Message-Passing Receiver Design for Joint Channel Estimation and Data Decoding in Uplink Grant-Free SCMA Systems

Fan Wei, Wen Chen, *Senior Member, IEEE*, Yongpeng Wu, *Senior Member, IEEE*, Jun Ma, and Theodoros A. Tsiftsis, *Senior Member, IEEE*

***Abstract*— The conventional grant-based network relies on the handshaking between the base station and active devices to achieve dynamic multi-user scheduling, which may result in large signaling overheads as well as system latency. To address those problems, a grant-free receiver design is considered in this paper based on sparse code multiple access (SCMA), one of the promising air interface technologies for 5G wireless networks. With the presence of unknown multipath fading, the proposed receiver performs joint channel estimation and data decoding without knowing the user activity in the network. Formulating a factor graph representation for the problem, we devise a message-passing receiver for the uplink SCMA that performs joint estimation iteratively. Motivated by the idea of approximate inference, we use expectation propagation to project the intractable distributions into Gaussian families such that a linear complexity decoder is obtained. The simulation results show that the proposed receiver can detect active devices in the network with a high accuracy and can achieve an improved bit-error-rate performance compared with existing methods.**

***Index Terms*— SCMA, grant-free, user activity detection, expectation propagation, joint channel estimation and data decoding.**

I. Introduction

With the explosive growing demand on network capacity, throughput and connected wireless devices, the mobile broadband network is evolved into fifth generation, in which the enhanced mobile broadband (eMBB), ultra-reliable low latency communication (URLLC), and massive machine type of communication (mMTC) are three typical application scenarios. Current air interface technologies, such as orthogonal frequency-division multiple access (OFDMA) cannot fulfill the requirements in the above scenarios as orthogonal multiple access (OMA) assigns the time-frequency resources to each user exclusively. Therefore, OMA is spectrum inefficient and cannot support large throughput as well as the massive connected users in the network. In contrast, non-orthogonal multiple access (NOMA) where each resource unit shared by multiple users is supposed to be more spectrum efficient. Sparse code multiple access (SCMA) [1] is a code domain NOMA and is designed based on the multi-dimensional sparse signal constellation. Due to the shaping gain of multi-dimensional constellation, SCMA has a better bit-error-rate (BER) performance compared with other NOMA schemes, such as low density signature (LDS) [2].

Manuscript received January 29, 2018; revised July 12, 2018; accepted October 22, 2018. Date of publication November 6, 2018; date of current version January 8, 2019. This work was supported in part by NSF, China, under Grant 61671294 and Grant 61701301, in part by the National Major Project under Grant 2017ZX03001002-005 and Grant 2018ZX03001009-002, in part by Shanghai Kewei under Grant 616JC1402900 and Grant 17510740700, in part by NSF Guangxi under Grant 2015GXNSFDA139003, and in part by the Guangxi Key Laboratory of Automatic Detecting Technology and Instruments under Grant YQ14115. This paper was presented in part at the IEEE GLOBECOM Workshops, Singapore, 2017. The associate editor coordinating the review of this paper and approving it for publication was R. Tandon. *(Corresponding author: Wen Chen).*

F. Wei, W. Chen, and Y. Wu are with the Shanghai Institute of Advanced Communications and Data Sciences, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: weifan89@sjtu.edu.cn; wenchen@sjtu.edu.cn; yongpeng.wu@sjtu.edu.cn).

J. Ma is with the School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guangxi 541004, China (e-mail: majun@guet.edu.cn).

T. A. Tsiftsis is with the School of Electrical and Information Engineering, Jinan University, Zhuhai 519070, China (e-mail: theo_tsiftsis@jnu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2018.2878571

A. Technical Literature Review

The multi-dimensional sparse constellation forms the codebook for SCMA. In [1], SCMA codebooks were heuristically designed based on the Cartesian product of quadrature amplitude modulation (QAM) symbols, and followed by a unitary rotation to achieve the signal space diversity. Later, constellation rotation and interleaving was introduced for the design of multi-dimensional codebooks in order to improve the minimum Euclidean distance between SCMA codewords and combat the multipath fading [3]. To lower the peak to average power ratio (PAPR), spherical codes [4] were used to design SCMA codebooks due to the constant energy of the codes. In addition to the multi-dimensional constellation construction, factor graph matrix design was considered in [5] to maximize the average sum rate for uplink SCMA.

On the receiver, multiuser detector based on message-passing algorithm (MPA) was used to decode the data. In spite of the sparsity of SCMA codewords, the decoding complexity grows exponentially with the number of collision users in each dimension. Consequently, low complexity SCMA decoder design is of practical interests. In [6], constellation with low number of projections

is constructed by searching the specific unitary rotation matrices. With a reduced number of projections in each dimension, the decoding complexity for higher order SCMA codebooks is thus reduced. Apart from constellation design, Chen *et al.* [7] proposed Monte-Carlo Markov-Chain method for SCMA decoding, in which the log-likelihood ratio (LLR) of the coded bits are computed by Gibbs sampling. The Gibbs sampling has a linear complexity but also results in low convergence rate. Later, sphere decoder [8], [9] was introduced in SCMA. By tightening the search scope, sphere decoder avoids exhaustive search for the possible transmitted points in each iteration. Low complexity decoder using expectation propagation (EP) was another alternative for SCMA decoding [10]. In EP, intractable distributions are projected into Gaussian distributions so that the cumbersome computing can be simplified. Moreover, low complexity hardware implementation of SCMA decoder using stochastic computing was discussed in [11].

B. Motivations

In the above works, perfect channel state information (CSI) is assumed at receiver and the users in the network are supposed to be active simultaneously. However, this may not be true in practical scenarios. Firstly, channel acquisition and tracking are always necessary and the receiver cannot obtain a perfect CSI usually. In addition, the distribution of active users in the network is always sparse in practice. In fact, according to the mobile traffic statistics [12], the ratio of simultaneous active users in a wireless network does not exceed to 10% even in the busy hours. As such, the base station needs to identify active users in the system before decoding their data. In LTE, the dynamic user scheduling is achieved through a request-grant procedure. However, the handshaking between base station and active users will cost the large signaling overheads as well as system latency, see [13] for the detailed discussions. In order to reduce the signaling overheads and latency, multiple access with grant-free is a promising technique in the next generation wireless networks. In [13], the time-frequency resource referred as contention transmission unit (CTU) was defined for the uplink grant-free SCMA. Later, proof of concept (PoC) had been conducted to verify the feasibility and effectiveness of the grant-free SCMA in the user-centric no-cell (UCNC) system [14]. The active users identification and data detection for other NOMA schemes (LDS, NOMA and etc.) were studied in [15]–[17] based on orthogonal matching pursuit, compressive sampling matching pursuit, and approximate message-passing, respectively.

C. Contributions

In this paper, we focus on the design of an iterative message-passing receiver for uplink SCMA that performs joint channel estimation, data decoding, and active users detection. Iterative receivers for joint channel estimation and data decoding has been studied in [18]–[20] for MIMO-OFDM systems with the assumption that receivers have a perfect knowledge of user activity in the network. Factor graph (FG) and belief propagation (BP) (or message-passing) are two efficient tools to address various practical algorithms [21]–[23], such as forward/backward algorithm, the Viterbi algorithm, the iterative turbo decoding algorithm, Kalman filter, and so on. In [18], by exploring the FG structure of receivers and merging the belief propagation (BP) together with mean field (MF) theory, the BP-MF algorithm for joint channel estimation and data detection was proposed in MIMO-OFDM system. Gaussian approximation with BP (BP-GA) was considered as another approach in (massive) MIMO-OFDM system via central-limit theorem and moment matching [19], [20].

In this paper, by formulating the factor graph of SCMA, we proposed a message-passing receiver by using the belief propagation [21]. As the joint detection of CSI and data symbols involves a mixture of continuous and discrete variables, a compact form of BP updating rule is unavailable. Using the idea of approximate inference, we approximate intractable distributions involved in BP to Gaussian with the minimized Kullback-Leibler (KL) divergence [24]. Unlike [20], the approximation is proceeded for each user individually based on expectation propagation (EP) message-passing [25]. For the active user detection, the CIRs for each user are modeled with the Student's t-distribution and we extract the sparse signals by variational Bayesian (VB) inference. The *maximum likelihood* (ML) estimation with pilot signals are used as the initialization for the algorithms.

In the remainder of this paper, we introduce the system model in Section II, where the factor graph is constructed for iterative message-passing receiver. In Section III-A, we consider the ML estimate of CIRs based on pilot signals only while in Section III-B, the joint channel estimation and data decoding for grant-free SCMA is discussed. The performance of our proposed receiver is evaluated in Section IV and the final conclusion is given in Section V.

Notations: Lowercase letters x, bold lowercase letters $\mathbf{x}$, and bold uppercase letters $\mathbf{X}$ denote scalars, column vectors, and matrices, respectively. We use $(\cdot)^*$ and $(\cdot)^T$ to denote complex conjugate and matrix transpose, respectively. $\mathcal{CN}(x;\tau,\upsilon)$ denotes the complex Gaussian distribution with mean τ and variance υ and $\delta(\cdot)$ denotes the Dirac delta function. Notations $\mathrm{diag}(\cdot)$ and $\mathrm{Bdiag}(\cdot)$ are used to denote the diagonal matrix and block diagonal matrices respectively. The notation $\xi \setminus k$ means the set ξ with element k being excluded and $\langle f(x)\rangle_{g(x)} = E_{g(x)}\{f(x)\}$.

II. System Model

A. SCMA Uplink Grant-Free Transmission

We consider an uplink grant-free SCMA system with K potential users in the network. The users can be active or inactive depending on the service requirements. In Fig. 1, the block diagram for uplink grant-free SCMA system is illustrated. For the active user k, an information stream $\mathbf{b}_k$ is firstly sent to the channel encoder, in which the coded bits $\mathbf{c}_k$ are generated. The SCMA encoder maps the coded bits into multi-dimensional SCMA codeword $\mathbf{x}_k$, where $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{Nk})^T$ is an N-dimensional signal constellation point essentially. The constellation contains M distinct multi-dimensional signal points, and each point is designed to be sparse such that only

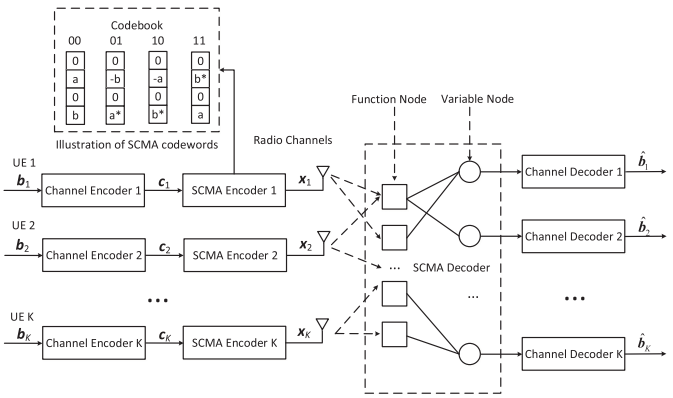


Fig. 1. Block diagram for uplink SCMA systems.

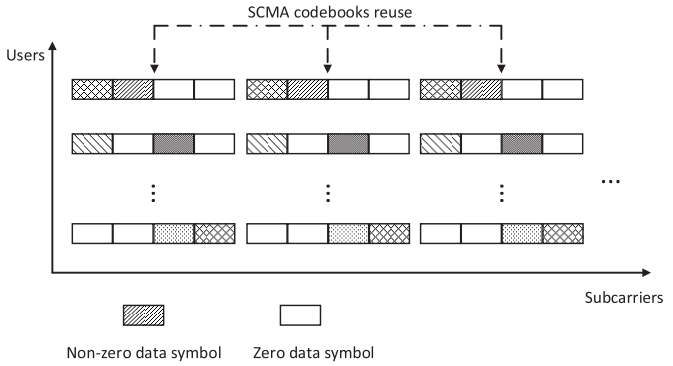


Fig. 2. Illustration of SCMA codebook reuse.

$d_v \ll N$ dimensions are used to transmit non-zero symbols while the remaining ones are set to be zeros. Each non-zero symbol is modulated into one OFDMA subcarrier shared by other users in the systems. Due to the sparsity of SCMA codewords, the number of overlapped users in each subcarrier equals $d_c \ll K$. Thus mitigated multiuser interferences are achieved.

With the identical codebook, SCMA receiver is able to detect multiple data streams as long as the codebook goes through different wireless channels. Specifically, each codebook occupies N subcarriers, which is referred as one *subcarrier block* in this paper and the codebook can be reused [13] in different *subcarrier blocks*. In Fig. 2, we illustrate the idea of codebook reuse, where X-axis denotes OFDMA subcarriers while Y-axis denotes the potential users in the system. Within this example, one subcarrier block contains 4 OFDMA subcarriers and is reused in more than 3 different subcarrier blocks. We assume that, in the subsequent of this paper, the SCMA codebook is reused in B subcarrier blocks for each user, and simultaneous B data streams can be transmitted consequently.

At the base station, the received signal at time slot t on the nth subcarrier can be written as

$$y_{tn} = \sum_{k=1}^{K} \alpha_{nk} x_{tnk} + z_n, \tag{1}$$

where x_{tnk} are the transmitted signals from user k which comprise both known pilot symbols and unknown SCMA codewords. In the time domain, the unknown channel coefficient $\alpha_k(\tau)$ follows the truncated tapped-delay [26, Sec. 13.5] model of length L,

$$\alpha_k(\tau) = \sum_{l=1}^{L} h_{kl}\delta[\tau - l/(BN)], \tag{2}$$

where $\delta(\cdot)$ is the Dirac delta function and h_{kl} denotes the l^{th} channel tap for user k, BN is the bandwidth of system. By Fourier transformation, the frequency response on the n^{th} subcarrier for user k can be written as

$$\alpha_{nk} = \sum_{l=1}^{L} h_{kl} \exp\big\{ -j2\pi nl/(BN)\big\}. \tag{3}$$

Finally, z_n is the additive complex Gaussian white noise with distribution $\mathcal{CN}(0, \sigma^2)$.

Writing the signals from N subcarriers in a matrix form, the received signal in one subcarrier block is given by

$$\mathbf{y}_t = \sum_{k=1}^{K} \mathrm{diag}(\mathbf{x}_{tk})\mathbf{F}\mathbf{h}_k + \mathbf{z}, \tag{4}$$

where $\mathbf{y}_t = (y_{t1}, y_{t2}, \ldots, y_{tN})^T$, $\mathbf{z} = (z_1, z_2, \ldots, z_N)^T$, and $\mathbf{h}_k = (h_{k1}, h_{k2}, \ldots, h_{kL})^T$. Matrix $\mathbf{F}$ is the $N \times L$ discrete Fourier transformation (DFT) where the $(n, l)^{th}$ entry $\mathrm{F}_{nl} = \exp\{-j2\pi nl/(BN)\}$. In each time slot, the transmitted codewords $\mathbf{x}_{tk} = (x_{t1k}, x_{t2k}, \ldots, x_{tNk})^T$ are chosen randomly from a predefined alphabet set $\mathcal{X}$ with $|\mathcal{X}| = M$. As the user activity detection is also an interest in this paper, we further introduce an augmented alphabet set $\mathcal{X}^+ = \mathcal{X} \cup \{0\}$ with size $|\mathcal{X}^+| = M + 1$, since an inactive user is thought of transmitting the all-zero codeword $\mathbf{x}_{tk} = \mathbf{0}$ equivalently.

B. Factor Graph Representation

From (4), the joint probability density function (PDF) of the variables is given by $p(\mathbf{C}, \mathbf{X}, \mathbf{H}, \mathbf{y})$ (time index t is dropped here for notation simplification), where $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K]$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K]$, and $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]$ are the collections of coded bits, SCMA codewords, and channel impulse responses (CIRs) from K users, respectively.

Based on the observation signal $\mathbf{y}$ as well as the pilot symbols, the SCMA decoder tries to find the *maximum a posteriori* (MAP) estimation for each coded bit c_{kl},

$$\hat{c}_{kl} = \arg\max p(c_{kl}|\mathbf{y}), \tag{5}$$

where c_{kl} is the l^{th} coded bits for user k and $p(c_{kl}|\mathbf{y})$ is given by

$$p(c_{kl}|\mathbf{y}) \propto \sum_{\mathbf{C}\setminus c_{kl}, \mathbf{X}} \int p(\mathbf{C}, \mathbf{X}, \mathbf{H}, \mathbf{y}) d\mathbf{H}. \tag{6}$$

The direct computation of (6) involves the multiple integration of continuous variables $\mathbf{H}$, and marginalization of discrete variables $\mathbf{C}$ and $\mathbf{X}$ which is prohibitively complex for large number of potential user K. In what follows, we formulate the factor graph representation of $p(\mathbf{C}, \mathbf{X}, \mathbf{H}, \mathbf{y})$ so that the probability of each variable can be calculated in a low complexity iterative way.

Based on the observation that $\mathbf{C} \rightarrow \mathbf{X} \rightarrow \mathbf{y}$ forms a Markov chain and CIRs $\mathbf{H}$ are independent of $\mathbf{C}$ and $\mathbf{X}$, the joint PDF can be factorized as follows

$$p(\mathbf{C}, \mathbf{X}, \mathbf{H}, \mathbf{y}) = p(\mathbf{C})p(\mathbf{X}|\mathbf{C})p(\mathbf{y}|\mathbf{X}, \mathbf{H})p(\mathbf{H}). \tag{7}$$

In (7), $p(\mathbf{C})$ denotes the *a priori* distribution of coded bits and $p(\mathbf{X}|\mathbf{C})$ is given by

$$p(\mathbf{X}|\mathbf{C}) = \prod_{k} p(\mathbf{x}_k|\mathbf{c}_k), \tag{8}$$

where each term $p(\mathbf{x}_k|\mathbf{c}_k)$ is the predefined mapping rule from coded bits to SCMA codewords, Gray mapping is preferred usually. Based on (4), the likelihood function $p(\mathbf{y}|\mathbf{X}, \mathbf{H})$ can be written as

$$p(\mathbf{y}|\mathbf{X}, \mathbf{H}) \propto \exp\left\{-\frac{1}{\sigma^2}\Big|\mathbf{y} - \sum_{k=1}^{K} \mathrm{diag}(\mathbf{x}_{tk})\mathbf{F}\mathbf{h}_k\Big|^2\right\}. \tag{9}$$

For channel $\mathbf{H}$, a hierarchical non-stationary zero mean complex Gaussian *priori* distribution is assumed in this paper in order to adapt the sparse signals [27]. Due to the independence of CIRs for different users and different taps, we have

$$\begin{aligned} p(\mathbf{H}, \mathbf{\Gamma}) &= \prod_{k} q(\mathbf{h}_k|\boldsymbol{\lambda}_k) \\ &= \prod_{k}\prod_{l} \mathcal{CN}(h_{kl}; 0, \lambda_{kl}^{-1}), \end{aligned} \tag{10}$$

where $\mathbf{\Gamma} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_K)$ and λ_{kl} is the precision parameter, and is modeled by Gamma distribution

$$p(\lambda_{kl}; a, b) = \mathrm{Gamma}(\lambda_{kl}|a, b). \tag{11}$$

After integrating the variable λ_{kl}, $p(h_{kl}; a, b)$ can be written as

$$\begin{aligned} p(h_{kl}; a, b) &= \int q(h_{kl}|\lambda_{kl})p(\lambda_{kl}; a, b)d\lambda_{kl} \\ &= \mathrm{St}(h_{kl}; \mu, \nu, \gamma), \end{aligned} \tag{12}$$

where the Student's t-distribution $\mathrm{St}(h_{kl}; \mu, \nu, \gamma)$ is given by

$$\begin{aligned} \mathrm{St}(h_{kl}; \mu, \nu, \gamma) = &\frac{\Gamma(\frac{\upsilon}{2}+1)}{\Gamma(\frac{\upsilon}{2})}\frac{2\gamma}{\pi\upsilon} \\ &\times \left[1 + \frac{2\gamma}{\upsilon}\big|h_{kl} - \mu\big|^2\right]^{-(\frac{\upsilon}{2}+1)}, \end{aligned} \tag{13}$$

with $\mu = 0$, $\gamma = \frac{a}{b}$, and $\upsilon = 2a$. In practice, a non-informative *priori* for a and b is assumed and we choose $a = 10^{-7}$ and $b = 10^{-7}$ in this paper. The Student's t-distribution exhibits heavy tails, which makes h_{kl} favour sparse solution such that for the inactive user, most of h_{kl}s in $\mathbf{H}$ are near zero values [27]. Notice that the distribution of the active users in a network is sparse and an inactive user is equivalent to have zero CIRs.

As the joint distribution can be factorized into several parts in (7), we formulate the factor graph representation of SCMA system in Fig. 3, where each part in (7) is represented with different square nodes for functions (e.g., f_{tn} or φ_{nk}) and circular nodes for variables (e.g., h_{kl}), respectively. The circular nodes are connected to the square nodes only if they are functions of the latter. For instance, given λ_{kl} follows $\mathrm{Gamma}(\lambda_{kl}|a, b)$, three circular nodes (i.e., λ_{kl}, a, b) are connected to each function node $p(\lambda_{kl})$. Notice that due to the sparse structure of SCMA codewords, each user choose d_v subcarriers to transmit and only d_c users are collided in one subcarrier. We define F_n to be the set of collision users in subcarrier n and V_k to be the set of subcarriers for user k to transmit subsequently. The received signal in (1) can thus be rewritten as

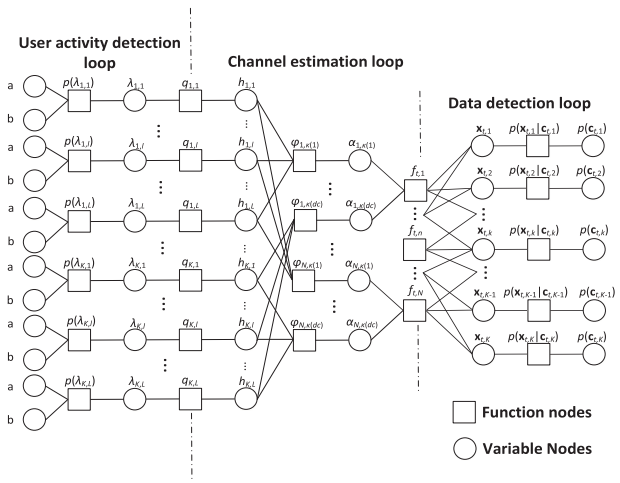


Fig. 3. Factor graph representation of the SCMA system.

$$y_{tn} = \sum_{k \in \mathrm{F}_n} \alpha_{nk} x_{tnk} + z_n. \tag{14}$$

The likelihood function $f_{t,n}$ in subcarrier n is a consequence of (14). Specifically, not all but d_c data symbols and channel coefficients are connected to each function node $f_{t,n}$ in Fig. 3, where $\kappa(\cdot)$ is used to denote the mapping from the user index within one subcarrier to the user index in the system. The function node φ_{nk} is formed in a similar approach as a consequence of (3). In what follows, we divide the factor graph into three loops for data detection, channel estimation, and active users detection, respectively.

III. Blind Detection for Grant-Free SCMA

Based on the factor graph formulation in the previous section, we develop an iterative message-passing receiver for the grant-free SCMA in this section. The algorithm is initialized with a pilot based channel estimation first. Thereafter, a low complexity iterative message-passing receiver that performs joint channel estimation, data decoding, and user activity detection is proposed.

A. Pilot Signals Based Channel Estimation

In this subsection, we consider the channel estimation problem for uplink SCMA using the pilot signals received from B subcarrier blocks. Assume that on the b^{th} subcarrier block, the received signals at time slot t can be written as

$$\mathbf{y}_{b,t} = \sum_{k} \mathbf{X}_{b,tk}^{p} \mathbf{F}_b \mathbf{h}_k + \mathbf{n}_b, \tag{15}$$

where $\mathbf{X}_{b,tk}^{p} = \mathrm{diag}(\mathbf{x}_{b,tk}^{p})$, and $\mathbf{x}_{b,tk}^{p}$ is the transmitted pilots symbols that have the same sparse structure as the SCMA codewords, and $\mathbf{h}_k$ is the CIRs for user k.

Let $\mathbf{y}_t = [\mathbf{y}_{1,t}^T, \mathbf{y}_{2,t}^T, \ldots, \mathbf{y}_{B,t}^T]^T$ be the collection of received signals from B subcarrier blocks, we have

$$\mathbf{y}_t = \sum_k \text{Bdiag}\{\mathbf{X}_{1,tk}^p, \mathbf{X}_{2,tk}^p, \ldots, \mathbf{X}_{B,tk}^p\}\check{\mathbf{F}}\mathbf{h}_k + \mathbf{n}, \quad (16)$$

where $\check{\mathbf{F}} = [\mathbf{F}_1^T, \mathbf{F}_2^T, \ldots, \mathbf{F}_B^T]^T$ is the $BN \times L$ DFT matrix. In another form, if we stack the CIRs from K users into a $KL \times 1$ column vector $\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \ldots, \mathbf{h}_K^T]^T$, the received signal $\mathbf{y}_t$ can be rewritten as

$$\begin{aligned}\mathbf{y}_t &= \begin{bmatrix} \mathbf{X}_{1,t}^p & & & \\ & \mathbf{X}_{2,t}^p & & \\ & & \ddots & \\ & & & \mathbf{X}_{B,t}^p \end{bmatrix} \begin{bmatrix} \text{Bdiag}(\mathbf{F}_1) \\ \text{Bdiag}(\mathbf{F}_2) \\ \cdots \\ \text{Bdiag}(\mathbf{F}_B) \end{bmatrix} \mathbf{h} + \mathbf{n} \\ &= \mathbf{X}_t\dot{\mathbf{F}}\mathbf{h} + \mathbf{n}, \end{aligned} \quad (17)$$

where $\mathbf{X}_{b,t}^p = [\mathbf{X}_{b,t1}^p, \mathbf{X}_{b,t2}^p, \ldots, \mathbf{X}_{b,tK}^p]$ is the $N \times KN$ pilot matrix and $\text{Bdiag}(\mathbf{F}_b)$ is the $KN \times KL$ block diagonal matrix with the matrix on the principal diagonal being $\mathbf{F}_b$ for the b^{th} subcarrier block, respectively.

Based on $\mathbf{y}_t$, the *maximum likelihood* (ML) estimation of CIRs $\mathbf{h}$ is given by

$$\begin{aligned}\hat{\mathbf{h}}_{ML} &= \arg\max \prod_t p(\mathbf{y}_t|\mathbf{h}) \\ &= \arg\min \sum_t \|\mathbf{y}_t - \mathbf{X}_t\dot{\mathbf{F}}\mathbf{h}\|^2 \\ &= \mathbf{G}^{-1}\mathbf{u}, \end{aligned} \quad (18)$$

where $\mathbf{G} = \dot{\mathbf{F}}^H(\sum_t \mathbf{X}_t^H\mathbf{X}_t)\dot{\mathbf{F}}$ and $\mathbf{u} = \dot{\mathbf{F}}^H\sum_t \mathbf{X}_t^H\mathbf{y}_t$. The production or summation is over all time slots that the block fading channels remain constant. The dimension of $\mathbf{G}$ equals $KL \times KL$. While the matrix inversion can be computed in an off-line manner, the multiplication of (18) has a computational complexity of $\mathcal{O}(K^2L^2)$, which grows quadratic with the number of potential users K. We further note that a low complexity computing for (18) is possible if the pilot symbols from collision users are designed to be orthogonal in each subcarrier (e.g., using the orthogonal Zadoff-Chu (ZC) sequence [28]. Due to the sparse structure of SCMA codewords, the number of required orthogonal pilots in one subcarrier is d_c, which is much smaller compared to the total number of potential users K).

In (18), the summation of the pilot symbols can be written as a block diagonal matrix,

$$\begin{aligned}\sum_t \mathbf{X}_t^H\mathbf{X}_t = \text{Bdiag}\Bigg\{ &\sum_t (\mathbf{X}_{1,t}^p)^H\mathbf{X}_{1,t}^p, \\ &\sum_t (\mathbf{X}_{2,t}^p)^H\mathbf{X}_{2,t}^p, \ldots, \sum_t (\mathbf{X}_{B,t}^p)^H\mathbf{X}_{B,t}^p \Bigg\}. \end{aligned} \quad (19)$$

If orthogonal pilot symbols are used such that for each subcarrier n (the requirement can be met for ZC sequence when the pilots numbers N_p is equal or greater than d_c),

$$\sum_t (x_{b,tni}^p)^* x_{b,tnj}^p = \begin{cases} 1, & i = j; \\ 0, & i \neq j, \end{cases} \quad (20)$$

$\sum_t (\mathbf{X}_{b,t}^p)^H\mathbf{X}_{b,t}^p$ is also a block diagonal matrix where the $(i,j)^{th}$ submatrix is given by

$$\sum_t (\mathbf{X}_{b,ti}^p)^H\mathbf{X}_{b,tj}^p = \begin{cases} \mathbf{E}_i, & i = j; \\ \mathbf{0}, & i \neq j. \end{cases} \quad (21)$$

In (21), $\mathbf{E}_i$ is a diagonal matrix with the entries on the principle diagonal being 1 or 0 depending on the sparse structure of SCMA codebook i, i.e., set V_i. As a consequence, matrix $\mathbf{G}$ can be rewritten as

$$\mathbf{G} = \sum_b \text{Bdiag}\{\mathbf{F}_b^H\mathbf{E}_1\mathbf{F}_b, \mathbf{F}_b^H\mathbf{E}_2\mathbf{F}_b, \ldots, \mathbf{F}_b^H\mathbf{E}_K\mathbf{F}_b\}. \quad (22)$$

The following proposition is a motivation of the sparse structure of matrix $\mathbf{G}$.

Proposition 1: A necessary condition for matrix $\mathbf{G}$ to be invertible is $Bd_v \geq L$.

Proof: Based on (22), matrix $\mathbf{G}$ reduces to a block diagonal matrix. To get the inverse matrix of $\mathbf{G}$, we must compute the inverse matrix of sub-block $\mathbf{F}_b' = \sum_b \mathbf{F}_b^H\mathbf{E}_k\mathbf{F}_b$, i.e., the k^{th} submatrix of $\mathbf{G}$. Note that the sparse matrix $\mathbf{E}_k$ has a rank of d_v and $\mathbf{F}_b'$ is an L-dimensional square matrix. As a result, the necessary condition for matrix $\mathbf{G}$ to be invertible is $\text{rank}(\mathbf{F}_b') = B\min\{L, d_v\} \geq L$. □

A heuristic understanding of this condition is that for each user, to obtain the convinced estimates for L taps of channel, one must conduct at least Bd_v independent measurements by sending the pilots symbols in Bd_v subcarriers simultaneously. Further, since the pilot signals are orthogonal in one subcarrier, each user can eliminate the interferences from other ones and obtain its own estimations. From Proposition 1, since we only need to focus on the sparse block diagonal matrix (22), the computational complexity of (18) thus reduces to $\mathcal{O}(KL^2)$, which grows linearly with the number of potential users.

B. Joint Channel Estimation and Data Decoding

In this subsection, we discuss the joint detection methods based on transmitted pilot as well as data symbols. Recall that in Section II, we have divided the factor graph of SCMA into three loops as shown in Fig 3. The message-passing rules for the three loops will be developed in the following.

1) Joint Detection for Data Decoding: We begin our discussion with the data detection loop first. To implement the reduced complexity decoding, the joint estimation shall be obtained in an iterative way instead of the direct MAP estimation in (6). For systems with the FG structure, belief propagation is an iterative algorithm that deals with the multi-variables problems. Based on the BP updating rule, the message sent from function node f_{tn} to variable node $\mathbf{x}_{tk}$ can be written as (we drop the subscript b here for notation simplification)

$$\begin{aligned} I_{f_{tn}\to\mathbf{x}_{tk}}^{(i)}(x_{tnk}) = &\sum_{\mathbf{x}_l : l\in\text{F}_n\backslash k} I_{\mathbf{x}_{tl}\to f_{tn}}^{(i-1)}(x_{tnl}) \\ &\cdot \int f_{tn}(\mathbf{X}_{tn}, \boldsymbol{\alpha}_n) \prod_{l\in\text{F}_n} I_{\alpha_{nl}\to f_{tn}}^{(i-1)}(\alpha_{nl}) d\boldsymbol{\alpha}_{\boldsymbol{n}}, \end{aligned} \quad (23)$$

where $I^{(i-1)}_{\mathbf{x}_{tl}\to f_{tn}}(x_{tnl})$ and $I^{(i-1)}_{\alpha_{nl}\to f_{tn}}(\alpha_{nl})$ are the extrinsic messages passed from nodes $\mathbf{x}_{tl}$ and α_{nl} to f_{tn} on the $(i-1)^{th}$ iteration, respectively. The likelihood function $f_{tn}(\mathbf{X}_{tn}, \boldsymbol{\alpha}_n)$ is given by

$$f_{tn}(\mathbf{X}_{tn}, \boldsymbol{\alpha}_n) \propto \exp\left\{-\frac{1}{\sigma^2}|y_{tn} - \sum_{k\in \mathrm{F}_n} \alpha_{nk}x_{tnk}|^2\right\}, \quad (24)$$

where $\mathbf{X}_{tn}$ and $\boldsymbol{\alpha}_n$ denote the transmitted data symbols and CIRs from collision users in subcarrier n, respectively.

From (23)-(24), it can be observed that the BP algorithm has reduced the computation order from K to d_c within each iteration. However, as the function node f_{tn} involves a mixture of discrete variables x_{tnk} and continuous variables α_{nk}, direct computation of (23) is still of high complexity due to the multiple integration as well as multiple marginalization.

To reduce the computational complexity of (23), the belief propagation mean field message passing is proposed based on variational Bayesian (VB) inference [18], [27]. Specifically, with BP-MF, the message-passing for x_{tnk} can be updated as

$$\begin{aligned}
&\ln I^{(i)}_{f_{tn}\to\mathbf{x}_{tk}}(x_{tnk}) \\
&\quad= \sum_{l\in \mathrm{F}_n\setminus k} b^{(i-1)}(\mathbf{x}_{tl}) \\
&\qquad\cdot \int \ln f_{tn}(\mathbf{X}_n, \boldsymbol{\alpha}_n) \prod_{l\in \mathrm{F}_n} b^{(i-1)}(\alpha_{nl}) \prod_{l\in \mathrm{F}_n} d\alpha_{nl} \\
&\quad\propto \ln \mathcal{CN}(x_{tnk}; \tau^{(i)}_{f_{tn}\to\mathbf{x}_{tk}}, \upsilon^{(i)}_{f_{tn}\to\mathbf{x}_{tk}}),
\end{aligned} \quad (25)$$

where $b^{(i-1)}(\mathbf{x}_{tl})$ and $b^{(i-1)}(\alpha_{nl})$ is the belief [21] of variable $\mathbf{x}_{tl}$ and α_{nl} on the $(i-1)^{th}$ iteration, respectively. Recall that in [21], the belief of a variable is defined as the production of all extrinsic messages to this variable. The mean $\tau^{(i)}_{f_{tn}\to\mathbf{x}_{tk}}$ and variance $\upsilon^{(i)}_{f_{tn}\to\mathbf{x}_{tk}}$ are given by

$$\tau^{(i)}_{f_{tn}\to\mathbf{x}_{tk}} = \frac{(\tau^{(i-1)}_{\alpha_{nk}})^*(y_{tn} - \sum_{l\in\mathbf{F}_n\setminus k}\tau^{(i-1)}_{\alpha_{nl}}\tau^{(i-1)}_{x_{tnl}})}{|\tau^{(i-1)}_{\alpha_{nk}}|^2 + \upsilon^{(i-1)}_{\alpha_{nk}}}, \quad (26)$$

$$\upsilon^{(i)}_{f_{tn}\to\mathbf{x}_{tk}} = \sigma^2(|\tau^{(i-1)}_{\alpha_{nk}}|^2 + \upsilon^{(i-1)}_{\alpha_{nk}})^{-1}, \quad (27)$$

where $\tau^{(i-1)}_{\alpha_{nk}}$ and $\upsilon^{(i-1)}_{\alpha_{nk}}$ are the mean and variance of α_{nk} with respect to belief $b^{(i-1)}(\alpha_{nk})$ on the $(i-1)^{th}$ iteration, and a similar definition holds for $\tau^{(i-1)}_{x_{tnk}}$.

Compared with BP updating, MF has a simple updating rule, in particular for conjugate-exponential models. However, for multiuser system, the interference cancellation structure in (26) only involves the mean values of interferences while the variances (i.e., the uncertainty of the estimation) are not being considered. As such, mean filed methods often perform poor in the estimation of LLRs for data symbols. Moreover, accurate initialization is often required for BP-MF otherwise it will trap in a local optimal point. As an alternative approach, the interferences are modeled as Gaussian distributions based on central-limit theorem [19], [20]. While central-limit theorem is effective in the massive MIMO-OFDM system, it may result in large performance degradation in SCMA since the number of collision users in each subcarrier is limited ($d_c \ll K$ in practice) due to the sparse structure of SCMA codewords.

In this paper, instead of using central-limit theorem, the distribution of each interference $\alpha_{nk}x_{tnk}$ is projected into Gaussian families separately using expectation propagation message-passing. Based on the idea of divide-and-conquer, EP can be viewed as a distributed inference where the big data is partitioned into smaller pieces and local inference is performed for each one separately. Meanwhile, expectation propagation belongs to a class of approximate inference that intractable distributions are always approximated to some simpler distributions by minimizing the Kullback-Leibler (KL) divergence

$$D_{KL}(p(x)||q(x)) = \int p(x)\log\frac{p(x)}{q(x)}dx, \quad (28)$$

where $p(x)$ is the true distribution and $q(x)$ is the approximated distribution.

Proposition 2: Let $u_{tnk} = \alpha_{nk}x_{tnk}$, given the extrinsic messages $I^{(i)}_{\mathbf{x}_{tk}\to f_{tn}}(x_{tnk})$ obtained on the previous step and the Gaussian distributed message $I^{(i-1)}_{\alpha_{nk}\to f_{tn}}(\alpha_{nk}) \sim \mathcal{CN}(\alpha_{nk}; \tau^{(i-1)}_{\alpha_{nk}\to f_{tn}}, \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}})$ (this will be justified later in (72)) computed on the previous iteration, the PDF of variable u_{tnk} can be written by (31), shown at the bottom of next page.

Proof: To get the distribution for message $I^{(i)}_{u_{tnk}\to f_{tn}}(u_{tnk})$ in (31), we first compute the cumulative probability distribution (CDF) of u_{tnk} conditioned on some given α_{nk}, *i.e.*,

$$\begin{aligned}
&p(\alpha_{nk}x_{tnk} \le u_{tnk}|\alpha_{nk}) \\
&= \langle p(\alpha_{nk}x_{tnk} \le u_{tnk}|\alpha_{nk}, x_{tnk})\rangle_{I^{(i)}_{\mathbf{x}_{tk}\to f_{tnk}}(x_{tnk})} \\
&= I^{(i)}_{\mathbf{x}_{tk}\to f_{tnk}}(x_{tnk}=0)1(u_{tnk}\ge 0) \\
&\quad + \sum_{x_{tnk}\neq 0} I^{(i)}_{\mathbf{x}_{tk}\to f_{tnk}}(x_{tnk})p(\alpha_{nk}x_{tnk} \le u_{tnk}|\alpha_{nk}, x_{tnk}),
\end{aligned} \quad (29)$$

where $1(u_{tnk} \ge 0) = p(u_{tnk} \ge \alpha_{nk}x_{tnk}|\alpha_{nk}, x_{tnk} = 0)$ is the indicator function depending whether $u_{tnk} \ge 0$ or not. Due to this indicator function, the CDF is discontinuous at $u_{tnk} = 0$. Let $F(u_{tnk}|\alpha_{nk}) = p(\alpha_{nk}x_{tnk} \le u_{tnk}|\alpha_{nk})$, we have

$$\begin{aligned}
&p(u_{tnk} = 0|\alpha_{nk}) \\
&\quad= \lim_{u_{tnk}\to 0^+} F(u_{tnk}|\alpha_{nk}) - \lim_{u_{tnk}\to 0^-} F(u_{tnk}|\alpha_{nk}) \\
&\quad= I^{(i)}_{\mathbf{x}_{tk}\to f_{tnk}}(x_{tnk}=0).
\end{aligned} \quad (30)$$

Averaging with respect to α_{nk}, we obtain the case for $x_{tnk} = 0$ in (31). For $x_{tnk} \neq 0$, the distribution of u_{tnk} can be obtained in a similar way as in [29], Chapter 6. □

The factor graph for subcarrier n with virtual node u_{tnk} is illustrated in Fig. 4. Note that this figure is used to illustrate the FG structure in subcarrier n, due to the *Proposition* 2, the message $I^{(i)}_{u_{tnk}\to f_{tn}}(u_{tnk})$ is not necessarily computed by BP. As can be seen from (31), the PDF of u_{tnk} is a mixture of Gaussian functions and is discontinuous at $u_{tnk} = 0$.

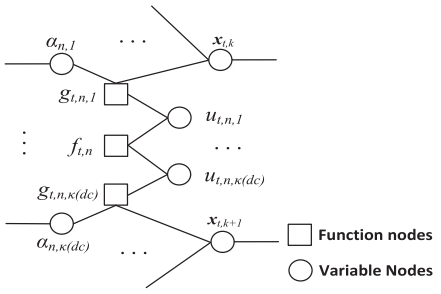


Fig. 4. Factor graph representation for subcarrier n, the virtual node is given by $u_{tnk} = g_{tnk}(x_{tnk}, \alpha_{nk}) = \alpha_{nk} x_{tnk}$.

The BP updating becomes cumbersome when such Gaussian mixtures are involved in each iteration. In order to reduce the complexity, we resort to EP message-passing to project $I^{(i)}_{u_{tnk}\to f_{tn}}(u_{tnk})$ into Gaussian distribution. Note that in most cases, the contribution to Gaussian mixture (31) comes from only one significant component, i.e., the signal point x_{tnk} that is most likely to be transmitted by the user, thus Gaussian distribution suffices to be a good approximation for (31).

Assume that the output information $I^{(i)}_{f_{tn}\to u_{tnk}}(u_{tnk})$ from function node f_{tn} follows the Gaussian distribution $\mathcal{CN}(u_{tnk}; \tau^{(i)}_{f_{tn}\to u_{tnk}}, \upsilon^{(i)}_{f_{tn}\to u_{tnk}})$ and has been computed on the previous step (the Gaussian distribution will be justified later in (42)), the belief of variable u_{tnk} can be calculated in (32), shown at the bottom of this page, where $\beta(x_{tnk})$, $\tau^{(i)}_{u_{tnk}}$ and $\upsilon^{(i)}_{u_{tnk}}$ are computed in (33)-(35), shown at the bottom of this page, and C is a normalization constant.

Notice that in (32), the belief of variable u_{tnk} comprises two pieces. Based on EP message-passing, approximate inference is proceeded for message of $I^{(i)}_{u_{tnk}\to f_{tn}}(u_{tnk})$. To start with, $b^{(i)}(u_{tnk})$ is projected into a Gaussian distribution $\hat{b}(u_{tnk})$ with the minimized KL-divergence $D_{KL}[b^{(i)}(u_{tnk})||\hat{b}(u_{tnk})]$. The result reduces to moment matching such that (see [25, Sec. 2]),

$$\hat{b}(u_{tnk}) = \mathcal{CN}(u_{tnk}; \hat{\tau}_{u_{tnk}}, \hat{\upsilon}_{u_{tnk}}), \tag{36}$$

$$\hat{\tau}_{u_{tnk}} = \sum_{x_{tnk}} \beta(x_{tnk}) \tau^{(i)}_{u_{tnk}}, \tag{37}$$

$$\hat{\upsilon}_{u_{tnk}} = \sum_{x_{tnk}} \beta(x_{tnk})(|\tau^{(i)}_{u_{tnk}}|^2 + \upsilon^{(i)}_{u_{tnk}}) - |\hat{\tau}_{u_{tnk}}|^2. \tag{38}$$

With $I^{(i)}_{f_{tn}\to u_{tnk}}(u_{tnk})$ follows Gaussian distribution, by EP message-passing, we have

$$\hat{I}^{(i)}_{u_{tnk}\to f_{tn}}(u_{tnk}) = \frac{\hat{b}(u_{tnk})}{I^{(i)}_{f_{tn}\to u_{tnk}}(u_{tnk})} \propto \mathcal{CN}(u_{tnk}; \hat{\tau}^{(i)}_{u_{tnk}\to f_{tn}}, \hat{\upsilon}^{(i)}_{u_{tnk}\to f_{tn}}), \tag{39}$$

where

$$\hat{\tau}^{(i)}_{u_{tnk}\to f_{tn}} = \hat{\upsilon}^{(i)}_{u_{tnk}\to f_{tn}} \Big(\frac{\hat{\tau}_{u_{tnk}}}{\hat{\upsilon}_{u_{tnk}}} - \frac{\tau^{(i)}_{f_{tn}\to u_{tnk}}}{\upsilon^{(i)}_{f_{tn}\to u_{tnk}}} \Big), \tag{40}$$

$$\hat{\upsilon}^{(i)}_{u_{tnk}\to f_{tn}} = \Big(\frac{1}{\hat{\upsilon}_{u_{tnk}}} - \frac{1}{\upsilon^{(i)}_{f_{tn}\to u_{tnk}}} \Big)^{-1}. \tag{41}$$

Now given the input messages to the function node f_{tn} $\hat{I}^{(i)}_{u_{tnl}\to f_{tn}}(u_{tnl}), l \in \mathrm{F}_n \backslash k$ follow independent Gaussian distributions, $u_{tnk} = y_{tn} - \sum_{l\in\mathbf{F}_n\backslash k} u_{tnl}$ is also a Gaussian variable such that the output messages on the next iteration can be updated as

$$I^{(i+1)}_{f_{tn}\to u_{tnk}}(u_{tnk}) \sim \mathcal{CN}(u_{tnk}; \tau^{(i+1)}_{f_{tn}\to u_{tnk}}, \upsilon^{(i+1)}_{f_{tn}\to u_{tnk}}), \tag{42}$$

where

$$\tau^{(i+1)}_{f_{tn}\to u_{tnk}} = y_{tn} - \sum_{l\in\mathrm{F}_n\backslash k} \hat{\tau}^{(i)}_{u_{tnl}\to f_{tn}}, \tag{43}$$

$$\upsilon^{(i+1)}_{f_{tn}\to u_{tnk}} = \sigma^2 + \sum_{l\in\mathrm{F}_n\backslash k} \hat{\upsilon}^{(i)}_{u_{tnl}\to f_{tn}}. \tag{44}$$

On the i^{th} iteration, since $I^{(i)}_{f_{tn}\to u_{tnk}}(u_{tnk})$ follows Gaussian distribution, the message sent from function node f_{tn} to variable node $\mathbf{x}_{tk}$ can be updated as

$$I^{(i)}_{f_{tn}\to \mathbf{x}_{tk}}(x_{tnk}) = \int I^{(i)}_{f_{tn}\to u_{tnk}}(u_{tnk}) I^{(i-1)}_{\alpha_{nk}\to f_{tn}}(\alpha_{nk}) d\alpha_{nk} \propto \exp\{-\Delta^{(i)}_{f_{tn}\to x_{tnk}}(x_{tnk})\}, \tag{45}$$

$$I^{(i)}_{u_{tnk}\to f_{tn}}(u_{tnk}) \propto \begin{cases} \sum_{x_{tnk}} I^{(i)}_{\mathbf{x}_{t,k}\to f_{tnk}}(x_{tnk}) |x_{tnk}| \mathcal{CN}(u_{tnk}; \tau^{(i-1)}_{\alpha_{nk}\to f_{tnk}} x_{tnk}, \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tnk}} |x_{tnk}|^2), & x_{tnk} \neq 0; \\ I^{(i)}_{\mathbf{x}_{tk}\to f_{tn}}(x_{tnk}), & x_{tnk} = 0. \end{cases} \tag{31}$$

$$b^{(i)}(u_{tnk}) = I^{(i)}_{u_{tnk}\to f_{tn}}(u_{tnk}) I^{(i)}_{f_{tn}\to u_{tnk}}(u_{tnk}) = \begin{cases} \sum_{x_{tnk}} \beta(x_{tnk}) \mathcal{CN}(u_{tnk}; \tau^{(i)}_{u_{tnk}}, \upsilon^{(i)}_{u_{tnk}}), & x_{tnk} \neq 0; \\ C^{-1} I^{(i)}_{\mathbf{x}_{tk}\to f_{tn}}(x_{tnk}) \mathcal{CN}(u_{tnk}; \tau^{(i)}_{f_{tn}\to u_{tnk}}, \upsilon^{(i)}_{f_{tn}\to u_{tnk}}), & x_{tnk} = 0. \end{cases} \tag{32}$$

$$\beta(x_{tnk}) = C^{-1} I^{(i)}_{\mathbf{x}_{t,k}\to f_{tnk}}(x_{tnk}) |x_{tnk}| \mathcal{CN}(\tau^{(i)}_{f_{tn}\to u_{tnk}}; \tau^{(i-1)}_{\alpha_{nk}\to f_{tnk}} x_{tnk}, \upsilon^{(i)}_{f_{tn}\to u_{tnk}} + \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tnk}} |x_{tnk}|^2), \tag{33}$$

$$\tau^{(i)}_{u_{tnk}} = \frac{\tau^{(i-1)}_{\alpha_{nk}\to f_{tnk}} \upsilon^{(i)}_{f_{tn}\to u_{tnk}} x_{tnk} + \tau^{(i)}_{f_{tn}\to u_{tnk}} \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tnk}} |x_{tnk}|^2}{\upsilon^{(i)}_{f_{tn}\to u_{tnk}} + \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tnk}} |x_{tnk}|^2}, \tag{34}$$

$$\upsilon^{(i)}_{u_{tnk}} = \frac{\upsilon^{(i)}_{f_{tn}\to u_{tnk}} \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tnk}} |x_{tnk}|^2}{\upsilon^{(i)}_{f_{tn}\to u_{tnk}} + \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tnk}} |x_{tnk}|^2}. \tag{35}$$

where

$$\Delta^{(i)}_{f_{tn}\to x_{tnk}}(x_{tnk}) = \frac{|\tau^{(i)}_{f_{tn}\to u_{tnk}} - \tau^{(i-1)}_{\alpha_{nk}\to f_{tn}} x_{tnk}|^2}{\upsilon^{(i)}_{f_{tn}\to u_{tnk}} + \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}}|x_{tnk}|^2} + \ln(\upsilon^{(i)}_{f_{tn}\to u_{tnk}} + \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}}|x_{tnk}|^2), \quad (46)$$

where we have assumed that the message $I^{(i-1)}_{\alpha_{nk}\to f_{tn}}(\alpha_{nk})$ follows $\mathcal{CN}(\alpha_{nk}; \tau^{(i-1)}_{\alpha_{nk}\to f_{tn}}, \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}})$ which will be given in (72). Notice that in computing (46), the LLRs for data symbols, not only the mean values but also the variances of interferences are involved.

With $I^{(i)}_{f_{tn}\to \mathbf{x}_{tk}}(x_{tnk})$ computed on the previous step, the message $I^{(i)}_{\mathbf{x}_{tk}\to f_{tn}}(x_{tnk})$ can be calculated according to the BP rule,

$$I^{(i)}_{\mathbf{x}_{tk}\to f_{tn}}(x_{tnk}) = p(\mathbf{x}_{tk}|\mathbf{c}_{tk}) \prod_{m\neq n} I^{(i)}_{f_{tm}\to \mathbf{x}_{tk}}(x_{tmk}) \propto \exp\Big\{- \sum_{m\in \mathrm{V}_k\backslash n} \Delta^{(i)}_{f_{tm}\to x_{tmk}}(x_{tmk})\Big\}. \quad (47)$$

where the mapping function $p(\mathbf{x}_{tk}|\mathbf{c}_{tk})$ denotes the SCMA encoder of user k. Finally, the belief of codewords x_{tk} is computed as

$$b^{(i)}(\mathbf{x}_{tk}) = p(\mathbf{x}_{tk}|\mathbf{c}_{tk}) \prod_{n\in \mathrm{V}_k} I^{(i)}_{f_{tn}\to \mathbf{x}_{tk}}(x_{tnk}) \quad (48)$$

2) Data Aided Channel Estimation and User Activity Detection: Now we begin our discussion on the channel estimation loop in Fig. 3. As in (45), since $I^{(i)}_{f_{tn}\to u_{tnk}}(u_{tnk})$ follows Gaussian distribution, the message passed from function node f_{tn} to the variable node α_{nk} can be updated as

$$\begin{aligned} I^{(i)}_{f_{tn}\to \alpha_{nk}}(\alpha_{nk}) &= \sum_{x_{tnk}} I^{(i)}_{\mathbf{x}_{tk}\to f_{tn}}(x_{tnk}) I^{(i)}_{f_{tn}\to u_{tnk}}(u_{tnk}) \\ &= \sum_{x_{tnk}} I^{(i)}_{\mathbf{x}_{tk}\to f_{tn}}(x_{tnk}) \\ &\quad \times \mathcal{CN}(\alpha_{nk}x_{tnk}; \tau^{(i)}_{f_{tn}\to u_{tnk}}, \upsilon^{(i)}_{f_{tn}\to u_{tnk}}), \end{aligned} \quad (49)$$

which is a mixture of Gaussian distributions. To obtain the compact form for $I^{(i)}_{f_{tn}\to \alpha_{nk}}(\alpha_{nk})$, again we project $I^{(i)}_{f_{tn}\to \alpha_{nk}}(\alpha_{nk})$ into Gaussian families.

The belief of α_{nk} is given by

$$\begin{aligned} b^{(i)}(\alpha_{nk}) &= I^{(i)}_{f_{tn}\to \alpha_{nk}}(\alpha_{nk}) I^{(i-1)}_{\alpha_{nk}\to f_{tn}}(\alpha_{nk}) \\ &= \sum_{x_{tnk}} \beta(x_{tnk}) \mathcal{CN}(\alpha_{nk}; \tilde{\tau}^{(i)}_{\alpha_{nk}}, \tilde{\upsilon}^{(i)}_{\alpha_{nk}}), \end{aligned} \quad (50)$$

where we have assumed $I^{(i-1)}_{\alpha_{nk}\to f_{tn}}(\alpha_{nk})$ follows Gaussian distribution and will be justified later in (72). The variables $\beta(x_{tnk})$, $\tilde{\tau}^{(i)}_{\alpha_{nk}}$, and $\tilde{\upsilon}^{(i)}_{\alpha_{nk}}$ are computed in (51)-(53), shown at the top of the next page, and C is a normalization constant.

We project $b^{(i)}(\alpha_{nk})$ into Gaussian distribution with the minimized KL-divergence $D_{KL}[b^{(i)}(\alpha_{nk})||\hat{b}(\alpha_{nk})]$,

$$\hat{b}(\alpha_{nk}) = \mathcal{CN}(\alpha_{nk}; \hat{\tau}_{\alpha_{nk}}, \hat{\upsilon}_{\alpha_{nk}}), \quad (54)$$

where by moment matching,

$$\hat{\tau}_{\alpha_{nk}} = \sum_{x_{tnk}} \beta(x_{tnk}) \tilde{\tau}^{(i)}_{\alpha_{nk}}, \quad (55)$$

$$\hat{\upsilon}_{\alpha_{nk}} = \sum_{x_{tnk}} \beta(x_{tnk}) (|\tilde{\tau}^{(i)}_{\alpha_{nk}}|^2 + \tilde{\upsilon}^{(i)}_{\alpha_{nk}}) - |\hat{\tau}_{\alpha_{nk}}|^2. \quad (56)$$

Since the extrinsic message $I^{(i-1)}_{\alpha_{nk}\to f_{tn}}(\alpha_{nk})$ follows Gaussian distribution $\mathcal{CN}(\alpha_{nk}; \tau^{(i-1)}_{\alpha_{nk}\to f_{tn}}, \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}})$, by EP message-passing, we have

$$\hat{I}^{(i)}_{f_{tn}\to \alpha_{nk}}(\alpha_{nk}) = \frac{\hat{b}(\alpha_{nk})}{I^{(i-1)}_{\alpha_{nk}\to f_{tn}}(\alpha_{nk})} \propto \mathcal{CN}(\alpha_{nk}; \hat{\tau}^{(i)}_{f_{tn}\to \alpha_{nk}}, \hat{\upsilon}^{(i)}_{f_{tn}\to \alpha_{nk}}), \quad (57)$$

where

$$\hat{\tau}^{(i)}_{f_{tn}\to \alpha_{nk}} = \hat{\upsilon}^{(i)}_{f_{tn}\to \alpha_{nk}} \Big(\frac{\hat{\tau}_{\alpha_{nk}}}{\hat{\upsilon}_{\alpha_{nk}}} - \frac{\tau^{(i-1)}_{\alpha_{nk}\to f_{tn}}}{\upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}}}\Big), \quad (58)$$

$$\hat{\upsilon}^{(i)}_{f_{tn}\to \alpha_{nk}} = \Big(\frac{1}{\hat{\upsilon}_{\alpha_{nk}}} - \frac{1}{\upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}}}\Big)^{-1}. \quad (59)$$

According to (42), $\hat{I}^{(i)}_{f_{tn}\to \alpha_{nk}}(\alpha_{nk})$ reduces to (60) for pilot signals,

$$\begin{aligned} \hat{I}^{(i)}_{f_{tn}\to \alpha_{nk}}(\alpha_{nk}) &= \mathcal{CN}(\alpha_{nk}x^p_{tnk}; \tau^{(i)}_{f_{tn}\to u_{tnk}}, \upsilon^{(i)}_{f_{tn}\to u_{tnk}}), \\ &\propto \mathcal{CN}(\alpha_{nk}; \hat{\tau}^{(i)}_{f_{tn}\to \alpha_{nk}}, \hat{\upsilon}^{(i)}_{f_{tn}\to \alpha_{nk}}), \end{aligned} \quad (60)$$

where

$$\hat{\tau}^{(i)}_{f_{tn}\to \alpha_{nk}} = \frac{\tau^{(i)}_{f_{tn}\to u_{tnk}}}{x^p_{tnk}}, \quad (61)$$

$$\hat{\upsilon}^{(i)}_{f_{tn}\to \alpha_{nk}} = \frac{\upsilon^{(i)}_{f_{tn}\to u_{tnk}}}{|x^p_{tnk}|^2}. \quad (62)$$

Now for both pilot and data signals, $\hat{I}^{(i)}_{f_{tn}\to \alpha_{nk}}(\alpha_{nk})$ follows the Gaussian distributions. According to the BP updating rule, we have

$$I^{(i)}_{\alpha_{nk}\to \varphi_{nk}}(\alpha_{nk}) = \prod_t \hat{I}^{(i)}_{f_{tn}\to \alpha_{nk}}(\alpha_{nk}) \propto \mathcal{CN}(\alpha_{nk}; \tau^{(i)}_{\alpha_{nk}\to \varphi_{nk}}, \upsilon^{(i)}_{\alpha_{nk}\to \varphi_{nk}}), \quad (63)$$

where

$$\tau^{(i)}_{\alpha_{nk}\to \varphi_{nk}} = \upsilon^{(i)}_{\alpha_{nk}\to \varphi_{nk}} \sum_t \frac{\hat{\tau}^{(i)}_{f_{tn}\to \alpha_{nk}}}{\hat{\upsilon}^{(i)}_{f_{tn}\to \alpha_{nk}}}, \quad (64)$$

$$\upsilon^{(i)}_{\alpha_{nk}\to \varphi_{nk}} = \Big(\sum_t \frac{1}{\hat{\upsilon}^{(i)}_{f_{tn}\to \alpha_{nk}}}\Big)^{-1}, \quad (65)$$

and the production in (63) is through all time slots that the block fading coefficient α_{nk} maintains unchanged.

As for the message updating in the reverse direction, $I^{(i)}_{\varphi_{nk}\to \alpha_{nk}}(\alpha_{nk})$ can be calculated with the assumption that $I^{(i-1)}_{h_{kl}\to \varphi_{nk}}(h_{kl})$ follows Gaussian distribution shown in (85)

$$I^{(i)}_{\varphi_{nk}\to \alpha_{nk}}(\alpha_{nk}) = \int \varphi_{nk}(\alpha_{nk}, \mathbf{h}_k) \prod_l I^{(i-1)}_{h_{kl}\to \varphi_{nk}}(h_{kl}) d\mathbf{h}_k \propto \mathcal{CN}(\alpha_{nk}; \tau^{(i)}_{\varphi_{nk}\to \alpha_{nk}}, \upsilon^{(i)}_{\varphi_{nk}\to \alpha_{nk}}), \quad (66)$$

$$\beta(x_{tnk}) = C^{-1} I^{(i)}_{\mathbf{x}_{tk}\to f_{tn}}(x_{tnk}) \mathcal{CN}(\tau^{(i-1)}_{\alpha_{nk}\to f_{tn}} x_{tnk}; \tau^{(i)}_{f_{tn}\to u_{tnk}}, \upsilon^{(i)}_{f_{tn}\to u_{tnk}} + \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}} |x_{tnk}|^2), \tag{51}$$

$$\tilde{\tau}^{(i)}_{\alpha_{nk}} = \frac{\tau^{(i-1)}_{\alpha_{nk}\to f_{tn}} \upsilon^{(i)}_{f_{tn}\to u_{tnk}} + \tau^{(i)}_{f_{tn}\to u_{tnk}} \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}} x^*_{tnk}}{\upsilon^{(i)}_{f_{tn}\to u_{tnk}} + \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}} |x_{tnk}|^2}, \tag{52}$$

$$\tilde{\upsilon}^{(i)}_{\alpha_{nk}} = \frac{\upsilon^{(i)}_{f_{tn}\to u_{tnk}} \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}}}{\upsilon^{(i)}_{f_{tn}\to u_{tnk}} + \upsilon^{(i-1)}_{\alpha_{nk}\to f_{tn}} |x_{tnk}|^2}. \tag{53}$$

where $\varphi_{nk}(\alpha_{nk}, \mathbf{h}_k) = \delta(\alpha_{nk} - \sum_l \mathrm{F}_{nl} h_{kl})$ and

$$\tau^{(i)}_{\varphi_{nk}\to\alpha_{nk}} = \sum_l \mathrm{F}_{nl} \tau^{(i-1)}_{h_{kl}\to\varphi_{nk}}, \tag{67}$$

$$\upsilon^{(i)}_{\varphi_{nk}\to\alpha_{nk}} = \sum_l \mathrm{F}^2_{nl} \upsilon^{(i-1)}_{h_{kl}\to\varphi_{nk}}. \tag{68}$$

With (63) and (66), the belief of α_{nk} can be calculated as

$$\begin{aligned} b^{(i)}(\alpha_{nk}) &= I^{(i)}_{\alpha_{nk}\to\varphi_{nk}}(\alpha_{nk}) I^{(i)}_{\varphi_{nk}\to\alpha_{nk}}(\alpha_{nk}) \\ &\propto \mathcal{CN}(\alpha_{nk}; \tau^{(i)}_{\alpha_{nk}}, \upsilon^{(i)}_{\alpha_{nk}}), \end{aligned} \tag{69}$$

where

$$\tau^{(i)}_{\alpha_{nk}} = \frac{\tau^{(i)}_{\alpha_{nk}\to\varphi_{nk}} \upsilon^{(i)}_{\varphi_{nk}\to\alpha_{nk}} + \tau^{(i)}_{\varphi_{nk}\to\alpha_{nk}} \upsilon^{(i)}_{\alpha_{nk}\to\varphi_{nk}}}{\upsilon^{(i)}_{\varphi_{nk}\to\alpha_{nk}} + \upsilon^{(i)}_{\alpha_{nk}\to\varphi_{nk}}}, \tag{70}$$

$$\upsilon^{(i)}_{\alpha_{nk}} = \frac{\upsilon^{(i)}_{\varphi_{nk}\to\alpha_{nk}} \upsilon^{(i)}_{\alpha_{nk}\to\varphi_{nk}}}{\upsilon^{(i)}_{\varphi_{nk}\to\alpha_{nk}} + \upsilon^{(i)}_{\alpha_{nk}\to\varphi_{nk}}}. \tag{71}$$

Given the belief of α_{nk}, it follows that the message $I^{(i)}_{\alpha_{nk}\to f_{tn}}(\alpha_{nk})$ is updated as

$$\begin{aligned} I^{(i)}_{\alpha_{nk}\to f_{tn}}(\alpha_{nk}) &= \frac{b^{(i)}(\alpha_{nk})}{\hat{I}^{(i)}_{f_{tn}\to\alpha_{nk}}(\alpha_{nk})} \\ &\propto \mathcal{CN}(\alpha_{nk}; \tau^{(i)}_{\alpha_{nk}\to f_{tn}}, \upsilon^{(i)}_{\alpha_{nk}\to f_{tn}}), \end{aligned} \tag{72}$$

where

$$\tau^{(i)}_{\alpha_{nk}\to f_{tn}} = \upsilon^{(i)}_{\alpha_{nk}\to f_{tn}} \Big(\frac{\tau^{(i)}_{\alpha_{nk}}}{\upsilon^{(i)}_{\alpha_{nk}}} - \frac{\hat{\tau}^{(i)}_{f_{tn}\to\alpha_{nk}}}{\hat{\upsilon}^{(i)}_{f_{tn}\to\alpha_{nk}}} \Big), \tag{73}$$

$$\upsilon^{(i)}_{\alpha_{nk}\to f_{tn}} = \Big(\frac{1}{\upsilon^{(i)}_{\alpha_{nk}}} - \frac{1}{\hat{\upsilon}^{(i)}_{f_{tn}\to\alpha_{nk}}} \Big)^{-1}. \tag{74}$$

In a similar approach to (66), we calculate the message from function node φ_{nk} to variable node h_{kl} as follows

$$I^{(i)}_{\varphi_{nk}\to h_{kl}}(h_{kl}) \propto \mathcal{CN}(h_{kl}; \tau^{(i)}_{\varphi_{nk}\to h_{kl}}, \upsilon^{(i)}_{\varphi_{nk}\to h_{kl}}), \tag{75}$$

where

$$\tau^{(i)}_{\varphi_{nk}\to h_{kl}} = \frac{1}{\mathrm{F}_{nl}} (\tau^{(i)}_{\alpha_{nk}\to\varphi_{nk}} - \sum_{j\neq l} \mathrm{F}_{nj} \tau^{(i-1)}_{h_{kj}\to\varphi_{nk}}), \tag{76}$$

$$\upsilon^{(i)}_{\varphi_{nk}\to h_{kl}} = \frac{1}{\mathrm{F}^2_{nl}} (\upsilon^{(i)}_{\alpha_{nk}\to\varphi_{nk}} + \sum_{j\neq l} \mathrm{F}^2_{nj} \upsilon^{(i-1)}_{h_{kj}\to\varphi_{nk}}). \tag{77}$$

Further, $I^{(i)}_{h_{kl}\to q_{kl}}(h_{kl})$ is updated as

$$\begin{aligned} I^{(i)}_{h_{kl}\to q_{kl}}(h_{kl}) &= \prod_n I^{(i)}_{\varphi_{nk}\to h_{kl}}(h_{kl}) \\ &\propto \mathcal{CN}(h_{kl}; \tau^{(i)}_{h_{kl}\to q_{kl}}, \upsilon^{(i)}_{h_{kl}\to q_{kl}}), \end{aligned} \tag{78}$$

where the production is through all nonzero subcarriers that user k transmits data and

$$\tau^{(i)}_{h_{kl}\to q_{kl}} = \upsilon^{(i)}_{h_{kl}\to q_{kl}} \sum_n \frac{\tau^{(i)}_{\varphi_{nk}\to h_{kl}}}{\upsilon^{(i)}_{\varphi_{nk}\to h_{kl}}}, \tag{79}$$

$$\upsilon^{(i)}_{h_{kl}\to q_{kl}} = \Big(\sum_n \frac{1}{\upsilon^{(i)}_{\varphi_{nk}\to h_{kl}}} \Big)^{-1}. \tag{80}$$

For the last part, we discuss the user activity detection loop. Notice that the distribution of active users in a network is sparse and the user activity detection can be treated as a sparse signals learning problem. Variational Bayesian inference is utilized here to learn the statistical properties of sparse signals. To start with, the message $I^{(i)}_{q_{kl}\to\lambda_{kl}}(\lambda_{kl})$ can be updated as

$$\begin{aligned} I^{(i)}_{q_{kl}\to\lambda_{kl}}(\lambda_{kl}) &= \exp\{\langle \ln q(h_{kl}|\lambda_{kl}) \rangle_{b^{(i-1)}(h_{kl})}\} \\ &\propto \mathrm{Gamma}(\lambda_{kl}; 2, |\tau^{(i-1)}_{h_{kl}}|^2 + \upsilon^{(i-1)}_{h_{kl}}), \end{aligned} \tag{81}$$

where $q(h_{kl}|\lambda_{kl})$ is given in (10) and the belief $b^{(i)}(h_{kl})$ is given in (88). In a similar approach, the message $I^{(i)}_{q_{kl}\to h_{kl}}(h_{kl})$ can be calculated as

$$\begin{aligned} I^{(i)}_{q_{kl}\to h_{kl}}(h_{kl}) &= \exp\{\langle \ln q(h_{kl}|\lambda_{kl}) \rangle_{b^{(i-1)}(\lambda_{kl})}\} \\ &\propto \mathcal{CN}(h_{kl}; 0, (\tau^{(i-1)}_{\lambda_{kl}})^{-1}), \end{aligned} \tag{82}$$

where $\tau^{(i)}_{\lambda_{kl}} = \frac{a+1}{b+|\tau^{(i-1)}_{h_{kl}}|^2+\upsilon^{(i-1)}_{h_{kl}}}$ and the belief $b^{(i)}(\lambda_{kl})$ is given by

$$\begin{aligned} b^{(i)}(\lambda_{kl}) &= p(\lambda_{kl}) I^{(i)}_{q_{kl}\to\lambda_{kl}}(\lambda_{kl}) \\ &\propto \mathrm{Gamma}(\lambda_{kl}; \hat{a}, \hat{b}), \end{aligned} \tag{83}$$

with $\hat{a} = a + 1$ and $\hat{b} = b + |\tau^{(i-1)}_{h_{kl}}|^2 + \upsilon^{(i-1)}_{h_{kl}}$. In (82), $\tau^{(i)}_{\lambda_{kl}}$ serves as the inverse of the channel power, when $(\tau^{(i)}_{\lambda_{kl}})^{-1} \to 0$, so does the CIR h_{kl}. Since the inactive users are equivalent to have zero CIRs, an inactive user k is judged when

$$\sum_l (\tau^{(i)}_{\lambda_{kl}})^{-1} < \delta, \tag{84}$$

and vice versa, where δ is a small enough number. Since a common channel model (10)-(12) is assumed, (84) is also valid for the user detection in BP-MF as well as BP-GA.

In the end, with (75) the extrinsic message $I^{(i)}_{h_{kl}\to\varphi_{nk}}(h_{kl})$ can be updated as

$$\begin{aligned} I^{(i)}_{h_{kl}\to\varphi_{nk}}(h_{kl}) &= \frac{b^{(i)}(h_{kl})}{I^{(i)}_{\varphi_{nk}\to h_{kl}}(h_{kl})} \\ &\propto \mathcal{CN}(h_{kl}; \tau^{(i)}_{h_{kl}\to\varphi_{nk}}, \upsilon^{(i)}_{h_{kl}\to\varphi_{nk}}), \end{aligned} \tag{85}$$

Algorithm 1 BP-GA-EP Message-Passing Algorithm

Initialization: $\forall t,n,k,l,\ \tau^{(0)}_{h_{kl}\to\varphi^{(0)}_{nk}} = \hat{h}^{ML}_{kl}, \upsilon^{(0)}_{h_{kl}\to\varphi_{nk}} = 1.0,\ \tau^{(0)}_{\alpha_{nk}\to f_{tn}} = 0, \upsilon^{(0)}_{\alpha_{nk}\to f_{tn}} = 1.0,\ \tau_{\lambda_{kl}} = 1.0$, and $\tau^{(0)}_{u_{tnk}\to f_{tn}} = 0, \upsilon^{(0)}_{u_{tnk}\to f_{tn}} = 10^6$.

Output: $b(\mathbf{x}_{tk})$, $\tau_{h_{kl}}$ and active users set by (84).

1: **Iteration:**
2: Update $\tau^{(i)}_{f_{tn}\to u_{tnk}}, \upsilon^{(i)}_{f_{tn}\to u_{tnk}}$ via (43), (44),$\forall t,n,k$.
3: Update $I^{(i)}_{f_{tn}\to\mathbf{x}_{tk}}(x_{tnk})$ and $I^{(i)}_{\mathbf{x}_{tk}\to f_{tn}}(x_{tnk})$ via (45) and (47), $\forall t,n,k$.
4: Update $\beta(x_{tnk}), \tau^{(i)}_{u_{tnk}}, \upsilon^{(i)}_{u_{tnk}}$ via (33)-(35), $\forall t,n,k$.
5: Update $\hat{\tau}^{(i)}_{u_{tnk}\to f_{tn}}, \hat{\upsilon}^{(i)}_{u_{tnk}\to f_{tn}}$ via (40), (41),$\forall t,n,k$.
6: Update $\beta(x_{tnk}), \tilde{\tau}^{(i)}_{\alpha_{nk}}, \tilde{\upsilon}^{(i)}_{\alpha_{nk}}$ via (51)-(53),$\forall t,n,k$.
7: Update $\hat{\tau}^{(i)}_{f_{tn}\to\alpha_{nk}}, \hat{\upsilon}^{(i)}_{f_{tn}\to\alpha_{nk}}$ via (58), (59) for data symbols and (61), (62) for pilot signals, $\forall t,n,k$.
8: Update $\tau^{(i)}_{\alpha_{nk}\to\varphi_{nk}}, \upsilon^{(i)}_{\alpha_{nk}\to\varphi_{nk}}$ and $\tau^{(i)}_{\varphi_{nk}\to\alpha_{nk}}, \upsilon^{(i)}_{\varphi_{nk}\to\alpha_{nk}}$ via (64), (65), and (67), (68), respectively, $\forall n,k$.
9: Update $\tau^{(i)}_{\alpha_{nk}}, \upsilon^{(i)}_{\alpha_{nk}}$ via (70), (71), $\forall n,k$.
10: Update $\tau^{(i)}_{\alpha_{nk}\to f_{tn}}, \upsilon^{(i)}_{\alpha_{nk}\to f_{tn}}$ via (73), (74), $\forall t,n,k$.
11: Update $\tau^{(i)}_{\varphi_{nk}\to h_{kl}}, \upsilon^{(i)}_{\varphi_{nk}\to h_{kl}}$ via (76), (77), $\forall n,k,l$.
12: Update $\tau^{(i)}_{h_{kl}\to q_{kl}}, \upsilon^{(i)}_{h_{kl}\to q_{kl}}$ via (79), (80), $\forall k,l$.
13: Update $\tau^{(i)}_{h_{kl}}, \upsilon^{(i)}_{h_{kl}}, \tau^{(i)}_{\lambda_{kl}}$ via (89), (90), and (83) $\forall k,l$.
14: Update $\tau^{(i)}_{h_{kl}\to\varphi_{nk}}, \upsilon^{(i)}_{h_{kl}\to\varphi_{nk}}$ via (86), (87), $\forall n,k,l$.
15: **until** iteration index reaches some preset numbers.

where

$$\tau^{(i)}_{h_{kl}\to\varphi_{nk}} = \upsilon^{(i)}_{h_{kl}\to\varphi_{nk}}\Big(\frac{\tau^{(i)}_{h_{kl}}}{\upsilon^{(i)}_{h_{kl}}} - \frac{\tau^{(i)}_{\varphi_{nk}\to h_{kl}}}{\upsilon^{(i)}_{\varphi_{nk}\to h_{kl}}}\Big), \tag{86}$$

$$\upsilon^{(i)}_{h_{kl}\to\varphi_{nk}} = \Big(\frac{1}{\upsilon^{(i)}_{h_{kl}}} - \frac{1}{\upsilon^{(i)}_{\varphi_{nk}\to h_{kl}}}\Big)^{-1}. \tag{87}$$

and the belief of h_{kl} is computed with (78) and (82),

$$\begin{aligned} b^{(i)}(h_{kl}) &= I^{(i)}_{q_{kl}\to h_{kl}}(h_{kl}) I^{(i)}_{h_{kl}\to q_{kl}}(h_{kl}) \\ &\propto \mathcal{CN}(h_{kl}; \tau^{(i)}_{h_{kl}}, \upsilon^{(i)}_{h_{kl}}), \end{aligned} \tag{88}$$

where

$$\tau^{(i)}_{h_{kl}} = \frac{\tau^{(i)}_{h_{kl}\to q_{kl}}}{1+\tau^{(i)}_{\lambda_{kl}}\upsilon^{(i)}_{h_{kl}\to q_{kl}}}, \tag{89}$$

$$\upsilon^{(i)}_{h_{kl}} = \frac{\upsilon^{(i)}_{h_{kl}\to q_{kl}}}{1+\tau^{(i)}_{\lambda_{kl}}\upsilon^{(i)}_{h_{kl}\to q_{kl}}}. \tag{90}$$

The proposed message-passing receiver is referred as BP-GA-EP (belief propagation based Gaussian approximation with expectation propagation), and is summarized in Algorithm 1. To deal with the deviated approximation problem in EP, damped updating [25] with scaling factor equals 0.5 is applied in step 4. Moreover, for the cases $\upsilon^{(i)}_{u_{tnk}\to f_{tn}} < 0$ and $\upsilon^{(i)}_{f_{tn}\to\alpha_{nk}} < 0$, a large positive number (e.g., 10^6) is set to replace the above variances.

TABLE I
COMPUTATIONAL COMPLEXITY ANALYSIS

	Initialization	Channel Estimation	Data Decoding
BP-GA-EP	$\mathcal{O}(KL^2)$	$\mathcal{O}(BNMd_c)$	$\mathcal{O}(BNMd_c)$
BP-MF	$\mathcal{O}(KL^2+BNd_c^3)$	$\mathcal{O}(BNMd_c)$	$\mathcal{O}(BNMd_c)$
BP-GA	$\mathcal{O}(KL^2)$	$\mathcal{O}(BNMd_c)$	$\mathcal{O}(BNMd_c)$

C. Complexity Analysis

The computational complexity analysis of joint detectors is discussed in this subsection. For BP-GA-EP, the computational consumption is dominated by (33)-(38) in data detection loop and (51)-(56) in channel estimation loop. The complexity in calculating $\beta(x_{tnk})$, $\tau_{u_{tnk}}$, and $\upsilon_{u_{tnk}}$ is on the order of $\mathcal{O}(BNMd_c)$ for each time slot. In parallel with data detection loop, the complexity in calculating $\beta(x_{tnk})$, $\tilde{\tau}_{\alpha_{nk}}$, and $\tilde{\upsilon}_{\alpha_{nk}}$ is also on the order of $\mathcal{O}(BNMd_c)$. For BP-MF, the interference cancellation structure in (26) dominates the computational consumption in data detection part while for channel estimation part, the updating rule is similar to (26). For both parts, the complexity is on the order of $\mathcal{O}(BNMd_c)$. Since mean field methods could only find a local optimal solution, a proper initialization is necessary. The initialization for channel estimation is based on (18) using pilot signals only, and followed by the soft estimation of data symbols with linear minimum mean square error (MMSE) receiver. For each time slot, the complexity for linear MMSE is on the order of $\mathcal{O}(BNd_c^3)$. With the Gaussian approximation, BP-GA has the same complexity with BP-GA-EP for both data detection and channel estimation, i.e., the complexity is identical to the order of $\mathcal{O}(BNMd_c)$, see (22) and (23) in [20]. The computational complexity is summarized in Table I.

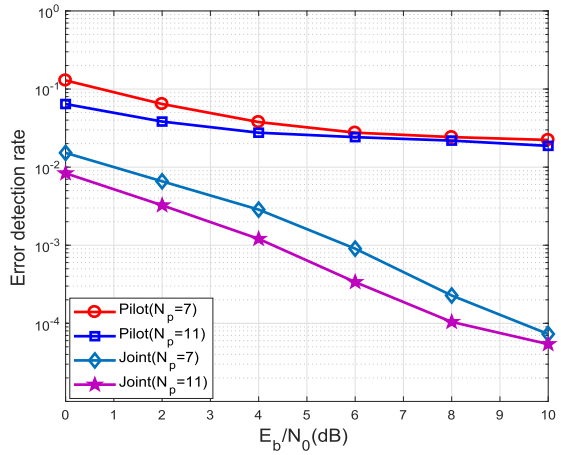
IV. SIMULATION RESULTS

A. Experimental Set Up

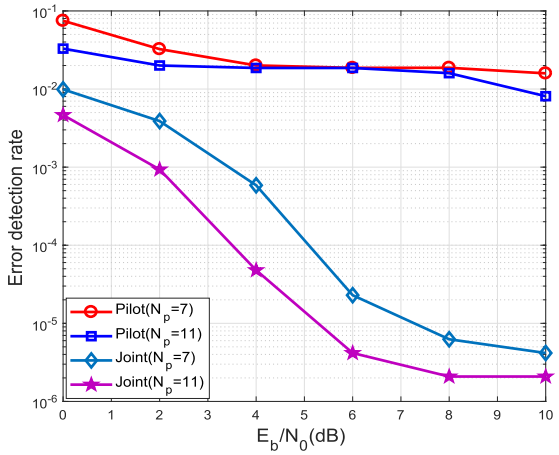
In this section, the performance of proposed message-passing receiver is evaluated through Monte Carlo simulations. The uplink SCMA system with $B=4, N=24, K=48$, $M=4, d_v=2$, and $d_c=4$ is considered. At each transmission, the active users are generated randomly, and we assume that the number of simultaneous active users follows Poisson distribution $\mathrm{Pois}(\lambda)$ [14]. The maximum channel taps L for each user is set to 6 with unknown power-delay-profiles following [30, Table I], truncated to $L=6$ samples. We assume the frequency-selective block-fading channel coefficient α_{nk} remains constant during N_s data symbols, and N_p pilot symbols are further multiplexed within each fading block for channel estimation. Zaddoff-Chu sequence [28] designed to be orthogonal in each subcarrier is used as pilot signals. We consider two communication scenarios in this paper: for short packet communication, the packet contains 256 coded bits whereas for long packet communication, the packet length equals 2048 coded bits.

B. Performance Evaluation

The error detection rate of the proposed grant-free receiver for user activity detection is shown in Fig. 5(a) and (b) for

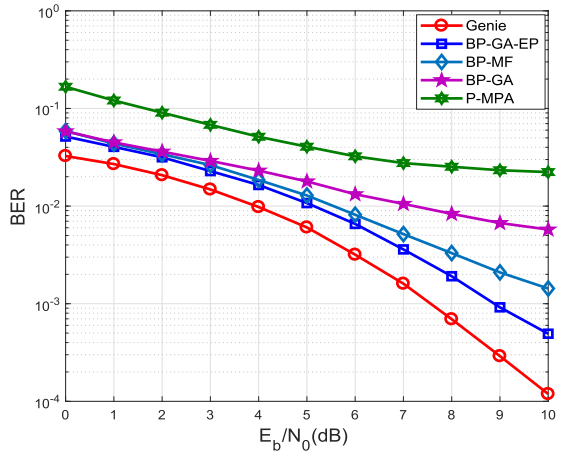


(a) Short packet, $N_s = 32$, $\lambda = 8.0$.

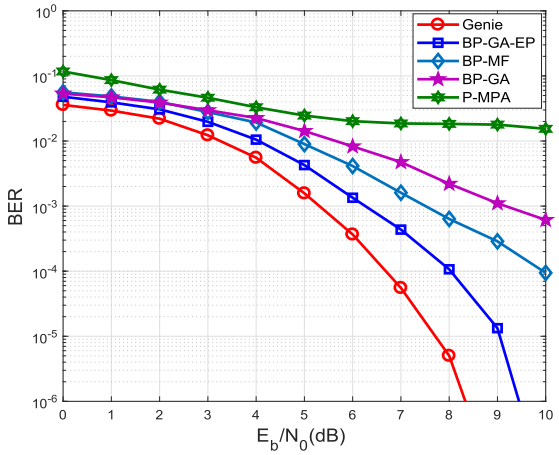


(b) Long packet, $N_s = 64$, $\lambda = 8.0$.

Fig. 5. Error detection rate for joint detector and pilot-aided only detector.



(a) Short packet, $N_p = 7$, $N_s = 32$, $\lambda = 8.0$.



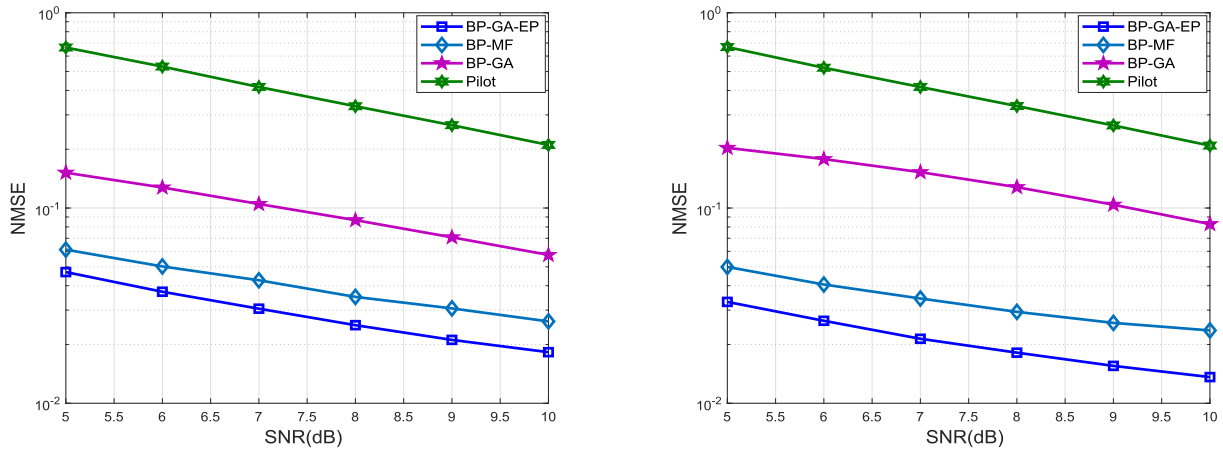
(b) Long packet, $N_p = 7$, $N_s = 64$, $\lambda = 8.0$.

Fig. 6. BER comparison for different decoders.

short and long packet communications, respectively. In both scenarios, we compare the accuracy of active user detection by joint detector and pilot-aided only detector. Monte Carlo simulation with 10000 random transmissions is carried out in order to get the average error detection rate. The curves in Fig. 5 ensure that the performance of joint detector is superior to the pilot-aided only one for several orders of magnitudes. The result was to be expected since the information of user activity is contained in not only pilot signals but also data signals. Therefore by exploring both received signals, the active users can be retrieved with a considerable accuracy based on the joint detector. In addition to the performance comparison for joint and pilot-aided detector, the impact of different pilot numbers ($N_p = 7$, $N_p = 11$) on the user error detection rate is also demonstrated in two figures. It can be observed that the error detection rate is almost on the same order of magnitudes when simply increasing the pilot numbers, while a noteworthy gain can be obtained by employing the joint detector. Consequently, the performance improvement is less significant when increasing the pilot numbers only.

In Fig. 6 and Fig. 7, the proposed BP-GA-EP algorithm is compared with BP-MF and BP-GA based joint message-passing receivers in terms of normalized minimum mean square error (NMSE) and bit-error-rate (BER), respectively. As the mean field method can only find a local optimal solution, the linear MMSE estimation for data symbols is served as the soft initialization based on ML estimation of CIRs $\hat{\mathbf{h}}_{ML}$ from the received pilot signals. For comparison, we also include the genie-aided receiver that has a perfect knowledge on both channel state information and user activity in the network. In addition, the pilot-aided receiver that use the pilot signals only for user detection and channel estimation is also simulated in this paper. For both genie and pilot-aided (P-MPA) receiver, conventional MPA algorithm is utilized for data decoding. For all the receivers but the genie, the transmitted bits are counted as errors entirely if the user is error detected [31]. This is reasonable since the best decoder cannot recover the transmitted data when it is acknowledged with the wrong information on user activity.

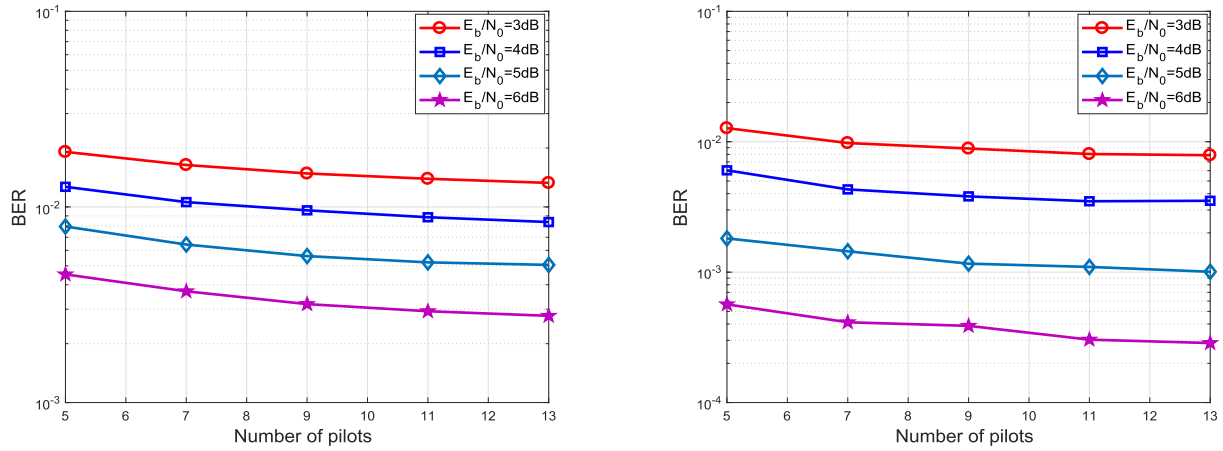
From Fig. 6, we observe that compared with genie-aided receiver, the state of the art receivers, ranging from BP-MF, BP-GA to P-MPA all suffer from significant performance losses. For P-MPA, the channel estimation and user detection are operated on the first phase, followed by the transmitted data symbols are being decoded on the second phase, through a utilizing of CSI as well as user activity information obtained



(a) Short packet, $N_p = 7$, $N_s = 32$, $\lambda = 8.0$.

(b) Long packet, $N_p = 7$, $N_s = 64$, $\lambda = 8.0$.

Fig. 7. NMSE comparison for different decoders.



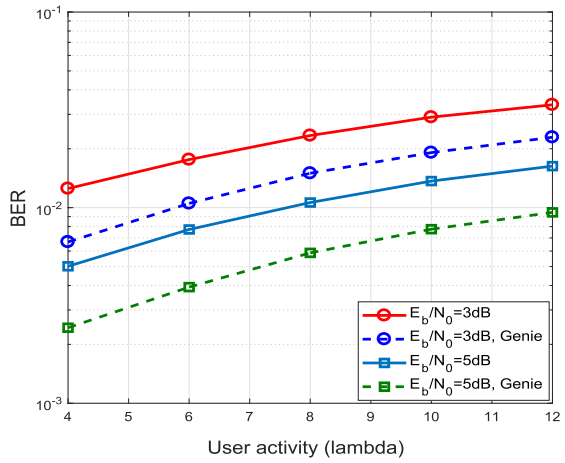
(a) Short packet, $N_s = 32$, $\lambda = 8.0$.

(b) Long packet, $N_s = 64$, $\lambda = 8.0$.

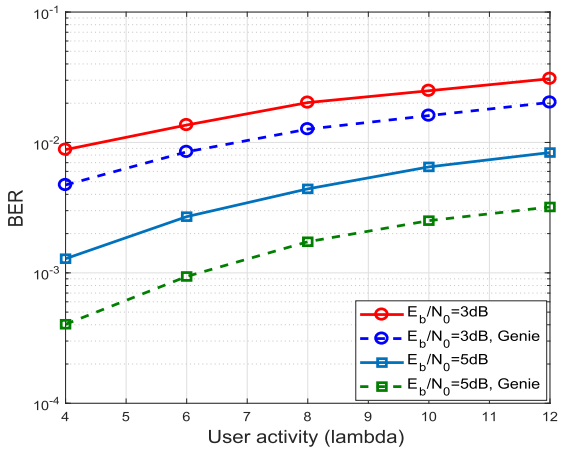
Fig. 8. BER performance of BP-GA-EP receiver with different number of pilots.

on the previous phase. Since the channel estimation and user detection rely on pilot signals only, the BER and NMSE performance of P-MPA are expected to be the worst among the overall decoders, which is confirmed by the simulations in Fig. 6 and 7. Compared with P-MPA, the BP-MF and BP-GA receivers perform joint detection and decoding by exploring the pilot and data signals concurrently. Thus they have the improved performance. The receiver based on mean field theory provides soft values for data symbols, however, the estimation is still unconvinced since the LLRs for data symbols only involves the mean values of interferences but omits the variances. As a result, it operates 3 dB away from the genie-aided receiver in the BER of 10^{-3} and 10^{-4} for short packet and long packet communication, respectively. Based on central-limit theory and moment matching, BP-GA receiver approximates the interferences in each OFDMA sub-carrier with Gaussian distribution and the estimations are combined with means and variances simultaneously. While the central-limit theorem is suitable for large scale MIMO-OFDM system, where the number of interferences may be tens or hundreds. For SCMA, however, the collision user number in one dimension is limited. As a result, direct Gaussian approximation suffers from a large performance loss, which can be observed in Fig. 6 and 7. The proposed BP-GA-EP receiver relies on Gaussian approximation as well. Nevertheless, different from BP-GA, the signal for each user is projected into Gaussian distribution individually with the minimized KL-divergence. Thus, the approximation is tight in BP-GA-EP. Simulation results in Fig. 6 and 7 show that the proposed BP-GA-EP algorithm has achieved the best performance among four decoders in either NMSE or BER, and is only more than 1 dB away compared to the genie-aided receiver at BER of 10^{-3} and 10^{-5} for short and long packet communication, respectively.

We further evaluate the impact of pilot numbers on BER performance with proposed receiver at different operational SNRs in Fig. 8. It is shown that for both short as well as long packet communications, the BER is getting improved with increasing the number of pilots. However, the improvement of performance is less significant, which is the consequence

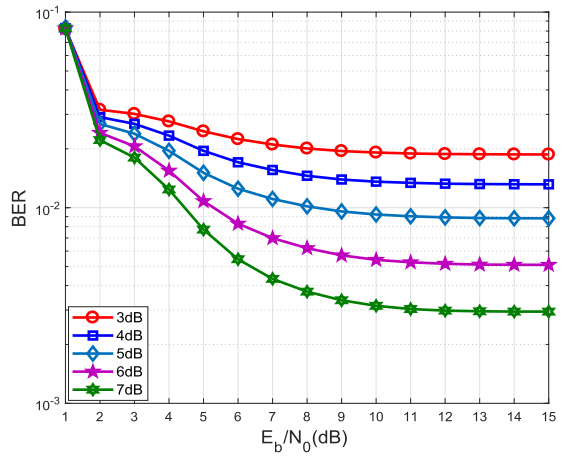


(a) Short packet, $N_p = 7$, $N_s = 32$.

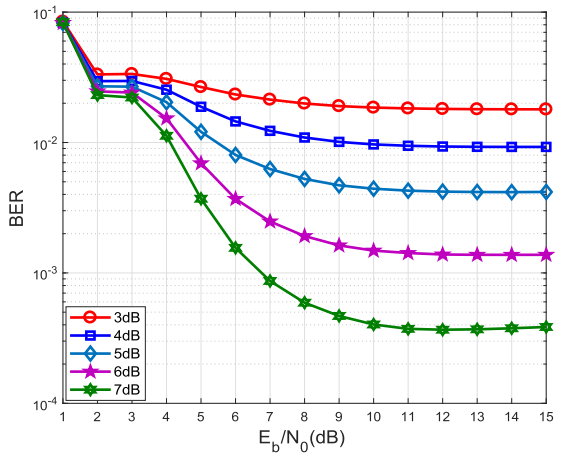


(b) Long packet, $N_p = 7$, $N_s = 64$.

Fig. 9. BER performance of BP-GA-EP receiver with different user activity.



(a) Short packet, $N_p = 7$, $N_s = 32$, $\lambda = 8.0$.



(b) Long packet, $N_p = 7$, $N_s = 64$, $\lambda = 8.0$.

Fig. 10. Convergence behaviour for BP-GA-EP receiver.

of the proposed joint pilot and data signals detection. In the joint detector, the pilots only provide an ML estimation for CIRs in (18), as a rough initialization for Algorithm 1. After that, the channel estimation is further refined in joint detection stage by data symbols. Since the number of data symbols is much larger than the number of pilots in one fading block, the quality of proposed detector depends mostly on the iterations within the joint detection stage. To that end, an increasing number of pilot symbols has only minor effects on the system performance. In other words, pilot overheads could be reduced by the proposed joint detector.

Finally, we demonstrate the impact of user activity on BER performance as well as the convergence behaviour of proposed BP-GA-EP receiver. In Fig. 9, the impact of user activity on BER performance is illustrated. For the short and long packet scenarios, the BER of BP-GA-EP increases with the rising of user activity but maintains a constant gap with genie-aided curves on several different SNRs. It should be noted that with increasing the of user activity, the multiuser interference in the system also increases. Therefore, the BER performance deteriorates for not only the proposed BP-GA-EP receiver but also for the genie. Nevertheless, a constant gap between the two algorithms always maintains. The stability of the proposed joint receiver is another concern of this paper. Recall that in Section III-B, we mentioned that to solve the deviated approximation problem in EP, a damped updating with scaling factor 0.5 is adopted. Further, the exceptions for negative variances are handled through the setting of some big numbers. In Fig. 10, we demonstrate the convergence behaviour of the proposed BP-GA-EP receiver. From the figure, it can be observed that the BP-GA-EP decoder achieves a stationary point in after 10 iterations on several operational SNRs. Thus, the proposed algorithm is ensured to be efficient and stable.

V. Conclusion

In this paper, we proposed a message-passing receiver for uplink grant-free SCMA. Due to the lack of information on CIRs and user activity, the proposed receiver performs joint channel estimation, data decoding, and active user detection semi-blindly with the aid of pilot signals. By exploring the received signals from transmitted pilot and data symbols concurrently, the joint receiver can identify active users with a much higher accuracy than pilot-aided only receiver. For data decoding, the FG of SCMA was formulated and intractable distributions were projected into Gaussian distributions with the minimized KL-divergence. We compared the proposed

BP-GA-EP receiver with the state of art decoders. Through simulations, it was demonstrated that the proposed algorithm performs the best among several receivers and the performance gap is only 1 dB or more compared with genie-aided receiver.

REFERENCES

[1] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 332–336.

[2] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1616–1626, Apr. 2008.

[3] D. Cai, P. Fan, X. Lei, Y. Liu, and D. Chen, "Multi-dimensional SCMA codebook design based on constellation rotation and interleaving," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.

[4] J. Bao, Z. Ma, M. A. Mahamadu, Z. Zhu, and D. Chen, "Spherical codes for SCMA codebook," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.

[5] Z. Li, W. Chen, F. Wei, F. Wang, X. Xu, and Y. Chen, "Joint codebook assignment and power allocation for SCMA based on capacity with Gaussian input," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Chengdu, China, Jul. 2016, pp. 1–6.

[6] A. Bayesteh, H. Nikopour, M. Taherzadeh, H. Baligh, and J. Ma, "Low complexity techniques for SCMA detection," in *Proc. IEEE Globecom Workshops*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[7] J. Chen, Z. Zhang, S. He, J. Hu, and G. E. Sobelman, "Sparse code multiple access decoding based on a Monte Carlo Markov chain method," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 639–643, May 2016.

[8] F. Wei and W. Chen, "Low complexity iterative receiver design for sparse code multiple access," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 621–634, Feb. 2017.

[9] M. Vameghestahbanati, E. Bedeer, I. Marsland, R. H. Gohary, and H. Yanikomeroglu, "Enabling sphere decoding for SCMA," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2750–2753, Dec. 2017.

[10] X. Meng, Y. Wu, Y. Chen, and M. Cheng, "Low complexity receiver for uplink SCMA system via expectation propagation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, San Francisco, CA, USA, Mar. 2017, pp. 1–5.

[11] K. Han, J. Hu, J. Chen, and H. Lu, "A low complexity sparse code multiple access detector based on stochastic computing," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 2, pp. 769–782, Feb. 2018.

[12] G. Szabo, D. Orincsay, B. P. Gero, S. Gyori, and T. Borsos, "Traffic analysis of mobile broadband networks," in *Proc. 3rd Int. Conf. Wireless Internet WICON*, Jun. 2007, Art. no. 18.

[13] K. Au *et al.*, "Uplink contention based SCMA for 5G radio access," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2014, pp. 900–905.

[14] J. Zhang *et al.*, "PoC of SCMA-based uplink grant-free transmission in UCNC for 5G," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1353–1362, Jun. 2017.

[15] 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.

[16] B. Wang, L. Dai, Y. Yuan, and Z. Wang, "Compressive sensing based multi-user detection for uplink grant-free non-orthogonal multiple access," in *Proc. IEEE 82nd Veh. Technol. Conf. (VTC Fall)*, Sep. 2015. pp. 1–5.

[17] 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.

[18] E. Riegler, G. E. Kirkelund, C. N. Manchon, M.-A. Badiu, and B. H. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 588–602, Jan. 2013.

[19] C. Novak, G. Matz, and F. Hlawatsch, "IDMA for the multiuser MIMO-OFDM uplink: A factor graph framework for joint data detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4051–4066, Aug. 2013.

[20] S. Wu, L. Kuang, Z. Ni, D. Huang, Q. Guo, and J. Lu, "Message-passing receiver for joint channel estimation and decoding in 3D massive MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8122–8138, Dec. 2016.

[21] 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.

[22] W. E. Ryan and S. Lin, *Channel Codes: Classical and Modern*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[23] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[24] F. Wei and W. Chen, "Message passing receiver design for uplink grant-free SCMA," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Singapore, Dec. 2017, pp. 1–6.

[25] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Artif. Intell.*, Aug. 2001, pp. 362–369.

[26] J. G. Proakis, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 1995.

[27] 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.

[28] *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) Physical Channels and Modulation, Release 13*, document TS 36.211 V13.2.0, 3GPP, 2016.

[29] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill, 2002.

[30] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.

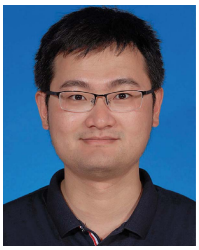
[31] G. Hannak, M. Mayer, A. Jung, G. Matz, and N. Goertz, "Joint channel estimation and activity detection for multiuser communication systems," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 2086–2091.

Fan Wei received the B.S. degree from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2011, and the M.S. degree from Xidian University, Xi'an, in 2014. He is currently pursuing the Ph.D. degree with the Shanghai Institute of Advanced Communications and Data Sciences, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include wireless communication, non-orthogonal multiple access, and massive random access.

Wen Chen (M'03–SM'11) received the B.S. and M.S. degrees from Wuhan University, China, in 1990 and 1993, respectively, and the Ph.D. degree from the University of Electro-Communications, Tokyo, Japan, in 1999. He was a Researcher of the Japan Society for the Promotion of Sciences (JSPS) from 1999 to 2001. In 2001, he joined the University of Alberta, Canada, starting as a Post-Doctoral Fellow with the Information Research Laboratory and continued as a Research Associate with the Department of Electrical and Computer Engineering. From 2014 to 2015, he was the Dean of the School of Electronics Engineering and Automations, Guilin University of Electronic Technology. Since 2006, he has been a Full Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China, where he has been the Director of the Institute for Signal Processing and Systems. Since 2016, he has been the Chairman of the Intellectual Property Management Corporation, Shanghai Jiao Tong University.

He received the Ariyama Memorial Research Prize in 1997, the JSPS Fellowship in 1999, and the PIMS Fellowship in 2001. He received the honors of a New Century Excellent Scholar in China in 2006 and a Pujiang Excellent Scholar in Shanghai in 2007, and the Best Service Award of the China Institute of Electronics in 2013, the Best Paper Awards of the Chinese Information Theory Society in 2014, the Innovate 5G Competition Award in 2015, and the IEEE VTS Chapter of the Year Award in 2018.

He has published 93 papers in IEEE journals and over 120 papers in IEEE conferences. His research interests cover multiple access, green communications, and co-operative communications with network coding. He is the IEEE VTS Shanghai Chapter Chair. He serves as the TPC chair and the general chair for many IEEE sponsored conferences. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE ACCESS.

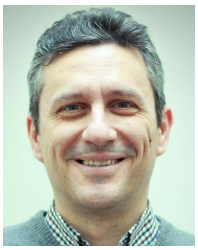


Yongpeng Wu (S'08–M'13–SM'17) received the B.S. degree in telecommunication engineering from Wuhan University, Wuhan, China, in 2007, and the Ph.D. degree in communication and signal processing from the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in 2013.

He is currently a Tenure-Track Associate Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. Previously, he was a Senior Research Fellow with the Institute for Communications Engineering, Technical University of Munich, Germany, and the Humboldt Research Fellow and a Senior Research Fellow with the Institute for Digital Communications, University Erlangen–Nuremberg, Germany. During his doctoral studies, he conducted cooperative research with the Department of Electrical Engineering, Missouri University of Science and Technology, USA. His research interests include massive MIMO/MIMO systems, physical layer security, signal processing for wireless communications, and multivariate statistical theory.

Dr. Wu received the IEEE Student Travel Grants from the IEEE International Conference on Communications (ICC) 2010, the Alexander von Humboldt Fellowship in 2014, the Travel Grants from the IEEE Communication Theory Workshop 2016, the Excellent Doctoral Thesis Awards of China Communications Society 2016, and an Exemplary Editor Award of the IEEE COMMUNICATION LETTERS 2017, and the Young Elite Scientist Sponsorship Program by CAST 2017. He was a TPC Member of various conferences, including GLOBECOM, ICC, VTC, and PIMRC. He was a Lead Guest Editor for the Special Issue Physical Layer Security for 5G Wireless Networks of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is currently an Editor of the IEEE ACCESS and the IEEE COMMUNICATIONS LETTERS. He was an Exemplary Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2015 and 2016.

Jun Ma received the B.S. degree in electronic instruments and measurement techniques and the M.S. degree in communication and information system from the Guilin University of Electronic Technology, China, in 2000 and 2005, respectively, and the Ph.D. degree in mechanical engineering from the Catholic University of America, Washington, DC, USA, in 2013. He is currently an Associate Professor and the Associate Dean of the School of Electronic Engineering and Automation, Guilin University of Electronic Technology. His main research interests, but not limited to, are signal processing, measurement and control technology, and applied and numerical harmonic analysis.



Theodoros A. Tsiftsis (S'02–M'04–SM'10) was born in Lamia, Greece, in 1970. He received the B.Sc. degree in physics from the Aristotle University of Thessaloniki, Greece, in 1993, the M.Sc. degree in digital systems engineering from Heriot-Watt University, Edinburgh, U.K., in 1995, the M.Sc. degree in decision sciences from the Athens University of Economics and Business, in 2000, and the Ph.D. degree in electrical engineering from the University of Patras, Greece, in 2006. He is currently a Professor with the School of Electrical and Information Engineering, Jinan University, Zhuhai, China. He has authored or co-authored over 170 technical papers in scientific journals and international conferences. His research interests include the broad areas of cooperative communications, cognitive radio, communication theory, wireless powered communication systems, and optical wireless communication systems.

He served as a Senior or Associate Editor for the Editorial Boards of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE COMMUNICATIONS LETTERS, *IET Communications*, and *IEICE Transactions on Communications*. He is currently an Area Editor of Wireless Communications II of the IEEE TRANSACTIONS ON COMMUNICATIONS and an Associate Editor of the IEEE TRANSACTIONS ON MOBILE COMPUTING. He also acts as a reviewer for several international journals and conferences.