

Blind Multiple Measurement Vector AMP Based on Expectation Maximization for Grant-Free NOMA

Takanori Hara^{ID}, *Graduate Student Member, IEEE*, and Koji Ishibashi^{ID}, *Senior Member, IEEE*

Abstract—We consider a new approach to perform active user detection and channel estimation in massive grant-free access without requiring prior knowledge of the wireless channels, such as information on large-scale fading coefficients. To this end, we propose a multiple measurement vector approximate message passing (MMV-AMP) with expectation-maximization (EM)-based hyperparameter update, i.e., EM-MMV-AMP. Moreover, we revisited the decision rule for active user detection for EM-MMV-AMP. The numerical results indicate that the performance of the proposed scheme is superior to those of conventional schemes.

Index Terms—Massive connectivity, grant-free, active user detection, channel estimation, approximate message passing, expectation maximization.

I. INTRODUCTION

FUTURE wireless communication systems need to cope with sporadic data traffic in Internet-of-Things networks, satisfying massive connectivity and low latency [1], [2]. However, it is infeasible to meet the demands of conventional multiple access schemes using a grant-based approach because they induce a huge control-signaling overhead to assign radio resources to users ahead of data transmission.

As a key solution to this impediment, *grant-free* random access, such as grant-free non-orthogonal multiple access (GF-NOMA), has been actively investigated [1]–[3]. In GF-NOMA systems, each active user can directly transmit their packet to the base station (BS) by omitting the exclusive assignment of radio resources to each user for data transmission, significantly reducing the signaling overhead. However, GF-NOMA systems must cope with active user detection (AUD) and channel estimation (CE) to retrieve transmitted data accurately from the overlapping signals.

To overcome this difficulty, schemes exploiting the inherent sparsity in sporadic traffic have been actively investigated. For instance, approaches based on message passing mechanisms have been proposed [4]–[8], performing AUD and CE with low computational complexity. However, many of these approaches, e.g., multiple measurement vector approximate message passing (MMV-AMP) [4]–[6], require large-scale fading (LSF) coefficients of all users as prior

knowledge of the wireless channels that is likely to be unavailable at the BS ahead of the uplink transmission. In addition, in [9], [10], the problem of AUD and CE was formulated into a low-rank sparse matrix recovery problem and solved by using a Riemannian optimization technique. As this method reduces the search dimension based on the low-rank structure, the BS must be equipped with antennas more than the number of active users. However, in terms of hardware cost, this may be impractical in GF-NOMA systems in which hundred users are active at each coherence time [1], requiring the same or more number of antennas.

Moreover, a covariance-based approach was proposed in [11], [12], which reformulates AUD into a maximum likelihood estimation problem including a sample covariance matrix and solves it using the coordinate descent (CD) method. Although this approach can achieve superior performance without prior knowledge of the channels, its computational complexity grows quadratically with the pilot sequence length, inducing a prohibitive one in massive grant-free (GF) access. Therefore, a practical and promising technique that can lower the computational complexity and avoid the need for prior information of the channels is strongly desired to support a large number of users efficiently.

In this letter, we propose a new algorithm called expectation-maximization-based MMV-AMP (EM-MMV-AMP) to jointly perform AUD and CE without prior knowledge of the wireless channels. This algorithm is based on the MMV-AMP algorithm [4] integrated with the expectation maximization (EM) algorithm in a manner similar to that in [13]. As a result, the proposed algorithm does not need to estimate the noise variance, unlike the generalized MMV-AMP (GMMV-AMP) algorithm [7]. Furthermore, we revisit the decision rule for AUD and propose an appropriate rule for the EM-MMV-AMP algorithm. Finally, we demonstrate the superior performance of AUD and CE based on EM-MMV-AMP via computer simulations.

II. SYSTEM MODEL

We consider an uplink GF-NOMA system comprising N potential single-antenna users and a common BS equipped with M antennas. For simplicity, it is assumed that each user has a unique pilot sequence with length $L < N$ and is perfectly synchronized in time and frequency. In addition, $K \ll N$ users are active in transmitting sequences within a coherence time.

Let $\mathbf{Y} \in \mathbb{C}^{L \times M}$ and ρ_{tx} denote the received signals at the BS and the transmitted power of active users, respectively. Then, the received signals can be expressed as

$$\mathbf{Y} = \sqrt{\xi} \sum_{n \in \mathcal{A}} \mathbf{a}_n \mathbf{h}_n^T + \mathbf{Z}, \quad (1)$$

Manuscript received December 28, 2021; accepted March 17, 2022. Date of publication March 22, 2022; date of current version June 10, 2022. This work was supported by the Ministry of Internal Affairs and Communications in Japan under Grant JPJ000254. The associate editor coordinating the review of this article and approving it for publication was F. Zhou. (*Corresponding author: Takanori Hara.*)

The authors are with the Advanced Wireless and Communication Research Center, The University of Electro-Communications, Tokyo 182-8585, Japan (e-mail: hara@awcc.uec.ac.jp; koji@ieee.org).

Digital Object Identifier 10.1109/LWC.2022.3161100

where $\xi = L\rho_{\text{tx}}$ is the total transmit energy of each active user, $\mathcal{A} \subset \{1, 2, \dots, N\}$, whose cardinality is K , is the set of active users, $\mathbf{a}_n \in \mathbb{C}^{L \times 1}$, $\mathbf{h}_n \in \mathbb{C}^{M \times 1}$, and $\mathbf{Z} \in \mathbb{C}^{L \times M}$ respectively denote the pilot sequence of user n , the channel between user n and the BS, and the noise matrix with each entry obeying a complex Gaussian distribution with zero mean and variance σ_n^2 .

The channel vector \mathbf{h}_n is modeled as $\mathbf{h}_n = \sqrt{\beta_n} \tilde{\mathbf{h}}_n$, $\forall n$, with LSF coefficient β_n and Rayleigh fading coefficient $\tilde{\mathbf{h}}_n \sim \mathcal{CN}(\tilde{\mathbf{h}}_n; \mathbf{0}_M, \mathbf{I}_M)$. Hence, \mathbf{h}_n follows a complex Gaussian distribution, i.e., $\mathbf{h}_n \sim \mathcal{CN}(\mathbf{h}_n; \mathbf{0}_M, \beta_n \mathbf{I}_M)$. In this letter, β_n is modeled as $\beta_n = -128.1 - 37.6 \log_{10}(d_n)$ [dB], where d_n denotes the distance measured in km between BS and user n . Hereafter, β_n is defined as the prior knowledge of the channels, unless otherwise specified.

Furthermore, (1) can be briefly expressed as

$$\mathbf{Y} = \sqrt{\xi} \mathbf{A} \mathbf{X} + \mathbf{Z}, \quad (2)$$

where $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N] \in \mathbb{C}^{L \times N}$ and $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T \in \mathbb{C}^{N \times M}$ with

$$\mathbf{x}_n = \begin{cases} \mathbf{h}_n, & n \in \mathcal{A}, \\ \mathbf{0}_M, & \text{otherwise.} \end{cases} \quad (3)$$

According to the fading model, each row of \mathbf{X} follows a Bernoulli-Gaussian distribution, i.e.,

$$p(\mathbf{x}_n) = (1 - \lambda) \delta_0(\mathbf{x}_n) + \lambda \mathcal{CN}(\mathbf{x}_n; \mathbf{0}_M, \beta_n \mathbf{I}_M), \quad \forall n, \quad (4)$$

where $\delta_0(\cdot)$ denotes the Dirac delta function, and $\lambda \triangleq K/N \in [0, 1]$ is the activity ratio. Throughout this letter, the activity ratio λ is assumed to be available at the BS.

III. PROPOSED SCHEME

This section first presents an overview of the GMMV-AMP algorithm [7], which estimates \mathbf{X} in (2) without prior knowledge of the channels, to clarify the difference between GMMV-AMP and EM-MMV-AMP. Subsequently, we describe the proposed algorithm and its computational complexity. Finally, we introduce the proper criterion for AUD.

A. Overview of the GMMV-AMP Algorithm

GMMV-AMP decouples the matrix estimation problem based on (2) into MN scalar estimation problems and assumes a prior distribution model of \mathbf{X} [13], given by

$$p(\mathbf{X}) = \prod_{m=1}^M \prod_{n=1}^N [(1 - \gamma_{n,m}) \delta_0(x_{n,m}) + \gamma_{n,m} \mathcal{CN}(x_{n,m}; \mu, \tau)], \quad (5)$$

where $\gamma_{n,m} \in [0, 1]$ presents the sparsity ratio, and μ and τ are the channel gain's mean and variance, respectively. Thus, this scheme executes the element-wise message passing procedure to obtain the minimum mean-squared error (MMSE) estimate of \mathbf{X} , i.e., the posterior mean.

Furthermore, to avoid the need for prior knowledge of the channels $\{\mu, \tau, \gamma_{n,m}, \forall n, m\}$ and the noise variance σ_n^2 , GMMV-AMP exploits the EM algorithm to learn the unknown hyperparameters, i.e., $\boldsymbol{\theta} = \{\mu, \tau, \sigma, \gamma_{n,m}, \forall n, m\}$. In this letter, the details of the GMMV-AMP algorithm are omitted owing to space constraints. Please refer to [7] for further details.

Here, we mention two significant differences between GMMV-AMP and EM-MMV-AMP. One is that GMMV-AMP decouples the matrix estimation problem into scalar problems. The other is that GMMV-AMP needs to learn the hyperparameters corresponding to the noise variance. It is worth noting that the first difference implies that GMMV-AMP does not directly reflect the row sparsity of \mathbf{X} and leads to the degradation of its estimation accuracy as compared to (EM-)MMV-AMP. This is confirmed in Section IV.

B. EM-MMV-AMP Algorithm

The proposed EM-MMV-AMP algorithm is based on the MMV-AMP algorithm [4] integrated with the EM algorithm to estimate desired signal \mathbf{X} and the LSF coefficients β_n without the decoupling considered in GMMV-AMP [7]. As with MMV-AMP, EM-MMV-AMP assumes the prior model in (4) and executes the row-wise message passing procedure based on the following calculations:

$$\mathbf{x}_n^{(t+1)} = \eta_{t,n} \left((\mathbf{R}^{(t)})^H \mathbf{a}_n + \mathbf{x}_n^{(t)} \right), \quad (6)$$

$$\mathbf{R}^{(t+1)} = \mathbf{Y} - \mathbf{A} \mathbf{X}^{(t+1)} + \frac{N}{L} \mathbf{R}^{(t)} \sum_{n=1}^N \frac{\eta'_{t,n} \left((\mathbf{R}^{(t)})^H \mathbf{a}_n + \mathbf{x}_n^{(t)} \right)}{N}, \quad (7)$$

where $t = 0, 1, \dots$ is the index of the iterations, and $\mathbf{X}^{(t)} = [\mathbf{x}_1^{(t)}, \dots, \mathbf{x}_N^{(t)}]^T$ and $\mathbf{R}^{(t)} \in \mathbb{C}^{L \times M}$ represent the estimate of \mathbf{X} and the corresponding residual at the t -th iteration, respectively. This procedure starts with $\mathbf{X}^0 = \mathbf{0}_{N \times M}$ and $\mathbf{R}^{(0)} = \mathbf{Y}$, and utilizes the MMSE denoiser $\eta_{t,n}(\cdot) : \mathbb{C}^{M \times 1} \mapsto \mathbb{C}^{M \times 1}$ and its first-order derivative, i.e., $\eta'_{t,n}(\cdot) \in \mathbb{C}^{M \times M}$.

For simplicity, we assume that the input to the MMSE denoiser as $\hat{\mathbf{x}}_{t,n} \triangleq (\mathbf{R}^{(t)})^H \mathbf{a}_n + \mathbf{x}_n^{(t)}$, can be modeled as follows:

$$\hat{\mathbf{x}}_{t,n} = \mathbf{x}_n + \tau_t \mathbf{v}_n, \quad \forall n, \quad (8)$$

where $\mathbf{v}_n \sim \mathcal{CN}(\mathbf{v}_n; \mathbf{0}_M, \mathbf{I}_M)$, and τ_t is a non-negative number that can be obtained by

$$\tau_t^2 = \frac{\|\mathbf{R}^{(t)}\|_F^2}{LM}. \quad (9)$$

Following this simple statistical model, $\eta_{t,n}(\hat{\mathbf{x}}_{t,n})$ is given by

$$\eta_{t,n}(\hat{\mathbf{x}}_{t,n}) = \psi_{t,n}(\lambda, \beta_n) \left(\frac{\beta_{t,n}}{\beta_{t,n} + \tau_t^2} \right) \hat{\mathbf{x}}_{t,n}, \quad \forall t, n, \quad (10)$$

where we have

$$\psi_{t,n}(\lambda, \beta_{t,n}) = \frac{1}{1 + \left(\frac{1-\lambda}{\lambda} \right) \left(\frac{\beta_{t,n} + \tau_t^2}{\tau_t^2} \right)^M \exp \left[-\frac{\beta_{t,n} \|\hat{\mathbf{x}}_{t,n}\|_2^2}{\tau_t^2 (\beta_{t,n} + \tau_t^2)} \right]}, \quad (11)$$

and $\beta_{t,n}, \forall n$, denotes the hyperparameter that is introduced to avoid the need for the knowledge of the LSF coefficient β_n and is updated using the EM algorithm. Owing to space constraints, the derivation of the MMSE denoiser is omitted. Please refer to [4, Appendix A] for further details.

Let $\boldsymbol{\beta}_t = (\beta_{t,1}, \dots, \beta_{t,N})$, to utilize the EM algorithm, we approximate the true marginal distribution with $\hat{\mathbf{x}}_t$ and τ_t^2 at the t -th iteration of MMV-AMP:

$$p_{X|Y}(\mathbf{x}_n | \mathbf{Y}, \hat{\mathbf{x}}_{t,n}, \tau_t^2, \boldsymbol{\beta}_t) \triangleq \frac{p_X(\mathbf{x}_n; \lambda, \boldsymbol{\beta}_t) \mathcal{CN}(\mathbf{x}_n; \hat{\mathbf{x}}_{t,n}, \tau_t^2 \mathbf{I}_M)}{\int_{\mathbf{x}} p_X(\mathbf{x}; \lambda, \boldsymbol{\beta}_t) \mathcal{CN}(\mathbf{x}; \hat{\mathbf{x}}_{t,n}, \tau_t^2 \mathbf{I}_M)}, \quad (12)$$

where $p_X(\mathbf{x}_n; \lambda, \boldsymbol{\beta}_t)$ is the prior distribution of \mathbf{x}_n given by

$$p_X(\mathbf{x}_n; \lambda, \boldsymbol{\beta}_t) = (1 - \lambda)\delta_0(\mathbf{x}_n) + \lambda\mathcal{CN}(\mathbf{x}_n; \mathbf{0}_M, \beta_{t,n}\mathbf{I}_M). \quad (13)$$

Plugging the prior model (13) into (12), the MMV-AMP approximated posterior can be obtained by

$$p_{X|Y}(\mathbf{x}_n|\mathbf{Y}, \hat{\mathbf{x}}_{t,n}, \tau_t^2, \boldsymbol{\beta}_t) = (1 - \pi_{t,n})\delta_0(\mathbf{x}_n) + \pi_{t,n}\mathcal{CN}(\mathbf{x}_n; \boldsymbol{\gamma}_{t,n}, \nu_{t,n}\mathbf{I}_M), \quad (14)$$

where we have

$$\pi_{t,n} \triangleq \psi_{t,n}(\lambda, \beta_{t,n}), \boldsymbol{\gamma}_{t,n} \triangleq \frac{\beta_{t,n}}{\beta_{t,n} + \tau_t^2} \hat{\mathbf{x}}_{t,n}, \nu_{t,n} \triangleq \frac{\beta_{t,n}\tau_t^2}{\beta_{t,n} + \tau_t^2}. \quad (15)$$

Eqs. (14) and (15) can be derived via the relations shown at the bottom of the next page.

In a manner similar to [13], we can obtain the update rule for β_n as follows:

$$\beta_{t+1,n} = \arg \max_{\beta_n > 0} \mathbb{E}[\ln p_X(\mathbf{x}_n; \lambda, \beta_n) | \mathbf{Y}; \boldsymbol{\beta}_t], \quad (18)$$

where $\mathbb{E}[\cdot | \mathbf{Y}; \boldsymbol{\beta}_t]$ is the expectation conditioned on \mathbf{Y} with the parameters $\boldsymbol{\beta}_t$. The solution of (18) necessarily satisfies

$$\int_{\mathbf{x}_n} p_{X|Y}(\mathbf{x}_n | \mathbf{Y}; \boldsymbol{\beta}_t) \frac{d}{d\beta_n} \ln p_X(\mathbf{x}_n; \lambda, \beta_n) = 0. \quad (19)$$

For the Bernoulli-Gaussian signal model, the derivative of the prior distribution is given by

$$\begin{aligned} \frac{d}{d\beta_n} \ln p_X(\mathbf{x}_n; \lambda, \beta_n) &= \left(\frac{\|\mathbf{x}_n\|_2^2}{\beta_n^2} - \frac{M}{\beta_n} \right) \frac{\lambda\mathcal{CN}(\mathbf{x}_n; \mathbf{0}_M, \beta_n\mathbf{I}_M)}{p_X(\mathbf{x}_n; \lambda, \beta_n)} \\ &= \begin{cases} \frac{\|\mathbf{x}_n\|_2^2}{\beta_n^2} - \frac{M}{\beta_n}, & \mathbf{x}_n \neq \mathbf{0}_M, \\ 0, & \mathbf{x}_n = \mathbf{0}_M. \end{cases} \end{aligned} \quad (20)$$

By substituting (14) and (20) into (19), we can obtain the update rule for β_n as follows:

$$\beta_{t+1,n} = \frac{1}{M} \|\boldsymbol{\gamma}_{t,n}\|_2^2 + \nu_{t,n}, \quad \forall n. \quad (21)$$

The derivation of (21) is based on the technique of [13] and the fact that $\int_{\mathbf{x}_n \neq \mathbf{0}_M} \|\mathbf{x}_n\|_2^2 p_{X|Y}(\mathbf{x}_n | \mathbf{Y}, \hat{\mathbf{x}}_{t,n}, \tau_t^2, \boldsymbol{\beta}_t) d\mathbf{x}_n = \pi_{t,n} (\|\boldsymbol{\gamma}_{t,n}\|_2^2 + \text{tr}(\nu_{t,n}\mathbf{I}_M)) = \pi_{t,n} (\|\boldsymbol{\gamma}_{t,n}\|_2^2 + M\nu_{t,n})$. Notice that the update in (21) is performed user-wise, whereas the GMMV-AMP algorithm updates each channel gain's variance.

The EM update mentioned above is affected by the initial values of the unknown parameters. Moreover, the difference between users' LSF coefficients might be significantly large due to the geometrical distribution of potential users, and some coefficients with even larger values dominate the EM update, which may converge to an undesirable local solution. Hence, we propose the following initialization strategy:

$$\beta_{0,n} = \frac{L}{N} \cdot \frac{\|\mathbf{x}_{\text{MF},n}\|_2^2}{M}, \quad (22)$$

where $\mathbf{x}_{\text{MF},n}$ denotes the n -th row of $\mathbf{X}_{\text{MF}} \triangleq \mathbf{A}^H \mathbf{Y}$.

The algorithm flow of the EM-MMV-AMP algorithm is summarized in Algorithm 1. The flow of Algorithm 1 basically follows the aforementioned update steps, such as Eqs. (6), (7), (9), (21), and (22). Concretely, the proposed algorithm initializes the hyperparameters, the estimate of \mathbf{X} , and the residual by procedures in line 1, and then repeats the operations in lines 3, 4, and 5 until the condition in line 7 is

Algorithm 1 EM-MMV-AMP

Input: Received signals $\mathbf{Y} \in \mathbb{C}^{L \times M}$, measurement matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$, maximum number of iterations T_{amp} , and termination threshold η_{th} .

- 1: Initialize the iteration index t to 1, $\boldsymbol{\beta}_0$ as in (22), and the matrices as $\mathbf{X}^{(0)} = \mathbf{0}_{N \times M}$, $\mathbf{R}^{(0)} = \mathbf{Y}$.
- 2: **repeat**
- 3: Update the estimate $\mathbf{X}^{(t)}$ and the residual $\mathbf{R}^{(t)}$ using (6) and (7), respectively.
- 4: Obtain τ_t^2 based on (9).
- 5: Update the hyperparameters $\beta_{t,n}$ using (21), $\forall n$.
- 6: $t = t + 1$.
- 7: **until** $t \geq T_{\text{amp}}$ or (*).
- 8: (*) $\begin{cases} \|\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}\|_{\text{F}} / \|\mathbf{X}^{(t)}\|_{\text{F}} < \eta_{\text{th}} & \text{(EM-MMV-AMP)} \\ (\tau_t - \tau_{t-1}) / \tau_{t-1} < \eta_{\text{th}} & \text{(MMV-AMP)} \end{cases}$

Output: The estimate of \mathbf{X} ; $\hat{\mathbf{X}} = \mathbf{X}^{(t)}$, the estimated large-scale fading coefficients $\beta_{t,n}$, and $\hat{\mathbf{x}}_{t,n}$, $\forall n$.

satisfied. Note that, as shown in line 8 of Algorithm 1, EM-MMV-AMP exploits a termination criterion that is different from MMV-AMP.

We here discuss the complexity of the proposed algorithm and related algorithms, such as GMMV-AMP and MMV-AMP. The proposed EM-MMV-AMP algorithm mainly executes (6), (7), and (21). As with MMV-AMP, the matrix multiplication $\mathbf{A}\mathbf{X}^{(t+1)}$ in (7) is a dominant term in the computational complexity of EM-MMV-AMP, resulting in their complexity order per iteration of $\mathcal{O}(LMN)$. On the other hand, the computational complexity per iteration of GMMV-AMP is $\mathcal{O}(LMN)$ according to [7, Table I]. In light of the above, the proposed algorithm is comparable to the related algorithms from a computational complexity perspective.

C. Active User Detection by EM-MMV-AMP

Unlike the conventional MMV-AMP, EM-MMV-AMP estimates not only \mathbf{X} in (2) but also the LSF coefficients β_n , followed by modification of the decision rule of active user detection. Therefore, we revisit and propose decision rules for the EM-MMV-AMP algorithm.

As a conventional rule, we consider the rule of [4]:

$$\begin{cases} n \in \hat{\mathcal{A}}, & \text{if } \|\hat{\mathbf{x}}_{t,n}\|_2 \geq \theta_{t,n}, \\ n \notin \hat{\mathcal{A}}, & \text{otherwise,} \end{cases} \quad (23)$$

where $\hat{\mathcal{A}}$ denotes the set of estimated active users and $\theta_{t,n} = M \ln(1 + \beta_{t,n}/\tau_t^2) / (1/\tau_t^2 - 1/(\beta_{t,n} + \tau_t^2))$. Notice that this mainly utilizes the quantities of the MMV-AMP algorithm, $\hat{\mathbf{x}}_{t,n}$, rather than the estimates of β_n . Hence, we consider two decision rules by using the estimates of β_n .

The first rule only relies on τ_t^2 :

$$\begin{cases} n \in \hat{\mathcal{A}}, & \text{if } \beta_{t,n} \geq 2\tau_t^2, \\ n \notin \hat{\mathcal{A}}, & \text{otherwise.} \end{cases} \quad (24)$$

The derivation of (24) is detailed in the Appendix. Note that this rule does not rely on the prior information of the channels.

The second rule can be utilized only when the BS knows the minimum value of β_n , i.e., the LSF coefficient of the cell-edge

TABLE I
SIMULATION PARAMETERS

# of potential users N	500, 1000
# of active users K	100
Sequence length L	100
Transmitted power of each symbol ρ_{tx}	23 [dBm]
Power spectrum density of noise	-169 [dBm/Hz]
System bandwidth	1 [MHz]
Minimum distance between BS and users	0.05 [km]
Maximum distance between BS and users	1 [km]
# of maximum iterations of the algorithms T_{amp}	200
Termination threshold of the algorithms η_{th}	10^{-5}

user. This rule is given by:

$$\begin{cases} n \in \hat{\mathcal{A}}, & \text{if } \beta_{t,n} \geq \beta_{\min}, \\ n \notin \hat{\mathcal{A}}, & \text{otherwise,} \end{cases} \quad (25)$$

with β_{\min} denoting the minimum value. As this rule can exploit the precise boundary between active users and others, it tends to avoid detecting non-active users as active ones.

IV. NUMERICAL RESULTS

In this section, we evaluate the performance of the proposed scheme via computer simulations in terms of normalized mean-squared error (NMSE) and the probabilities of miss detection (MD) and false alarm (FA). The simulation parameters are listed in Table I, and the matrix \mathbf{A} is obtained by normalizing each column comprising L randomly selected and reordered rows of an N -point discrete Fourier transform matrix for all simulations. The NMSE is defined as $\text{NMSE} \triangleq \mathbb{E}[\|\hat{\mathbf{X}} - \mathbf{X}\|_{\text{F}}^2 / \|\mathbf{X}\|_{\text{F}}^2]$. Moreover, MD is the event in which an active user is detected as a non-active user, whereas FA is the vice versa.

In this letter, we consider both the MMV-AMP [4] and GMMV-AMP [7] as state-of-the-art (SotA) schemes. For comparison, in GMMV-AMP, μ is fixed to zero and τ of each user differs, i.e., $\tau \rightarrow \tau_{n,m}, \forall n, m$. In other words, the EM updates of GMMV-AMP are modified as follows:

$$\mu^{(t+1)} \leftarrow 0, \tau_{n,m}^{(t+1)} \leftarrow |A_{n,m}^{(t)}|^2 + B_{n,m}^{(t)}, \forall n, m, \quad (26)$$

where $A_{n,m}^{(t)}$ and $B_{n,m}^{(t)}$ are the quantities calculated using Eqs. (24) in [7]. Furthermore, the hyperparameters $\gamma_{n,m}, \forall n, m$ are initialized as the known activity ratio λ .

A. NMSE Performance

Fig. 1 shows the NMSE performance of the proposed scheme when $N = 500$ and 1000. As a benchmark, this figure includes the performance of the MMSE estimation with the perfect knowledge of active users and LSF coefficients, denoted by ‘‘Oracle MMSE.’’ This result indicates that the proposed initialization strategy (22) appropriately works for both $N = 500$ and 1000, enabling EM-MMV-AMP to approach the ideal performance as the number of antennas at the BS

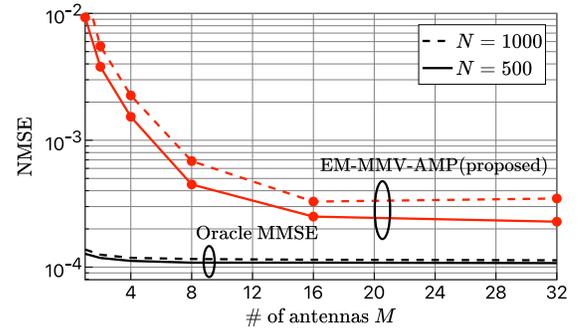


Fig. 1. NMSE performance for $N = 500$ and 1000.

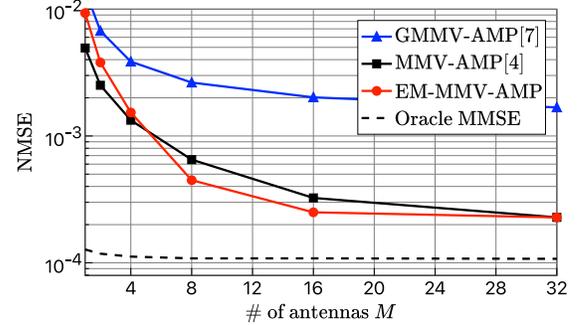


Fig. 2. NMSE performance of the proposed and conventional schemes.

increases. Next, we investigate and discuss the performance focusing on the case of $N = 500$.

We compare the proposed and SotA schemes, as shown in Fig. 2. As shown in the figure, EM-MMV-AMP outperforms GMMV-AMP [7] owing to the full use of row sparsity in \mathbf{X} . Furthermore, the performance of the proposed scheme is comparable to that of MMV-AMP, exploiting the LSF coefficients of all users [4]. In particular, when $M \geq 8$, the proposed scheme slightly outperforms the conventional scheme because the rows of $\hat{\mathbf{X}}$ corresponding to non-active users tend to approach zero.

B. MD and FA Probabilities

To confirm the accuracy of AUD of the proposed scheme, we compared the decision rules considered in Section III-D. Fig. 3 shows the probabilities of MD and FA for (23) and (24) with the performance of the conventional MMV-AMP. According to Fig. 3, the FA probability of the proposed scheme with (23) remains above 10^{-1} , because both the estimates of \mathbf{x}_n and β_n for a non-active user are close to zero. On the other hand, the decision rule of (24) lowers both the probabilities of MD and FA even though BS does not know β_n in advance. For instance, the proposed scheme can attain probabilities below 10^{-2} when the number of antennas is greater than 16. However, MMV-AMP requires at least 20 antennas at the BS to achieve FA probability below 10^{-2} .

$$\mathcal{CN}(\mathbf{x}_n; \mathbf{0}_M, \beta_{t,n} \mathbf{I}_M) \mathcal{CN}(\mathbf{x}_n; \hat{\mathbf{x}}_{t,n}, \tau_t^2 \mathbf{I}_M) = \mathcal{CN}(\mathbf{0}_M; \hat{\mathbf{x}}_{t,n}, (\beta_{t,n} + \tau_t^2) \mathbf{I}_M) \mathcal{CN}(\mathbf{x}_n; \boldsymbol{\gamma}_{t,n}, \nu_{t,n} \mathbf{I}_M) \quad (16)$$

$$\pi_{t,n} = \frac{\lambda \mathcal{CN}(\mathbf{0}_M; \hat{\mathbf{x}}_{t,n}, (\beta_{t,n} + \tau_t^2) \mathbf{I}_M)}{\int_{\mathbf{x}} p_X(\mathbf{x}; \lambda, \boldsymbol{\beta}_t) \mathcal{CN}(\mathbf{x}; \hat{\mathbf{x}}_{t,n}, \tau_t^2 \mathbf{I}_M)} = \frac{\lambda \mathcal{CN}(\mathbf{0}_M; \hat{\mathbf{x}}_{t,n}, (\beta_{t,n} + \tau_t^2) \mathbf{I}_M)}{(1 - \lambda) \mathcal{CN}(\mathbf{0}_M; \hat{\mathbf{x}}_{t,n}, \tau_t^2 \mathbf{I}_M) + \lambda \mathcal{CN}(\mathbf{0}_M; \hat{\mathbf{x}}_{t,n}, (\beta_{t,n} + \tau_t^2) \mathbf{I}_M)} \quad (17)$$

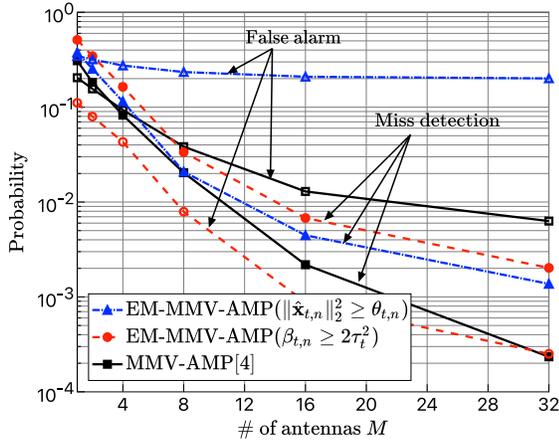


Fig. 3. MD and FA probabilities of the proposed scheme with (23) and (24).

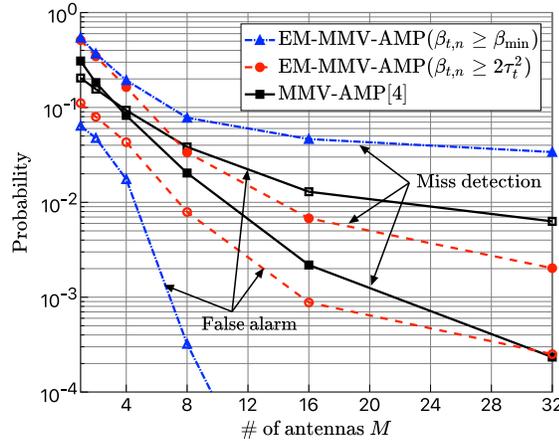


Fig. 4. MD and FA probabilities of the proposed scheme with (24) and (25).

Finally, we present the probabilities of MD and FA when the decision rule in (25) is adopted. The performance of the proposed scheme with AUD based on (25) is shown in Fig. 4. This result indicates that the decision rule of (25) can significantly reduce the FA probability because of the knowledge of the minimum value of β_n , whereas its MD probability is higher than 10^{-2} even when $M = 32$. In light of the above, we conclude that the best decision rule for AUD of EM-MMV-AMP is (24) among the three rules considered in this letter.

V. CONCLUSION

We considered a new approach to perform AUD and CE in massive GF access without requiring prior knowledge of the wireless channels. To this end, we proposed the EM-MMV-AMP algorithm that integrates the EM update with the conventional MMV-AMP. Furthermore, we revisited the EM update and active user decision rules for EM-MMV-AMP to achieve superior performance. Numerical results indicate that the performance of the proposed scheme is comparable to those of conventional schemes in terms of NMSE and probabilities of MD and FA.

APPENDIX

The decision rule of (24) can be derived using certain approximations. First, according to (8), $\boldsymbol{\gamma}_{t,n}$ can be

approximated as

$$\boldsymbol{\gamma}_{t,n} \approx \frac{\beta_{t,n}\tau_t}{\beta_{t,n} + \tau_t^2} \mathbf{v}_n, \quad n \notin \mathcal{A}. \quad (27)$$

In addition, the squared norm of \mathbf{v}_n , i.e., $\|\mathbf{v}_n\|_2^2$, is assumed to be M because $\mathbf{v}_n \sim \mathcal{CN}(\mathbf{v}_n; \mathbf{0}_M, \mathbf{I}_M)$. Thus, we can obtain the following relationship

$$\|\boldsymbol{\gamma}_{t,n}\|_2^2 \approx M \left(\frac{\beta_{t,n}\tau_t}{\beta_{t,n} + \tau_t^2} \right)^2. \quad (28)$$

Hence, the update of $\beta_{t,n}$ for $n \notin \mathcal{A}$ satisfies the following inequality:

$$\begin{aligned} \beta_{t+1,n} &\approx \left(\frac{\beta_{t,n}\tau_t}{\beta_{t,n} + \tau_t^2} \right)^2 + \frac{\beta_{t,n}\tau_t^2}{\beta_{t,n} + \tau_t^2} \\ &= \frac{2\beta_{t,n}^2 + \beta_{t,n}\tau_t^2}{(\beta_{t,n} + \tau_t^2)^2} \tau_t^2 \\ &< 2\tau_t^2 \left(\because 2\beta_{t,n}^2 + \beta_{t,n}\tau_t^2 < 2(\beta_{t,n} + \tau_t^2)^2 \right). \end{aligned} \quad (29)$$

Based on the above, we can obtain the decision rule of (24).

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, Sep. 2018.
- [2] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.
- [3] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 3rd Quart., 2020.
- [4] 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.
- [5] K. Senel and E. G. Larsson, "Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6164–6175, Dec. 2018.
- [6] T. Hara and K. Ishibashi, "Grant-free non-orthogonal multiple access with multiple-antenna base station and its efficient receiver design," *IEEE Access*, vol. 7, pp. 175717–175726, 2019.
- [7] 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, Jan. 2020.
- [8] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, and W. Yu, "Grant-free access via bilinear inference for cell-free MIMO with low-coherence pilots," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7694–7710, Nov. 2021.
- [9] X. Shao, X. Chen, and R. Jia, "A dimension reduction-based joint activity detection and channel estimation algorithm for massive access," *IEEE Trans. Signal Process.*, vol. 68, pp. 420–435, Dec. 2019.
- [10] X. Shao, X. Chen, C. Zhong, and Z. Zhang, "Joint activity detection and channel estimation for mmW/THz wideband massive access," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2020, pp. 1–6.
- [11] S. Haghghatshoar, P. Jung, and G. Caire, "Improved scaling law for activity detection in massive MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Vail, CO, USA, Jun. 2018, pp. 381–385.
- [12] T. Hara, H. Iimori, and K. Ishibashi, "Hyperparameter-free receiver for grant-free NOMA systems with MIMO-OFDM," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 810–814, Apr. 2021.
- [13] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.