# Iterative Receiver Design for FTN Signaling Aided Sparse Code Multiple Access

Weijie Yuan, Nan Wu, Andrew Zhang, *Senior Member, IEEE,* Xiaojing Huang, *Senior Member, IEEE,* Yonghui Li, *Fellow, IEEE,* and Lajos Hanzo, *Fellow, IEEE,*

***Abstract*—The sparse code multiple access (SCMA) is a promising candidate for bandwidth-efficient next generation wireless communications, since it can support more users than the number of resource elements. On the same note, faster-than-Nyquist (FTN) signaling can also be used to improve the spectral efficiency. Hence in this paper, we consider a combined uplink FTN-SCMA system in which the data symbols corresponding to a user are further packed using FTN signaling. As a result, a higher spectral efficiency is achieved at the cost of introducing intentional inter-symbol interference (ISI). To perform joint channel estimation and detection, we design a low complexity iterative receiver based on the factor graph framework. In addition, to reduce the signaling overhead and transmission latency of our SCMA system, we intrinsically amalgamate it with grant-free scheme. Consequently, the active and inactive users should be distinguished. To address this problem, we extend the aforementioned receiver and develop a new algorithm for jointly estimating the channel state information, detecting the user activity and for performs data detection. In order to further reduce the complexity, an energy minimization based approximation is employed for restricting the user state to Gaussian. Finally, a hybrid message passing algorithm is conceived. Our Simulation results show that the FTN-SCMA system relying on the proposed receiver design has a higher throughput than conventional SCMA scheme at a negligible performance loss.**

***Index Terms*—Sparse code multiple access, faster-than-Nyquist signaling, grant-free, channel estimation, hybrid message passing, high spectral efficiency**

## I. Introduction

The rapid proliferation of wireless applications requires higher spectral efficiency since the available bandwidth becomes limited. Conventional orthogonal multiple access (OMA) schemes such as time division multiple access (TDMA), code division multiple access (CDMA) and orthogonal frequency-division multiple access (OFDMA) assign orthogonal resource elements to different users [1]–[4]. Although OMA avoids multiuser interference under non-dispersive channel conditions, the challenges of high throughput and massive number of connections make it inferior for the next-generation wireless communications. By contrast, non-orthogonal multiple access (NOMA) is capable of increasing the spectral efficiency and addressing the aforementioned problems [5]. Amongst several NOMA technologies [6]–[9], the sparse code multiple access (SCMA) has attracted significant attention, due to its capability of achieving extra shaping gain [10].

To elaborate, the SCMA encoder directly maps the bits to sparse codewords. After multi-dimensional modulation and low density spreading, the bits streams corresponding to different users are directly mapped to sparse codewords of a predesigned codebook and then multiplexed over several orthogonal resource elements. Numerous authors considered the signal design of SCMA at the transmitter side. For example in [11], the authors investigated the SCMA codebook design based on systematic construction methods. To maximize the minimum codeword distance, a multi-dimensional codebook is designed based on a constellation rotation and interleaving method in [12]. In [13], capacity based codebook design is proposed to achieve the maximum sum rate. However, supporting more users than the number of resource elements result in a rank-deficient system, hence the complexity of the optimal receiver increases exponentially with the number of interferring users. To tackle this problem, several factor graph (FG) and message passing algorithm (MPA) based multiuser detectors were proposed by exploiting the low density codewords of SCMA. In [14], a low-complexity detection algorithm is proposed based on discretization and the fast Fourier transform (FFT). A list-sphere-decoding-based algorithm is devised in [15], but it only considered the signal falling within a hypersphere. The authors of [16] developed a partial marginalization based message passing detector for uplink SCMA. In [17], a Monte Carlo Markov Chain (MCMC) based SCMA decoder was proposed for a large SCMA codebook. In [18], the authors proposed a guaranteed convergence message passing algorithm for MIMO-SCMA systems. The authors of [19] proposed a modified MPA receiver, namely the max-log MPA that relies on message updating in the log domain to avoid multiplication operations.

On the other hand, an increased spectral efficiency can also

This work was supported by the "National Science Foundation of China (NSFC)" (Grant No.61571041,61971041), "A Foundation for the Author of National Excellent Doctoral Dissertation of P. R. China (FANEDD)" (Grant No. 201445). L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/Noo4558/1, EP/PO34284/1, COALESCE, of the Royal Society's Global Challenges Research Fund Grant as well as of the European Research Council's Advanced Fellow Grant QuantCom. (Corresponding Author: Nan Wu)

W. Yuan was with the School of Information and Electronics, Beijing Institute of Technology, 100081, China. He is now with the School of Electrical Engineering and Telecommunications, University of New South Wales, NSW 2052, Australia. (email: weijie.yuan@unsw.edu.au)

N. Wu is with the School of Information and Electronics, Beijing Institute of Technology, 100081, China. (e-mail:wunan@bit.edu.cn).

A. Zhang and X. Huang are with the Faculty of Engineering and Information Technology, University of Technology Sydney, NSW 2007, Australia. (e-mail: andrew.zhang, xiaojing.huang@uts.edu.au).

Y. Li is with the Centre of Excellence in Telecommunications, School of Electrical and Information Engineering, University of Sydney, NSW 2006, Australia. (e-mail: yonghui.li@sydney.edu.au).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK (e-mail:lh@ecs.soton.ac.uk)

be achieved by faster-than-Nyquist (FTN) signaling. Mazo proved that in conjunction with an appropriate packing ratio, FTN transmission is capable of preserving the same bit error rate (BER) performance as classic Nyquist signaling [20]. Hence, it becomes a promising candidate for future communications applications. However, due to the nonorthogonality of the shaping pulse with respect to the symbol interval, having both long intersymbol interference (ISI) as well as colored noise at the receiver side becomes unavoidable [21]. As a result, a prohibitively high complexity may be imposed. Hence, a reduced-complexity BCJR detector is developed in [22] for FTN signaling in AWGN channels, which considered only $M$ states based on a minimum-phase model. Nevertheless, the complexity still increases exponentially with the number of ISI taps. The authors of [23], [24] added a cyclic prefix (CP) to tackle the extra ISI imposed by FTN by taking advantage of single carrier frequency-domain equalization (FDE). However, the effects of colored noise were not considered in [23] and the CP will also degrade the efficiency while almost eliminate the gain of FTN signaling. A Forney-style factor graph based detector was proposed in [25] to handle the colored noise imposed by FTN signaling for transmission over AWGN channels. An extension to doubly selective channels is considered in [26] and [27], where Gaussian message passing and variational inference techniques are employed to detect the symbols, respectively. Nevertheless, the detection of data symbols in FTN-SCMA systems is still challenging due to the interference imposed by the non-orthogonal waveforms and nonorthogonal multiple access.

Note that in the aforementioned receiver design contributions, the channel information is assumed to be perfectly known, but in practice training sequences are used for accurate channel estimation [28]. However, a joint channel estimation and data detection is capable of avoiding the use of long training sequences, whilst improving the BER performance [29]. From this perspective, researchers developed several low-complexity near-optimal joint estimation methods [30], [31]. Hence, joint channel estimation and MPA-aided detection is very attractive. A virtual zero-padding aided belief propagation (BP) algorithm was devised in [32] for joint iterative channel estimation, detection and decoding. In [33], a belief propagation aided variational expectation maximization based method was derived for MIMO-OFDM systems. Both algorithms were designed based on Nyquist signaling. In [34] and [35], time domain based BP and frequency domain based generalized approximate message passing (GAMP) joint channel estimation and decoding algorithms were proposed for FTN signaling. Nevertheless, the existing approaches do not consider systems that use both nonorthogonal waveform and nonorthogonal multiple access.

Hence, we solve this open problem by conceiving a low-complexity joint channel estimation and detection algorithm based on factor graphs (FG) and message passing algorithm designed for the uplink of FTN-SCMA systems. To model the colored noise imposed by FTN signaling, we employ the classic auto regressive (AR) process. Then the joint distribution of data symbols, channel taps and noise samples can be factorized into several local functions and represented by a factor graph. Even when based on on FG, the conventional MPA relying on the maximum *a posteriori* (MAP) criterion still has an excessive complexity order. We circumvent the problem by resorting to the expectation propagation (EP) method that restricts the message gleaned from the channel decoder to the Gaussian distribution. Compared to direct approximation via moment matching [16], the EP method aims for minimizing a specified relative entropy related to the true marginal and the trail distribution [36]. In EP, the extrinsic information fed to the channel decoder is also considered in the approximation, which enhances the BER performance. However, since the modulus of the channel coefficient does not equal to 1, the Gaussian form of messages is unavailable. To tackle this problem, we commence from the variational framework of [37] and construct a modified factor node, which facilitates the employment of variational message passing. Correspondingly, only the means and variances have to be updated iteratively and hence the complexity order of the proposed FTN-SCMA receiver only grows linearly with the number of users.

Moreover, we intrinsically amalgamate our FTN-SCMA system with a grant-free transmission scheme. It has been shown that even in busy hours, only a small percentage of users is active [38]. In the operational OMA uplink scenarios, a request-grant procedure is used: the base station (BS) schedules the uplink transmissions after receiving an explicit request from the users [39]. This procedure leads to a substantial communication overhead and excessive latency, especially for massive connectivity associated with a huge number of devices. Therefore the uplink grant-free transmission scheme is expected to significantly reduce both the communication overhead and transmission latency [10]. By contrast, in grant-free transmission, the active users directly send their signals to the BS without requiring access the grants. In order to decode the information bits from the simultaneously connected users, BS has to detect the user activity based on the received signal. Motivated by the sparsity of active users, compressive sensing (CS) based multiuser detection method was proposed in [40]. A two-stage algorithm which detects user activity using CS first and then performs channel estimation and detection was proposed in [41]. An AMP-expectation maximization (EM) solution was proposed in [42], which solved the active user detection and channel estimation problem jointly. In [43] and [44], the authors characterized the user activity and then constructed a factor graph for performing joint detection and channel estimation. In contrast to [43] and [44], in this paper we use a binary variable for representing active/inactive users. By formulating the corresponding factor graph, we propose a modified message passing algorithm for iteratively calculating the distribution of user activity. Additionally, to further reduce the receiver complexity, we use EP for approximating the binary variable by a Gaussian distribution. Accordingly, the proposed receiver exhibits a low complexity.

In summary, the main contributions of this paper are as follows.

- We intrinsically amalgamate FTN signaling with a SCMA system to transmit more data symbols using the same radio resources. As a result, a higher spectral efficiency is achieved.

- To mitigate both the colored noise and ISI imposed by FTN signaling and the inter user interference inflicted by SCMA, we design a novel receiver based on an AR model and a message passing algorithm that jointly perform channel estimation and detection. Since all messages are represented in Gaussian closed form, the proposed receiver only has a linearly increasing complexity versus the number of users.
- Finally, we amalgamate our FTN-SCMA scheme with a grant-free system which requires detection of the active users. Hence, we develop a joint user activity detection, channel estimation and decoding algorithm. With the use of the EP approximation of binary variables representing the user states, we reconstruct a specific factor node, which allows us to represent all messages in parametric forms for maintaining a low complexity.

Our simulation results show that the FTN-SCMA system relying on the proposed receiver is capable of increasing the data rate without significantly affecting the BER performance. The resultant grant-free SCMA system reliably distinguishes the active/inactive users.

The remainder of this paper is organized as follows. In Section II, we introduce the model of our FTN-SCMA system. Section III presents our low complexity algorithm proposed for joint channel estimation and decoding. In Section IV, our grant-free transmission concept and the proposed joint user activity detection, channel estimation and decoding algorithms are described. Our simulation results are provided in Section V. Finally, we draw conclusions in Section VI.

*Notations:* We use a boldface letter to denote a vector. The superscript $T$ and $-1$ denote the transpose and the inverse operations, respectively; $\mathcal{G}(m_x, v_x)$ denotes a Gaussian distribution of variable $x$ with mean $m_x$ and variance $v_x$; $\mathbb{B}^N$ denotes a N-dimensional binary number space and $\mathbb{C}^N$ denotes a N-dimensional complex number space; $\odot$ denotes the componentwise product; $|\cdot|$ denotes the modulus of a complex number or the cardinality of a set; $\|\cdot\|_2$ denotes the $\ell^2$ norm; $\propto$ represents equality up to a constant normalization factor; $\mathbf{x}\backslash x$ denotes all variables in $\mathbf{x}$ except $x$.

## II. System Model

We consider an SCMA uplink system with $K$ users and $J$ resource elements. In a NOMA system, $K > J$ is assumed and we denote $\lambda = \frac{K}{J}$ as the normalized user-load. In SCMA encoding, the coded bit streams of different users are directly mapped to $J$-dimensional SCMA codewords, i.e. $\varphi : \mathbf{c}_k \in \mathbb{B}^{\log_2 M} \rightarrow \mathbf{x}_k\mathbb{C}^J$, where $M$ is the size of the predefined SCMA codebook. For brevity we denote the codeword of user $k$ at time instant $n$ as $\mathbf{x}_k^n = [x_{k1}^n, ..., x_{kJ}^n]^T$. Due to the sparse structure of SCMA codewords, only $D < J$ elements of $\mathbf{x}_k^n$ are non-zero. Usually we use a matrix $\mathbf{F} = [\mathbf{f}_1, ..., \mathbf{f}_K]$ to capture the sparse structure of SCMA codewords. For the $k$th user, $\mathbf{f}_k$ is a $J$ dimensional vector having binary entries of $f_{kj} = 1$ if and only if the $j$th resource element is occupied by user $k$. Given this definition, the nonzero entries in the $j$th column of $F$ represent the users who occupy the $j$th resource element, while the nonzero entries in the $k$th row denote the resource elements that are occupied by user $k$.

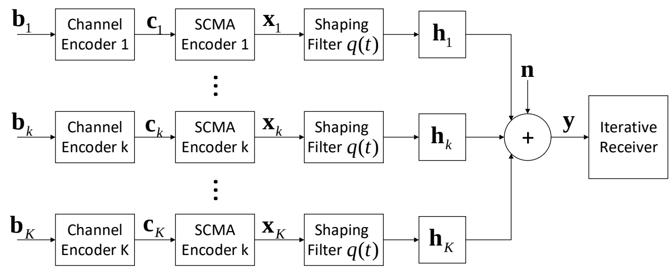



Fig. 1. Transmitter of the FTN-SCMA system considered.

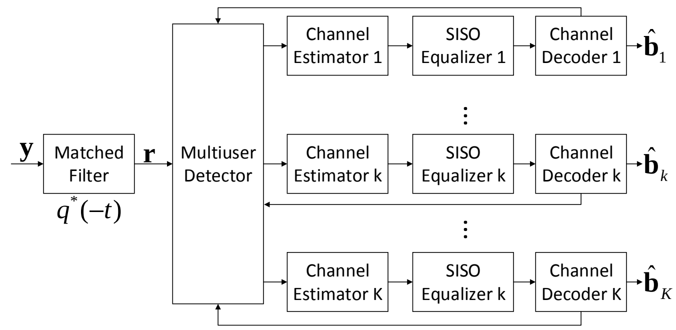



Fig. 2. Receiver structure of the FTN-SCMA system considered.

After SCMA encoding, the SCMA codewords are passed through a shaping filter $q(t)$ having the symbol period $T = \tau T_0$, where $T_0$ is the symbol interval of the Nyquist signaling and $\tau$ is the FTN packing factor. The modulated signal corresponding to user $k$ over the $j$th resource element is formulated as

$$s_{kj}(t) = \sum_n x_{kj}^n q(t - n\tau T_0). \tag{1}$$

In Nyquist signaling, $\tau = 1$ guarantees inter symbol interference (ISI) free transmission. By contrast, in FTN signaling, we use $0 < \tau < 1$ to transmit more data symbols in the same time period at the cost of introducing intentional ISI. Then the signal corresponding to user $k$ is transmitted through channel $\mathbf{h}_k = [h_{k1}, ..., h_{kJ}]^T$. The block diagram of the transmitter is shown in Fig. 1.

Assuming perfect synchronization between the users and the base station, the signal received at the BS can be expressed as,

$$\mathbf{y}(t) = \sum_{k=1}^{K} \mathbf{h}_k \odot \mathbf{s}_k(t) + \mathbf{n}(t), \tag{2}$$

where $\mathbf{s}_k(t) = [s_{k1}(t), ..., s_{kJ}(t)]^T$ denotes the modulated signals of user $k$ transmitted over all $J$ resources and $\mathbf{n}_t$ is the additive white Gaussian noise with power spectral density of $N_0$. As shown in Fig. 2, the received signal is filtered by a matched filter $q^*(-t)$. Without loss of generality, we denote $g(t) = q(t) * q^*(-t)$. Then the signal is given by

$$\mathbf{r}(t) = \sum_{k=1}^{K} \mathbf{h}_k \odot \sum_n x_{kj}^n g(t - n\tau T_0) + \boldsymbol{\omega}(t). \tag{3}$$

After sampling at rate $1/\tau T_0$, the samples at the $n$th time slot can be expressed as

$$\mathbf{r}^n = \sum_{k=1}^{K} \mathbf{h}_k \odot \tilde{\mathbf{s}}_k^n + \boldsymbol{\omega}^n, \tag{4}$$

where the $j$th entry in $\tilde{\mathbf{s}}_k^n$ is given as[1]

$$\tilde{s}_{kj}^n = \sum_{i=-L}^{L} g_i x_{kj}^{n-i}, \tag{5}$$

and $g_{n-i} = \int q(t - n\tau T_0) q^*(t - i\tau T_0) \mathrm{d}t$. In (4), $\boldsymbol{\omega}^n$ denotes the noise samples for all resource elements at time instant $n$, formulated as $\boldsymbol{\omega}^n = \int \mathbf{n}(t) q^*(t - n\tau T_0)$. Since the signal rate is above the Nyquist rate, the autocorrelation function of the noise sample $\omega_j^n, \forall\, j$ becomes

$$\mathbb{E}[\omega_j^n \omega_j^m] = N_0 g_{n-m}, \tag{6}$$

which indicates that in FTN systems the noise at the receiver side is colored. To avoid increasing the receiver complexity by using the whitening process, in the following section, we will propose an autoregressive model aided factor graph construction to overcome the colored noise and perform channel estimation and decoding.

## III. Joint Channel Estimation and Decoding Algorithm for FTN-SCMA Systems

### A. Approximation of Colored Noise

According to [45], the colored noise can be approximated by a $P$th-order autoregressive (AR) model as

$$\omega_j^n = \sum_{p=1}^{P} a_p \omega_j^{n-p} + \delta_j^n, \tag{7}$$

where $a_p$ denotes the AR process parameter and $\delta_j^n$ is the noise term with zero mean and variance $\sigma_\delta^2$. The values of $\{a_p\}$ are determined by solving the Yule-Walker equations [46].

### B. Probabilistic Model and Factor Graph Representation

Assuming that each user transmits a total of $N$ SCMA codewords and that $N$ samples are received at the BS. Our goal is to determine the *a posteriori* distribution (marginal) of the transmitted symbol $x_{kj}^n$ based on all observations at the base station $\mathbf{r}$. Then such the resultant marginal is transformed into extrinsic log likelihood ratios (LLR) and fed to the channel decoder. The marginal distribution of $x_{kj}^n$ is given by

$$p(x_{kj}^n|\mathbf{r}) \propto \int_{\mathbf{h},\boldsymbol{\omega},\mathbf{X}\backslash x_{kj}^n} p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r}), \tag{8}$$

where $\mathbf{X}$, $\mathbf{h}$ and $\mathbf{n}$ denote the transmitted symbols, channel response and colored noise samples, respectively. Instead of direct marginalization, here we further factorize the joint distribution $p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r})$ and resort to a low-complexity factor graph construction to solve the problem.

[1]In theory, the number of ISI taps induced by FTN is infinite. However in practice, we can choose sufficiently large number of taps, i.e. $2L + 1$ taps.

According to Bayes's theorem, $p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r})$ is factorized as

$$p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r}) \propto p(\mathbf{X}) \cdot p(\mathbf{h}) \cdot p(\boldsymbol{\omega}) \cdot p(\mathbf{r}|\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}). \tag{9}$$

Since the transmitted symbols and channel coefficients are independent of each other, we have

$$p(\mathbf{X})p(\mathbf{h}) = \prod_{k,j} \left[ p(h_{kj}) \prod_n p(x_{kj}^n) \right], \tag{10}$$

where $p(x_{kj}^n)$ is obtained from the output LLR of the channel decoder. The *a priori* distribution $p(\boldsymbol{\omega})$ can be factorized based on the AR model as

$$p(\boldsymbol{\omega}) \propto \prod_j \prod_n \underbrace{\exp\left( -\frac{\omega_j^n - \sum_{p=1}^{P} a_p \omega_j^{n-p}}{2\sigma_\delta^2} \right)}_{\psi_j^n}. \tag{11}$$

Conditioned on $\omega_j^n$, the observations $r_j^n$ at different instants $n$ are independent. As shown in [18], using an auxiliary variable helps us to reduce the computation load. Therefore we factorize $p(\mathbf{r}|\mathbf{X}, \mathbf{h}, \boldsymbol{\omega})$ as

$$p(\mathbf{r}|\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}) \propto \prod_{j,n} \underbrace{\delta(r_j^n - \sum_{k=1}^{K} \left[ h_{kj} \tilde{s}_{kj}^n \right] - \omega_j^n)}_{f_j^n} \cdot \underbrace{\delta(\tilde{s}_{kj}^n - \sum_{i=-L}^{L} g_i x_{kj}^{n-i})}_{\phi_{kj}^n}. \tag{12}$$

Based on the factorization (10)-(12), the joint distribution $p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}|\mathbf{r})$ can be represented by a factor graph, as shown in Fig. 3, where the message passing algorithm is executed to determine the unknown variables. Note that in the factorization (12), the local function contains different number of variables. Hence, the resultant factor graph is an irregular one.

### C. Message Passing Receiver Design

The conventional message passing algorithm (MPA) consists of two kinds of messages. Following the sum product algorithm, the message passed from factor vertex $f$ to variable node $x$ is given by

$$\mu_{f\to x}(x) \propto \int f(\mathbf{x}) \prod_{x' \in \mathcal{S}(f)\backslash\{x\}} \mu_{x'\to f}(x') \mathrm{d}x', \tag{13}$$

and the message forwarded from $x$ to $f$ is defined as

$$\mu_{x\to f}(x) \propto \prod_{f' \in \mathcal{S}(x)\backslash\{f\}} \mu_{f'\to x}(x), \tag{14}$$

where $\mathcal{S}(f)$ and $\mathcal{S}(x)$ denotes the set of variable vertices connected to $f$ and the set of factor vertices connected to $x$, respectively. The belief (marginal) of variable $x$ is then given by $b(x) = \prod_{f\in\mathcal{S}(x)} \mu_{f\to x}(x)$. Next, we consider the derivations of messages on the factor graph in Fig. 3.

In "turbo" equalization, the channel decoder and the equalizer iteratively exchange extrinsic information. Since the decoding issues are beyond the scope of this paper, we simply

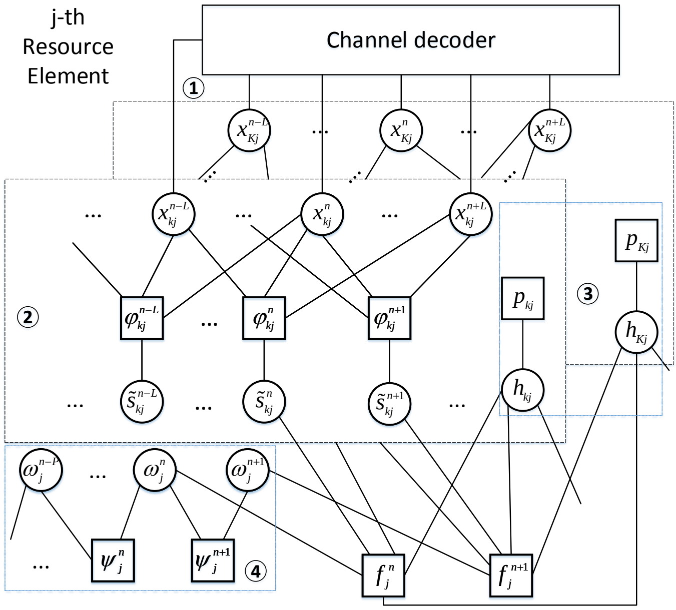



Fig. 3. Factor graph representation of the $j$th resource element, where the shorthand notations $p_{kj} = p(h_{kj})$. The factor graph is separated into four parts, i.e. decoding part denoted by ①, equalization part denoted by ②, channel estimation part denoted by ③, and colored noise part denoted by ④.

mention that optimal BP decoding [47] is used by the channel decoder. After decoding, the output LLRs become

$$L^a(c_{n,m}) = \frac{p(c_{n,m}=0)}{p(c_{n,m}=1)}, \tag{15}$$

where the subscripts $n$ and $m$ denote the $n$th coded bit and the $m$th constellation point, respectively. Then the LLRs are transformed to the *a priori* distribution of $x_{kj}^n$, which is used for the next global "turbo" iteration, i.e., we have:

$$p(x_{kj}^n) = \sum_{i=1}^{M} p_i \delta(x_{kj}^n - \chi_i), \tag{16}$$

where $\chi_i$ represents a constellation point of the SCMA encoder, $p_i$ is the associated probability and $M$ is the modulation order. Although the discrete distribution $p(x_{kj}^n)$ can indeed be used as the incoming message, the complexity of the conventional MPA receiver relying on the MAP criterion will increase exponentially with the number of interfering symbols. Here we resort to the Kullback-Leibler divergence based method of [36], also known as expectation propagation (EP) to approximate the incoming message by a Gaussian distribution. Explicitly, we aim for finding the Gaussian distribution that minimizes the Kullback-Leibler divergence [48],

$$b_G(x_{kj}^n) = \arg\min_{b_G} \int b_G(x_{kj}^n) \ln \frac{b_G(x_{kj}^n)}{b(x_{kj}^n)} \mathrm{d}x_{kj}^n, \tag{17}$$

where $b_G$ belongs to the family of Gaussian distributions and $b(x_{kj}^n)$ is the marginal distribution of the variable $x_{kj}^n$. The minimization formulated in (17) is equivalent to matching the moments of $b(x_{kj}^n)$. Assuming that the outgoing message has a mean and variance of $m_{x_{kj}^n}^e$ and $v_{x_{kj}^n}^e$, it is easy to obtain the mean and variance of $b_G(x_{kj}^n)$ as $m_{x_{kj}^n}$ and $v_{x_{kj}^n}$, respectively. Then the Gaussian approximation of $p(x_{kj}^n)$ has the mean and variance

$$m_{x_{kj}^n}^0 = v_{x_{kj}^n}^0 \left( \frac{m_{x_{kj}^n}}{v_{x_{kj}^n}} - \frac{m_{x_{kj}^n}^e}{v_{x_{kj}^n}^e} \right), \tag{18}$$

$$v_{x_{kj}^n}^0 = \left( \frac{1}{v_{x_{kj}^n}} - \frac{1}{v_{x_{kj}^n}^e} \right)^{-1}. \tag{19}$$

Having determined $m_{x_{kj}^n}^0$ and $v_{x_{kj}^n}^0$, we can now calculate the message in the equalization part. Again, we assume that the message $\mu_{\tilde{s}_{kj}^n \to \phi_{k,j}^n} = \mu_{f_j^n \to \tilde{s}_{kj}^n}$ has been obtained as

$$\mu_{\tilde{s}_{kj}^n \to \phi_{k,j}^n} = \mathcal{G}(m_{\tilde{s}_{kj}^n \to \phi_{k,j}^n}, v_{\tilde{s}_{kj}^n \to \phi_{k,j}^n}). \tag{20}$$

Then the message $\mu_{\phi_{k,j}^n \to x_{kj}^{n+l}}$ can be written as:

$$m_{\phi_{k,j}^n \to x_{kj}^{n+l}} = m_{\tilde{s}_{kj}^n \to \phi_{k,j}^n} - \sum_{i=-L, i\neq l}^{L} g_i m_{x_{kj}^{n+i} \to \phi_{k,j}^n}, \tag{21}$$

$$v_{\phi_{k,j}^n \to x_{kj}^{n+l}} = v_{\tilde{s}_{kj}^n \to \phi_{k,j}^n} + \sum_{i=-L, i\neq l}^{L} g_i^2 v_{x_{kj}^{n+i} \to \phi_{k,j}^n}. \tag{22}$$

Usually, calculating $\mu_{x_{kj}^n \to \phi_{k,j}^n}$ for different factor nodes $\phi_{k,j}^{n+l}|_{l=-L}^{L}$ following (14) requires that the product of messages is calculated $(2L+1)$ times. Motivated by the fact that we have $\mu_{x_{kj}^n \to \phi_{k,j}^n} \cdot \mu_{\phi_{k,j}^n \to x_{kj}^n} = b_G(x_{kj}^n)$, the objective message can be calculated at a linear complexity as $\mu_{x_{kj}^n \to \phi_{k,j}^n} = b_G(x_{kj}^n)/\mu_{\phi_{k,j}^n \to x_{kj}^n}$ with the aid of:

$$v_{x_{kj}^n \to \phi_{k,j}^n} = \left( \frac{1}{v_{x_{kj}^n}} - \frac{1}{v_{\phi_{k,j}^n \to x_{kj}^n}} \right)^{-1}, \tag{23}$$

$$m_{x_{kj}^n \to \phi_{k,j}^n} = v_{x_{kj}^n \to \phi_{k,j}^n} \left( \frac{m_{x_{kj}^n}}{v_{x_{kj}^n}} - \frac{m_{\phi_{k,j}^n \to x_{kj}^n}}{v_{\phi_{k,j}^n \to x_{kj}^n}} \right). \tag{24}$$

After obtaining all messages $\mu_{\phi_{k,j}^{n+l} \to x_{kj}^n}|_{l=-L}^{L}$, the mean and variance of the extrinsic message provided for the channel decoder are given by

$$v_{x_{kj}^n}^e = \left( \sum_{l=-L}^{L} 1/v_{\phi_{k,j}^{n+l} \to x_{kj}^n} \right)^{-1}, \tag{25}$$

$$m_{x_{kj}^n}^e = v_{x_{kj}^n}^e \left( \sum_{l=-L}^{L} \frac{m_{\phi_{k,j}^{n+l} \to x_{kj}^n}}{v_{\phi_{k,j}^{n+l} \to x_{kj}^n}} \right). \tag{26}$$

Based on $m_{x_{kj}^n}^e$ and $v_{x_{kj}^n}^e$, the extrinsic LLRs are calculated as

$$L^e(c_{n,m}) = \ln \frac{\sum_{\chi_i \in \mathcal{A}_m^0} \exp\left( -\frac{-|\chi_i - m_{x_{kj}^n}^e|^2}{v_{x_{kj}^n}^e} \right) p_{m' \neq m}}{\sum_{\chi_i \in \mathcal{A}_m^1} \exp\left( -\frac{-|\chi_i - m_{x_{kj}^n}^e|^2}{v_{x_{kj}^n}^e} \right) p_{m' \neq m}}, \tag{27}$$

where $p_{m' \neq m} = \prod_{m' \neq m} p(c_{n,m'} = s_{i,m'})$ and $s_{i,m} = \{0,1\}$ denotes the $m$th label of the constellation point $\chi_i$, while $\mathcal{A}_m^0$ and $\mathcal{A}_m^0$ are the subsets of all constellation points $\chi_i$ satisfying $s_{i,m} = 0$ or 1. Then the extrinsic LLRs are fed to the channel decoder for determining the data bits $\hat{\mathbf{b}}_k$ of the users.

Next, let us consider the message updating in the colored noise part. Since the non-orthogonality of FTN signaling does not affect the first order moment of noise samples, the mean of noise samples $\mathbb{E}[\omega_j^n]=0$ and we only focus our attention on the evolution of its variance. According to (11), the variance $v_{\psi_j^n\to\omega_j^n}$ is expressed as

$$v_{\psi_j^n\to\omega_j^n}=\sigma_\delta^2+\sum_{p=1}^{P}(a^p)^2 v_{\omega_j^{n-p}\to\psi_j^n}. \tag{28}$$

It should be noted that the colored noise represents a causal system where the sample at instant $n$ only depends on the previous noise samples. Therefore the message forwarded from $\omega_j^n$ to $f_j^n$ is identical to $\mu_{\psi_j^n\to\omega_j^n}$, i.e. we have $v_{\omega_j^n\to f_j^n}=v_{\psi_j^n\to\omega_j^n}$.

For the channel estimation part, the message $\mu_{h_{kj}\to f_j^n}$ is readily determined according to the SPA rules as

$$\mu_{h_{kj}\to f_j^n}=p(h_{kj})\prod_{n'\neq n}\mu_{f_j^{n'}\to h_{kj}}, \tag{29}$$

where $p(h_{kj})$ is usually coarsely evaluated by using a sequence of pilot symbols, which can be modeled as a Gaussian distributed variable with a mean of $m_{h_{kj}}^0$ and variance of $v_{h_{kj}}^0$. We assume that $\mu_{f_j^{n'}\to h_{kj}}$ has also been obtained in the Gaussian form as $\mu_{f_j^{n'}\to h_{kj}}=(m_{f_j^{n'}\to h_{kj}},v_{f_j^{n'}\to h_{kj}})$. Hence $\mu_{h_{kj}\to f_j^n}$ has a mean and variance of

$$m_{h_{kj}\to f_j^n}=v_{h_{kj}\to f_j^n}\left(\frac{m_{h_{kj}}^0}{v_{h_{kj}}^0}+\sum_{n'\neq n}\frac{m_{f_j^{n'}\to h_{kj}}}{v_{f_j^{n'}\to h_{kj}}}\right) \tag{30}$$

$$v_{h_{kj}\to f_j^n}=\left(\frac{1}{v_{h_{kj}}^0}+\sum_{n'\neq n}\frac{1}{v_{f_j^{n'}\to h_{kj}}}\right)^{-1}. \tag{31}$$

The belief $b(h_{kj})$ is obtained by adding the terms $m_{f_j^{n'}\to h_{kj}}/v_{f_j^{n'}\to h_{kj}}$ and $1/v_{f_j^{n'}}$ with index $n'=n$ into (30) and (31). Then the maximum *a posteriori* (MAP) estimator can be used for determining the estimate of the channel coefficient by $\hat{h}_{kj}=\arg\max_{h_{kj}} b(h_{kj})$. Since $b(h_{kj})$ obeys a Gaussian distribution, the MAP estimate $\hat{h}_{kj}$ is the mean of $b(h_{kj})$

Above we have derived closed form Gaussian messages in four parts of the factor graph. However, they are based on the fact that the messages gleaned from $f_j^n$ to its connected variable vertices obey Gaussian distributions. In what follows, we will calculate the messages related to vertex $f_j^n$. Following (13), the message $\mu_{f_j^n\to\tilde{s}_{kj}^n}$ is expressed as

$$\begin{aligned}\mu_{f_j^n\to\tilde{s}_{kj}^n}&\propto\int\delta(r_j^n-\sum_{k=1}^{K}[h_{kj}\tilde{s}_{kj}^n]-\omega_j^n)\mu_{\omega_j^n\to f_j^n}\prod_k\mu_{h_{kj}\to f_j^n}\\&\qquad\prod_{k'\neq k}\mu_{\tilde{s}_{k'j}^n\to f_j^n}\mathrm{d}h_{kj}\mathrm{d}\omega_j^n\mathrm{d}\tilde{s}_{k'j}^n\\&\propto\int\exp\left(-\frac{|r_j^n-\sum_{k=1}^{K}[m_{h_{kj}\to f_j^n}\tilde{s}_{kj}^n]|^2}{v_{\omega_j^n\to f_j^n}+\sum_{k=1}^{k}|\tilde{s}_{kj}^n|^2 v_{h_{kj}\to f_j^n}}\right)\prod_{k'\neq k}\mu_{\tilde{s}_{k'j}^n\to f_j^n}\mathrm{d}\tilde{s}_{k'j}^n.\end{aligned} \tag{32}$$

From (32), we can see that when calculating message $\mu_{f_j^n\to\tilde{s}_{kj}^n}$, the variable $\tilde{s}_{kj}^n$ appears in both the numerator and denominator of the exponential term, which makes the conventional MPA unsuitable. Here we resort to the variational message passing (VMP) method of [49] where the message forwarded from factor vertex $f$ to variable vertex $x$ is formulated as

$$\mu_{f\to x}(x)\propto\exp\left(\int\ln f(\mathbf{x})\prod_{x'\in\mathcal{S}(f)\backslash\{x\}}\mu_{x'\to f}(x')\mathrm{d}x'\right). \tag{33}$$

By employing the VMP rules for the message $\mu_{f_j^n\to\tilde{s}_{kj}^n}$, we have

$$\begin{aligned}\mu_{f_j^n\to\tilde{s}_{kj}^n}\propto\exp\Bigg(\int-\frac{|r_j^n-\sum_{k=1}^{K}[h_{kj}\tilde{s}_{kj}^n]|^2}{v_{\omega_j^n\to f_j^n}}\prod_k\mu_{h_{kj}\to f_j^n}\\\cdot\prod_{k'\neq k}\mu_{\tilde{s}_{k'j}^n\to f_j^n}\mathrm{d}h_{kj}\mathrm{d}\tilde{s}_{k'j}^n\Bigg).\end{aligned} \tag{34}$$

After straightforward manipulations, the message (34) can be obtained in Gaussian form with a mean of

$$m_{f_j^n\to\tilde{s}_{kj}^n}=\frac{(r_j^n-\sum_{k'=1,k\neq k}^{K}m_{h_{k'j}\to f_j^n}m_{\tilde{s}_{k'j}^n\to f_j^n})m_{h_{kj}\to f_j^n}}{|m_{h_{kj}\to f_j^n}|^2+v_{h_{kj}\to f_j^n}}, \tag{35}$$

and variance of

$$v_{f_j^n\to\tilde{s}_{kj}^n}=\frac{v_{\omega_j^n\to f_j^n}}{|m_{h_{kj}\to f_j^n}|^2+v_{h_{kj}\to f_j^n}}. \tag{36}$$

The message $\mu_{f_j^n\to h_{kj}}$ can be calculated similarly by the VMP, whose mean and variance are

$$m_{f_j^n\to h_{kj}}=\frac{(r_j^n-\sum_{k'=1,k\neq k}^{K}m_{h_{k'j}\to f_j^n}m_{\tilde{s}_{k'j}^n\to f_j^n})m_{\tilde{s}_{kj}^n\to f_j^n}}{|m_{\tilde{s}_{kj}^n\to f_j^n}|^2+v_{\tilde{s}_{kj}^n\to f_j^n}}, \tag{37}$$

$$v_{f_j^n\to h_{kj}}=\frac{v_{\omega_j^n\to f_j^n}}{|m_{\tilde{s}_{kj}^n\to f_j^n}|^2+v_{\tilde{s}_{kj}^n\to f_j^n}}. \tag{38}$$

### D. Algorithmic Summary

Using appropriate approximations, all messages on the factor graph are represented in parametric forms, which significantly reduces the computational complexity of the conventional MPA receiver. Compared to the existing advanced MPA receiver, the computational complexity associated with the introduction of an auxiliary variable and modified message updating rules only increases linearly with the number of users and resource elements. The details of the proposed joint channel estimation and FTN-SCMA detector are presented in Algorithm. I.

## IV. User Activity Detection in Grant-free FTN-SCMA Systems

In the state-of-the-art grant-based uplink transmission, before transmitting information, a user first sends a scheduling request to the base station and then waits for the scheduling grant from the base station. This kind of 'handshaking' process introduces excessive signaling overhead and latency, especially for massive access networks. In a grant-free system, a user do not have to wait for the base station to assign a resource

**Algorithm 1** Joint Channel Estimation and Decoding Algorithm for FTN-SCMA System

**Initialization**:
At the first turbo iteration, initialize all undetermined messages as Gaussian distribution with zero mean and unit variance;
Using pilot sequence to coarsely estimate the mean $m^0_{h_{kj}}$ and variance $v^0_{h_{kj}}$ of channel coefficient.
**for** iter=1:$N_{iter}$ **do**
  Compute the means and variances of messages in equalization part according to (21)-(24);
  Compute the message from factor vertex $f^n_j$ to variable vertices $x^n_{kj}$ and $h_{kj}$ according to (35)-(38);
  Compute the variance $v_{\psi^n_j \to \omega^n_j}$ according to (28);
  Compute the message from $h_{kj}$ to factor vertex $f^n_j$ via (30) and (31);
  Compute the mean and variance of message to channel decoder according to (25) and (26);
  Convert the outgoing messages to LLR and feed them to the channel decoder;
  Perform BP decoding;
  Convert the extrinsic LLRs to Gaussian messages using (18) and (19);
**end for**
Determine the estimate of channel coefficient by MAP estimator.

element before sending their signals to the BS. Therefore the signal latency is significantly reduced. In the existing works, a precision parameter is used for estimating the channel's sparsity. However, this results in more short loops in the factor graph and increases the receiver complexity. In this section, we propose an algorithm for FTN-SCMA systems that directly determines the user activity while performing channel estimation and decoding.

Let us use a binary variable $\xi_k = \{0, 1\}$ to represent the user activity, i.e. $\xi_k = 1$ indicates that user $k$ is active and vice versa. Then the $n$th sample at the $j$th resource element is expressed by

$$r^n_j = \sum_{k=1}^{K} h_{kj}\xi_k \tilde{s}^n_{kj} + \omega^n_j. \tag{39}$$

The *a priori* distribution of $\xi_k$ is a Bernoulli distribution given by

$$p^0(\xi_k) = p_1^{\xi_k}(1-p_1)^{1-\xi_k}, \tag{40}$$

where $p_1$ is the *a priori* knowledge of user activity based on existing data.

### A. Probability Based Active User Detection Algorithm

To determine the activity of user $k$, we have to calculate its probability $\gamma_k$ of being active based on the received samples. To this end, we modify the factor graph structure by including the probability $\gamma_k$ of $\xi_k = 1$. The corresponding part of the factor graph is illustrated in Fig. 4. Here we introduce a weighting factor $\gamma_{kj}$ that represents the probability of $\xi_{kj} = 1$

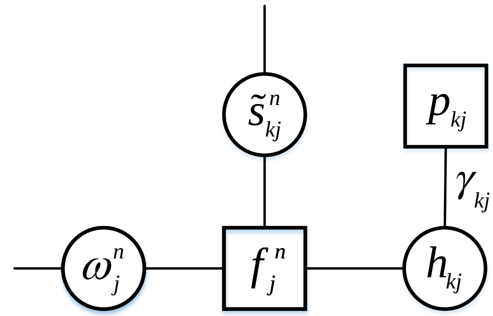



Fig. 4. Modified factor graph structure including user activity.

on the specific edge connected to the vertices $h_{kj}$ to represent the weight on this edge.

Based on Fig. 4, the message passed from $f^n_j$ to $h_{kj}$ has a mean of $m_{f^n_j \to h_{kj}}$ and variance of $v_{f^n_j \to h_{kj}}$ according to (37) and (38). Hence we arrive at the intrinsic message for $\xi_{kj}h_{kj}$ with a mean and variance of

$$\overrightarrow{m}_{\xi_{kj}h_{kj}} = \overrightarrow{v}_{\xi_{kj}h_{kj}} \sum_n \frac{m_{f^n_j \to h_{kj}}}{v_{f^n_j \to h_{kj}}} \tag{41}$$

$$\overrightarrow{v}_{\xi_{kj}h_{kj}} = \left( \sum_n \frac{1}{v_{f^n_j \to h_{kj}}} \right)^{-1}. \tag{42}$$

The distribution of $\xi_{kj}$ is obtained by integrating $h_{kj}$ over the joint distribution, formulated as

$$\begin{aligned} p(\xi_{kj}) \propto & \int \exp\left( -\frac{(\xi_{kj}h_{kj} - \overrightarrow{m}_{\xi_{kj}h_{kj}})^2}{\overrightarrow{v}_{\xi_{kj}h_{kj}}} \right) \\ & \cdot \exp\left( -\frac{(h_{kj} - m^0_{h_{kj}})^2}{v^0_{h_{kj}}} \right) \mathrm{d}h_{kj} \\ \propto & \exp\left( -\frac{(\xi_{kj}m^0_{h_{kj}} - \overrightarrow{m}_{\xi_{kj}h_{kj}})^2}{\xi^2_{kj}v^0_{h_{kj}} + \overrightarrow{v}_{\xi_{kj}h_{kj}}} \right). \end{aligned} \tag{43}$$

Then the probability $\gamma_{kj}$ is updated as

$$\begin{aligned} \gamma_{kj} &= \frac{p(\xi_{kj} = 1)}{p(\xi_{kj} = 0) + p(\xi_{kj} = 1)} \\ &= \frac{1}{1 + \frac{p(\xi_{kj}=0)}{p(\xi_{kj}=1)}}. \end{aligned} \tag{44}$$

Once the probability $\xi_{kj} = 1$ has been obtained, the probability $\tilde{\gamma}_k$ conditioned on all received signal samples is readily determined as

$$\tilde{\gamma}_k = \frac{\prod_j \gamma_{kj}}{\prod_j \gamma_{kj} + \prod_j (1 - \gamma_{kj})}. \tag{45}$$

Considering the *a priori* probability $p_1$, we get $\gamma_k$ as follows,

$$\gamma_k = \frac{p_1 \tilde{\gamma}_k}{p_1 \tilde{\gamma}_k + (1 - p_1)(1 - \tilde{\gamma}_k)}. \tag{46}$$

To determine the value of $\xi_k$, we set a threshold $\beta$ according to empirical evidence. Then we say that user $k$ is deemed to be active if $\gamma_k \geq \beta$ and vice versa.

**Algorithm 2** User Activity Detection Algorithm I

1. Run Algorithm 1;
2. Calculate the intrinsic message to $h_{kj}$ according to (41) and (42);
3. Determine the probability $\gamma_{kj}$ by (44);
4. Calculate $\gamma_k$ according to (46) and decide $\xi_k$;
5. Approximate the message $\mu_{h_{kj}\to f_j^n}$ to Gaussian and continue running algorithm 1.

The extrinsic message passed on from $h_{kj}$ to $f_j^n$ is still obtained by $\mu_{h_{kj}\to f_j^n} = \mu_{p_{kj}\to h_{kj}} \prod_{n'\neq n} \mu_{f_j^{n'}\to h_{kj}}$. Specifically, when calculating $\mu_{p_{kj}\to h_{kj}}$, we combine $\xi_{kj}$ and $h_{kj}$ to create a new variable,

$$\mu_{p_{kj}\to h_{kj}} \propto \left( \gamma_{kj} e^{-\frac{(h_{kj}-m^0_{h_{kj}})^2}{v^0_{h_{kj}}}} + (1-\gamma_{kj}) e^{-\frac{[m^0_{h_{kj}}]^2}{v^0_{h_{kj}}}} \right). \tag{47}$$

It is plausibly that $\mu_{p_{kj}\to h_{kj}}$ is a Gaussian mixture distribution (GMD) and $\mu_{h_{kj}\to f_j^n}$ is also a GMD. In conjunction with the message passing receiver of Section III, we approximate $\mu_{h_{kj}\to f_j^n}$ by a Gaussian distribution having a mean and variance as

$$m_{h_{kj}\to f_j^n} = \mathbb{E}_{\mu_{h_{kj}\to f_j^n}}[h_{kj}] \tag{48}$$

$$v_{h_{kj}\to f_j^n} = \mathbb{E}_{\mu_{h_{kj}\to f_j^n}}[h_{kj}^2] - m^2_{h_{kj}\to f_j^n}. \tag{49}$$

In Algorithm 2, the user activity detection based on the message passing algorithm is described. To sum up, we can see that the algorithm can be readily extended from the algorithm proposed in Section III and we only have to do small modification in the factor graph. However, since $\xi_{kj}$ has to be calculated separately, the receiver complexity increases. Furthermore, the derivation of messages is not straightforward from the perspective of probabilistic factorization. In the following subsection, we will propose another active user detection method having a reduced complexity .

### *B. Message Passing Based Active User Detection Algorithm*

To create a concise form of the message passing receiver in the factor graph framework of Fig. 3, we add $\xi_k$ as a new variable vertex to the factor graph. According to (39), we use a dirac Delta function $\delta(\bar{s}^n_{kj} - \xi_k \tilde{s}^n_{kj})$ to represent the multiplication relationship of $\bar{s}^n_{kj} = \xi_k \tilde{s}^n_{kj}$. Accordingly, the joint likelihood function (12) is revised as

$$p(\mathbf{r}|\mathbf{X},\mathbf{h},\boldsymbol{\omega},\boldsymbol{\xi}) \propto \prod_{j,n} \underbrace{\delta(r_j^n - \sum_{k=1}^{K}\left[h_{kj}\bar{s}^n_{kj}\right] - \omega_j^n)}_{f_j^n}$$
$$\cdot\, \delta(\bar{s}^n_{kj} - \xi_k \tilde{s}^n_{kj}) \cdot \underbrace{\delta(\tilde{s}^n_{kj} - \sum_{i=-L}^{L} g_i x_{kj}^{n-i})}_{\phi^n_{kj}}, \tag{50}$$

and the factor graph is modified as shown in Fig. 5. Since $\xi_k$ is a binary variable, it has a discrete distribution. Then upon

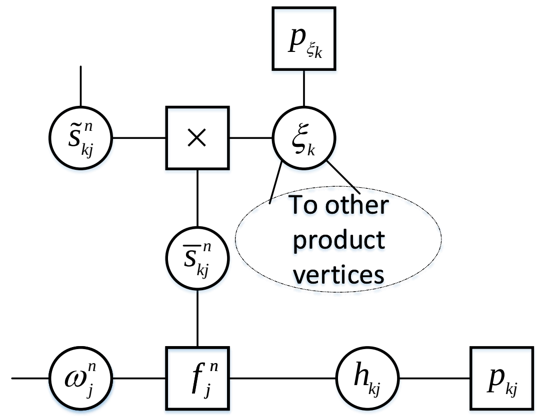



Fig. 5. Modified factor graph structure. The product node $\times^n_{kj}$ represents the constraint $\delta(\bar{s}^n_{kj} - \xi_k \tilde{s}^n_{kj})$.

considering the message updating, if the messages arriving from different product vertices to $\xi_k$ are Gaussian, the resultant message $\mu_{\xi_k\to\times^n_{kj}}$ follows a Gaussian mixture distribution, which is unsuitable for deriving Gaussian messages. To tackle this problem, we approximate the message forwarded from $\xi_k$ to the product vertex $\times^n_{kj}$ by a Gaussian distribution via expectation propagation.

Following the classic SPA rules, the belief of $\xi_k$ is $b(\xi_k) = \mu_{\xi_k\to p_{\xi_k}} p(\xi_k)$. Assuming that $\mu_{\xi_k\to p_{\xi_k}}$ is Gaussian with a mean of $m_{\xi_k\to p_{\xi_k}}$ and variance of $v_{\xi_k\to p_{\xi_k}}$, the mean and variance of $b(\xi_k)$ becomes

$$m_{\xi_k} = \frac{p_1 \exp(-\frac{1-2m_{\xi_k\to p_{\xi_k}}}{v_{\xi_k\to p_{\xi_k}}})}{p_1\left[\exp(-\frac{1-2m_{\xi_k\to p_{\xi_k}}}{v_{\xi_k\to p_{\xi_k}}}) - 1\right] + 1}, \tag{51}$$

$$v_{\xi_k} = m_{\xi_k} - m^2_{\xi_k}. \tag{52}$$

A plausible observation is that in (51) the absolute value of the exponential term dominates the value of $m_{\xi_k}$, and $v_{\xi_k}$ becomes smaller when $\xi_k$ approaches 0 or 1. Hence, after running several iterations, the belief of $\xi_k$ becomes more 'concentrated'. Having determined $m_{\xi_k}$ and $v_{\xi_k}$, we can now readily determine the Gaussian approximation of the message $\mu_{\xi_k\to\times^n_{kj}}$, which is

$$\mu_{\xi_k\to\times^n_{kj}} \sim \mathcal{G}\left( \frac{m_{\xi_k} v_{\xi_k} - m_{\times^n_{kj}\to\xi_k} v_{\times^n_{kj}\to\xi_k}}{v_{\xi_k} - v_{\times^n_{kj}\to\xi_k}}, \frac{v_{\xi_k} v_{\times^n_{kj}\to\xi_k}}{v_{\xi_k} - v_{\times^n_{kj}\to\xi_k}} \right). \tag{53}$$

Since we have $\mu_{\xi_k\to\times^n_{kj}}$ and $\mu_{\tilde{s}^n_{kj}\to\times^n_{kj}} = \mu_{\phi^n_{kj}\to\tilde{s}^n_{kj}}$, the mean and variance of message $\mu_{\bar{s}^n_{kj}\to f_j^n}$ for the product vertex are given by

$$m_{\bar{s}^n_{kj}\to f_j^n} = m_{\xi_k\to\times^n_{kj}} m_{\phi^n_{kj}\to\tilde{s}^n_{kj}} \tag{54}$$

$$m_{\bar{s}^n_{kj}\to f_j^n} = v_{\xi_k\to\times^n_{kj}} m^2_{\phi^n_{kj}\to\tilde{s}^n_{kj}} + (m^2_{\xi_k\to\times^n_{kj}} + v_{\xi_k\to\times^n_{kj}}) v_{\phi^n_{kj}\to\tilde{s}^n_{kj}}. \tag{55}$$

The detailed derivations of (54) and (55) are given in Appendix A. By contrast, the message $\mu_{\bar{s}^n_{kj}\to\times^n_{kj}}$ is the same as the message calculated in (35) and (36). Next, we calculate the messages forwarded from $\times_{kj}^n$ to $\tilde{s}_{kj}^n$ and $\xi_k$. Again, a similar problem as that in Section III occurs: even if $\mu_{\tilde{s}_{kj}^n \to \times_{kj}^n}$ is Gaussian, it is not possible to formulate Gaussian distributed messages for $\xi_k$. To overcome this challenge, we adopt the Kullback-Leibler divergence. By grouping the message $\mu_{\bar{s}_{kj}^n \to \times_{kj}^n}$ and the constraint into a new factor node, the joint distribution $p(\tilde{s}_{kj}^n, \xi_k)$ is formulated as

$$p(\tilde{s}_{kj}^n, \xi_k) \propto \exp\left(-\frac{(m_{\bar{s}_{kj}^n \to \times_{kj}^n} - \xi_k \tilde{s}_{kj}^n)^2}{v_{\bar{s}_{kj}^n \to \times_{kj}^n}}\right) \mu_{\xi_k \to \times_{kj}^n} \cdot \mu_{\tilde{s}_{kj}^n \to \times_{kj}^n}. \tag{56}$$

According to the variational inference framework, we consider using $b(\xi_k)b(\tilde{s}_{kj}^n)$ to approximate (56). The Kullback Leibler divergence is given by [48]

$$\begin{aligned}\mathrm{KLD}(\xi_k, \tilde{s}_{kj}^n) &= \int b(\xi_k)b(\tilde{s}_{kj}^n) \ln \frac{b(\xi_k)b(\tilde{s}_{kj}^n)}{p(\tilde{s}_{kj}^n, \xi_k)} \mathrm{d}\xi_k \mathrm{d}\tilde{s}_{kj}^n \\ &= -\int b(\xi_k)\left[\int \ln p(\tilde{s}_{kj}^n, \xi_k) b(\tilde{s}_{kj}^n) \mathrm{d}\tilde{s}_{kj}^n\right] \mathrm{d}\xi_k \\ &\quad + \int b(\xi_k) \ln b(\xi_k) \mathrm{d}\xi_k + C,\end{aligned} \tag{57}$$

where $C$ denotes a constant. To minimize the KLD, it may be shown that

$$b(\xi_k) = \exp\left(\int \ln p(\tilde{s}_{kj}^n, \xi_k) b(\tilde{s}_{kj}^n)\right). \tag{58}$$

Substituting (56) into (58) yields

$$\frac{b(\xi_k)}{\mu_{\xi_k \to \times_{kj}^n}} \propto \exp\left(-\xi_k^2 \frac{m_{\tilde{s}_{kj}^n}^2 + v_{\tilde{s}_{kj}^n}}{v_{\bar{s}_{kj}^n \to \times_{kj}^n}} + 2\xi_k \frac{m_{\bar{s}_{kj}^n \to \times_{kj}^n} m_{\tilde{s}_{kj}^n}}{v_{\bar{s}_{kj}^n \to \times_{kj}^n}}\right), \tag{59}$$

where $v_{\tilde{s}_{kj}^n} = (v^{-1}_{\tilde{s}_{kj}^n \to \times_{kj}^n} + v^{-1}_{\times_{kj}^n \to \tilde{s}_{kj}^n})^{-1}$ and $m_{\tilde{s}_{kj}^n} = v_{\tilde{s}_{kj}^n}(m_{\tilde{s}_{kj}^n \to \times_{kj}^n} v^{-1}_{\tilde{s}_{kj}^n \to \times_{kj}^n} + m_{\times_{kj}^n \to \tilde{s}_{kj}^n} v^{-1}_{\times_{kj}^n \to \tilde{s}_{kj}^n})$. Therefore the message $\mu_{\times_{kj}^n \to \xi_k}$ is determined to be Gaussian with a mean and variance of

$$m_{\times_{kj}^n \to \xi_k} = \frac{m_{\bar{s}_{kj}^n \to \times_{kj}^n} m_{\tilde{s}_{kj}^n}}{m_{\tilde{s}_{kj}^n}^2 + v_{\tilde{s}_{kj}^n}} \tag{60}$$

$$v_{\times_{kj}^n \to \xi_k} = \frac{v_{\bar{s}_{kj}^n \to \times_{kj}^n}}{m_{\tilde{s}_{kj}^n}^2 + v_{\tilde{s}_{kj}^n}}. \tag{61}$$

Similarly, we have the Gaussian message $\mu_{\times_{kj}^n \to \tilde{s}_{kj}^n}$ represented as

$$\mu_{\times_{kj}^n \to \tilde{s}_{kj}^n} \propto \mathcal{G}\left(\frac{m_{\xi_k \to \times_{kj}^n} m_{\xi_k}}{m_{\xi_k}^2 + v_{\xi_k}}, \frac{v_{\xi_k \to \times_{kj}^n}}{m_{\xi_k}^2 + v_{\xi_k}}\right), \tag{62}$$

where $m_{\xi_k}$ and $v_{\xi_k}$ are the mean and variance of $b(\xi_k)$. For the other product vertices, the messages to $\xi_k$ can also be obtained in Gaussian form following the update rules derived in (59) - (61) based on the message $\mu_{\xi_k \to \times_{kj}^n}$ determined in the previous iteration. Having $\mu_{\times_{kj}^n \to \xi_k}$ in the Gaussian form, the mean and variance of the Gaussian message $\mu_{\xi_k \to p_{\xi_k}}$ can now be calculated by straightforward manipulations. The value of $\xi_k$ is given by the MAP estimate of $b(\xi_k)$, which is shown in (51). We also set a threshold $\beta$ and compare it with $m_{\xi_k}$ to decide whether user $k$ is active or inactive. The details of the proposed user activity detection algorithm are summarized in Algorithm. 3.

**Algorithm 3** User Activity Detection Algorithm II

1: Run Algorithm 1;
2: Approximate the message from $\xi_k$ to the product vertex to Gaussian by EP according to (51) and (53);
3: Calculate the mean $m_{\bar{s}_{kj}^n \to f_j^n}$ and variance $v_{\bar{s}_{kj}^n \to f_j^n}$ using (54) and (55);
4: Determine the messages from the product vertex to $\xi_k$ and $\tilde{s}_{kj}^n$ using (60)-(62);
5: Calculate the message $\mu_{\xi_k \to p_{\xi_k}}$ and estimate $\xi_k$ using (51);
6: Continue running Algorithm 1.

## V. Simulation Results

In this section, we evaluate the performance of the proposed algorithm via simulations. We consider an SCMA system with $J = 4$ resource elements that supports $K = 6$ users, yielding a normalized user-load of 1.5. The codebook is defined according to [13] with size $M = 4$ and an indicator matrix $\mathbf{F}$ of

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{63}$$

Each user transmits a sequence of data bits, which is coded using a rate-1/2 irregular low density parity check (LDPC) code having a length of 8640 bits and then mapped to a sequence of SCMA codewords. The LDPC code is designed for FTN signaling based on [50], which has a cycle distribution of [2]

$$\begin{aligned}f(X) = &236520X^8 + 5756880X^{10} + 170094120X^{12} \\ &+ 5121774960X^{14}.\end{aligned} \tag{64}$$

Standard BP decoding is used for channel decoder and the maximum number of iterations used for decoding is 50. We set the number of transmitted symbols corresponding to each user as $N = 4320$. The transmitted symbols pass through root raised cosine shaping filters with a roll-off factor of $\alpha = 0.5$ and FTN packing factor of $\tau = 0.8$.[3] The number of ISI taps imposed by FTN signaling is assumed to be $L = 10$. The channel obeys uncorrelated Rayleigh fading whose impulse response is generated according to Jake's model. The coarse estimate of channel coefficients is obtained by using 8 pilots symbols. Hence the pilot overhead is as low as $8/4320 \approx 2 \cdot 10^{-3}$, corresponding to semi-blind estimation. The maximum number of 'turbo' iterations is $N_{iter} = 10$. All results are averaged over 1000 independent Monte Carlo trials.

In Fig. 6, we compare the proposed Algorithm 1 both to the MPA-Gauss and to the MMSE-MPA methods in terms of

[2]The design of the LDPC code for FTN signaling depends on searching for codes with large girth of their Tanner graph. We have compared the BER performance of the specifically designed LDPC code and of the rate-2/5 CCSDS standard LDPC code having a length of 10240 for our FTN system associated with $\tau = 0.8$. As a result, 0.1dB performance gain can be observed for the specifically designed LDPC code, even though its code rate is higher than that of the CCSDS code.

[3]We assume the same shaping filter is employed for different resource elements at the transmitter side.

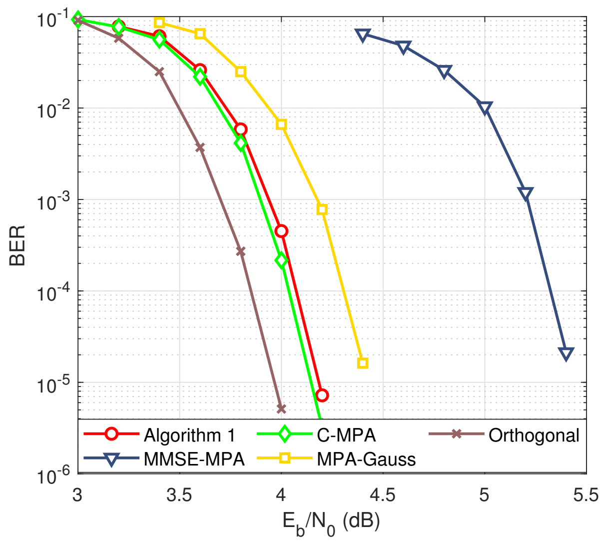



Fig. 6. BER performance of different algorithms for FTN-SCMA system.

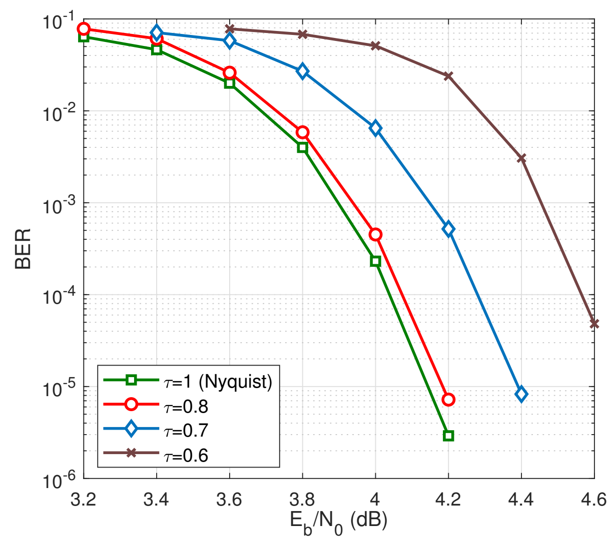



Fig. 7. BER performance of the proposed algorithm for different $\tau$ values.

their bit error rate (BER). As a reference, the performance of the conventional MPA (C-MPA) receiver relying on the MAP criterion is also illustrated. The 'MPA-Gauss' method refers to directly approximating the *a priori* distribution of $p(x^n_{kj})$ by the Gaussian distribution. The 'MMSE-MPA' method relies on a combination of the MMSE equalizer and the SCMA decoder. It is observed that the proposed algorithm outperforms all the other three algorithms and has almost the same performance as the conventional MPA receiver. MMSE-MPA method suffers from significant performance loss due to error propagation. Moreover, using an MMSE equalizer imposes a cubic complexity order, which is prohibitively high in practical applications. Compared to MPA-Gauss, the proposed algorithm achieves a performance gain, since EP exploits the extrinsic information fed to the channel decoder. Furthermore, the performance of the joint channel estimation and detection method based on an OMA system with using Nyquist signaling is also plotted. We see that the performance loss of the proposed algorithm is as small low 0.2 dB. Meanwhile, 50% more users are supported and 25% higher data rate is achieved. More explicitly, using the same resources, a total of 87.5% more information can be transmitted via our FTN-SCMA system advocated at a negligible performance loss.

Fig. 7 depicts the BER versus $E_b/N_0$ of the proposed algorithm parameterized by different packing factors $\tau$, where $\tau = 1$ corresponds to the Nyquist signaling case. It is seen that the iterative receiver proposed for FTN-SCMA system is capable of achieving a similar performance to classic Nyquist signaling for $\tau \geq 0.8$. Moreover, as the packing ratio becomes smaller, more severe interference is imposed and the performance gap between the FTN signaling and Nyquist signaling becomes higher. Since the actual number of ISI taps induced by FTN is infinite, the number of interfered symbols $L$ used in model (5) may not be sufficient for describing the ISI induced by FTN, hence causing a performance loss. In Fig. 8, we illustrate the BER curves for various values

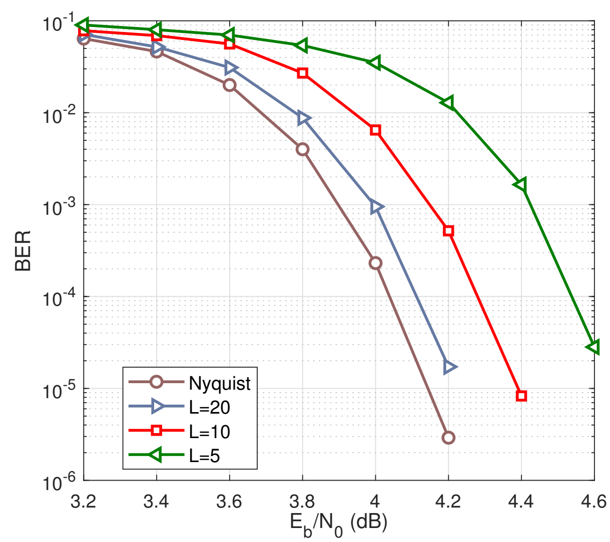



Fig. 8. BER performance of the proposed algorithm for different $L$ values.

of $L$, when the packing ratio $\tau$ is fixed as 0.7. Observed that when $L$ increases to 20, the performance gap between FTN and Nyquist signaling becomes negligible, which means that using a smaller packing ratio is still possible at the cost of more complex equalization. This implies that we can strike a compromise between the transmission rate and the receiver complexity. Nevertheless, there is a lower bound for the packing factor due to the Mazo limit [20].

To further show the advantage of FTN signaling, we evaluate the BER performance of the proposed message passing receiver for different packing factor $\tau$ and roll-off factor $\alpha$ pairs in Fig. 9, while the bandwidth efficiency is fixed. Compared to classic Nyquist signaling using $\alpha = 0.2$, it can be seen that FTN signaling associated with $\tau = 0.8$ and $\alpha = 0.5$ achieves better BER performance. When we further reduce the packing factor and increase the roll-off factor, the performance

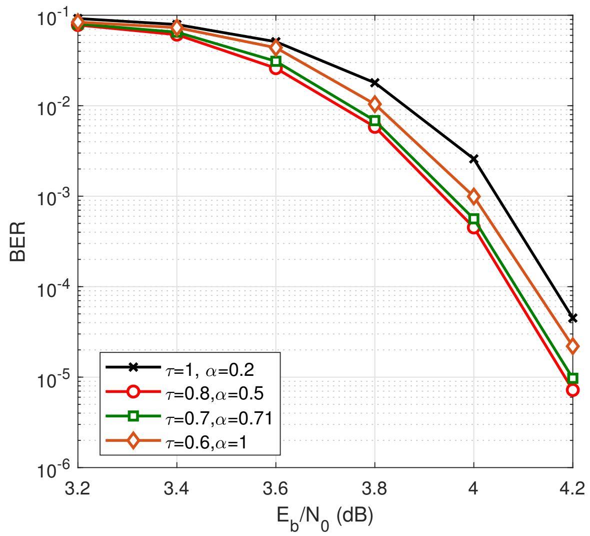



Fig. 9. BER performance of the proposed algorithm for different values of $\tau$ and $\alpha$ give fixed bandwidth efficiency.

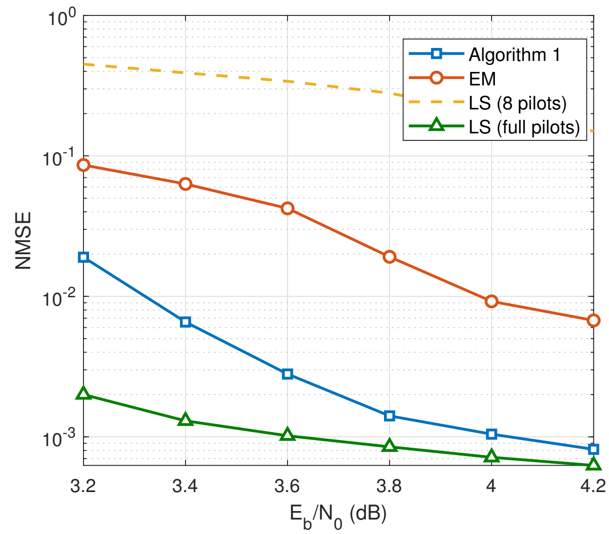



Fig. 10. NMSEs of different algorithms versus $E_b/N_0$.

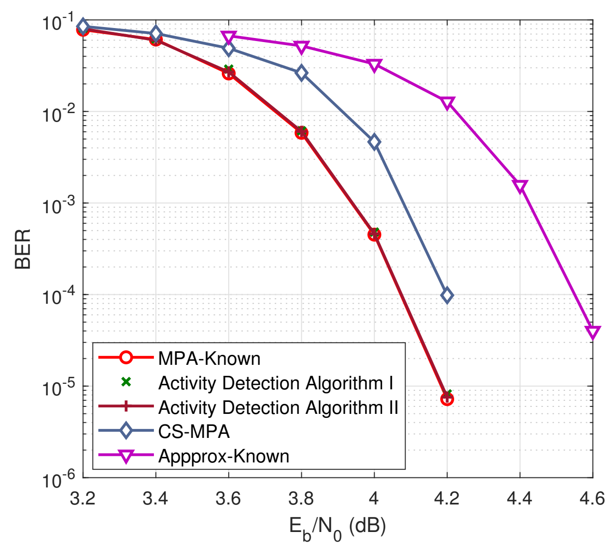



Fig. 11. BER performance of the proposed active user detection algorithms and existing methods.

degrades slightly. This is because the ISI imposed by FTN signaling cannot be fully eliminated. However, it can be seen that all three FTN signaling scenarios outperform their classic Nyquist counterparts at a given bandwidth efficiency. This is due to the fact that FTN signaling has the ability to exploit the excess bandwidth [21], which confirms the benefits of FTN signaling.

In Fig. 10, the normalized mean squared error (NMSE) of the estimated channel coefficients versus $E_b/N_0$ is illustrated. The NMSE is defined as

$$\mathrm{NMSE}_h = \frac{\sum_{k=1}^{K} \|\mathbf{h}_k - \hat{\mathbf{h}}_k\|^2}{\sum_{k=1}^{K} \|\mathbf{h}_k\|^2}, \tag{65}$$

where $\hat{\mathbf{h}}_k$ is the channel estimate obtained in Section III. The NMSE of the least square (LS) channel estimation method using 8 and full pilot symbols is also depicted for comparison. Observed from Fig. 10 that the proposed channel estimation algorithm is efficient, which can approach the performance of the LS algorithm based on all pilot symbols. Compared to the coarse estimate using a fewer 8 pilot symbols, the proposed algorithm significantly improves the channel estimation performance. Furthermore, the performance of an advanced joint channel estimation and decoding algorithm based on expectation maximization (EM) is also presented here. Since EM neglects all uncertainties of latent variables in the iterative process, it suffers from a performance loss.

Next, we evaluate the performance of the proposed active user detection algorithms in a grant-free system. In Fig. 11, the BER performance of the proposed algorithm versus $E_b/N_0$ is illustrated, where the probability that a user is active is $p_1 = 0.3$. For comparison, we also present the performance for the algorithm proposed in Section III with known active users (denoted by 'MPA-Known'), the algorithm that regards all users as active users (denoted by 'Approx-known') and the two-stage CS-MPA algorithm [51] that first uses compressive sensing for active user detection and then performs MPA multiuser detection. We can observe that Approx-known suffers from a significant performance degradation. Since the two-stage method only provides a hard decision concerning the users' activities to the equalizer of Fig. 2, it also experiences considerable performance loss. Compared to the optimal case that all users' activities are known, the proposed algorithms designed under our factor graph framework is capable of achieving a nearly optimal performance. Since the user activity detection Algorithm II has a lower complexity than Algorithm I, it is more attractive for practical grant-free systems.

Fig. 12 depicts the NMSE of channel estimates based on the joint channel estimation, decoding and active user detection algorithm of Fig. 2 parameterized by the occurrence probability $p_1$ of active users. We see that the performance degrades as $p_1$ becomes higher. This can be explained by the fact that a higher $p_1$ leads to having more active users in FTN-

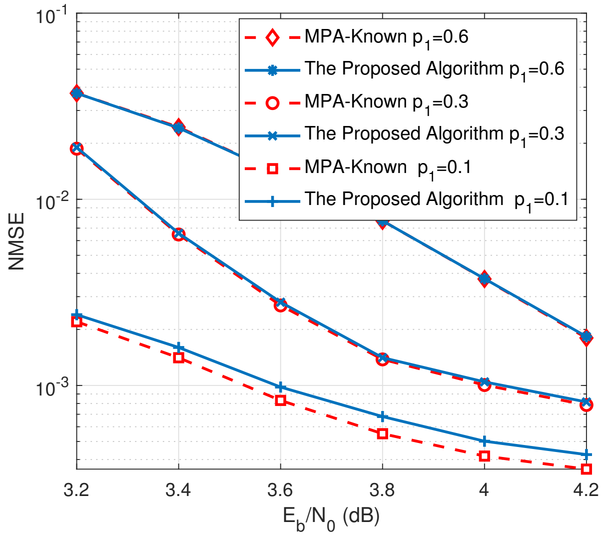



Fig. 12. NMSE of channel estimate with different active probability $p_1$.

SCMA systems and both the inter-user and the inter-symbol interferences become more severe. Finally, we also illustrate the performance of the MPA-Known algorithm for different $p_1$ as a performance bound. It can be observed that when $p_1$ is high, the proposed joint estimation algorithm is capable of approaching the bound. When $p_1$ decreases, although a small performance gap emerges, the proposed algorithm still performs well. This observation indicates that the proposed activity detection algorithm is more efficient when the number of active users increases.

## VI. Conclusions

In this paper, we considered an uplink SCMA system that utilized FTN signaling for increasing the spectral efficiency. Using an AR model, the correlated noise samples are approximated by an AR process. Then, based on the factorization of the joint *a posteriori* distribution, a factor graph based hybrid message passing receiver was proposed for estimating both the channel coefficients and the FTN data symbols. We extended the factor graph model and proposed a pair of novel user activity detection methods for a grant-free transmission scheme. Consequently, the proposed receiver carries out joint iterative channel estimation, decoding and active user detection in FTN-SCMA systems. Our simulation results showed that the FTN-SCMA system relying on the proposed receiver is capable of increasing the data rate by 80% of conventional orthogonal communications systems.

## Appendix A
## Derivations of (54) and (55)

For the product vertex $\times^n_{kj}$, the message $\mu_{\bar{s}^n_{kj}\to f^n_j}$ can be characterized by the distribution of $\bar{s}^n_{kj} = \xi_k\tilde{s}^n_{kj}$ associated with the random variables $\xi_k$ and $\tilde{s}^n_{kj}$ obeying the distributions $\mu_{\xi_k\to\times^n_{kj}}$ and $\mu_{\tilde{s}^n_{kj}\to\times^n_{kj}}$. Since $\mu_{\xi_k\to\times^n_{kj}}$ and $\mu_{\tilde{s}^n_{kj}\to\times^n_{kj}}$ are both Gaussian distributied, we can calculate the density of $\bar{s}^n_{kj}$ as

$$\begin{aligned} f(\bar{s}^n_{kj}) &= \int f(\tilde{s}^n_{kj}) f(\xi_k)\delta(\bar{s}^n_{kj} - \tilde{s}^n_{kj}\xi_k)\mathrm{d}\tilde{s}^n_{kj}\mathrm{d}\xi_k \\ &= \int \frac{1}{|\xi_k|} f\left(\frac{\bar{s}^n_{kj}}{\xi_k}\right) f(\xi_k)\mathrm{d}\xi_k \\ \propto \int \frac{1}{|\xi_k|} &\exp\left(-\frac{\left(\frac{\bar{s}^n_{kj}}{\xi_k} - m_{\tilde{s}^n_{kj}\to\times^n_{kj}}\right)^2}{v_{\tilde{s}^n_{kj}\to\times^n_{kj}}} - \frac{(\xi_k - m_{\xi_k\to\times^n_{kj}})^2}{v_{\xi_k\to\times^n_{kj}}}\right)\mathrm{d}\xi_k. \end{aligned} \tag{66}$$

However, the above integral does not have a closed-form analytical expression. Since our goal is to derive a Gaussian message, we in turn aim for determining the mean and variance of $\mu_{\bar{s}^n_{kj}\to f^n_j}$ based on the incoming messages.

It is widely recognized that, for two independent random variables $x$ and $y$, based on the Mellin Transform [52], the $n$th-order moment of $xy$ satisfies

$$\mathbb{E}[(xy)^n] = \mathbb{E}(x^n)\mathbb{E}(y^n). \tag{67}$$

Thus the first two order moments of $\mu_{\bar{s}^n_{kj}\to f^n_j}$ are given by

$$\mathbb{E}[\bar{s}^n_{kj}] = \mathbb{E}[\tilde{s}^n_{kj}]\mathbb{E}[\xi_k] = m_{\xi_k\to\times^n_{kj}} m_{\phi^n_{kj}\to\tilde{s}^n_{kj}}, \tag{68}$$

$$\begin{aligned} \mathbb{E}[(\bar{s}^n_{kj})^2] &= \mathbb{E}[(\tilde{s}^n_{kj})^2]\mathbb{E}[\xi_k^2] \\ &= (m^2_{\xi_k\to\times^n_{kj}} + v_{\xi_k\to\times^n_{kj}})(m^2_{\phi^n_{kj}\to\tilde{s}^n_{kj}} + v_{\phi^n_{kj}\to\tilde{s}^n_{kj}}), \end{aligned} \tag{69}$$

and the variance $v_{\bar{s}^n_{kj}\to f^n_j} = \mathbb{E}[(\bar{s}^n_{kj})^2] - \mathbb{E}^2[\bar{s}^n_{kj}]$, which are given by (54) and (55).

## References


[1] D. D. Falconer, F. Adachi, and B. Gudmundson, "Time division multiple access methods for wireless personal communications," *IEEE Commun. Mag.*, vol. 33, no. 1, pp. 50–57, 1995.

[2] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. Prentice Hall PTR New Jersey, 1996, vol. 2.

[3] T. Ojanpera and R. Prasad, *Wideband CDMA For Third Generation Mobile Communications: Universal Personal Communications*. Artech House, Inc., 1998.

[4] M. Morelli, C.-C. J. Kuo, and M.-O. Pun, "Synchronization techniques for orthogonal frequency division multiple access (OFDMA): A tutorial review," *Proc. IEEE*, vol. 95, no. 7, pp. 1394–1427, 2007.

[5] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Ten Brink, I. Gaspar, N. Michailow, A. Festag *et al.*, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 97–105, 2014.

[6] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, 2016.

[7] N. Bonello, R. Zhang, S. Chen, and L. Hanzo, "Channel code division multiple access and its multilevel structured LDPC based instantiation," in *Proc. IEEE 68th Veh. Technol. Conf.*, Sep. 2008, pp. 1–5.

[8] L. Li, Z. Ma, L. Wang, P. Fan, and L. Hanzo, "Cutoff rate of sparse code multiple access in downlink broadcast channels," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3328–3342, Aug 2017.

[9] Y. Liu, L. Yang, and L. Hanzo, "Spatial modulation aided sparse code-division multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1474–1487, March 2018.

[10] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, 2015.

[11] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Proc. IEEE Veh. Technol. Conf.* IEEE, 2014, pp. 1–5.

[12] L. Li, Z. Ma, P. Z. Fan, and L. Hanzo, "High-dimensional codebook design for the SCMA down link," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 10 118–10 122, Oct 2018.

[13] K. Xiao, B. Xia, Z. Chen, B. Xiao, D. Chen, and S. Ma, “On capacity-based codebook design and advanced decoding for sparse code multiple access systems,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3834–3849, June 2018.

[14] C. Zhang, Y. Luo, and Y. Chen, “A low-complexity SCMA detector based on discretization,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2333–2345, 2018.

[15] L. Yang, X. Ma, and Y. Siu, “Low complexity MPA detector based on sphere decoding for SCMA,” *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1855–1858, 2017.

[16] H. Mu, Z. Ma, M. Alhaji, P. Fan, and D. Chen, “A fixed low complexity message pass algorithm detector for up-link SCMA system,” *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 585–588, 2015.

[17] 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, 2016.

[18] W. Yuan, N. Wu, Q. Guo, Y. Li, C. Xing, and J. Kuang, “Iterative receivers for downlink MIMO-SCMA: Message passing and distributed cooperative detection,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3444–3458, 2018.

[19] 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.

[20] J. E. Mazo, “Faster-than-Nyquist signaling,” *Bell System Technical Journal*, vol. 54, no. 8, pp. 1451–1462, 1975.

[21] J. B. Anderson, F. Rusek, and V. Öwall, “Faster-than-Nyquist signaling,” *Proc. IEEE*, vol. 101, no. 8, pp. 1817–1830, 2013.

[22] A. Prlja and J. B. Anderson, “Reduced-complexity receivers for strongly narrowband intersymbol interference introduced by faster-than-Nyquist signaling,” *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2591–2601, 2012.

[23] S. Sugiura, “Frequency-domain equalization of faster-than-Nyquist signaling,” *IEEE Wireless Commun. Lett.*, vol. 2, no. 5, pp. 555–558, 2013.

[24] S. Sugiura and L. Hanzo, “Frequency-domain-equalization-aided iterative detection of faster-than-nyquist signaling,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 2122–2128, May 2015.

[25] P. Şen, T. Aktaş, and A. Ö. Yilmaz, “A low-complexity graph-based LMMSE receiver designed for colored noise induced by FTN-signaling,” in *Proc. 2014 IEEE Wireless Commun. Net. Conf.* IEEE, 2014, pp. 642–647.

[26] N. Wu, W. Yuan, H. Wang, Q. Shi, and J. Kuang, “Frequency-domain iterative message passing receiver for faster-than-Nyquist signaling in doubly selective channels,” *IEEE Wireless Commun. Lett.*, vol. 5, no. 6, pp. 584–587, 2016.

[27] W. Yuan, N. Wu, H. Wang, and J. Kuang, “Variational inference-based frequency-domain equalization for faster-than-Nyquist signaling in doubly selective channels,” *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1270–1274, 2016.

[28] Y. Li, “Simplified channel estimation for OFDM systems with multiple transmit antennas,” *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 67–75, 2002.

[29] A. Kocian and B. H. Fleury, “EM-based joint data detection and channel estimation of DS-CDMA signals,” *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1709–1720, 2003.

[30] M. Abuthinien, S. Chen, and L. Hanzo, “Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems,” *IEEE Signal Process. Lett.*, vol. 15, pp. 202–205, 2008.

[31] T. Cui and C. Tellambura, “Joint data detection and channel estimation for OFDM systems,” *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 670–679, 2006.

[32] Q. Guo, L. Ping, and D. Huang, “A low-complexity iterative channel estimation and detection technique for doubly selective channels,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, 2009.

[33] 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, 2016.

[34] N. Wu, W. Yuan, Q. Guo, and J. Kuang, “A hybrid BP-EP-VMP approach to joint channel estimation and decoding for FTN signaling over frequency selective fading channels,” *IEEE Access*, vol. 5, pp. 6849–6858, 2017.

[35] Q. Shi, N. Wu, X. Ma, and H. Wang, “Frequency-domain joint channel estimation and decoding for faster-than-Nyquist signaling,” *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 781–795, 2018.

[36] T. P. Minka, “Expectation propagation for approximate Bayesian inference,” in *Proc. Conf. Uncertainty Artif. Intell.* Morgan Kaufmann Publishers Inc., 2001, pp. 362–369.

[37] M. J. Wainwright, M. I. Jordan *et al.*, “Graphical models, exponential families, and variational inference,” *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[38] 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*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007, p. 18.

[39] J. Zhang, L. Lu, Y. Sun, Y. Chen, J. Liang, J. Liu, H. Yang, S. Xing, Y. Wu, J. Ma *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, 2017.

[40] B. Shim and B. Song, “Multiuser detection via compressive sensing,” *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 972–974, 2012.

[41] 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.* IEEE, 2015, pp. 1–5.

[42] 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, 2017.

[43] 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.*, pp. 1–1, 2018.

[44] F. Wei, W. Chen, Y. Wu, J. Ma, and T. A. Tsiftsis, “Message-passing receiver design for joint channel estimation and data decoding in uplink grant-free SCMA systems,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 167–181, 2019.

[45] J. D. Gibson, B. Koo, and S. D. Gray, “Filtering of colored noise for speech enhancement and coding,” *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1732–1742, 1991.

[46] S. L. Marple and S. L. Marple, *Digital spectral analysis: with applications*. Prentice-Hall Englewood Cliffs, NJ, 1987, vol. 5.

[47] M. El-Hajjar and L. Hanzo, “EXIT charts for system design and analysis,” *IEEE Commun. Survey. Tutor.*, vol. 16, no. 1, pp. 127–153, 2014.

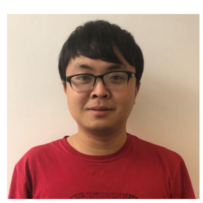
[48] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The Annals Math. Stat.*, vol. 22, no. 1, pp. 79–86, 1951.

[49] J. Winn and C. M. Bishop, “Variational message passing,” *J. Machine Learning Res.*, vol. 6, no. Apr, pp. 661–694, 2005.

[50] I. E. Bocharova, B. D. Kudryashov, and R. Johannesson, “Searching for binary and nonbinary block and convolutional LDPC codes,” *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 163–183, 2015.

[51] 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, 2016.

[52] A. M. Mathai, *A handbook of generalized special functions for statistical and physical sciences*. Oxford University Press, USA, 1993.



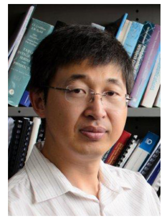
**Weijie Yuan** (S’15) received the Ph.D. degree from the Beijing Institute of Technology, China, and the Ph.D. degree from the University of Technology Sydney, Australia, in 2019. He is currently a Research Associate with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia. He has served as a Research Assistant with the University of Sydney, a Visiting Associate Fellow with the University of Wollongong, and a Visiting Fellow with the University of Southampton, from 2017 to 2019. In 2016, he was a Visiting Ph.D. Student with the Institute of Telecommunications, Vienna University of Technology, Austria. His research interest includes statistical inference on graphical models. He has served as a TPC member for several conferences.

**Nan Wu** (M'11) received the B.S., M.S., and Ph.D. degrees from the Beijing Institute of Technology (BIT), Beijing, China, in 2003, 2005, and 2011, respectively. From 2008 to 2009, he was a Visiting Ph.D. Student with the Department of Electrical Engineering, Pennsylvania State University, USA. He is currently a Professor with the School of Information and Electronics, BIT. His research interests include signal processing in wireless communication networks. He was a recipient of the National Excellent Doctoral Dissertation Award by MOE of China in 2013. He serves as an Editorial Board Member of the IEEE Access, the International Journal of Electronics and Communications, the KSII Transactions on Internet and Information Systems, and the IEICE Transactions on Communications.

**J. Andrew Zhang** (M'04-SM'11) received the B.Sc. degree from Xi'an JiaoTong University, China, in 1996, the M.Sc. degree from Nanjing University of Posts and Telecommunications, China, in 1999, and the Ph.D. degree from the Australian National University, in 2004.

Currently, Dr. Zhang is an Associate Professor in the School of Electrical and Data Engineering, University of Technology Sydney, Australia. He was a researcher with Data61, CSIRO, Australia from 2010 to 2016, the Networked Systems, NICTA, Australia from 2004 to 2010, and ZTE Corp., Nanjing, China from 1999 to 2001. Dr. Zhang's research interests are in the area of signal processing for wireless communications and sensing, and autonomous vehicular networks. He has published more than 150 papers in leading international Journals and conference proceedings, and has won 5 best paper awards for his work. He is a recipient of CSIRO Chairman's Medal and the Australian Engineering Innovation Award in 2012 for exceptional research achievements in multi-gigabit wireless communications.

**Xiaojing Huang** (M'99-SM'11) received the B.Eng., M.Eng., and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1983, 1986, and 1989, respectively. He was a Principal Research Engineer with the Motorola Australian Research Center, Botany, NSW, Australia, from 1998 to 2003, and an Associate professor with the University of Wollongong, Wollongong, NSW, Australia, from 2004 to 2008. He had been a Principal Research Scientist with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Sydney, NSW, Australia, and the Project Leader of the CSIRO Microwave and mm-Wave Backhaul projects since 2009. He is currently a Professor of Information and Communications Technology with the School of Electrical and Data Engineering and the Program Leader for Mobile Sensing and Communications with the Global Big Data Technologies Center, University of Technology Sydney (UTS), Sydney, NSW, Australia. With over 30 years of combined industrial, academic, and scientific research experience, he has authored over 300 book chapters, refereed journal and conference papers, major commercial research reports, and filed 31 patents. Prof. Huang was a recipient of the CSIRO Chairman's Medal and the Australian Engineering Innovation Award in 2012 for exceptional research achievements in multigigabit wireless communications.



**Yonghui Li** (M04-SM09-F'19) received his PhD degree in November 2002 from Beijing University of Aeronautics and Astronautics. From 1999 2003, he was affiliated with Linkair Communication Inc, where he held a position of project manager with responsibility for the design of physical layer solutions for the LAS-CDMA system. Since 2003, he has been with the Centre of Excellence in Telecommunications, the University of Sydney, Australia. He is now a Professor in School of Electrical and Information Engineering, University of Sydney. He is the recipient of the Australian Queen Elizabeth II Fellowship in 2008 and the Australian Future Fellowship in 2012.

His current research interests are in the area of wireless communications, with a particular focus on MIMO, millimeter wave communications, machine to machine communications, coding techniques and cooperative communications. He holds a number of patents granted and pending in these fields. He is now an editor for IEEE transactions on communications and IEEE transactions on vehicular technology. He also served as a guest editor for several special issues of IEEE journals, such as IEEE JSAC special issue on Millimeter Wave Communications. He received the best paper awards from IEEE International Conference on Communications (ICC) 2014, IEEE PIMRC 2017 and IEEE Wireless Days Conferences (WD) 2014.

**Lajos Hanzo** (M91-SM92-F04) (http://www. mobile.ecs.soton.ac.uk) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 111 PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1756 research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Councils Advanced Fellow Grant and the Royal Societys Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE ComSoc and VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk Lajos has 34 000+ citations.