Distributed Pilot Design for Massive Connectivity in Cellular Networks

Xihan Chen [1],An Liu [1], Wei Yu [2], Hei Victor Cheng [2], Kaiming Shen [2], and Min-Jian Zhao [1]
[1] *College of ISEE, Zhejiang University, Hangzhou, China*
[2] *Department of Electrical and Computer Engineering, University of Toronto, ON M5S 3G4, Canada*
Emails: [1]{chenixhan, anliu, mjzhao}@zju.edu.cn, [2]{weiyu, hei.cheng, kshen}@utoronto.ca

***Abstract*—Massive connectivity is regarded as a key requirement for future networks to support new communication paradigms, where the human-type communications coexist with machine-type communications. Owing to the limited coherence time but the huge number of potential devices, it is impossible to allocate mutually orthogonal pilot sequence for all potential devices, which may impose severe interference on the device activity detection and channel estimation. Existing nonorthogonal pilot design methods for conventional cellular network are not suitable for the massive connectivity regime. To overcome this challenge, we first formulate the pilot sequences design as an optimization problem to minimize the average mean square error (MSE) of channel estimation under the individual power constraint. The proposed optimization problem is nonconvex and highly coupled. By exploiting some approximation techniques, we convert the problem into a more tractable form and subsequently develop a distributed algorithm based on the matrix fractional programming (FP) and the alternating direction method of multipliers (ADMM) methods. Simulations validates that the proposed scheme not only achieves significant gains in channel estimation over state-of-the-art baseline schemes, but also improves the device activity detection performance.**

I. Introduction

During the last decade, the explosive growth in the number, type and functionality of smart mobile devices has spurred the development of new mobile services. Specifically, massive connectivity is regarded as a key requirement for future networks to support new communication paradigms, where the human-type communications (HTC) coexist with machine-type communications (MTC). Different from conventional HTC, the distinguishing features of MTC are the huge number of potential devices and the sporadic traffic pattern of each devices, which pose very stringent requirements on access schemes and scheduling overhead [1], [2]. Therefore, the existing grant-based access scheme and resource allocation strategies for the conventional cellular networks cannot be applied for massive connectivity scenario. In other words, the grant-free random access scheme is more amenable and favorable in massive connectivity regime.

Under the grant-free random access scheme, each device is preassigned with a unique pilot sequence to enable accurate device activity detection and channel estimation. Owing to the limited coherence time but the huge number of potential devices, it is impossible to allocate mutually orthogonal pilot sequence for all devices. As a result, the nonorthogonality of the pilot sequences may impose severe interference on the device activity detection and channel estimation, which further affects the availability and reliability of communications between the devices and the BS [3]. To overcome this issue, a growing body of literatures have recently proposed various methods. By exploiting the sparsity in the device activity pattern, the joint device detection and channel estimation problem can be formulated as a compressed sensing problem, which can be efficiently solved by applying approximate message passing (AMP) technique [3], [4]. Further, to accommodate larger number of devices within limited coherence time, covariance based approach have been proposed [5] for cases in which massive MIMO has been deployed at the BS.

It should be emphasized that most existing joint device activity detection and channel estimation algorithms only consider the independent and identically distributed (i.i.d.) random Gaussian pilots. But this may be far from optimal, especially with the coexistence of HTCs and MTCs. For example, HTC and MTC devices can have different transmit power and access probabilities and may require different types of pilots.

This paper considers the optimal pilot design problem for grant-free nonorthogonal multiple-access (GF-NOMA). As observed in earlier work, in the massive multiple-input multiple-output (MIMO) regime, perfect device activity detection can often be guaranteed with high probability, but channel estimation error remains [6]. Hence, this paper focuses on minimizing the channel estimation error in the massive MIMO regime by optimizing the pilot sequences. Interestingly, although the objective function is chosen to be the channel estimation error, simulations show that the device detection performance can also be improved by the proposed pilot sequence design. Specifically, we propose a distributed algorithm to optimize the pilot sequences for the uplink transmission of massive connectivity scenario, to alleviate the performance bottleneck by the sporadic traffic pattern and the multi-device interference. The proposed scheme achieves better performance than the baselines by exploiting the distinct characteristics of the potential devices.

II. System Model

A. Network Architecture

Consider the uplink communication of a single-cell massive connectivity scenario, as shown in Fig. 1, where the BS is equipped with massive M antennas to serve K potential devices. As in [7], we also consider that all potential devices consist of K_h HTC devices and $K_m = K - K_h$ machine-type devices. Without loss of generality, we assume that the indexes $k \in \mathcal{K}_h \triangleq \{1, \cdots, K_h\}$ refers to HTC devices, while $k \in \mathcal{K}_m \triangleq \{K_h + 1, \cdots, K\}$ refers to machine-type devices. Both HTC and MTC devices are equipped with single antennas. To further capture the traffic nature of such heterogeneous scenario, we thereby make the following reasonable assumptions: (i) all devices are fully synchronized; (ii) each device accesses the channel with probability τ_k in an i.i.d fashion [4]; (iii) the average access probability of HTC devices is much larger than that of MTC device, while the number of MTC devices is much larger than the number of HTC devices, i.e., $K_m \gg K_h$ [7]. Specifically, we focus on a coherence time interval of channel where channel is assumed to be constant. Each coherence time interval is divided into S time slots, where S is chosen to meet the latency requirement.

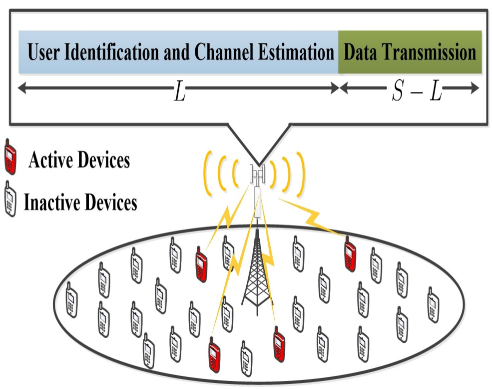

Fig. 1. An illustration of massive connectivity scenario.

Without loss of generality, the uplink channel of each device is assumed to be a block-fading channel under a narrowband assumption. But the proposed algorithm can be easily modified to cover the wideband system as well. In this case, the received signal at BS of the s-th time slot is given by

$$\mathbf{y}(s) = \sum_{k=1}^{K} a_k \boldsymbol{h}_k x_k(s) + \boldsymbol{w}(s), \forall t = 1, \cdots, S, \tag{1}$$

where a_k is the activity indicator for device k, $\boldsymbol{h}_k \in \mathbb{C}^M$ is the channel vector between device k and the BS, $x_k(s) \in \mathbb{C}$ is the transmit signal by the k-th device in the s-th time slot, $\boldsymbol{w}(s)$ is the additive complex Gaussian noise vector with i.i.d entries distributed as $\mathcal{CN}(0, \sigma_w^2)$.

For convenience, we define $\mathbf{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^T \in \mathbb{C}^{K \times M}$, $\mathbf{X} = [\boldsymbol{x}(1), \ldots, \boldsymbol{x}(S)]^T \in \mathbb{C}^{S \times K}$ with $\boldsymbol{x}(s) = [x_1(s), \cdots x_K(s)]^T \in \mathbb{C}^K$, $\mathbf{W} = [\boldsymbol{w}(1), \cdots, \boldsymbol{w}(S)]^T \in \mathbb{C}^{S \times M}$, $\mathbf{Y} = [\mathbf{y}(1), \ldots \mathbf{y}(S)]^T \in \mathbb{C}^{S \times M}$, and $\mathbf{A} = \mathrm{diag}\,(a_1, \cdots, a_K) \in \mathbb{R}^{K \times K}$. Using the above notations, the received signal $\mathbf{Y}$ at the BS is expressed in matrix form as

$$\mathbf{Y} = \mathbf{XAH} + \mathbf{W}. \tag{2}$$

In this paper, we model the channel vector as a Gaussian distribution $\boldsymbol{h}_k = \sqrt{\beta_k}\boldsymbol{v}_k, \forall k$, where β_k reflects the pathloss and shadowing component, and $\boldsymbol{v}_k \sim \mathcal{CN}(0, \mathbf{I}_{M \times M}) \in \mathbb{C}^M$ is the Rayleigh fading component. Namely, we can obtain $\boldsymbol{h}_k \sim \mathcal{CN}(0, \beta_k \mathbf{I}_{M \times M})$. Following the pioneering works on the massive connectivity scenario [3], [4], we also focus on the scenario where the devices are stationary, so that the path-loss and shadowing can be estimated and stored at the BS as prior information. In other words, the BS knows the channel statistics of each user rather than real-time channel state information (CSI).

B. Grant-Free Nonorthogonal Multiple-Access Scheme

In this paper, we adopt the GF-NOMA scheme to enable multiple devices share the allotted spectrum in the most effective manner. Under the GF-NOMA scheme, the transmission in each particular block is divided into two phases. In the first phase, each device is assigned with a specific pilot sequence and transmits its respective pilot sequence to the BS simultaneously. In subsequent phase, the BS jointly identifies the active devices and estimates the channels for the active devices. Based on the knowledge of device activities and channels obtained in the first phase, the active devices begin to send data as in a conventional multiple-access channel with a fixed number of transmitters and a single receiver [8]. Hence, this paper will mostly focus on the design of the pilot sequences in the first phase. In this case, the received signal at BS over the first phase is given by

$$\mathbf{Y}_{1:L} = \mathbf{\Phi AH} + \mathbf{W}_{1:L}, \tag{3}$$

where $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_K] \in \mathbb{C}^{L \times K}$ with $\boldsymbol{\phi}_k = [\phi_{k,1}, \cdots, \phi_{k,L}]^T \in \mathbb{C}^L$ is the matrix for pilot sequences, and L is the length of the pilot sequence.

III. Pilot Optimization Formulation for MTC

A. Problem Formulation

From Theorem 4 in [4], it is proved that even though assigning random Gaussian pilot sequences to all devices, accurate device activity detection is also guaranteed with high probability in the massive MIMO regime for a massive connectivity MTC, but channel estimation error remains. Hence, we assume that accurate device activity detection is performed for specific device activity pattern in this paper, and focus on minimizing the average channel estimation error by optimizing the pilot sequences. As will be verified in the simulations, such a metric will also help to improve the device detection performance because the average channel estimation error also depends on the probability distribution of device activity.

Under the assumption that accurate device activity detection is performed for specific device activity pattern $\boldsymbol{a} \triangleq [a_1, \cdots, a_K]$, we define the MMSE estimate for channel as $\hat{\mathbf{H}}(\boldsymbol{a})$ and corresponding MSE

$$\mathbf{V} \triangleq \mathbb{E}_{\boldsymbol{a}} \left[\mathrm{MSE}(\hat{\mathbf{H}}; \boldsymbol{a}) \right] = \sum_{r=1}^{R} \rho_r \mathrm{MSE}(\hat{\mathbf{H}}; r), \tag{4}$$

where $\mathrm{MSE}(\hat{\mathbf{H}}; \boldsymbol{a}) \triangleq \mathbb{E}_{\mathbf{H}} \left[||\hat{\mathbf{H}}(\boldsymbol{a}) - \mathbf{H}||^2 \right]$ is the MSE for device activity pattern $\boldsymbol{a}$, $\rho_r \triangleq \prod_{k=1}^{K} \tau_k^{a_{r,k}} (1 - \tau_k)^{1 - a_{r,k}}$ is the probability of r-th device activity pattern, and $a_{r,k} \in \{0,1\}$ is the k-th bit of r in binary form[1], $R = 2^N$ is the total number of all possible device activity pattern, $\text{MSE}(\hat{\mathbf{H}};r) \triangleq \mathbb{E}_{\mathbf{H}}\left[||\hat{\mathbf{H}}(r) - \mathbf{H}||^2\right]$ is the MSE for r-th device activity pattern.

It is noted that for fixed device activity pattern r, the corresponding MMSE estimate of $\hat{\mathbf{H}}(r)$ is given by

$$\left(\mathbf{I}_{M\times M} \bigotimes \mathbf{\Phi}\mathbf{A}(r)\mathbf{G}\right)\left(\mathbf{I}_{M\times M} \bigotimes \mathbf{Q}(r)\right)^{-1} \text{vec}(\mathbf{Y}_{1:L}),$$

where $\mathbf{G} \triangleq \text{diag}(\beta_1, \cdots, \beta_K) \in \mathbb{R}^{K\times K}$ is the large-scale channel strength matrix, and $\mathbf{Q}(r) \triangleq \sigma_w^2\mathbf{I}_{L\times L} + \mathbf{\Phi}\mathbf{A}(r)\mathbf{G}\mathbf{A}(r)\mathbf{\Phi}^H \in \mathbb{C}^{L\times L}$ is the covariance matrix of the received signal at BS for the r-th device activity pattern. Using standard Kronecker product properties, we obtain a closed-form expression for $\text{MSE}(\hat{\boldsymbol{h}};r)$ as

$$\text{MSE}(\hat{\boldsymbol{h}};r) \triangleq \left(\overline{\mathbf{R}}(r) - \overline{\mathbf{R}}(r)\mathbf{\Phi}^H\mathbf{Q}^{-1}(r)\mathbf{\Phi}\overline{\mathbf{R}}(r)\right),$$

where $\overline{\mathbf{R}}(r) \triangleq \mathbf{A}(r)\mathbf{G}\mathbf{A}(r) \in \mathbb{C}^{K\times K}$ is the covariance matrix of effective channel at BS for the r-th device activity pattern.

Using the above notations, the pilot sequences design for massive connectivity scenario can be formulated as the following power-constrainted average MMSE minimization problem (PC-AMP):

$$\begin{aligned} &\min_{\mathbf{\Phi}} \sum_{r=1}^{R} \rho_r \text{Tr}\left\{\overline{\mathbf{R}}(r) - \overline{\mathbf{R}}(r)\mathbf{\Phi}^H\mathbf{Q}^{-1}(r)\mathbf{\Phi}\overline{\mathbf{R}}(r)\right\} \\ &\text{s.t. } \text{Tr}(\boldsymbol{\phi}_k\boldsymbol{\phi}_k^H) \le P_k, \forall k = 1, \cdots, K, \end{aligned} \tag{5}$$

where P_k is the individual power budget at each user. Since the matrix $\overline{\mathbf{R}}(r)$ is independent of the optimization variable $\mathbf{\Phi}$, problem (5) can be equivalently written as

$$\begin{aligned} &\max_{\mathbf{\Phi}} \sum_{r=1}^{R} \rho_r \text{Tr}\left\{\overline{\mathbf{R}}(r)\mathbf{\Phi}^H\mathbf{Q}^{-1}(r)\mathbf{\Phi}\overline{\mathbf{R}}(r)\right\} \\ &\text{s.t. } \text{Tr}(\boldsymbol{\phi}_k\boldsymbol{\phi}_k^H) \le P_k, \forall k = 1, \cdots, K, \end{aligned} \tag{6}$$

B. Problem Approximation

From (6), we can find that there are $R = 2^K$ sum terms in the expression of average MSE, which leads to exponential computational complexity. To address the above issue, we must resort to approximation techniques to make problem (6) more tractable. In this paper, we propose the following approximation

$$\begin{aligned} &\sum_{r=1}^{R} \rho_r \text{Tr}\left\{\overline{\mathbf{R}}(r)\mathbf{\Phi}^H\mathbf{Q}^{-1}(r)\mathbf{\Phi}\overline{\mathbf{R}}(r)\right\} \\ &= \sum_{i=1}^{2^{K_h}} \rho_i^h \sum_{j=1}^{2^{K_m}} \rho_j^m \text{Tr}\left\{\overline{\mathbf{R}}(r_{i,j})\mathbf{\Phi}^H\mathbf{Q}^{-1}(r_{i,j})\mathbf{\Phi}\overline{\mathbf{R}}(r_{i,j})\right\} \\ &\approx \sum_{i=1}^{I_h} \rho_i^h \text{Tr}\left\{\overline{\mathbf{R}}(i)\mathbf{\Phi}^H\overline{\mathbf{Q}}^{-1}(i)\mathbf{\Phi}\overline{\mathbf{R}}(i)\right\}, \end{aligned} \tag{7}$$

where $(r_{i,j})_2 \triangleq (i)_2(j)_2$, $\rho_i^h = \prod_{k=1}^{K_h}\tau_k^{a_{i,k}}(1-\tau_k)^{1-a_{i,k}}$ with $\sum_{i=1}\rho_i^h = 1$ is the probability of activity pattern for human-type device, $\rho_j^m = \prod_{k=K_h+1}^{K}\tau_k^{a_{j,k}}(1-\tau_k)^{1-a_{j,k}}$ with $\sum_{j=1}\rho_j^m = 1$ is the probability of activity pattern for machine-type device, $I_h = 2^{K_h}$ is the total number of HTC device activity pattern, $\widetilde{\mathbf{A}}(i) = \text{diag}\left(a_{i,1}, \cdots, a_{i,K^h}, \tau_{K^h+1}, \cdots, \tau_K\right)$ is the approximated activity pattern for all potential devices, $\overline{\mathbf{R}}(i) \triangleq \widetilde{\mathbf{A}}(i)\mathbf{G}\widetilde{\mathbf{A}}(i)$ is the covariance matrix of effective channel vector at BS for the i-th activity pattern for HTC device, and $\overline{\mathbf{Q}}(i) \triangleq \sigma_w^2\mathbf{I}_{L\times L} + \mathbf{\Phi}\overline{\mathbf{R}}(i)\mathbf{\Phi}^H$ is the covariance matrix of the received signal at BS for the i-th activity pattern for HTC device.

Note that the idea behind the above approximation is due to the fact that the average access probability of HTC devices is much larger than that of MTC device, while the number of MTC devices is much larger than the number of HTC devices. In this case, we approximate the activity indicator for MTC devices under any device activity pattern by the average access probability $\tau_k, \forall k \in \mathcal{K}_m$. In the special case when there only exists MTC device in the system, the approximated MSE in (7) is equivalent to the MSE achieved by applying linear MMSE estimator. Based on (7), the approximated optimization problem is given by:

$$\begin{aligned} &\max_{\mathbf{\Phi}} \sum_{i=1}^{I_h} \rho_i^h \text{Tr}\left\{\overline{\mathbf{R}}(i)\mathbf{\Phi}^H\overline{\mathbf{Q}}^{-1}(i)\mathbf{\Phi}\overline{\mathbf{R}}(i)\right\} \\ &\text{s.t. } \text{Tr}(\boldsymbol{\phi}_k\boldsymbol{\phi}_k^H) \le P_k, \forall k = 1, \cdots, K. \end{aligned} \tag{8}$$

IV. Distributed Pilot Design

There are several challenges in finding stationary solutions of problem (8), elaborated as follows. First, problem (8) is typically a function of multiple-ratio fractional programming (FP), where the optimization variable $\mathbf{\Phi}$ appear in both numerator and the denominator. In particular, solving the above multiple-ratio FP is always NP-hard. Second, the problem size will grow exponentially as the number of HTC devices increases. In other words, the traditional centralized optimization algorithms cannot be applied to here.

To the best of our knowledge, there lacks an efficient and distributed algorithm to handle such high-dimension non-convex optimization problem (8). In this section, we propose a distributed FP-ADMM algorithm to find stationary solutions to problem (8). We shall first outline the proposed FP-ADMM algorithm. Then, we elaborate the implementation details.

A. Algorithm Outline

In this subsection, we propose a distributed algorithm that solves problem (8) to a stationary solution. To this end, we propose to integrate two existing algorithms. The first one is the matrix quadratic transform based algorithm developed in [9], which is used for fractional programming (FP) with multiple ratios. By applying the above matrix FP algorithm, it addresses the issue of nonlinear fractional and high coupling. The second one is the well-known ADMM algorithm for large-scale optimization [10]. The key advantage of ADMM algorithms is well-suited for distributed and parallel

[1]The binary form of r is given by $(r)_2 = a_{r,1}a_{r,2}\cdots a_{r,K}$,and $a_{r,k}$ is a constant for the fixed device activity pattern

implementation, which is attractive for the considered massive connectivity scenario. The proposed FP-ADMM algorithm is summarized in Algorithm 1.

Algorithm 1 FP-ADMM Algorithm for problem (8)

Initialization: Let $j = 0$, $\|\phi_k\|^2 = P_k, \forall k, \mathbf{Z}_i = \mathbf{\Phi}, \forall i$, and $\mathbf{E}_i = \mathbf{0}, \forall i$. Define the maximum number of iterations J and the penalty parameter μ.

Step 1 (Update P): For $i = 1, ..., I_h$**, let**

$$\mathbf{P}^{j+1}(i) = \overline{\mathbf{Q}}^{-1}(i)\mathbf{\Phi}\overline{\mathbf{R}}_g(i). \tag{9}$$

Step 2 (Update $\mathbf{Z}_i$**): Let** $\mathbf{Z}_i^{j+1} = \mathbf{Z}_i^*, \forall i$, where $\mathbf{Z}_i^*$ is given in (22).

Step 3 (Update $\mathbf{\Phi}$**): For** $k = 1, ..., K$**, let**

$$\phi_k^{j+1} = \phi_k^*(\lambda_k), \tag{10}$$

where $\phi_k^*(\lambda_k)$ is given in (24).

Step 4 (Update $\mathbf{E}_i$**): Let** $\mathbf{E}_i^{j+1} = \mathbf{E}_i^j + \mathbf{Z}_i^{j+1} - \mathbf{\Phi}^{j+1}, \forall i$.

Let $j = j + 1$. If $j = J$, **terminate** the algorithm and **output** $\mathbf{\Phi}^J$. Otherwise, **go to** Step 1.

B. Matrix FP Method

In this part, we first apply the matrix FP method to reformulate the problem (8) as a convex optimization problem.

Lemma 1. *Define* $\Omega \triangleq \left\{\mathbf{\Phi} | \, Tr(\phi_k \phi_k^H) \leq P_k\right\}$. *Then* $\mathbf{\Phi}^*$ *solves the problem in (8) if and only if it solves*

$$\max_{\mathbf{\Phi}\in\Omega} f\left(\mathbf{\Phi}\right) \triangleq \sum_{i=1}^{I_h} \rho_i^h Tr \ \left\{2\Re\left\{\overline{\mathbf{R}}(i)\mathbf{\Phi}^H\mathbf{P}(i)\right\} - \mathbf{P}^H(i)\overline{\mathbf{Q}}(i)\mathbf{P}(i)\right\}, \tag{11}$$

where $\mathbf{P}(i) = \mathbf{P}^*(i) \in \mathbb{C}^{L\times K}$ *is the auxiliary variable introduced for each matrix ratio term given by*

$$\mathbf{P}^*(i) \triangleq \overline{\mathbf{Q}}^{-1}(i)\mathbf{\Phi}\overline{\mathbf{R}}(i). \tag{12}$$

Proof: Lemma 1 can be proven by a similar approach as in [11]. It is clear that $f\left(\mathbf{\Phi}, \mathbf{P}\right)$ is concave over $\mathbf{P}$ while fixing $\mathbf{\Phi}$ and also analytic in complex region. Applying the first-order optimal condition, we obtain the optimal $\mathbf{P}^*(i)$ as $\mathbf{Q}^{-1}(i)\mathbf{\Phi}\overline{\mathbf{R}}(i)$. By simply substituting this optimal $\mathbf{P}^*(i)$ back into problem (11), we can have the equivalent problem (8). As a result, problem (8) is equivalent to problem (11). ∎

It is noteworthy that the numerator and denominator in (8) are now decoupled in the reformulated problem (11).

C. Consensus ADMM Algorithm for Solving problem (11)

In this part, we apply the consensus ADMM algorithm for solving problem (11). We emphasize that the difficulty in designing the distributed algorithm comes from the variable coupling in $f\left(\mathbf{\Phi}\right)$, and the large number of all possible device activity pattern. To overcome this difficulty, we first introduce auxiliary variables and then design a special variable splitting scheme in order to separate the variable coupling in $f\left(\mathbf{\Phi}\right)$. Following this, we elaborate the implementation details of the consensus ADMM-based algorithm.

We first introduce auxiliary variable $\mathbf{Z}_i \triangleq [\mathbf{Z}_{i,1}, \cdots, \mathbf{Z}_{i,K}] \in \mathbb{C}^{L\times K}$ to reformulate problem (11) into a consensus form, which is given by

$$\begin{aligned} \max_{\mathbf{\Phi},\mathbf{Z}} & \sum_{i=1}^{I_h} f_i\left(\mathbf{Z}_i\right) \\ \text{s.t. } & \text{Tr}(\phi_k\phi_k^H) \leq P_k, \forall k = 1, \cdots, K, \\ & \mathbf{Z}_i = \mathbf{\Phi}, \forall i = 1, \cdots, I_h, \end{aligned} \tag{13}$$

where

$$f_i\left(\mathbf{Z}_i\right) \triangleq \rho_i^h \text{Tr} \ \left\{2\Re\left\{\overline{\mathbf{R}}(i)\mathbf{Z}_r^H\mathbf{P}(i)\right\} - \mathbf{P}^H(i)\overline{\mathbf{Q}}(i, \mathbf{Z}_i)\mathbf{P}(i)\right\}, \tag{14}$$

is the cost function for the i-th device activity pattern, and

$$\overline{\mathbf{Q}}(i, \mathbf{Z}_i) \triangleq \sigma_w^2\mathbf{I}_{LM} + \mathbf{Z}_i\overline{\mathbf{R}}(i)\mathbf{Z}_i^H, \tag{15}$$

is the covariance matrix of the total received signal at BS for the i-th device activity pattern.

By moving the additional equality constraints $\mathbf{Z}_i = \mathbf{\Phi}, \forall i$, into the objective function of problem (13), we can obtain the corresponding augmented Lagrangian function and it can be expressed as

$$L(\boldsymbol{\phi}, \boldsymbol{z}, \boldsymbol{e}; \boldsymbol{p}) = \sum_{i=1}^{I_h} f_i\left(\mathbf{Z}_i\right) - \mu \left\|\mathbf{Z}_i - \mathbf{\Phi} + \mathbf{E}_i\right\|^2, \tag{16}$$

where $\mu \geq 0$ is the constant penalty parameter for adjusting the convergence speed of ADMM, $\mathbf{E}_i \triangleq [\mathbf{E}_{i,1}, \cdots, \mathbf{E}_{i,K}] \in \mathbb{C}^{L\times K}$ is the Lagrangian multiplier corresponding to the equality constraint $\mathbf{Z}_i = \mathbf{\Phi}$.

Then the consensus ADMM iterative update equation for problem (13) can be subsequently obtained as

$$\mathbf{Z}_i^{j+1} = \arg \max_{\mathbf{Z}_i} f_i\left(\mathbf{Z}_i\right) - \mu \left\|\mathbf{Z}_i - \mathbf{\Phi}^j + \mathbf{E}_i^j\right\|^2, \tag{17}$$

$$\mathbf{\Phi}^{j+1} = \arg \min_{\mathbf{\Phi}\in\Omega} \mu \sum_{i=1}^{I_h} \left\|\mathbf{Z}_i^{j+1} - \mathbf{\Phi} + \mathbf{E}_i^j\right\|^2, \tag{18}$$

$$\mathbf{E}_i^{j+1} = \mathbf{E}_i^j + \mathbf{Z}_i^{j+1} - \mathbf{\Phi}^{j+1}, \forall i = 1, \cdots, I_h. \tag{19}$$

where j is the iteration index.

There are several advantages of the above consensus ADMM algorithm. First, the convergence of the above consensus ADMM algorithm is guaranteed when the optimization problem is convex and the penalty parameter μ is in positive regime [12]. Note that the reformulated problem (11) is a convex optimization problem after applying the matrix FP method. Second, each $\mathbf{Z}_i$ can be updated in parallel manner and the corresponding subproblem can be solved efficiently in closed form. In a nutshell, each step of the consensus ADMM algorithm is easily computable and amenable for distributed implementation. The choice of the initial point and the update equation for each variable is elaborated below.

1) Choice of Initial Point: For pilot sequence $\boldsymbol{\phi}$, we choose the initial point to be $\|\phi_k\|^2 = P_k, \forall k$, i.e., each device transmits at the maximum power. For all $\mathbf{Z}_i$ and $\mathbf{E}_i$, we choose the initial point to be $\mathbf{Z}_i = \mathbf{\Phi}, \forall i$, and $\mathbf{E}_i = \mathbf{0}, \forall i$, respectively.

2) Solving Subproblem (17): Since the first term in $\mathbf{Q}(i, \mathbf{Z}_i)$ is independent of $\mathbf{Z}_i$, we rewrite problem (17) as

$$\max_{\mathbf{Z}_i} \text{Tr} \left\{2\Re\left\{\overline{\mathbf{R}}(i)\mathbf{Z}_i^H\mathbf{P}(i)\right\} -\mathbf{P}^H(i)\mathbf{Z}_i\overline{\mathbf{R}}(i)\mathbf{Z}_i^H\mathbf{P}(i)\right\}\right\} - \mu\left\|\mathbf{Z}_i - \mathbf{\Phi}^j + \mathbf{E}_i^j\right\|^2. \quad (20)$$

It is seen that problem (20) is an unconstrained quadratic optimization, which can be efficiently solved by applying the first-order optimal condition. After some tedious calculations and appropriate rearrangement, it is shown that the first-order optimal condition can be expressed in a compact form as

$$\mathbf{B}_i\mathbf{Z}_i^*\mathbf{C} + \mathbf{D}\mathbf{Z}_i^* = \mathbf{F}_i, \quad (21)$$

where $\mathbf{B}_i \triangleq \mathbf{P}(i)\mathbf{P}^H(i)$,$\mathbf{C}_i \triangleq \overline{\mathbf{R}}(i)$, $\mathbf{D} \triangleq \mu\mathbf{I}_{L\times L}$ and $\mathbf{F}_i \triangleq \mathbf{P}(i)\overline{\mathbf{R}}(i) + \mu\left(\mathbf{\Phi}^j - \mathbf{E}_i^j\right)$. By vectorizing the both side of equation (21), we can obtain the optimal $\mathbf{Z}_i^*$ as

$$\text{vec}\left(\mathbf{Z}_i^*\right) = \mathbf{J}_i^{-1}\text{vec}\left(\mathbf{F}_i\right), \quad (22)$$

where $\mathbf{J}_i \triangleq \mathbf{C}_i^T \bigotimes \mathbf{B}_i + \mathbf{I}_K \bigotimes \mathbf{D}$.

3) Solving Subproblem (18): The subproblem (18) is given by

$$\min_{\mathbf{\Phi}} \sum_{i=1}^{I_h}\left\|\mathbf{Z}_i^{j+1} - \mathbf{\Phi} + \mathbf{E}_i^j\right\|^2 \quad (23)$$
$$\text{s.t. } \text{Tr}(\boldsymbol{\phi}_k\boldsymbol{\phi}_k^H) \leq P_k, \forall k = 1, \cdots, K, .$$

Clearly, problem (23) is an quadratically constrained quadratic optimization problem. Thus, it can be solved by dealing with its dual problem. To this end, by introducing Lagrange multiplier λ_k for the corresponding constraint $\text{Tr}(\boldsymbol{\phi}_k\boldsymbol{\phi}_k^H) \leq P_k$, we define the Lagrangian associated with problem (23) as

$$L(\mathbf{\Phi}, \lambda) \triangleq \sum_{i=1}^{I_h}\left\|\mathbf{Z}_i^{j+1} - \mathbf{\Phi} + \mathbf{E}_i^j\right\|^2 + \sum_{k=1}^{K}\lambda_k\left(\text{Tr}(\boldsymbol{\phi}_k\boldsymbol{\phi}_k^H) - P_k\right)$$
$$= \sum_{k=1}^{K}\left(\sum_{i=1}^{I_h}\left\|\mathbf{Z}_{i,k}^{j+1} - \boldsymbol{\phi}_k + \mathbf{E}_{i,k}^j\right\|^2 + \lambda_k\left(\text{Tr}(\boldsymbol{\phi}_k\boldsymbol{\phi}_k^H) - P_k\right)\right)$$
$$= \sum_{k=1}^{K} L_k(\boldsymbol{\phi}_k, \lambda_k)$$

Since $L_k(\boldsymbol{\phi}_k, \lambda_k)$ with respect to $\boldsymbol{\phi}_k$ for fixed Lagrange multiplier is an unconstrained quadratic optimization problem, it can be efficiently solved by checking its first-order optimal condition, which yields the optimal $\boldsymbol{\phi}_k^*$ as

$$\boldsymbol{\phi}_k^*(\lambda) = \frac{1}{I_h + \lambda_k}\sum_{i=1}^{I_h}\left(\mathbf{Z}_{i,k}^{j+1} + \mathbf{E}_{i,k}^j\right), \quad (24)$$

where λ is chosen to be zero if $||\boldsymbol{\phi}_k^*(0)||^2 \leq P_k$ and chosen to satisfy $||\boldsymbol{\phi}_k^*(\lambda)||^2 = P_k$ otherwise.

D. Implementation Consideration

Using graphic processing units (GPUs) implemented at the BS, we can apply parallel processing to accelerate the computation speed. But in order to maximize the efficiency of GPU, the algorithm needs to have some good distributed features such as those provided by the proposed algorithm. For example, from (17) and (19), we can see that the $\mathbf{Z}_i$ and $\mathbf{E}_i$ can be updated in a parallel manner, which further speeds up the calculations under the GPU implementation.

V. SIMULATION RESULTS AND DISCUSSIONS

In this section, we evaluate the numerical performance of the proposed pilot design in a cell of radius 250 m. There are 20 HTC devices and 80 MTC devices randomly distributed in the cell. The path-loss for device k is modeled as $15.3 + 37.6\log 10\left(\text{d}_k\right)$ in dB [13], and d_k is the distance between BS and device k in meters. The shadowing is assumed to be a Gaussian distribution with zero mean and variance $\sigma_S^2 = 8$ [3]. The background noise is -169 dBm/Hz over 10 MHz. At any given time slot, the access probability of human-type devices follows a uniform distribution over the interval $[0.4, 0.6]$, while the access probability of machine-type devices follows a uniform distribution over the interval $[0.1, 0.2]$. Unless otherwise specified, we consider $M = 64$ antennas and the length of pilot sequence is $L = 30$. The transmit power constraint for each device is $P_k = 23$ dBm. Two baseline schemes are considered for comparison: the random scheme [4] and the linear MMSE scheme [14] obtained by approximating the activity indicator for all potential devices by their average access probability.

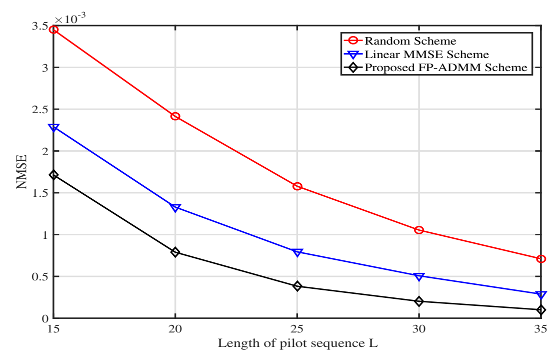

Fig. 2. NMSE performance versus the length of pilot sequence L.

In Fig. 2, we plot the NMSE performance versus the length of pilot sequence L for the different schemes. We can see that the proposed FP-ADMM scheme achieves significant gain over all the other competing schemes. This is due to the fact that the proposed FP-ADMM can make full use of the statistical information of devices with higher average access probability. When the length of pilot sequence L increases, the performance gap between the proposed FP-ADMM scheme and other schemes becomes larger. In Fig. 3, we plot the cumulative distribution function (CDF) of NMSE for the

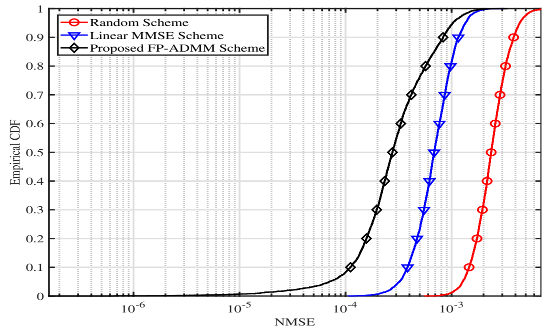

Fig. 3. Cumulative distribution function of the NMSE.

different schemes. The MSE achieved by the proposed FP-ADMM outperforms other schemes, and concentrates on a small range.

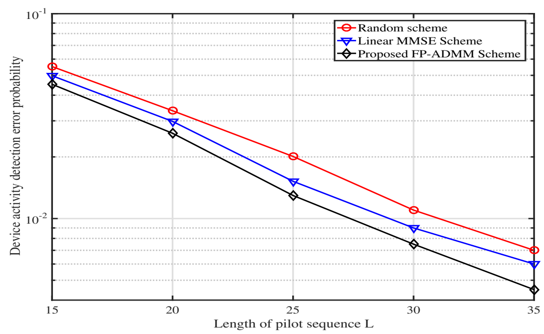

Fig. 4. Device activity detection error probability versus the length of the pilot sequences

In Fig. 4, we plot the device activity detection error probability versus the length of the pilot sequences. Specifically, the error probability of device activity detection corresponds to two types of error events, i.e., missed detection and false alarm, respectively. The missed detection occurs when an active device is viewed as in the inactive status, while the false alarm occurs when a inactive device is declared active. It is observed that as L increases, the error probability of all schemes decreases and gradually gets closer to zero. In other words, near-perfect device activity detection is guaranteed in the massive MIMO regime, which is consistent with the theoretical analysis of [4]. This indicates that our proposed FP-ADMM scheme not only achieve better channel estimation over all the other competing schemes, but also guarantee superior performance of the device activity detection.

Table I compares the performance of different schemes in terms of the average mutual coherence. The mutual coherence is defined as the degree of orthogonality between pilot sequences of different devices, i.e., $\kappa_{j,n} \triangleq \frac{|\phi_j^H \phi_n|}{|\phi_j||\phi_n|}$. According to the strength of channel pathloss, we further divide all potential devices into four classes: 1) HTC devices with large pathloss; 2) HTC devices with small pathloss; 3) MTC devices with large pathloss; 4) MTC devices with small pathloss. In the Table I, we let κ_L^h, κ_S^h, κ_L^m, and κ_S^m denote the average mutual coherence of corresponding class of devices, respectively. It is observed that different from other competing schemes, the proposed FP-ADMM scheme allocates more orthogonal pilots to the HTC devices with large pathloss, while assigning more nonorthogonal pilots to the MTC devices with small pathloss. This demonstrates that the proposed FP-ADMM schme can achieve better pilots allocation according to the distinic characteristics of devices.

	κ_L^h	κ_S^h	κ_L^m	κ_S^m
FP-ADMM Scheme	0.1442	0.4673	0.3028	0.7175
Linear Scheme	0.1984	0.4146	0.3422	0.6343
Random Scheme	0.5030	0.5006	0.5134	0.5389

TABLE I
THE AVERAGE MUTUAL COHERENCE COMPARISON

VI. CONCLUSIONS

In this paper, we consider the distributed design of pilot sequences for massive connectivity in cellular networks. We formulate the optimization of pilot sequences as the PC-AMP, and propose a distributed FP-ADMM algorithm to find stationary solutions of this non-convex problem. Simulations verify that the proposed FP-ADMM algorithm achieves better channel estimation MSE over existing solutions.

REFERENCES

[1] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep 2016.

[2] L. Liu, E. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. 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, April 2018.

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

[5] P. J. S. Haghighatshoar and G. Caire, "Improved scaling law for activity detection in massive MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June 2019, pp. 381–385.

[6] L. Liu and W. Yu, "Massive connectivity with Massive MIMO – Part II: Achievable rate characterization," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2947–2959, June 2018.

[7] K. Senel, E. Bjornson, and E. Larsson, "Human and machine type communications can coexist in uplink massive MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, April 2018.

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

[9] K. Shen and W. Yu, "Fractional programming for communication systems – Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, Mar 2018.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn*, vol. 3, no. 1, pp. 1–122, 2011.

[11] K. Shen and W. Yu, "Fractional programming for communication systems – Part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, Mar 2018.

[12] K. Huang and N. D.Sidiropulos, "Consensus-admm for general quadratically constrained quadratic programming," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5297–5310, Oct 2016.

[13] *Technical Specification Group Radio Access Network; Further Advancements for E-UTRA Physical Layer Aspects*, 3GPP TR 36.814. [Online]. Available: http://www.3gpp.org

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