AMP-based Joint Activity Detection and Channel Estimation for OFDM-based Grant-Free Access

Zhiyan Li, Ying Cui, and Danny H.K. Tsang
IoT Thrust, HKUST(GZ), Guangzhou, China

***Abstract*—To realize orthogonal frequency division multiplexing (OFDM)-based grant-free access for wide-band systems under frequency-selective fading, the existing device activity detection and channel estimation methods need substantial accuracy improvement or computation time reduction. In this paper, we aim to resolve this issue. First, we present an exact time domain signal model for OFDM-based grant-free access under frequency-selective fading. Then, we build a new factor graph and propose an approximate message passing (AMP) algorithm for joint device activity detection and channel estimation, unleashing AMP's full potential. Next, we analyze the state evolution of the AMP algorithm and the error probability and mean square error (MSE) of the estimated device activities and active devices' channels, respectively. Finally, numerical results validate the theoretical results and show the superior performance of the proposed AMP algorithm.**

I. Introduction

As the global IoT device count continues to surge, massive machine-type communication (mMTC) has emerged as one of primary use cases for both 5G and B5G wireless networks. However, supporting a large number of energy-limited and intermittently active IoT devices transmitting small packets faces significant challenges. Grant-free access has been proposed as a solution to this issue, with device activity detection and channel estimation being critical components.

Most existing work investigates activity detection and channel estimation for narrow band systems under flat fading [2]–[5]. For example, the block coordinate descent (BCD) method is proposed to obtain stationary points of the maximum likelihood estimation (MLE) [4] and maximum a posterior estimation (MAPE) [5] problems for device activities. The MLE-based and MAPE-based device activity detection methods via BCD achieve high detection accuracy at the cost of increased computational complexity. Given very accurate device activity detection results, the channel estimations of active devices are readily obtained via classic channel estimation methods. The approximate message passing (AMP) method is proposed for joint activity detection and channel estimation by approximately computing the minimum mean square error (MMSE) estimation of the effective channels of all devices with low computational complexity at the sacrifice of accuracy [2], [3].

Notably, the abovementioned state-of-the-art methods for narrow-band systems with flat fading are no longer suitable for wide-band systems under frequency-selective fading. Orthogonal frequency division multiplexing (OFDM) is widely adopted for wide-band systems to cope with channel-frequency selectivity. Activity detection and channel estimation for OFDM-based grant-free access for wide-band systems under frequency-selective fading have drawn increasing attention [6]–[9]. For instance, in [6], leveraging an exact time domain signal model under the assumption that the pilot length and the number of subcarriers are identical, the authors propose MLE and MAPE-based device activity detection methods that achieve superior detection accuracy with very high computational complexity. Besides, in [7]–[9], by approximating the frequency domain signal model for simplicity, the authors propose approximate versions of the message passing (MP) methods [7], [8] and AMP method [9] for joint device activity detection and channel estimation. MP and its variations have higher computational complexity than AMP. Although the MP and AMP-based methods in [7]–[9] successively reduce the computational complexity, they yield significant accuracy loss, mainly due to the approximation of the complex channel model under OFDM and frequency selective fading.

Four primary limitations exist in the existing work for OFDM-based grant-free access under frequency selective fading. Firstly, the existing AMP algorithms [2], [3], [9] return the estimates of the effective channels that may not be the best among all iterates, limiting the potential accuracy of AMP. Secondly, the AMP algorithm in [9] incurs additional accuracy loss due to the channel model approximation. Thirdly, the exact time domain signal model in [6] is valid only for the case of equal pilot length and number of subcarriers, and the MLE and MAPE-based methods are computationally expensive for practical systems. Finally, the performance analysis of activity detection and channel estimation for OFDM-based grant-free access under frequency selective fading remains unresolved.

In this paper, we would like to address the above limitations. Specifically, we investigate the joint device activity detection and channel estimation for OFDM-based grant-free access in a wide-band system under frequency selective fading. First, we extend the exact time domain signal model in [6] to incorporate more flexible setups for the pilot length and number of subcarriers. Then, we build a new factor graph and propose an AMP algorithm for joint device activity detection and channel estimation. Unlike those in [7]–[9], the proposed AMP algorithm precisely captures the dependency of the actual frequency domain channels and selects the best estimate among the iterates, thereby unleashing the full potential of AMP. Besides, in contrast to [3], [7]–[9], the proposed AMP algorithm successfully avoids computing matrix inversions, thereby reducing the computational complexity for updating the estimate and residual in each iteration. What's more, we

explicitly show that AMP approximately obtains the MAP-based activity detection for the first time. Next, we analyze the state evolution of the proposed AMP algorithm and the error probability and mean square error (MSE) of the estimated device activities and channels of active devices, respectively. Finally, numerical results verify the theoretical analysis and demonstrate the superior performance of the proposed AMP algorithm in accuracy and computation time.

Notation: Use bold uppercase (e.g., $\mathbf{X}$) to denote matrices. $\mathbf{X}_{i,m:n}$ denotes a row vector consisting of the elements from columns m to n in the i-th row of matrix $\mathbf{X}$, $\mathbf{X}_{i:j,m}$ denote a column vector consisting of the elements from rows i to j in the m-th column, and $\mathbf{X}_{i:j,m:n}$ denote a block containing the elements from rows i to j and columns m to n. $\mathbf{I}$ denotes the unit matrix. Use calligraphy uppercase letters (e.g., $\mathcal{N}$) to denote sets. The transpose and conjugate transpose operators are denoted by $(\cdot)^T$ and $(\cdot)^H$, respectively. Use $\mathrm{tr}(\cdot)$ to denote the trace. Use $f_{\mathcal{CN}}(x;a,b)$ to denote the complex Gaussian probability density function (PDF) with mean a and variance b. Use $\delta(\cdot)$ to denote the Delta function.

II. System model

We study a single-cell cellular network which consists of one M-antenna BS and N single-antenna IoT devices and operates on a wide band. Let $\mathcal{M} \triangleq \{1,2,\cdots,M\}$ and $\mathcal{N} \triangleq \{1,2,\cdots,N\}$. The wireless channel is characterized by large-scale fading and small-scale fading. For all $n \in \mathcal{N}$, let $\beta_n > 0$ denote the large-scale fading power of the channel between device n and the BS. Suppose that $\beta_n, n \in \mathcal{N}$ are known to the BS [3], [6]. We consider the block fading and frequency-selective fading models for the small-scale fading with P channel taps. Denote $\mathcal{P} \triangleq \{1,2,\cdots,P\}$. For all $n \in \mathcal{N}, p \in \mathcal{P}$, and $m \in \mathcal{M}$, let $g_{n,p,m}$ denote the small-scale fading coefficient of the p-th tap of the channel between device n and the BS's antenna m. Suppose that the small-scale fading coefficients $g_{n,p,m}, n \in \mathcal{N}, p \in \mathcal{P}, m \in \mathcal{M}$ are unknown to the BS and follow i.i.d. $\mathcal{CN}(0,1)$. For all $n \in \mathcal{N}, p \in \mathcal{P}$, and $m \in \mathcal{M}$, the overall wireless channel between device n and the BS's antenna m can be represented as $h_{n,p,m} = \sqrt{\beta_n} g_{n,p,m} \in \mathbb{C}$.

We investigate the massive access scenario stemming from mMTC. In particularly, only a small fraction of devices among massive potential ones activate and access the BS within each coherence block. For all $n \in \mathcal{N}$, $a_n \in \{0,1\}$ denotes the activity state of device n, where $a_n = 1$ indicates that device n is active and $a_n = 0$ otherwise. In the considered scenario, $\sum_{n\in\mathcal{N}} a_n = N_a \ll N$, i.e., $\boldsymbol{a} \triangleq (a_n)_{n\in\mathcal{N}} \in \{0,1\}^N$ is sparse. As in [3], suppose that $a_n, n \in \mathcal{N}$ are i.i.d. Bernoulli random variables with probabilities $\rho_n \in (0,1), n \in \mathcal{N}$. For ease of exposition, suppose that $\rho_n, n \in \mathcal{N}$ are known to the BS.[1] The probability mass function (PMF) of $\boldsymbol{a}$ is given by:

$$p(\boldsymbol{a}) = \prod_{n=1}^{N} p(a_n) = \prod_{n=1}^{N} \rho_n^{a_n}(1-\rho_n)^{1-a_n}. \quad (1)$$

[1]The proposed method can be readily extended to the case of unknown $\beta_n, \rho_n, n \in \mathcal{N}$ leveraging the expectation maximization method as in [9].

For all $n \in \mathcal{N}, m \in \mathcal{M}$, and $p \in \mathcal{P}$, let $x_{n,p,m} = a_n h_{n,p,m}$ represent the p-th effective channel between device n and the BS's antenna m. By (1) and $h_{n,p,m} \sim \mathcal{CN}(0,\beta_n)$, we have:

$$p(x_{n,p,m}|a_n) = a_n f_{\mathcal{CN}}(x_{n,p,m};0,\beta_n) + (1-a_n)\delta(x_{n,p,m}).$$

For all $n \in \mathcal{N}$, $x_{n,p,m} = a_n h_{n,p,m}, p \in \mathcal{P}, m \in \mathcal{M}$ are unconditionally dependent but independently conditioned on a_n. By noting that a_n and $h_{n,p,m}$ are unknown, we will estimate active devices' channel conditions from the estimate of $x_{n,p,m}, p \in P, m \in M$, which will be illustrated shortly.

We consider an OFDM-based massive grant-free access scheme to prompt efficient uplink transmission over the wide-band system in an arrive-and-go manner [6]. Denote K as the number of subcarriers, and $\mathcal{K} \triangleq \{1,2,\cdots,K\}$. Each device $n \in \mathcal{N}$ is allocated a unique length-L pilot sequence. Suppose $L = KQ \ll N$, for some $T \in \mathbb{N}_+$. Denote $\mathcal{L} \triangleq \{1,2,\cdots,L\}$ and $\mathcal{Q} \triangleq \{1,2,\cdots,Q\}$. For all $n \in \mathcal{N}$, device n inserts its L pilot symbols into the K subcarriers, forming Q OFDM pilot symbols, denoted by $\tilde{\mathbf{s}}_{q,n} \in \mathbb{C}^{K\times 1}, q \in \mathcal{Q}$. For all $q \in \mathcal{Q}$ and $n \in \mathcal{N}$, the time domain representation of $\tilde{\mathbf{s}}_{q,n}$ is given by its normalized inverse discrete Fourier transform (IDFT):

$$\mathbf{s}_{q,n} = \mathbf{F}^H \tilde{\mathbf{s}}_{q,n} \in \mathbb{C}^{K\times 1}, \quad q \in \mathcal{Q}, n \in \mathcal{N}. \quad (2)$$

Here, unitary matrix $\mathbf{F} \triangleq (F_{\ell,\ell'})_{\ell,\ell'\in\mathcal{L}} \in \mathbb{C}^{L\times L}$ represents the normalized discrete Fourier transform (DFT) matrix, where $F_{\ell,\ell'} \triangleq \frac{1}{\sqrt{L}} e^{-\frac{j2\pi(\ell-1)(\ell'-1)}{L}}$. Each device $n \in \mathcal{N}$ appends a cyclic prefix of no smaller than $P-1$ to $\mathbf{s}_{q,n}$, for all $q \in \mathcal{Q}$, and transmits the Q OFDM pilot symbols with cyclic prefixes. After removing the respective cyclic prefixes, the received signal for the q-th OFDM pilot symbol at antenna $m \in \mathcal{M}$ of the BS, $\mathbf{y}_{q,m} \triangleq (y_{q,k,m})_{k\in\mathcal{K}} \in \mathbb{C}^K$, can be written as: [2]

$$\mathbf{y}_{q,m} = \sum_{n\in\mathcal{N}} a_n \mathbf{H}_{n,m} \mathbf{F}^H \tilde{\mathbf{s}}_{q,n} + \mathbf{n}_{q,m}, q \in \mathcal{Q}, m \in \mathcal{M}, \quad (3)$$

where

$$\mathbf{H}_{n,m} \triangleq \begin{bmatrix} h_{n,1,m} & h_{n,K,m} & \cdots & h_{n,2,m} \\ h_{n,2,m} & h_{n,1,m} & \cdots & h_{n,3,m} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n,K,m} & h_{n,K-1,m} & \cdots & h_{n,1,m} \end{bmatrix} \in \mathbb{C}^{K\times K},$$

and $\mathbf{n}_{q,m} \triangleq (n_{q,k,m})_{k\in\mathcal{K}} \in \mathbb{C}^K$ is the additive white Gaussian noise (AWGN) with $n_{q,k,m}, q \in \mathcal{Q}, k \in \mathcal{K}, m \in \mathcal{M}$ following i.i.d. $\mathcal{CN}(0,\sigma^2)$. Here, for notation convenience, we let $h_{n,k,m} = 0$, $k \in \mathcal{K}\backslash\mathcal{P}, n \in \mathcal{N}$, $m \in \mathcal{M}$. For all $n \in \mathcal{N}, m \in \mathcal{M}$, $\mathbf{H}_{n,m}$ is a circulant matrix that is specified by the first column or row vector.

To facilitate device activity detection and channel estimation, we present an equivalent model for the received signals. By left-multiplying a DFT matrix to $\mathbf{y}_{q,m}$ in (3), we have:

$$\tilde{\mathbf{y}}_{q,m} = \mathbf{F}\mathbf{y}_{q,m} = \sum_{n\in\mathcal{N}} a_n \mathbf{F}\mathbf{H}_{n,m}\mathbf{F}^H \tilde{\mathbf{s}}_{q,n} + \mathbf{F}\mathbf{n}_{q,m}$$

$$= \sum_{n\in\mathcal{N}} a_n \mathrm{diag}(\tilde{\mathbf{s}}_{q,n}) \mathbf{F} (\mathbf{H}_{n,m})_{:,1} + \tilde{\mathbf{n}}_{q,m}, q \in \mathcal{Q}, m \in \mathcal{M},$$

where the last equality is because $\mathbf{F}\mathbf{H}_{n,m}\mathbf{F}^H =$

[2]Throughout this paper, we assume that subcarrier spacing, symbol time, and cyclic prefix are appropriately chosen such that inter-symbol interference and inter-carrier interference can be avoided.

$\text{diag}(\mathbf{F}(\mathbf{H}_{n,m})_{:,1}) \in \mathbb{C}^{K\times K}$ is a diagonal matrix. Then, by left-multiplying the normalized IDFT matrix to $\tilde{\mathbf{y}}_{q,m}$, we rewrite $\mathbf{y}_{q,m}$ in (3) as:

$$\mathbf{y}_{q,m} = \mathbf{F}^H \tilde{\mathbf{y}}_{q,m} = \sum_{n\in\mathcal{N}} a_n \tilde{\mathbf{A}}_{q,n} \mathbf{h}_{n,m} + \mathbf{n}_{q,m}, q \in \mathcal{Q}, m \in \mathcal{M},$$

where $\tilde{\mathbf{A}}_{q,n} \triangleq (\mathbf{F}^H \text{diag}(\tilde{\mathbf{s}}_{q,n})\mathbf{F})_{:,1:P} \in \mathbb{C}^{K\times P}$, $\mathbf{h}_{n,m} \triangleq (h_{n,p,m})_{p\in\mathcal{P}} \in \mathbb{C}^P$, and the last equality is due to $\mathbf{F}^H\mathbf{F} = \mathbf{I}_K$ and $h_{n,k,m} = 0$, $k \in \mathcal{K}\backslash\mathcal{P}, n \in \mathcal{N}$, $m \in \mathcal{M}$. Define $\mathbf{Y} \in \mathbb{C}^{L\times M}$ with $\mathbf{Y}_{(t-1)K+1:tK,m} = \mathbf{y}_{q,m}$, $\mathbf{N} \in \mathbb{C}^{L\times M}$ with $\mathbf{N}_{(t-1)K+1:tK,m} = \mathbf{n}_{q,m}$, $\mathbf{A} \in \mathbb{C}^{L\times NP}$ with $\mathbf{A}_{(t-1)K+1:tK,(n-1)P+1:nP} = \tilde{\mathbf{A}}_{q,n}$, and $\mathbf{X} \in \mathbb{C}^{NP\times M}$ with $\mathbf{X}_{(n-1)P+1:nP,m} = a_n \mathbf{h}_{n,m}$. Notice that $x_{n,p,m}$ is the $((n-1)P+p,m)$-th element of $\mathbf{X}$. More compactly, the received signals for the Q OFDM symbols over the M antennas at the BS can be expressed as $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N}$.

Remark 1 (Comparisons for Signal Models): The considered signal model for pilot length $L = KQ$ under frequency-selective fading is more complex than those for pilot length L under flat fading [3] and pilot length $L = K$ under frequency-selective fading [6]. Specifically, the considered signal model degrades to the one in [3] when $P = 1$ and $K = 1$ and to the one in [6] when $Q = 1$.

III. MAP Detection, MMSE Estimation, and AMP Algorithm

A. MAP Detection and MMSE Estimation

One can apply the MAP detector for device activity detection, which maximizes the a posterior probability $p(\mathbf{a}|\mathbf{Y})$. The MAP detection problem is formulated as follows:

$$\max_{\hat{\mathbf{a}}(\mathbf{Y})\in\{0,1\}^N} p(\hat{\mathbf{a}}(\mathbf{Y})|\mathbf{Y}), \tag{4}$$

where $\hat{\mathbf{a}}(\mathbf{Y}) = (\hat{a}_n(\mathbf{Y}))_{n\in\mathcal{N}}$ represents the detector of $\mathbf{a}$. Since the problem in (4) is NP-hard, we approximate it with N simpler scalar optimization problems, for $\hat{a}_n(\mathbf{Y}), n \in \mathcal{N}$, respectively, i.e., $\max_{a_n\in\{0,1\}} p(a_n|\mathbf{Y})$. For all $n \in \mathcal{N}$, the optimal solution of the scalar optimization problem, denoted by $\hat{a}_n^*(\mathbf{Y})$, is given by:

$$\hat{a}_n^*(\mathbf{Y}) \triangleq \begin{cases} 0, & \theta_n < 0, \\ 1, & \theta_n \geq 0, \end{cases} \tag{5}$$

where

$$\theta_n \triangleq \log\left(\frac{p(a_n = 1|\mathbf{Y})}{p(a_n = 0|\mathbf{Y})}\right). \tag{6}$$

To calculate $\hat{a}_n^*(\mathbf{Y})$ given in (5), we need to first obtain $p(a_n|\mathbf{Y})$ which determines the threshold θ_n.

To perform channel estimation, we apply the Bayesian MMSE estimator for the effective channels $\mathbf{X}$, which minimizes the Bayesian MSE. Notice that for all $n \in N$ with $a_n = 1$,we have $h_{n,p,m} = x_{n,p,m}, p \in \mathcal{P}, m \in \mathcal{M}$. The MMSE estimation problem of $\mathbf{X}$ is formulated as follows:

$$\min_{\hat{\mathbf{X}}(\mathbf{Y})} \mathbb{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}(\mathbf{Y})\|_F^2|\mathbf{Y}\right], \tag{7}$$

where $\hat{\mathbf{X}} = (\hat{x}_{n,p,m}(\mathbf{Y}))_{n\in\mathcal{N},p\in\mathcal{P},m\in\mathcal{M}}$ represents the estimator of $\mathbf{X}$. Let $\hat{\mathbf{X}}^*(\mathbf{Y}) \triangleq (\hat{x}^*_{n,p,m}(\mathbf{Y}))_{n\in\mathcal{N},p\in\mathcal{P},m\in\mathcal{M}}$ denote the optimal solution of the problem in (7). The problem in (7) can be equivalently separated to NPM scalar optimization problems, for $\hat{x}_{n,p,m}(\mathbf{Y}), n \in \mathcal{N}, p \in \mathcal{P}, m \in \mathcal{M}$, respectively, that can be solved in parallel with the optimal solutions:

$$\hat{x}^*_{n,p,m}(\mathbf{Y}) \triangleq \int x_{n,p,m} p(x_{n,p,m}|\mathbf{Y}) dx_{n,p,m}. \tag{8}$$

To calculate $\hat{x}^*_{n,p,m}(\mathbf{Y})$, we need to first obtain $p(x_{n,p,m}|\mathbf{Y})$.

Lemma 1: For all $n \in \mathcal{N}, p \in \mathcal{P}, m \in \mathcal{M}$, we have

$$p(a_n|\mathbf{Y}) = \frac{\sum_{\mathbf{a}_{\bar{n}}} \int p(\mathbf{X},\mathbf{a},\mathbf{Y}) d\mathbf{X}}{\sum_{\mathbf{a}} \int p(\mathbf{X},\mathbf{a},\mathbf{Y}) d\mathbf{X}}, \tag{9}$$

$$p(x_{n,p,m}|\mathbf{Y}) = \frac{\sum_{\mathbf{a}} \int p(\mathbf{X},\mathbf{a},\mathbf{Y}) d\mathbf{X}_{\overline{n,p,m}}}{\int \sum_{\mathbf{a}} \int p(\mathbf{X},\mathbf{a},\mathbf{Y}) d\mathbf{X}_{\overline{n,p,m}} dx_{n,p,m}}, \tag{10}$$

where

$$p(\mathbf{X},\mathbf{a},\mathbf{Y}) = \prod_{l=1}^{L}\prod_{m=1}^{M} p(y_{l,m}|\mathbf{X}_{:,m}) \prod_{n=1}^{N} p(a_n) \prod_{m=1}^{M}\prod_{p=1}^{P} p(x_{n,p,m}|a_n). \tag{11}$$

Here, $p(y_{l,m}|\mathbf{X}_{:,m}) = f_{\mathcal{CN}}(y_{l,m}; \mathbf{A}_{l,:}\mathbf{X}_{:,m}, \sigma^2)$.

Lemma 1 indicates that multiple integrals have to be computed to successfully obtain $p(a_n|\mathbf{Y})$ in (9) and $p(x_{n,p,m}|\mathbf{Y})$ in (10) before calculating $\hat{a}_n^*(\mathbf{Y})$ in (5) and $\hat{x}^*_{n,p,m}(\mathbf{Y})$ in (8), respectively. The computation cost for the multiple integrals is significant, especially for large N and M, rendering the exact computation of $\hat{a}_n^*(\mathbf{Y})$ and $\hat{x}^*_{n,p,m}(\mathbf{Y})$ nearly impractical.

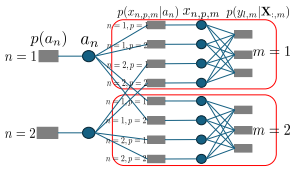


Fig. 1: Illustration of the factor graph of $p(\mathbf{X},\mathbf{a},\mathbf{Y})$ in (11).

TABLE I: Message definitions.

$\mu^{(t)}_{f_{l,m}\to x_{n,p,m}}(x_{n,p,m})$	message from $p(y_{l,m}\|\mathbf{X}_{:,m})$ to $x_{n,p,m}$
$\mu^{(t)}_{x_{n,p,m}\to f_{l,m}}(x_{n,p,m})$	message from $x_{n,p,m}$ to $p(y_{l,m}\|\mathbf{X}_{:,m})$
$\mu^{(t)}_{x_{n,p,m}\to q_{n,p,m}}(x_{n,p,m})$	message from $x_{n,p,m}$ to $p(x_{np,m}\|a_n)$
$\mu^{(t)}_{q_{n,p,m}\to x_{n,p,m}}(x_{n,p,m})$	message from $p(x_{n,p,m}\|a_n)$ to $x_{n,p,m}$
$\mu^{(t)}_{a_n\to q_{n,p,m}}(a_n)$	message from a_n to $p(x_{n,p,m}\|a_n)$
$\mu^{(t)}_{q_{n,p,m}\to a_n}(a_n)$	message from $p(x_{n,p,m}\|a_n)$ to a_n

B. AMP Algorithm

To tackle the aforementioned challenge, we propose an AMP algorithm to approximately compute $\hat{a}_n^*(\mathbf{Y})$ and $\hat{x}^*_{n,p,m}(\mathbf{Y})$ with low computational complexity. Firstly, we employ the MP method to compute the multiple integrals in (9) and (10) in a parallel and distributed manner. Specifically, we first graphically represent the same factorable integrand of the multiple integrals in (9) and (10), i.e., $p(\mathbf{X},\mathbf{a},\mathbf{Y})$, with a factor graph according to (11), as illustrated in Fig. 1. Then, we iteratively compute the exchanging messages between each variable node (circle) and each function node (rectangle) in the factor graph. The definitions of the messages are given in Table I. We first initialize $\mu^{(0)}_{x_{n,p,m}\to f_{l,m}}(x_{n,p,m})$,

$\mu^{(0)}_{x_{n,p,m}\to q_{n,p,m}}(x_{n,p,m})$, and $\mu^{(0)}_{a_n\to q_{n,p,m}}(a_n)$. At each iteration t, we update the messages according to:

$$\mu^{(t)}_{f_{l,m}\to x_{n,p,m}}(x_{n,p,m}) \propto \int p(y_{l,m}|\mathbf{X}_{:,m}) \times \prod_{(i-1)\times P+j\neq(n-1)\times P+p}^{(N-1)\times P+P} \mu^{(t)}_{x_{i,j,m}\to f_{l,m}}(x_{i,j,m}) d\mathbf{X}_{:,m\setminus x_{n,p,m}}, \tag{12}$$

$$\mu^{(t)}_{q_{n,p,m}\to x_{n,p,m}}(x_{n,p,m}) \propto \sum_{a_n=0}^{1} \mu^{(t)}_{a_n\to q_{n,p,m}}(a_n) p(x_{n,p,m}|a_n), \tag{13}$$

$$\mu^{(t)}_{q_{n,p,m}\to a_n}(a_n) \propto \int p(x_{n,p,m}|a_n) \mu^{(t)}_{x_{n,p,m}\to q_{n,p,m}}(x_{n,p,m}) dx_{n,p,m}, \tag{14}$$

$$\mu^{(t+1)}_{x_{n,p,m}\to q_{n,p,m}}(x_{n,p,m}) \propto \prod_{l=1}^{L} \mu^{(t)}_{f_{l,m}\to x_{n,p,m}}(x_{n,p,m}), \tag{15}$$

$$\mu^{(t+1)}_{a_n\to q_{n,p,m}}(a_n) \propto p(a_n) \prod_{m=1}^{M}\prod_{j\neq p}^{P} \mu^{(t)}_{q_{n,j,m}\to a_n}(a_n), \tag{16}$$

$$\mu^{(t+1)}_{x_{n,p,m}\to f_{l,m}}(x_{n,p,m}) \propto \sum_{a_n=0}^{1} \mu^{(t)}_{a_n\to q_{n,p,m}}(a_n) \times p(x_{n,p,m}|a_n) \prod_{b\neq l}^{L} \mu^{(t)}_{f_{b,m}\to x_{I,p,m}}(x_{n,p,m}). \tag{17}$$

Note that the aforementioned messages represent probability measures. Therefore, the notation "$\propto$" implies that the right-hand side of it differs from the left-hand side by a normalized constant to ensure that the integral result on the left-hand side equals 1. Besides, note that the messages in (12), (13), and (14) are updated in parallel, and then the messages in (15), (16), and (17) are updated in parallel. Finally, when the updates terminate, we compute the approximations of $\hat{a}^*_n(\mathbf{Y})$ and $\hat{x}^*_{n,p,m}(\mathbf{Y})$ at the t-th iteration, which are given by:

$$\hat{a}^{(t)}_n = \begin{cases} 0, & \theta^{(t)}_n < 0, \\ 1, & \theta^{(t)}_n \geq 0, \end{cases} \tag{18}$$

$$\hat{x}^{(t)}_{n,p,m} \triangleq C \int x_{n,p,m} \mu^{(t)}_{q_{n,p,m}\to x_{n,p,m}}(x_{n,p,m}) \times \mu^{(t)}_{x_{n,p,m}\to q_{n,p,m}}(x_{n,p,m}) dx_{n,p,m}, \tag{19}$$

where

$$\theta^{(t)}_n \triangleq \log\left(\frac{\rho_n \prod_{m=1}^{M}\prod_{p=1}^{P} \mu^{(t)}_{q_{n,p,m}\to a_n}(a_n=1)}{(1-\rho_n)\prod_{m=1}^{M}\prod_{p=1}^{P} \mu^{(t)}_{q_{n,p,m}\to a_n}(a_n=0)} \right). \tag{20}$$

Here, C represents the normalized constant . If the message updates converge, $\hat{a}^{(t)}_n$ converges to $\hat{a}^*_n(\mathbf{Y})$ ($\theta^{(t)}_n$ converges to θ_n) and $\hat{x}^{(t)}_{n,p,m}$ converges to $\hat{x}^*_{n,p,m}(\mathbf{Y})$, as $t\to\infty$.

It is well-known that the MP method, which establishes the concept for computing multiple integrals in a parallel and distributed manner, is still not tractable. It motivates a further approximation of the MP method in the asymptotic region, i.e., $N, L \to \infty$ and $\frac{N}{L} \to c$ for some $c > 0$, leading to the AMP method. Since $\hat{a}^{(t)}_n$ can be readily obtained once $\theta^{(t)}_n$ is computed, the proposed AMP algorithm iteratively updates $\theta^{(t)}_n$ and $\hat{x}^{(t)}_{n,p,m}$. Let $A_{l,n,p}$ denote the element in l-th row and $((n-1)P+p)$-th column of $\mathbf{A}$, for all $l\in\mathcal{L}, n\in\mathcal{N}, p\in\mathcal{P}$. Moreover, the columns of $\mathbf{A}$ are assumed to have unit l_2-norm.

Theorem 1: In the asymptotic region, the MP method given by (12)-(17), $\theta^{(t)}_n$ in (20), and $\hat{x}^{(t)}_{n,p,m}$ in (19) are approximated by:

$$\theta^{(t)}_n = \log\left(\frac{\rho_n}{(1-\rho_n)} \times \frac{\prod_{m=1}^{M}\prod_{p=1}^{P} f_{\mathcal{CN}}(0; \hat{x}^{(t)}_{n,p,m} + \sum_{l=1}^{L} A_{l,n,p} z^{(t)}_{f_{l,m}}, \tau^{(t)} + \beta_n)}{\prod_{m=1}^{M}\prod_{p=1}^{P} f_{\mathcal{CN}}(0; \hat{x}^{(t)}_{n,p,m} + \sum_{l=1}^{L} A_{l,n,p} z^{(t)}_{f_{l,m}}, \tau^{(t)})} \right), \tag{21}$$

$$\hat{x}^{(t+1)}_{n,p,m} = \eta\left(\hat{x}^{(t)}_{n,p,m} + \sum_{l=1}^{L} A_{l,n,p} z^{(t)}_{f_{l,m}}, \tau^{(t)}, \theta^{(t)}_n \right), \tag{22}$$

$$z^{(t+1)}_{f_{l,m}} = y_{l,m} - \sum_{n=1}^{N}\sum_{p=1}^{P} A_{l,n,p} \hat{x}^{(t+1)}_{n,p,m} + \frac{1}{L} z^{(t)}_{f_{l,m}} \sum_{n=1}^{N}\sum_{p=1}^{P} \eta'\left(\hat{x}^{(t)}_{n,p,m} + \sum_{l=1}^{L} A_{l,n,p} z^{(t)}_{f_{l,m}}, \tau^{(t)}, \theta^{(t)}_n \right), \tag{23}$$

where $\eta'(x,y,z) \triangleq \frac{\partial \eta(x,y,z)}{\partial x}$, and

$$\eta(x,y,z) \triangleq \frac{\exp(z)\beta_n x}{(1+\exp(z))(y+\beta_n)}, \tag{24}$$

$$\tau^{(t)} \triangleq \frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M} \left| z^{(t)}_{f_{l,m}} \right|^2. \tag{25}$$

Let $\boldsymbol{\theta}^{(t)} \triangleq \left(\theta^{(t)}_n\right)_{n\in\mathcal{N}}$, $\hat{\mathbf{X}}^{(t)} \triangleq (\hat{x}^{(t)}_{n,p,m})_{n\in\mathcal{N},p\in\mathcal{P},m\in\mathcal{M}}$, and $\mathbf{Z}^{(t)} \triangleq (z^{(t)}_{f_{l,m}})_{l\in\mathcal{L},m\in\mathcal{M}}$. $\mathbf{Z}^{(t)}$ can be interpreted as the residual, and $\tau^{(t)}$ can be represented as the residual variance. All elements of $\boldsymbol{\theta}^{(t)}$, $\hat{\mathbf{X}}^{(t)}$, and $\mathbf{Z}^{(t)}$ are computed in parallel.

AMP has no convergence guarantee. In most existing work on AMP, when $|\tau^{(t)} - \tau^{(t-1)}|$ is sufficiently small, AMP stops, and the latest iterate $\hat{\mathbf{X}}^{(t)}$ is returned as the estimate of $\mathbf{X}$. Since the MSE $\mathbb{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}^{(t)}\|_F^2 | \mathbf{Y}\right]$ is not monotonically decreasing with t, $\hat{\mathbf{X}}^{(t)}$ may not be the best estimate among $\hat{\mathbf{X}}^{(1)}, ..., \hat{\mathbf{X}}^{(t)}$. This motivates us to track the best estimate obtained so far. Since $\mathbb{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}^{(t)}\|_F^2 | \mathbf{Y}\right]$ requires computing multiple integrals which is as computationally expensive as computing the MMSE estimator given in (8), we use the objective function of the GROUP LASSO problem:

$$f(\mathbf{X}) \triangleq \frac{1}{2}\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \sum_{i=1}^{NP} \|\mathbf{X}_{i,:}\|_2 \tag{26}$$

to measure the quality of $\hat{\mathbf{X}}^{(t)}$. Specifically, we keep tracking $f^{(t)}_{best} \triangleq \max\{f(\hat{\mathbf{X}}^{(1)}), ..., f(\hat{\mathbf{X}}^{(t)})\}$ and the estimate $\hat{\mathbf{X}}^{(t)}_{best} \triangleq \hat{\mathbf{X}}^{(t_{best})}$ with t_{best} satisfying $f(\hat{\mathbf{X}}^{(t_{best})}) = f^{(t)}_{best}$.

Algorithm 1 The AMP Algorithm for Estimating Effective Channels

Input: received signals $\mathbf{Y}$ and measurement matrix $\mathbf{A}$.
Output: threshold $\boldsymbol{\theta}_{best}^{(t)}$ and estimate of effective channels $\hat{\mathbf{X}}_{best}^{(t)}$
Initialize: set $\hat{\mathbf{X}}^{(0)} = \mathbf{0}$, $\hat{\mathbf{X}}_{best}^{(t)} = \hat{\mathbf{X}}^{(0)}$, $\mathbf{Z}^{(0)} = \mathbf{Y}$, $f_{best}^{(t)} = f(\hat{\mathbf{X}}^{(0)})$, $t = 0$.
repeat
1: Calculate $\tau^{(t)}$ according to (24).
2: Calculate $\theta_n^{(t)}, n \in \mathcal{N}$ in parallel according to (21).
3: Calculate $\hat{x}_{n,p,m}^{(t+1)}, n \in \mathcal{N}, p \in \mathcal{P}, m \in \mathcal{M}$ in parallel according to (22).
4: Calculate $z_{l,m}^{(t+1)}, l \in \mathcal{L}, m \in \mathcal{M}$ in parallel according to (23).
5: **If** $f(\hat{\mathbf{X}}^{(t+1)}) > f_{best}^{(t)}$
6: set $f_{best}^{(t+1)} = f(\hat{\mathbf{X}}^{(t+1)})$, $\boldsymbol{\theta}_{best}^{(t)} = \boldsymbol{\theta}^{(t)}$ and $\hat{\mathbf{X}}_{best}^{(t+1)} = \hat{\mathbf{X}}^{(t)}$.
7: **else**
8: set $f_{best}^{(t+1)} = f_{best}^{(t)}$.
9: **end**
10: set $t = t + 1$.
until some stopping criteria are satisfied.

The corresponding $\boldsymbol{\theta}_{best}^{(t)} \triangleq \boldsymbol{\theta}^{(t_{best})}$ may produce device activity detection with good accuracy. We also use $\hat{\mathbf{X}}_{best}^{(t)} \triangleq \left(\hat{x}_{n,p,m}^{best,(t)}(\mathbf{Y})\right)_{n\in\mathcal{N},p\in\mathcal{P},m\in\mathcal{M}}$ and $\boldsymbol{\theta}_{best}^{(t)} \triangleq \left(\theta_n^{best,(t)}\right)_{n\in\mathcal{N}}$. This procedure can successfully boost the accuracy of AMP, which will be shown in Section V. When some stopping criteria are satisfied, e.g., the preset number of iterations is reached, $|\tau^{(t)} - \tau^{(t-1)}|$ is sufficiently small, we can stop the AMP algorithm.

The details of the AMP algorithm are presented in Algorithm 1. The per-iteration computational complexities of Step 1, Step 2, and Step 3 are $\mathcal{O}(LM)$, $\mathcal{O}(NPM)$, and $\mathcal{O}(LNPM)$, respectively. Therefore, the total per-iteration computational complexity of Algorithm 1 is $\mathcal{O}(LNPM)$. Finally, we detect the device actives and estimate the active devices' channels. Specifically, for all $n \in \mathcal{N}$, the activity detection for device n is given by:

$$\hat{a}_n = \begin{cases} 0, & \theta_n^{best,(t)} < 0, \\ 1, & \theta_n^{best,(t)} \geq 0. \end{cases} \tag{27}$$

For all $n \in \mathcal{N}$ with $\hat{a}_n = 1$, the channel estimate of device n is given by:

$$\hat{h}_{n,p,m} = \hat{x}_{n,p,m}^{best,(t)}, p \in \mathcal{P}, m \in \mathcal{M}. \tag{28}$$

Remark 2 (Comparisons of AMP Algorithms): The factor graph shown in Fig. 1 is different from those in [2], [3]. Firstly, we incorporate a_n as a factor to handle the dependence of $x_{n,p,m}, p \in \mathcal{P}, m \in \mathcal{M}$, for all $n \in N$. Secondly, we treat each scalar $x_{n,p,m}$ as a factor, for all $n \in \mathcal{N}, p \in \mathcal{P}, m \in \mathcal{M}$, to avoid computing matrix inverses to reduce the computational complexity. Because of these differences, the resulting AMP method distinct from those in [2], [3]. Therefore, it is expected that the proposed AMP algorithm achieves higher accuracy and lower computation complexity. Finally, we explicitly show that AMP approximately obtains the MAP-based device activity detection.

IV. Performance Analysis

$\tau^{(t)}$ in (24) is also referred as the "state" of AMP, which can be theoretically predicted by a so-called state evolution.

Theorem 2: In the asymptotic region, $\tau^{(t)}$ in (24) approximately evolves according to:

$$\tau^{(t+1)} = \sigma^2 + \frac{NP}{L}\left(\frac{\rho\beta\tau^{(t)}}{\beta+\tau^{(t)}} + \phi_t\left(\tau^{(t)}\right)\right), \tag{29}$$

where $\tau^{(0)} = \sigma^2 + \frac{NP}{L}\rho\beta$, and

$$\phi_t\left(\tau^{(t)}\right) = \frac{1}{MP}\mathbb{E}_{\mathbf{r}}\left[\frac{g(\mathbf{r},\tau^{(t)})\left(1-g(\mathbf{r},\tau^{(t)})\right)\beta^2}{\left(\beta+\tau^{(t)}\right)^2}\mathbf{r}^H\mathbf{r}\right].$$

Here $\mathbf{r} \triangleq (r_{p,m})_{p\in\mathcal{P},m\in\mathcal{M}} \triangleq \mathbf{x}+\tau^{(t)}\mathbf{v} \in \mathbb{C}^{PM\times 1}$ is a random vector, and $g(\mathbf{r},\tau^{(t)})$ is given by (30), where $\mathbf{x} \in \mathbb{C}^{PM\times 1}$ follows the distribution $(1-\rho)\delta(\mathbf{x}) + \rho\mathcal{CN}(\mathbf{x};\mathbf{0},\beta\mathbf{I})$, and $\mathbf{v} \in \mathbb{C}^{PM\times 1} \sim \mathcal{CN}(\mathbf{v};\mathbf{0},\mathbf{I})$ is independent of $\mathbf{x}$.

Theorem 2 provides the basis for the performance analysis of device activity detection and channel estimation.

Now, we analyze the error probability of the activity detection for device $n \in \mathcal{N}$, $\hat{a}_n^{(t)}$, at the t-th iteration, $p_n^{\mathrm{e}(t)} \triangleq \Pr(\hat{a}_n^{(t)} \neq a_n)$.

Theorem 3: In the asymptotic region, we have:

$$p_n^{\mathrm{e}(t)} = \rho_n\frac{\underline{\Gamma}\left(PM, b_{t,n}(\beta_n,\tau^{(t)})PM\right)}{\Gamma(PM)} + (1-\rho_n)\frac{\bar{\Gamma}\left(PM, c_{t,n}(\beta_n,\tau^{(t)})PM\right)}{\Gamma(PM)}, \tag{31}$$

where $\bar{\Gamma}(\cdot,\cdot)$, $\underline{\Gamma}(\cdot,\cdot)$, and $\Gamma(\cdot)$ denote the upper incomplete Gamma function, lower incomplete Gamma function, and Gamma function, respectively, $\tau^{(t)}$ is given in Theorem 2, $b_{t,n}(\beta_n,\tau^{(t)}) \triangleq \frac{\tau^{(t)}}{\beta_n}\log\left(1+\frac{\beta_n}{\tau^{(t)}}\right)$, and $c_{t,n}(\beta_n,\tau^{(t)}) \triangleq \frac{\beta_n+\tau^{(t)}}{\beta_n}\log\left(1+\frac{\beta_n}{\tau^{(t)}}\right)$.

Since $a > \log(1+a) > \frac{a}{1+a}$ for $a > 0$, we have $b_{t,n}(\beta_n,\tau^{(t)}) < 1$ and $c_{t,n}(\beta_n,\tau^{(t)}) > 1$. According to the property of the Gamma functions, Theorem 3 shows that the error probability for the detection of each device activity decreases with M and P, and $p_n^{\mathrm{e}(t)} \to 0$ as $M \to \infty$.

For each active device $n \in \mathcal{N}$, we analyze the MSE of the channel estimation at the t-th iteration, $\frac{1}{PM}\mathbb{E}\left[\|\mathbf{h}_n - \hat{\mathbf{h}}_n\|_2^2\right]$, where $\mathbf{h}_n \triangleq (h_{n,p,m})_{p\in\mathcal{P},m\in\mathcal{M}}$ and $\hat{\mathbf{h}}_n \triangleq (\hat{h}_{n,p,m})_{p\in\mathcal{P},m\in\mathcal{M}}$.

Theorem 4: In the asymptotic region, we have:

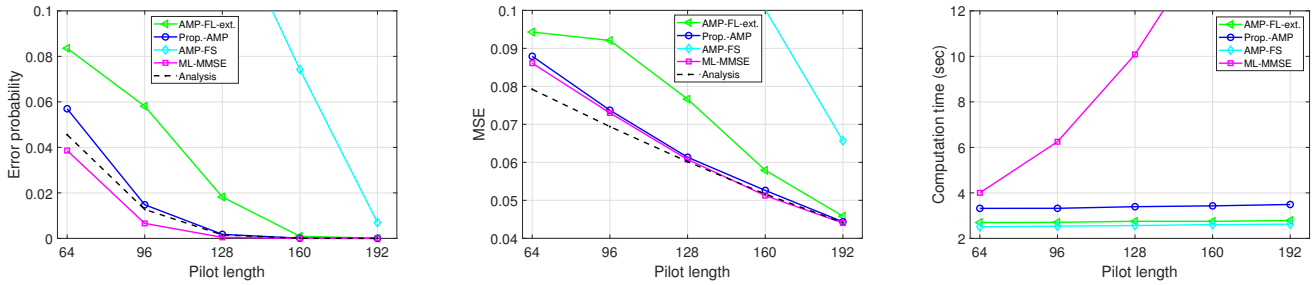
$$\frac{1}{PM}\mathbb{E}\left[\|\mathbf{h}_n - \hat{\mathbf{h}}_n\|_2^2\right] = \frac{\beta_n\tau^{(t)}}{\beta_n+\tau^{(t)}}. \tag{32}$$

Theorem 4 indicates that the MSE for the estimation of each channel component does not change with M and P.

Remark 3 (Comparisons of Analytical Results): The analytical results given by Theorems 2, 3, and 4 recover those for flat fading in [3] when $P = 1$.

V. Numerical results

In this section, we evaluate the accuracy and computation time of Algorithm 1, referred to as *Prop.-AMP*. We construct three baseline schemes, i.e., *ML-MMSE* [6], *AMP-FL-ext.* [3], and *AMP-FS* [9]. Specifically, in *ML-MMSE*, MLE is employed for device activity detection, and *MMSE* estimation



(a) Error probability versus pilot length. (b) MSE versus pilot length. (c) Computation time versus pilot length.

Fig. 2: Error probability, MSE, and computation time versus pilot length.

$$g(\mathbf{r},\tau^{(t)}) \triangleq \frac{\rho_n \prod_{m=1}^{M}\prod_{p=1}^{P} f_{\mathcal{CN}}(0;r_{p,m},\tau^{(t)}+\beta_n)}{\rho_n \prod_{m=1}^{M}\prod_{p=1}^{P} f_{\mathcal{CN}}(0;r_{p,m},\tau^{(t)}+\beta_n)+(1-\rho_n)\prod_{m=1}^{M}\prod_{p=1}^{P} f_{\mathcal{CN}}(0;r_{p,m},\tau^{(t)})} \tag{30}$$

is utilized for channel estimation for devices that are detected active. *AMP-FL-ext.* is the extension of AMP-MMV for flat fading [3] to frequency-selective fading by treating the NP channel taps as NP virtual users and disregarding the correlations among the channel taps of each actual device. After obtaining the activity estimates for these NP virtual devices, $\hat{\mathbf{b}} \in \{0,1\}^{NP}$, the activity estimate of each real device is set as $\hat{a}_n = \frac{\sum_{p\in\mathcal{P}} \hat{b}_{(n-1)P+p}}{P}, n \in \mathcal{N}$. Note that, *AMP-FS* neglects the correlations among subcarriers. In the simulation, set $N = 1000$, $\beta_n = 1, n \in \mathcal{N}$, $\sigma^2 = 0.1$, $M = 64$, $P = 3$, $K = 32$, $Q = L/K$, and $\rho_n = 0.1, n \in \mathcal{N}$. We generate pilots according to i.i.d. $\mathcal{CN}(0, \mathbf{I}_L)$ and normalize the norms to $\sqrt{L}$ [2], [3]. Then the pilots of each user are inserted into these K subcarriers, forming Q OFDM pilot symbols. We generate 500 realizations for $\alpha_n, n \in \mathcal{N}$, $h_{n,p,m}$, $n \in \mathcal{N}$, $p \in \mathcal{P}$, $m \in \mathcal{M}$ and Gaussian pilots and compute the average error probability and MSE over the 500 realizations. Moreover, the AMP-based algorithms run 20 iterations.

Fig. 2 illustrates the error probability of activity detection, MSE of channel estimation, and the computational time versus the pilot length. Fig. 2 shows that the error probability and MSE decrease with the pilot length. The proposed AMP algorithm achieves substantially smaller error probability (up to 74.5%) and MSE (up to 20%) with longer computation times (30% longer) than AMP-FL-ext. and AMP-FS. The significant gains in accuracy come from the fact that the proposed AMP algorithm precisely captures the dependency of the actual frequency domain channels and selects the best estimate among the iterates. Furthermore, the proposed AMP algorithm achieves slightly larger error probability and MSE with significantly shorter computation time (84% shorter) than ML-MMSE. The gains in accuracy and computation time demonstrate the practical value of the proposed AMP algorithm for OFDM-based grant-free access under frequency selective fading. Finally, it can be observed that the analytical error probability and MSE of the proposed AMP algorithm are close to the numerical ones.

VI. Conclusion

This paper investigated AMP-based joint device activity detection and channel estimation for OFDM-based grant-free access under frequency-selective fading. The proposed AMP algorithm captures the dependency of the actual frequency domain channels and selects the best estimate among the iterates to improve the accuracy. Besides, the proposed AMP algorithm avoids computing matrix inversions to reduce the computational complexity. The state evolution of the proposed AMP algorithm, the error probability of device activity detection, and the MSE of channel estimation were analyzed. Numerical results demonstrated the superior performance of the proposed AMP algorithm in terms of accuracy and computation time.

References

[1] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sept. 2018.

[2] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

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

[4] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2925–2951, May 2021.

[5] D. Jiang and Y. Cui, "ML and MAP device activity detections for grant-free massive access in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3893–3908, Jun. 2022.

[6] W. Jiang, Y. Jia and Y. Cui, "Statistical Device Activity Detection for OFDM-Based Massive Grant-Free Access," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3805-3820, Jun. 2023.

[7] F. Lehmann, "Joint User Activity Detection, Channel Estimation, and Decoding for Multiuser/Multiantenna OFDM Systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8263-8275, Sept. 2018.

[8] Y. Zhu et al., "OFDM-Based Massive Grant-Free Transmission Over Frequency-Selective Fading Channels," *IEEE Trans. Wireless Commun.*, vol. 70, no. 7, pp. 4543-4558, July 2022.

[9] M. Ke, Z. Gao, Y. Wu, X. Gao and R. Schober, "Compressive Sensing-Based Adaptive Active User Detection and Channel Estimation: Massive Access Meets Massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 764-779, Sept. 2020.