Covariance Based Joint Activity and Data Detection for Massive Random Access with Massive MIMO

Zhilin Chen*, Foad Sohrabi*, Ya-Feng Liu§, and Wei Yu*
*Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada
§LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
Emails: *{zchen, fsohrabi, weiyu}@ece.utoronto.ca, §yafliu@lsec.cc.ac.cn

***Abstract*—This paper considers a grant-free massive random access scenario for machine-type communications in which the devices are sporadically active with small payloads to send to a base-station (BS). Each active device transmits the identification information as well as the data symbol by selecting a signature sequence from a pre-assigned pilot sequence set, and the BS detects both the device activity and the data by detecting which sequences are transmitted. This paper makes an observation that in the massive multiple-input multiple-output (MIMO) regime, where the BS is equipped with a large number of antennas, a covariance based detection scheme that solves a maximum likelihood estimation problem for detecting both device activity and data, is more effective than the approximate message passing (AMP) based compressed sensing approach for sequence detection. A main contribution of this paper is an analytic framework capable of accurately predicting the performance of the proposed scheme in terms of the probabilities of false alarm and missed detection for the covariance based approach. The analysis is based on the asymptotic properties of the maximum likelihood estimator under a non-standard condition. Simulation results validate the analysis, and demonstrate that as compared to the AMP based approach, the covariance based approach achieves lower error probabilities by leveraging the multiple antennas at the BS for reliable detection, especially when the pilot signature length is short, as is often the case for low-latency machine-type communications.**

I. Introduction

Massive machine-type communications (mMTC) is an important application area for the fifth generation (5G) cellular technologies [1]. A main challenge of mMTC is to enable scalable and efficient uplink random access for a large pool of devices, among which only a small fraction are active, to send small payload data to the base-station (BS).

This paper investigates the grant-free random access problem [2] in which each active device needs to send a pre-assigned unique signature sequence for user identification and channel probing, then directly transmits the data without waiting for the grant signal from the BS. Toward this send, [3], [4] propose a two-phase detection scheme in which the BS first detects the device activity together with estimating the channels based on the signature sequences transmitted by the devices, then subsequently decodes the data in a second phase based on the estimated channel. However, if the payload data contain only a few bits (e.g., for status update in mMTC), it may be more efficient to combine the two phases and to embed the data symbol in the signature sequence itself [5].

This paper considers such a grant-free massive random access scenario for mMTC with very small data payloads as investigated in [5], in which each device maintains a unique set of pre-assigned 2^J signature sequences. When a device is active, it sends J bits of data by transmitting one sequence from the set. By detecting which sequences are received, the BS acquires both the identity of the active devices as well as the J-bit messages from each of the active devices.

Due to the sparse nature of the device activity as well as the data symbol, this sequence detection problem can be formulated as a compressed sensing problem, for which various efficient numerical algorithms, such as approximate message passing (AMP), have been proposed [3]–[5]. In contrast, this paper makes an observation that if the BS is equipped with a large number of antennas, then an alternative covariance based strategy, first suggested in [6] for sporadic device detection, would have better performance. The work of [6] formulates the sequence detection problem as a maximum likelihood estimation problem, whose solution depends on the received signal through certain covariance matrix only. In this paper, we adopt this covariance based approach also for the scenario with data embedding, and show that as compared to the AMP based compressed sensing approach, the covariance based approach can exploit the multiple BS antennas more effectively, especially when the signature length is short, which is often the case for low-latency mMTC.

Although the covariance based approach has been used in [6] for device activity detection, a performance analysis is still not yet available. As a main contribution of this paper, we provide an accurate performance analysis, in terms of the probabilities of false alarm and missed detection, for the joint device activity detection and data decoding scheme using the covariance based approach. This is accomplished by exploiting the asymptotic properties of the maximum likelihood estimator in the massive multiple-input multiple-output (MIMO) regime, and by establishing a relationship between the distribution of the estimation error and the associated Fisher information matrix. Due to the non-standard boundary condition, the analysis involves solving a convex quadratic programming problem. We obtain a simple closed-form solution to the quadratic program using a reasonable approximation, which in turn provides insight into the distribution of the estimation error.

The massive random access problem has a rich history. Many of the well-known protocols are based on the (slotted)

ALOHA or its variations; see [7]–[10] and references therein. In ALOHA, each transmitter sends the data as one packet repeatedly, and the receiver decodes the packet if there is no collision. Although there are extensive studies on the ALOHA based schemes, most of the works (e.g., [7], [8]) abstract the physical channel model into a collision channel model and focus on the contention resolution design. Different from the works mentioned above, this paper studies random access with physical channel model containing both fading and noise. Other closely related works in the context of mMTC include [11], [12], where compressed sensing techniques are used for channel and data estimation, and our previous works [3], [4], [13] that mainly focus on the device activity detection.

II. System Model and Problem Formulation

A. System Model

Consider an uplink single-cell system consisting of one BS with M antennas and N single-antenna devices. Assume that the channel model is block fading, i.e., the channel is static in each coherence block while varies among different blocks. Due to the sporadic traffic, only $K \ll N$ devices are active in each block, and each active device has a message of J bits to transmit, where J is assumed to be a small number in the context of mMTC. To enable efficient random access, we consider a grant-free scheme where device activity detection and data detection are performed simultaneously as in [5].

For the purpose of device identification and information transmission, assume that each device n maintains a unique sequence set of size $Q \triangleq 2^J$ as

$$\mathcal{S}_n = \left\{\mathbf{s}_n^1, \mathbf{s}_n^2, \ldots, \mathbf{s}_n^Q\right\}, \tag{1}$$

where $\mathbf{s}_n^q = [s_{n1}^q, s_{n2}^q, \ldots, s_{nL}^q]^T \in \mathbb{C}^{L\times 1}, 1 \le q \le Q$ is a sequence of length L, where L is smaller than the coherence block length. When device n is active and needs to transmit J bits data to the BS, the device transmits one sequence from $\mathcal{S}_n$. The BS then performs the user activity detection and data decoding simultaneously by detecting which sequences are transmitted based on the received signal, which is a superposition of the transmitted signals from all the active devices. In this paper, we assume that all sequences are generated from independent and identically distributed (i.i.d.) complex Gaussian distribution with zero mean and unit variance. Since the size of $\mathcal{S}_n$ increases exponentially as J increases, such a sequence selection scheme is suitable for small values of J.

Let $a_n^q \in \{1, 0\}$ indicate whether or not sequence q of device n is transmitted. Since at most one sequence is selected by each device, a_n^q satisfies $\sum_{q=1}^Q a_n^q \in \{0, 1\}$, where $\sum_{q=1}^Q a_n^q = 0$ indicates that device n is inactive, and $\sum_{q=1}^Q a_n^q = 1$ indicates that device n is active. Let $g_n\mathbf{h}_n$ denote the channel vector between device n and the BS, where $\mathbf{h}_n \in \mathbb{C}^{M\times 1}$ is the Rayleigh fading component over multiple antennas following i.i.d. complex Gaussian distribution with zero mean and unit variance, and g_n is the large-scale fading component depending on the user location. Assume that the sequences selected by all the active devices are transmitted synchronously. The received signal $\mathbf{Y} \in \mathbb{C}^{L\times M}$ at the BS can be expressed as

$$\begin{aligned}\mathbf{Y} &= \sum_{n=1}^N \sum_{q=1}^Q a_n^q \mathbf{s}_n^q g_n \mathbf{h}_n^T + \mathbf{W} \\ &= \sum_{n=1}^N \begin{bmatrix}\mathbf{s}_n^1 & \ldots & \mathbf{s}_n^Q\end{bmatrix} \begin{bmatrix} a_n^1 g_n & & \\ & \ddots & \\ & & a_n^Q g_n \end{bmatrix} \begin{bmatrix}\mathbf{h}_n^T \\ \vdots \\ \mathbf{h}_n^T\end{bmatrix} + \mathbf{W} \\ &\triangleq \sum_{n=1}^N \mathbf{S}_n \mathbf{D}_n \mathbf{H}_n + \mathbf{W}, \end{aligned} \tag{2}$$

where $\mathbf{S}_n \triangleq [\mathbf{s}_n^1, \ldots, \mathbf{s}_n^Q] \in \mathbb{C}^{L\times Q}$ is a stack of all sequences of device n, $\mathbf{D}_n \triangleq \mathrm{diag}\{a_n^1 g_n, \ldots, a_n^Q g_n\} \in \mathbb{C}^{Q\times Q}$ is a diagonal matrix with at most one non-zero diagonal entry since $\sum_{q=1}^Q a_n^q \in \{0, 1\}$, $\mathbf{H}_n \triangleq [\mathbf{h}_n, \ldots, \mathbf{h}_n]^T \in \mathbb{C}^{Q\times M}$ is the n-th channel matrix with repeated rows, and $\mathbf{W} \in \mathbb{C}^{L\times M}$ is the effective i.i.d. Gaussian noise whose variance σ_w^2 is the background noise power normalized by the device transmit power. By further concatenating all sequence matrices of N users as $\mathbf{S} \triangleq [\mathbf{S}_1, \ldots, \mathbf{S}_N] \in \mathbb{C}^{L\times NQ}$, the received signal in (2) can be written in a more compact form as

$$\begin{aligned}\mathbf{Y} &= \begin{bmatrix}\mathbf{S}_1 & \ldots & \mathbf{S}_N\end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & & \\ & \ddots & \\ & & \mathbf{D}_N \end{bmatrix} \begin{bmatrix}\mathbf{H}_1 \\ \vdots \\ \mathbf{H}_N\end{bmatrix} + \mathbf{W} \\ &\triangleq \mathbf{S}\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H} + \mathbf{W}, \end{aligned} \tag{3}$$

where $\boldsymbol{\Gamma}^{\frac{1}{2}} \triangleq \mathrm{diag}\{\mathbf{D}_1, \ldots, \mathbf{D}_N\} \in \mathbb{C}^{NQ\times NQ}$ is a diagonal matrix since it is a block diagonal matrix where each block $\mathbf{D}_n$ is also diagonal, and $\mathbf{H} \triangleq [\mathbf{H}_1^T, \ldots, \mathbf{H}_N^T]^T \in \mathbb{C}^{NQ\times M}$.

B. Problem Formulation

Our goal is to detect the binary variable a_n^q, which indicates both the activity of device n and its data if it is active. One potential approach is to treat (3) as a compressive sensing problem with multiple measurement vectors by using the row sparsity of $\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}$. Once $\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}$ is recovered from $\mathbf{Y}$, a_n^q can be determined by the rows of $\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}$. However, such an approach usually requires an algorithmic complexity that scales with M since $\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}$ is of size $NQ\times M$, which may not be preferred for a large M. Moreover, the channel state information contained in $\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}$ is not necessary for data detection in the considered scheme. Therefore, instead of recovering $\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}$, we consider to estimate diagonal matrix $\boldsymbol{\Gamma}$, from which a_n^q can be also determined.

We formulate the estimation of $\boldsymbol{\Gamma}$ as a maximum likelihood estimation problem as suggested in [6]. Let $\boldsymbol{\gamma} \in \mathbb{C}^{NQ\times 1}$ denote the diagonal entries of $\boldsymbol{\Gamma}$, i.e., $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_N^T]^T$, where $\boldsymbol{\gamma}_n = [a_n^1 g_n^2, \ldots, a_n^Q g_n^2]^T \in \mathbb{C}^{Q\times 1}$. We will use $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$ interchangeably due to the correspondence. From (3) we observe that given $\boldsymbol{\gamma}$, the columns of $\mathbf{Y}$, each denoted as $\mathbf{y}_m \in \mathbb{C}^{L\times 1}, 1 \le m \le M$, can be seen as independent samples from a multivariate complex Gaussian distribution as

$$\mathbf{y}_m \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{S}\boldsymbol{\Gamma}^{\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{S}^H + \sigma_w^2\mathbf{I}\right), \tag{4}$$

where $\mathbf{\Lambda} \triangleq \text{diag}\{\mathbf{E}, \ldots, \mathbf{E}\} \in \mathbb{R}^{NQ \times NQ}$ is a block diagonal matrix with each block $\mathbf{E} \in \mathbb{R}^{Q \times Q}$ representing the all-one matrix, and $\mathbf{I}$ is an identity matrix. Note that the all-one matrix $\mathbf{E}$ comes from the fact that $\mathbf{H}$ in (3) contains repeated rows. By noticing that each diagonal block $\mathbf{D}_n$ in $\mathbf{\Gamma}^{\frac{1}{2}}$ has at most one non-zero entry in the diagonal, it can be verified that $\mathbf{D}_n^{\frac{1}{2}} \mathbf{E} \mathbf{D}_n^{\frac{1}{2}} = \mathbf{D}_n$, which implies that the covariance matrix in (4) can be simplified as $\mathbf{S}\mathbf{\Gamma}^{\frac{1}{2}}\mathbf{\Lambda}\mathbf{\Gamma}^{\frac{1}{2}}\mathbf{S}^H + \sigma_w^2\mathbf{I} = \mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma_w^2\mathbf{I}$.

Based on (4), we express the likelihood of $\mathbf{Y}$ given $\boldsymbol{\gamma}$ as

$$\begin{aligned} p(\mathbf{Y}|\boldsymbol{\gamma}) &= \prod_{m=1}^{M} \frac{1}{|\pi\mathbf{\Sigma}|} \exp\left(-\mathbf{y}_m^H \mathbf{\Sigma}^{-1} \mathbf{y}_m\right) \\ &= \frac{1}{|\pi\mathbf{\Sigma}|^M} \exp\left(-\operatorname{Tr}\left(\mathbf{\Sigma}^{-1}\mathbf{Y}\mathbf{Y}^H\right)\right) \end{aligned} \tag{5}$$

where $\mathbf{\Sigma} \triangleq \mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma_w^2\mathbf{I}$, $|\cdot|$ denotes the determinant of a matrix, and $\text{Tr}(\cdot)$ denotes the trace of a matrix. The maximization of the log-likelihood $\log p(\mathbf{Y}|\boldsymbol{\gamma})$ can be casted as the minimization of $-\log p(\mathbf{Y}|\boldsymbol{\gamma})$ expressed as follows

$$\begin{aligned} \underset{\boldsymbol{\gamma}}{\text{minimize}} \quad & \log|\mathbf{\Sigma}| + \frac{1}{M}\operatorname{Tr}\left(\mathbf{\Sigma}^{-1}\mathbf{Y}\mathbf{Y}^H\right) && \text{(6a)} \\ \text{subject to} \quad & \boldsymbol{\gamma} \geq 0, && \text{(6b)} \\ & |\boldsymbol{\gamma}_n|_0 \leq 1, \ n = 1, 2, \ldots, N, && \text{(6c)} \end{aligned}$$

where $|\cdot|_0$ denotes the number of non-zero entries of a vector. Note that constraint $\boldsymbol{\gamma} \geq 0$ is due to $a_n^q g_n^2 \geq 0$, which defines a natural parameter space of $\boldsymbol{\gamma}$ that guarantees the positive definiteness of covariance matrix $\mathbf{\Sigma}$. The constraint $|\boldsymbol{\gamma}_n|_0 \leq 1$ comes from the sequence selection, i.e., $\sum_{q=1}^{Q} a_n^q \in \{0, 1\}$.

The solution to (6) depends on $\mathbf{Y}$ through $\frac{1}{M}\mathbf{Y}\mathbf{Y}^H \in \mathbb{C}^{L \times L}$, which is the sampled covariance of the received signal averaged over different antennas, whose size scales with L instead of M. For this reason, the formulation (6) is called the covariance based approach.

It is worth mentioning that the use of maximum likelihood for parameter estimation with multivariate Gaussian observations has appeared in various contexts. For example, in the direction of arrival (DOA) estimation, a similar optimization problem is formulated in [14] for angle estimation. More other related examples include the sparse approximation in [15] and the sparse user activity detection in [6], where the maximum likelihood is used for sparse signal recovery. The considered problem slightly differs from previously mentioned problems in that the parameter $\boldsymbol{\gamma}$ exhibits an extra block structure as indicated in (6c). However, in despite of the difference, the algorithms developed previously in [6], [15] are still very useful to solve (6), as we discuss in the next section.

III. Joint Device Activity and Data Detection

To perform joint device activity and data detection at the BS, the main task is to solve the optimization problem (6). Without considering the constraint (6c), the problem can be relaxed to find the optimal $\boldsymbol{\gamma}$ in the NQ-dimensional parameter space $[0, +\infty)^{NQ}$. Although the relaxed problem is still not convex in general due to the fact that $\log|\mathbf{\Sigma}|$ is concave whereas $\text{Tr}\left(\mathbf{\Sigma}^{-1}\mathbf{Y}\mathbf{Y}^H\right)$ is convex, there already exist various algorithms that perform well in practice. For example, the authors of [15] propose a multiple sparse Bayesian learning (M-SBL) algorithm based on expectation maximization (EM) that estimates $\boldsymbol{\gamma}$ iteratively. The authors of [6] suggest a coordinate descent algorithm that randomly updates each coordinate of $\boldsymbol{\gamma}$ until convergence. Note that although the relaxed problem is non-convex, the global optimality of M-SBL or coordinate descent for such a problem can be justified if $\mathbf{\Gamma}^{\frac{1}{2}}\mathbf{H}$ or $\mathbf{S}$ satisfies certain conditions; see [15] and [6].

Algorithm 1 Coordinate descent to estimate $\boldsymbol{\gamma}$

Initialize $\hat{\boldsymbol{\gamma}} = \mathbf{0}, \hat{\mathbf{\Sigma}} = \sigma_w^2\mathbf{I}, \hat{\mathbf{\Sigma}}^{-1} = \sigma_w^{-2}\mathbf{I}$.
for $i = 1, 2, \ldots$ **do**
  Randomly select a permutation $i_1, i_2, \ldots, i_{NQ}$ of the coordinate indices $\{1, 2, \ldots, NQ\}$ of $\hat{\boldsymbol{\gamma}}$
  for $n = 1$ to NQ **do**
    $\delta = \max\{\frac{\mathbf{s}_{i_n}^H \hat{\mathbf{\Sigma}}^{-1} \frac{\mathbf{Y}\mathbf{Y}^H}{M} \hat{\mathbf{\Sigma}}^{-1} \mathbf{s}_{i_n} - \mathbf{s}_{i_n}^H \hat{\mathbf{\Sigma}}^{-1} \mathbf{s}_{i_n}}{(\mathbf{s}_{i_n}^H \hat{\mathbf{\Sigma}}^{-1} \mathbf{s}_{i_n})^2}, -\hat{\gamma}_{i_n}\}$
    $\hat{\gamma}_{i_n} \leftarrow \hat{\gamma}_{i_n} + \delta$
    $\hat{\mathbf{\Sigma}}^{-1} \leftarrow \hat{\mathbf{\Sigma}}^{-1} - \delta \frac{\hat{\mathbf{\Sigma}}^{-1} \mathbf{s}_{i_n} \mathbf{s}_{i_n}^H \hat{\mathbf{\Sigma}}^{-1}}{1 + \delta \mathbf{s}_{i_n}^H \hat{\mathbf{\Sigma}}^{-1} \mathbf{s}_{i_n}}$
  end for
end for
Output $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_1, \ldots, \hat{\gamma}_{NQ}]^T$.

In this paper, we adopt the coordinate descent approach from [6] to first solve problem (6) without constraint (6c). After an estimate $\hat{\boldsymbol{\gamma}}$ is obtained, we then use a coordinate selection operation with thresholding to enforce the block-wise constraint and to detect a_n^q. The coordinate descent algorithm is given in Algorithm 1. As compared to [6], we add random index permutation and rank-1 update to further improve the efficiency. To perform the coordinate selection and thresholding, let l_{th} be a predefined threshold, and $\hat{\gamma}_n^q$ be the q-th entry in the n-th block of $\hat{\boldsymbol{\gamma}}$, we determine a_n^q by

$$a_n^q = \begin{cases} 1, & \text{if} \quad \hat{\gamma}_n^q \geq l_{th}, \hat{\gamma}_n^q = \max_{i=1}^{Q} \hat{\gamma}_n^i, \\ 0, & \text{else}, \end{cases} \tag{7}$$

where the condition $\hat{\gamma}_n^q = \max_{i=1}^{Q} \hat{\gamma}_n^i$ guarantees that $\sum_{q=1}^{Q} a_n^q \in \{0, 1\}$ is satisfied, and the purpose of the threshold l_{th} is to balance the missed detection and the false alarm since the estimate $\hat{\boldsymbol{\gamma}}$ may not be sparse.

It is worth mentioning that instead of enforcing (6c) after an estimate of $\boldsymbol{\gamma}$ is obtained, we can also enforce (6c) during the updating of each block, which leads to a block coordinate descent. However, the performance improvement is not substantial, especially when M is large. In the simulation, we observe that even without considering (6c) in solving (6), the estimate $\hat{\boldsymbol{\gamma}}$ given by coordinate descent is approximately sparse, hence constraint (6c) is already approximately satisfied. We will further explain this via the *consistency* of the maximum likelihood estimator in the next section.

IV. Asymptotic Performance Analysis

As the main contribution of this paper, we provide a way to analyze the performance of the considered random access

scheme in the massive MIMO regime. The difficulty of such an analysis lies in the characterization of the error probability of device activity detection or data decoding, which requires the distribution information of the estimation error $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$. The method used in this paper is to consider the maximum likelihood estimator $\hat{\boldsymbol{\gamma}}^{ML}$ as a surrogate of $\hat{\boldsymbol{\gamma}}$, and to exploit the asymptotic properties of the maximum likelihood estimator to analyze the estimation error. In the following, we first discuss $\hat{\boldsymbol{\gamma}}^{ML}$ in the regime $M \to \infty$, and later establish the connection between the distribution of $\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma}$ and the associated Fisher information matrix as $M \to \infty$.

Since it is difficult to directly analyze the estimation error $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$ given by Algorithm 1, as an alternative, we consider the global optimal solution, i.e., the maximum likelihood estimator $\hat{\boldsymbol{\gamma}}^{ML}$, to the original maximum likelihood problem (6) without considering (6c). After dropping (6c), the feasible parameter space of $\boldsymbol{\gamma}$ is $[0, +\infty)^{NQ}$, which is easier to deal with. Later we show that (6c) is automatically satisfied if $M \to \infty$. Note that $\hat{\boldsymbol{\gamma}}^{ML}$ can be regarded as a good surrogate of $\hat{\boldsymbol{\gamma}}$ because: (i) $\hat{\boldsymbol{\gamma}}$ is at least a local minimizer; (ii) as suggested in [6] $\hat{\boldsymbol{\gamma}}$ is also a global minimizer if $\mathbf{S}$ satisfies certain conditions.

Based on the standard estimation theory [16], as the number of i.i.d. samples increases, the maximum likelihood estimator $\hat{\boldsymbol{\gamma}}^{ML}$ is *consistent*, i.e.,

$$\hat{\boldsymbol{\gamma}}^{ML} \to \boldsymbol{\gamma}, \quad \text{as } M \to \infty, \tag{8}$$

where $\boldsymbol{\gamma}$ is the true parameter, and $\to$ denotes convergence in probability. Thus $\hat{\boldsymbol{\gamma}}^{ML}$ should concentrate around the true $\boldsymbol{\gamma}$ and become an approximate sparse vector for large M, which suggests that (6c) is satisfied approximately when M is large.

To further analyze the distribution of the estimation error $\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma}$, the *asymptotic normality* of the maximum likelihood estimator [16] can be exploited. This property states that $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma})$ tends to a Gaussian distribution as the number of i.i.d. observations increases, i.e.,

$$M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma}) \to \mathcal{N}(0, M\mathbf{J}^{-1}(\boldsymbol{\gamma})), \quad \text{as } M \to \infty, \tag{9}$$

where $\mathbf{J}(\boldsymbol{\gamma})$ is the Fisher information matrix, whose (i,j)-th entry is defined as

$$[\mathbf{J}(\boldsymbol{\gamma})]_{ij} = -\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{Y}|\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}\right]. \tag{10}$$

The probability $p(\mathbf{Y}|\boldsymbol{\gamma})$ in the above is given in (5), and the expectation is taken with respect to $\mathbf{Y}$. Note that the asymptotic normality involves $\mathbf{J}^{-1}(\boldsymbol{\gamma})$ and is closely related to the Cramer-Rao bound [16]. However, such asymptotic normality holds only when the true value of $\boldsymbol{\gamma}$ is an interior point of the parameter space $[0, +\infty)^{NQ}$. In the considered problem, most of the entries in $\boldsymbol{\gamma}$ are zero, indicating that the true value of $\boldsymbol{\gamma}$ is always on the boundary of $[0, +\infty)^{NQ}$. Under such a case, the distribution of $\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma}$ is no longer multivariate Gaussian with covariance $\mathbf{J}(\boldsymbol{\gamma})^{-1}$. Instead, $\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma}$ depends on $\mathbf{J}(\boldsymbol{\gamma})$ through a more complicated way. Such boundary case has been studied in [17] for general estimation problems. Based on Theorem 2 in [17], the asymptotic distribution can be characterized in the following proposition.

Proposition 1. *Let $\mathbf{x} \in \mathbb{R}^{NQ\times 1}$ be a random vector sampled from the multivariate Gaussian distribution $\mathcal{N}\left(\mathbf{0}, M\mathbf{J}^{-1}(\boldsymbol{\gamma})\right)$. Let $\boldsymbol{\mu} \in \mathbb{R}^{NQ\times 1}$ be the solution to the following constrained quadratic programming problem*

$$\underset{\boldsymbol{\mu}}{\text{minimize}} \quad \frac{1}{M}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{J}(\boldsymbol{\gamma})(\mathbf{x}-\boldsymbol{\mu}) \tag{11a}$$

$$\text{subject to} \quad \mu_i \geq 0,\ i \in \mathcal{I}, \tag{11b}$$

where μ_i is the i-th entry of $\boldsymbol{\mu}$, and $\mathcal{I}$ is an index set corresponding to the zero entries of $\boldsymbol{\gamma}$, i.e., $\mathcal{I} \triangleq \{i|\gamma_i = 0\}$. Note that $\boldsymbol{\mu}$ is random due to the randomness of $\mathbf{x}$. Then $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma})$ has asymptotically the same distribution as $\boldsymbol{\mu}$ as $M \to \infty$.

The above proposition provides a way to accurately compute the asymptotic distribution of $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma})$. A special case is when the index set $\mathcal{I}$ is empty, i.e., all entries of $\boldsymbol{\gamma}$ are strictly larger than zero, the solution to (11) is $\boldsymbol{\mu} = \mathbf{x}$, indicating that the distribution of $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma})$ becomes Gaussian as stated in (9). When some entries of $\boldsymbol{\gamma}$ are zero, as the case in the considered activity detection and data decoding problem, (11) does not admit a closed-form solution in general. However, the proposition is still very useful in the sense that the distribution of $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma})$ can be empirically calculated by solving the convex optimization problem (11) efficiently, rather than solving the original non-convex problem (6).

Besides employing a numerical algorithm to solve (11), it is also possible to get an approximate analytic solution if $\mathbf{J}(\boldsymbol{\gamma})$ exhibits some special structure. To this end, we first derive an explicit expression of $\mathbf{J}(\boldsymbol{\gamma})$ in the following proposition.

Proposition 2. *Consider the likelihood function in* (5)*, where $\boldsymbol{\gamma}$ is the parameter to be estimated. The Fisher information matrix of $\boldsymbol{\gamma}$ defined in* (10) *can be expressed as*

$$\mathbf{J}(\boldsymbol{\gamma}) = M\left(\mathbf{P} \odot \mathbf{P}^*\right), \tag{12}$$

where $\mathbf{P} \triangleq \mathbf{S}^H\left(\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{S}$, $\odot$ is the element-wise product, and $(\cdot)^*$ is the conjugate operation.

Proof. Please see Appendix A. □

We observe from (12) that $\mathbf{J}(\boldsymbol{\gamma})$ is diagonally dominant, if all sequences in $\mathbf{S}$ are randomly generated. This is because with random sequences the columns and rows of $\mathbf{S}$ are approximately mutually orthogonal, which leads to diagonally dominant matrices $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma_w^2\mathbf{I}$ and $\mathbf{P}$, and also $\mathbf{J}(\boldsymbol{\gamma})$. Note that even though all these are coarse approximations, they still help us get insights about the estimation error. By approximating $\mathbf{J}(\boldsymbol{\gamma})$ as a diagonal matrix, the coordinates in (11) are decoupled, and a simple solution is

$$\mu_i = \begin{cases} x_i, & \text{if} \quad i \notin \mathcal{I}, \\ x_i^+, & \text{if} \quad i \in \mathcal{I}, \end{cases} \tag{13}$$

where $x_i^+ \triangleq \max\{x_i, 0\}$, and x_i is the i-th entry of $\mathbf{x}$. Note that by Propositions 1 and 2, $\mathbf{x}$ follows $\mathcal{N}(\mathbf{0}, (\mathbf{P} \odot \mathbf{P}^*)^{-1})$. The solution in (13) indicates that with such an approximation, the estimation error on the non-zero entries of $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma})$

is Gaussian with variance $[(\mathbf{P}\odot\mathbf{P}^*)^{-1}]_{ii}, i\notin\mathcal{I}$, whereas the estimation error on the zero entries of $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML}-\boldsymbol{\gamma})$ is half Gaussian with variance $[(\mathbf{P}\odot\mathbf{P}^*)^{-1}]_{ii}, i\in\mathcal{I}$ plus a point mass at 0. Therefore, for given M the error $\hat{\boldsymbol{\gamma}}^{ML}-\boldsymbol{\gamma}$ depends on a Gaussian distribution with variance $M^{-1}[(\mathbf{P}\odot\mathbf{P}^*)^{-1}]_{ii}$, whose value drops linearly as M increases since $(\mathbf{P}\odot\mathbf{P}^*)^{-1}$ does not depends on M.

It is worth mentioning that in the asymptotic analysis so far we implicitly assume that the Fisher information matrix $\mathbf{J}(\boldsymbol{\gamma})$ is non-singular (invertible), i.e., $\mathrm{Rank}(\mathbf{P}\odot\mathbf{P}^*)=NQ$. By using $\mathrm{Rank}(\mathbf{A}\odot\mathbf{B})\leq\mathrm{Rank}(\mathbf{A})\,\mathrm{Rank}(\mathbf{B})$ for arbitrary matrices $\mathbf{A}$ and $\mathbf{B}$, we obtain a necessary condition on L and NQ to fulfill the non-singularity

$$NQ=\mathrm{Rank}(\mathbf{P}\odot\mathbf{P}^*)\leq\mathrm{Rank}(\mathbf{P})^2\leq L^2, \qquad (14)$$

from which $NQ\leq L^2$ is necessary to guarantee the existence of $\mathbf{J}^{-1}(\boldsymbol{\gamma})$. In other words, at most L^2 device sequences can be supported for the asymptotic analysis discussed above. When the Fisher information matrix is singular, some more involved techniques are required to reformulate the estimation problem.

V. Simulation Results

We consider a single cell of radius 1000m containing $N=400$ potential devices, among which $K=80$ devices are active. We consider the worst case that all devices are located in the cell-edge such that the large-scale fading components g_n are the same for an ease of demonstration. The power spectrum density of the background noise is -169dBm/Hz, and the transmit power of each device is set as 25dBm.

The performance metrics are the probabilities of false alarm and missed detection, which are defined as follows. The probability of false alarm corresponds to the event that a device is inactive but declared active. The probability of missed detection corresponds to two types of error events: a device is active but is declared to be inactive, *or* a device is active but the data is not correctly decoded although the device is declared active. Note that the probability of missed detection used here slightly differs from its standard definition. Different probabilities of false alarm and missed detection can be obtained by adjusting the value of the threshold l_{th} in (7).

We first validate the asymptotic analysis in Fig. 1 with $J=1$ and $M=128$, or 256. We compare the simulated performance obtained by Algorithm 1, and the theocratical performance predicted by Propositions 1 and 2. We observe that the simulated and theocratical curves match very well. There are slightly larger differences when the probability of missed detection or probability of false alarm is very small, which might be due to the mismatch in the tail distributions.

In Fig. 2 we compare the covariance based method used in this paper with an AMP based method from compressed sensing that has been used to solve the random access problem in [3]–[5], as the signature length L increases. We set $J=2$. Since there are two types of detection errors, to conveniently show the error behavior with L, we properly select the threshold l_{th} to achieve a point where probability of false alarm and probability of missed detection are equal, which is

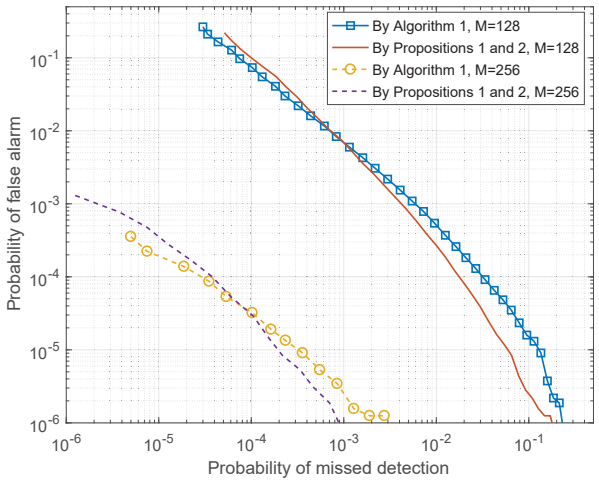

Fig. 1. Comparison of the simulated results and the analysis in terms of probability of false alarm and probability of missed detection.

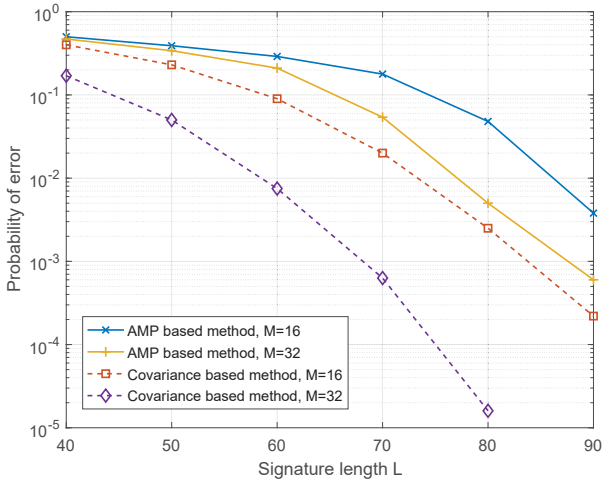

Fig. 2. Performance comparison between the covariance based method and the AMP based method with different L and $J=2$.

represented as "probability of error" in Fig. 2. We observe that increasing L substantially decreases the error probability for the covariance based method. However, for the AMP based method, the benefit of increasing L becomes obvious only when L exceeds some threshold, e.g., $L=60$ when $M=32$. This can be explained by the phase transition in AMP [18], which requires L to be sufficient large, depending on the problem size. We observe that the covariance based method consistently outperforms the AMP based method. Fig. 2 also shows that the covariance based method is more suitable for small L in which case AMP may not work well.

Fig. 3 compares the covariance based method and the AMP based method as the number of antennas at the BS increases. We consider the signature length $L=60$, or 80, and show the behavior of the detection error against M. We observe that for the covariance based method the error drops effectively as

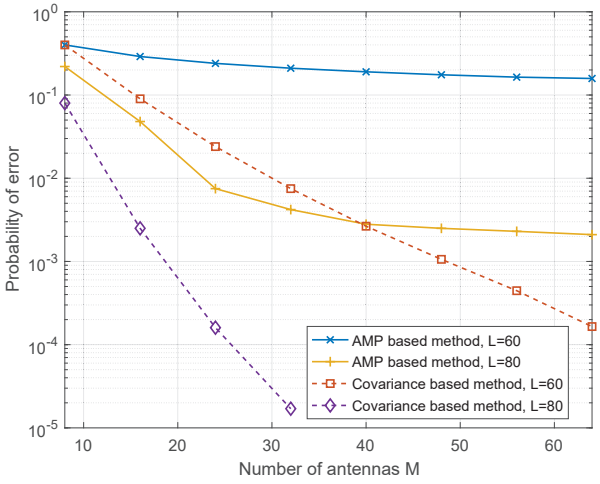

Fig. 3. Performance comparison between the covariance based method and the AMP based method with different M and $J = 2$.

M increases, whereas for the AMP based method, the error becomes saturated when M exceeds some point, e.g., $M = 32$ when $L = 80$. This can be explained by the state evolution of the AMP in [4] which requires that L grows faster than M to fully exploit the benefit of multiple antennas.

VI. Conclusion

This paper studies a grant-free random access scheme for mMTC with sporadically active devices. Each active device transmits its identification and payload by selecting a signature sequence from a pre-assigned set, and the BS detects the device activity and the data by detecting their sequences. The paper formulates the detection problem as a maximum likelihood estimation problem based on the covariance matrix of the received signal, and employs the coordinate descent algorithm to solve the problem. The main contribution of this paper is a method to analyze the probabilities of false alarm and missed detection in the massive MIMO regime by exploiting the asymptotic properties of the maximum likelihood estimator. Simulation results validate the analysis, and also show that as compared to the AMP based method from compressive sensing, the covariance based method is much better at making use of the multiple antennas to improve reliability especially when the signature length is short.

Appendix A
Derivation of the Fisher Information Matrix

Let $\mathcal{L}(\boldsymbol{\gamma}) \triangleq \log p(\mathbf{Y}|\boldsymbol{\gamma})$. Based on (5), the second order derivative of $\mathcal{L}(\boldsymbol{\gamma})$ with respect to γ_i and γ_j is

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j} = -M \frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial \gamma_i \partial \gamma_j} - \frac{\partial^2 \left(\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1} \mathbf{Y}\mathbf{Y}^H \right) \right)}{\partial \gamma_i \partial \gamma_j}, \quad (15)$$

where the first term in the right hand side can be derived as

$$M \frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial \gamma_i \gamma_j} = -M \, \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1} \mathbf{s}_j \mathbf{s}_j^H \boldsymbol{\Sigma}^{-1} \mathbf{s}_i \mathbf{s}_i^H \right), \quad (16)$$

and the second term can be computed as

$$\begin{aligned} \frac{\partial^2 \left(\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1} \mathbf{Y}\mathbf{Y}^H \right) \right)}{\partial \gamma_i \partial \gamma_j} &= -\frac{\partial \, \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1} \mathbf{s}_i \mathbf{s}_i^H \boldsymbol{\Sigma}^{-1} \mathbf{Y}\mathbf{Y}^H \right)}{\partial \gamma_j} \\ &= \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1} \mathbf{s}_j \mathbf{s}_j^H \boldsymbol{\Sigma}^{-1} \mathbf{s}_i \mathbf{s}_i^H \boldsymbol{\Sigma}^{-1} \mathbf{Y}\mathbf{Y}^H \right) \\ &\quad + \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1} \mathbf{s}_i \mathbf{s}_i^H \boldsymbol{\Sigma}^{-1} \mathbf{s}_j \mathbf{s}_j^H \boldsymbol{\Sigma}^{-1} \mathbf{Y}\mathbf{Y}^H \right). \end{aligned} \quad (17)$$

Combining (16) and (17), and taking the expectation with respect to $\mathbf{Y}$ by using $\mathbf{E}\left[\mathbf{Y}\mathbf{Y}^H\right] = \mathbf{E}\left[\sum_{m=1}^{M} \mathbf{y}_m \mathbf{y}_m^H\right] = M\boldsymbol{\Sigma}$, we get the (i, j)-th entry of the Fisher information matrix as

$$-\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j} \right] = M(\mathbf{s}_i^H \boldsymbol{\Sigma}^{-1} \mathbf{s}_j)(\mathbf{s}_j^H \boldsymbol{\Sigma}^{-1} \mathbf{s}_i), \quad (18)$$

based on which $\mathbf{J}(\boldsymbol{\gamma})$ can be written in a matrix form as (12).

References

[1] ITU-R, "ITU-R M.[IMT-2020.TECH PERF REQ] - minimum requirements related to technical performance for IMT2020 radio interface(s)," Report ITU-R M.2410-0, Nov. 2017.

[2] L. Liu, E. G. Larsson, W. Yu, P. Popovski, Č. Stefanović, 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.

[3] 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.

[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.*, 2018.

[6] S. Haghighatshoar, P. Jung, and G. Caire, "Improved scaling law for activity detection in massive MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 381–385.

[7] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 477–487, Feb. 2011.

[8] K. R. Narayanan and H. D. Pfister, "Iterative collision resolution for slotted ALOHA: An optimal uncoordinated transmission policy," in *Proc. Int. Symp. Turbo Codes Iterative Inf. Process. (ISTC)*, Aug. 2012, pp. 136–139.

[9] E. Paolini, Č. Stefanović, G. Liva, and P. Popovski, "Coded random access: Applying codes on graphs to design random access protocols," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 144–150, Jun. 2015.

[10] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2528–2532.

[11] H. F. Schepker, C. Bockelmann, and A. Dekorsy, "Exploiting sparsity in channel and data estimation for sporadic multi-user communication," in *Proc. Int. Symp. Wireless Commun. Sys. (ISWCS)*, Aug. 2013, pp. 1–5.

[12] G. Wunder, H. Boche, T. Strohmer, and P. Jung, "Sparse signal processing concepts for efficient 5G system design," *IEEE Access*, vol. 3, pp. 195–208, Feb. 2015.

[13] W. Yu, "On the fundamental limits of massive connectivity," in *Proc. Inf. Theory Appl. (ITA) Workshop*, Feb. 2017, pp. 1–6.

[14] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample ML techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, S. S. Haykin, J. Litva, and T. J. Shepherd, Eds. New York: Springer-Verlag, 1993, pp. 99–151.

[15] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[17] S. G. Selfand and K.-Y. Liang, "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions," *J. Am. Stat. Assoc.*, vol. 82, no. 398, pp. 605–610, 1987.

[18] D. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.