# Massive CSI Acquisition for Dense Cloud-RANs With Spatial-Temporal Dynamics

Xuan Liu, Yuanming Shi, *Member, IEEE*, Jun Zhang, *Senior Member, IEEE*, and Khaled B. Letaief, *Fellow, IEEE*

***Abstract*— Dense cloud radio access networks (cloud-RANs) provide a promising way to enable scalable connectivity and handle diversified service requirements for massive mobile devices. To fully exploit the performance gains of dense cloud-RANs, channel state information of both the signal link and interference links is required. However, with limited radio resources for training, the channel estimation problem in dense cloud-RANs becomes a high-dimensional estimation problem, i.e., the number of measurements will be typically smaller than the dimension of the channel. In this paper, we shall develop a generic high-dimensional structured channel estimation framework for dense cloud-RANs, which is based on a convex structured regularizing formulation. Observing that the wireless channel possesses ample exploitable statistical characteristics, we propose to convert the available spatial and temporal prior information into appropriate convex regularizers. Simulation results demonstrate that exploiting the spatial and temporal dynamics can achieve good estimation performance even with limited training resources. The alternating direction method of multipliers algorithm is further adopted to solve the resultant large-scale high-dimensional channel estimation problems. The proposed framework thus enjoys modeling flexibility, low training overhead, and computation cost scalability.**

***Index Terms*— Cloud-RANs, CSI, high-dimensional structured estimation, structured regularizers, ADMM, spatial and temporal dynamics, and massive device connectivity.**

## I. Introduction

CLOUD radio access network (Cloud-RAN) [2] has recently been proposed as a revolutionary architecture for future cellular networks to meet the exponential growth of mobile data traffic. In this new architecture, all the baseband signal processing will be shifted to a central location in a datacenter cloud, called the baseband unit (BBU) pool, while conventional powerful base stations are replaced by low-power low-cost remote radio heads (RRHs). These RRHs are connected to the BBU pool through high-bandwidth and low-latency transport links [3]. Such an approach has significant cost advantages and can reduce both the capital expenditure (CAPEX) (e.g., via low-cost site construction) and operational expenditure (OPEX) (e.g., via centralized cooling). Thanks to the centralized signal processing and resource allocation, it can significantly improve both spectrum efficiency and energy efficiency [4], [5].

The benefits of full cooperation in Cloud-RANs, however, highly depend on the quality of the available channel state information (CSI) [6]. In particular, to provide transmitter-side CSI in frequency-division duplex (FDD) systems, which dominate the current cellular networks, the channel state needs to be first estimated at the mobile users (MUs) via downlink training and then to be fed back to the transmitters [7]. However, in dense Cloud-RANs with massive devices, as the BBU pool can typically support hundreds of RRHs [2], [8], obtaining such massive CSI will deplete the radio resources, which is regarded as the "curse of dimensionality" of Cloud-RAN [9]. Specifically, Hassibi and Hochwald [10] showed that the training length for a MIMO channel should be equal to the number of transmit antennas to make channel estimation feasible with the least-squares estimator [11], which thus cannot be applied in dense Cloud-RAN with a large number of antennas at cooperative RRHs. It was shown in [12] that the training length can be reduced by exploiting the channel spatial correlation with the minimum mean-square error (MMSE) estimator [12]. Recently, temporal correlation has been proven to be necessary to reduce the training overhead for the FDD massive MIMO system [13], [14], with the Kalman filter [11] to track the channel time variations.

However, to make it tractable, either the MMSE estimator or the Kalman filter usually makes the Gaussian distribution assumption for the underlying channel model. Furthermore, in the context of dense wireless cooperative networks, such spatial and temporal channel statistics may be difficult to obtain. For instance, as the dimension of the channel covariance matrix grows quadratically with the total number of transmit antennas, the overhead for feeding back such a high dimensional covariance matrix will be prohibitive, even though channel statistics change relatively slow compared

Manuscript received March 26, 2017; revised September 3, 2017 and December 27, 2017; accepted January 16, 2018. Date of publication February 13, 2018; date of current version April 8, 2018. This work was supported in part by the National Nature Science Foundation of China under Grant 61601290, in part by the Hong Kong Research Grant Council under Grant 16211815, and in part by the Shanghai Sailing Program under Grant 16YF1407700. This paper was presented at the IEEE International Conference on Communications, Paris, France, May 2017 [1]. The associate editor coordinating the review of this paper and approving it for publication was H.-C. Wu. *(Corresponding author: Yuanming Shi.)*

X. Liu is with the School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: xuan.liu@sydney.edu.au).

Y. Shi is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: shiym@shanghaitech.edu.cn).

J. Zhang is with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong (e-mail: eejzhang@ust.hk).

K. B. Letaief is with Hamad Bin Khalifa University, Doha, Qatar, and also with The Hong Kong University of Science and Technology, Hong Kong (e-mail: eekhaled@ust.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2018.2797969

with instantaneous channel information [14]. Hence, reducing the prior information overhead and modeling assumptions while guaranteeing a good estimation performance is necessary to overcome the curse of dimensionality for the channel estimation problem in dense Cloud-RANs.

In this paper, we are interested in the high-dimensional channel estimation problem in dense Cloud-RANs, where the training length is smaller than the dimension of the channel. This problem is thus an ill-posed inverse problem [15]. It has been well recognized that exploiting low-complexity structures of the underlying signals (e.g., sparsity, low-rankness) can make the underdetermined inverse problems solvable. Convex optimization provides a powerful tool to achieve this goal with computational efficiency (i.e., polynomial time solvable) and statistical efficiency (i.e., good estimation performance) by designing appropriate regularizing functions [15], [16] to incorporate the low-complexity signal structures. This motivates recent works on the compressed sensing based wireless channel estimation technique [17] by exploiting the channel sparsity in the angular and frequency domain. In particular, the required training length is only in the order of $\mathcal{O}(s \log N)$ for a MIMO channel with the sparsity level $s$ by exploiting its spatial sparsity via $\ell_1$-norm minimization [15], which substantially reduces the training overhead compared with the least-squares estimator [10]. The training overhead can be further reduced if additional structured information is available, e.g., the partial support information of sparse channels [18].

However, the spatial sparsity modeling assumption is questionable in dense wireless networks [19]. Therefore, the conventional compressed sensing based approach [17] may lose its effectiveness in terms of training overhead reduction. This motivates us to take both spatial and temporal dynamics into consideration to enhance the channel estimation performance, thereby reducing the training overhead. In particular, we shall exploit the unique property of heterogeneous channel gains in the spatial domain for dense Cloud-RANs to reduce the training overhead with few modeling assumptions [5]. The heterogeneity of large-scale fading has already been exploited for coordinated beamforming [9] and topological interference alignment [20], [21] to reduce the CSI signaling overhead. We further exploit the temporal correlations [13], [14] and temporal sparsity to reduce the training overhead. The temporal sparsity in dense Cloud-RANs with massive devices is based on the fact that only a few number of mobile devices are active for Internet-of-Things (IoT) devices and machine-type mobile devices using random access [22].

Although there have been numerous research efforts on CSI acquisition, there is still a lack of a systematic approach for CSI training overhead reduction, especially in such dense wireless networks as Cloud-RANs. In contrast to the previous works, in this paper, we propose a novel modeling and algorithmic framework for high-dimensional structured channel estimation via the following regularized optimization formulation

$$\hat{\boldsymbol{H}} = \arg \min_{\boldsymbol{M} \in \mathbb{C}^{m \times n}} \mathcal{L}(\boldsymbol{M}; \mathcal{Z}) + \lambda \mathcal{R}(\boldsymbol{M}), \quad (1)$$

where $\mathcal{Z}$ is a collection of observations, $\lambda \geq 0$ is the user specified parameter, $\mathcal{L}$ and $\mathcal{R}$ are convex functions. Specifically, function $\mathcal{L}$ measures the compatibility of the channel estimate $\boldsymbol{M}$ with the observations $\mathcal{Z}$ using the least-squares criteria, and function $\mathcal{R}$ encodes the available spatial and temporal prior information into the regularizing functions [15], [16], which aims at enforcing structures in the solution and improving the estimation performance. We mainly focus on the convex structured regularizers due to the computational and statistical efficiency. In Section III, we shall propose to adopt the weighted $\ell_1$-norm [23] to exploit the spatial prior information, which consists of only the dominated large-scale fading coefficients. A quadratic regularizing function [11, Sec. 2.4] will be proposed to exploit the temporal correlation prior information in Section IV, which measures the closeness of the current channel estimate to the previous estimated channel via temporal correlation. We further present the group-structured regularizer [16] in Section V to exploit the temporal sparsity for compressive channel estimation with massive device connectivity .

To improve the computational efficiency and make the algorithms scale well to large network sizes, the generic large-scale conic programming solver SCS [24] is adopted to solve the resultant convex high-dimensional channel estimation problems. It is an alternating direction method of multipliers (ADMM) (i.e., operator splitting method) based first-order algorithm [25], and is amenable to parallel computation. This is achieved by equivalently reformulating the convex regularized optimization problem (1) into a standard conic program form using the epigraph form [26] and Smith form [27]. Therefore, the proposed unified massive CSI acquisition framework possesses features including modeling flexibly via exploiting the spatial and temporal dynamics, as well as computation scalability. That is, it can scale well to large network sizes in terms of the reduction of prior information overhead and the computation cost. Simulation results shall demonstrate the effectiveness of the proposed CSI estimation method, and the important role of the spatial and temporal prior information for channel training.

### A. Contributions

The major contributions of the paper are summarized as follows:

1) We propose a high-dimensional structured channel estimation framework to unify the benefits of exploiting the spatial and temporal dynamics for massive CSI acquisition. It enjoys the modeling flexibility and algorithmic efficiency by developing various structured convex regularizers.
2) To encode the heterogenous large-scale fading in the spatial domain, the weight $\ell_1$-norm regularizer is developed. A composite convex regularizer is further proposed to exploit the spatial-temporal channel dynamics. We also present a group-structured regularizer for training to exploit the temporal sparsity for massive device connectivity.
3) The operator splitting method [24] is adopted to solve the convex high-dimensional channel estimation

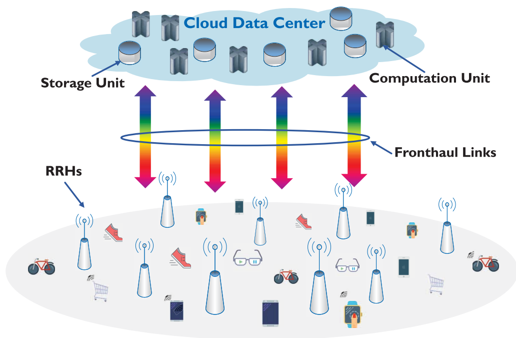



Fig. 1. The architecture of Cloud-RAN with massive mobile devices, in which, all the RRHs are connected to a single BBU pool through high-capacity and low-latency optical fronthaul links.

problems in parallel. To achieve this goal, we propose to transform the original convex programs into a standard cone program via epigraph reformulation, followed by the Smith form reformulation [27].

4) The proposed high-dimensional structured channel estimation framework provides a robust way for massive CSI acquisition with different spatial and temporal dynamics patterns. Typical examples on exploiting spatial and temporal prior information are simulated to demonstrate the effectiveness of the proposed methods.

This work may serve the general purpose of developing practical modeling and algorithmic strategies for massive CSI acquisition in dense wireless networks. Our previous works [3], [5], [8], [9], [20], [27] mainly focus on large-scale network resource management problems with different CSI assumptions.

## II. System Model and Problem Statements

### A. System Model

Consider a frequency division duplexing (FDD) Cloud-RAN with $L$ multi-antenna RRHs and $K$ single-antenna mobile users (MUs) as shown in Fig. 1, where the $l$-th RRH is equipped with $N_l$ antennas. Denote $N = \sum_{l=1}^{L} N_l$ as the total number of transmit antennas at all the RRHs. In Cloud-RAN, all the baseband units (BBUs) are shifted to a single cloud datacenter, i.e., forming a BBU pool, where the centralized signal processing is performed. The RRHs are connected to the BBU pool via high-capacity and low-latency optical fronthaul links. Cloud-RAN is thus a cost-effective network architecture that leverages recent advances in cloud computing and network function virtualization to improve energy and spectral efficiency of wireless networks. With the shared computation resources at the cloud data center and distributed low-cost low-power remote radio heads (RRHs), Cloud-RAN provides an ideal platform to harness the benefits of network cooperation and coordination [4], [5], [28].

In this paper, we first focus on the downlink channel estimation problem in Section III and Section IV, and then discuss how to extend to the uplink channel estimation in Section V. We adopt a block fading channel model with coherence time $T_0$, i.e., the channel coefficients keep constant for a time period $T_0$ and change to a new realization in the next block [7]. In particular, at the $\tau$-th block, the channel propagation from the $l$-th RRH to the $k$-th MU is denoted as $\boldsymbol{h}_{kl}[\tau] \in \mathbb{C}^{N_l}$. Let $\boldsymbol{h}_k[\tau] = [\boldsymbol{h}_{k1}[\tau]^T, \ldots, \boldsymbol{h}_{kL}[\tau]^T]^T \in \mathbb{C}^N$ be the channel vector from all the RRHs to MU $k$ at block $\tau$. The channel matrix from all the RRHs to all the MUs at block $\tau$ is denoted as $\boldsymbol{H}[\tau] = [\boldsymbol{h}_1[\tau], \ldots, \boldsymbol{h}_K[\tau]] \in \mathbb{C}^{N \times K}$.

### B. Linear Measurements Model

We consider the downlink training problem. Let $T$ be the training length for each block, and $0 < T < T_0$. At the $\tau$-th block, RRH $l$ sends a $(T \times N_l)$-dimensional measurement matrix $\boldsymbol{Q}_l[\tau]$, and MU $k$ receives a $T$-dimensional complex observation vector $\boldsymbol{y}_k[\tau] \in \mathbb{C}^T$

$$\boldsymbol{y}_k[\tau] = \boldsymbol{Q}[\tau]\boldsymbol{h}_k[\tau] + \boldsymbol{n}_k[\tau], \quad \forall \tau \in \mathcal{J}, \tag{2}$$

where $\mathcal{J} = \{0, 1 \ldots, J-1\}$, $\boldsymbol{Q}[\tau] = [\boldsymbol{Q}_1[\tau], \ldots, \boldsymbol{Q}_L[\tau]] \in \mathbb{C}^{T \times N}$, and $\boldsymbol{n}_k[\tau] \in \mathbb{C}^T$ is the additive noise vector. Let $\boldsymbol{N}[\tau] = [\boldsymbol{n}_1[\tau], \ldots, \boldsymbol{n}_K[\tau]] \in \mathbb{C}^{T \times K}$. The linear observations $\boldsymbol{Y}[\tau] = [\boldsymbol{y}_1[\tau], \ldots, \boldsymbol{y}_K[\tau]] \in \mathbb{C}^{T \times K}$ for channel estimation can be rewritten as

$$\boldsymbol{Y}[\tau] = \boldsymbol{Q}[\tau]\boldsymbol{H}[\tau] + \boldsymbol{N}[\tau], \quad \forall \tau \in \mathcal{J}. \tag{3}$$

The main purpose of downlink CSI acquisition is to enable the centralized precoding/beamforming design at the BBU pool. For FDD systems, channel estimation can be either performed at MUs and then the MUs feed back the quantized channel estimate to the RRHs, or the MUs directly feed back the observations (2) to the RRHs and then perform channel estimation at the BBU pool, which may lead to non-linear observations [29] for channel estimation due to quantization. Note that either the received quantized channel estimate or the observations at RRHs need to be further delivered to the BBU pool through the fronthaul links, where the signal processing (e.g., precoding, estimation) is performed. As our main focus in this paper is the channel estimation algorithm design, we ignore the impact of limited feedback and limited fronthaul link capacity, and assume that the BBU pool can perfectly access the linear observations $\boldsymbol{Y}[\tau]$'s. In this paper, we are interested in the scenario with $T < N$. We thus propose to directly feed back all the observations $\boldsymbol{Y}[\tau]$ to RRHs and then recover the channel $\boldsymbol{H}[\tau]$ jointly at the BBU pool. This helps reduce the feedback overhead, and can further reduce the estimation overhead by exploiting the low-dimensional structure in $\boldsymbol{H}[\tau]$ jointly. The effect of quantization of the observations will be evaluated via simulations. As the proposed methodologies can be also applied to the time division duplex (TDD) mode, where the uplink training is performed. Considering the amount of feedback, in the scenario with a large number of MUs, uplink training is more attractive.

### C. Channel Estimation and Modeling Assumptions

*1) High-Dimensional Structured Channel Estimation:* In this paper, we are interested in the high-dimensional channel estimation problem in dense Cloud-RANs, where the training length $T$ is smaller than the dimension of the channel. Therefore, it is an ill-posed inverse problem [15]. Fortunately, the wireless channel propagation often possesses spatial and temporal dynamics, thereby providing prior information to improve estimation performance. We thus propose a general framework to convert the available prior information into convex structured regularizing functions [16], yielding a convex optimization solution to the underdetermined inverse problem for channel estimation. We, therefore, mainly focus on the estimation algorithm design based on the available spatial and temporal prior information.

*2) Channel Modeling With Spatial-Temporal Dynamics:* Channel modeling assumptions often lead to the use of different strategies to estimate channel efficiently. Channel spatial and temporal dynamics play a critical role for high-dimensional structured channel estimation [17]. To simplify the presentation, we assume that the channel spatial and temporal statistics keep the same during $J$ blocks. Specifically, the instantaneous channel matrix $\boldsymbol{H}[\tau]$ of block $\tau$ can be modeled as the following Hadamard product form

$$\boldsymbol{H}[\tau] = \boldsymbol{D} \circ \boldsymbol{G}[\tau], \quad \forall \tau \in \mathcal{J}, \tag{4}$$

where $\boldsymbol{A} \circ \boldsymbol{B}$ denotes the element-wise product of $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{G}[\tau] \in \mathbb{C}^{N\times K}$ and $\boldsymbol{D} = [D_{ij}] \in \mathbb{C}^{N\times K}$ represent the small-scale and large-scale fading coefficient matrix, respectively. Here, entry $D_{ij}$ models the path-loss from transmit antenna $i$ to mobile user $j$. As RRHs in a Cloud-RAN are distributed at different locations, we do not assume any spatial correlations. Furthermore, we assume that the time variation of the channel matrices $\boldsymbol{H}[\tau]$'s across different blocks follow the following first-order stationary Gauss-Markov process [13], [14]

$$\boldsymbol{H}[\tau] = \boldsymbol{C}\boldsymbol{H}[\tau-1] + \boldsymbol{V}[\tau], \quad \forall \tau \in \mathcal{J}, \tag{5}$$

where $\boldsymbol{C} \in \mathbb{C}^{N\times N}$ is the temporal correlation coefficient matrix and is assumed to be the same during all the $J$ blocks, and $\boldsymbol{V}[\tau] \in \mathbb{C}^{N\times K}$ is an innovation process. Here, channel matrices $\boldsymbol{H}[\tau]$'s are written as (4), while (5) represents the temporal relationship between current channel $\boldsymbol{H}[\tau]$ and previous channel $\boldsymbol{H}[\tau-1]$.

To make the estimation approach robust, we only make few channel modeling assumptions on the spatial and temporal prior information for channel estimation, while keeping the estimation performance competitive. Specifically, for channel modeling in the temporal domain, we only assume that the temporal correlation matrix $\boldsymbol{C}$ is known. We also further exploit the channel temporal sparsity for massive device connectivity. In this scenario, out of large number of devices, only a few number of machine-type mobile devices are active with sporadic traffic [22].

Although sparsity plays a key role for high-dimensional structured estimation, spatial sparsity assumption for wireless channel is still questionable in dense wireless networks [19]. Furthermore, the channel distribution information is difficult to measure and model. Most of the available distribution models are basically empirical and may lead to mismatch with the real channel distributions. In this paper, we only assume that the large-scale fading channel coefficients are available, which only depend on the location information. The success of this proposal is based on the heterogeneity of large-scale fading [9], [21]. The recent work on network connectivity information based localization approach [30] can be applied in dense wireless networks for location estimation. As only connectivity information is required, this approach thus scales well to large network sizes.

## III. High-Dimensional Structured Channel Estimation With Spatial Dynamics

In this section, we propose the weight $\ell_1$-norm minimization approach for high-dimensional structured channel estimation by exploiting the spatial prior information. To simplify the notation, we omit the block index $\tau$ in this section.

### A. Channel Estimation With Structured Regularizers

Given the observation $\boldsymbol{Y}$ and the measurement matrix $\boldsymbol{Q}$ at one channel block, a popular way for estimating the channel matrix $\boldsymbol{H}$ is based on the following least-squares criterion,

$$\hat{\boldsymbol{H}} = \arg\min_{\boldsymbol{M}} \|\boldsymbol{Y} - \boldsymbol{Q}\boldsymbol{M}\|_F^2, \tag{6}$$

which is, however, meaningful only when $T \geq N$ [11]. For the high-dimensional estimation problem considered in this paper, we investigate how the prior information can help improve the performance of channel estimation. Specifically, we propose the following generic high-dimensional channel estimation framework via the regularizing formulation

$$\hat{\boldsymbol{H}} = \arg\min_{\boldsymbol{M}} \ \|\boldsymbol{Y} - \boldsymbol{Q}\boldsymbol{M}\|_F^2 + \sum_i \nu_i f_i(\boldsymbol{M}), \tag{7}$$

where the first term serves to measure the fitness of the estimate $\boldsymbol{M}$ to the observation $\boldsymbol{Y}$ based on the least-squares criterion as in (6), and the second term serves to encode different types of available prior information into the solution using different structured regularizing functions $f_i(\boldsymbol{M})$'s, while $\nu_i$'s ($\nu_i > 0$) are the corresponding weights to be chosen. In this paper, we mainly focus on the following problems

- Design algorithmically and statistically efficient regularizing functions $f_i$'s to encode the available spatial prior information.
- Demonstrate that high-dimensional channel estimation can be done with a reasonable amount of prior information.

The main observation that motivates the proposed channel estimation framework is that, in wireless networks, the channel is not arbitrary and actually possesses additional exploitable statistical characteristics. To make the estimation algorithm computationally tractable, we restrict the regularizing functions to convex functions, which have been proven powerful in high-dimensional problems [15], [16] to achieve both statistical and computational efficiency.

Furthermore, there is an alternative estimation framework based on the following constrained formulation

$$\mathscr{P}: \ \hat{\boldsymbol{H}} = \arg\min_{\boldsymbol{M}} \ \sum_i \lambda_i f_i(\boldsymbol{M}) \quad \text{subject to } \|\boldsymbol{Y} - \boldsymbol{Q}\boldsymbol{M}\|_F \leq \epsilon, \tag{8}$$

where $\epsilon > 0$ is an upper bound on the noise $\|\boldsymbol{N}\|_F$ and is assumed to be known as a *prior*. Lagrange multipliers indicate that solving the constrained program $\mathscr{P}$ is equivalent to solving the regularized problem (7) under convexity and mild regularity conditions [31]. We thus focus on channel estimator design based on the constrained formulation $\mathscr{P}$. Note that the denoising error $\epsilon$ in problem $\mathscr{P}$ is bounded by the noise variance with high probability via the denoising procedure [32]. Basically, the noise variance can be estimated either using the pilot sequences embedded in the signal or using the transmitted data in a non-data aided manner [33].

### B. Prior Information With Spatial Dynamics

We exploit the spatial domain prior information for training, including the statistical distribution information and the heterogeneity of large-scale fading coefficients.

*1) Channel Distribution Information:* A unique property of distributed networks such as Cloud-RAN is that signals coming from different RRHs are with different pathlosses, i.e., some links will be strong while some will be weak. This is different from a point-to-point multi-antenna system, and will be critical for the channel estimation design for Cloud-RANs. As $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^T \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B})$, the signal model (3) can be vectorized as

$$\underbrace{\text{vec}(\boldsymbol{Y})}_{\boldsymbol{y}} = \underbrace{(\boldsymbol{I}_K \otimes \boldsymbol{Q})}_{\tilde{\boldsymbol{Q}}} \underbrace{\text{vec}(\boldsymbol{H})}_{\boldsymbol{h}} + \underbrace{\text{vec}(\boldsymbol{N})}_{\boldsymbol{n}}, \tag{9}$$

where $\boldsymbol{y} \in \mathbb{C}^{TK}$, $\tilde{\boldsymbol{Q}} \in \mathbb{C}^{TK \times NK}$, $\boldsymbol{h} \in \mathbb{C}^{NK}$ and $\boldsymbol{n} \in \mathbb{C}^{TK}$. Assume that only the spatial prior information (i.e., large-scale fading coefficients) is available. We first consider the optimal performance with full prior information. Assume that $\boldsymbol{h} \in \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{C}^{NK}$ and $\boldsymbol{\Sigma} \in \mathbb{C}^{NK \times NK}$ with $\boldsymbol{\Sigma} \succeq \boldsymbol{0}$; and $\boldsymbol{n} \in \mathcal{CN}(\boldsymbol{\theta}, \boldsymbol{\Theta})$, where $\boldsymbol{\theta} \in \mathbb{C}^{TK}$ and $\boldsymbol{\Theta} \in \mathbb{C}^{TK \times TK}$ with $\boldsymbol{\Theta} \succeq \boldsymbol{0}$. The MMSE estimate for the channel vector $\boldsymbol{h}$ is given by [11]

$$\begin{aligned}\hat{\boldsymbol{h}}_{\mathsf{mmse}} &= \mathbb{E}[\boldsymbol{h}|\boldsymbol{y}] = \int \boldsymbol{h} f(\boldsymbol{h}|\boldsymbol{y}) d\boldsymbol{h} \\ &= \boldsymbol{\mu} + \boldsymbol{\Sigma} \tilde{\boldsymbol{Q}}^{\mathsf{H}} (\tilde{\boldsymbol{Q}} \boldsymbol{\Sigma} \tilde{\boldsymbol{Q}}^{\mathsf{H}} + \boldsymbol{\Theta})^{-1} (\boldsymbol{y} - \tilde{\boldsymbol{Q}} \boldsymbol{\mu} - \boldsymbol{\theta}). \end{aligned} \tag{10}$$

The MMSE estimator can provide the optimal performance but it requires the knowledge of the channel covariance matrix $\boldsymbol{\Sigma} \in \mathbb{C}^{NK \times NK}$, which is difficult to obtain in dense Cloud-RANs. Furthermore, most of the distribution models are empirical and may yield mismatch with the real channel propagations especially in the ultra-dense wireless networks, where the exact channel model should be derived from the Maxwell's equations [34]. In the following, we will show that with the help of the new estimator (8), we can achieve the performance of the MMSE estimator with much less prior information. For the conventional MMSE estimator, the number of orthogonal pilot symbols, i.e., the training length $T$, must scale linearly with the total number of transmit antennas $N$ at all the RRHs, i.e., $\mathcal{O}(N)$. However, the number of available orthogonal pilot symbols is limited by the channel coherence bandwidth and coherence time [35]. As $N$ becomes very large in ultra-dense Cloud-RANs, we do not have sufficient pilots to support channel estimation even without the consideration of the channel signaling overhead issue. In this paper, we are thus particularly interested in the scenario with $T < N$.

*2) Heterogenous Large-Scale Fading:* Due to the pathloss and shadowing, in dense Cloud-RANs, the wireless channel propagation gains are heterogeneous. For example, the results in [9] showed that only obtaining the dominated channel links is enough to achieve performance close to the one with full CSI. Thus, significant CSI overhead reduction can be achieved. Furthermore, topological interference alignment is a promising proposal to manage the interference in partially connected dense wireless networks [20]. However, the spatial sparsity modeling assumption for the purpose of channel estimation is still questionable [19]. Instead of using $\|\text{vec}(\boldsymbol{H})\|_0$ as the sparsity measurement, the soft sparsity measurement $\|\text{vec}(\boldsymbol{H})\|_1/\|\text{vec}(\boldsymbol{H})\|_2$ [36] has recently been proposed as an effective way to address the issue of sparsity modeling. However, estimating the soft sparsity level is also challenging.

To address the above modeling issues, as well as reduce the overhead of spatial prior information, we propose to only acquire the dominated large-scale fading coefficients $D_{ij}$'s indexed by the following set

$$\mathcal{D} = \{(i, j)|D_{ij} \geq D_0\}, \tag{11}$$

where $D_0 \geq 0$ is a predefined threshold. This is based on the heterogeneity of large-scale fading coefficients due to the pathloss and shadowing. Furthermore, the large-scale fading coefficients can be obtained by the location information of all the nodes in the networks. Localization can normally be achieved by measuring the pairwise Euclidean distance matrix [37], which, however, yields significant overhead in dense Cloud-RANs. To scale well to large network sizes, we can adopt the 1-bit matrix completion approach for localization, which is only based on the partial network connectivity information [30].

Based on the large-scale fading prior information $\mathcal{D}$ (11), we propose to use the following weighted $\ell_1$-norm [23] to encode the available spatial prior information,

$$f_1(\boldsymbol{M}) = \|\boldsymbol{W} \circ \boldsymbol{M}\|_1, \tag{12}$$

where $\|\boldsymbol{A}\|_1 = \sum_{i=1}^m \sum_{j=1}^n |A_{ij}|$ is the entry-wise $\ell_1$-norm for the $m \times n$ matrix, and the weight matrix $\boldsymbol{W} = [W_{ij}]$ is defined as follows

$$W_{ij} = \begin{cases} 1/D_{ij}, & \text{if } (i, j) \in \mathcal{D}; \\ 1/(2D_0), & \text{otherwise.} \end{cases} \tag{13}$$

Intuitively, smaller large-scale fading will yield smaller channel estimate $M_{ij}$, due to the large penalty weight $W_{ij}$. That is, the channel coefficient with small large-scale fading coefficient has a high probability to yield a small instantaneous channel coefficient. The large-scale fading coefficients provide the prior information in the form of probabilities that each channel coefficient is non-zero. This spatial prior information thus provides the prior distribution of the channel coefficients, which can be leveraged by using the non-uniform weights in (12). Specifically, we assume that partially spatial prior information (i.e., large-scale fading coefficients indexed by $\mathcal{D}$) is available. The threshold $D_0$ is determined by the amount of available large-scale fading coefficients.

### C. High-Dimensional Channel Estimation via Weighted $\ell_1$-Norm Minimization

Based on the convex structured regularizer (12), we arrive at the following high-dimensional channel estimation approach via weighted $\ell_1$-norm minimization:

$$\mathscr{P}_1: \quad \hat{\boldsymbol{H}} = \arg\min_{\boldsymbol{M}} \ \|\boldsymbol{W} \circ \boldsymbol{M}\|_1 \quad \text{subject to } \|\boldsymbol{Y} - \boldsymbol{Q}\boldsymbol{M}\|_F \leq \epsilon, \tag{14}$$

where $\epsilon > 0$ is an upper bound on the noise and is assumed to be known as a *prior*. Although problem $\mathscr{P}_1$ can be reformulated as the regularized optimization formulation (1), it is difficult to determine the optimal regularizer parameter. From the Hadamard product channel representation (4), the weighted $\ell_1$-norm regularizing function serves to enforce the structure in the solution of algorithm $\mathscr{P}_1$ such that the instantaneous channel coefficients are encouraged to be small if the corresponding large-scale coefficients are small. Note that without weights, it will become the compressed sensing (CS) based estimator, i.e., it only tries to exploit the sparsity structure in the channel. As will be shown later, the CS approach will not work for dense Cloud-RANs, as the channel is not truly sparse [19]. This demonstrates a typical channel modeling issue in dense Cloud-RANs.

## IV. Massive CSI Acquisition With Spatial-Temporal Dynamics

In this section, we exploit both the spatial and temporal dynamics to further improve the channel estimation. This is achieved by the proposal of the composite structured regularizer to encode the spatial and temporal prior information.

### A. Prior Information With Temporal Correlations

In this subsection, we shall demonstrate that simultaneously exploiting the spatial and temporal dynamics can improve the channel estimation performance if second-order statistical information of the channel is available. Exploiting channel correlation both in time and space is known to be effective for training [13], [14]. Specifically, the Gauss-Markov channel model (5) can be vectorized as

$$\boldsymbol{h}[\tau] = \tilde{\boldsymbol{C}}\boldsymbol{h}[\tau - 1] + \boldsymbol{v}[\tau], \quad \forall \tau \in \mathcal{J}, \tag{15}$$

where $\tilde{\boldsymbol{C}} = (\boldsymbol{I}_K \otimes \boldsymbol{C}) \in \mathbb{C}^{NK \times NK}$ and $\boldsymbol{v}[\tau] = \text{vec}(\boldsymbol{V}[\tau]) \in \mathbb{C}^{NK}$. Combining (9) and (15), we arrive at the following state-space system [11]

$$\mathcal{S}: \begin{cases} \boldsymbol{h}[\tau] = \tilde{\boldsymbol{C}}\boldsymbol{h}[\tau - 1] + \boldsymbol{v}[\tau] \\ \boldsymbol{y}[\tau] = \tilde{\boldsymbol{Q}}_i[\tau]\boldsymbol{h}[\tau] + \boldsymbol{n}[\tau], \end{cases} \tag{16}$$

where $\tau \in \mathcal{J}$. An optimal channel estimator for (16) is given by the Kalman filter [11]. To make it computationally tractable, we further make the following assumptions:

- All the variables, i.e., $\boldsymbol{h}, \boldsymbol{n}, \boldsymbol{v}$, in the state-space system $\mathcal{S}$ are jointly Gaussian distributed.
- The channel, observation noise and innovation process covariance matrices $\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{h}[\tau]\boldsymbol{h}[\tau]^{\mathsf{H}}]$, $\boldsymbol{\Theta} = \mathbb{E}[\boldsymbol{n}[\tau]\boldsymbol{n}[\tau]^{\mathsf{H}}]$, and $\boldsymbol{\Xi} = \mathbb{E}[\boldsymbol{v}[\tau]\boldsymbol{v}[\tau]^{\mathsf{H}}]$, $\forall \tau \in \mathcal{J}$, are available.

The Kalman filter [11] for the channel estimate $\boldsymbol{h}[\tau]$ based on the observations $\boldsymbol{y}[0], \ldots, \boldsymbol{y}[J-1]$, is presented in Algorithm 1 with the following notations:

- $\hat{\boldsymbol{h}}_{\tau|i}$ with $i = \tau$ or $\tau - 1$, denotes the MMSE estimate of $\boldsymbol{h}[\tau]$ based on $\boldsymbol{y}[0]$ through $\boldsymbol{y}[i]$;
- The estimation error covariance is denoted as $\boldsymbol{\Delta}_{\tau|i} = \mathbb{E}[\boldsymbol{\delta}_{\tau|i}\boldsymbol{\delta}_{\tau|i}^{\mathsf{H}}]$ with $\boldsymbol{\delta}_{\tau|i} = \boldsymbol{h}[\tau] - \hat{\boldsymbol{h}}_{\tau|i}$ and $i = \tau$ or $i = \tau - 1$.

**Algorithm 1** Kalman Filter Based Channel Estimation

**Initialization**: $\hat{\boldsymbol{h}}_{0|-1} = \boldsymbol{0}$; $\boldsymbol{\Delta}_{0|-1} = \boldsymbol{\Sigma}$.
**while** $\tau = 0, 1, \ldots, J - 1$, **do**
**Measurement update**:

- Calculate Kalman gain matrix:
$$\boldsymbol{K}_\tau = \boldsymbol{\Delta}_{\tau|\tau-1}\tilde{\boldsymbol{Q}}[\tau]^{\mathsf{H}}(\tilde{\boldsymbol{Q}}[\tau]\boldsymbol{\Delta}_{\tau|\tau-1}\tilde{\boldsymbol{Q}}[\tau]^{\mathsf{H}} + \boldsymbol{\Theta})^{-1}.$$
- Update estimate with measurement $\boldsymbol{y}_i$:
$$\hat{\boldsymbol{h}}_{\tau|\tau} = \hat{\boldsymbol{h}}_{\tau|\tau-1} + \boldsymbol{K}_\tau(\boldsymbol{y}[\tau] - \tilde{\boldsymbol{Q}}[\tau]\hat{\boldsymbol{h}}_{\tau|\tau-1}).$$
- Compute MSE matrix
$$\boldsymbol{\Delta}_{\tau|\tau} = (\boldsymbol{I} - \boldsymbol{K}_\tau\tilde{\boldsymbol{Q}}[\tau])\boldsymbol{\Delta}_{\tau|\tau-1}.$$

**Time update**:

- Prediction: $\hat{\boldsymbol{h}}_{\tau|\tau-1} = \boldsymbol{C}\hat{\boldsymbol{h}}_{\tau-1}$.
- Prediction MSE matrix:
$$\boldsymbol{\Delta}_{\tau|\tau-1} = \boldsymbol{C}\boldsymbol{\Delta}_{\tau-1|\tau-1}\boldsymbol{C}^{\mathsf{H}} + \boldsymbol{\Xi}.$$

**end while**.

However, as we have seen that the Kalman filter requires the prior information on the second-order statistics of the channel, the observation noise, and the innovation process. In last Section III, we have presented a way to reduce the requirements of the spatial prior information compared to the statistical distribution based MMSE estimator. In this section, we propose to further reduce the temporal prior information overhead compared to the Kalman filter. This is achieved by only exploiting the large-scale fading coefficients $\mathcal{D}$ (11) and temporal correlation coefficient $\boldsymbol{C}$. Furthermore, Kalman filter based channel estimation has the cubic complexity to perform matrix inversion at each iteration. We shall propose a convex regularized optimization approach to reduce both the prior information overhead and the computational complexity. The computational scalability issue will be addressed in Section VI via the parallel first-order method.

### B. Spatial-Temporal Structured Channel Estimation

In this subsection, we shall further demonstrate that the channel estimation performance can be improved with information of only the large-scale fading and temporal correlation coefficients, instead of the second-order statistical prior information used in the Kalman Filter approach in Algorithm 1. From the Gauss-Markov model (5), we can use the previously estimated channel $\hat{\boldsymbol{H}}[\tau - 1]$ to predict the current channel $\boldsymbol{H}[\tau]$, i.e.,

$$\mathcal{H}[\tau] = \boldsymbol{C}\hat{\boldsymbol{H}}[\tau - 1], \tag{17}$$

which provides a guess for the current channel at the $\tau$-th block. Therefore, we propose to use the following squared $\ell_2$-norm with a quadratic form as a convex regularizing function to incorporate the available temporal prior information (17),

$$f_2(\boldsymbol{M}; \tau) = \|\boldsymbol{M} - \boldsymbol{\mathcal{H}}[\tau]\|_F^2, \tag{18}$$

for the $\tau$-th block. Note that the previous estimated channel $\hat{\boldsymbol{H}}[\tau-1]$ always available as a prior, as we can store it at the BBU pool.

We thus arrive at the following channel estimation problem by encoding the spatial and temporal prior information into the following composite convex regularizer function,

$$\mathscr{P}_2(\tau): \ \hat{\boldsymbol{H}}[\tau] = \arg\min_{\boldsymbol{M}} \ \|\boldsymbol{W} \circ \boldsymbol{M}\|_1 + \lambda \|\boldsymbol{M} - \boldsymbol{\mathcal{H}}[\tau]\|_F^2 \quad \text{subject to } \|\boldsymbol{Y}[\tau] - \boldsymbol{Q}[\tau]\boldsymbol{M}\|_F \leq \epsilon, \tag{19}$$

where the quadratic regularizing function (i.e., the squared $\ell_2$-norm) serves to enforce the fidelity of the prior information $\boldsymbol{\mathcal{H}}[\tau]$ and weight $\lambda$ indicates how confident we are about the fidelity of the optimal solution $\hat{\boldsymbol{H}}[\tau]$ to the given matrix $\boldsymbol{\mathcal{H}}[\tau]$ [11, Section 2.4]. Here, $\epsilon > 0$ is an upper bound on the noise and is assumed to be known as a *prior*. Although problem $\mathscr{P}_2$ can be reformulated as the regularized optimization formulation (1), it is difficult to determine the optimal regularizer parameter.

We now present the proposed spatial-temporal structured channel estimation algorithm in Algorithm 2.

**Algorithm 2** Spatial-Temporal Structured Channel Estimation Algorithm $\mathscr{P}_2$

**Initialization**: $\hat{\boldsymbol{H}}[-1] = \boldsymbol{0}$, $\boldsymbol{\mathcal{H}}[0] = \boldsymbol{0}$;
**while** $\tau = 0, 1, \ldots, J-1$, **do**
- **Measurement update**: Solve problem $\mathscr{P}_2(\tau)$ and obtain $\hat{\boldsymbol{H}}[\tau]$.
- **Time update**: Let $\tau = \tau + 1$, compute the matrix $\boldsymbol{\mathcal{H}}[\tau]$ (17) that serves as the temporal prior information.

**end while**.

Note that, to update the measurement, both Kalman filter and the proposed Algorithm 2 have cubic complexity to perform matrix inversion and to solve convex optimization problem $\mathscr{P}_2$ using interior-point method [26], respectively. To improve the computational efficiency, in Section VI, we propose to use a first-order method to solve the large-scale optimization problem $\mathscr{P}_2$ to scale well to large problem sizes.

We summarize the required prior information to compute different channel estimation algorithms in Table I.

TABLE I

PRIOR INFORMATION REQUIRED FOR DIFFERENCE ALGORITHMS

| Algorithms | Prior Information |
|---|---|
| MMSE estimator (10) | Probability distribution |
| Estimator $\mathscr{P}_1$ | Large-scale fading matrix $\boldsymbol{D}$, noise bound $\epsilon$ |
| Kalman filter estimator | Second-order statistics |
| Estimator $\mathscr{P}_2$ | Large-scale fading matrix $\boldsymbol{D}$, temporal correlation matrix $\boldsymbol{C}$, noise bound $\epsilon$ |
| Estimator $\mathscr{P}_3$ | Sparsity of mobile devices activity |

## V. HIGH-DIMENSIONAL UPLINK CHANNEL ESTIMATION WITH MASSIVE DEVICES

In this section, we consider the uplink channel estimation for dense Cloud-RANs with massive mobile devices. It is straightforward to extend the results in Section III and Section IV for downlink channel estimation to the scenario of uplink channel estimation when all the devices are active. However, for ultra-dense Cloud-RAN with massive sporadic traffic type devices (i.e., in each time slot only a few devices are active out of all the massive devices), it is critical to provide radio access for massive mobile devices by exploiting the sparsity of the mobile devices activity. In this scenario, the results in Section III and Section IV may not be effective. More structured sparsity thus needs to be exploited. In this section, we propose to exploit the temporal sparsity for uplink training, thereby supporting massive device connectivity.

### A. Prior Information With Temporal Sparsity

Massive device connectivity is a key requirement for 5G cellular networks, e.g., dense Cloud-RANs. As the communication for machine-type mobile devices using random access and Internet-of-Things (IoT) devices is sporadic, only a few devices are active out of all the massive devices [22]. For uplink transmission with channel coherence time $T_0$, we consider the joint active user identification and channel estimation problem. Specifically, for any given channel coherence block, the received signal at all the RRHs is given by

$$\boldsymbol{y}(\ell) = \sum_{i=1}^{K} \boldsymbol{h}_i q_i(\ell) + \boldsymbol{n}(\ell), \quad \ell = 1, \ldots, T, \tag{20}$$

where $T \leq T_0$ is the training length, $\boldsymbol{h}_i \in \mathbb{C}^N$ is the channel vector from mobile device $i$ to all the RRHs with $N$ antennas in total, $q_i(\ell) \in \mathbb{C}$ is the pilot symbol transmitted from mobile device $i$ at time slot $\ell$, $\boldsymbol{y}(\ell) \in \mathbb{C}^N$ is the received signal at all RRHs, $\boldsymbol{n}(\ell) \in \mathbb{C}^N$ is the additive noise.

Let $\boldsymbol{Y} = [\boldsymbol{y}(1), \ldots, \boldsymbol{y}(T)]^{\mathsf{H}} \in \mathbb{C}^{T\times N}$ be the received signal across $N$ antennas, $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^{\mathsf{H}} \in \mathbb{C}^{K\times N}$ be the channel matrix from all the mobile devices to all the RRHs, and $\boldsymbol{Q} = [\boldsymbol{q}(1), \ldots, \boldsymbol{q}(T)]^{\mathsf{H}} \in \mathbb{C}^{T\times K}$ be the known pilot matrix with $\boldsymbol{q}(\ell) = [q_1(\ell), \ldots, q_K(\ell)]^{\mathsf{H}} \in \mathbb{C}^K$. We then rewrite (20) as

$$\boldsymbol{Y} = \boldsymbol{Q}\boldsymbol{\Pi}\boldsymbol{H} + \boldsymbol{N}, \tag{21}$$

where $\boldsymbol{\Pi} = \mathrm{diag}(\pi_1, \ldots, \pi_K) \in \mathbb{R}^{K\times K}$ is the diagonal activity matrix with $\pi_i = 1$ indicating device $i$ active and $\pi_i = 0$ representing device $i$ inactive, and $\boldsymbol{N} = [\boldsymbol{n}(1), \ldots, \boldsymbol{n}(T)] \in \mathbb{C}^{T\times N}$ is the additive noise matrix. Our goal is to estimate $\boldsymbol{H}$ and $\boldsymbol{\Pi}$ simultaneously, provided that active diagonal matrix $\boldsymbol{\Pi}$ is sparse. This is based on the fact that the number of active mobile users has temporal sparsity.

The model in (21) provides a unified framework for high-dimensional structured estimation for the unknown matrices

$\boldsymbol{\Pi}$ and $\boldsymbol{H}$. Specifically, when $\boldsymbol{\Pi}$ is an unknown diagonal matrix, problem (21) represents the problems including phase retrieval [38] and blind deconvolution [39]. If $\boldsymbol{\Pi}$ is an unknown permutation matrix, problem (21) is the linear regression problem with an unknown permutation [40], which is critical to reduce the control signaling overhead to support ultra-low latency short packet communications [41].

### *B. Group-Structured Sparsity Estimation*

Let $\boldsymbol{M} = \boldsymbol{\Pi}\boldsymbol{H} \in \mathbb{C}^{K\times N}$ with $\boldsymbol{\Pi}$ as an unknown sparse diagonal matrix. The effective matrix $\boldsymbol{M}$ thus has a group-structured sparsity in rows of matrix $\boldsymbol{M}$ [16]. Then the linear measurement model (21) can be rewritten as

$$\boldsymbol{Y} = \boldsymbol{Q}\boldsymbol{M} + \boldsymbol{N}. \tag{22}$$

To induce the group row sparsity in $\boldsymbol{M}$, we introduce the following group norm, i.e., mixed $\ell_1/\ell_2$-norm [16]

$$\mathcal{R}(\boldsymbol{M}) := \sum_{i=1}^{K} \|\boldsymbol{m}_i\|_2, \tag{23}$$

where $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_K]^{\mathsf{H}} \in \mathbb{C}^{K\times N}$ with $\boldsymbol{m}_i \in \mathbb{C}^N$. This is also known as the group Lasso norm in statistics.

The associated group sparse matrix estimation problem for joint active user identification and channel estimation can be formulated as following group norm regularized optimization problem:

$$\mathscr{P}_3 : \hat{\boldsymbol{M}} = \underset{\boldsymbol{M}\in\mathbb{C}^{K\times N}}{\arg\min} \|\boldsymbol{Y} - \boldsymbol{Q}\boldsymbol{M}\|_F^2 + \lambda\mathcal{R}(\boldsymbol{M}), \tag{24}$$

where $\lambda \geq 0$ is a user defined regularization parameter to control the tradeoffs between quality of fit and sparsity. Given the estimate matrix $\hat{\boldsymbol{M}}$, the activity matrix can be recovered as $\hat{\boldsymbol{\Pi}} = \mathrm{diag}(\hat{\pi}_1, \ldots, \hat{\pi}_n)$ with $\hat{\pi}_i = 1$ if $\|\hat{\boldsymbol{m}}_i\|_2 \geq \gamma_0$ for a small enough $\gamma_0$ ($\gamma_0 \geq 0$). The estimated $|\mathcal{A}| \times N$ channel matrix for the active users is thus given by $\hat{\boldsymbol{M}}(\mathcal{A})$ with $\mathcal{A} = \{i|\hat{\pi}_i = 1\}$.

Temporal sparsity for massive device connectivity and channel estimation has been investigated in the frameworks of Bayesian compressive sensing [42], approximate message passing [43] and orthogonal matching pursuit [22]. However, most of the results rely on either the prior distribution information or sparsity level prior information of signal $\boldsymbol{M}$, which is impractical. For the group Lasso based formulation $\mathscr{P}_3$, we may either use the approach in [36] to estimate the sparsity level, thereby designing the optimal regularizer, or adopt the sparsity level independent approach to determine the regularizer [24]. Although problem $\mathscr{P}_3$ is convex and thus can be solved in polynomial time, it is still critical to design the scalable convex algorithms in the scenario of massive devices.

*Remark 1:* If we assume that all the mobile users are active, we can apply the proposed downlink channel estimation approaches $\mathscr{P}_1$ and $\mathscr{P}_2$ to solve the uplink channel estimation problem. Specifically, for uplink channel estimation, the training matrices are sent by the mobile users, while the observations are received by the RRHs. However, in this section, we further consider a more critical scenarios with massive machine-type mobile devices with sporadic data traffic, i.e., only a few mobile devices are active out of all the devices. We thus propose a group sparse estimation approach $\mathscr{P}_3$ to simultaneously detect the user activity and estimate the channel coefficients for the active mobile users.

## VI. Large-Scale Convex Optimization via Operator Splitting Theory

In this section, we present the generic first-order optimization algorithm based on the operator splitting method (i.e., the ADMM algorithm) to solve the large-scale convex optimization problems $\mathscr{P}_1$, $\mathscr{P}_2$ and $\mathscr{P}_3$. This is achieved by reformulating the original problems into the convex programs and then the standard conic programming form, followed by the operator splitting method to solve the transformed standard conic program. The presented large-scale first-order convex algorithm enjoys both the capability of infeasibility detection and the ability to scale to large problem sizes with parallel computing.

### *A. Conic Reformulation via Epigraph Form and Smith Form*

In this subsection, we propose to transform the original convex optimization problems $\mathscr{P}_1$, $\mathscr{P}_2$, and $\mathscr{P}_3$ into standard conic optimization form $\mathscr{P}_{\text{cone}}$ (29). For the convex problems $\mathscr{P}_1$, $\mathscr{P}_2$ and $\mathscr{P}_3$, one may use the interior-point method (a second-order method), which is implemented in most advanced off-the-shelf solvers like SeDuMi [44]. Nevertheless, the computational burden of the second-order method makes it inapplicable in dense Cloud-RAN. For instance, for a network with $L = 100$ single antenna RRHs and $K = 100$ single-antenna MUs, the dimension of the channel matrix is $d = 10^4$ and the corresponding computational complexity for computing problem $\mathscr{P}_1$, $\mathscr{P}_2$ or $\mathscr{P}_3$, grows cubically with dimension $d$ using the interior-point method [26]. To make the estimation algorithm scalable to large problem sizes, we adopt the first-order method to solve it with modest accuracy within a reasonable time. The general idea is to first reformulate the original problems into the second-order cone programs (SOCPs) based on the principle of epigraph reformulation. Then we reformulate them as the standard cone programs, which are then solved by the ADMM algorithm solver SCS [24].

*1) SOCP Formulation via Epigraph Form:* We present the basic ideas of transforming the original problems $\mathscr{P}_1$, $\mathscr{P}_2$ and $\mathscr{P}_3$ into a standard conic optimization form. Specifically, we can rewrite the weighted matrix $\ell_1$-norm $\|\boldsymbol{W} \circ \boldsymbol{M}\|_1$ in problem $\mathscr{P}_1$ as

$$\|\boldsymbol{W} \circ \boldsymbol{M}\|_1 = \sum_{n=1}^{N}\sum_{k=1}^{K} W_{nk}(|\mathrm{Re}(M_{nk})| + |\mathrm{Im}(M_{nk})|), \tag{25}$$

where $\boldsymbol{W} = [W_{nk}] \in \mathbb{R}^{N\times K}$ and $\boldsymbol{M} = [M_{nk}] \in \mathbb{C}^{N\times K}$. Let $x_{nk,1} = \mathrm{Re}(M_{nk})$ and $x_{nk,2} = \mathrm{Im}(M_{nk})$, and based on the principle of the epigraph reformulation for convex optimization problem, we introduce slack variables $t_{nk,1}$ and $t_{nk,2}$ for the associated optimization variables $x_{nk,1}$ and $x_{nk,2}$, respectively. We thus reformulate problem $\mathscr{P}_1$ as

$$\begin{aligned} \underset{\boldsymbol{M},\boldsymbol{t}}{\text{minimize}} \quad & \sum_{n=1}^{N}\sum_{k=1}^{K} W_{nk}(t_{nk,1}+t_{nk,2}) \\ \text{subject to} \quad & \|\boldsymbol{Y}[\tau]-\boldsymbol{Q}[\tau]\boldsymbol{M}\|_F \leq \epsilon, \\ & |x_{nk,1}| \leq t_{nk,1}, |x_{nk,2}| \leq t_{nk,2}, \end{aligned} \tag{26}$$

which is then rewritten as the following form:

$$\begin{aligned} \underset{\boldsymbol{M},\boldsymbol{t}}{\text{minimize}} \quad & \sum_{n=1}^{N}\sum_{k=1}^{K} W_{nk}(t_{nk,1}+t_{nk,2}) \\ \text{subject to} \quad & \|\boldsymbol{Y}[\tau]-\boldsymbol{Q}[\tau]\boldsymbol{M}\|_F \leq \epsilon, \\ & -\boldsymbol{t} \preceq \boldsymbol{x} \preceq \boldsymbol{t}, \end{aligned} \tag{27}$$

where $\boldsymbol{t} \in \mathbb{R}^{2NK}$, $\boldsymbol{x} = [\text{Re}(M_{nk}), \text{Im}(M_{nk})] \in \mathbb{R}^{2NK}$, and $\preceq$ represents componentwise inequality. Similarly, problem $\mathscr{P}_2$ is recast as the following SOCP problem:

$$\begin{aligned} \underset{\boldsymbol{M},\boldsymbol{t},u,s}{\text{minimize}} \quad & \sum_{n=1}^{N}\sum_{k=1}^{K} W_{nk}(t_{nk,1}+t_{nk,2}) + \lambda s \\ \text{subject to} \quad & \|\boldsymbol{Y}[\tau]-\boldsymbol{Q}[\tau]\boldsymbol{M}\|_F \leq \epsilon, \\ & \|\boldsymbol{M}-\boldsymbol{\mathcal{H}}[\tau]\|_F \leq u, \\ & \|(1-s, 2u)\|_2 \leq 1+s, \\ & -\boldsymbol{t} \preceq \boldsymbol{x} \preceq \boldsymbol{t}, \end{aligned} \tag{28}$$

where $u \in \mathbb{R}$ and $s \in \mathbb{R}$. Following the same idea in [45], problem $\mathscr{P}_3$ can also be equivalently reformulated as the SOCP. Please refer to [24, Sec. 6.2] for the SOCP reformulation with the special case of individual sparsity Lasso, i.e., $\mathcal{R}(\boldsymbol{m}) = \|\boldsymbol{m}\|_1$ with $\boldsymbol{m} \in \mathbb{C}^{K\times 1}$ for the single receiver antenna case.

*2) Conic Formulation via Smith Form:* Based on the Smith form reformulation [27] supported by the software CVX for disciplined convex programming [46], all the SOCP formulations, e.g., (27) and (28), can be easily and automatically transform in turn into the following equivalent conic optimization form:

$$\begin{aligned} \mathscr{P}_{\text{cone}} : \underset{\boldsymbol{\nu},\boldsymbol{\mu}}{\text{minimize}} \quad & \boldsymbol{c}^T\boldsymbol{\nu} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{\nu}+\boldsymbol{\mu} = \boldsymbol{b} \\ & (\boldsymbol{\nu},\boldsymbol{\mu}) \in \mathbb{R}^n \times \mathcal{C}, \end{aligned} \tag{29}$$

where $\boldsymbol{\nu} \in \mathbb{R}^n$ and $\boldsymbol{\mu} \in \mathbb{R}^m$ are the optimization variables, $\mathcal{C} = \{0\}^r \times \mathcal{S}^{m_1} \times \cdots \times \mathcal{S}^{m_q}$ with $\mathcal{S}^p$ as the standard second-order cone of dimension $p$,

$$\mathcal{S}^p = \{(y,\boldsymbol{z}) \in \mathbb{R}\times\mathbb{R}^{p-1} | \|\boldsymbol{z}\| \leq y\}, \tag{30}$$

and $\mathcal{S}^1$ is defined as the cone of nonnegative reals, i.e., $\mathbb{R}_+$. Here, each $\mathcal{S}^i$ has dimension $m_i$ such that $(r+\sum_{i=1}^{q} m_i) = m$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{c} \in \mathbb{R}^n$. The equivalence means that the optimal solutions or the certificates of infeasibility of the original problem can be extracted from the solutions of the equivalent cone program $\mathscr{P}_{\text{cone}}$.

TABLE II
SIMULATION PARAMETERS

| Parameter | Value |
|---|---|
| Path-loss at distance $d_{kl}$ (km) | 148.1+37.6 $\log_2(d_{kl})$ dB |
| Standard deviation of log-norm shadowing $\sigma_s$ | 8 dB |
| Small-scale fading distribution $\boldsymbol{g}_{kl}$ | $\mathcal{CN}(\boldsymbol{0},\boldsymbol{I})$ |
| Noise power $\sigma_k^2$ (10 MHz bandwidth) | -92dBm |
| Transmit antenna power gain | 9 dBi |

### *B. Large-Scale Conic Optimization via Operator Splitting*

In this subsection, we adopt the operator splitting technique to solve the large-scale transformed conic optimization problem $\mathscr{P}_{\text{con}}$ (29) in parallel. To unify the capability of detecting infeasibility and computing the optimal solutions into a single system, the homogeneous self-dual embedding was proposed by adding extra variables into the Karush-Kuhn-Tucker (KKT) conditions [24] for the primal-dual programs of $\mathscr{P}_{\text{cone}}$. The second-order algorithms based on the interior-point method were then implemented in the well developed software packages, e.g., SDPT3 [47], to solve the homogeneous self-dual embedding automatically. However, the second-order algorithms are still computationally expensive and thus are not computationally feasible in dense Cloud-RANs. We thus present a novel first-order algorithm software package SCS based on the principle of operator splitting method, i.e., the ADMM algorithm [24], to solve the large-scale homogeneous self-dual embedding. The final iterative algorithm is presented as follows [24], [27]:

$$\tilde{\boldsymbol{x}}^{[i+1]} = (\boldsymbol{I}+\boldsymbol{J})^{-1}(\boldsymbol{x}^{[i]}+\boldsymbol{y}^{[i]}) \tag{31}$$
$$\boldsymbol{x}^{[i+1]} = \Pi_{\mathcal{V}}(\tilde{\boldsymbol{x}}^{[i+1]}-\boldsymbol{y}^{[i]}) \tag{32}$$
$$\boldsymbol{y}^{[i+1]} = \boldsymbol{y}^{[i]}-\tilde{\boldsymbol{x}}^{[i+1]}+\boldsymbol{x}^{[i+1]}, \tag{33}$$

where $\Pi_{\mathcal{V}}(\boldsymbol{x})$ denotes the Euclidean projection of $\boldsymbol{x}$ onto the convex set $\mathcal{V}$ with $\mathcal{V} = \mathbb{R}^n \times \mathcal{C}^* \times \mathbb{R}_+$ and $\mathcal{C}^*$ being the dual cone of $\mathcal{C}$, $\boldsymbol{x} = [\boldsymbol{\nu},\boldsymbol{\eta},\tau] \in \mathbb{R}^{m+n+1}$, $\boldsymbol{y} = [\boldsymbol{\lambda},\boldsymbol{\mu},\kappa] \in \mathbb{R}^{m+n+1}$ with $\tau \geq 0$ and $\kappa \geq 0$ being variables to encode different outcomes including the certificate of infeasibility and optimal solutions, $\tilde{\boldsymbol{x}}$ is the replicating variables of $\boldsymbol{x}$. Here, the $(m+n+1)\times(m+n+1)$ data matrix $\boldsymbol{J}$ is given by

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}^T & \boldsymbol{c} \\ -\boldsymbol{A} & \boldsymbol{0} & \boldsymbol{b} \\ -\boldsymbol{c}^T & -\boldsymbol{b}^T & \boldsymbol{0} \end{bmatrix}. \tag{34}$$

The subspace projection (31) can be efficiently computed using cached factorization approach [24], while the cone projection (32) can be computed in parallel with closed-forms using the proximal algorithms [48, Sec. 6]. In this paper, we shall use the numeric-based modeling framework CVX [46] to automatically transform the SOCP problems, e.g., (27) or (28), to the standard conic program form $\mathscr{P}_{\text{cone}}$, and then call the first-order optimization solver SCS to solve it efficiently.

## VII. Simulation Results

In this section, we simulate the proposed high-dimensional structured channel estimation algorithms in dense Cloud-RANs. We consider the following channel model [3]

$$\boldsymbol{h}_{kl}[\tau] = \underbrace{10^{-L(d_{kl})/20}\sqrt{\varphi_{kl}s_{kl}}}_{D_{kl}}\,\boldsymbol{g}_{kl}[\tau], \quad \forall \tau \in \mathcal{J}, \qquad (35)$$

where $L(d_{kl})$ is the pathloss at distance $d_{kl}$ in kilometer, $s_{kl}$ is the shadowing coefficient, $\varphi_{kl}$ is the antenna gain, $\boldsymbol{g}_{kl}$ is the small-scale fading coefficient and $D_{kl}$ is the associated large-scale fading coefficient. We adopt the standard cellular channel parameters as shown in Table II. To model the channel time variations (5) and (15), we adopt the model in [13] and [14] such that the corresponding innovation process is given by $\boldsymbol{v}_{kl}[\tau] = \sqrt{1-\eta^2}D_{kl}\boldsymbol{v}_{kl}[\tau]$ with $\boldsymbol{v}_{kl}[\tau] \sim \mathcal{CN}(\boldsymbol{0}_{N_l}, \boldsymbol{I}_{N_l})$, and $\boldsymbol{C} = \eta\boldsymbol{I}_N$ with $\eta = J_0(2\pi f_D\kappa)$ as the temporal correlation coefficient based on the Jakes' model, where $J_0$ is the 0-th order Bessel function of the first kind, $f_D$ denotes the maximum Doppler frequency and $\kappa$ represents the channel instantiation interval. We set $\eta = 0.9881$ based on the parameters provided in [14]. In this section, we assume that all the random variables, including small-scale fading coefficients, large-scale fading coefficients, innovation process and observation noise are mutually independently, due to the large distance between different antennas in Cloud-RANs. We use the unitary training matrix with equal power $p \geq 0$ per training symbol, i.e.,

$$\boldsymbol{Q}[\tau] \in \left\{\boldsymbol{U} \in \mathbb{C}^{T\times N}, \boldsymbol{U}\boldsymbol{U}^{\mathsf{H}} = p\cdot\boldsymbol{I}_T\right\}, \quad \forall \tau \in \mathcal{J}, \qquad (36)$$

which is optimal in i.i.d. Rayleigh fading channels [14]. We generate the unitary matrix from the orthogonal group using the method described in [49]. All the optimization problems $\mathscr{P}_1$, $\mathscr{P}_2$ and $\mathscr{P}_3$ are computed using the modeling framework CVX [46] with the large-scale convex optimization solver SCS [24]. The performance metric is given by the mean squared error [14], i.e., $\text{MSE} = \mathbb{E}[\|\hat{\boldsymbol{H}} - \boldsymbol{H}\|_F^2]/(NK)$.

### A. Channel Estimation With Spatial Prior Information

Consider a network with $L = 50$ single-antenna RRHs and $K = 30$ single-antenna MUs uniformly and independently distributed in the square region $[-5000, 5000] \times [-5000, 5000]$ meters. Each point of the simulation results is averaged over $10^4$ randomly generated network realizations (i.e., one small scaling fading and observation noise realization for each large-scale fading realization). The performance of different channel estimation algorithms with different available prior information assumptions is illustrated in Fig. 2 with the training length as $T = 10$. Transmit SNR is defined as the transmit power at per transmit antenna of per training symbol over the noise power, and "10% $\boldsymbol{D}$" means that there are only 10% largest large-scale fading coefficients available.

From this figure, we can see that the gap between the proposed estimator $\mathscr{P}_1$ and the MMSE estimator (which requires the Gaussian distribution assumption) is very small. In particular, estimator $\mathscr{P}_1$ with only 10% dominated large-scale fading coefficients can achieve almost the same performance as the one with full large-scale fading coefficients. This

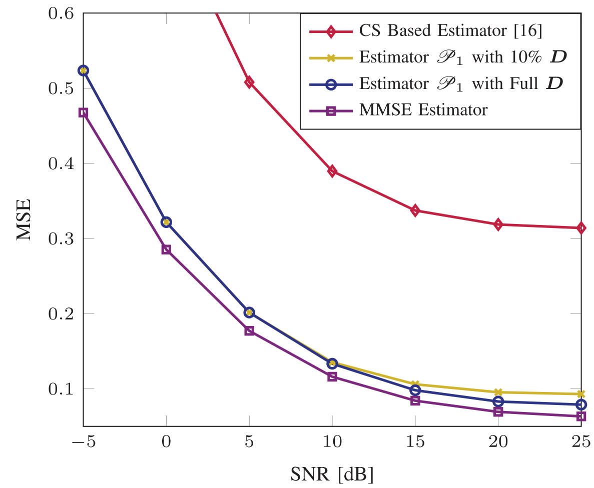



Fig. 2. Mean-squared error versus transmit SNR with $T = 10$.

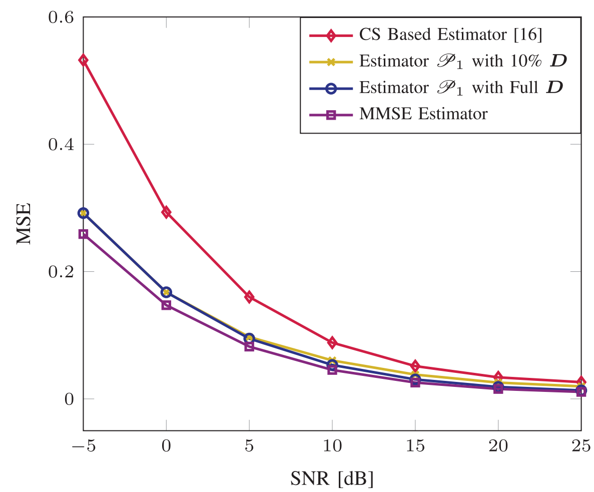



Fig. 3. Mean-squared error versus transmit SNR with $T = 30$.

substantially reduces the acquisition overhead of spatial prior information. We emphasize that the proposed estimator $\mathscr{P}_1$ does not require the Gaussian distribution assumption for all the variables in the system, but it requires the bound of the observation noise as a prior.

Furthermore, compared with the compressed sensing (CS) based channel estimation algorithm [17] (i.e., estimator $\mathscr{P}_1$ with equal weights, so it only exploits the channel sparsity without any prior information on the large-scale fading coefficients), it is clearly demonstrated that such approach does not work well in dense Cloud-RANs, while the estimation performance can be improved at the price of exploiting a reasonable amount of prior information with the proposed method. The main reason for the poor performance of the CS based estimator is that the channel spatial sparsity assumption in Cloud-RANs is still questionable [19]. Similar conclusion can be obtained with the training length $T = 30$ as shown in Fig. 3. In this case, the performance loss of the CS-based estimator is smaller. Besides, Fig. 4 demonstrates that setting the noise bound in the estimator $\mathscr{P}_1$ as the noise variance can achieve good performance.

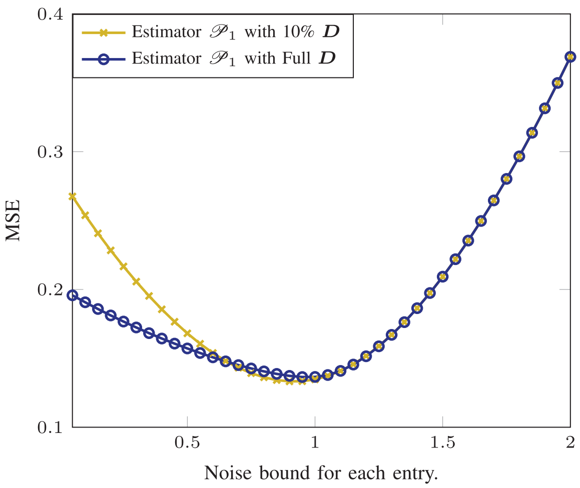



Fig. 4. Mean-squared error versus noise bound for each entry (the underlying noise variance is 1 for each entry).

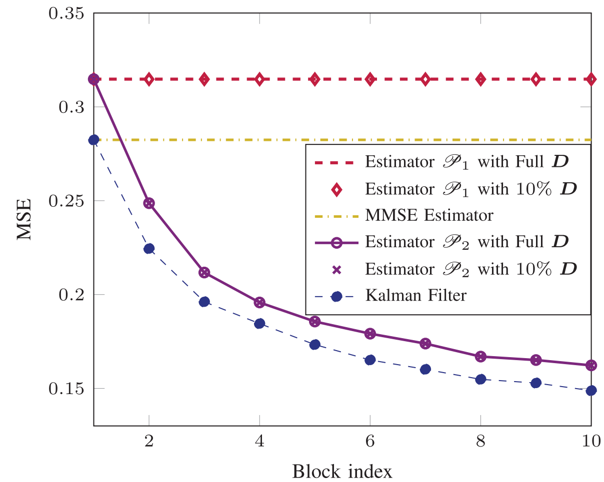



Fig. 5. Mean-squared error versus fading block index.

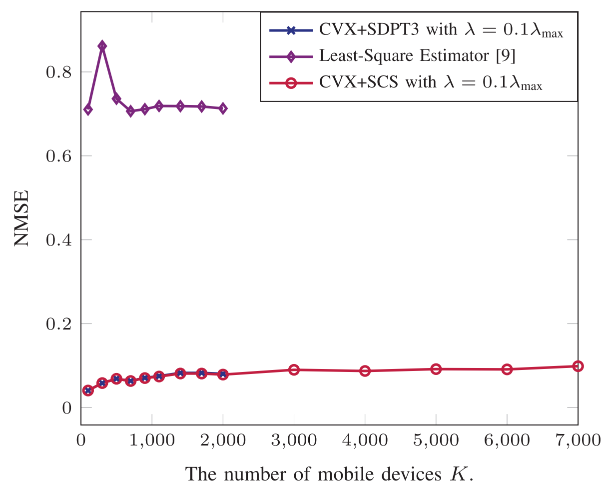



Fig. 6. Normalized mean-squared error versus the number of mobile devices.

### *B. Channel Estimation With Spatial-Temporal Prior Information*

Consider a network with $L = 25$ 2-antennas RRHs and $K = 5$ single-antenna MUs uniformly and independently distributed in the square region $[-10000, 10000] \times [-10000, 10000]$ meters. The training time is set to be $T = 5$ and $J$ is set to be 10 such that the spatial and temporal statistics keep the same during 10 blocks. The transmit SNR is fixed and set to be 0 dB. Each point of the simulation results in Fig. 5 is averaged over $10^4$ randomly generated network realizations (i.e., one small scaling fading, observation noise and innovation process realization for each large-scale fading realization).

For both the proposed estimators $\mathscr{P}_1$ and $\mathscr{P}_2$, from Fig. 5, we can see that knowing only the 10% dominated large-scale fading coefficients is enough to achieve the performance with the one with full large-scale fading coefficients. In particular, we set $\lambda = 1$ for estimator $\mathscr{P}_2$ to balance the spatial and temporal prior information. Note that, in principle, the performance of estimator $\mathscr{P}_2$ can be further improved by choosing an optimal parameter $\lambda$. However, this is not trivial and we leave it as our future work. From our experiments, in this simulated setting, we observe that $\lambda = 1$ yields good performance via cross validation.

Fig. 5 shows that the estimation performance can be significantly improved by further exploiting the temporal prior information for both the Kalman filter and the proposed estimator $\mathscr{P}_2$. In particular, as the block index increases, the performance improves, i.e., the estimation error decreases. The main reason is that both algorithms can exploit all the previous observations to make a good guess for the current channel. For the temporal prior information, the main difference between the proposed estimator $\mathscr{P}_2$ and Kalman filter is that the former only needs to know the temporal correlation coefficient $\eta$, while the Kalman filter further requires the innovation process covariance and the Gaussian distribution assumption for the state-space system. It is observed that, compared with the Kalman filter, estimator $\mathscr{P}_2$ can achieve good estimation performance with reduced spatial and temporal prior information, i.e., it only requires 10% dominated large-scale fading coefficients, the temporal correlation coefficient $\eta$ and a bound for the observation noise $\epsilon$.

### *C. High-Dimensional Uplink Channel Estimation for Massive Device Connectivity*

We consider a simple scenario with one single-antenna RRH with $K$ single-antenna mobile devices, in which the number of active users is $s = 0.05K$ and active users are generated at random. The pilot length $T$ is set to be $2s\log(K/s)$ (the optimal pilot length as obtained in [15] via convex geometry) for uplink training. For estimator $\mathscr{P}_3$, the channel matrix $\boldsymbol{H}$, pilot matrix $\boldsymbol{Q}$ and additive noise matrix $\boldsymbol{N}$ are generated independently from the distributions $\boldsymbol{H} = [h_{ij}]$ with $h_{ij} \sim \mathcal{CN}(0, 1)$, $\boldsymbol{Q} = [q_{ij}]$ with $q_{ij} \sim \mathcal{CN}(0, 1)$, and $\boldsymbol{N} = [n_{ij}]$ with $n_{ij} \sim \mathcal{CN}(0, \sigma^2)$ and $\sigma = 0.1$. We choose $\lambda = 0.1\lambda_{\max}$ for all problem instances with $\lambda_{\max} = \|\boldsymbol{Q}^{\mathsf{H}}\boldsymbol{Y}\|_\infty$ being the smallest value of $\lambda$ such that the solution to problem $\mathscr{P}_3$ is zero [24]. The normalized mean squared error (NMSE) is defined as $\text{NMSE} = \|\hat{\boldsymbol{M}} - \boldsymbol{M}_0\|_F^2 / \|\boldsymbol{M}_0\|_F^2$ with $\boldsymbol{M}_0 = \boldsymbol{\Pi}\boldsymbol{H}$ as the ground truth. Each point of the simulation results is averaged over

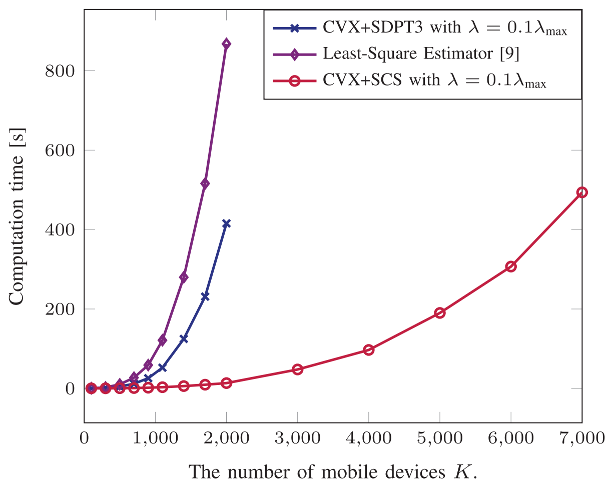



Fig. 7. Computation time versus the number of mobile devices.

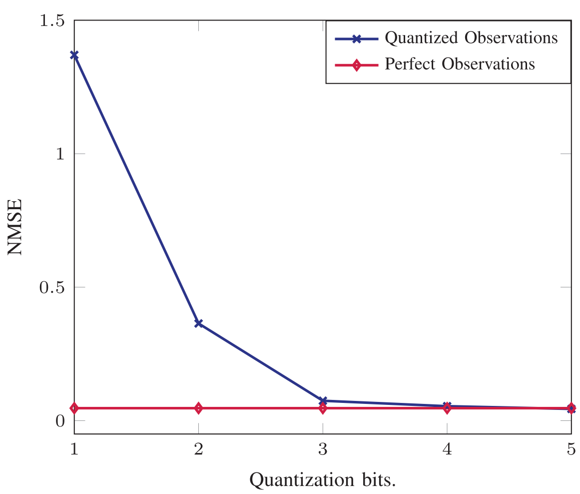



Fig. 8. The estimation error versus number of quantization bits.

50 generated network realizations, i.e., $\boldsymbol{H}$, $\boldsymbol{Q}$ and $\boldsymbol{N}$. We compute the estimator $\mathscr{P}_3$ using CVX+SDPT3 and CVX+SCS, where the solver SDPT3 is based on the second-order algorithm, i.e., interior-point method, while SCS is the ADMM based first-order algorithm solver presented in Section VI.

Fig. 6 shows that the proposed estimator $\mathscr{P}_3$ significantly outperforms the conventional least-square estimator [10] (computed via CVX+SDPT3), i.e., $\lambda = 0$. This indicates that it is critical to exploit the temporal sparsity of the active mobile devices to improve the performance for uplink channel estimation. Furthermore, both the second-order algorithm and the first-order algorithm achieve the same performance, while the first-order algorithm significantly reduces the computation time as demonstrated in Fig. 7, and thus it scales well to large network sizes.

Fig. 8 further investigates the estimation performance of the estimator $\mathscr{P}_3$ with quantization errors in the observations. Specifically, let $K = 200$ and other settings being the same as in Fig. 6 and Fig. 7. Each entry $y_i$ in the observation vector is quantized into the interval $[-10, 10]$ for the real part and imaginary part accordingly, which is divided into $2^q$ sub-intervals with equal length with $q$ as the number of quantization bits. Therefore, with the quantizer $\mathcal{Q}_i$, the quantized noise corrupted measurements are $z_i = \mathcal{Q}_i(y_i)$. Based on the quantized observations $z_i$, we apply estimator $\mathscr{P}_3$ to perform uplink channel estimation. From Fig. 8, we see that, as the number of quantization bits increase, the estimator approaches the performance of the unquantized case, and 4 quantization bits are sufficient for this case.

## VIII. Conclusions

In this paper, we proposed an efficient and flexible high-dimensional channel estimation framework for dense Cloud-RANs by exploiting spatial and temporal prior information. Novel structured regularizers and scalable first-order algorithms were presented to demonstrated the modeling flexibility and algorithmic efficiency of the proposed framework. The simulation results showed that the proposed high-dimensional channel estimation algorithms with substantially reduced prior information can achieve almost the same performance as the ones with full spatial and temporal prior information. In particular, it was demonstrated that the temporal sparsity plays a key role to provide radio access for massive mobile devices.

Several future research directions are listed as follows:

- While the convex formulations achieve good performance in simulations, it is interesting to characterize the fundamental estimation performance with the given prior information, thereby demonstrating the feasibility and robustness of high-dimensional channel estimation.
- To further speed up the present ADMM algorithms for real-time implementation of the convex high-dimensional estimation framework, it is interesting to adopt the deep neural networks to learn and approximate the algorithms, thereby improving the convergence rates [50].
- It is also interesting to develop the nonconvex regularizers to further encode more structured prior information and establish statistical and algorithmic guarantees [51].

## References

[1] X. Liu, Y. Shi, J. Zhang, and K. B. Letaief, "Massive CSI acquisition in dense cloud-RAN with spatial and temporal prior information," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[2] K. Chen and R. Duan, "C-RAN the road towards green RAN," China Mobile Res. Inst., Beijing, China, White Paper, 2011, vol. 2.

[3] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[4] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: system architectures, key techniques, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2282–2308, 3rd Quart., 2016.

[5] Y. Shi, J. Zhang, K. B. Letaief, B. Bai, and W. Chen, "Large-scale convex optimization for ultra-dense cloud-RAN," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 84–91, Jun. 2015.

[6] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[7] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[8] Y. Shi, J. Zhang, and K. B. Letaief, "Scalable coordinated beamforming for dense wireless cooperative networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp. 3603–3608.

[9] Y. Shi, J. Zhang, and K. B. Letaief, "Optimal stochastic coordinated beamforming for wireless cooperative networks with CSI uncertainty," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 960–973, Feb. 2015.

[10] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[11] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[12] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.

[13] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.

[14] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[15] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, pp. 805–849, Dec. 2012.

[16] M. J. Wainwright, "Structured regularizers for high-dimensional problems: Statistical and computational issues," *Annu. Rev. Statist. Appl.*, vol. 1, pp. 233–253, Jan. 2014.

[17] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[18] J.-C. Shen, J. Zhang, E. Alsusa, and K. B. Letaief, "Compressed CSI acquisition in FDD massive MIMO: How much training is needed?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4145–4156, Jun. 2016.

[19] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[20] Y. Shi, J. Zhang, and K. B. Letaief, "Low-rank matrix completion for topological interference management by Riemannian pursuit," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4703–4717, Jul. 2016.

[21] S. A. Jafar, "Topological interference management through index coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 529–568, Jan. 2014.

[22] 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, 2015.

[23] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, "A unified framework for high-dimensional analysis of $M$-estimators with decomposable regularizers," *Statist. Sci.*, vol. 27, no. 4, pp. 538–557, 2012.

[24] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *J. Optim. Theory Appl.*, vol. 169, no. 3, pp. 1042–1068, 2016.

[25] 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, Jan. 2011.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[27] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4729–4743, Sep. 2015.

[28] T. Q. S. Quek, M. Peng, O. Simeone, and W. Yu, *Cloud Radio Access Networks: Principles, Technologies, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[29] Y. Plan and R. Vershynin, "The generalized Lasso with non-linear observations," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1528–1537, Mar. 2016.

[30] S. A. Bhaskar, "Localization from connectivity: A 1-bit maximum likelihood approach," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2939–2953, Oct. 2016.

[31] R. T. Rockafellar, *Convex Analysis*, vol. 28. Princeton, NJ, USA: Princeton univ. Press, 1997.

[32] V. Chandrasekaran and M. I. Jordan, "Computational and statistical tradeoffs via convex relaxation," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 13, pp. E1181–E1190, 2013.

[33] A. Das and B. D. Rao, "SNR and noise variance estimation for MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3929–3941, Aug. 2012.

[34] S.-H. Lee and S.-Y. Chung, "Capacity scaling of wireless ad hoc networks: Shannon meets Maxwell," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1702–1715, Mar. 2012.

[35] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[36] M. E. Lopes, "Unknown sparsity in compressed sensing: Denoising and inference," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5145–5166, Sep. 2016.

[37] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: Essential theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 12–30, Nov. 2015.

[38] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4814–4826, Sep. 2015.

[39] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, p. 115002, 2015.

[40] A. Pananjady, M. J. Wainwright, and T. A. Courtade. (2016). "Linear regression with an unknown permutation: Statistical and computational limits." [Online]. Available: https://arxiv.org/abs/1608.02902

[41] 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, Aug. 2016.

[42] X. Xu, X. Rao, and V. K. N. Lau, "Active user detection and channel estimation in uplink CRAN systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 2727–2732.

[43] G. Hannak, M. Mayer, A. Jung, G. Matz, and N. Goertz, "Joint channel estimation and activity detection for multiuser communication systems," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, London, U.K., Jun. 2015, pp. 2086–2091.

[44] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, nos. 1–4, pp. 625–653, 1999.

[45] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, nos. 1–3, pp. 193–228, Nov. 1998.

[46] *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.0 (Beta)*, CVX Research, Inc., Austin, TX, USA, 2013.

[47] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3—A MATLAB software package for semidefinite programming, version 1.3," *Optim. Methods Softw.*, vol. 11, nos. 1–4, pp. 545–581, 1999.

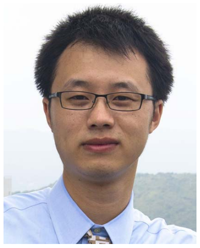
[48] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[49] F. Mezzadri, "How to generate random matrices from the classical compact groups," *Notices AMS*, vol. 54, pp. 592–604, Sep. 2007.

[50] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Barcelona, Spain, 2016, pp. 10–18.

[51] P.-L. Loh and M. J. Wainwright, "Regularized $M$-estimators with nonconvexity: Statistical and algorithmic theory for local optima," *J. Mach. Learn. Res.*, vol. 16, pp. 559–616, Mar. 2015.

**Xuan Liu** received the B.S. degree in telecommunication from Jilin University, Changchun, China, in 2000, the M.S. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2012, and the Ph.D. degree in electronic and information engineering from The University of Sydney, Australia, in 2017. She is currently a Post-Doctoral Fellow with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology. Her research interests include network information theory, large-scale optimization, big data analysis, and Internet of Things.



**Yuanming Shi** (S'13–M'15) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2011, and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, in 2015. Since 2015, he has been with the School of Information Science and Technology, ShanghaiTech University, as a tenure-track Assistant Professor. He visited the University of California at Berkeley, Berkeley, CA, USA, from 2016 to 2017. He was a recipient of the 2016 IEEE Marconi Prize Paper Award in wireless communications and the 2016 Young Author Best Paper Award by the IEEE Signal Processing Society. He was involved in optimization theory and algorithms, deep and machine learning, high-dimensional probability and statistics, with applications in intelligent IoT, large-scale data analysis, and distributed edge computing.

**Jun Zhang** (S'06–M'10–SM'15) received the B.Eng. degree in electronic engineering from the University of Science and Technology of China in 2004, the M.Phil. degree in information engineering from The Chinese University of Hong Kong in 2006, and the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin in 2009. He is currently a Research Assistant Professor with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology. He has co-authored the book *Fundamentals of LTE* (Prentice-Hall, 2010). His research interests include dense wireless cooperative networks, mobile edge computing, cloud computing, and big data analytics systems.

Dr. Zhang was a co-recipient of the 2016 Marconi Prize Paper Award in wireless communications, the 2014 Best Paper Award for the *EURASIP Journal on Advances in Signal Processing*, the IEEE GLOBECOM Best Paper Award in 2017, the IEEE ICC Best Paper Award in 2016, and the IEEE PIMRC Best Paper Award in 2014. A paper he co-authored received the 2016 Young Author Best Paper Award of the IEEE Signal Processing Society. He was also a recipient of the 2016 IEEE ComSoc Asia-Pacific Best Young Researcher Award. He frequently serves on the technical program committees of major IEEE conferences in wireless communications, such as ICC, Globecom, WCNC, and VTC. He served as a MAC Track Co-Chair for the IEEE WCNC 2011. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and was a Guest Editor of the special section on Mobile Edge Computing for Wireless Networks in the IEEE ACCESS.

**Khaled B. Letaief** (S'85–M'86–SM'97–F'03) received the B.S. (Hons.), M.S., and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1984, 1986, and 1990, respectively. From 1990 to 1993, he was a Faculty Member with The University of Melbourne, Melbourne, VIC, Australia. Since 1993, he has been with The Hong Kong University of Science and Technology (HKUST), where he is known as one of HKUST's most distinguished professors. While at HKUST, he has held numerous administrative positions, including the Head of the Electronic and Computer Engineering Department, the Director of the Huawei Innovation Laboratory, and the Director of the Hong Kong Telecom Institute of Information Technology. While at HKUST, he has also served as the Chair Professor and the Dean of Engineering. Under his leadership, the School has also dazzled in international rankings (rising from No. 26 in 2009 to 14 in the world in 2015 according to QS World University Rankings). In 2015, he joined Hamad Bin Khalifa University as a Provost to help establish a research-intensive university in Qatar in partnership with esteemed universities that include Northwestern University, Cornell University, and Texas A&M. He has authored over 580 journal and conference papers and given keynote talks and courses all over the world. He has 15 patents, including 11 U.S. patents. He is an internationally recognized leader in wireless communications with research interest in green communications, Internet of Things, Cloud-RANs, and 5G systems. He served as a consultant for different organizations, including Huawei, ASTRI, ZTE, Nortel, PricewaterhouseCoopers, and Motorola. In addition to his active research and professional activities, he has been a dedicated teacher committed to excellence in teaching and scholarship. He is a fellow of HKIE. He received the Michael G. Gale Medal for Distinguished Teaching (highest university-wide teaching award and only one recipient/year is honored for his/her contributions). He was also a recipient of many other distinguished awards and honors, including the 2007 IEEE Joseph LoCicero Publications Exemplary Award, the 2009 IEEE Marconi Prize Award in wireless communications, the 2010 Purdue University Outstanding Electrical and Computer Engineer Award, the 2011 IEEE Harold Sobol Award, the 2016 IEEE Marconi Prize Paper Award in wireless communications, and over 11 IEEE best paper awards. He is recognized as a long-time volunteer with dedicated service to professional societies and in particular IEEE, where he has served in many leadership positions including as Treasurer, Vice President for Conferences, and Vice President for Technical Activities of the IEEE Communications Society. He is also recognized by Thomson Reuters as an ISI Highly Cited Researcher (which places him among the top 250 preeminent individual researchers in the field of computer science and engineering) and is currently the President of the IEEE Communications Society in 2018. He has served on the editorial board of other prestigious journals. He is the founding Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.