

Iterative Joint Channel Estimation, User Activity Tracking, and Data Detection for FTN-NOMA Systems Supporting Random Access

Weijie Yuan, Nan Wu, *Member, IEEE*, Qinghua Guo, *Senior Member, IEEE*,
Derrick Wing Kwan Ng, *Senior Member, IEEE*, Jinhong Yuan, *Fellow, IEEE*, and Lajos Hanzo, *Fellow, IEEE*,

Abstract—Given the requirements of increased data rate and massive connectivity in the Internet-of-things (IoT) applications of the fifth-generation communication systems (5G), non-orthogonal multiple access (NOMA) was shown to be capable of supporting more users than OMA. As a further potential enhancement, the faster-than-Nyquist (FTN) signaling is also capable of increasing the symbol rate. Since NOMA and FTN signaling impose non-orthogonalities from different perspectives, it is possible to achieve further increased spectral efficiency by exploiting both. Hence we investigate the FTN-NOMA uplink in the context of random access. Although random access schemes reduce the signaling overheads as well as latency, they require the base station to identify active users before performing data detection. As both inter-symbol and inter-user interferences exist, performing optimal detection requires a prohibitively high complexity. Moreover, in typical mobile communication environments, the channel envelope of users fluctuates violently, which imposes challenges on the receiver design. To tackle this problem, we propose a joint user activity tracking and data detection algorithm based on the factor graph framework, which relies on a sophisticated amalgam of expectation maximization (EM) and hybrid message passing algorithms. The complexity of the algorithm advocated only increases linearly with the number of active users. Our simulation results show that the proposed algorithm is effective in tracking user activity and detecting data symbols in dynamic random access systems.

Index Terms—Machine-type communications, non-orthogonal multiple access, faster-than-Nyquist signaling, dynamic networks, random access, hybrid message passing, birth and survive probability

I. INTRODUCTION

Wireless communications has played an increasingly important role in modern society. The rapid development of communication technologies has fueled the roll-out of the Internet-of-things (IoT) in 5G wireless systems [1]–[3]. In particular, low-cost mobile devices with considerable computational capabilities facilitate the massive deployment of machine-type communications (MTC) or machine-to-machine (M2M) communications. Unlike conventional human-type communications, MTC involves a massive number of devices. Thus, massive connectivity should be supported in next-generation mobile communications. In conventional orthogonal multiple access (OMA) [4], [5], each user is assigned with an orthogonal radio resource element for avoiding inter-user interference. However, due to the limited radio resource elements and an increasing number of devices, traditional OMA may not be able to support massive access in MTC.

Recent investigations on non-orthogonal multiple access (NOMA) have shown that it is capable of supporting massive connectivity [6]–[8]. By introducing controllable interference, multiple users can share the same orthogonal radio resource elements, which allows a communication system to support more users relying on the same amounts of orthogonal radio resource elements as OMA. Therefore, NOMA has been recognized as a promising 5G candidate. The existing NOMA techniques can be classified into power-domain based [9], [10] and code-domain based methods [11], [12]. In power-domain based NOMA, users are multiplexed over several resources by employing superposition coding with proper power [13]. In code-domain based NOMA, the signals of users are spread over the radio resource elements with predefined codebooks. In general, the code-domain based scheme achieves better performance by exploiting the shaping gain [14]. The code-domain NOMA schemes include the low density signature (LDS) based scheme [15], sparse code multiple access (SCMA) [16], pattern division multiple access (PDMA) [17], and multi-user shared access (MUSA) [18]. To elaborate, LDS adopts sparse sequences to spread the data symbols of users over all radio resource elements. Besides, SCMA can be regarded as

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), the Australia Research Council Discovery Project (DP190101363) and Linkage Projects (LP 160100708 and LP170101196). D. W. K. Ng is supported by funding from the UNSW Digital Grid Futures Institute, UNSW, Sydney, under a cross-disciplinary fund scheme and by the Australian Research Council's Discovery Project (DP190101363). L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/N004558/1, 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, China. He is now with the School of Electrical Engineering and Telecommunications, University of New South Wales, NSW 2052, Australia (e-mail: 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).

Q. Guo is with the School of Electrical, Computer and Telecommunications Engineering, The University of Wollongong, NSW 2522, Australia. (e-mail:qguo@uow.edu.au).

D. W. K. Ng and J. Yuan are with the School of Electrical Engineering and Telecommunications, University of New South Wales, NSW 2052, Australia (e-mail:{w.k.ng, j.yuan}@unsw.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)

an extension of the LDS based scheme, in which the bit-to-symbol mapping and spreading operations are amalgamated. As for MUSA, the transmitted symbols of a user are multiplied with the same spreading sequence. On the other hand, PDMA is similar to the LDS based schemes where the low density sequences of LDS are replaced by non-orthogonal patterns. In this paper, we apply a simple LDS-based NOMA scheme since we mainly aim for designing iterative receivers for FTN-NOMA systems. On the other hand, although NOMA is capable of accommodating massive number of devices and of increasing the spectral efficiency, the demand for further increasing the spectral efficiency is still growing. In accordance with IMT-2020, the data rate of 5G is required to be ten times higher than that in 4G. Faster-than-Nyquist (FTN) signaling, proposed by Mazo in 1970s [19], has attracted substantial interests, since it can transmit at a symbol-rate beyond the Nyquist rate. It was shown that the symbols' minimum Euclidean distance is preserved, provided that the FTN rate is not substantially higher than the Nyquist rate [20]. However, FTN signaling intentionally introduces inter-symbol interference (ISI).

Then the judicious question arrives, whether amalgamating the NOMA and FTN principles allows us to improve the bandwidth-efficiency of both in isolation? More explicitly, having increased the Nyquist-rate by FTN signaling, can we glean further gain by additionally multiplexing the FTN users employing the popular NOMA principles? Hence naturally, both inter-user interference (IUI) as well as inter-symbol interference (ISI) would co-exist in the FTN-NOMA systems. **It is widely recognized that the optimal receiver's complexity grows exponentially both with the number of ISI taps and with the number of users.** Hence, several authors have developed low complexity receivers for FTN signaling [21]–[24] and NOMA [25]–[30]. To elaborate, in [21], an M-algorithm based Bahl-Cocke-Jelinek-Raviv (BCJR) detector was proposed by Prlja and Anderson for eliminating the ISI imposed by FTN signaling. Based on the Fast Fourier Transform (FFT), frequency domain (FD) equalizers were proposed in [22] and [23] for FTN signaling. However, the insertion of cyclic prefixes reduces the effective throughput. As a further result, in [24], the state space model of the received signal was represented by a Forney-style factor graph, which conveniently lent it self to the employment of the classic Gaussian message passing algorithm to detect the FTN symbols. Upon invoking multi-user detection (MUD) for NOMA, the conventional minimum mean squared error (MMSE) detector suffers from an excessive complexity due to the inversion of high-dimensional matrices. To circumvent this problem, successive interference cancellation (SIC) based MUDs were developed for attaining near-optimal performance at low complexity [25]. However, in SIC, the users are detected one by one, hence resulting in certain delay in massive-connectivity scenarios. Moreover, a frequent assumption is that the information from the high-power users has been perfectly removed from the received signals, which is impractical. Motivated by the power of iterative 'turbo' receivers, several iterative methods based on the classic message passing algorithms (MPA) were proposed for MUD. To elaborate, in [26], a partial marginalization

based MPA detector was proposed for the NOMA uplink. The authors of [27] developed a Monte Carlo Markov Chain based MUD for large-scale networks. A factor graph based linear MMSE (LMMSE) detector exhibiting rapid convergence was proposed in by Liu et al. [28] for multiple-input multiple-output (MIMO)-NOMA systems. Then, by exploiting the 'convexified' variational free energy, the authors of [29] proposed a guaranteed-convergence-MPA receiver for MIMO-NOMA systems. In [30], an energy minimization based detector relying on 'potential-cliques' factorization was proposed and the algorithm's convergence behaviors was analyzed. Despite this research-momentum, there is no literature considering the receiver design problem of the proposed systems.

As a further challenge, the large number of connected users in MTC makes the conventional 'handshaking'-based access control impractical owing to its high communication overhead and due to the resultant signal latency [31]. As a remedy, random access control schemes dispensing with handshaking have received intense attention in NOMA scenarios [32]. As a benefit of random access schemes, users can instantly transmit their signals to a base station (BS) in any available time slots without waiting for access grant by the BS. **In order to correctly decode the information corresponding to different users, the BS has to uniquely and unambiguously identify the active users before data detection can take place.** Moreover, it was shown by Hong et al. that only a fraction of users are active concurrently in wireless networks even during the rush-hour [33]. Therefore, the sparsity of the user-activity may be beneficially exploited to reduce the receiver's complexity. Motivated by this, Wang et al. [34] employed compressive sensing (CS) for identifying the active users and then designed a powerful MPA based MUD. To mitigate the performance degradation imposed by suboptimal two-stage user-activity and MUD, an MUD based on a sophisticated combination of approximate message passing and the least square (LS) algorithm was developed by Wei et al. in [35].

Despite the above rigorous attempts to improve the receiver design of NOMA relying on random user access, all contributions have stipulated the idealized simplifying assumption that the channel is perfectly known. Naturally, in practical applications, the channel fluctuates quite violently and it has to be estimated before detection. Channel estimation is an extensively investigated problem in wireless communications, which typically relies a sequence of pilot symbols [36]. Furthermore, by taking the equalization problem into account, several joint channel estimation and data detection schemes were proposed in the literature [37]–[39]. The benefits are two-fold: only a few pilot symbols are needed, which reduces the signaling overhead; the performance may approach the 'perfect-knowledge' performance. By invoking the heuristic iterative receiver, an MPA and factor graph based joint channel estimation and decoding algorithm was proposed in [40]. Its extension to FTN signaling was then considered in [41], where the generalized approximate MPA was employed. For NOMA relying on random access, the authors of [42] and [43] assumed the user activity to be static and constructed factor graph models for simultaneously solving the channel estimation as well as data and user activity detection problem.

In practice, user activity in networks fluctuates over time. Some active users may become inactive in the next few time slots, while several sleeping users may become active. Hence, the BS has to determine the user activity in real time, tracking the user activity. The channel envelope is also fluctuating, especially in high mobility scenarios. Therefore, joint user activity tracking, channel estimation and data detection has to be designed for FTN-NOMA systems relying on random access. Instead of relying on precision parameter of [42] and [43] to capture the user activity, we introduce a binary variable to represent the user activity. By employing the autoregressive moving average (ARMA) model of [44] for approximating the correlated noise samples imposed by FTN signaling, we arrive at the factorization problem of the joint *a posteriori* distribution. Then a factor graph is conceived for representing the factorization and a message passing algorithm (MPA) can be used for user-activity detection. Since the binary variable representing user activity makes the MPA updating on the multiplier node unavailable, we construct the log density based on the beliefs obtained and then derive an expectation maximization (EM) [45] algorithm for estimating the user activity. Upon approximating the data symbols by Gaussian variables, only the corresponding means and (co)variances are updated, which results in closed-form solutions of the messages and beliefs of the variables. Overall, the complexity of the proposed EM-MPA algorithm only scales linearly with the number of users. Our simulation results will demonstrate the efficiency of using FTN signaling in NOMA systems and show that the data rate is increased by 87.5% over that of its orthogonal counterpart. Moreover, the proposed algorithms are efficient in tracking the user activity and in detection in dynamic environments.

In a nutshell, our contributions in this paper are as follows.

- In contrast to [42], [43] assuming static user activity in random access, we study dynamically fluctuating user activity and channel envelopes. We intrinsically amalgamate FTN signaling with random access NOMA systems and demonstrate that their throughput gains may be combined.
- Based on an ARMA model of the correlated noise samples, we can fully factorize the joint *a posteriori* distribution. The factor graph representation is then derived and a bespoke message passing algorithm is constructed for determining the beliefs of variables. To avoid leaving an exponentially escalating complexity, expectation propagation method is employed to obtain the Gaussian approximations of the data symbols.
- An EM-MPA algorithm is proposed for iteratively estimating both the user activity as well as the channel coefficients, and for detecting the data symbols. Since all messages defined over the factor graph and the solutions of the M-step of the EM are obtained in closed forms, the complexity of the proposed algorithm only increases linearly with the number of users.

The remainder of this paper is organized as follows. In Section II, the model of the FTN-NOMA system relying on random user access is introduced. Section III formulates our statistical model and derives its factor graph representation.

The EM-MPA receiver proposed for low complexity joint user-activity tracking and data detection is presented in Section IV. The results of our Monte Carlo simulations are provided in Section V. Finally, our conclusions are drawn in Section VI.

Notations: We use a boldface letter to denote a vector. The superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^{-1}$ denote the transpose, conjugate, and the inverse operations, respectively; $\mathcal{G}(\mathbf{m}_x, \mathbf{V}_x)$ denotes the Gaussian distribution of variable \mathbf{x} having a mean vector of \mathbf{m}_x and covariance matrix of \mathbf{V}_x ; \mathbb{E} represents the expectation operator; $J_0(\cdot)$ denotes the zeroth-order Bessel function of the first kind; $\text{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with the diagonal elements \mathbf{a} ; \odot is the component-wise product; $|\cdot|$ represents the modulus of a complex number or the cardinality of a set; $\|\cdot\|_2$ denotes the ℓ_2 norm; $k!$ is the factorial of k ; \mathbf{I}_n denotes an identity matrix of dimension n ; $\mathbf{0}_n$ represents a n dimensional column vector with entry 0; the superscript n in $(\cdot)^{[n]}$ denotes the time slot index n while the subscript in $(\cdot)_n$ denotes the user and/or resource element index; \propto represents both sides of the equation are multiplicatively connected to a constant; $\mathbf{x}\setminus x$ denotes all variables in \mathbf{x} except x ; the big O notation \mathcal{O} asymptotically describes the order of computational complexity.

II. SYSTEM MODEL

Throughout this paper, we consider a practical uplink scenario supporting a large number of users, which is more than the available orthogonal radio resource elements. LDS-based NOMA relying on random access is employed, which is eminently suitable for future IoT applications supporting massive access. Additionally, it is intrinsically amalgamated with FTN signaling for further improving the data rate. Since the user states and channel envelopes vary with time, the BS has to keep tracking both the channel coefficient as well as the user activities based on the received signals. For ease of exposition, we focus on baseband signals.

A. Faster than Nyquist Signaling

Considering a sequence of data symbols $\{x^{[n]}\}$ from a transmitter, the ISI-free Nyquist signaling is formulated as

$$s(t) = \sum_n x^{[n]} q(t - nT_0), \quad (1)$$

where $q(t)$ is a bandlimited signaling pulse and the symbol duration is T_0 . Using the same signaling pulse, the FTN signaling is characterized by [20]

$$s(t) = \sum_n x^{[n]} q(t - n\tau T_0), \quad (2)$$

where $0 < \tau \leq 1$ defines the ‘packing factor’. Naturally, setting $\tau < 1$ transmits more data symbols within the same bandwidth and time period at the cost of imposing intentional ISI. Then the modulated signal is transmitted through a time-variant fading channel $h(t)$ and it is impaired by additional Gaussian noise $w(t)$.

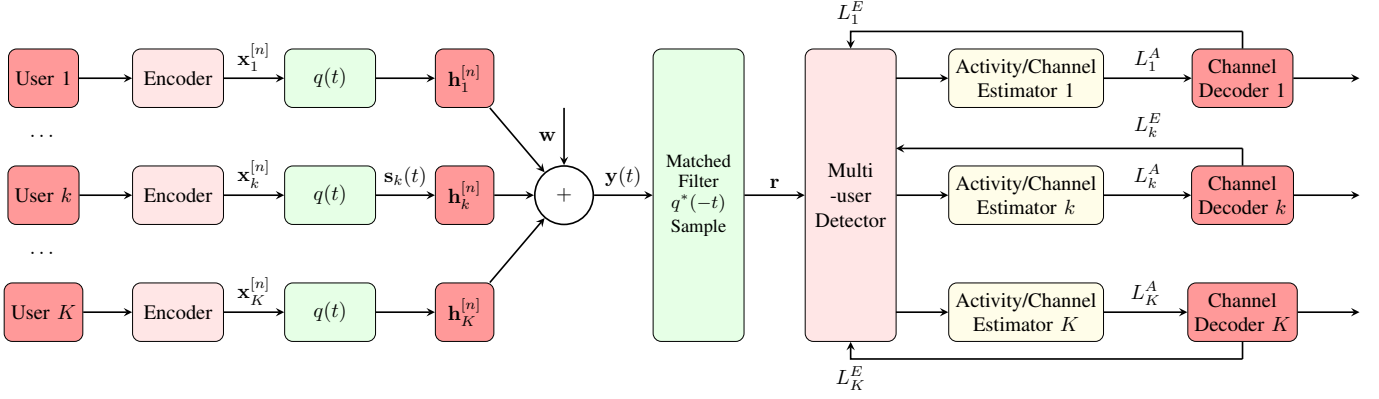


Fig. 1. Block diagram model of an FTN-NOMA system. L_k^A and L_k^E denote the extrinsic log likelihood ratio feed to and output from the channel decoder of user k , respectively.

The received signal $y(t)$ is then detected by a matched filter $q^*(-t)$ and sampled at the sampling rate of $\frac{1}{\tau T_0}$, yielding

$$r^{[n]} = h^{[n]} \sum_{i=-L}^L g_i x^{[n-i]} + \omega^{[n]}, \quad (3)$$

where L denotes the number of one-sided ISI taps induced by FTN signaling, $h^{[n]}$ is the channel coefficient at time instant n , while the variables g_i and $\omega^{[n]}$ denote the FTN-induced ISI taps and noise samples, respectively, obeying

$$g_i = \int q(t - i\tau T_0) q^*(t) dt, \quad (4)$$

$$\omega^{[n]} = \int w(t) q^*(t - n\tau T_0) dt. \quad (5)$$

However, temporally correlated noise emerges, since the signaling pulse is no longer orthogonal to the Nyquist symbol period T_0 , yielding $\mathbb{E}(\omega^{[n]} \omega^{[n+i]}) = N_0 g_i$, where N_0 is the power spectral density of $w(t)$.

B. Non-orthogonal Multiple Access

Again, the NOMA uplink system is considered where K simultaneous users transmit their information to the BS relying on J orthogonal radio resource elements, where $J < K$ and $\rho = \frac{K}{J}$ is the normalized user-load. The coded bit stream corresponding to the k th user is first mapped to a sequence of data symbols and then spread over J resource-slots using a low-density signature (LDS) sequence $\mathbf{x}_k^{[n]} = [x_{k,1}^{[n]}, \dots, x_{k,J}^{[n]}]^T$. For any two different users, the positions of non-zero entries in the LDS sequences $\mathbf{x}_k^{[n]}$ are different for avoiding potential packet collision.

Then the signal received within the j th resource element at the BS at time instant n can be expressed as

$$y_j^{[n]} = \sum_{k=1}^K h_{k,j}^{[n]} x_{k,j}^{[n]} + w_j^{[n]}, \quad \forall j \in \{1, \dots, J\}, \quad (6)$$

where $h_{k,j}^{[n]}$ denotes the channel gain of the j th resource element between the k th user and the BS, while $w_j^{[n]}$ denotes the noise samples.

C. FTN-NOMA System with Random Access

To employ FTN signaling in a NOMA system, the transmitted sequences, $\mathbf{x}_k^{[n]}$ of different users pass through the shaping filter $q(t)$, having a symbol period¹ of τT_0 , yielding

$$\mathbf{s}_k(t) = \sum_n \mathbf{x}_k^{[n]} q(t - n\tau T), \quad (7)$$

where $\mathbf{s}_k(t) = [s_{k,1}(t), \dots, s_{k,J}(t)]^T$. As shown in Fig. 1, the transmitted signals of all users are multiplexed over J radio resource elements and passed through the time-variant fading channel $\mathbf{h}_k(t) = [h_{k,1}(t), \dots, h_{k,J}(t)]^T$. Assuming perfect synchronization between the BS and the users, the signal received at the BS obeys:

$$\mathbf{y}(t) = \sum_{k=1}^K \text{diag}\{\mathbf{h}_k(t)\} \cdot \mathbf{s}_k(t) + \mathbf{w}(t), \quad (8)$$

where $\mathbf{y}(t)$ and $\mathbf{w}(t)$ are both J -dimensional vectors with the j th entries being the received signal and noise at the j th resource element, respectively. Then $\mathbf{y}(t)$ is processed by a matched filter $q^*(-t)$ which yields the corresponding discrete time model:

$$r_j^{[n]} = \sum_{k=1}^K h_{k,j}^{[n]} \sum_{i=-L}^L g_i x_{k,j}^{[n-i]} + \omega_j^{[n]}. \quad (9)$$

Moreover, given that random access is used for reducing both the signaling overhead and the latency, we use \mathcal{K}^+ to denote the set of active users. Then $r_j^{[n]}$ can be expressed as

$$r_j^{[n]} = \sum_{k \in \mathcal{K}^+} h_{k,j}^{[n]} \sum_{i=-L}^L g_i x_{k,j}^{[n-i]} + \omega_j^{[n]}. \quad (10)$$

To distinguish the active/inactive users, we introduce a binary variable $\lambda_k^{[n]} = \{0, 1\}$ to denote the activity state of user k at time instant n , where $\lambda_k^{[n]} = 1$ represents an active user and

¹For simplicity, we assume that the same shaping filter $q(t)$ and packing factor τ are used for all the users.

vice versa². Consequently, we arrive at a general form of $r_j^{[n]}$, given as

$$r_j^{[n]} = \sum_{k=1}^K \lambda_k^{[n]} h_{k,j}^{[n]} \sum_{i=-L}^L g_i x_{k,j}^{[n-i]} + \omega_j^{[n]}. \quad (11)$$

Note that the signal received at the BS is the summation of all users' transmitted signals. If all users are inactive, the receiver's noise does not exceed the threshold used for identifying if the received signal contains desired information. In the following sections, we will design low-complexity receivers for active user identification and data symbol detection based on a graphical model, which can fully exploit the sparsities of both the spreading sequence and of the user activity.

III. STATISTICAL FORMULATION AND FACTOR GRAPH REPRESENTATION

We aim for inferring the channel taps and decoding all users' information bits from the received signal samples. From a statistical inference perspective, this corresponds to determining the *a posteriori* distributions of the corresponding variables.

A. Bayesian Estimator

Again, in FTN signaling, the noise samples of different time slots are correlated, which imposes challenges on the receiver design. According to [46], the correlation of additive noise can be incorporated into the general auto-regressive moving average (ARMA) model. For the noise sample $\omega_j^{[n]}$, the ARMA model having an MA order of n_a and AR order of n_b is given by [46]

$$\omega_j^{[n]} = \epsilon_j^{[n]} + \sum_{i=1}^{n_a} a_i \epsilon_j^{[n-i]} - \sum_{i=1}^{n_b} b_i \omega_j^{[n-i]}, \quad (12)$$

where $\epsilon_j^{[n]}$ is the random Gaussian impairment having zero-mean and variance of σ_ϵ^2 , while a_i and b_i are the MA coefficient and AR coefficient, respectively. To derive a more compact form, we set $e_j^{[n]} = \sum_{i=0}^{n_a} a_i \epsilon_j^{[n-i]}$, which is still a zero-mean Gaussian variable with variance of $\sigma_e^2 = \sigma_\epsilon^2 \sum a_i^2$. Then the covariance function of $\omega_j^{[n]}$ is given by

$$N_0 g_n = \begin{cases} \sum_{i=1}^{n_b} b_i g_{n-i} & \text{for } n > 0, \\ \sigma_e^2 + \sum_{i=1}^{n_b} b_i g_{n-i} & \text{for } n = 0. \end{cases} \quad (13)$$

Since the coefficients $\{g_n\}$ are already known, the parameters $\{b_i\}$ and σ_e^2 can be estimated by solving equation (13). Then the correlated noise samples ω_j^n can be iteratively approximated by the ARMA process.

By stacking all transmitted symbols, received samples, channel taps, user states and noise samples into vectors, i.e., \mathbf{x} , \mathbf{r} , \mathbf{h} , $\boldsymbol{\lambda}$, and $\boldsymbol{\omega}$, the joint *a posteriori* distribution reads $p(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\omega} | \mathbf{r})$. In particular, for any unknown variable z , we aim for deriving its marginal distribution $p(z | \mathbf{r})$ and estimating

²Here, we propose a generalized model that the user activity varies per symbol. In practical, the user activity is more likely to be static in a sequence of symbols. In this case, the state transition of λ for two time instants in one sequence is $\lambda_k^{[n]} = \lambda_k^{[n-1]}$.

it via the MMSE or maximum *a posteriori* (MAP) estimators formulated as

$$z = \mathbb{E}_z(z | \mathbf{y}) \text{ or } z = \arg \max_z p(z | \mathbf{r}). \quad (14)$$

Nevertheless, the derivation of the marginal distribution is usually intractable due to the associated high-dimensional integration. The factor graph framework is capable of circumventing this problem by exploiting the conditional independence of variables given the observations. In what follows, we will formulate the probabilistic model and the factor graph representation.

B. Probabilistic Model

A factor graph provides a convenient graphical model representing the factorization of a function. Since \mathbf{x} , \mathbf{h} , $\boldsymbol{\lambda}$, and $\boldsymbol{\omega}$ are independent of each other, the joint distribution $p(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\omega} | \mathbf{r})$ can be factorized as

$$p(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\omega} | \mathbf{r}) \propto p(\mathbf{x}) \cdot p(\mathbf{h}) \cdot p(\boldsymbol{\lambda}) \cdot p(\boldsymbol{\omega}) \cdot p(\mathbf{r} | \mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\omega}). \quad (15)$$

• Factorization of $p(\mathbf{x})$:

The transmitted symbols of different users at different instants are independent, hence $p(\mathbf{x})$ can be fully factorized as

$$p(\mathbf{x}) = \prod_{k,j,n} p(x_{k,j}^{[n]}), \quad (16)$$

where $p(x_{k,j}^{[n]})$ is the *a priori* distribution of $x_{k,j}^{[n]}$, which can be determined based on the log-likelihood ratios (LLRs) output by the channel decoder.

• Factorization of $p(\mathbf{h})$:

According to Bello's model [47], the time-varying channel taps obey a wide-sense stationary random process and their time-domain (TD) correlation is characterized by the Doppler rate $f_D \tau T_0$, where f_D is the Doppler spread. However, it is quite a challenge to model the time evolution of channel taps exactly. Since it is physically plausible that the adjacent sample tends to have the highest correlation for practical correlation functions [48], we characterize the channel taps $h_{k,j}^{[n]}$ by a Gauss-Markov model, given by

$$h_{k,j}^{[n]} = \alpha h_{k,j}^{[n-1]} + \varepsilon^{[n]}, \quad (17)$$

where the coefficient α obeys the zero-order Bessel function of the first kind [48]

$$\alpha = \mathbb{E}[h_{k,j}^{[n]} (h_{k,j}^{[n]})^*] = J_0(2\pi f_D \tau T_0), \quad (18)$$

and $\varepsilon^{[n]}$ is a zero-mean Gaussian distributed variable with variance $1 - |\alpha|^2$. Consequently, the joint *a priori* distribution of \mathbf{h} may be expressed as

$$p(\mathbf{h}) = \prod_{k,j} p(h_{k,j}^0) \prod_n p(h_{k,j}^{[n]} | h_{k,j}^{[n-1]}), \quad (19)$$

where $p(h_{k,j}^{[0]}) \propto \mathcal{G}(m_{h_{k,j}^{[0]}} | v_{h_{k,j}^{[0]}})$ is obtained by performing channel estimation via a few pilot symbols and $p(h_{k,j}^{[n]} | h_{k,j}^{[n-1]})$ is a Gaussian distribution of $h_{k,j}^{[n]}$ as $\mathcal{G}(\alpha h_{k,j}^{[n-1]}, 1 - |\alpha|^2)$.

- Factorization of $p(\boldsymbol{\lambda})$:

In a dynamically fluctuating environment, the evolution of the user-activity state $\boldsymbol{\lambda}$ can be modeled by a Markov chain, i.e., the current activities of the users depend on the previous states. Hence, the pdf $p(\boldsymbol{\lambda})$ can be factorized as

$$p(\boldsymbol{\lambda}) = \prod_{k=1}^K p(\lambda_k^0) \cdot \prod_n p(\lambda_k^{[n]} | \lambda_k^{[n-1]}). \quad (20)$$

Depending on the previous state of user k , the state transition function $p(\lambda_k^{[n]} | \lambda_k^{[n-1]})$ has different expressions. If at time instant $n - 1$ user k is inactive, the probability that user k becomes active at instant n is given by the user-birth probability of $p_{b_k}^{[n]}$. Therefore the probability that the user remains inactive is $1 - p_{b_k}^{[n]}$. By contrast, if $\lambda_k^{[n-1]} = 1$, the probability that user k is inactive at instant n is its mortality probability, given by $p_{m_k}^{[n]}$. Then $p(\lambda_k^{[n]} | \lambda_k^{[n-1]})$ can be expressed as

$$p(\lambda_k^{[n]} | \lambda_k^{[n-1]}) = \begin{cases} (p_{b_k}^{[n]})^{1-\lambda_k^{[n-1]}} (1 - p_{m_k}^{[n]})^{\lambda_k^{[n-1]}} & \lambda_k^{[n]} = 1, \\ (1 - p_{b_k}^{[n]})^{1-\lambda_k^{[n-1]}} (p_{m_k}^{[n]})^{\lambda_k^{[n-1]}} & \lambda_k^{[n]} = 0. \end{cases} \quad (21)$$

It is widely acknowledged that the Poisson distribution accurately model the probability of a given number of events occurring within a fixed time interval or space. Therefore, we assume that at each time instant n the number N of new active users satisfying $k \in \mathcal{K}^+$ and $k \notin \mathcal{K}^+$ at time instant $n - 1$ obeys the Poisson distribution of

$$p(N) = \frac{\Lambda^N}{N!} e^{-\Lambda}, \quad (22)$$

where Λ is the average number of events. Hence, we set $p_{b_k}^{[n]} = p_b = \Lambda/K$ as the birth probability of a user. Next, we consider the mortality probability $p_{m_k}^{[n]}$. In general, an active user has a high probability of being active in the next few time instants. However, establishing an accurate model of the transition probability requires a large amount of data and it is beyond the scope of this paper. Here we employ a fair scheme assuming that $p_{m_k}^{[n]} = 0.5$.

- Factorization of $p(\boldsymbol{\omega})$:

Based on the ARMA model (12) of the noise sample $\omega_j^{[n]}$, we can factorize $p(\boldsymbol{\omega})$ as

$$p(\boldsymbol{\omega}) = \prod_j p(\omega_j^0) \prod_n p(\omega_j^{[n]} | \omega_j^{[n-n_b]}, \dots, \omega_j^{[n-1]}), \quad (23)$$

where $p(\omega_j^0) \propto \mathcal{G}(0, \sigma_e^2)$ and $p(\omega_j^{[n]} | \omega_j^{[n-n_b]}, \dots, \omega_j^{[n-1]})$ formulated as

$$p(\omega_j^{[n]} | \omega_j^{[n-n_b]}, \dots, \omega_j^{[n-1]}) \propto \exp\left(-\frac{(\omega_j^{[n]} - \mathbf{b}^T \boldsymbol{\omega}_j^{[n]})^2}{\sigma_e^2}\right), \quad (24)$$

with the notations $\mathbf{b} = [b_1, \dots, b_{n_b}]^T$ and $\boldsymbol{\omega}_j^{[n]} = [\omega_j^{[n-1]}, \dots, \omega_j^{[n-n_b]}]^T$. We can write furthermore the evolution model of $\boldsymbol{\omega}_j^{[n+1]}$ as

$$\boldsymbol{\omega}_j^{[n+1]} = \mathbf{B}_1 \boldsymbol{\omega}_j^{[n]} + \mathbf{b}_1 \omega_j^{[n]}, \quad (25)$$

where $\mathbf{B}_1 = \begin{bmatrix} \mathbf{0}_{n_b-1}^T & 0 \\ \mathbf{I}_{n_b-1} & \mathbf{0}_{n_b-1} \end{bmatrix}$ and $\mathbf{b}_1 = [1, \mathbf{0}_{n_b-1}^T]^T$.

- Factorization of $p(\mathbf{r} | \mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\omega})$:

Based on (11), we use a Dirac delta function $\delta(\cdot)$ for representing the relationship between the received signal sample and the unknown variables. Therefore, $p(\mathbf{r} | \mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\omega})$ is factorized as

$$p(\mathbf{r} | \mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\omega}) = \prod_{j,n} \delta(r_j^{[n]} - \sum_{k=1}^K \lambda_k^{[n]} h_{k,j}^{[n]} \sum_{i=-L}^L g_i x_{k,j}^{[n-i]} - \omega_j^{[n]}). \quad (26)$$

Note that in (11), when calculating $r_j^{[n]}$, $2KL$ multiplications are required. According to the associative law, we introduce an auxiliary variable $s_{k,j}^{[n]} = \sum_{i=-L}^L g_i x_{k,j}^{[n-i]} = \mathbf{g}^T \tilde{\mathbf{x}}_{k,j}^{[n]}$ to ensure that the number of multiplications is reduced to K , where \mathbf{g} and $\tilde{\mathbf{x}}_{k,j}^{[n]}$ denote the vectors $[g_{-L}, \dots, g_L]^T$ and $[\tilde{x}_{k,j}^{[n+L]}, \dots, \tilde{x}_{k,j}^{[n-L]}]$, respectively. Consequently, we can formulate the factorization of the joint likelihood function as

$$p(\mathbf{r} | \mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\omega}) = \prod_{j,n} \delta(r_j^{[n]} - \sum_{k=1}^K \lambda_k^{[n]} h_{k,j}^{[n]} s_{k,j}^{[n]} - \omega_j^{[n]}) \cdot \delta(s_{k,j}^{[n]} - \mathbf{g}^T \tilde{\mathbf{x}}_{k,j}^{[n]}), \quad (27)$$

where $\tilde{\mathbf{x}}_{k,j}^{[n]}$ follows a similar evolution model as in (25),

$$\tilde{\mathbf{x}}_{k,j}^{[n]} = \mathbf{B}_2 \tilde{\mathbf{x}}_{k,j}^{[n-1]} + \mathbf{b}_2 x_{k,j}^{[n+L]}, \quad (28)$$

where $\mathbf{B}_2 = \begin{bmatrix} \mathbf{0}_{2L}^T & 0 \\ \mathbf{I}_{2L} & \mathbf{0}_{2L} \end{bmatrix}$ and $\mathbf{b}_2 = [1, \mathbf{0}_{2L}^T]^T$.

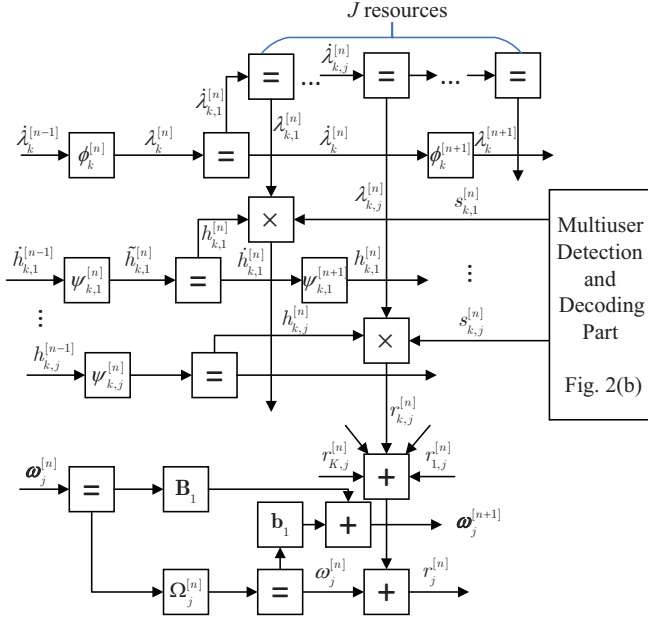
C. Factor Graph Representation

Based on the factorizations of the joint *a priori* distributions and likelihood function (16)-(27), we now have the factorization of $p(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\omega} | \mathbf{r})$ and represent it by a factor graph, as depicted in Fig. 2. On this factor graph, the factor nodes denoted by squares represent the functions while the variables are denoted by edges. An edge is connected to a factor node if and only if the variable is an argument of the function. Equality factor nodes represented by the symbol \square for variable ‘cloning’ are introduced to enforce the condition that a variable may only appear in a maximum of two functions.

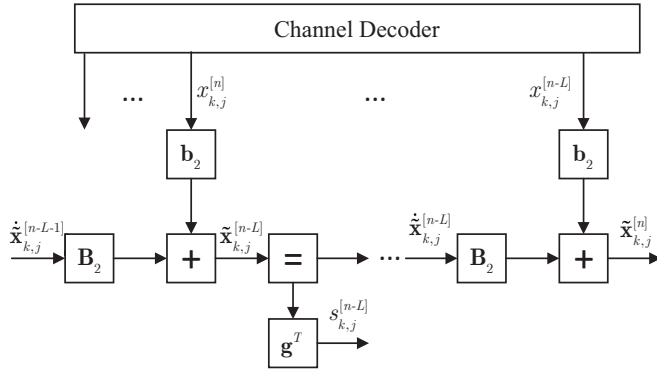
IV. EM-MPA BASED ITERATIVE RECEIVER DESIGN

In this section, we aim for designing a low complexity receiver for joint user activity tracking as well as channel estimation and data detection. An amalgamated expectation maximization (EM) aided message passing algorithm (MPA) method is proposed on the factor graph to determine the beliefs of all unknown variables. To simplify the notations, we use $\overrightarrow{\mu}(x)$ and $\overleftarrow{\mu}(x)$ to denote the messages of the variable x that flow in the same and in the opposite the direction as the edge.

We use the block diagram of Fig. 3 for portraying the proposed EM-MPA method. In an iterative receiver extrinsic soft information is exchanged between the detector and the channel decoder. The soft outputs of the channel decoder are approximated using a Gaussian distribution by invoking the



(a) Factor graph representation of the joint distribution (15). The notations $\phi_k^{[n]}$, $\psi_{k,j}^{[n]}$, and $\Omega_j^{[n]}$ represent the function $p(\lambda_k^{[n]}|\lambda_k^{[n-1]})$, $p(h_{k,j}^{[n]}|h_{k,j}^{[n-1]})$ and $p(\omega_j^{[n]}|\omega_j^{[n-1]}, \dots, \omega_j^{[n-b]})$; For sequence based user activity identification, the factor node $\phi_k^{[n]}$ denotes an equality node if indices $\cdot^{[n-1]}$ and $\cdot^{[n]}$ belong to the same sequence of symbols.



(b) Subgraph for the multiuser detection and decoding part.

Fig. 2. Factor graph for joint user activity tracking and data detection.

expectation propagation (EP) method, which will be detailed in Section IV-B. The receiver is constituted of three blocks, namely the channel estimation, detection and user-activity identification. The message derivations in each block are based on the conventional MPA rules, as shown in Section IV-C, while the messages exchanged among the three blocks are determined by the EM algorithm, as discussed in Section IV-D.

A. Brief Review of the Message Passing Algorithm

The MPA aims for efficiently computing the marginals of variables on the factor graph. For each edge on the factor graph, there are messages gleaned from and entered into the adjacent factor nodes. According to the MPA rules, the message passing from a factor node to a connected factor node via an edge is computed by integrating all other variables in this factor [49]. Considering the factor node $\psi_{k,1}^{[n]}$ connected with $\hat{h}_{k,1}^{[n-1]}$ and $\tilde{h}_{k,1}^{[n]}$ in Fig. 2 as an example, the message $\bar{\mu}(\tilde{h}_{k,1}^{[n]})$ is given by

$$\bar{\mu}(\tilde{h}_{k,1}^{[n]}) = \int p(\tilde{h}_{k,1}^{[n]}|\hat{h}_{k,1}^{[n-1]}) \bar{\mu}(\hat{h}_{k,1}^{[n-1]}) d\hat{h}_{k,1}^{[n-1]}. \quad (29)$$

Having determined all messages on the factor graph, the belief of a variable x is simply calculated by the product of the forward and backward messages, which is formulated as

$$b(x) = \bar{\mu}(x) \overleftarrow{\mu}(x). \quad (30)$$

B. Detection and Decoding Part

Based on L_k^e output by the channel decoder, we can calculate the corresponding *a priori* distributions $p(x_{k,j}^{[n]})$ of user k 's transmitted symbols, which is given by the intrinsic information $\bar{\mu}(x_{k,j}^{[n]})$. Since $x_{k,j}^{[n]}$ is discretely distributed, the MPA exhibits an excessive complexity. Instead of applying the central limit theorem, we project $\bar{\mu}(x_{k,j}^{[n]})$ onto the Gaussian distribution by using the expectation propagation method.

As an approximate inference method, EP aims for finding a Gaussian belief $\tilde{b}(x_{k,j}^{[n]})$ for approximating the original belief $b(x_{k,j}^{[n]})$ by minimizing the relative entropy between the two distributions. Assuming that the extrinsic message $\overleftarrow{\mu}(x_{k,j}^{[n]})$ obeys the Gaussian distribution of $\mathcal{G}(\overleftarrow{m}_{x_{k,j}^{[n]}}, \overleftarrow{v}_{x_{k,j}^{[n]}})$, we can readily obtain the mean and variance of $b(x_{k,j}^{[n]})$ by moment matching and then determine the Gaussian belief $\tilde{b}(x_{k,j}^{[n]}) \propto \mathcal{G}(m_{x_{k,j}^{[n]}}, v_{x_{k,j}^{[n]}})$. Hence we have the message $\bar{\mu}(x_{k,j}^{[n]})$ expressed as

$$\bar{\mu}(x_{k,j}^{[n]}) = \frac{\tilde{b}(x_{k,j}^{[n]})}{\overleftarrow{\mu}(x_{k,j}^{[n]})} \propto \mathcal{G}(\overrightarrow{m}_{x_{k,j}^{[n]}}, \overrightarrow{v}_{x_{k,j}^{[n]}}). \quad (31)$$

Having obtained $\overrightarrow{m}_{x_{k,j}^{[n]}}$ and $\overrightarrow{v}_{x_{k,j}^{[n]}}$, we can readily drive the mean vector and covariance matrix of $\bar{\mu}(\tilde{x}_{k,j}^{[n-L]})$ as

$$\overrightarrow{m}_{\tilde{x}_{k,j}^{[n-L]}} = \mathbf{b}_2 \overrightarrow{m}_{x_{k,j}^{[n]}} + \mathbf{B}_2 \overrightarrow{m}_{\tilde{x}_{k,j}^{[n-L-1]}}, \quad (32)$$

$$\overrightarrow{V}_{\tilde{x}_{k,j}^{[n-L]}} = \mathbf{b}_2 \overrightarrow{v}_{x_{k,j}^{[n]}} \mathbf{b}_2^T + \mathbf{B}_2 \overrightarrow{V}_{\tilde{x}_{k,j}^{[n-L-1]}} \mathbf{B}_2^T. \quad (33)$$

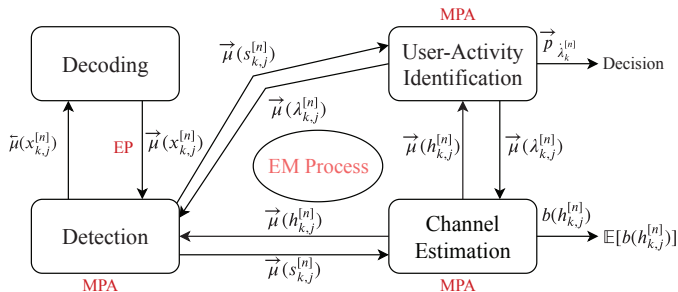


Fig. 3. Block diagram of the proposed EM-MPA.

By employing the MPA rules, we can determine the mean vector and covariance matrix of message $\vec{\mu}(\dot{\mathbf{x}}_{k,j}^{[n-L]})$. Therefore the message output by the multiuser detector and by the FEC decoder can be assumed to be Gaussian and formulated as

$$\vec{\mu}(s_{k,j}^{[n-L]}) \propto \mathcal{G}(\mathbf{g}_T \vec{\mathbf{m}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}}, \mathbf{g}_T \vec{\mathbf{V}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}} \mathbf{g}). \quad (34)$$

Finally, we are interested in calculating the extrinsic message $\overleftarrow{\mu}(x_{k,j}^{[n]})$, whose mean and variance obey

$$\overleftarrow{m}_{x_{k,j}^{[n]}} = \mathbf{b}_2^T (\overleftarrow{\mathbf{m}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}} - \mathbf{B}_2 \overleftarrow{\mathbf{m}}_{\dot{\mathbf{x}}_{k,j}^{[n-L-1]}}), \quad (35)$$

$$\overleftarrow{v}_{x_{k,j}^{[n]}} = \mathbf{b}_2^T (\overleftarrow{\mathbf{V}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}} + \mathbf{B}_2 \overleftarrow{\mathbf{V}}_{\dot{\mathbf{x}}_{k,j}^{[n-L-1]}} \mathbf{B}_2^T) \mathbf{b}_2, \quad (36)$$

where $\overleftarrow{\mathbf{m}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}}$ and $\overleftarrow{\mathbf{V}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}}$ are expressed as

$$\overleftarrow{\mathbf{m}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}} = \overleftarrow{\mathbf{V}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}} (\overleftarrow{\mathbf{W}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}} \overleftarrow{\mathbf{m}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}} + \mathbf{g}_{s_{k,j}^{[n-L]}}^T \overleftarrow{m}_{s_{k,j}^{[n-L]}}), \quad (37)$$

$$\overleftarrow{\mathbf{V}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}} = (\overleftarrow{\mathbf{W}}_{\dot{\mathbf{x}}_{k,j}^{[n-L]}} + \mathbf{g}_{s_{k,j}^{[n-L]}}^T \mathbf{g}^T)^{-1}. \quad (38)$$

C. MPA for Message Calculations in Fig. 2(a)

This subsection aims for calculating the messages in the channel estimation block and the user-activity identification block of Fig. 3.

1) *Channel Estimation*: Let us commence the discussion with the message calculations related to channel estimation. Provided that the message $\vec{\mu}(\dot{h}_{k,j}^{[n-1]})$ is available in Gaussian form, by applying the MPA rules, the message $\vec{\mu}(\dot{h}_{k,j}^{[n]})$ is expressed as

$$\begin{aligned} \vec{\mu}(\dot{h}_{k,j}^{[n]}) &\propto \mathcal{G}(\overrightarrow{m}_{\dot{h}_{k,j}^{[n]}}, \overrightarrow{v}_{\dot{h}_{k,j}^{[n]}}) \\ &\propto (\alpha \overrightarrow{m}_{\dot{h}_{k,j}^{[n-1]}} + 1 + |\alpha|^2 (\overrightarrow{v}_{\dot{h}_{k,j}^{[n-1]}} - 1)). \end{aligned} \quad (39)$$

Then, we are able to derive the message $\vec{\mu}(\dot{h}_{k,j}^{[n]})$, given by

$$\overrightarrow{v}_{\dot{h}_{k,1}^{[n]}} = \frac{\overrightarrow{v}_{\dot{h}_{k,1}^{[n-1]}} + \overleftarrow{v}_{\dot{h}_{k,1}^{[n]}}}{\overrightarrow{v}_{\dot{h}_{k,1}^{[n-1]}} \overleftarrow{v}_{\dot{h}_{k,1}^{[n]}}}, \quad (40)$$

$$\overrightarrow{m}_{\dot{h}_{k,1}^{[n]}} = \overrightarrow{v}_{\dot{h}_{k,1}^{[n-1]}} \left(\frac{\overrightarrow{m}_{\dot{h}_{k,1}^{[n-1]}}}{\overrightarrow{v}_{\dot{h}_{k,1}^{[n-1]}}} + \frac{\overleftarrow{m}_{\dot{h}_{k,1}^{[n]}}}{\overleftarrow{v}_{\dot{h}_{k,1}^{[n]}}} \right). \quad (41)$$

In (40) and (41), we have assumed that $\overleftarrow{\mu}(h_{k,j}^{[n]})$ follows $\mathcal{G}(\overleftarrow{m}_{h_{k,j}^{[n]}} + \overleftarrow{v}_{h_{k,j}^{[n]}})$.

Next, we consider the process of colored noise. Since the ARMA process given by (12) is causal, the messages are only propagated forward along the arrow's direction. Provided that the means of the noise parameters are 0, we can readily derive the corresponding messages as follows

$$\overrightarrow{v}_{\omega_j^{[n]}} = \mathbf{b}_1^T \overrightarrow{\mathbf{V}}_{\omega_j^{[n]}} \mathbf{b}_1, \quad (42)$$

$$\overrightarrow{\mathbf{V}}_{\omega_j^{[n+1]}} = \mathbf{B}_1 \overrightarrow{\mathbf{V}}_{\omega_j^{[n]}} \mathbf{B}_1^T + \mathbf{b}_1 \overrightarrow{v}_{\omega_j^{[n]}} \mathbf{b}_1^T. \quad (43)$$

Consequently, we can calculate the message $\overleftarrow{\mu}(r_{k,j}^{[n]}) \propto \mathcal{G}(\overleftarrow{m}_{r_{k,j}^{[n]}} + \overleftarrow{v}_{r_{k,j}^{[n]}})$ with

$$\overleftarrow{m}_{r_{k,j}^{[n]}} = r_j^{[n]} - \sum_{k' \neq k} \overleftarrow{m}_{r_{k',j}^{[n]}} \quad (44)$$

$$\overleftarrow{v}_{r_{k,j}^{[n]}} = \overrightarrow{v}_{\omega_j^{[n]}} + \sum_{k' \neq k} \overrightarrow{v}_{r_{k',j}^{[n]}} \quad (45)$$

where the messages $\overleftarrow{\mu}(r_{k',j}^{[n]})$, $\forall k' \neq k$ will be determined in (56) and (57).

2) *User-activity Identification*: For the discrete random variable representing a user-activity state at time instant $n-1$, we have $\dot{\lambda}_k^{[n-1]}$, and the message $\vec{\mu}(\dot{\lambda}_k^{[n-1]})$ is the belief of user k 's state at time instant $n-1$, given by

$$\vec{\mu}(\dot{\lambda}_k^{[n-1]}) = (\overrightarrow{p}_{\dot{\lambda}_k^{[n-1]}})^{\lambda_k^{[n-1]}} \cdot (1 - \overrightarrow{p}_{\dot{\lambda}_k^{[n-1]}})^{1 - \lambda_k^{[n-1]}}, \quad (46)$$

with $\overrightarrow{p}_{\dot{\lambda}_k^{[n-1]}}$ being the probability of $\lambda_k^{[n-1]} = 1$. Motivated by the fact that $\vec{\mu}(\dot{\lambda}_k^{[n-1]})$ is fully characterized by $\overrightarrow{p}_{\dot{\lambda}_k^{[n-1]}}$, passing the user-activity probability $\overrightarrow{p}_{\dot{\lambda}_k^{[n-1]}}$ instead of the message can simplify the expressions. Hence, we arrive at the forward message $\overrightarrow{p}(\lambda_k^{[n]})$ as

$$\overrightarrow{p}_{\lambda_k^{[n]}} = (1 - p_{m_k}^{[n]}) \overrightarrow{p}_{\dot{\lambda}_k^{[n]}} + p_{b_k}^{[n]} (1 - \overrightarrow{p}_{\dot{\lambda}_k^{[n]}}). \quad (47)$$

The equality node is equivalent to the product of messages. Therefore the message updating concerning $\dot{\lambda}_k^{[n]}$ is derived as

$$\overrightarrow{p}_{\dot{\lambda}_k^{[n]}} = \frac{\overrightarrow{p}_{\dot{\lambda}_k^{[n-1]}} \overleftarrow{p}_{\dot{\lambda}_k^{[n-1]}}}{1 - \overrightarrow{p}_{\dot{\lambda}_k^{[n-1]}} - \overleftarrow{p}_{\dot{\lambda}_k^{[n-1]}}}. \quad (48)$$

The message $\overrightarrow{p}_{\dot{\lambda}_k^{[n]}}$ forward to the multiplier node can be obtained similarly.

Note that the above message calculations depend on the assumption of having known backward messages from the multiplier node. According to the conventional update rules, the messages derived at the multiplier node $\boxed{\times}$ are non-Gaussian. Hence we invoke the expectation maximization algorithm for the multiplier node.

D. Modified EM Algorithm for '×' Node

The conventional EM algorithm aims for finding the maximum likelihood estimates of parameters in a statistical model that has unobserved latent variables [45]. In the E-step, the expected log-likelihood function is evaluated based on the beliefs of the latent variables. Then in the M-step, the estimates of the parameters are found by maximizing the expected log-likelihood. Without loss of generality, we consider the multiplier node connected with $r_{k,j}^{[n]}$ and the joint distribution $p(\lambda_{k,j}^{[n]}, s_{k,j}^{[n]}, h_{k,j}^{[n]} | r_{k,j}^{[n]})$.

We first define $\lambda_{k,j}^{[n]}$ as the unknown parameter and $\{h_{k,j}^{[n]}, s_{k,j}^{[n]}, r_{k,j}^{[n]}\}$ as the complete data set associated with incomplete data $r_{k,j}^{[n]}$ and latent variables $h_{k,j}^{[n]}, s_{k,j}^{[n]}$. Assuming

that the beliefs $b(s_{k,j}^{[n]})$ and $b(h_{k,j}^{[n]})$ are available, the expectation of the complete-data log augmented density is calculated as

$$\begin{aligned} q(\lambda_{k,j}^{[n]}) &\propto \int b(s_{k,j}^{[n]})b(h_{k,j}^{[n]}) \ln p(\lambda_{k,j}^{[n]}, h_{k,j}^{[n]}, s_{k,j}^{[n]} | r_{k,j}^{[n]}) dh_{k,j}^{[n]} ds_{k,j}^{[n]} \\ &= - \int \frac{(\overleftarrow{m}_{r_{k,j}^{[n]}} - \lambda_{k,j}^{[n]} s_{k,j}^{[n]} h_{k,j}^{[n]})^2}{\overleftarrow{v}_{r_{k,j}^{[n]}}} \cdot b(s_{k,j}^{[n]})b(h_{k,j}^{[n]}) dh_{k,j}^{[n]} ds_{k,j}^{[n]} \\ &\quad + \ln \overrightarrow{\mu}(\lambda_{k,j}^{[n]}) + C, \end{aligned} \quad (49)$$

where C is a constant that is irrelevant to $\lambda_{k,j}^{[n]}$. Maximizing (49) yields the estimate $\hat{\lambda}_{k,j}^{[n]}$ which is equivalent to the solution of $\frac{\partial q(\lambda_{k,j}^{[n]})}{\partial \lambda_{k,j}^{[n]}} = 0$. However, note that (49) only considers the j th resource element, while the user activity applies to all radio resource elements. The maximization should be performed by obtaining the necessary information from all J radio resource elements, i.e., with respect to the variable λ_k^n . To this end, the multiplier node will feed back the message $\overleftarrow{\mu}(\lambda_{k,j}^{[n]})$ to the equality node of Fig. 2(a). Having $q(\lambda_{k,j}^{[n]})$ in hand, we have

$$\overleftarrow{\mu}(\lambda_{k,j}^{[n]}) \propto \frac{\exp[q(\lambda_{k,j}^{[n]})]}{\overrightarrow{\mu}(\lambda_{k,j}^{[n]})}. \quad (50)$$

Similar to the forward message $\overrightarrow{p}(\lambda_k^{[n]})$, we can simply pass on the normalized probability $\overleftarrow{p}_{\lambda_{k,j}^{[n]}}$ since $\overleftarrow{\mu}(\lambda_{k,j}^{[n]})$ depends on the normalized probability of $\lambda_{k,j}^{[n]} = 1$, which is given by

$$\overleftarrow{p}_{\lambda_{k,j}^{[n]}} = \frac{(1 - \overrightarrow{p}_{\lambda_{k,j}^{[n]}})q(1)}{(1 - \overrightarrow{p}_{\lambda_{k,j}^{[n]}})q(1) + \overrightarrow{p}_{\lambda_{k,j}^{[n]}}q(0)}. \quad (51)$$

The probability $\overleftarrow{p}_{\lambda_{k,j}^{[n]}}$ is used for calculating $\overleftarrow{p}_{\lambda_{k,j-1}^{[n]}}$, then $\overleftarrow{p}_{\lambda_{k,1}^{[n]}}$ and finally $\overrightarrow{p}_{\lambda_k^{[n]}}$.

To obtain the beliefs of $s_{k,j}^{[n]}$ and $h_{k,j}^{[n]}$, we apply the concept of EM again, so that $s_{k,j}^{[n]}$ becomes the unknown parameter and $h_{k,j}^{[n]}$ remains the latent variable. In this way, the belief of $s_{k,j}^{[n]}$ is updated as follows,

$$\begin{aligned} b(s_{k,j}^{[n]}) &\propto \\ &\overrightarrow{\mu}(s_{k,j}^{[n]}) \cdot \exp\left(\int b(h_{k,j}^{[n]}) \ln p(r_{k,j}^{[n]}, s_{k,j}^{[n]} | h_{k,j}^{[n]}, \hat{\lambda}_{k,j}^{[n]}) dh_{k,j}^{[n]}\right), \end{aligned} \quad (52)$$

where $\lambda_{k,j}^{[n]}$ is replaced by the estimate $\hat{\lambda}_{k,j}^{[n]}$ obtained from the maximization of (49), expressed as

$$\hat{\lambda}_{k,j}^{[n]} = \frac{\overleftarrow{m}_{r_{k,j}^{[n]}} m_{s_{k,j}^{[n]}} m_{h_{k,j}^{[n]}} + \overleftarrow{v}_{r_{k,j}^{[n]}} (1 - 2\overrightarrow{p}_{\lambda_{k,j}^{[n]}})}{(|m_{s_{k,j}^{[n]}}|^2 + v_{s_{k,j}^{[n]}})(|m_{h_{k,j}^{[n]}}|^2 + v_{h_{k,j}^{[n]}})}. \quad (53)$$

Since we have $b(s_{k,j}^{[n]}) = \overrightarrow{\mu}(s_{k,j}^{[n]}) \cdot \overleftarrow{\mu}(s_{k,j}^{[n]})$, it is natural to define the second term on the righthand side of (52) as $\overleftarrow{\mu}(s_{k,j}^{[n]})$. Assuming that $b(h_{k,j}^{[n]})$, it follows that $\mathcal{G}(m_{h_{k,j}^{[n]}}, v_{h_{k,j}^{[n]}})$, $\overleftarrow{\mu}(s_{k,j}^{[n]})$

can be modeled by a Gaussian PDF having the mean and variance of

$$\overleftarrow{m}_{s_{k,j}^{[n]}} = \frac{\overleftarrow{m}_{r_{k,j}^{[n]}} m_{h_{k,j}^{[n]}}}{|m_{h_{k,j}^{[n]}}|^2 + v_{h_{k,j}^{[n]}}}, \quad (54)$$

$$\overleftarrow{v}_{s_{k,j}^{[n]}} = \frac{\overleftarrow{v}_{r_{k,j}^{[n]}}}{|m_{h_{k,j}^{[n]}}|^2 + v_{h_{k,j}^{[n]}}}. \quad (55)$$

Consequently the belief $b(s_{k,j}^{[n]})$ is readily obtained. By exchanging the roles of $s_{k,j}^{[n]}$ and $h_{k,j}^{[n]}$, we have the updating rules of the message $\overleftarrow{\mu}(h_{k,j}^{[n]})$ and belief $b(h_{k,j}^{[n]})$. After obtaining the beliefs $b(s_{k,j}^{[n]})$ and $b(h_{k,j}^{[n]})$, we can now determine $Q(\lambda_{k,j}^{[n]})$ in the next iteration following (49).

Let us now consider the forward message $\overrightarrow{\mu}(r_{k,j}^{[n]})$. Since the variables $\lambda_{k,j}^{[n]}$, $s_{k,j}^{[n]}$ and $h_{k,j}^{[n]}$ are independent, the moments of $r_{k,j}^{[n]}$ are given by the product of the moments of the above three variables. Consequently, we have

$$\overrightarrow{m}_{r_{k,j}^{[n]}} = \hat{\lambda}_{k,j}^{[n]} m_{h_{k,j}^{[n]}} m_{s_{k,j}^{[n]}}}, \quad (56)$$

$$\begin{aligned} \overrightarrow{v}_{r_{k,j}^{[n]}} &= \hat{\lambda}_{k,j}^{[n]} ((1 - \hat{\lambda}_{k,j}^{[n]}) |m_{h_{k,j}^{[n]}}|^2 |m_{s_{k,j}^{[n]}}|^2 + |m_{h_{k,j}^{[n]}}|^2 v_{s_{k,j}^{[n]}} \\ &\quad + |m_{s_{k,j}^{[n]}} v_{h_{k,j}^{[n]}} + v_{h_{k,j}^{[n]}} v_{s_{k,j}^{[n]}}|^2). \end{aligned} \quad (57)$$

Above, we have obtained $\overrightarrow{p}_{\lambda_k^{[n]}}$ and $b(h_{k,j}^{[n]})$. Then we can compare $\overrightarrow{p}_{\lambda_k^{[n]}}$ to a specific threshold to decide whether the user k is active, while the estimate of the channel tap $h_{k,j}^{[n]}$ is given by

$$\hat{h}_{k,j}^{[n]} = \mathbb{E}_h [b(h_{k,j}^{[n]})] = m_{h_{k,j}^{[n]}}. \quad (58)$$

E. Algorithm Summary

We have derived the beliefs concerning the unknown variables and message updtings on factor graph in closed forms. The detector and decoder exchange extrinsic information and iteratively update their corresponding messages. To start the iterative algorithm, the user state, channel taps and noise samples are predicted based on state evolution functions and the previously obtained beliefs. Then the proposed EM-MPA algorithm is employed for updating the messages and for calculating the extrinsic information for the channel decoder. After channel decoding, the extrinsic information generated is transformed to a Gaussian prior via EP approximation. The priors are again feedback to the detector for refining the parameters using our EM-MPA algorithm. This process will be repeated for several times for refining the beliefs of channel taps and user states. By using appropriate estimators, we arrive at the estimates of channel information and user activities.

In each iteration, the complexity of the proposed algorithm is dominated by the integrations. For a conventional MPA receiver, the computations of the integration for obtaining the messages involve multiple variables and the discrete priors make the optimal detection excessively complex due to the associated multi-dimensional search, leading to an exponential order of complexity as a function of the number of users and the number of ISI taps. Explicitly, we have a complexity order

of $\mathcal{O}(|\chi|^{2|\mathcal{K}^+|L})$ with χ representing the set of constellation points. The block orthogonal matching pursuit (BOMP) algorithm of [42] is a prominent method of sparse signal estimation, which has a complexity increasing with the cube of the number of active users. By contrast, in the proposed algorithm, the messages are represented in a parametric form. Hence the integrations are simplified to addition and multiplication operations, which dramatically reduces the complexity of the receiver. Explicitly, the complexity only increases linearly with the number of active users and with the FTN-induced ISI length, according to $\mathcal{O}(2|\mathcal{K}^+|L)$. This clearly shows the superiority of the proposed algorithm in handling massive MTC scenarios relying on random access. On the other hand, bearing in mind the complexity escalation imposed by the further increase of active users, we can specifically design the LDS codewords for further exploring the sequence sparsity, which will reduce the IUI and control the complexity of the receiver.

V. SIMULATION RESULTS

In this section, our simulation results are discussed. We adopt a 1/2-rate low density parity check (LDPC) code having a length of 2,160, which has been optimized for FTN signaling according to the criteria in [50]. The number of users and radio resource elements is $K = 180$ and $J = 120$, respectively, leading to the normalized user-load of $\rho = 150\%$. The low-density signature scheme of [51] is employed as the spreading matrix and Quadrature Phase Shift Keying (QPSK) is used for constellation mapping. We assume that 5 sequences of symbols are transmitted for each user and each sequence has a length of 1080. The symbols corresponding to different users are shaped by a root raised cosine (RRC) filter conceived with FTN packing ratio of $\tau = 0.8$ and Nyquist roll-off factor of $\alpha = 0.4$. According to [20], the interfered symbols that are ‘far’ from the current transmitted symbol have negligible effects. Therefore, we set the number of ISI contributions engendered by FTN signaling to $L = 10$. The channel obeys Rayleigh fading and the taps are generated by Jake’s model with a fading rate of $f_D\tau T_0 = 0.005$. The initial channel information is obtained via coarse channel estimation using as 5 pilots. The parameter Λ is set to 20, which indicates that approximately 11% of users are active. Again, the mortality probability was set to $p_{m_k}^{[n]} = 0.5$. The user activity is assumed to remain static for a sequence of 1,080 symbols. The threshold of 0.5 is employed for user activity identification. Finally, we set the maximum number of ‘turbo’ iterations to $L_{\text{turbo}} = 10$ and the maximum number of decoding iterations to $L_{\text{dec}} = 20$.

We first compare the BER performance of the proposed FTN-NOMA system and the FTN-OMA, Nyquist-NOMA and orthogonal systems in Fig. 4. For fair comparison, we assume that all in these four cases the number of active users is fixed to $u = 20$ and their activity-states are known at the BS. In other words, we do not have to estimate the user states. For the orthogonal system, interference free transmission is achieved. Hence, a simple linear minimum mean squared error estimation is employed at the BS. For the other three non-orthogonal systems, the proposed EM-MPA algorithm is both

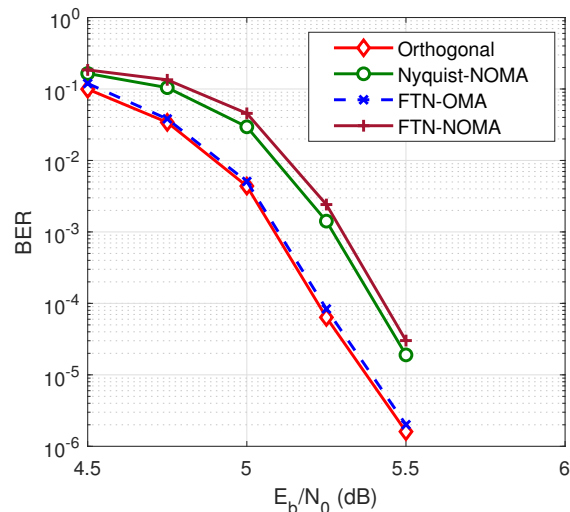


Fig. 4. BER performance comparison of orthogonal and non-orthogonal transmission schemes.

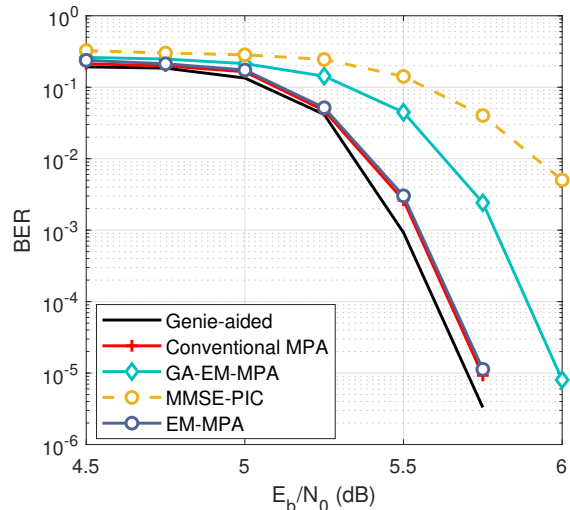


Fig. 5. BER performance of the proposed algorithm and existing ones with known user activities.

used for eliminating the interferences and for estimating the channels. As observed from Fig. 4, the orthogonal system has the best BER performance. The FTN-NOMA system attains similar performance to the orthogonal system with a receiver complexity of $\mathcal{O}(2L)$. The Nyquist-NOMA system exhibits a 0.2 dB performance loss, but the number of users supported is increased by $\frac{180-120}{120} \times 100\% = 50\%$, at the cost of a computational complexity of $\mathcal{O}(|\mathcal{K}^+|)$. The proposed FTN-NOMA scheme suffers from a slight E_b/N_0 degradation compared to the orthogonal system due to the introduction of both ISI and IUI. Nevertheless, the proposed FTN-NOMA scheme is capable of increasing the data rate by $\rho \cdot \frac{1}{\tau} - 100\% = 87.5\%$, given the same number of radio resource elements and a slightly higher receiver complexity of $\mathcal{O}(2|\mathcal{K}^+|L)$.

In Fig. 5, we compare the proposed receiver design to some existing reference algorithms in terms of its BER performance, when the user-activities are perfectly known. We also

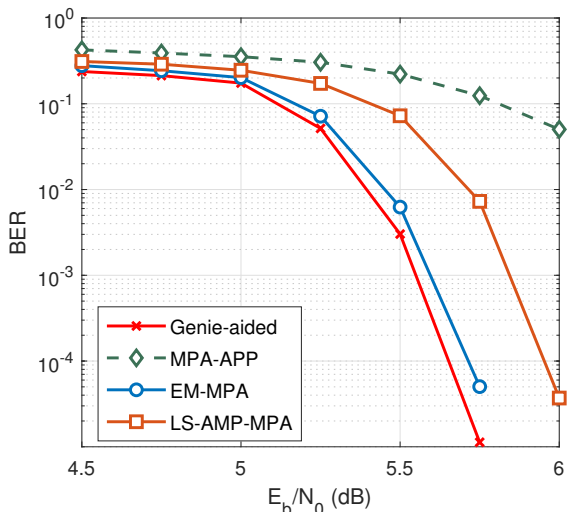


Fig. 6. BER performance of the proposed algorithm and existing ones with unknown user activities.

include the ‘Genie-aided’ method as the ultimate bound, which assumes having perfectly known channel information. The MMSE-PIC method estimates the channel information using the MMSE method and iteratively detects the data symbols using parallel interference cancellation (PIC), which is seen to suffer from error propagation. Compared to the conventional MPA algorithm relying on the MAP criterion, the proposed EM-MPA algorithm slightly degrades the performance, but its complexity is low. Finally, the performance for the GA-EM-MPA method is also shown, which approximates all non-Gaussian messages by Gaussian distribution using the moment matching. The proposed FTN-NOMA algorithm outperforms the GA-EM-MPA method, since the EP exploits the extrinsic information gleaned from the MUD.

In Fig. 6, we consider a practical random access scenario. The BER performances versus E_b/N_0 of the proposed algorithm as well as of the MPA-APP, BOMP-MPA and LS-AMP-MPA methods are illustrated. The BOMP-MPA and LS-AMP-MPA are two-step methods which firstly identify the active users and then perform MPA based MUD. It can be observed from Fig. 6 that BOMP-MPA achieves a beneficial performance gain compared to the LS-AMP-MPA method by exploiting the sparsity of the spreading matrix. However, the BOMP method requires the number of active users $|\mathcal{K}^+|$ to be known at the BS, which limits its application in practical MTC relying on random access. Moreover, since the two-stage methods only provide the estimates of user-activities for data detection, considerable performance loss can be seen compared to the proposed EM-MPA algorithm. The FTN-NOMA system relying on the MPA-APP detection method that assumes all users to be active suffers from a significant performance degradation. To show the efficiency of the proposed user activity identification method, we also include the curve corresponding to the EM-MPA algorithm in Fig. 5 (EM-MPA-Ideal) here.

Fig. 7 and Fig. 8 depict the equivalent spectral efficiency versus E_b/N_0 parameterized by different values of the packing

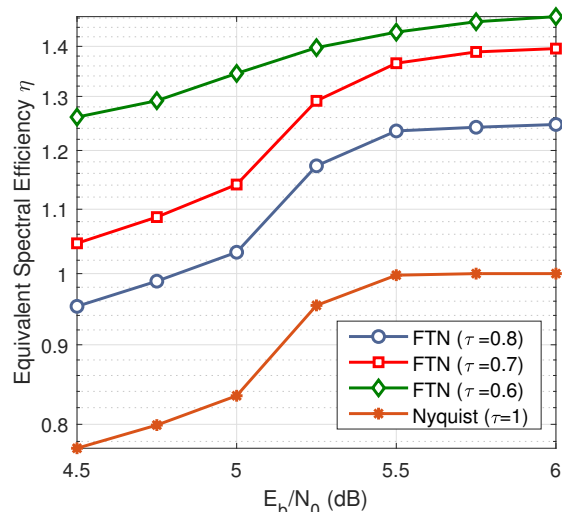


Fig. 7. Equivalent spectral efficiency for different values of packing factor τ .

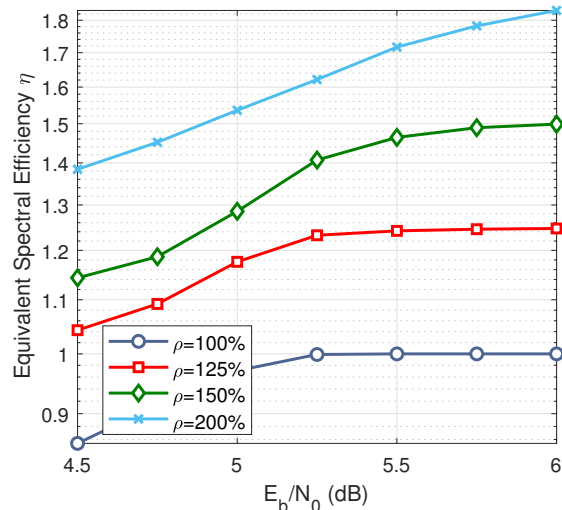


Fig. 8. Equivalent spectral efficiency for different values of normalized overloaded factor ρ .

factor τ and normalized user-load ρ , respectively. The equivalent spectral efficiency is defined as

$$\eta = \frac{\rho}{\tau} * (1 - \text{BER}), \quad (59)$$

which indicates the tradeoff between the spectral efficiency and the BER performance. Observe from Fig. 7 that when the packing factor decreases, the equivalent spectral efficiency gap between the FTN signaling and the classic Nyquist scenario becomes wider. This is because more data symbols can be transmitted within the same time period. However, we can see the gain of η becomes marginal when we further reduce the packing factor τ . This is because a lower value of τ lead to more severe interference and result in worse BER performance. Fig. 8 illustrates equivalent spectral efficiency for four cases with the number of users being $K = 120, 160, 180,$ and 240 , leading to different levels of normalized user-load, $\rho = 100\%, 125\%, 150\%, 200\%$. As expected, increasing

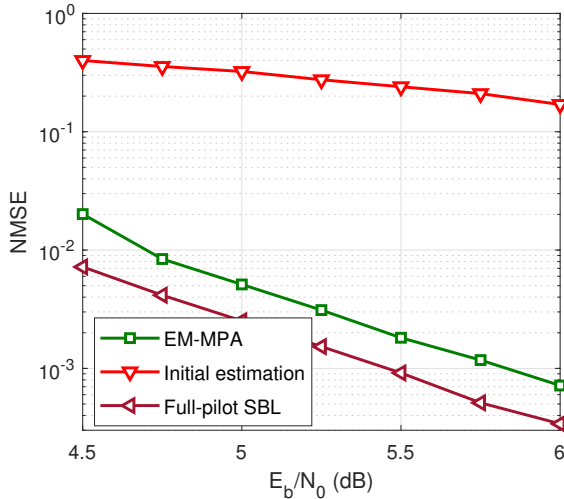


Fig. 9. NMSE of channel estimation versus E_b/N_0 .

the user-load results in higher equivalent spectral efficiency, showing its potential in supporting massive connectivity in future IoT applications. Observe from both 2 figures that we can indeed further increase the throughput at the cost of some BER performance degradation and increased receiver complexity. In our future work, we aim for jointly designing the FTN-NOMA system to strike an improved throughput versus decoding performance tradeoff.

We plot the normalized minimum mean squared errors (NMSE) of the channel estimate based on the proposed algorithm as well as on the GA-MPA and on the full pilot based sparse Bayesian learning (SBL) [52] method in Fig. 9. The initial channel estimation result using 5 pilots is also depicted. Compared to the pilot based initial channel estimation, the GA-EM-MPA and the proposed algorithms are observed to have improved performance by exploiting the data symbols for channel estimation. Moreover, the comparison of the GA-MPA and of the the proposed EM-MPA algorithm's curves demonstrates the benefits of using the EP method for Gaussian approximation. It is also worth noting that the NMSE performance of the proposed algorithm has only 0.3 dB performance loss compared to the full-pilot based method, which demonstrates the powerful capability of the proposed algorithm.

To show the user activity tracking accuracy, we depict the cumulative distribution functions (CDFs) of the user identification errors for different values of Λ in Fig. 10. We conclude that a higher value of Λ leads to better user-activity identification. Considering the extreme case, when all users are active, the identification error will drop to 0. Moreover, it is seen that for all four levels of user-activity, only a very small fraction of the users' states are mis-determined. Finally, we consider the impact of the number of active users on the decoding performance for the FTN-NOMA system. The BER performance versus the parameter Λ at $E_b/N_0 = 5.75$ dB is plotted in Fig. 11, where the performance of the EM-MPA-Ideal method that assumes perfect user-activity identification is

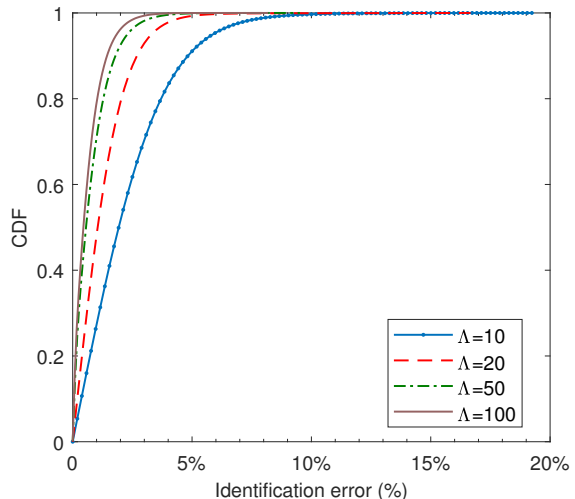


Fig. 10. CDFs of the user identification errors for different Λ .

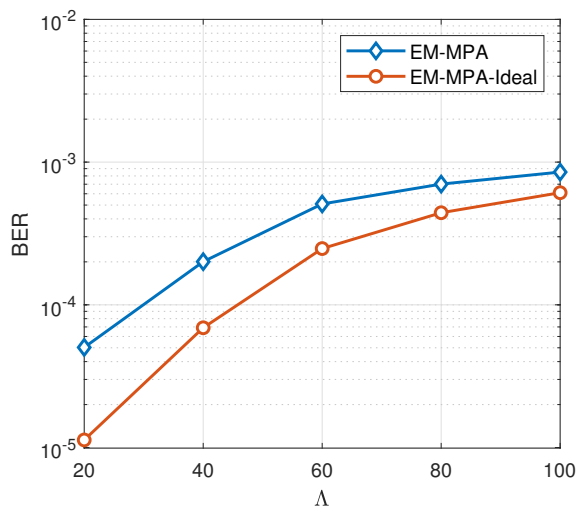


Fig. 11. Impact of number of active users on BER performance.

depicted as a performance upper bound. It is interesting to see that when Λ increases, the proposed algorithm is approaching the bound, because user identification becomes more accurate for more active users. A higher Λ results in more active users to be supported by the FTN-NOMA system. Hence, the IUI becomes more severe, degrading the BER performance, which can be observed for both the proposed method and the EM-MPA-Ideal method. We can also observe that the proposed algorithm has a modest performance erosion compared to the EM-MPA-Ideal method, which indicates that user-activity can be accurately identified.

VI. CONCLUSIONS

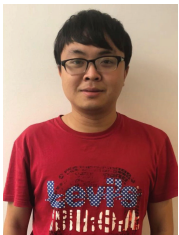
In this paper, an FTN-NOMA system relying on random access was presented, which can support massive connectivity and high throughput in machine-type communications. Dynamically fluctuating environments were considered, where both the user states and channel coefficients varied with

time. By employing an AMRA process for modeling the colored noise, we formulated the factor graph representation of the system considered based on the factorization of the *a posteriori distribution*. Then a new expectation maximization - message passing algorithm combination was proposed for joint FTN symbol detection, channel estimation and user-activity tracking. It was shown that the complexity of the proposed receiver increases linearly with the number of active users, which is significantly reduced compared to that of the conventional message passing receiver. Our simulation results demonstrated the compelling benefits of the FTN-NOMA system considered and of the proposed algorithm. The proposed method advocated can accurately identify active users and decode the information bits.

REFERENCES

- [1] C. X. Mavroumoustakis, G. Mastorakis, and J. M. Batalla, *Internet of Things (IoT) in 5G mobile technologies*. Springer, 2016, vol. 8.
- [2] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of things in the 5G era: Enablers, architecture, and business models," *IEEE J. Sel. Areas. Commun.*, vol. 34, no. 3, pp. 510–527, Mar. 2016.
- [3] M. Alioto, *Enabling the Internet of Things: From Integrated Circuits to Integrated Systems*. Springer, 2017.
- [4] H. G. Myung, J. Lim, and D. J. Goodman, "Single carrier FDMA for uplink wireless transmission," *IEEE Veh. Technol. Mag.*, vol. 1, no. 3, pp. 30–38, May 2006.
- [5] H. Holma and A. Toskala, *LTE for UMTS: OFDMA and SC-FDMA based radio access*. John Wiley & Sons, 2009.
- [6] 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, Sep. 2015.
- [7] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Mar. 2017.
- [8] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, "On the performance gain of NOMA over OMA in uplink communication systems," *IEEE Trans. Commun.*, pp. 1–1, early access, 2019.
- [9] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys & Tutor.*, vol. 20, no. 3, pp. 2294–2323, 2018.
- [10] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys & Tutor.*, vol. 19, no. 2, pp. 721–742, Feb. 2016.
- [11] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017.
- [12] S.-Y. Lien, S.-L. Shieh, Y. Huang, B. Su, Y.-L. Hsu, and H.-Y. Wei, "5G new radio: Waveform, frame structure, multiple access, and initial access," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 64–71, Jun. 2017.
- [13] Z. Wei, L. Zhao, J. Guo, D. W. K. Ng, and J. Yuan, "Multi-beam NOMA for hybrid mmwave systems," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1705–1719, Feb. 2019.
- [14] M. Moltafet, N. M. Yamchi, M. R. Javan, and P. Azmi, "Comparison study between PD-NOMA and SCMA," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1830–1834, Feb. 2017.
- [15] 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.
- [16] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. 2013 IEEE Int. Symp. Person., Indoor, Mobile Radio Commun. (PIMRC)*. IEEE, 2013, pp. 332–336.
- [17] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple access—a novel nonorthogonal multiple access for Fifth-Generation radio networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3185–3196, Apr. 2017.
- [18] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for Internet of things," in *Proc. IEEE 83rd Veh. Technol. Conf.*, May 2016, pp. 1–5.
- [19] J. E. Mazo, "Faster-than-Nyquist signaling," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1451–1462, Aug. 1975.
- [20] J. B. Anderson, F. Rusek, and V. Öwall, "Faster-than-Nyquist signaling," *Proc. of the IEEE*, vol. 101, no. 8, pp. 1817–1830, Aug. 2013.
- [21] A. Prlija 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, Sep. 2012.
- [22] S. Sugiura, "Frequency-domain equalization of faster-than-Nyquist signaling," *IEEE Wireless Commun. Lett.*, vol. 2, no. 5, pp. 555–558, Sep. 2013.
- [23] 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 2014.
- [24] 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. Netw. Conf. (WCNC)*. IEEE, Apr. 2014, pp. 642–647.
- [25] B. Ling, C. Dong, J. Dai, and J. Lin, "Multiple decision aided successive interference cancellation receiver for NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 498–501, Nov. 2017.
- [26] 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, Nov. 2015.
- [27] 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.
- [28] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and C. Huang, "Gaussian message passing for overloaded massive MIMO-NOMA," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 210–226, Jan. 2018.
- [29] 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, May 2018.
- [30] W. Yuan, N. Wu, C. Yan, Y. Li, X. Huang, and L. Hanzo, "A low-complexity energy-minimization-based SCMA detector and its convergence analysis," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12398–12403, Dec. 2018.
- [31] Z. Sun, Z. Wei, L. Yang, J. Yuan, X. Cheng, and L. Wan, "Exploiting transmission control for joint user identification and channel estimation in massive connectivity," *IEEE Transactions on Communications*, vol. 67, no. 9, pp. 6311–6326, Sep. 2019.
- [32] J. Choi, "NOMA-based random access with multichannel ALOHA," *IEEE J. Sel. Areas. Commun.*, vol. 35, no. 12, pp. 2736–2743, Dec. 2017.
- [33] J.-P. Hong, W. Choi, and B. D. Rao, "Sparsity controlled random multiple access with compressed sensing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 998–1010, Feb. 2014.
- [34] 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.
- [35] C. Wei, H. Liu, Z. Zhang, J. Dang, and L. Wu, "Near-optimum sparse channel estimation based on least squares and approximate message passing," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 754–757, Nov. 2017.
- [36] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*. John Wiley & Sons, 2005.
- [37] K. Yen and L. Hanzo, "Genetic algorithm assisted joint multiuser symbol detection and fading channel estimation for synchronous CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 6, pp. 985–998, 2001.
- [38] M. Jiang, J. Akhtman, and L. Hanzo, "Iterative joint channel estimation and multi-user detection for multiple-antenna aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2904–2914, 2007.
- [39] P. Sun, Z. Wang, and P. Schniter, "Joint channel-estimation and equalization of single-carrier systems via bilinear AMP," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2772–2785, Oct. 2018.
- [40] 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.
- [41] 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, Feb. 2018.

- [42] Y. Zhang, Q. Guo, Z. Wang, J. Xi, and N. Wu, "Block sparse Bayesian learning based joint user activity detection and channel estimation for grant-free NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9631–9640, Oct. 2018.
- [43] 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, Jan. 2018.
- [44] B. C. Levy, *Principles of signal detection and parameter estimation*. Springer Science & Business Media, 2008.
- [45] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, Jan. 1977.
- [46] R. Murray-Smith and A. Girard, "Gaussian process priors with ARMA noise models," in *Irish Signals. Sys. Conf.*, 2001, pp. 147–152.
- [47] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Sys.*, vol. 11, no. 4, pp. 360–393, Apr. 1963.
- [48] C. Kominakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.
- [49] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 2004.
- [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, Jan. 2015.
- [51] 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.
- [52] E. Heo, N. Kim, and H. Park, "Sparse structure-based channel estimation for uplink SCMA system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8037–8046, Sep. 2017.



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.



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.



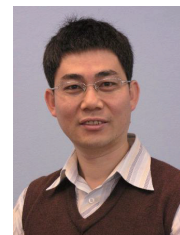
Qinghua Guo (S'07–M'08–SM'18) received the B.E. degree in electronic engineering and the M.E. degree in signal and information processing from Xidian University, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Kowloon, Hong Kong, in 2008. He is currently an Associate Professor with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW, Australia, and an Adjunct Associate Professor with

the School of Engineering, The University of Western Australia, Perth, WA, Australia. His research interests include signal processing and telecommunications. He was a recipient of the Australian Research Council's inaugural Discovery Early Career Researcher Award in 2012.



Derrick Wing Kwan Ng (S'06–M'12–SM'17) received the bachelor degree with first-class honors and the Master of Philosophy (M.Phil.) degree in electronic engineering from the Hong Kong University of Science and Technology (HKUST) in 2006 and 2008, respectively. He received his Ph.D. degree from the University of British Columbia (UBC) in 2012. He was a senior postdoctoral fellow at the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany. He is now working as a Senior

Lecturer and a Scientia Fellow at the University of New South Wales, Sydney, Australia. His research interests include convex and non-convex optimization, physical layer security, wireless information and power transfer, and green (energy-efficient) wireless communications. Dr. Ng received the Best Paper Awards at the IEEE TCGCC Best Journal Paper Award 2018, INISCOM 2018, IEEE International Conference on Communications (ICC) 2018, IEEE International Conference on Computing, Networking and Communications (ICNC) 2016, IEEE Wireless Communications and Networking Conference (WCNC) 2012, the IEEE Global Telecommunication Conference (Globecom) 2011, and the IEEE Third International Conference on Communications and Networking in China 2008. He has been serving as an editorial assistant to the Editor-in-Chief of the IEEE Transactions on Communications from Jan. 2012 to Dec. 2018. In addition, he is listed as a Highly Cited Researcher by Clarivate Analytics in 2018 and 2019.



Jinhong Yuan (M'02–SM'11–F'16) received the B.E. and Ph.D. degrees in electronics engineering from the Beijing Institute of Technology, Beijing, China, in 1991 and 1997, respectively. From 1997 to 1999, he was a Research Fellow with the School of Electrical Engineering, University of Sydney, Sydney, Australia. In 2000, he joined the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia, where he is currently a Professor and Head of Telecommunication Group with the School. He has

published two books, five book chapters, over 300 papers in telecommunications journals and conference proceedings, and 50 industrial reports. He is a co-inventor of one patent on MIMO systems and two patents on low-density-parity-check codes. He has co-authored four Best Paper Awards and one Best Poster Award, including the Best Paper Award from the IEEE International Conference on Communications, Kansas City, USA, in 2018, the Best Paper Award from IEEE Wireless Communications and Networking Conference, Cancun, Mexico, in 2011, and the Best Paper Award from the IEEE International Symposium on Wireless Communications Systems, Trondheim, Norway, in 2007. He is an IEEE Fellow and currently serving as an Associate Editor for the IEEE Transactions on Wireless Communications. He served as the IEEE NSW Chapter Chair of Joint Communications/Signal Processions/Ocean Engineering Chapter during 2011–2014 and served as an Associate Editor for the IEEE Transactions on Communications during 2012–2017. His current research interests include error control coding and information theory, communication theory, and wireless communications.



Lajos Hanzo (M'91–SM'92–F'04) (https://en.wikipedia.org/wiki/Lajos_Hanzo) received his 5-year Master degree in Electronics in 1976 and his doctoral degree in 1983. In 2004 the University of Southampton bestowed the Doctor of Science (DSc) higher research degree upon him. During his 44-year career in Telecommunications, he has held various research and academic positions in Hungary, Germany, and U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, U.K., where he holds

the Chair of Telecommunications. He has successfully supervised 119 Ph.D. students, coauthored 18 John Wiley/IEEE Press books on mobile radio communications, totaling in excess of 10 000 pages, published 1329 research papers at IEEE Xplore, acted as TPC and the General Chair of IEEE conferences, presented keynote lectures, and has been awarded a number of

distinctions. He is currently directing an 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), U.K., the European Research Council's Advanced Fellow Grant, and the Royal Society, UK. He is an enthusiastic supporter of industrial and academic liaison, and he offers a range of industrial courses. He is a Fellow of the Royal Academy of Engineering (FREng), Fellow of the Institute of Engineering & Technology (FIET), Fellow of the European Signal Processing Association (EURASIP). In 2009, he received an Honorary Doctorate from the Technical University of Budapest, and again in 2015 by from University of Edinburgh. From 2008 to 2012, he was the Editor-in-Chief of the IEEE Press and a Chaired Professor at Tsinghua University, Beijing. His research is funded by the European Research Council's Advanced Fellow Grant. He is a Governor of the IEEE VTS.