carlo runs.

To evaluate the proposed Bayesian algorithms, we define the activity detection success rate (ADSR) and the mean square error (MSE) as follows

$$\mathrm{ADSR} = \frac{\hat{K}_a}{K_a}, \tag{25}$$

$$\mathrm{MSE} = \frac{\left\| \hat{\mathbf{h}} - \mathbf{h} \right\|_2^2}{KM}, \tag{26}$$

where $\hat{K}_a$ denotes the number of successfully identified active users and $\hat{\mathbf{h}}$ denotes the estimated channel vector.

Fig. 2 shows the recovery accuracy of the considered algorithms under varying signal-to-noise ratios (SNR). Clearly, the proposed BSBL and GAMP-BSBL algorithms significantly outperform the other four counterparts, because they not only exploit the spatial correlation of user activity, but also incorporate the empirical prior distribution of channel vector. Moreover, the GAMP-BSBL algorithm exhibits similar performance to the BSBL algorithm, which coincides with the fact that the approximated posterior distribution strongly approaches the actual posterior probability.

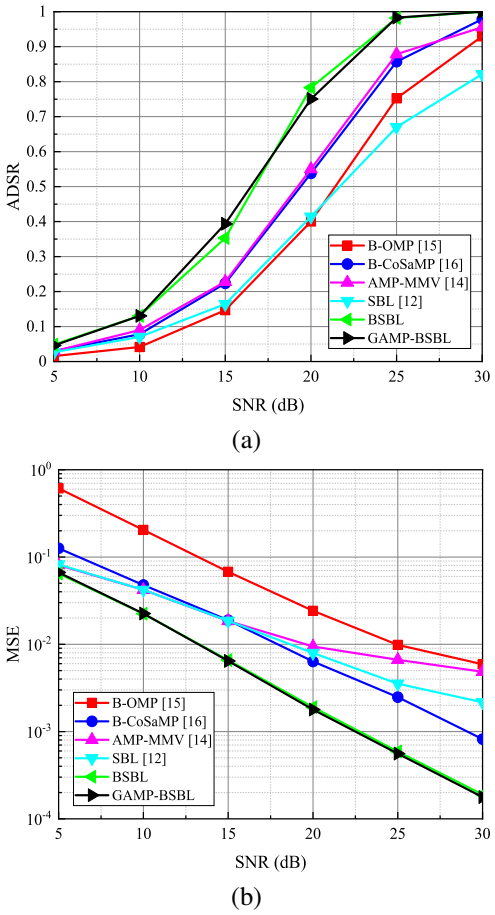
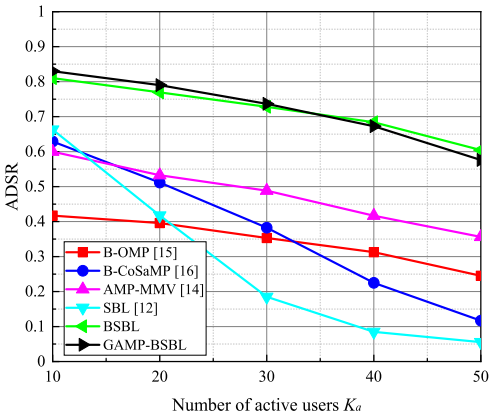


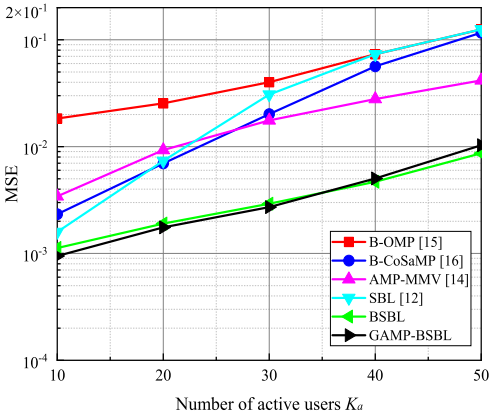
Fig. 2. Comparison of the considered algorithms versus SNR when $K_a = 20$. (a) ADSR. (b) MSE.

Fig. 3 presents the recovery accuracy of the considered algorithms under varying number of active users. The performance of all algorithms deteriorates as the user activity probability increases, but the proposed BSBL and GAMP-BSBL algorithms still outperform the remaining CS methods. Note that the B-OMP, B-CoSaMP, and AMP-MMV methods require the accurate number of active users. However, the proposed Bayesian algorithms enable automatic learning of block sparse solutions without the need for user activity factor as prior information.

Table I illustrates the number of multiplications and the running time of the considered algorithms. Specifically, we record the computational runtime used to perform one iteration when the user sparsity level is $K_a = 20$. It is obvious that SBL and BSBL have a higher computational complexity as compared with the B-OMP and B-CoSaMP. However, the proposed GAMP-BSBL algorithm significantly improves the



(a)



(b)

Fig. 3. Comparison of the considered algorithms versus the number of active users when SNR = 20 dB. (a) ADSR. (b) MSE.

TABLE I
COMPUTATIONAL COMPLEXITY

Algorithm	Computational cost for single iteration	Runtime (ms)
B-OMP [15]	$K+2KM+2NKM^2+NtM^2+2Nt^2M^3+t^3M^3$	4.592
B-CoSaMP [16]	$3K+4KM+2NKM^2+3NK_aM^2+10NK_a^2M^3+9K_a^3M^3$	29.953
AMP-MMV [14]	$M+22K+2NM+7KM+(K+N+1)M^2+2NKM$	5.482
SBL [12]	$NM+2KM+2NKM^2+4NK^2M^3+K^3M^3$	544.332
BSBL	$NM+KM+(2N+1)KM^2+4NK^2M^3+K^3M^3$	555.967
GAMP-BSBL	$9NM+7KM+(6N+1)KM^2$	8.862

cost efficiency by circumventing the matrix inversion process. Moreover, GAMP-BSBL achieves better detection accuracy at a slight computational cost compared to the AMP-MMV method, as presented in Table I and Fig. 2.

VI. CONCLUSIONS

In this paper, we developed two Bayesian learning algorithms for joint activity detection and channel estimation in MIMO-enabled GF-RA system. The developed BSBL and GAMP-BSBL algorithms exploit the underlying block sparse structure induced by multiantenna reception, and do not require noise level and user activity probability as explicit information. The simulation results validate the superiority of the proposed algorithms over the standard CS methods both in recovery accuracy and cost efficiency.

REFERENCES

[1] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. A. Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun*., vol. 39, no. 3, pp. 615-636, Mar. 2021.

[2] S. Zhang, Y. Cui, and W. Chen, "Joint device activity detection, channel estimation and signal detection for massive grant-free access via BiGAMP," *IEEE Trans. Signal Process*., vol. 71, pp. 1200-1215, 2023.

[3] W. Chen, H. Xiao, L. Sun, and B. Ai, "Joint activity detection and channel estimation in massive MIMO systems with angular domain enhancement," *IEEE Trans. Wireless Commun*., vol. 21, no. 5, pp. 2999-3011, May 2022.

[4] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Trans. Signal Process*., vol. 68, pp. 764-779, 2020.

[5] X. Zhang, P. Fan, J. Liu, and L. Hao, "Bayesian learning-based multiuser detection for grant-free NOMA systems," *IEEE Trans. Wireless Commun*., vol. 21, no. 8, pp. 6317-6328, Aug. 2022.

[6] Y. Mei et al., "Compressive sensing-based joint activity and data detection for grant-free massive IoT access," *IEEE Trans. Wireless Commun*., vol. 21, no. 3, pp. 1851-1869, Mar. 2022.

[7] Y. Wang, X. Zhu, E. G. Lim, Z. Wei, and Y. Jiang, "Grant-free communications with adaptive period for IIoT: Sparsity and correlationbased joint channel estimation and signal detection," *IEEE Internet Things J*., vol. 9, no. 6, pp. 4624-4638, Mar. 2022.

[8] B. Wang, L. Dai, Y. Zhang, T. Mir, and J. Li, "Dynamic compressive sensing-based multi-user detection for uplink grant-free NOMA," *IEEE Commun. Lett*., vol. 20, no. 11, pp. 2320-2323, Nov. 2016.

[9] A. Alshukaili and K. A. Hamdi, "Sparse recovery algorithms implementations for short packet communications," in *Proc. IEEE Veh. Technol. Conf*, Helsinki, Finland, Jun. 2022, pp. 1-5.

[10] Y. Du *et al*., "Joint channel estimation and multiuser detection for uplink grant-free NOMA," *IEEE Wireless Commun. Lett*., vol. 7, no. 4, pp. 682-685, Aug. 2018.

[11] C. Wei, H. Liu, Z. Zhang, J. Dang, and L. Wu, "Approximate message passing-based joint user activity and data detection for NOMA," *IEEE Commun. Lett*., vol. 21, no. 3, pp. 640-643, Mar. 2017.

[12] X. Zhang, F. Labeau, L. Hao, and J. Liu, "Joint active user detection and channel estimation via Bayesian learning approaches in MTC communications," *IEEE Trans. Veh. Technol*., vol. 70, no. 6, pp. 6222-6226, Jun. 2021.

[13] X. Zhang, P. Fan, L. Hao, and X. Quan, "Generalized approximate message passing based Bayesian learning detectors for uplink grant-free NOMA," *IEEE Trans. Veh. Technol*., vol. 72, no. 11, pp. 15057-15061, Nov. 2023.

[14] L. Liu and W. Yu, "Massive connectivity with massive MIMO-Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process*., vol. 66, no. 11, pp. 2933-2946, Jun. 2018.

[15] Y. Zhang, Q. Guo, Z. Wang, J. Xi, and N. Wu, "Block sparse Bayesian learning based joint user activity detection and channel estimation for grant-free NOMA systems," *IEEE Trans. Veh. Technol*., vol. 67, no. 10, pp. 9631-9640, Oct. 2018.

[16] X. Zhang, W. Xu, Y. Cui, L. Lu, and J. Lin, "On recovery of block sparse signals via block compressive sampling matching pursuit," *IEEE Access*., vol. 7, pp. 175554-175563, 2019.

[17] L. Wu, Z. Wang, P. Sun, and Y. Yang, "Temporal correlation enhanced sparse activity detection in MIMO enabled grant-free NOMA," *IEEE Trans. Veh. Technol*., vol. 71, no. 3, pp. 2887-2899, Mar. 2022.

[18] 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.

[19] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "A GAMP-based low complexity sparse Bayesian learning algorithm," *IEEE Trans. Signal Process*., vol. 66, no. 2, pp. 294-308, Jan. 2018.

[20] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag*., vol. 25, no. 6, pp. 131-146, Nov. 2008.