Privacy-Preserving Channel Estimation in Cell-Free Hybrid Massive MIMO Systems

Jun Xu, Xiaodong Wang, *Fellow, IEEE*, Pengcheng Zhu, *Member, IEEE*, and Xiaohu You, *Fellow, IEEE*

***Abstract*— We consider a cell-free hybrid massive multiple-input multiple-output (MIMO) system with K users and M access points (APs), each with N_a antennas and $N_r < N_a$ radio frequency (RF) chains. When $K \ll MN_a$, efficient uplink channel estimation and data detection with reduced number of pilots can be performed based on low-rank matrix completion. However, such a scheme requires the central processing unit (CPU) to collect received signals from all APs, which may enable the CPU to infer the private information of user locations. We therefore develop and analyze privacy-preserving channel estimation schemes under the framework of differential privacy (DP). As the key ingredient of the channel estimator, two joint differentially private noisy matrix completion algorithms based respectively on Frank-Wolfe iteration and singular value decomposition are presented. We provide an analysis on the tradeoff between the privacy and the channel estimation error. In particular, we show that the estimation error can be mitigated while maintaining the same privacy level by increasing the payload size with fixed pilot size; and the scaling laws of both the privacy-induced and privacy-independent error components in terms of payload size are characterized. Simulation results are provided to further demonstrate the tradeoff between privacy and channel estimation performance.**

***Index Terms*— Cell-free, hybrid massive MIMO, channel estimation, location privacy, joint differentially private, matrix completion, Frank-Wolfe, singular value decomposition.**

I. Introduction

DUE to the high spectral and energy efficiencies, the cell-free massive MIMO has emerged as a promising wireless technology, where a large number of access points (APs) are distributed over a geographical area, and collaboratively serve users using the same time-frequency resource [1]. To reduce the high cost associated with equipping each antenna with a radio frequency (RF) chain that contains a high-resolution analog-to-digital converter (ADC) [2], hybrid analog/digital architectures are typically employed where with analog combining based on switches or phase shifters, antennas are randomly connected to a reduced number of RF chains and ADCs [3].

To enable cell-free hybrid massive MIMO systems, it is crucial to obtain accurate channel state information (CSI). In [3], a semi-blind channel estimation method based on low-rank matrix completion was proposed for hybrid massive MIMO, with the salient feature that the number of pilots is proportional to the number of users, instead of the number of antennas; and the estimation error reduces with the increase of the data payload size. In order to apply such channel estimation scheme in a cell-free system, each AP needs to send its observed received signal to a central processing unit (CPU), which performs channel estimation and data detection for all users. However, this may lead to the leakage of users' location information to the CPU, since the large scale fading of channels are determined by the locations of users and APs according to the path loss law.

Nowadays, the privacy awareness of the public has been significantly increased when using smart mobile devices and services. From the view point of users, privacy in 5G network can be divided into three main categories: data privacy, location privacy and identity privacy [4]. Locations are usually regarded as one of the most important sensitive information for most people, the leakage of which may pose threats to other sensitive information (e.g., home address, work place) and even personal safety [5]. Hence, it is crucial to provide high-quality services without disclosing the users' location privacy in 5G mobile networks. In [6], the location privacy of users is considered in the context of mobile traffic offloading system, where a system administrator coordinates assigning each mobile user one of multiple offloading stations. Each user sends its preference over all available offloading stations. Since such preference data can be used to infer the user location, it is required to be kept private from the system administrator. In this paper, we consider the same location privacy issue in the context of physical layer signal processing, i.e., channel estimation.

Three popular techniques for maintaining privacy include anonymization, data encryption and differential privacy [7]. Among them, anonymization strategies do not guarantee complete level of protection from adversaries; cryptographic techniques are computationally expensive; whereas differential privacy is easy to implement and provides provable privacy guarantee. Therefore in this paper, we aim to apply differential privacy to achieve privacy-preserving channel estimation.

Manuscript received June 3, 2020; revised October 25, 2020 and January 7, 2021; accepted January 12, 2021. Date of publication January 29, 2021; date of current version June 10, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFE0113400; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180011; and in part by the National Natural Science Foundation of China under Grant 61871122 and Grant 61971127. The associate editor coordinating the review of this article and approving it for publication was J. Yang. *(Corresponding author: Pengcheng Zhu.)*

Jun Xu, Pengcheng Zhu, and Xiaohu You are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: xujunseu@seu.edu.cn; p.zhu@seu.edu.cn; xhyu@seu.edu.cn).

Xiaodong Wang is with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: wangx@ee.columbia.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3053770.

Digital Object Identifier 10.1109/TWC.2021.3053770

TABLE I
TABLE OF NOTATIONS

Notation	Description
M, K, N_a, N_r	The number APs, users, antennas and RF chains in each AP
τ_{c}, τ_{p}, τ_{d}	The number of time slots within a coherence interval, for channel estimation and data transmission
$\mathcal{M}$, $\mathcal{K}$, $\mathcal{T}_{\mathrm{c}}$, $\mathcal{T}_{\mathrm{p}}$, $\mathcal{T}_{\mathrm{d}}$	The set of APs, users, time slots within a coherence, for channel estimation and data transmission
$\mathbf{h}_{k,m}$, $\mathbf{H}_m$, $\mathbf{H}$	The channel between user k and AP m, all users and AP m, all users and all APs
$\beta_{k,m}$, $\mathbf{g}_{k,m}$	the large-scale and small-scale fading of the channel between user k and AP m
σ^2, μ^2, ν^2	the variance of received noise, perturbation noise in Algorithm 1 and 2
ϵ, δ	Two privacy parameters that represent privacy level of DP
$\mathbf{S}$, $\mathbf{P}$, $\mathbf{D}$	The matrix of transmitted signal, pilot signal, data signal
Ω_m, Ω	The set of nonzero indices at AP m and all APs
$\mathbf{R}_m$, $\mathbf{X}_m$, $\mathbf{Y}_m$, $\mathbf{N}_m$	The received signal, noise-free received signal, observed signal and received noise at AP m
$\mathbf{R}$, $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{N}$	The stacked received signal, noise-free received signal, observed signal and received noise at all APs
$\widehat{\mathbf{X}}$, $\widehat{\mathbf{X}}_m$, $\widehat{\mathbf{H}}_m$, $\widehat{\mathbf{D}}$	The estimates of $\mathbf{X}$, $\mathbf{X}_m$, $\mathbf{H}_m$, $\mathbf{D}$
$\mathbf{X}_m^{(n)}$, T	The output at AP m in the n-th iteration and the number of iterations of Algorithm 1
L	The bound on $\|\mathbf{Y}_m\|_{\mathcal{F}}$

Differential privacy (DP) is a probabilistic framework based on the notion of indistinguishability [8]. In particular, observing an output of a differentially private algorithm, one cannot infer whether any specific user contributed to the data. In this framework, privacy is mainly achieved by randomizing the released statistics. DP has been accepted as a standard privacy model and widely adopted in many fields, such as recommender system [9], deep learning [10], distributed optimization [11], data mining [12], ridesharing services [5], etc. In addition, applications of DP in communication networks include data-driven caching in information-centric networks [13], and big data analytics in edge computing [14], [15]. However, so far there is no work addressing DP for physical-layer signal processing. A particular challenge is that unlike the above-mentioned higher-layer applications, the physical-layer is much more sensitive to the perturbation noise added to achieve DP.

In this paper, we aims to design privacy-preserving channel estimation algorithms for cell-free hybrid massive MIMO systems. The major contributions are summarized as follows:

- To the best of our knowledge, this is the first work that integrates DP with physical-layer signal processing.
- We propose two privacy-preserving channel estimators based on Frank-Wolfe (FW) iteration and singular value decomposition (SVD), respectively.
- We show that both channel estimation algorithms are joint differentially private. We also analyze the estimation error bounds for the two algorithms, and characterize the scaling laws of the estimation error in terms of data payload size.
- Through extensive simulations, we illustrate the tradeoff between privacy and channel estimation and data detection performance for the two algorithms.

The remainder of this paper is organized as follows. Section II describes the cell-free hybrid massive MIMO system under consideration and provides some background on DP. Two privacy-preserving channel estimation algorithms are proposed in Section III. Section IV presents the analysis on the privacy and channel estimation performance of the two algorithms. Simulation results are provided in Section V. Finally, Section VI concludes the paper.

Notations: Boldface letters denote matrices (upper case) or vectors (lower case). The transpose, conjugate transpose and trace operators are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ and $\mathrm{tr}\,(\cdot)$ respectively. $\|\cdot\|_{\mathcal{F}}$, $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{nuc}}$ denote the Frobenius norm, spectral norm and nuclear norm of a matrix, respectively. Assuming the singular values of a matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ are $\lambda_1, \cdots, \lambda_{\min(m,n)}$ in descending order, then we have $\|\mathbf{A}\|_{\mathcal{F}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{A}^2(i,j)} = \sqrt{\sum_{i=1}^{\min(m,n)}\lambda_i^2}$; $\|\mathbf{A}\|_2 = \lambda_1$; $\|\mathbf{A}\|_{\mathrm{nuc}} = \sum_{i=1}^{\min(m,n)}\lambda_i$. $\mathrm{Diag}\,(\mathbf{d})$ returns a diagonal matrix whose diagonal elements are given by a vector $\mathbf{d}$. $\mathbf{I}_M$, $\otimes$ and $\mathbb{E}\{\cdot\}$ respectively represent the $M \times M$ identity matrix, the Kronecker product and the expectation operator. $\mathcal{N}_{\mathrm{c}}(\mu, \sigma^2)$ and $\mathcal{N}(\mu, \sigma^2)$ respectively denote the complex and real circularly symmetric Gaussian distribution with mean μ and variance σ^2. $f(n) = \Theta(g(n))$ means f is bounded below by g asymptotically; $f(n) = O(g(n))$ means f is bounded above by g asymptotically; $f(n) = \omega(g(n))$ means f dominates g asymptotically. The descriptions of some notations used in this paper are summarized in Table I.

II. System Descriptions and Background

A. Signal Model

We consider a cell-free massive MIMO system, in which M distributed APs each equipped with N_a antennas collaboratively serve K single-antenna users using the same time-frequency resource, as shown in Fig. 1. We denote $\mathcal{M} = \{1, \ldots M\}$ and $\mathcal{K} = \{1, \ldots K\}$ as the sets of APs and users respectively. Each AP employs an analog structure with N_r RF chains to combine the incoming signal in the RF band. Each RF chain contains a high-resolution ADC and forwards the data stream to the baseband processor that performs only simple signal processing. All APs are connected to a CPU through perfect backhaul links, which performs computation-intensive signal processing.

We assume a block flat-fading channel between each user-AP pair. The channel coefficients remain constant during a coherence interval with τ_{c} time slots. Let $\mathcal{T}_{\mathrm{c}} = \{1, \cdots, \tau_{\mathrm{c}}\}$ denote the set of time slots within a coherence. Throughout the paper, we assume that $MN_a > \tau_{\mathrm{c}}$, which can be satisfied in massive MIMO. The first τ_{p} time slots denoted by

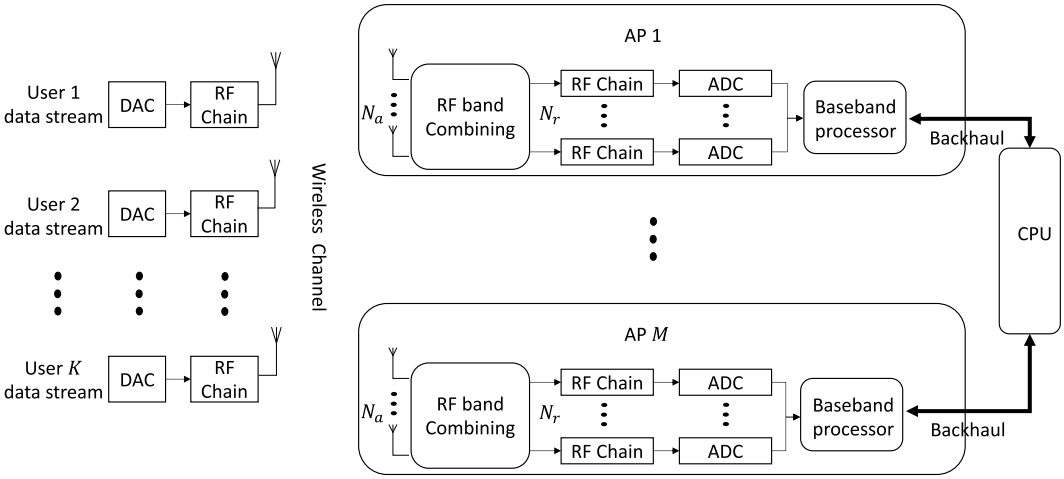


Fig. 1. The cell-free multiuser massive MIMO uplink system. Left: K single-antenna users. Right: M APs each with a hybrid structure consisting of N_a antennas and N_r RF chains. Each AP has a baseband processor with limited computational capability and all APs are connected to a CPU via perfect backhual links.

$\mathcal{T}_{\rm p} = \{1, \cdots, \tau_{\rm p}\}$ are used for uplink channel estimation, and the remaining $\tau_{\rm d} = \tau_{\rm c} - \tau_{\rm p}$ time slots denoted by $\mathcal{T}_{\rm d} = \{\tau_{\rm p+1}, \cdots, \tau_{\rm c}\}$ are used for uplink data transmission. The channel vector from user k to AP m can be modeled as

$$\mathbf{h}_{k,m} = \sqrt{\beta_{k,m}}\mathbf{g}_{k,m} \in \mathbb{C}^{N_a \times 1}, \tag{1}$$

where $\beta_{k,m}$ and $\mathbf{g}_{k,m} \in \mathcal{N}_{\rm c}\left(0, \mathbf{I}_{N_a}\right)$ respectively represent the large-scale and small-scale fading.

Let $\mathbf{s}\left[t\right] \sim \mathcal{N}_{\rm c}\left(0, \mathbf{I}_K\right)$ denote the transmitted signal from K users at time slot t, i.e., $\mathbf{s}\left[t\right]$ corresponds to pilots for $t \in \mathcal{T}_{\rm p}$ and data symbols for $t \in \mathcal{T}_{\rm d}$. The received signal $\mathbf{r}_m\left[t\right] \in \mathbb{C}^{N_a \times 1}$ across N_a antennas at AP m is given by

$$\mathbf{r}_m\left[t\right] = \mathbf{H}_m\mathbf{s}\left[t\right] + \mathbf{n}_m\left[t\right], \quad \forall t \in \mathcal{T}_{\rm c}, \ \forall m \in \mathcal{M}, \tag{2}$$

where $\mathbf{H}_m = \left[\mathbf{h}_{1,m}, \cdots, \mathbf{h}_{K,m}\right] \in \mathbb{C}^{N_a \times K}$ denotes the channel matrix between AP m and all users. $\mathbf{n}_m\left[t\right] \sim \mathcal{N}_{\rm c}\left(0, \sigma^2\mathbf{I}_{N_a}\right)$ is the received noise sample at AP m at time slot t. Denote $\mathbf{R}_m \triangleq \left[\mathbf{r}_m\left[1\right], \cdots, \mathbf{r}_m\left[\tau_{\rm c}\right]\right] \in \mathbb{C}^{N_a \times \tau_{\rm c}}$, $\mathbf{N}_m \triangleq \left[\mathbf{n}_m\left[1\right], \cdots, \mathbf{n}_m\left[\tau_{\rm c}\right]\right] \in \mathbb{C}^{N_a \times \tau_{\rm c}}$, $\mathbf{P} \triangleq \left[\mathbf{s}\left[1\right], \cdots, \mathbf{s}\left[\tau_{\rm p}\right]\right] \in \mathbb{C}^{K \times \tau_{\rm p}}$, $\mathbf{D} \triangleq \left[\mathbf{s}\left[\tau_{\rm p}+1\right], \cdots, \mathbf{s}\left[\tau_{\rm c}\right]\right] \in \mathbb{C}^{K \times \tau_{\rm d}}$, $\mathbf{S} = \left[\mathbf{P}, \mathbf{D}\right] \in \mathbb{C}^{K \times \tau_{\rm c}}$. Then (2) can be rewritten as

$$\mathbf{R}_m = \mathbf{H}_m\mathbf{S} + \mathbf{N}_m, \quad \forall m \in \mathcal{M}. \tag{3}$$

By stacking the signals from all APs and denoting $\mathbf{R} = \left[\mathbf{R}_1^{\rm T}, \cdots, \mathbf{R}_M^{\rm T}\right]^{\rm T} \in \mathbb{C}^{MN_a \times \tau_{\rm c}}$, $\mathbf{H} = \left[\mathbf{H}_1^{\rm T}, \cdots, \mathbf{H}_M^{\rm T}\right]^{\rm T} \in \mathbb{C}^{MN_a \times K}$ and $\mathbf{N} = \left[\mathbf{N}_1^{\rm T}, \cdots, \mathbf{N}_M^{\rm T}\right]^{\rm T} \in \mathbb{C}^{MN_a \times \tau_{\rm c}}$, we then have

$$\mathbf{R} = \mathbf{H}\mathbf{S} + \mathbf{N}. \tag{4}$$

Note that $\mathrm{rank}\left(\mathbf{H}\mathbf{S}\right) \leq K$ and we assume that $MN_a \gg K$ and $\tau_{\rm c} \gg K$, hence $\mathbf{R}$ is a noisy version of a low-rank matrix.

Then $\mathbf{r}_m\left[t\right]$ will pass through analog structures and be combined in the RF band. In this paper, we consider the analog combining based on switches, where each RF chain is randomly connected to one of N_a antennas through a switch at each time slot [3].[1] Such antenna selection can capture many advantages of massive MIMO and has low power consumption [19]. Denote Ω_m as the set of indices (n, t) such that the n-th element of $\mathbf{r}_m\left[t\right]$, $\mathbf{R}_m(n, t)$ is observed. Denote $\mathbf{Y}_m = \left(\mathbf{R}_m\right)_{\Omega_m} \in \mathbb{C}^{N_a \times \tau_{\rm c}}$ as the sampled version of $\mathbf{R}_m$ such that

$$\mathbf{Y}_m\left(n, t\right) = \begin{cases} \mathbf{R}_m\left(n, t\right), & \text{if } (n, t) \in \Omega_m, \\ 0, & \text{if } (n, t) \notin \Omega_m. \end{cases} \tag{5}$$

Note that each column of $\mathbf{Y}_m$ has exactly N_r non-zero elements.

Traditionally when there is no privacy concern, to fully exploit the low-rank structure of $\mathbf{R}$ in (4) and the computing power of CPU, each AP m will send its received signal $\mathbf{Y}_m$ to the CPU. The CPU will then estimate the channels $\{\mathbf{H}_m, m \in \mathcal{M}\}$ and the user data payload $\mathbf{D}$, based on $\{\mathbf{Y}_m, m \in \mathcal{M}\}$ and the pilots $\mathbf{P}$. However, note that the large-scale fading coefficient $\beta_{k,m}$ in (1) contains the path loss information which is in turn determined by the distance between the user-AP pair. Since the locations of APs are fixed, the location of user k can be accurately estimated if its distances from more than three APs are known. Hence if each AP m directly sends $\mathbf{Y}_m$ to the CPU, then the location information of users might be compromised once the CPU obtains accurate estimates of the channels $\{\mathbf{H}_m, m \in \mathcal{M}\}$. Hence, in order to protect the location privacy of users, each AP m cannot send its $\mathbf{Y}_m$ directly to the CPU. Instead, the APs and the CPU should collaborate in a privacy-preserving way such that the estimate of channel $\mathbf{H}_m$ is only available to AP m but not to the CPU or other APs.

[1] Such frequent antenna switching requires time-limited pulse shaping, which may lead to lower bandwidth efficiency than time-orthogonal pulse shaping. Note that the spatial modulation (SM) MIMO systems adopt the similar antenna switching mechanism [16]–[18]. It was shown in [16], [17] that the single-RF SM MIMO exhibits higher bandwidth efficiency than the conventional full-RF MIMO when the number of antennas is large, which is typical in massive MIMO.

This paper focuses on the design and analysis of such privacy-preserving channel estimation schemes. In the next subsection, we provide a general overview of the notion of differential privacy (DP) and basic approaches to achieving DP.

B. Background on Differential Privacy (DP)

Recall that $\mathbf{Y}_m$ and $\mathbf{H}_m$ denote the received signal and the channel at AP m, respectively. Denote $\widehat{\mathbf{H}}_m$ as an estimate of $\mathbf{H}_m$, $\mathbf{Y} = \left[\mathbf{Y}_1^{\mathrm{T}}, \cdots, \mathbf{Y}_M^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{MN_a \times \tau_c}$ and $\widehat{\mathbf{H}} = \left[\widehat{\mathbf{H}}_1^{\mathrm{T}}, \cdots, \widehat{\mathbf{H}}_M^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{MN_a \times K}$. Let $\mathcal{A} : \mathbb{C}^{MN_a \times \tau_c} \rightarrow \mathbb{C}^{MN_a \times K}$ be a randomized channel estimation algorithm which takes $\mathbf{Y}$ as input, and outputs $\widehat{\mathbf{H}}$. In addition, $\mathcal{A}_{-m} : \mathbb{C}^{MN_a \times \tau_c} \rightarrow \mathbb{C}^{(M-1)N_a \times K}$ outputs $\widehat{\mathbf{H}}_{-m}$, which denotes the estimated channels for all APs other than AP m. Similarly denote $\mathbf{Y}_{-m} \in \mathbb{C}^{(M-1)N_a \times \tau_c}$ as $\mathbf{Y}$ with $\mathbf{Y}_m$ removed. Recall that our goal is to devise the channel estimation algorithm $\mathcal{A}$ such that $\widehat{\mathbf{H}}_m$ is available only to AP m, but not to the CPU or other APs.

Definition 1 (Standard DP [20] and Joint DP [21]): Given $\epsilon, \delta > 0$, $\mathcal{A}$ is (ϵ, δ)-differentially private if for any AP m, any two possible values $\mathbf{Y}_m, \mathbf{Y}'_m$ of the received signal at AP m, any value $\mathbf{Y}_{-m}$ of the received signal at all other APs, and any subset $S \in \mathbb{C}^{MN_a \times K}$, we have

$$P\left[\mathcal{A}\left(\left[\mathbf{Y}_m, \mathbf{Y}_{-m}\right]\right) \in S\right] \leq e^{\epsilon} P\left[\mathcal{A}\left(\left[\mathbf{Y}'_m, \mathbf{Y}_{-m}\right]\right) \in S\right] + \delta, \quad (6)$$

where the probability $P[\cdot]$ is over the randomness of $\mathcal{A}$. Moreover, $\mathcal{A}$ is (ϵ, δ)-joint differentially private if for any subset $S \in \mathbb{C}^{(M-1)N_a \times K}$, we have

$$P\left[\mathcal{A}_{-m}\left(\left[\mathbf{Y}_m, \mathbf{Y}_{-m}\right]\right) \in S\right] \leq e^{\epsilon} P\left[\mathcal{A}_{-m}\left(\left[\mathbf{Y}'_m, \mathbf{Y}_{-m}\right]\right) \in S\right] + \delta. \quad (7)$$

The meaning of the above standard DP is that a change of the received signal $\mathbf{Y}_m$ at any AP m has a negligible impact on the estimated channel $\widehat{\mathbf{H}}$. Hence, one cannot infer much about the private data $\mathbf{Y}_m$ from the output $\widehat{\mathbf{H}}$ and the data $\mathbf{Y}_{-m}$, which however, implies that the estimated $\widehat{\mathbf{H}}_m$ of AP m should not depend strongly on $\mathbf{Y}_m$. Obviously such channel estimator will not be of practical value. On the other hand, joint DP means that the estimate $\widehat{\mathbf{H}}_m$ for any particular AP m can depend strongly on its data $\mathbf{Y}_m$; but the output of all other APs $\widehat{\mathbf{H}}_{-m}$ and $\mathbf{Y}_{-m}$ do not reveal much about $\mathbf{Y}_m$, which is a meaningful notion of privacy in the context of channel estimation and will be adopted hereafter. Here, ϵ and δ represent the worst-case privacy loss and smaller values of them imply stronger privacy guarantee.

Next we review two important properties of DP, which hold for both standard DP and joint DP.

Lemma 1 (Post-Processing [22]): Let $\mathcal{A} : \mathcal{T} \rightarrow \mathcal{S}$ and $\mathcal{B} : \mathcal{S} \rightarrow \mathcal{Q}$ be randomized algorithms. Define an algorithm $\mathcal{A}' : \mathcal{T} \rightarrow \mathcal{Q}$ by $\mathcal{A}' = \mathcal{B}(\mathcal{A})$. If $\mathcal{A}$ satisfies (ϵ, δ)-(joint) DP, then $\mathcal{A}'$ also satisfies (ϵ, δ)-(joint) DP.

Hence any operation performed on the output of a (joint) differentially private algorithm, without accessing the raw data, remains (joint) differentially private with the same level of privacy.

Lemma 2 (T-Fold Composition [8], [22]): We assume that there are T independent randomized algorithms $\mathcal{A}_1, \cdots, \mathcal{A}_T$, where each algorithm $\mathcal{A}_t : \mathcal{D} \rightarrow \mathcal{R}_t$ is (ϵ, δ)-(joint) differentially private. For all $\epsilon, \delta, \delta' > 0$, an algorithm $\mathcal{A}$ defined as $\mathcal{A}(\mathcal{D}) = (\mathcal{A}_1(\mathcal{D}), \cdots, \mathcal{A}_T(\mathcal{D}))$ satisfies $(\epsilon', T\delta + \delta')$-(joint) DP with

$$\epsilon' = \epsilon\sqrt{2T\ln(1/\delta')} + T\epsilon\left(e^{\epsilon} - 1\right). \quad (8)$$

In particular, given target privacy parameters $0 < \epsilon' < 1$ and $\delta' > 0$, $\mathcal{A}$ satisfies $(\epsilon', T\delta + \delta')$-(joint) DP if each algorithm is $\left(\epsilon'/\sqrt{8T\ln(1/\delta')}, \delta\right)$-(joint) differentially private.

In the context of channel estimation, if $\mathbf{Y}$ is accessed by CPU T times, each denoted by $\mathcal{A}_t(\mathbf{Y}), t = 1, \cdots, T$, then the information released to CPU is $\mathcal{A}(\mathbf{Y}) = (\mathcal{A}_1(\mathbf{Y}), \cdots, \mathcal{A}_T(\mathbf{Y}))$. To make $\mathcal{A}(\mathbf{Y})$ joint differentially private, we need to guarantee that each access $\mathcal{A}_t(\mathbf{Y})$ is joint differentially private. In addition, the difference in the privacy level between $\mathcal{A}(\mathbf{Y})$ and $\mathcal{A}_t(\mathbf{Y})$ stated in this lemma will be useful in the design of the privacy-preserving channel estimator.

The following important lemma provides us a way to achieve joint DP by standard DP.

Lemma 3 (Billboard Lemma [23]): Suppose $\mathcal{A} : \mathcal{D} = (\mathcal{D}_1, \cdots, \mathcal{D}_M) \rightarrow \mathcal{R}$ is (ϵ, δ)-differentially private, where $\mathcal{D}_m$ denotes the data of AP m. If a randomized algorithm $\mathcal{B}$ has M components with the m-th component $\mathcal{B}_m(\mathcal{D}_m, \mathcal{A}(\mathcal{D}))$, where $\mathcal{B}_m : \mathcal{D}_m \times \mathcal{R} \rightarrow \mathcal{Q}_m, \forall m \in \mathcal{M}$, then $\mathcal{B}$ is (ϵ, δ)-joint differentially private.

Next we review the definition of ℓ_2-sensitivity and a well-known approach to achieving standard DP.

Definition 2 (ℓ_2-Sensitivity [22]): Let $f(\mathbf{Y})$ be an arbitrary function on the received signal $\mathbf{Y} = \left[\mathbf{Y}_1^{\mathrm{T}}, \cdots, \mathbf{Y}_M^{\mathrm{T}}\right]^{\mathrm{T}}$. Then its ℓ_2-sensitivity is defined as the maximum difference in the function values when the received signals differ only at one AP, i.e.,

$$\Delta_f \triangleq \max_{1 \leq m \leq M} \max_{\substack{\mathbf{Y}_m \neq \mathbf{Y}'_m \\ \mathbf{Y}_i = \mathbf{Y}'_i, \forall i \neq m}} \left\| f(\mathbf{Y}) - f(\mathbf{Y}') \right\|_{\mathcal{F}}. \quad (9)$$

Lemma 4 (Gaussian Mechanism [22]): Assuming that the information released to CPU during channel estimation is

$$\mathcal{A}(\mathbf{Y}) = f(\mathbf{Y}) + \mathbf{G}, \quad (10)$$

where $f(\mathbf{Y})$ has the ℓ_2-sensitivity Δ_f and

$$\mathbf{G}(i,j) \overset{\text{i.i.d.}}{\sim} \mathcal{N}_{\mathrm{c}}\left(0, \frac{\Delta_f^2 2\ln(1.25/\delta)}{\epsilon^2}\right). \quad (11)$$

Then the released $\mathcal{A}(\mathbf{Y})$ satisfies (ϵ, δ)-DP.

The above lemma helps us to calibrate the Gaussian perturbation noise to achieve (ϵ, δ)-DP. It can be seen that larger perturbation noise is required to achieve stronger privacy level, i.e., smaller ϵ and/or δ, which is intuitive because larger noise variance increases the uncertainty about the released information and hence improves privacy.

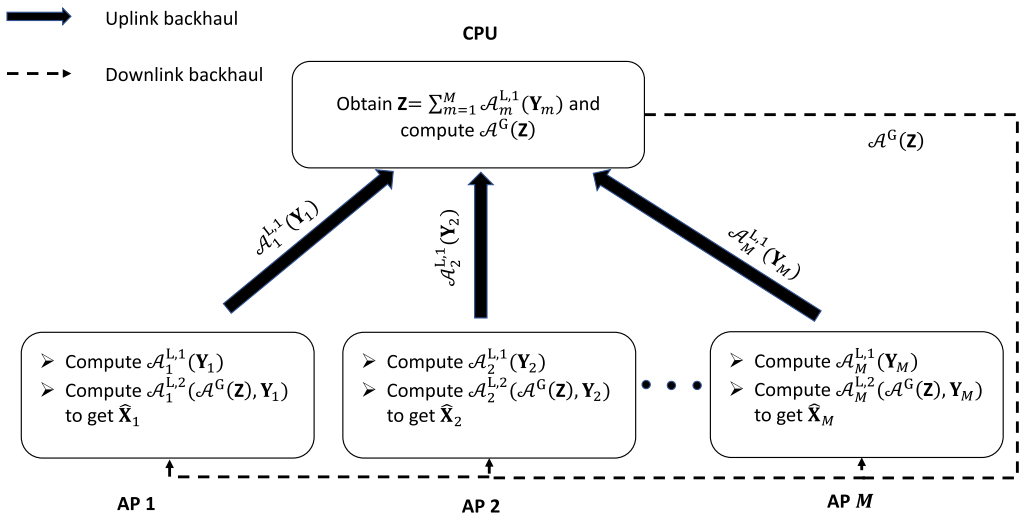


Fig. 2. The global-local computation model of the proposed privacy-preserving noisy matrix completion algorithm.

III. Privacy Preserving Channel Estimation

In this section, we first show that the key component of the privacy-preserving channel estimator is privacy-preserving matrix completion. We then outline two such matrix completion algorithms.

A. Channel Estimation Based on Matrix Completion

Recall that in (4), $\mathbf{X} = \mathbf{HS} \in \mathbb{C}^{MN_a \times \tau_c}$ is low-rank, i.e., $\mathrm{rank}\,(\mathbf{X}) \leq K$, and $\mathbf{Y}$ is an incomplete observation of $\mathbf{X}$ corrupted by channel noise $\mathbf{N}$. We first design a privacy-preserving algorithm $\mathcal{A}$ to solve a noisy matrix completion problem, which takes $\mathbf{Y}$ as input and outputs a low-rank matrix $\widehat{\mathbf{X}} = \left[\left(\widehat{\mathbf{X}}_1\right)^{\mathrm{T}}, \cdots, \left(\widehat{\mathbf{X}}_M\right)^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{MN_a \times \tau_c}$ as an estimate of $\mathbf{X} = \left[\mathbf{X}_1^{\mathrm{T}}, \cdots, \mathbf{X}_M^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\mathbf{X}_m = \mathbf{H}_m\mathbf{S}$. Then, the estimation of channel $\mathbf{H}_m$ can be performed locally at AP m based on $\widehat{\mathbf{X}}_m$.

To satisfy (ϵ, δ)-joint DP, the matrix completion algorithm $\mathcal{A}$ consists of a global component $\mathcal{A}^{\mathrm{G}}$ at the CPU and M local components $\mathcal{A}_m^{\mathrm{L}} = \left(\mathcal{A}_m^{\mathrm{L},1}, \mathcal{A}_m^{\mathrm{L},2}\right), m = 1, \cdots, M$ at M APs. The local component at AP m first performs a privacy-preserving computation $\mathcal{A}_m^{\mathrm{L},1}$ on private $\mathbf{Y}_m$ to get $\mathcal{A}_m^{\mathrm{L},1}(\mathbf{Y}_m)$, and then transmits it to the CPU through backhaul links. The global component $\mathcal{A}^{\mathrm{G}}$ aggregates the received information $\mathbf{Z} = \sum_{m=1}^{M} \mathcal{A}_m^{\mathrm{L},1}(\mathbf{Y}_m)$, and computes $\mathcal{A}^{\mathrm{G}}(\mathbf{Z})$, which is then broadcast to all APs through backhaul links. Based on the public $\mathcal{A}^{\mathrm{G}}(\mathbf{Z})$ and private $\mathbf{Y}_m$, the local component at AP m then performs a computation $\mathcal{A}_m^{\mathrm{L},2}$ to get a complete matrix $\widehat{\mathbf{X}}_m = \mathcal{A}_m^{\mathrm{L},2}\left(\mathcal{A}^{\mathrm{G}}(\mathbf{Z}), \mathbf{Y}_m\right)$. Such a global-local computation model for privacy-preserving noisy matrix completion is depicted in Fig. 2.

With $\widehat{\mathbf{X}}_m$, each AP m can proceed to estimate its channel $\mathbf{H}_m$ as follows. Recall that $\widehat{\mathbf{X}}_m \equiv \left[\widehat{\mathbf{X}}_{m,\mathrm{p}}, \widehat{\mathbf{X}}_{m,\mathrm{d}}\right]$, where $\widehat{\mathbf{X}}_{m,\mathrm{p}} \in \mathbb{C}^{N_a \times \tau_p}$ and $\widehat{\mathbf{X}}_{m,\mathrm{d}} \in \mathbb{C}^{N_a \times \tau_d}$ denote the estimates of $\mathbf{X}_{m,\mathrm{p}} = \mathbf{H}_m\mathbf{P}$ and $\mathbf{X}_{m,\mathrm{d}} = \mathbf{H}_m\mathbf{D}$ respectively. Then the estimate of $\mathbf{H}_m$ is given by

$$\widehat{\mathbf{H}}_m = \widehat{\mathbf{X}}_{m,\mathrm{p}}\mathbf{P}^{\dagger}, \quad \forall m \in \mathcal{M}, \tag{12}$$

where $\mathbf{P}^{\dagger}$ denotes the pseudo-inverse of $\mathbf{P}$ which is pre-stored in each AP. Finally, for data detection, the optimal scheme is to compute $\widehat{\mathbf{D}} = \widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{X}}_{\mathrm{d}}$ at the CPU. However, since the privacy constraint prevents the CPU from having access to $\{\widehat{\mathbf{H}}_m\}$, we let each AP compute the local statistic

$$\widehat{\mathbf{D}}_m = \widehat{\mathbf{H}}_m^{\dagger}\widehat{\mathbf{X}}_{m,\mathrm{d}}, \quad \forall m \in \mathcal{M}. \tag{13}$$

Then $\left\{\widehat{\mathbf{D}}_m\right\}$ are sent to the CPU which performs data detection based on the combined statistic

$$\widehat{\mathbf{D}} = \frac{1}{M}\sum_{m=1}^{M} \widehat{\mathbf{D}}_m. \tag{14}$$

Next we provide two privacy-preserving noisy matrix completion algorithms based on the FW algorithm and SVD algorithm, respectively, both of which are in the form of the global-local computation model.

B. FW-Based Privacy-Preserving Matrix Completion

Recall that in the absence of noise, we have $\mathbf{Y} = (\mathbf{X})_{\Omega}$ and $\mathrm{rank}\,(\mathbf{X}) \leq K$. However, the rank constraint is nonconvex. A popular approach to noisy matrix completion is based on the following least-squares formulation with the nuclear norm constraint [24]–[27]

$$\widehat{\mathbf{X}} = \arg\min_{\|\mathbf{X}\|_{\mathrm{nuc}} \leq K} \|(\mathbf{X})_{\Omega} - \mathbf{Y}\|_{\mathcal{F}}^2, \tag{15}$$

which is a convex problem. The Frank-Wolfe (FW) algorithm is an iterative procedure to solve (15), given by

$$\mathbf{X}^{(n)} = \left(1 - \eta^{(n)}\right)\mathbf{X}^{(n-1)} - \frac{\eta^{(n)}K}{\lambda^{(n)}}\mathbf{J}^{(n-1)}\mathbf{v}^{(n)}\left(\mathbf{v}^{(n)}\right)^{\mathrm{H}}, \tag{16}$$

where $\eta^{(n)}$ is the step size at the n-th iteration; $\mathbf{X}^{(0)} = \mathbf{0}_{MN_a \times \tau_c}$; $\mathbf{J}^{(n-1)} = \left(\mathbf{X}^{(n-1)}\right)_{\Omega} - \mathbf{Y}$. $\lambda^{(n)}$ and $\mathbf{v}^{(n)} \in \mathbb{C}^{\tau_c \times 1}$ are respectively the largest singular value and right-singular vector of $\mathbf{J}^{(n-1)}$. Hence each FW update is in terms of a rank-one matrix with nuclear norm at most K. After T iterations, the rank of $\mathbf{X}^{(T)}$ is at most T. In addition, it is known that

$\mathbf{X}^{(T)}$ can approach the optimal solution to (15) when T is large [25].

Recall that $MN_a > \tau_c$, thus $\left(\lambda^{(n)}\right)^2$ and $\mathbf{v}^{(n)}$ are the largest eigenvalue and eigenvector of $\left(\mathbf{J}^{(n-1)}\right)^{\mathrm{H}}\mathbf{J}^{(n-1)} = \sum_{m=1}^{M}\left(\mathbf{J}_m^{(n-1)}\right)^{\mathrm{H}}\mathbf{J}_m^{(n-1)}$, respectively, where

$$\mathbf{J}_m^{(n-1)} = \left(\mathbf{X}_m^{(n-1)}\right)_{\Omega_m} - \mathbf{Y}_m, \tag{17}$$

which can be computed at AP m. Hence, (16) can be rewritten into a global-local structure as

$$\mathbf{X}_m^{(n)} = \left(1-\eta^{(n)}\right)\mathbf{X}_m^{(n-1)} - \frac{\eta^{(n)}K}{\lambda^{(n)}}\mathbf{J}_m^{(n-1)}\mathbf{v}^{(n)}\left(\mathbf{v}^{(n)}\right)^{\mathrm{H}}, \quad \forall m \in \mathcal{M}. \tag{18}$$

At the $(n-1)$th iteration, each AP m computes

$$\mathbb{J}_m^{(n-1)} = \left(\mathbf{J}_m^{(n-1)}\right)^{\mathrm{H}}\mathbf{J}_m^{(n-1)}, \tag{19}$$

and sends it to the CPU. Then the CPU computes the largest eigenvalue $\lambda^{(n)}$ and corresponding eigenvector $\mathbf{v}^{(n)}$ of $\mathbb{W}^{(n-1)} = \sum_{m=1}^{M}\mathbb{J}_m^{(n-1)}$. At the n-th iteration, $\lambda^{(n)}$ and $\mathbf{v}^{(n)}$ are broadcast to all APs and then each AP m computes $\mathbf{X}_m^{(n)}$ according to (18). However, because $\mathbb{J}_m^{(n-1)}$ contains the private data $\mathbf{Y}_m$, it should not be sent directly from AP m to the CPU.

To preserve the privacy, we add Gaussian noise with calibrated variance to perturb $\mathbb{J}_m^{(n-1)}$. Each AP m sends the perturbed matrix $\widehat{\mathbb{J}}_m^{(n-1)} = \mathbb{J}_m^{(n-1)} + \mathbf{G}_m^{(n-1)}$ to the CPU, where $\mathbf{G}_m^{(n-1)}$ is a $\tau_c \times \tau_c$ Hermitian noise matrix, whose upper triangular and diagonal elements are respectively i.i.d. $\mathcal{N}_c\left(0,\mu^2\right)$ and $\mathcal{N}\left(0,\mu^2\right)$ samples, and the lower triangular elements are complex conjugates of the upper triangular counterparts. To calibrate the perturbation noise variance μ, we need an upper bound on $\|\mathbf{Y}_m\|_{\mathcal{F}}, \forall m \in \mathcal{M}$, which can be obtained as follows. Note that $\|\mathbf{Y}_m\|_{\mathcal{F}} \le \|\mathbf{R}_m\|_{\mathcal{F}} \le \|\mathbf{H}_m\mathbf{S}\|_{\mathcal{F}} + \|\mathbf{N}_m\|_{\mathcal{F}} \le \|\mathbf{H}_m\|_{\mathcal{F}}\|\mathbf{S}\|_{\mathcal{F}} + \|\mathbf{N}_m\|_{\mathcal{F}}$. In the massive MIMO regime, due to channel hardening, we have

$$\|\mathbf{H}_m\|_{\mathcal{F}}^2 \to \mathbb{E}\left\{\|\mathbf{H}_m\|_{\mathcal{F}}^2\right\} = N_a\sum\nolimits_{k=1}^{K}\beta_{k,m}. \tag{20}$$

Moreover, since each data symbol in $\mathbf{S}$ has unit power and each noise sample in $\mathbf{N}_m$ has variance σ^2, by the law of large numbers, we have

$$\|\mathbf{S}\|_{\mathcal{F}}^2 \to K\tau_c, \tag{21a}$$

$$\|\mathbf{N}_m\|_{\mathcal{F}}^2 \to N_a\tau_c\sigma^2. \tag{21b}$$

Hence, when $K\tau_c$ and $N_a\tau_c$ are sufficiently large, we can bound $\|\mathbf{Y}_m\|_{\mathcal{F}}, \forall m \in \mathcal{M}$ by

$$L = \max_{m\in\mathcal{M}}\sqrt{K\tau_c N_a\sum\nolimits_{k=1}^{K}\beta_{k,m}} + \sqrt{N_a\tau_c\sigma^2}. \tag{22}$$

Note that each AP m sends a perturbed matrix which contains the private signal $\mathbf{Y}_m$ to the CPU, for a total of T iterations. Hence, the information released to the CPU can be regarded as a T-fold composition. Then by Lemma 2 and 4, in order to achieve (ϵ,δ)-joint DP, the perturbation noise is calibrated as

$$\mu = 16L^2\sqrt{\frac{T}{M}\ln\left(\frac{2.5T}{\delta}\right)\ln\left(\frac{2}{\delta}\right)}/\epsilon. \tag{23}$$

Next the CPU computes the largest eigenvalue $\left(\widehat{\lambda}^{(n)}\right)^2$ and eigenvector $\widehat{\mathbf{v}}^{(n)}$ of $\widehat{\mathbb{W}}^{(n-1)} = \sum_{m=1}^{M}\widehat{\mathbb{J}}_m^{(n-1)}$. To control the error introduced by the perturbation noise, $\widehat{\lambda}^{(n)}$ is replaced by [25]

$$\widetilde{\lambda}^{(n)} = \widehat{\lambda}^{(n)} + \sqrt{\mu}\left(M\tau_c\right)^{1/4}. \tag{24}$$

Finally, the privacy-preserving implementation of (18) can be written as (25) shown at the bottom of the page, where the operator $\Xi_{L,\Omega_m}(\mathbf{M}) = \min\left\{\frac{L}{\|(\mathbf{M})_{\Omega_m}\|_{\mathcal{F}}}, 1\right\}\mathbf{M}$ ensures $\left\|\left(\mathbf{X}_m^{(n)}\right)_{\Omega_m}\right\|_{\mathcal{F}} \le L$.

The proposed FW-based privacy-preserving matrix completion algorithm is summarized in Algorithm 1.

Algorithm 1 FW-Based Privacy-Preserving Matrix Completion Algorithm

Input: Privacy parameters (ϵ,δ), number of users K, sampled matrix $\mathbf{Y}_m$ in each AP m, bound L on $\|\mathbf{Y}_m\|_{\mathcal{F}}$, number of APs M, number of time slots τ_c, number of iterations T

Output: $\widehat{\mathbf{X}}_m, \forall m \in \mathcal{M}$

Set $\mathbf{X}^{(0)} = \mathbf{0}_{MN_a\times\tau_c}$, $\eta^{(1)} = 1$, $\eta^{(n)} = \frac{1}{T}, n = 2,\cdots,T$ and μ given by (23)

for $n = 1 : T$ **do**

  $\mathcal{A}_m^{\mathrm{L},1}$: Each AP m computes $\mathbb{J}_m^{(n-1)}$ according to (19) and (17) and sends $\widehat{\mathbb{J}}_m^{(n-1)} = \mathbb{J}_m^{(n-1)} + \mathbf{G}_m^{(n-1)}$ to CPU

  $\mathcal{A}^{\mathrm{G}}$: CPU receives $\widehat{\mathbb{J}}_m^{(n-1)}$ from each AP m and computes $\widehat{\mathbb{W}}^{(n-1)} = \sum_{m=1}^{M}\widehat{\mathbb{J}}_m^{(n-1)}$, then

  - computes the largest eigenvalue $\left(\widehat{\lambda}^{(n)}\right)^2$ and eigenvector $\widehat{\mathbf{v}}^{(n)}$ of $\widehat{\mathbb{W}}^{(n-1)}$
  - computes $\widetilde{\lambda}^{(n)}$ according to (24)

  CPU sends $\left(\widehat{\mathbf{v}}^{(n)}, \widetilde{\lambda}^{(n)}\right)$ to all APs

  $\mathcal{A}_m^{\mathrm{L},2}$: Each AP m computes $\mathbf{X}_m^{(n)}$ according to (25)

end

$$\mathbf{X}_m^{(n)} = \Xi_{L,\Omega_m}\left(\left(1-\eta^{(n)}\right)\mathbf{X}_m^{(n-1)} - \frac{\eta^{(n)}K}{\widetilde{\lambda}^{(n)}}\mathbf{J}_m^{(n-1)}\widehat{\mathbf{v}}^{(n)}\left(\widehat{\mathbf{v}}^{(n)}\right)^{\mathrm{H}}\right), \quad \forall m \in \mathcal{M} \tag{25}$$

C. SVD-Based Privacy-Preserving Matrix Completion Algorithm

Here we consider another approach to low-rank matrix completion based on singular value decomposition (SVD) that yields a solution with bounded error [28]. It first trims $\mathbf{Y}$ to obtain $\widetilde{\mathbf{Y}}$ by setting to zero all rows with more than $2N_r\tau_{\rm c}/N_a$ non-zero entries, where recall that N_r is the number of RF chains at each AP, i.e., the number of non-zero elements in each column of $\mathbf{Y}$. Then it performs SVD on $\widetilde{\mathbf{Y}}$ and denotes $\mathbf{V}_K \in \mathbb{C}^{\tau_{\rm c}\times K}$ as the matrix consisting of the right singular vectors corresponding to the K largest singular values. Finally, the completed matrix is given by

$$\widehat{\mathbf{X}} = \frac{N_a}{N_r}\widetilde{\mathbf{Y}}\mathbf{V}_K\mathbf{V}_K^{\rm H}. \tag{26}$$

It is clear that $\mathrm{rank}\left(\widehat{\mathbf{X}}\right) = K$.

Note that $\mathbf{V}_K$ can also be obtained as the eigenvectors corresponding to the K largest eigenvalues of $\widetilde{\mathbf{Y}}^{\rm H}\widetilde{\mathbf{Y}} = \sum_{m=1}^{M}\widetilde{\mathbf{Y}}_m^{\rm H}\widetilde{\mathbf{Y}}_m$, where $\widetilde{\mathbf{Y}}_m^{\rm H}\widetilde{\mathbf{Y}}_m$ is computed at AP m and then sent to the CPU. However, because $\mathbb{J}_m = \widetilde{\mathbf{Y}}_m^{\rm H}\widetilde{\mathbf{Y}}_m$ contains the private data $\mathbf{Y}_m$, it should not be sent directly from AP m to the CPU. Similarly, we also add Gaussian noise to perturb it. Hence, the released data from each AP m to the CPU is $\widehat{\mathbb{J}}_m = \mathbb{J}_m + \mathbf{G}_m$, where $\mathbf{G}_m$ is a $\tau_{\rm c}\times\tau_{\rm c}$ Hermitian noise matrix, whose upper triangular and diagonal elements are respectively i.i.d. $\mathcal{N}_{\rm c}\left(0,\nu^2\right)$ and $\mathcal{N}\left(0,\nu^2\right)$ samples, and the lower triangular elements are complex conjugates of the upper triangular counterparts. Hence, the variance of total perturbation noise at the CPU is $M\nu^2$. By Lemma 4, in order to achieve (ϵ,δ)-joint DP, ν is calibrated as

$$\nu = L^2\sqrt{\frac{2}{M}\ln\left(\frac{1.25}{\delta}\right)}/\epsilon. \tag{27}$$

Then the CPU computes the K largest eigenvectors of $\widehat{\mathbf{V}}_K$ of the matrix $\widehat{\mathbb{W}} = \sum_{m=1}^{M}\widehat{\mathbb{J}}_m$ and broadcasts it to all APs. Hence, (26) can be implemented in a privacy-preserving form as

$$\widehat{\mathbf{X}}_m = \frac{N_a}{N_r}\widetilde{\mathbf{Y}}_m\widehat{\mathbf{V}}_K\widehat{\mathbf{V}}_K^{\rm H}, \quad \forall m\in\mathcal{M}. \tag{28}$$

The proposed SVD-based privacy-preserving matrix completion algorithm is summarized in Algorithm 2.

D. Computational Complexity and Communication Overhead

For both Algorithm 1 and Algorithm 2, the computations $\mathcal{A}_m^{\rm L,1}$ and $\mathcal{A}_m^{\rm L,2}$ at each AP m involve only additions and multiplications. The computation $\mathcal{A}^{\rm G}$ at the CPU in Algorithm 1 involves computing the largest eigen component of an $\tau_{\rm c}\times\tau_{\rm c}$ matrix in each iteration, for a total of T iterations; whereas in Algorithm 2, $\mathcal{A}^{\rm G}$ involves computing the K largest eigen components of an $\tau_{\rm c}\times\tau_{\rm c}$ matrix only once. The complexities of both Algorithm 1 and Algorithm 2 mainly depend on the computation of eigen components in $\mathcal{A}^{\rm G}$. If we adopt the traditional eigen-decomposition methods to obtain all eigen components, then the complexity of Algorithm 1 is almost T times than that of Algorithm 2. Furthermore, the complexity of Algorithm 1 can be improved by adopting other methods to compute the largest eigen component, such as classical power method and novel Oja's algorithm [29], which also achieve good performance.

Algorithm 2 SVD-Based Privacy-Preserving Matrix Completion Algorithm

Input: Privacy parameters (ϵ,δ), number of users K, sampled matrix $\mathbf{Y}_m$ in each AP m, bound L on $\|\mathbf{Y}_m\|_{\mathcal{F}}$, number of APs M, number of time slots $\tau_{\rm c}$

Output: $\widehat{\mathbf{X}}_m, \forall m\in\mathcal{M}$

1 Set ν given by (27)

2 $\mathcal{A}_m^{\rm L,1}$: AP m trims $\mathbf{Y}_m$ to obtain $\widetilde{\mathbf{Y}}_m$, then computes $\mathbb{J}_m = \widetilde{\mathbf{Y}}_m^{\rm H}\widetilde{\mathbf{Y}}_m$ and sends $\widehat{\mathbb{J}}_m = \mathbb{J}_m + \mathbf{G}_m$ to the CPU

3 $\mathcal{A}^{\rm G}$: CPU receives $\widehat{\mathbb{J}}_m$ from each AP m and computes the K largest eigenvectors $\widehat{\mathbf{V}}_K$ of $\widehat{\mathbb{W}} = \sum_{m=1}^{M}\widehat{\mathbb{J}}_m$. Then $\widehat{\mathbf{V}}_K$ is broadcast to all APs

4 $\mathcal{A}_m^{\rm L,2}$: AP m computes $\widehat{\mathbf{X}}_m$ according to (28)

We next compare the communication overheads of the two algorithms. For Algorithm 1, in each iteration, each AP sends the $\tau_{\rm c}\times\tau_{\rm c}$ matrix $\widehat{\mathbb{J}}_m^{(n)}$ to the CPU; and the CPU broadcasts the scalar $\widehat{\lambda}^{(n)}$ and the $\tau_{\rm c}\times 1$ vector $\widehat{\mathbf{v}}^{(n)}$ to all APs, for a total of T iterations. For Algorithm 2, each AP sends the $\tau_{\rm c}\times\tau_{\rm c}$ matrix $\widehat{\mathbb{J}}_m$ to the CPU only once; and the CPU broadcasts the $\tau_{\rm c}\times K$ matrix $\widehat{\mathbf{V}}_K$ to all APs only once. Hence the ratio of the broadcast overhead from the CPU to all APs between Algorithms 1 and 2 is $T:K$; and the ratio of the unicast overhead from each AP to the CPU is $T:1$.

IV. Privacy and Estimation Error Tradeoff Analysis

In this section, we first show that both Algorithms 1 and 2 are (ϵ,δ)-joint differentially private. We then provide their channel estimation error bounds in terms of the privacy parameters.

A. Privacy Analysis

Theorem 1: Algorithm 1 is (ϵ,δ)-joint differentially private.

Proof: First, we prove that in each iteration, the information released to the CPU is differentially private. Specifically, the received signal by the CPU at the n-th iteration is

$$\sum\nolimits_{m=1}^{M}\widehat{\mathbb{J}}_m^{(n-1)} = \sum\nolimits_{m=1}^{M}\mathbb{J}_m^{(n-1)} + \sum\nolimits_{m=1}^{M}\mathbf{G}_m^{(n-1)}, \tag{29}$$

where $\sum_{m=1}^{M}\mathbf{G}_m^{(n-1)}$ is the total perturbation noise from the APs, which is a Hermitian matrix. Its upper triangular and diagonal elements are respectively i.i.d. $\mathcal{N}_{\rm c}\left(0,M\mu^2\right)$ and $\mathcal{N}\left(0,M\mu^2\right)$ samples, and the lower triangular elements are complex conjugates of the upper triangular counterparts, where

$$M\mu^2 = \frac{\left(4L^2\right)^2 2\ln\left(1.25/\frac{\delta}{2T}\right)}{\left(\epsilon/\sqrt{8T\ln\left(2/\delta\right)}\right)^2}. \tag{30}$$

Recall that $\sum_{m=1}^{M}\mathbb{J}_m^{(n-1)}=\sum_{m=1}^{M}\left(\mathbf{J}_m^{(n-1)}\right)^{\mathrm{H}}\mathbf{J}_m^{(n-1)}$ with $\mathbf{J}_m^{(n-1)}=\left(\mathbf{X}_m^{(n-1)}\right)_{\Omega_m}-\mathbf{Y}_m$; and for each AP $m\in\mathcal{M}$, we have $\|\mathbf{Y}_m\|_{\mathcal{F}}\le L$ and $\left\|\left(\mathbf{X}_m^{(n-1)}\right)_{\Omega_m}\right\|_{\mathcal{F}}\le L$. Hence, the ℓ_2-sensitivity of the signal $\sum_{m=1}^{M}\mathbb{J}_m^{(n-1)}$ is $4L^2$. Then according to Lemma 4, the information released to the CPU at the n-th iteration, $\sum_{m=1}^{M}\widehat{\mathbb{J}}_m^{(n)}$, is $\left(\epsilon/\sqrt{8T\ln(2/\delta)},\delta/2T\right)$-differentially private. The CPU computes the dominant eigen components of the differentially private $\sum_{m=1}^{M}\widehat{\mathbb{J}}_m^{(n)}$ and by Lemma 1, the obtained $\widetilde{\lambda}^{(n)}$ and $\widehat{\mathbf{v}}^{(n)}$ are also $\left(\epsilon/\sqrt{8T\ln(2/\delta)},\delta/2T\right)$-differentially private. From Line 6 in Algorithm 1, each AP m computes $\mathbf{X}_m^{(n)}$ using differentially private $\widetilde{\lambda}^{(n)}$ and $\widehat{\mathbf{v}}^{(n)}$ and its local signal as input. By Lemma 3, $\mathbf{X}_m^{(n)}$ is then $\left(\epsilon/\sqrt{8T\ln(2/\delta)},\delta/2T\right)$-joint differentially private.

Finally, because $\widehat{\mathbf{X}}_m=\mathbf{X}_m^{(T)}$ is the result of a T-fold composition of $\left(\epsilon/\sqrt{8T\ln(2/\delta)},\delta/2T\right)$-joint differentially private algorithms, by Lemma 2, $\widehat{\mathbf{X}}_m,\forall m\in\mathcal{M}$ satisfy (ϵ,δ)-joint DP. □

Similar privacy analysis can be carried out for Algorithm 2 and we arrive at the following result.

Theorem 2: Algorithm 2 is (ϵ,δ)-joint differentially private.

Hence in both matrix completion algorithms, by adding Gaussian noise with calibrated variance to perturb the released information from the APs to the CPU, joint differential privacy can be achieved. However, the perturbation noise inevitably increases the matrix completion error and therefore the channel estimation error. Next, we will provide an analysis on the estimation error bounds for Algorithms 1 and 2. In particular, for both algorithms, we will show that the estimation error decreases with the increase of the data payload size τ_{d}.

B. Error Bounds and Scaling Laws for Channel Estimation

The error bounds for the FW and SVD-based noisy matrix completions are provided in [24] and [28] respectively, both of which do not consider privacy. [25] gives the error bounds for both FW and SVD-based differentiately private matrix completions, but the matrix is noise-free. Here we analyze the error bounds for differentiately private noisy matrix completions.

1) Error Bound and Scaling Law of Algorithm 1: The key step of Algorithm 1 in (25) can be rewritten in terms of the whole matrix as (31) shown at the bottom of this page. Following Lemma D.5 in [25], the projection operator $\Xi_{L,\Omega}$ does not introduce additional error. Hence, we will ignore it in the following analysis. Comparing (31) and (16), we can see that the privacy-preserving FW algorithm replaces $\mathbf{O}^{(n)}=-\frac{K}{\lambda^{(n)}}\mathbf{J}^{(n-1)}\mathbf{v}^{(n)}\left(\mathbf{v}^{(n)}\right)^{\mathrm{H}}$ in the original FW algorithm with $\mathbf{Q}^{(n)}=-\frac{K}{\widetilde{\lambda}^{(n)}}\mathbf{J}^{(n-1)}\widehat{\mathbf{v}}^{(n)}\left(\widehat{\mathbf{v}}^{(n)}\right)^{\mathrm{H}}$. To quantify the additional error caused by this replacement, we first provide the following two lemmas, which are generalizations of Lemma D.4 in [25] and Theorem 1 in [30] to the case of noisy matrix completion.

Lemma 5: The following bound holds with high probability[2]

$$\left\langle\mathbf{Q}^{(n)},\frac{1}{|\Omega|}\mathbf{J}^{(n-1)}\right\rangle_{\mathcal{F}}-\left\langle\mathbf{O}^{(n)},\frac{1}{|\Omega|}\mathbf{J}^{(n-1)}\right\rangle_{\mathcal{F}}\le O\left(\frac{K}{|\Omega|}\sqrt{\mu}\left(M\tau_{\mathrm{c}}\right)^{1/4}\right),\quad\forall n. \tag{32}$$

Proof: See Appendix A.

Lemma 6: If the update of $\mathbf{X}^{(n)}$ in (16) is modified as

$$\mathbf{X}^{(n)}=\left(1-\eta^{(n)}\right)\mathbf{X}^{(n-1)}+\eta^{(n)}\mathbf{W}^{(n)}, \tag{33}$$

where $\mathbf{W}^{(n)}$ satisfies $\left\|\mathbf{W}^{(n)}\right\|_{\mathrm{nuc}}\le K,\forall n$ and

$$\left\langle\mathbf{W}^{(n)},\frac{1}{|\Omega|}\mathbf{J}^{(n-1)}\right\rangle_{\mathcal{F}}-\left\langle\mathbf{O}^{(n)},\frac{1}{|\Omega|}\mathbf{J}^{(n-1)}\right\rangle_{\mathcal{F}}\le\gamma,\quad\forall n, \tag{34}$$

where $\langle\mathbf{A},\mathbf{B}\rangle_{\mathcal{F}}=\mathrm{tr}\left(\mathbf{A}^{\mathrm{H}}\mathbf{B}\right)$ is the Frobenius inner product. Then for $n=T$, we have

$$\frac{1}{|\Omega|}\left\|\left(\mathbf{X}^{(T)}\right)_{\Omega}-\mathbf{Y}\right\|_{\mathcal{F}}^2\le 4\gamma+\frac{K^2}{|\Omega|}+\frac{K^2}{|\Omega|\,T}+\sigma^2. \tag{35}$$

Proof: See Appendix B.

Using Lemmas 6 and 5, we can then obtain the error bound for Algorithm 1 as follows.

Theorem 3: Denote $\widehat{\mathbf{X}}=\mathbf{X}^{(T)}$ as the output of Algorithm 1. Then the following error bound on the observed entries in Ω holds with high probability

$$\frac{1}{|\Omega|}\left\|\left(\widehat{\mathbf{X}}\right)_{\Omega}-(\mathbf{X})_{\Omega}\right\|_{\mathcal{F}}^2=O\left(\frac{4K}{|\Omega|}\sqrt{\mu}\left(M\tau_{\mathrm{c}}\right)^{1/4}+\frac{K^2}{|\Omega|}+\frac{K^2}{|\Omega|T}+2\sigma^2\right). \tag{36}$$

Furthermore, the generalization error $\mathcal{E}\left(\widehat{\mathbf{X}}\right)=\frac{1}{MN_a\tau_{\mathrm{c}}}\left\|\widehat{\mathbf{X}}-\mathbf{X}\right\|_{\mathcal{F}}^2$ is bounded as (37) shown at the bottom of next page with high probability, where $\widetilde{O}(\cdot)$ hides poly-logarithmic terms in MN_a and τ_{c}. When the number of iterations is chosen as $T=O\left(\frac{K^{4/3}}{(|\Omega|(MN_a+\tau_{\mathrm{c}}))^{1/3}}\right)$, we can obtain the generalization error bound as (38) shown at the bottom of the next page.

[2]"with high probability" means with probability $1-1/{\tau_{\mathrm{c}}}^{\Theta(1)}$.

$$\mathbf{X}^{(n)}=\Xi_{L,\Omega}\left(\left(1-\eta^{(n)}\right)\mathbf{X}^{(n-1)}-\frac{\eta^{(n)}K}{\widetilde{\lambda}^{(n)}}\mathbf{J}^{(n-1)}\widehat{\mathbf{v}}^{(n)}\left(\widehat{\mathbf{v}}^{(n)}\right)^{\mathrm{H}}\right) \tag{31}$$

Proof: Since $\mathbf{Q}^{(n)} = -\frac{K}{\widetilde{\lambda}^{(n)}}\mathbf{J}^{(n-1)}\widehat{\mathbf{v}}^{(n)}\left(\widehat{\mathbf{v}}^{(n)}\right)^{\mathrm{H}}$ is rank-one, we have $\left\|\mathbf{Q}^{(n)}\right\|_{\mathrm{nuc}} = \left\|\mathbf{Q}^{(n)}\right\|_{\mathcal{F}}$, and

$$\begin{aligned}
&\left\|\mathbf{Q}^{(n)}\right\|_{\mathcal{F}}^2 \\
&= \mathrm{tr}\left(\left(\mathbf{Q}^{(n)}\right)^{\mathrm{H}}\mathbf{Q}^{(n)}\right) \\
&= \frac{K^2}{\left(\widetilde{\lambda}^{(n)}\right)^2}\mathrm{tr}\left(\widehat{\mathbf{v}}^{(n)}\left(\widehat{\mathbf{v}}^{(n)}\right)^{\mathrm{H}}\mathbf{J}^{(n-1)^{\mathrm{H}}}\mathbf{J}^{(n-1)}\widehat{\mathbf{v}}^{(n)}\left(\widehat{\mathbf{v}}^{(n)}\right)^{\mathrm{H}}\right) \\
&= K^2\frac{\left(\widehat{\mathbf{v}}^{(n)}\right)^{\mathrm{H}}\mathbf{J}^{(n-1)^{\mathrm{H}}}\mathbf{J}^{(n-1)}\widehat{\mathbf{v}}^{(n)}}{\left(\widetilde{\lambda}^{(n)}\right)^2}\left(\widehat{\mathbf{v}}^{(n)}\right)^{\mathrm{H}}\widehat{\mathbf{v}}^{(n)} \\
&= K^2\frac{\left\|\mathbf{J}^{(n-1)}\widehat{\mathbf{v}}^{(n)}\right\|_{\mathcal{F}}^2}{\left(\widetilde{\lambda}^{(n)}\right)^2}. \qquad (39)
\end{aligned}$$

According to Lemma D.2 in [25], the following holds with high probability

$$\left\|\mathbf{J}^{(n-1)}\widehat{\mathbf{v}}^{(n)}\right\|_{\mathcal{F}} \le \widehat{\lambda} + O\left(\sqrt{\mu}\left(M\tau_{\mathrm{c}}\right)^{1/4}\right). \qquad (40)$$

Hence, according to the definition of $\widetilde{\lambda}^{(n)}$ in (24), with high probability, we have $\left\|\mathbf{Q}^{(n)}\right\|_{\mathrm{nuc}} \le K, \forall n$. Then by Lemma 5 and Lemma 6, the following holds with high probability

$$\begin{aligned}
&\frac{1}{|\Omega|}\left\|\left(\widehat{\mathbf{X}}\right)_{\Omega} - \mathbf{Y}\right\|_{\mathcal{F}}^2 \\
&\quad = O\left(\frac{4K}{|\Omega|}\sqrt{\mu}\left(M\tau_{\mathrm{c}}\right)^{1/4} + \frac{K^2}{|\Omega|} + \frac{K^2}{|\Omega|T} + \sigma^2\right). \qquad (41)
\end{aligned}$$

Due to the triangular inequality property, we have

$$\begin{aligned}
&\left\|\left(\widehat{\mathbf{X}}\right)_{\Omega} - (\mathbf{X})_{\Omega}\right\|_{\mathcal{F}}^2 \\
&\le \left(\left\|\left(\widehat{\mathbf{X}}\right)_{\Omega} - \mathbf{Y}\right\|_{\mathcal{F}} + \left\|\mathbf{Y} - (\mathbf{X})_{\Omega}\right\|_{\mathcal{F}}\right)^2 \\
&\le 2\left\|\left(\widehat{\mathbf{X}}\right)_{\Omega} - \mathbf{Y}\right\|_{\mathcal{F}}^2 + 2\left\|\mathbf{Y} - (\mathbf{X})_{\Omega}\right\|_{\mathcal{F}}^2. \qquad (42)
\end{aligned}$$

Note that $\left\|\mathbf{Y} - (\mathbf{X})_{\Omega}\right\|_{\mathcal{F}}^2 = \left\|(\mathbf{N})_{\Omega}\right\|_{\mathcal{F}}^2 \to |\Omega|\,\sigma^2$ when $|\Omega| = MN_r\tau_{\mathrm{c}}$ is large. Hence, with high probability (36) holds.

Note that (36) gives the error bound on observed entries in Ω. Using Theorem 1 in [31], we can then generalize the error bound to the entire matrix $\widehat{\mathbf{X}}$ given by (37). It can be seen that the third term in (37) decreases with T, while the last term increases with T. By setting these two terms as the same order, we obtain $T = O\left(\frac{K^{4/3}}{(|\Omega|(MN_a+\tau_{\mathrm{c}}))^{1/3}}\right)$, and the corresponding generalization error bound in (38). □

Remark 1: To achieve stronger privacy level, i.e., smaller ϵ and/or δ, the perturbation noise variance μ in (23) will increase, which in turn increases the matrix completion error according to (37).

Note that the matrix completion error in (38) has two sources: the first term is due to the perturbation noise added to achieve DP, and the other terms are the error inherent to the FW algorithm. Both error sources contribute to the channel estimation error through (12). The following result shows that the channel estimation errors caused by both sources decreases with the increase of the payload size τ_d, at different speed, for fixed pilot size τ_{p}.

Corollary 1: For fixed privacy parameters ϵ and δ, and fixed pilot length τ_{p}, the estimation error of the proposed privacy-preserving channel estimator that employs Algorithm 1 scales with the data payload size τ_{d} as

$$\left\|\widehat{\mathbf{H}}_m - \mathbf{H}_m\right\|_{\mathcal{F}}^2 = O\left(\tau_{\mathrm{d}}^{-1/3}\right). \qquad (43)$$

Moreover, the portion of the estimation error due to the perturbation noise added to preserve privacy scales as $O\left(\tau_{\mathrm{d}}^{-5/12}\right)$.

Proof: Note that from (12), for a given pilot matrix $\mathbf{P}$, the channel estimation error at AP m satisfies

$$\begin{aligned}
\left\|\widehat{\mathbf{H}}_m - \mathbf{H}_m\right\|_{\mathcal{F}}^2 &= \left\|\left(\widehat{\mathbf{X}}_{m,\mathrm{p}} - \mathbf{X}_{m,\mathrm{p}}\right)\mathbf{P}^{\dagger}\right\|_{\mathcal{F}}^2 \\
&\le \left\|\widehat{\mathbf{X}}_{m,\mathrm{p}} - \mathbf{X}_{m,\mathrm{p}}\right\|_{\mathcal{F}}^2\left\|\mathbf{P}^{\dagger}\right\|_{\mathcal{F}}^2. \qquad (44)
\end{aligned}$$

Now assuming that the matrix completion error is uniform among the entries of $\mathbf{X}$, then we have

$$\left\|\widehat{\mathbf{X}}_{m,\mathrm{p}} - \mathbf{X}_{m,\mathrm{p}}\right\|_{\mathcal{F}}^2 = \frac{1}{M}\frac{\tau_{\mathrm{p}}}{\tau_{\mathrm{p}}+\tau_{\mathrm{d}}}\left\|\widehat{\mathbf{X}} - \mathbf{X}\right\|_{\mathcal{F}}^2. \qquad (45)$$

For fixed privacy parameters ϵ and δ, and fixed pilot length τ_{p}, when $T = O\left(\frac{K^{4/3}}{(|\Omega|(MN_a+\tau_{\mathrm{c}}))^{1/3}}\right)$, we have $L = O\left(\sqrt{\tau_{\mathrm{d}}}\right)$ and $\mu = O\left(\tau_{\mathrm{d}}^{2/3}\right)$ hiding all other parameters and the logarithmic term $\sqrt{\ln\left(\tau_{\mathrm{d}}^{-2/3}\right)}$ according to (22) and (23), respectively. The term in (38) that has μ is due to the perturbation noise, and scales as $O\left(\tau_{\mathrm{d}}^{-5/12}\right)$ since $|\Omega| = MN_r\tau_{\mathrm{c}}$. In addition, the last term in (38) scales as $O\left(\tau_{\mathrm{d}}^{-1/3}\right)$, which dominates the matrix completion error. Then by (44) and (45), the statements of the corollary hold. □

2) Error Bound and Scaling Law of Algorithm 2: The key step of Algorithm 2 in (28) can be rewritten in terms of the whole matrix as

$$\widehat{\mathbf{X}} = \frac{N_a}{N_r}\widetilde{\mathbf{Y}}\widehat{\mathbf{V}}_K\widehat{\mathbf{V}}_K^{\mathrm{H}}. \qquad (46)$$

Compared to the original SVD-based matrix completion in (26), (46) uses $\widehat{\mathbf{\Pi}}_K = \widehat{\mathbf{V}}_K\widehat{\mathbf{V}}_K^{\mathrm{H}}$ instead of $\mathbf{\Pi}_K = \mathbf{V}_K\mathbf{V}_K^{\mathrm{H}}$.

$$\mathcal{E}\left(\widehat{\mathbf{X}}\right) = \widetilde{O}\left(\frac{4K}{|\Omega|}\sqrt{\mu}\left(M\tau_{\mathrm{c}}\right)^{1/4} + \frac{K^2}{|\Omega|} + \frac{K^2}{|\Omega|T} + 2\sigma^2 + \sqrt{\frac{T\left(MN_a+\tau_{\mathrm{c}}\right)}{|\Omega|}}\right) \qquad (37)$$

$$\mathcal{E}\left(\widehat{\mathbf{X}}\right) = \widetilde{O}\left(\frac{4K}{|\Omega|}\sqrt{\mu}\left(M\tau_{\mathrm{c}}\right)^{1/4} + \frac{K^2}{|\Omega|} + 2\sigma^2 + \frac{2K^{2/3}\left(MN_a+\tau_{\mathrm{c}}\right)^{1/3}}{|\Omega|^{2/3}}\right) \qquad (38)$$

Denote the K-th and $(K+1)$-th singular values of $\widetilde{\mathbf{Y}}$ as λ_K and λ_{K+1}, respectively. When there is a large gap between λ_K^2 and λ_{K+1}^2, the space spanned by the K largest eigenvectors of the noise-perturbed version of $\widetilde{\mathbf{Y}}^{\mathrm{H}}\widetilde{\mathbf{Y}}$, i.e., $\widehat{\mathbb{W}}$ is very close to the space spanned by the K largest right singular vectors of $\widetilde{\mathbf{Y}}$ [32]. In massive MIMO with sufficiently large MN_a, such a large gap holds and then we have the following matrix completion error bound for Algorithm 2.

Theorem 4: If $\lambda_K^2 - \lambda_{K+1}^2 = \omega\left(\nu\sqrt{M\tau_{\mathrm{c}}}\right)$, then the following error bound on the output $\widehat{\mathbf{X}}$ of Algorithm 2 shown in (47), shown at the bottom of the page, holds with high probability.

Proof: Denoting $\mathbf{\Psi} = \frac{N_a}{N_r}\widetilde{\mathbf{Y}}$, we can write

$$\begin{aligned}\left\|\widehat{\mathbf{X}} - \mathbf{X}\right\|_{\mathcal{F}} &= \left\|\mathbf{\Psi}\widehat{\mathbf{\Pi}}_K - \mathbf{X}\right\|_{\mathcal{F}} \\ &\le \left\|\mathbf{\Psi}\mathbf{\Pi}_K - \mathbf{X}\right\|_{\mathcal{F}} + \left\|\mathbf{\Psi}\widehat{\mathbf{\Pi}}_K - \mathbf{\Psi}\mathbf{\Pi}_K\right\|_{\mathcal{F}} \\ &\le \left\|\mathbf{\Psi}\mathbf{\Pi}_K - \mathbf{X}\right\|_{\mathcal{F}} + \left\|\mathbf{\Psi}\right\|_{\mathcal{F}}\left\|\widehat{\mathbf{\Pi}}_K - \mathbf{\Pi}_K\right\|_{\mathcal{F}}.\end{aligned} \tag{48}$$

First, according to Theorem 7 of [32], if $\lambda_K^2 - \lambda_{K+1}^2 = \omega\left(\nu\sqrt{M\tau_{\mathrm{c}}}\right)$, then with high probability

$$\left\|\widehat{\mathbf{\Pi}}_K - \mathbf{\Pi}_K\right\|_2 = O\left(\frac{\nu\sqrt{M\tau_{\mathrm{c}}}}{\omega\left(\nu\sqrt{M\tau_{\mathrm{c}}}\right)}\right), \tag{49}$$

and hence

$$\left\|\widehat{\mathbf{\Pi}}_K - \mathbf{\Pi}_K\right\|_{\mathcal{F}} = O\left(\frac{\nu\sqrt{KM\tau_{\mathrm{c}}}}{\omega\left(\nu\sqrt{M\tau_{\mathrm{c}}}\right)}\right). \tag{50}$$

Furthermore, we have

$$\left\|\mathbf{\Psi}\right\|_{\mathcal{F}} = \frac{N_a}{N_r}\left\|\widetilde{\mathbf{Y}}\right\|_{\mathcal{F}} \le \frac{N_a}{N_r}\left\|\mathbf{Y}\right\|_{\mathcal{F}} \le \frac{N_a}{N_r}\sqrt{ML}. \tag{51}$$

Next, using Theorem 1.1 in [28], with high probability, there exist constants c_0 and c_1 such that

$$\begin{aligned}&\frac{1}{\sqrt{MN_a\tau_{\mathrm{c}}}}\left\|\mathbf{\Psi}\mathbf{\Pi}_K - \mathbf{X}\right\|_{\mathcal{F}} \\ &\le c_0 x_{\max}\left(\frac{K^2MN_a^3}{{N_r}^2\tau_{\mathrm{c}}^3}\right)^{1/4} + c_1\frac{\sqrt{KN_a}}{N_r\sqrt{M\tau_{\mathrm{c}}}}\left\|\left(\widetilde{\mathbf{N}}\right)_{\Omega}\right\|_2,\end{aligned} \tag{52}$$

where $x_{\max} = \max_{(i,j)}\left|\mathbf{X}\left(i,j\right)\right|$ and $\left(\widetilde{\mathbf{N}}\right)_{\Omega}$ denotes the matrix obtained from $\left(\mathbf{N}\right)_{\Omega}$ after the trimming step. By using the Theorem 1.3 in [28], with high probability, there exists a constant c_2 such that

$$\left\|\left(\widetilde{\mathbf{N}}\right)_{\Omega}\right\|_2 \le c_2\sigma\left(MN_r\log\tau_{\mathrm{c}}\right)^{1/2}. \tag{53}$$

Plugging (50)-(53) and $x_{\max} \le L$ into (48), we obtain (47). □

Remark 2: Since $L = O\left(\sqrt{\tau_{\mathrm{d}}}\right)$ and $\nu = O\left(\tau_{\mathrm{d}}\right)$, the sum of the first two terms in (47) scales with τ_{d} as $O(\tau_{\mathrm{d}}^{-1/4})$, and the third term in (47) represents the completion error caused by the perturbation noise, that scales as $O\left(\frac{\tau_{\mathrm{d}}^{3/2}}{\omega\left(\tau_{\mathrm{d}}^{3/2}\right)}\right)$, which dominates the matrix completion error. Hence using (44) and (45), we arrive at the following corollary regarding the scaling of the channel estimation error by Algorithm 2 at AP m with respect to the data payload size τ_{d}.

Corollary 2: For fixed privacy parameters ϵ and δ, and fixed pilot length τ_{p}, the estimation error of the proposed privacy-preserving channel estimator that employs Algorithm 2 scales with the data payload size τ_{d} as

$$\left\|\widehat{\mathbf{H}}_m - \mathbf{H}_m\right\|_{\mathcal{F}} = O\left(\frac{\tau_{\mathrm{d}}^{3/2}}{\omega\left(\tau_{\mathrm{d}}^{3/2}\right)}\right). \tag{54}$$

In summary, we see that for both Algorithms 1 and 2, the channel estimator error consists of a privacy independent component, that is due to the channel noise and matrix completion error, and a privacy-induced component, that is due to the perturbation noise. A higher privacy level leads to a higher privacy-induced channel estimation error, and vice versa. As the payload size increases, both error components decrease. However, for Algorithm 1, the channel estimation error is dominated by the privacy-independent component; whereas for Algorithm 2, it is dominated by the privacy-induced component.

V. Simulation Results

A. Simulation Setup

We consider a cell-free massive MIMO system covering a hexagonal region with radius $R = 1\mathrm{km}$, where APs and users are randomly and uniformly distributed. The channel model in (1) is adopted to generate channel matrices $\mathbf{H}_m$ with the large-scale fading $\beta_{k,m}$ modeled as

$$\beta_{k,m} = 10^{-\frac{\mathrm{PL}\left(d_{k,m}\right) + \sigma_{\mathrm{sh}}z_{k,m}}{10}}, \tag{55}$$

where $\mathrm{PL}\left(d_{k,m}\right)\left(\mathrm{dB}\right)$ is the path loss between AP m and user k with distance $d_{k,m}$; $\sigma_{\mathrm{sh}}\left(\mathrm{dB}\right)$ is the standard deviation of shadow fading and $z_{k,m} \sim \mathcal{N}_{\mathrm{c}}\left(0,1\right)$. $\tau_{\mathrm{p}} = K$ orthonormal pilot sequences are used, resulting in an orthonormal pilot matrix $\mathbf{P}$. Data symbols are independently drawn from the QPSK constellation with unit average power. We consider two settings of the number of users: $K = 5$ and $K = 25$. All simulation parameters are shown in Table II. All methods are implemented by MATLAB R2020a on a MacBook Pro with 2.6 GHz 6-Core Intel Core i7 and 16GB RAM. The channel estimation performance is evaluated by the normalized mean squared error (NMSE) defined as $\mathrm{NMSE} = \mathbb{E}\left\{\left\|\widehat{\mathbf{H}} - \mathbf{H}\right\|_{\mathcal{F}}^2 / \left\|\mathbf{H}\right\|_{\mathcal{F}}^2\right\}$. The data detection performance is evaluated by symbol error rate (SER). Both NMSE and SER are obtained through Monte-Carlo simulations with fixed

$$\frac{1}{\sqrt{MN_a\tau_{\mathrm{c}}}}\left\|\widehat{\mathbf{X}} - \mathbf{X}\right\|_{\mathcal{F}} = O\left(\left(\frac{L^4K^2MN_a^3}{N_r^2\tau_{\mathrm{c}}^3}\right)^{1/4} + \frac{\sqrt{KN_aN_r\ln\tau_{\mathrm{c}}\sigma^2}}{N_r\sqrt{\tau_{\mathrm{c}}}} + \frac{\sqrt{N_a}L\nu\sqrt{KM\tau_{\mathrm{c}}}}{N_r\sqrt{\tau_{\mathrm{c}}}\omega\left(\nu\sqrt{M\tau_{\mathrm{c}}}\right)}\right) \tag{47}$$

TABLE II
BASIC SIMULATION PARAMETERS

Parameter	Meaning	Value
M	The number of APs	100
K	The number of users	5, 25
N_a	The number of antennas at each AP	4
N_r	The number of RF chains at each AP	2, 3
$\sigma_{\rm sh}$	The standard deviation of shadow fading	8 dB
$\mathrm{PL}\,(d)$	The path loss with distance $d\,(\mathrm{m})$	$36.8 + 36.7 \log_{10}(d)$ dB
σ^2	The variance of received noise sample	10^{-13} Watts

large-scale fadings $\{\beta_{k,m}\}$ and a minimum of 500 independent fast channel realizations $\{\mathbf{g}_{k,m}\}$.

In Algorithm 1, we approximate the rank constraint $\mathrm{rank}\,(\mathbf{X}) \leq K$ with the nuclear norm constraint $\|\mathbf{X}\|_{\rm nuc} \leq K$. However, the true $\|\mathbf{X}\|_{\rm nuc}$ is far smaller than K due to the large-scale fading. Hence, in our implementation, we replace the rank bound K in (25) with an appropriate nuclear norm bound r. According to (20) and (21a), for massive MIMO, we have

$$\begin{aligned}\|\mathbf{X}\|_{\mathcal{F}} &= \|\mathbf{HS}\|_{\mathcal{F}} \leq \|\mathbf{H}\|_{\mathcal{F}} \|\mathbf{S}\|_{\mathcal{F}} \\ &= \sqrt{K\tau_{\rm c} N_a \sum\nolimits_{m=1}^{M} \sum\nolimits_{k=1}^{K} \beta_{k,m}}. \end{aligned} \tag{56}$$

Since $\|\mathbf{X}\|_{\rm nuc} \leq \sqrt{\mathrm{rank}\,(\mathbf{X})}\,\|\mathbf{X}\|_{\mathcal{F}}$ [33], we can bound $\|\mathbf{X}\|_{\rm nuc}$ by

$$r = \sqrt{K^2\tau_{\rm c} N_a \sum\nolimits_{m=1}^{M} \sum\nolimits_{k=1}^{K} \beta_{k,m}}. \tag{57}$$

For $K = 5$, we choose $r \in [0.001, 0.01]$ by cross-validation, e.g., we uniformly choose 10 values of r and run Algorithm 1 using them. The r value that has the lowest NMSE is then chosen for the simulations. For $K = 25$, we choose $r \in [0.01, 0.1]$ by cross-validation. Moreover, for the number of iterations T, we choose $T \in \{K, 2K, \cdots, 5K\}$ by cross-validation.

For comparison, we consider the following three channel estimators:

(1) Non-private FW (NPFW): To show the performance upper bound of Algorithm 1 when privacy is not considered, we set $\mu = 0$ and $T = 200$.

(2) Non-private SVD (NPSVD): To show the performance upper bound of Algorithm 2 when privacy is not considered, we set $\nu = 0$.

(3) Pilot-only (PO): We also consider the pilot-only method, where each AP m estimates its channel matrix $\mathbf{H}_m$ locally based on its received pilots $\{\mathbf{Y}_m(:,t), t \in \mathcal{T}_p\}$ only. Specifically, each AP m first computes its least squares (LS) estimate as $\widehat{\mathbf{H}}_m = \mathbf{Y}_m(:, 1:\tau_{\rm p})\mathbf{P}^{\rm H}$. Then the local linear minimum mean-squared error (LMMSE) estimate of data symbols $\mathbf{D}_m$ is computed as follows. Denoting $\mathbf{y}_m[t]$ as the N_r-dimensional vector consisting of the non-zero elements of $\mathbf{Y}_m(:,t)$. Then from (5) we can write $\mathbf{y}_m[t] = \mathbf{C}_m[t]\,\mathbf{r}_m[t]$, where $\mathbf{C}_m[t] \in \mathbb{C}^{N_r \times N_a}$ and $\mathbf{C}_m[t](i,j) = 1$ if the j-th antenna is connected to the i-th RF chain, and it is 0 otherwise. Then LMMSE estimate is given by $\widehat{\mathbf{D}}_m[t] = \left(\mathbf{F}_m^{\rm H}\mathbf{F}_m + \sigma^2\mathbf{I}_K\right)^{-1}\mathbf{F}_m^{\rm H}\mathbf{y}_m[t], \tau_{\rm p}+1 \leq t \leq \tau_{\rm c}$ with $\mathbf{F}_m = \mathbf{C}_m[t]\,\widehat{\mathbf{H}}_m$. At last, $\left\{\widehat{\mathbf{D}}_m\right\}$ are sent to the CPU which performs data detection based on the combined statistic according to (14). Since the PO method does not send any private signal to the CPU, it is perfectly privacy-preserving.

B. Results

Fig. 3(a) and Fig. 3(b) respectively show the NMSE performance of channel estimation and the corresponding SER performance of data detection versus the privacy parameter ϵ with $K = 5$, $\tau_{\rm c} = 100$, and $\delta = 0.1$ for different methods. It can be seen that the performance of both Algorithm 1 and 2 improve as ϵ increases, which means the privacy level degrades. Algorithm 1 significantly outperforms Algorithm 2 under both private and non-private cases. Moreover, despite of the added perturbation noise to achieve privacy, both algorithms outperform the PO method, by exploiting the received data payload signal.

Fig. 4(a) and Fig. 4(b) respectively show the NMSE performance of channel estimation and the corresponding SER performance of data detection versus the payload size $\tau_{\rm d}$ with $K = 5$, fixed privacy parameter $\epsilon = 1$ and $\delta = 0.1$ for different methods. Since the PO method makes use of the received pilot signal only for channel estimation, its performance remains the same as $\tau_{\rm d}$ increases. On the contrary, the performances of Algorithm 1 and 2 improve as $\tau_{\rm d}$ increases. Hence, the proposed methods can utilize the received data payload signal to improve the accuracy of channel estimation and data detection, while maintaining the same privacy level at the same time. Algorithm 1 significantly outperforms Algorithm 2 under both private and non-private cases for all payload size.

We also show the performances of five methods with a larger number of users. For $K = 25$, Fig.5 and Fig. 6 respectively show the performances of five methods versus privacy parameter ϵ and payload size $\tau_{\rm d}$. It can be seen that both the channel estimation and data detection performances of both Algorithm 1 and Algorithm 2 still improve as ϵ or $\tau_{\rm d}$ increases. Algorithm 1 still significantly outperforms Algorithm 2 and the PO method for all considered ϵ and $\tau_{\rm d}$. However, Algorithm 2 is less effective when the number of users is high. In addition, by comparing Fig. 3 and Fig. 5, as well as Fig. 4 and Fig. 6, it can be seen that both the channel estimation and data detection performance get worse with a larger number of users.

The SERs in the above figures are in the range of 0.05-0.1, which mainly because the $N_r = \frac{1}{2}N_a$, i.e., missing data is 50%. Note that for many applications, such SER is satisfactory since by employing powerful error correction codes, such

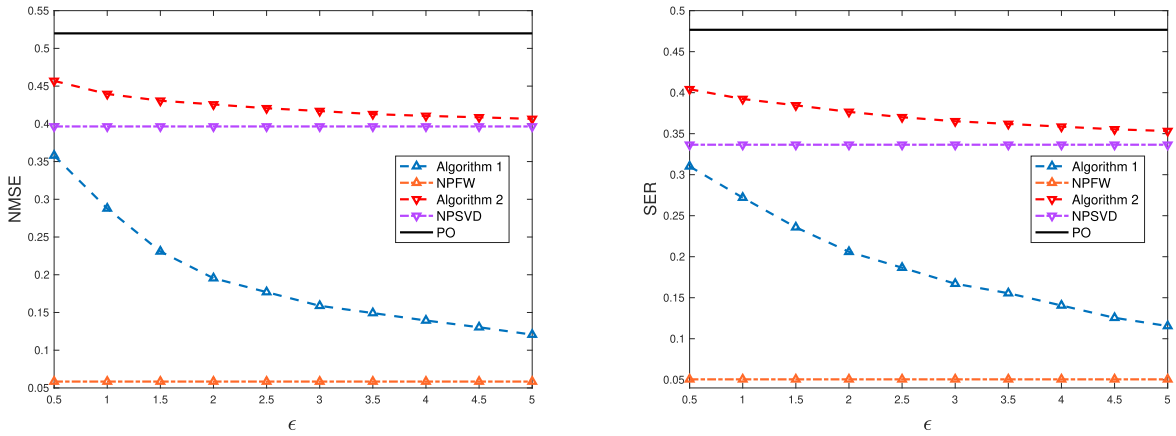


Fig. 3. The performance versus ϵ with $K = 5$, $\tau_{\mathrm{c}} = 100$ and $\delta = 0.1$. (a) NMSE of channel estimation. (b) SER of data detection.

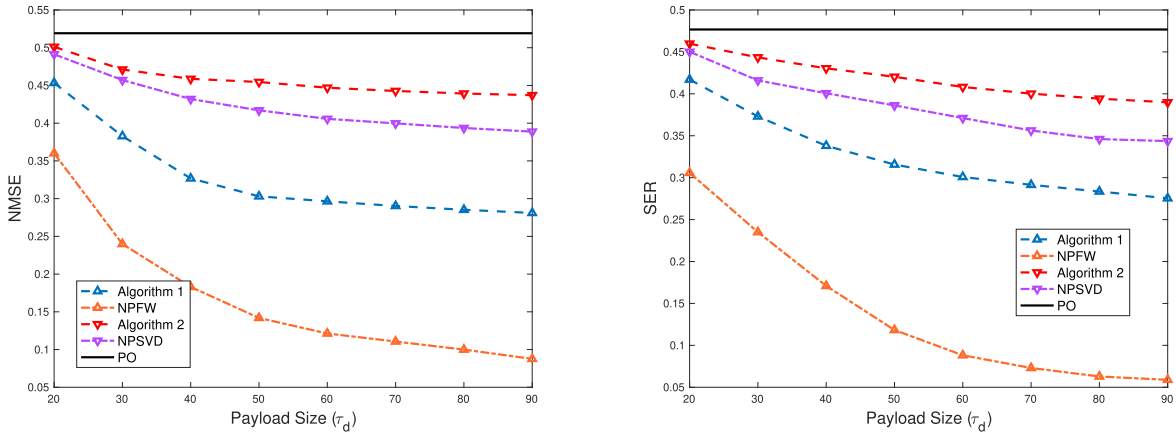


Fig. 4. The performance versus payload size τ_{d} with $K = 5$, $\epsilon = 1$ and $\delta = 0.1$. (a) NMSE of channel estimation. (b) SER of data detection.

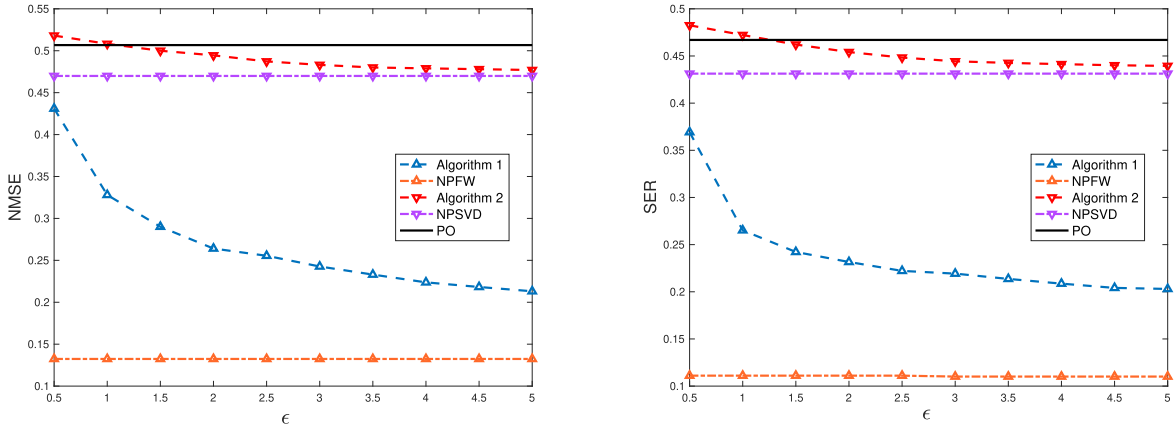


Fig. 5. The performance versus ϵ with $K = 25$, $\tau_{\mathrm{c}} = 100$ and $\delta = 0.1$. (a) NMSE of channel estimation. (b) SER of data detection.

as low-density parity-check (LDPC) code, or concatenated codes, a information bit error rate (BER) of, i.e., 10^{-4} can be achieved [34], [35]. Considering a reduced number of missing entries with $N_r = 3$, Fig.7(a) and Fig. 7(b) show the SER performance of data detection versus the privacy parameter ϵ with $K = 5$, $\tau_{\mathrm{c}} = 100$, $\delta = 0.1$; and the payload size τ_{d} with $K = 5$, $\epsilon = 1$, $\delta = 0.1$, respectively. It can be seen that the SER performances of both Algorithm 1 and Algorithm 2 still improve as ϵ or τ_{d} increases. In addition, by comparing Fig. 7(a) and Fig. 3(b), as well as Fig. 7(b) and Fig. 4(b), it can be seen that the data detection performances are greatly improved with a reduced number of missing entries.

Finally, we compare the time complexity of Algorithm 1 and 2 for different number of users in Table III. For both Algorithm 1 and 2, we adopt traditional eigen-decomposition methods to obtain all eigen components.

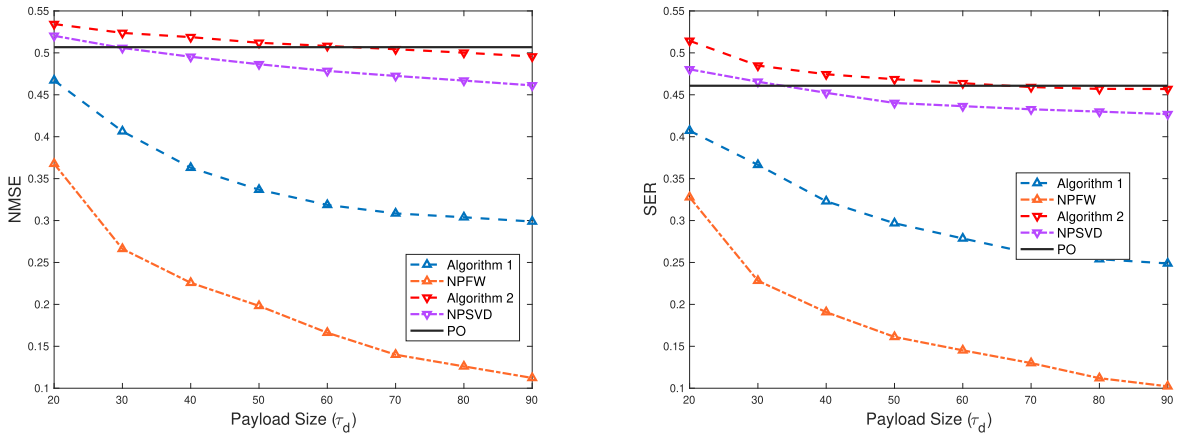


Fig. 6. The performance versus payload size τ_{d} with $K = 25$, $\epsilon = 1$ and $\delta = 0.1$. (a) NMSE of channel estimation. (b) SER of data detection.

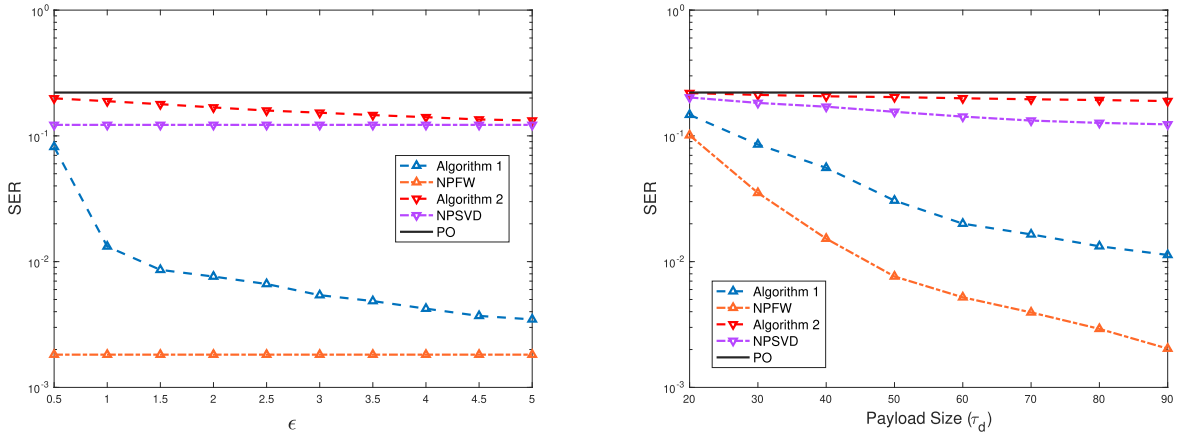


Fig. 7. The SER of data detection versus (a) ϵ with $K = 5$, $\tau_{\mathrm{c}} = 100$, $\delta = 0.1$ and $N_r = 3$. (b) payload size τ_{d} with $K = 5$, $\epsilon = 1$, $\delta = 0.1$ and $N_r = 3$.

TABLE III
THE CONSUMED TIME OF ALGORITHM 1 AND 2

The number of users K	5	10	15	20	25
The consumed time of Algorithm 1	0.3125s	0.6951s	0.9031s	1.4264s	1.9546s
The consumed time of Algorithm 2	0.0112s	0.0109s	0.0.0109s	0.0110s	0.0111s

It can be seen that Algorithm 1 consumes more time with the larger number of users, which is because a larger number of iterations is required to converge. In contrast, the consumed time of Algorithm 2 remains constant with the increase of the number of users.

VI. CONCLUSION

This paper considers a cell-free hybrid massive MIMO system, where the number of users is typically much smaller than the total number of antennas. Efficient uplink channel estimation and data detection with reduced number of pilots can be performed based on low-rank matrix completion. However, such a scheme requires the CPU to collect received signals from all APs, which may enable the CPU to infer the private information of user locations. To solve this problem, we develop and analyze privacy-preserving channel estimation schemes under the framework of differential privacy. The key ingredient of such a channel estimator is a joint differentially private noisy matrix completion algorithm, which consists of a global component implemented at the CPU and local components implemented at APs. Two joint differentially private channel estimators based respectively on FW and SVD are proposed and analyzed. In particular, we have shown that for both algorithms the estimation error can be mitigated while maintaining the same privacy level by increasing the payload size with fixed pilot size; and the scaling laws of both the privacy-induced and privacy-independent error components in terms of payload size are characterized. Simulation results corroborate the theoretical analysis and clearly demonstrate the tradeoff between privacy and channel estimation performance. Finally, we note that frequent antenna switching is required for uniformly sampling the received signal, so that the channel estimation accuracy can be guaranteed. It is important to investigate the effect of antenna switching on other performance metrics in our future work, such as bandwidth efficiency

and energy efficiency. Another future direction is to apply alternative matrix completion algorithms that do not require uniform sampling to allow for less frequent antenna switching, so that bandwidth-efficient time-orthogonal shaping filters can be employed.

APPENDIX A
PROOF OF LEMMA 6

Define a function $\Gamma_\Omega(\mathbf{X}) = \frac{1}{2|\Omega|}\|(\mathbf{X})_\Omega - \mathbf{Y}\|_{\mathcal{F}}^2$ on the feasible set $\mathcal{D}: \|\mathbf{X}\|_{\text{nuc}} \le K$. The curvature parameter C_Γ of the above function can be defined as

$$C_\Gamma = \max_{\mathbf{X}_a,\mathbf{S}\in\mathcal{D}} \frac{2}{\kappa^2}(\Gamma_\Omega(\mathbf{X}_b) - \Gamma_\Omega(\mathbf{X}_a) - \langle \mathbf{X}_b - \mathbf{X}_a, \nabla\Gamma_\Omega(\mathbf{X}_a)\rangle_{\mathcal{F}}), \quad (58)$$

where $\kappa \in [0,1]$; $\mathbf{X}_b = \mathbf{X}_a + \kappa(\mathbf{S} - \mathbf{X}_a)$; $\nabla\Gamma_\Omega(\mathbf{X}_a) = \frac{1}{|\Omega|}((\mathbf{X}_a)_\Omega - \mathbf{Y})$ is the gradient of $\Gamma_\Omega(\mathbf{X})$ at $\mathbf{X}_a$. It then follows from the definition in (58) that for any $\mathbf{X}_a$ and $\mathbf{S}$

$$\Gamma_\Omega(\mathbf{X}_b) \le \Gamma_\Omega(\mathbf{X}_a) + \langle \mathbf{X}_b - \mathbf{X}_a, \nabla\Gamma_\Omega(\mathbf{X}_a)\rangle_{\mathcal{F}} + \frac{C_\Gamma\kappa^2}{2}. \quad (59)$$

According to (33), we have

$$\mathbf{X}^{(n)} = \mathbf{X}^{(n-1)} + \eta^{(n)}\left(\mathbf{W}^{(n)} - \mathbf{X}^{(n-1)}\right). \quad (60)$$

Recall that $\left\|\mathbf{W}^{(n)}\right\|_{\text{nuc}} \le K, \forall n$, $\eta^{(1)} = 1$ and $\eta^{(n)} = \frac{1}{T}, n = 2,\cdots,T$, thus we have $\left\|\mathbf{X}^{(n)}\right\|_{\text{nuc}} \le K, \forall n$. By letting $\mathbf{X}_b = \mathbf{X}^{(n)}$, $\mathbf{X}_a = \mathbf{X}^{(n-1)}$, $\mathbf{S} = \mathbf{W}^{(n)}$ and $\kappa = \eta^{(n)}$ in (59), we have

$$\Gamma_\Omega\left(\mathbf{X}^{(n)}\right) \le \Gamma_\Omega\left(\mathbf{X}^{(n-1)}\right) + \eta^{(n)}\left\langle \mathbf{W}^{(n)} - \mathbf{X}^{(n-1)}, \nabla\Gamma_\Omega\left(\mathbf{X}^{(n-1)}\right)\right\rangle_{\mathcal{F}} + \frac{C_\Gamma\left(\eta^{(n)}\right)^2}{2}. \quad (61)$$

Note that $\nabla\Gamma_\Omega\left(\mathbf{X}^{(n-1)}\right) = \frac{1}{|\Omega|}\mathbf{J}^{(n-1)}$. If $\mathbf{W}^{(n)}$ satisfies (34), we have

$$\left\langle \mathbf{W}^{(n)} - \mathbf{X}^{(n-1)}, \nabla\Gamma_\Omega\left(\mathbf{X}^{(n-1)}\right)\right\rangle_{\mathcal{F}} \le \left\langle \mathbf{O}^{(n)} - \mathbf{X}^{(n-1)}, \nabla\Gamma_\Omega\left(\mathbf{X}^{(n-1)}\right)\right\rangle_{\mathcal{F}} + \gamma. \quad (62)$$

Note that $\mathbf{O}^{(n)} = \arg\min_{\|\mathbf{O}\|_{\text{nuc}}\le K}\left\langle \mathbf{O}, \nabla\Gamma_\Omega\left(\mathbf{X}^{(n-1)}\right)\right\rangle_{\mathcal{F}}$ according to [25]. Therefore, we have $\mathbf{O}^{(n)} = \arg\max_{\|\mathbf{O}\|_{\text{nuc}}\le K}\left\langle \mathbf{X}^{(n-1)} - \mathbf{O}, \nabla\Gamma_\Omega\left(\mathbf{X}^{(n-1)}\right)\right\rangle_{\mathcal{F}}$. We define $\Upsilon(\Theta) = \Gamma_\Omega(\Theta) - \Gamma_\Omega\left(\widehat{\mathbf{X}}\right)$, where $\widehat{\mathbf{X}}$ is given by (15). The convexity of $\Gamma_\Omega(\mathbf{X})$ implies [30]

$$\left\langle \mathbf{X}^{(n-1)} - \mathbf{O}^{(n)}, \nabla\Gamma_\Omega\left(\mathbf{X}^{(n-1)}\right)\right\rangle_{\mathcal{F}} \ge \Upsilon\left(\mathbf{X}^{(n-1)}\right). \quad (63)$$

Plugging (62) and (63) into (61), we have

$$\Gamma_\Omega\left(\mathbf{X}^{(n)}\right) \le \Gamma_\Omega\left(\mathbf{X}^{(n-1)}\right) - \eta^{(n)}\Upsilon\left(\mathbf{X}^{(n-1)}\right) + \eta^{(n)}\left(\gamma + \frac{C_\Gamma\eta^{(n)}}{2}\right). \quad (64)$$

Letting $n = T$ and subtracting $\Gamma_\Omega\left(\widehat{\mathbf{X}}\right)$ from both sides, we have

$$\begin{aligned}
&\Upsilon\left(\mathbf{X}^{(T)}\right)\\
&\le \Upsilon\left(\mathbf{X}^{(T-1)}\right) - \eta^{(T)}\Upsilon\left(\mathbf{X}^{(T-1)}\right) + \eta^{(T)}\left(\gamma + \frac{C_\Gamma\eta^{(T)}}{2}\right)\\
&= \left(1-\eta^{(T)}\right)\Upsilon\left(\mathbf{X}^{(T-1)}\right) + \eta^{(T)}\left(\gamma + \frac{C_\Gamma\eta^{(T)}}{2}\right)\\
&= \sum_{n=T}^{1}\left(\prod_{j=n+1}^{T}\left(1-\eta^{(j)}\right)\right)\eta^{(n)}\left(\gamma + \frac{C_\Gamma\eta^{(n)}}{2}\right)\\
&\quad + \prod_{n=1}^{T}\left(1-\eta^{(n)}\right)\Upsilon\left(\mathbf{X}^{(0)}\right). \quad (65)
\end{aligned}$$

Recall that $\eta^{(1)} = 1$ and $\eta^{(n)} = \frac{1}{T}, n = 2,\cdots,T$, thus we have $\prod_{n=1}^{T}\left(1-\eta^{(n)}\right) = 0$ and $0 \le \prod_{j=n+1}^{T}\left(1-\eta^{(j)}\right) \le 1, n = 1,\cdots,T$. Then (65) can be written as

$$\Upsilon\left(\mathbf{X}^{(T)}\right) \le \sum_{n=1}^{T}\eta^{(n)}\left(\gamma + \frac{C_\Gamma\eta^{(n)}}{2}\right). \quad (66)$$

Recall that $\Upsilon\left(\mathbf{X}^{(T)}\right) = \Gamma_\Omega\left(\mathbf{X}^{(T)}\right) - \Gamma_\Omega\left(\widehat{\mathbf{X}}\right)$, we have

$$\begin{aligned}
\Gamma_\Omega\left(\mathbf{X}^{(T)}\right) &\le \gamma + \frac{C_\Gamma}{2} + \frac{T-1}{T}\left(\gamma + \frac{C_\Gamma}{2T}\right) + \Gamma_\Omega\left(\widehat{\mathbf{X}}\right)\\
&\le 2\gamma + \frac{C_\Gamma}{2} + \frac{C_\Gamma}{2T} + \Gamma_\Omega\left(\widehat{\mathbf{X}}\right). \quad (67)
\end{aligned}$$

Note that

$$\begin{aligned}
\Gamma_\Omega\left(\widehat{\mathbf{X}}\right) &= \frac{1}{2|\Omega|}\left\|\left(\widehat{\mathbf{X}}\right)_\Omega - \mathbf{Y}\right\|_{\mathcal{F}}^2\\
&\overset{(a)}{\le} \frac{1}{2|\Omega|}\|(\mathbf{X})_\Omega - \mathbf{Y}\|_{\mathcal{F}}^2\\
&\overset{(b)}{=} \frac{1}{2|\Omega|}\|(\mathbf{N})_\Omega\|_{\mathcal{F}}^2 \overset{(c)}{\to} \frac{\sigma^2}{2}, \quad (68)
\end{aligned}$$

where (a) is due to (15); (b) is because $\mathbf{Y} = (\mathbf{X})_\Omega + (\mathbf{N})_\Omega$; (c) is satisfied when $|\Omega| = MN_r\tau_c$ is large, which holds in massive MIMO. Plugging (68) into (67), we have

$$\Gamma_\Omega\left(\mathbf{X}^{(T)}\right) \le 2\gamma + \frac{C_\Gamma}{2} + \frac{C_\Gamma}{2T} + \frac{\sigma^2}{2}. \quad (69)$$

Note that C_Γ is upper bounded by $\frac{K^2}{|\Omega|}$ [36]. Hence, we obtain (35).

APPENDIX B
PROOF OF LEMMA 5

First, we introduce the following lemma.

Lemma 7 (Theorem 4 of [32]): Let $\mathbf{v} \in \mathbb{C}^{n\times 1}$ be the largest right singular vector of matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ $(m > n)$ and let $\widehat{\mathbf{v}}$ be the largest eigenvector of matrix $\mathbf{B} = \mathbf{A}^{\text{H}}\mathbf{A} + \mathbf{C}$, where $\mathbf{C} \in \mathbb{C}^{n\times n}$ is a Hermitian matrix whose upper triangular and diagonal elements are i.i.d samples from $\mathcal{N}_c\left(0,\sigma^2\right)$ and $\mathcal{N}\left(0,\sigma^2\right)$ respectively. Then with high probability

$$\|\mathbf{A}\widehat{\mathbf{v}}\|_{\mathcal{F}}^2 \ge \|\mathbf{A}\mathbf{v}\|_{\mathcal{F}}^2 - O\left(\sigma\sqrt{n}\right). \quad (70)$$

Recall that $\mathbf{v}^{(n)}$ and $\widehat{\mathbf{v}}^{(n)}$ are respectively the largest right singular vector and eigenvector of $\mathbf{J}^{(n-1)}$ and $\widehat{\mathbb{W}}^{(n-1)} = \sum_{m=1}^{M} \widehat{\mathbb{J}}_m^{(n-1)} = \left(\mathbf{J}^{(n-1)}\right)^{\mathrm{H}} \mathbf{J}^{(n-1)} + \sum_{m=1}^{M} \mathbf{G}_m^{(n-1)}$, where $\sum_{m=1}^{M} \mathbf{G}_m^{(n)}$ is a Hermitian matrix whose upper triangular and diagonal elements are i.i.d samples from $\mathcal{N}_{\mathrm{c}}\left(0, M\mu^2\right)$ and $\mathcal{N}\left(0, M\mu^2\right)$ respectively. According to Lemma 7, with high probability, we have

$$\left\|\mathbf{J}^{(n-1)}\widehat{\mathbf{v}}^{(n)}\right\|_{\mathcal{F}}^2 \geq \left\|\mathbf{J}^{(n-1)}\mathbf{v}^{(n)}\right\|_{\mathcal{F}}^2 - O\left(\mu\sqrt{M\tau_{\mathrm{c}}}\right). \quad (71)$$

We define $\alpha = \left\langle \mathbf{O}^{(n)}, \frac{1}{|\Omega|}\mathbf{J}^{(n-1)} \right\rangle_{\mathcal{F}}$ and $\widehat{\alpha} = \left\langle \mathbf{Q}^{(n)}, \frac{1}{|\Omega|}\mathbf{J}^{(n-1)} \right\rangle_{\mathcal{F}}$. Then with high probability, the following holds

$$\begin{aligned}
\widehat{\alpha} &= \left\langle \mathbf{Q}^{(n)}, \frac{1}{|\Omega|}\mathbf{J}^{(n-1)} \right\rangle_{\mathcal{F}} \overset{(a)}{=} -\frac{K\left\|\mathbf{J}^{(n-1)}\widehat{\mathbf{v}}^{(n)}\right\|_{\mathcal{F}}^2}{\widetilde{\lambda}^{(n)}|\Omega|} \\
&\leq -\frac{K\left(\left\|\mathbf{J}^{(n-1)}\mathbf{v}^{(n)}\right\|_{\mathcal{F}}^2 - O\left(\mu\sqrt{M\tau_{\mathrm{c}}}\right)\right)}{\widetilde{\lambda}^{(n)}|\Omega|} \\
&\overset{(b)}{=} \frac{K\left(\frac{\lambda^{(n)}|\Omega|}{K}\alpha + O\left(\mu\sqrt{M\tau_{\mathrm{c}}}\right)\right)}{\widetilde{\lambda}^{(n)}|\Omega|}, \quad (72)
\end{aligned}$$

where (a) follows from the Frobenius inner product; (b) follows from the definition of α. Then we have

$$\begin{aligned}
&\widehat{\alpha} - \alpha \\
&\leq \left(\frac{\lambda^{(n)}}{\widetilde{\lambda}^{(n)}} - 1\right)\alpha + O\left(\frac{K\mu\sqrt{M\tau_{\mathrm{c}}}}{\widetilde{\lambda}^{(n)}|\Omega|}\right) \\
&\overset{(a)}{=} \left(\frac{\widehat{\lambda}^{(n)} - \lambda^{(n)} + \sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}}{\widehat{\lambda}^{(n)} + \sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}}\right)\frac{\lambda^{(n)}K}{|\Omega|} \\
&\quad + O\left(\frac{K\mu\sqrt{M\tau_{\mathrm{c}}}}{\left(\widehat{\lambda}^{(n)} + \sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}\right)|\Omega|}\right) \\
&= \left(\widehat{\lambda}^{(n)} - \lambda^{(n)} + \sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}\right)\frac{\lambda^{(n)}}{\widehat{\lambda}^{(n)} + \sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}}\frac{K}{|\Omega|} \\
&\quad + O\left(\frac{K\mu\sqrt{M\tau_{\mathrm{c}}}}{\left(\widehat{\lambda}^{(n)} + \sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}\right)|\Omega|}\right), \quad (73)
\end{aligned}$$

where (a) follows from the definition of $\widetilde{\lambda}^{(n)}$ in (24) and $\alpha = -\lambda^{(n)}K/|\Omega|$.

By Corollary 2.3.6 from [37], we have $\left\|\widehat{\mathbb{W}}^{(n-1)} - \left(\mathbf{J}^{(n-1)}\right)^{\mathrm{H}}\mathbf{J}^{(n-1)}\right\|_2 = O\left(\mu\sqrt{M\tau_{\mathrm{c}}}\right)$ with high probability. Recall that $\widehat{\lambda}^{(n)}$ and $\lambda^{(n)}$ are respectively the largest eigenvalues of $\widehat{\mathbb{W}}^{(n-1)}$ and $\left(\mathbf{J}^{(n-1)}\right)^{\mathrm{H}}\mathbf{J}^{(n-1)}$. Then we have $\left|\widehat{\lambda}^{(n)} - \lambda^{(n)}\right| = O\left(\sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}\right)$ according to Weyl's inequality [32], which implies that

$$\widehat{\lambda}^{(n)} - \lambda^{(n)} + \sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4} = O\left(\sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}\right), \quad (74)$$

and therefore

$$\frac{\lambda^{(n)}}{\widehat{\lambda}^{(n)} + \sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}} = O(1). \quad (75)$$

Hence, the first term in (73) is $O\left(\frac{K}{|\Omega|}\sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}\right)$. Since $\widehat{\lambda}^{(n)} \geq 0$, the second term in (73) is $O\left(\frac{K}{|\Omega|}\sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}\right)$. In conclusion, we have $\widehat{\alpha} - \alpha \leq O\left(\frac{K}{|\Omega|}\sqrt{\mu}(M\tau_{\mathrm{c}})^{1/4}\right)$ with high probability.

REFERENCES

[1] X. Wang, A. Ashikhmin, and X. Wang, "Wirelessly powered cell-free IoT: Analysis and optimization," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8384–8396, Sep. 2020.

[2] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[3] S. Liang, X. Wang, and L. Ping, "Semi-blind detection in hybrid massive MIMO systems via low-rank matrix completion," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5242–5254, Nov. 2019.

[4] M. Liyanage, J. Salo, A. Braeken, T. Kumar, S. Seneviratne, and M. Ylianttila, "5G privacy: Scenarios and solutions," in *Proc. IEEE 5G World Forum (5GWF)*, Jul. 2018, pp. 197–203.

[5] W. Tong, J. Hua, and S. Zhong, "A jointly differentially private scheduling protocol for ridesharing services," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 10, pp. 2444–2456, Oct. 2017.

[6] Y. Zhang, Y. Mao, and S. Zhong, "Joint differentially private Gale–shapley mechanisms for location privacy protection in mobile traffic offloading systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2738–2749, Oct. 2016.

[7] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 746–789, 1st Quart., 2020.

[8] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *Proc. IEEE 51st Annu. Symp. Found. Comput. Sci.*, Oct. 2010, pp. 51–60.

[9] F. McSherry and I. Mironov, "Differentially private recommender systems: Building privacy into the netflix prize contenders," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 627–636.

[10] M. Abadi *et al.*, "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 308–318.

[11] Z. Huang, S. Mitra, and N. Vaidya, "Differentially private distributed optimization," in *Proc. Int. Conf. Distrib. Comput. Netw.*, Jan. 2015, pp. 1–10.

[12] N. Mohammed, R. Chen, B. Fung, and P. S. Yu, "Differentially private data release for data mining," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 493–501.

[13] X. Zhang *et al.*, "Data-driven caching with users' local differential privacy in information-centric networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[14] M. Du, K. Wang, Y. Chen, X. Wang, and Y. Sun, "Big data privacy preserving in multi-access edge computing for heterogeneous Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 62–67, Aug. 2018.

[15] C. Xu, J. Ren, D. Zhang, and Y. Zhang, "Distilling at the edge: A local differential privacy obfuscation framework for IoT data analytics," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 20–25, Aug. 2018.

[16] M. Arisaka and S. Sugiura, "Energy-versus-bandwidth-efficiency trade-off in spatially modulated massive MIMO downlink," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 197–200, Feb. 2019.

[17] K. Ishibashi and S. Sugiura, "Effects of antenna switching on band-limited spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 345–348, Aug. 2014.

[18] E. Soujeri and G. Kaddoum, "The impact of antenna switching time on spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 5, no. 3, pp. 256–259, Jun. 2016.

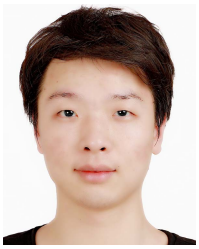
[19] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.

[20] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. 3rd Theory Cryptogr. Conf.*, 2006, pp. 265–284.

[21] M. Kearns, M. Pai, A. Roth, and J. Ullman, "Mechanism design in large games: Incentives and privacy," in *Proc. 5th Conf. Innov. Theor. Comput. Sci.*, 2014, pp. 403–410.

[22] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[23] J. Hsu, Z. Huang, A. Roth, T. Roughgarden, and Z. S. Wu, "Private matchings and allocations," *SIAM J. Comput.*, vol. 45, no. 6, pp. 1953–1984, Jan. 2016.
[24] R. M. Freund, P. Grigas, and R. Mazumder, "An extended Frank-Wolfe method with 'in-face' directions, and its application to low-rank matrix completion," *SIAM J. Optim.*, vol. 27, no. 1, pp. 319–346, 2017.
[25] P. Jain, O. Thakkar, and A. Thakurta, "Differentially private matrix completion revisited," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 2215–2224.
[26] W. Zheng, A. Bellet, and P. Gallinari, "A distributed Frank-Wolfe framework for learning low-rank matrices with the trace norm," *Mach. Learn.*, vol. 107, nos. 8–10, pp. 1457–1475, 2018.
[27] H.-T. Wai, J. Lafond, A. Scaglione, and E. Moulines, "Decentralized Frank-Wolfe algorithm for convex and nonconvex problems," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 5522–5537, Nov. 2017.
[28] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," *J. Mach. Learn. Res.*, vol. 11, pp. 2057–2078, Mar. 2010.
[29] P. Jain, C. Jin, S. M. Kakade, P. Netrapalli, and A. Sidford, "Streaming PCA: Matching matrix bernstein and near-optimal finite sample guarantees for Oja's algorithm," in *Proc. Conf. Learn. Theory*, vol. 2016, pp. 1147–1164.
[30] M. Jaggi, "Revisiting Frank-Wolfe: Projection-free sparse convex optimization," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. 427–435.
[31] N. Srebro and A. Shraibman, "Rank, trace-norm and max-norm," in *Proc. 18th Int. Conf. Comput. Learn. Theory*, 2005, pp. 545–560.
[32] C. Dwork, K. Talwar, A. Thakurta, and L. Zhang, "Analyze Gauss: Optimal bounds for privacy-preserving principal component analysis," in *Proc. 46th Annu. ACM Symp. Theory Comput.*, 2014, pp. 11–20.
[33] X. Peng, C. Lu, Z. Yi, and H. Tang, "Connections between nuclear-norm and Frobenius-norm-based representations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 218–224, Jan. 2018.
[34] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1311–1321, Aug. 2004.
[35] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
[36] K. L. Clarkson, "Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm," *ACM Trans. Algorithms*, vol. 6, no. 4, pp. 1–30, Aug. 2010.
[37] T. Tao, *Topics in Random Matrix Theory*, vol. 132. Providence, RI, USA: AMS, 2012.



Jun Xu was born in Anhui, China, in 1993. He received the B.S. degree from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2016, where he is currently pursuing the Ph.D. degree in information and communication engineering with the National Mobile Communications Research Laboratory.

From October 2019 to October 2020, he was a Visiting Scholar with the Department of Electrical Engineering, Columbia University, New York, NY, USA. His research interests include distributed massive MIMO, channel estimation, as well as physical layer security and privacy.

Xiaodong Wang (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Princeton University.

He is currently a Professor of Electrical Engineering with Columbia University, New York, NY, USA. His research interests fall in the general areas of computing, signal processing and communications, and has published extensively in these areas. Among his publications is the book *Wireless Communication Systems: Advanced Techniques for Signal Reception* (Prentice Hall, 2003). His current research interests include wireless communications, statistical signal processing, and genomic signal processing. He received the 1999 NSF CAREER Award, the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award, and the 2011 IEEE Communication Society Award for Outstanding Paper on New Communication Topics. He has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON INFORMATION THEORY. He is listed as an ISI Highly-Cited Author.

Pengcheng Zhu (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Shandong University, Ji'nan, China, in 2001 and 2004, respectively, and the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, China, in 2009. He is currently a Professor with the National Mobile Communications Research Laboratory, Southeast University. His research interests lie in the areas of wireless communications and mobile networks, including B5G/6G mobile communication systems, physical layer security and privacy, distributed MIMO, and mmWave communications.

Xiaohu You (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 1982, 1985, and 1988, respectively.

Since 1990, he has been with the National Mobile Communications Research Laboratory, Southeast University, where he holds the ranks of Professor and the Director. He is the Chief of the Technical Group of the China 3G/B3G Mobile Communication Research and Development Project. His research interests include mobile communications, adaptive signal processing, and artificial neural networks, with applications to communications and biomedical engineering. He was a recipient of the Excellent Paper Prize from the China Institute of Communications in 1987, the Elite Outstanding Young Teacher Award from Southeast University in 1990, 1991, and 1993, respectively, and the National Technological Invention Award of China in 2011. He was also a recipient of the 1989 Young Teacher Award of the Fok Ying Tung Education Foundation, State Education Commission of China.