

Beyond MMSE: Rank-1 Subspace Channel Estimator for Massive MIMO Systems

Bin Li, Ziping Wei, Shaoshi Yang, *Senior Member, IEEE*, Yang Zhang, Jun Zhang, *Senior Member, IEEE*, Chenglin Zhao, Sheng Chen, *Life Fellow, IEEE*

Abstract—To glean the benefits offered by massive multi-input multi-output (MIMO) systems, channel state information must be accurately acquired. Despite the high accuracy, the computational complexity of classical linear minimum mean squared error (MMSE) estimator becomes prohibitively high in the context of massive MIMO, while the other low-complexity methods degrade the estimation accuracy seriously. In this paper, we develop a novel rank-1 subspace channel estimator to approximate the maximum likelihood (ML) estimator, which outperforms the linear MMSE estimator, but incurs a surprisingly low computational complexity. Our method first acquires the highly accurate angle-of-arrival (AoA) information via a constructed space-embedding matrix and the rank-1 subspace method. Then, it adopts the *post-reception* beamforming to acquire the unbiased estimate of channel gains. Furthermore, a fast method is designed to implement our new estimator. Theoretical analysis shows that the extra gain achieved by our method over the linear MMSE estimator grows according to the rule of $\mathcal{O}(\log_{10} M)$, while its computational complexity is *linearly* scalable to the number of antennas M . Numerical simulations also validate the theoretical results. Our new method substantially extends the accuracy-complexity region and constitutes a promising channel estimation solution to the emerging massive MIMO communications.

Index Terms—Massive MIMO, channel estimation, MMSE estimator, rank-1 subspace, *post-reception* beamforming, Cramer-Rao lower bound, low complexity

I. INTRODUCTION

Massive multi-input multi-output (MIMO), as well as its evolution, has been recognized as a key technology in 5G and 6G communications [1]–[8]. It is capable of substantially improving both the spectrum and energy efficiency [2], thus enabling a wide variety of emerging applications [4], [9], [10]. However, the vast potential benefits of massive MIMO can be attained only if the channel state information (CSI) is accurately acquired. Given the prohibitive complexity of the maximum likelihood (ML) estimator, the task of accurate CSI acquisition is usually accomplished by another classical linear

minimum mean squared error (MMSE) channel estimator [3], [11]–[16]. The linear MMSE estimator exploits the covariance matrix of wireless channels and is capable of attaining high accuracy. Unfortunately, the *a priori* statistical covariance information tends to be unavailable, particularly in realistic dynamic environments. Moreover, the linear MMSE estimator may also incur an excessively high computational complexity and a significant processing delay, and hence becomes impractical to low-cost and low-power devices. For these reasons, the linear least-square (LS) channel estimator [1] has been widely applied. However, the LS estimator inevitably degrades the CSI accuracy and thus the channel capacity, despite its benefit of low-complexity implementation [17], [18].

Recently, a variety of other channel estimators have been proposed for massive MIMO systems, in order to achieve more balanced tradeoff between estimation accuracy and computational complexity. These estimators can be divided into three types. The first type of channel estimators exploit the low-rank property [18]–[21], as massive MIMO channels are in general spatially correlated [19]. By approximately computing the inverse of a large matrix in the linear MMSE estimator, a weighted polynomial expansion channel (WPEACH) estimator was designed [17], [22], which results in a substantial reduction in computational complexity. However, in time-varying environments, the updating of the weight coefficients is highly complex [18]. The authors of [20] proposed a singular value projection (SVP) method, which has lower mean squared error (MSE) than the LS estimator and lower complexity than the linear MMSE estimator. The second type of estimators resort to the sparsity property of massive MIMO channels operating in millimeter-wave (mmWave) frequencies [23]–[29]. For example, an iterative LS estimator with sparse message passing (LSE-SMP) was proposed to acquire CSI [28], [29], and it attains the minimum variance unbiased estimate.

The third type of channel estimators first exploit the array signal processing techniques [30]–[32] that are widely used in radar systems, such as multiple signal classification (MUSIC) and fast Fourier transform (FFT), to estimate the angle of arrival (AoA), and then invoke the acquired accurate AoA information to assist the fading gain estimation. With the accurate information of AoAs, it is expected that the signal-to-noise ratio (SNR) in fading gain estimation will be improved [33], [34]. When the MUSIC algorithm is applied, the high resolution AoA estimate can be attained, yet its complexity is formidable to massive MIMO systems. Besides, the subspace computation is infeasible for uniform linear array (ULA), as the covariance matrix becomes unavailable in the multi-user case. For this reason, FFT has been widely applied. As a result, this type of channel estimators are capable of effectively

This work was supported in part by the Beijing Municipal Natural Science Foundation under Grant L234025 and Grant Z220004, in part by the National Natural Science Foundation of China under Grant 62071247, in part by the Beijing Municipal Science & Technology Commission under Grant Z221100007722036, and in part by the Fundamental Research Funds for the Central Universities under Grant 2023ZCJH02. (*Corresponding author: Shaoshi Yang, Chenglin Zhao.*)

B. Li, Z. Wei, S. Yang and C. Zhao are with School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China (E-mails: Binli@bupt.edu.cn, weizp@bupt.edu.cn, shaoshi.yang@bupt.edu.cn, clzhao@bupt.edu.cn).

Y. Zhang is with China Satellite Network Group Co., Ltd, Beijing, China (E-mail: zhangyang@bupt.cn).

J. Zhang is with Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (E-mail: zhangjun@njupt.edu.cn).

S. Chen is with School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (E-mail: sqc@ecs.soton.ac.uk).

reducing the complexity. However, due to its leakage effect, the angular resolution is seriously limited and the estimated AoAs even become *biased*. Therefore, the existing angular-domain estimators may potentially suffer from either unaffordable complexity or degraded accuracy.

The designs of the state-of-the-art channel estimators for massive MIMO are to strike an attractive compromise between the two conflicting objectives, namely, the computational complexity and the estimation accuracy. Most existing methods can reduce the computational complexity to some extent [17], [20], [34], compared with the linear MMSE estimator; but none of them is capable of outperforming the linear MMSE estimator. Despite its great significance in both theoretical analysis and practical applications, designing a channel estimator whose achievable performance is superior to that of the linear MMSE estimator remains an important open problem, especially when the low computation complexity is also required.

Against the above backdrop, in this work we propose a novel rank-1 subspace channel estimator for massive MIMO. To the best of our knowledge, it is the first time that an approximate ML estimator is designed, which can outperform the classical linear MMSE estimator, yet requiring a dramatically reduced computational complexity. To attain this, it firstly acquires the AoA information by leveraging a space-embedding Hankel matrix and exploiting the rank-1 subspace, which attains the near-optimal AoA estimates and breaks the resolution limit of the FFT method. Second, a *post-reception* beamforming scheme is suggested to estimate fading gains, relying on the ML criterion. More importantly, we provide the theoretical result on the lower-bound error of this two-stage estimator, showing that the achieved gain over the linear MMSE estimator increases obeying the rule of $\mathcal{O}(\log_{10} M)$, where M is the number of antennas at the base station (BS). In order to reduce the complexity, we further integrate a fast method, by using the low rank property of channel matrix of massive MIMO communications. It thus achieves the highest-ever complexity reduction, which is only on the order of $\mathcal{O}(KP^2M)$, compared with $\mathcal{O}(M^3)$ for the MMSE estimator, whereby $K < M$ is the number of users and $P \ll M$ is the number of paths.

Note that, although the single-snapshot MUSIC was studied in MIMO radar systems [35]–[37], this is the first time it is introduced to massive MIMO communications. By incorporating a post-reception beamforming scheme and the ML-based estimation of fading gains, the single-snapshot MUSIC approximates the ML estimator. Moreover, we maximally simplify the computational decomposition of large matrix, thus making complex subspace methods applicable to massive MIMO systems. As such, our method overcomes two challenges in classical CSI estimators, by significantly improving the accuracy and reducing the complexity. The main contributions of this work are summarized as follows.

- 1) We develop a two-stage channel estimator for massive MIMO systems by leveraging the rank-1 subspace method. We first obtain the estimation of unknown AoAs, relying on a constructed Hankel matrix and an estimated pseudo-spectrum. Thus, the high-resolution AoAs are acquired, resulting in the estimation error on

the order of $\mathcal{O}(1/M^{1+\varepsilon})$ for $0 < \varepsilon < 1$. Moreover, we adopt the post-reception beamforming to estimate unknown channel gains, which effectively improves the SNR and, unlike the classical FFT method, attains the *unbiased* channel gain estimation.

- 2) We prove that the Cramér-Rao lower bound (CRLB) of our two-stage CSI estimator is improved by $\mathcal{O}(1/M)$, compared with the linear MMSE estimator. For the target MSE, our near-ML estimator thus attains an extra gain that increases according to $\mathcal{O}(\log_{10} M)$. As a result, the channel estimation accuracy is dramatically improved, which in turn enhances the communication coverage and the channel capacity.
- 3) We design a fast method to implement this rank-1 subspace channel estimator, in the case of low-rank channel matrix. By utilizing the inherent low-rank property of a large Hankel matrix, we first approximate it via three small matrices. Then, the singular value decomposition (SVD) of the large Hankel matrix is replaced with the SVD of these small matrices. This reduces the computational complexity of our estimator from $\mathcal{O}(KM^3)$ to $\mathcal{O}(KP^2M)$ ($P \ll M$). Moreover, we show that our fast method attains the high-resolution AoA estimation, without compromising the accuracy.
- 4) The theoretical performance analysis of our new estimator is consolidated by comprehensive numerical simulation results, which demonstrate that our method indeed outperforms the linear MMSE estimator and the sparsity-based methods, while dramatically reducing the computational complexity by 2~3 orders of magnitude. In other words, the substantial performance gain is attained by imposing a much lower complexity. Thus, our new CSI estimator substantially extends the attainable complexity-accuracy region of CSI estimation for massive MIMO systems.

The rest of our paper is organized as follows. In Section II, we analyze the AoA-aided post-reception beamforming for acquiring the channel gains in the single-user case, for emphasizing the critical role of accurate AoA estimation. Section III details our proposed rank-1 subspace channel estimator, which involves the successive estimation of AoAs and channel gains. In Section IV, we theoretically analyze the estimation accuracy of our method. In Section V a fast method is developed to implement our estimator, and it remarkably reduces the computational complexity without degrading the accuracy. Numerical simulations are provided in Section VI to validate our method, and we conclude this paper in Section VII.

This paper adopts the following notation conventions. Bold-face lower-case and capital letters denote vectors and matrices, respectively. $\mathbf{X}^*$, $\mathbf{X}^T$ and $\mathbf{X}^H$ represent the conjugate, the transpose and the conjugate transpose of $\mathbf{X}$, respectively, while $\mathbf{X}^{-1}$ and $\mathbf{X}^\dagger = \mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}$ are the inverse and the Moore-Penrose pseudo-inverse of $\mathbf{X}$, respectively. $\text{vec}(\mathbf{H})$ is the vectorization of $\mathbf{H}$, $\otimes$ is the Kronecker product, and $\lceil \cdot \rceil$ is the ceiling operator. $\mathbf{I}_M$ is the $M \times M$ identity matrix, and $\|\cdot\|_F$ is the Frobenius norm, while $\mathbb{E}[\cdot]$ denotes the expectation operator and $\text{var}[\cdot]$ stands for variance.

II. CHANNEL ESTIMATION AIDED WITH AOA

In the uplink training of massive MIMO systems, if the AoA information is available to the BS, it can be utilized in the post-reception beamforming to improve the channel estimation accuracy. In this section, we first analyze the AoA-aided post-reception beamforming for acquiring the channel gains in a single-user case, upon assuming perfect AoA information. This provides some insight on how the post-reception beamforming improves the channel estimation accuracy. We next derive a necessary condition on the estimated AoA information in order to attain the *unbiased* channel estimation, and thereby show that classical FFT-based angular-domain channel estimators may fail to satisfy this condition.

We consider the up-link channel estimation with a single-antenna user equipment (UE), which transmits the pilot signal of length $B \geq 1$, while the BS equipped with M antennas estimates the unknown channel gains. Hence, the received signal $\mathbf{Y} \in \mathbb{C}^{M \times B}$ is given by

$$\mathbf{Y} = \mathbf{h}\mathbf{x}^H + \mathbf{N}, \quad (1)$$

where $\mathbf{x} \in \mathbb{C}^{B \times 1}$ is the random pilot signal (e.g., i.i.d. Gaussian sequence) with a covariance matrix $\sigma_x^2 \mathbf{I}_B$, $\mathbf{N} \in \mathbb{C}^{M \times B}$ is the additive white Gaussian noise (AWGN) matrix whose entries have zero mean and variance σ_n^2 , and $\mathbf{h} \in \mathbb{C}^{M \times 1}$ is the wireless channel vector. Without loss of generality, we use the Saleh-Valenzuela (S-V) channel model to capture both the spatial sparsity and the low-rank property of massive MIMO channels [24], [38]–[41]. The low-rank channel occurs widely in massive MIMO communications, especially when the BS deployed in high buildings suppresses local scattering [19], [20] and the signal propagation of multiple clustered users shares the same reflectors. Then, the channel vector between the UE and the BS is given by [18]–[20], [39]

$$\mathbf{h} = \sum_{p=0}^{P-1} \alpha_p \mathbf{a}_M(\theta_p) = \sum_{p=0}^{P-1} \mathbf{c}_p, \quad (2)$$

where P is the number of paths; $\mathbf{c}_0 \triangleq \alpha_0 \mathbf{a}_M(\theta_0)$ is the line-of-sight (LoS) component with the complex-valued gain α_0 and the AoA θ_0 ; while $\mathbf{c}_p \triangleq \alpha_p \mathbf{a}_M(\theta_p)$, $1 \leq p \leq P-1$, denotes the p th non-line-of-sight (NLoS) component that has the complex-valued gain α_p and the AoA θ_p [20], [42]. As shown in [43], when operating in the high frequency band (e.g., millimeter-wave band), the number of paths typically satisfy $P \leq 5$ in realistic environments. If ULA is used, the steering vector $\mathbf{a}_M(\theta) \in \mathbb{C}^{M \times 1}$ corresponding to the AoA θ reads

$$\mathbf{a}_M(\theta) \triangleq \left[1 \quad \exp\left(\frac{j2\pi d \sin \theta}{\lambda}\right) \quad \cdots \quad \exp\left(\frac{j2\pi(M-1)d \sin \theta}{\lambda}\right) \right]^T. \quad (3)$$

Here, λ is the carrier wavelength and d is the antenna spacing.

If we further define the received signal vector as

$$\mathbf{y} \triangleq \mathbf{Y}\mathbf{x} \in \mathbb{C}^{M \times 1}, \quad (4)$$

and assume that the accurate AoA information of P paths, $\{\hat{\theta}_p\}_{p=0}^{P-1}$, is available, then we estimate the channel gains by

invoking a post-reception beamforming as

$$\hat{\alpha}_p = \frac{1}{M} \mathbf{a}_M^H(\hat{\theta}_p) \mathbf{y}, \quad 0 \leq p \leq P-1. \quad (5)$$

Theorem 1. *When the uplink AoAs of a single-antenna UE are exactly known, i.e., $\hat{\theta}_p = \theta_p$ for $0 \leq p \leq P-1$, and the number of BS antennas M is large, the channel gains estimated by using the post-reception beamforming (5) are unbiased, i.e., $\mathbb{E}[\hat{\alpha}_p] = \alpha_p$, with the estimation error bounded by $\frac{\sigma_n^2}{MB\sigma_x^2}$, for $0 \leq p \leq P-1$.*

Proof. See Appendix A. $\square$

It can be seen that in this idealized scenario with perfect P AoAs, the estimation accuracy for the channel gains approaches the optimum, achieving a CRLB of $\frac{\sigma_n^2}{MB\sigma_x^2}$. In practice, however, the BS does not have the perfect AoA information and can only estimate these P AoAs. Naturally, the accuracy of the estimated AoAs, $\{\hat{\theta}_p\}_{p=0}^{P-1}$, affects the accuracy of estimating the channel gains, when using the post-reception beamforming (5).

Theorem 2. *To achieve the unbiased estimation of the channel gains via (5), a necessary condition is that the estimation error of the AoA information must satisfy:*

$$|\hat{\theta}_p - \theta_p| = \mathcal{O}(1/M^{1+\epsilon}), \quad \exists \epsilon > 0, \forall p. \quad (6)$$

Proof. See Appendix B. $\square$

As mentioned, there exist some angular-domain channel estimators, which first acquire the AoAs and then estimate the channel gains [34]. In terms of the achievable MSE, such angular-domain estimators [34] are superior to the LS estimator, but inferior to the linear MMSE estimator. The main reason is that, in the case of ULA, the information available at the BS is limited and thus inadequate for classical subspace methods to acquire high-resolution AoAs¹. Thus, only the FFT method can be applicable to acquiring AoAs, but it attains low-resolution and even biased AoAs, due to the limited FFT length (recall that the angular resolution is $\Delta\theta = 1/M$ in the FFT method). As a result, the channel gain estimation upon using the post-reception beamforming is biased unfortunately. Then, the FFT-based AoA estimation, with an error decaying rate obeying $\mathcal{O}(1/M)$, is insufficient to achieve the unbiased channel estimation via the post-reception beamforming. This will also be demonstrated later by our numerical simulations, where the FFT-based method exhibits the MSE floor in the high-SNR region.

The above discussion emphasizes the significant importance of the viable, high-resolution AoA estimates for the accurate CSI acquisition in massive MIMO systems, which partially motivates our work presented in this paper.

III. RANK-1 SUBSPACE CHANNEL ESTIMATOR

In this section, we first introduce a new approach to acquire high-resolution AoA information, which then assists us to

¹ Another 2D MUSIC method has been studied in [33], which also becomes inapplicable to the case considered here, i.e., one ULA serves multiple UEs, where the rank of the covariance matrix $\mathbf{R}_y = \frac{1}{M} \mathbf{y}\mathbf{y}^H$ is one.

obtain the unbiased and accurate channel estimation via the post-reception beamforming. In the case of ULA with K UEs, the received signal reduces to an $M \times 1$ spatial-domain vector, since the other dimension of the signal matrix vanishes after removing the multi-user interference (MUI), as will be seen later. Thus, classical subspace methods (e.g. standard MUSIC) relying on the covariance matrix becomes invalid. To overcome this difficulty, we design a novel rank-1 subspace channel estimator, by leveraging the *space-embedding* technique in analogy to the time-embedding technique used in modeling dynamical systems [44]–[46]. Note that the spatial smoothing MUSIC technique is also applicable in this context.

A. Single-User Uplink Channel Estimation

We begin with the single-user case ($K = 1$). Let the length of pilot signal be $B \geq K = 1$. The elements of the received signal vector $\mathbf{y} = [y_0 \ y_1 \ \cdots \ y_{M-1}]^H$ of (4) can be expressed for $0 \leq m \leq M - 1$ as

$$y_m = \sum_{p=0}^{P-1} \alpha_p \exp\left(\frac{j2\pi m d \sin \theta_p}{\lambda}\right) + n_m. \quad (7)$$

It can be seen that the SNR is increased by B times, since we have $\text{var}[n_m] = \sigma_n^2/B\sigma_x^2$. By defining the steering matrix as

$$\mathbf{A}_M(\boldsymbol{\theta}) = [\mathbf{a}_M(\theta_0) \ \mathbf{a}_M(\theta_1) \ \cdots \ \mathbf{a}_M(\theta_{P-1})] \in \mathbb{C}^{M \times P}, \quad (8)$$

with $\boldsymbol{\theta} = [\theta_0 \ \theta_1 \ \cdots \ \theta_{P-1}]^H$, the received signal vector can be restructured as

$$\mathbf{y} = \mathbf{A}_M(\boldsymbol{\theta})\boldsymbol{\alpha} + \mathbf{n}, \quad (9)$$

where we have $\boldsymbol{\alpha} = [\alpha_0 \ \alpha_1 \ \cdots \ \alpha_{P-1}]^H \in \mathbb{C}^{P \times 1}$.

To extract the unknown AoA information $\boldsymbol{\theta}$, we adopt a novel rank-1 subspace method which allows for high-resolution AoA estimation [35], [47]. To be specific, we first construct a space-embedding Hankel matrix $\bar{\mathbf{Y}} \in \mathbb{C}^{L \times (M-L)}$ from the received signal vector $\mathbf{y}$, i.e.,

$$\bar{\mathbf{Y}} \triangleq \begin{bmatrix} y_0 & y_1 & \cdots & y_{M-L-1} \\ y_1 & y_2 & \cdots & y_{M-L} \\ \vdots & \vdots & \ddots & \vdots \\ y_{L-1} & y_L & \cdots & y_{M-1} \end{bmatrix} = \bar{\mathbf{A}}_L \boldsymbol{\Lambda} \bar{\mathbf{A}}_{M-L}^H + \bar{\mathbf{N}}, \quad (10)$$

where we have the diagonal matrix $\boldsymbol{\Lambda} \triangleq \text{diag}\{\alpha_0, \alpha_1, \dots, \alpha_{P-1}\}$, the stack length L satisfies $L \geq P$ and $M - L \geq P$, while $\bar{\mathbf{A}}_L \in \mathbb{C}^{L \times P}$ is the submatrix consisting of the first L rows of $\mathbf{A}_M(\boldsymbol{\theta})$ given in (8). Since the above Hankel matrix $\bar{\mathbf{Y}}$ admits the Vandermonde decomposition, one can directly obtain the signal space and noise subspace by performing SVD on $\bar{\mathbf{Y}}$ [35].

Thus, like other subspace-based methods, e.g., the MUSIC method [31], [32], [48], in order to accomplish high-resolution AoA estimation we first compute the SVD of $\bar{\mathbf{Y}}$ as

$$\bar{\mathbf{Y}} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^H = \mathbf{U}_P \boldsymbol{\Sigma}_P \mathbf{V}_P^H + \mathbf{U}_{-P} \boldsymbol{\Sigma}_n \mathbf{V}_{-P}^H, \quad (11)$$

where we denote the left singular matrix $\mathbf{U} \in \mathbb{C}^{L \times L}$, the diagonal singular value matrix $\boldsymbol{\Sigma}_P = \text{diag}\{\sigma_1(\bar{\mathbf{Y}}), \sigma_2(\bar{\mathbf{Y}}), \dots, \sigma_P(\bar{\mathbf{Y}})\}$ and $\boldsymbol{\Sigma}_n =$

$\text{diag}\{\sigma_{P+1}(\bar{\mathbf{Y}}), \sigma_{P+2}(\bar{\mathbf{Y}}), \dots, \sigma_L(\bar{\mathbf{Y}})\}$, with the singular values $\sigma_1(\bar{\mathbf{Y}}) \geq \dots \geq \sigma_P(\bar{\mathbf{Y}}) \geq \sigma_{P+1}(\bar{\mathbf{Y}}) = \dots = \sigma_L(\bar{\mathbf{Y}})^2$, and the right singular matrix is $\mathbf{V} \in \mathbb{C}^{(M-L) \times L}$, while $\mathbf{U}_P \in \mathbb{C}^{L \times P}$ is the submatrix consisting of the first P columns of $\mathbf{U}$ and it corresponds to the signal subspace, and $\mathbf{V}_P \in \mathbb{C}^{(M-L) \times P}$ is the submatrix consisting of the first P columns of $\mathbf{V}$. Furthermore, $\mathbf{U}_{-P} \in \mathbb{C}^{L \times (L-P)}$ is the submatrix consisting of the last $(L - P)$ columns of $\mathbf{U}$ and it corresponds to the noise subspace, and $\mathbf{V}_{-P} \in \mathbb{C}^{(M-L) \times (L-P)}$ is the submatrix consisting of the last $(L - P)$ columns of $\mathbf{V}$. Note that the first P singular values of matrix $\bar{\mathbf{Y}}$ are significantly larger than the remaining $L - P$ singular values (e.g., in the high-SNR case). Thus, we refer $\bar{\mathbf{Y}}$ as an approximate low-rank matrix, i.e., $\text{rank}(\bar{\mathbf{Y}}) = \text{rank}(\boldsymbol{\Lambda}) \approx P$.

For each angle $\theta \in [-\pi, \pi]$, the spatial pseudo-spectrum $P(\theta)$ is then estimated via [35]:

$$P(\theta) = \frac{1}{\mathbf{a}_L^H(\theta)(\mathbf{I}_L - \mathbf{U}_P \mathbf{U}_P^H) \mathbf{a}_L(\theta)}, \quad (12)$$

where $\mathbf{a}_L(\theta) \in \mathbb{C}^{L \times 1}$ denotes the subvector consisting of the first L elements of $\mathbf{a}_M(\theta)$. Upon evaluating the spatial pseudo-spectrum with the total N discretized angles $\{-\pi + n\Delta\theta, n = 0, 1, \dots, N - 1\}$, where $\Delta\theta = 2\pi/N$ is the angle resolution and N is a sufficiently large integer, multiple AoAs of unknown channel paths can be extracted, e.g., by identifying the P highest peaks in $P(\theta)$:

$$\hat{\boldsymbol{\theta}} \triangleq \{\hat{\theta}_p \mid \hat{\theta}_p = \text{peak}[P(\theta)], 0 \leq p \leq P - 1\}. \quad (13)$$

After obtaining the unknown AoAs $\{\hat{\theta}_p\}_{p=0}^{P-1}$, the estimated channel gains $\{\hat{\alpha}_p\}_{p=0}^{P-1}$ can be computed using the post-reception beamforming (5). In fact, based on the obtained highly accurate AoA information our estimator essentially achieves the ML estimation:

$$\hat{\alpha}_p = \arg \max_{\alpha_p \in \mathbb{C}} p \left(\frac{1}{M} \mathbf{a}_M^H(\hat{\theta}_p) (\mathbf{h} - \mathbf{y}) | \alpha_p \right), \quad (14)$$

where the likelihood is given by

$$\begin{aligned} p \left(\frac{1}{M} \mathbf{a}_M^H(\hat{\theta}_p) (\mathbf{h} - \mathbf{y}) | \alpha_p \right) &\approx p \left(\alpha_p - \frac{1}{M} \mathbf{a}_M^H(\hat{\theta}_p) \mathbf{y} \right) \\ &\approx p \left(\frac{1}{M} \mathbf{a}_M^H(\hat{\theta}_p) \mathbf{n} \right), \end{aligned} \quad (15)$$

since $\frac{1}{M} \mathbf{a}_M^H(\hat{\theta}_p) \mathbf{a}_M^H(\theta_p) \approx 1$ and $\frac{1}{M} \mathbf{a}_M^H(\hat{\theta}_p) \mathbf{a}_M^H(\theta_{p'}) \approx 0$, $p \neq p'$, for large M and accurate $\hat{\theta}_p$. That is, it is a Gaussian distribution, and solving the optimization problem (14) leads to the ML solution of (5). Finally, we arrive at the equivalent angular-domain channel estimate

$$\hat{\mathbf{h}} = \sum_{p=0}^{P-1} \hat{\alpha}_p \mathbf{a}_M(\hat{\theta}_p). \quad (16)$$

B. Multiple-User Uplink Channel Estimation

Now consider the multi-user MIMO (MU-MIMO) case, where the BS equipped with M antennas serves K single-

antenna users. Given the pilot signal $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \cdots \mathbf{x}_K] \in \mathbb{C}^{B \times K}$, the received signal matrix $\mathbf{Y} \in \mathbb{C}^{M \times B}$ is given by

$$\mathbf{Y} = \mathbf{H}\mathbf{X}^H + \mathbf{N}, \quad (17)$$

where again $\mathbf{N}$ is the channel AWGN matrix, and the channel matrix $\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \cdots \mathbf{h}_K] \in \mathbb{C}^{M \times K}$ with

$$\mathbf{h}_k = \sum_{p=0}^{P-1} \alpha_{p,k} \mathbf{a}_M(\theta_{p,k}), \quad 1 \leq k \leq K. \quad (18)$$

By exploiting the orthogonal property for the pilot sequences of different users, namely, $\mathbf{x}_k^H \mathbf{x}_k = 1$ and $\mathbf{x}_k^H \mathbf{x}_{k'} = 0$ for $k, k' \in \{1, 2, \dots, K\}$ and $k \neq k'$, we first obtain the received signal vector of the k th user:

$$\mathbf{y}_k = \mathbf{Y} \mathbf{x}_k \in \mathbb{C}^{M \times 1}, \quad 1 \leq k \leq K. \quad (19)$$

For each signal vector $\mathbf{y}_k$, the MUI has been removed. Thus, we will adopt the rank-1 subspace method in Section III-A to acquire the highly accurate AoAs of user k .

To be specific, a space-embedding matrix $\bar{\mathbf{Y}}_k$ is constructed similar to (10). On this basis, we can estimate the unknown channel vector of the k th user as $\hat{\mathbf{h}}_k = \sum_{p=0}^{P-1} \hat{\alpha}_{p,k} \mathbf{a}_M(\theta_{p,k})$. By repeating the above single-user estimator for K times, the whole channel matrix can be estimated as $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1 \ \hat{\mathbf{h}}_2 \cdots \hat{\mathbf{h}}_K]$.

C. Complexity Analysis

For the single-user case, the computational complexity of our rank-1 subspace channel estimator consists of three parts: 1) the SVD of a space-embedding matrix $\bar{\mathbf{Y}} \in \mathbb{C}^{L \times (M-L)}$, which requires the computational complexity on the order of $\mathcal{O}(M^3)$, where we have assumed $L = \mathcal{O}(M)$, e.g., $L = \frac{M}{2}$; 2) the calculation of pseudo-spectrum, which imposes the complexity of $\mathcal{O}(NLP)$ (note that this procedure involves only matrix multiplication and has highly efficient parallel acceleration, see [48], [49]); 3) the estimation of the unknown channel gains, which requires the computational complexity of $\mathcal{O}(PM)$. Therefore, the overall complexity of our proposed estimator is on the order of $\mathcal{O}(M^3)$.

For the MU-MIMO case with K users, the computational complexity of our estimator is obviously on the order of $\mathcal{O}(KM^3)$. In massive MIMO systems, M is very large, and therefore this computational complexity is still too high. In Section V, we further design a fast rank-1 subspace CSI estimator, relying on the randomized low-rank approximation technique [32], [50]–[52], with which the computational complexity of our channel estimator can be reduced dramatically, whilst the AoA estimation accuracy suffers no degradation.

IV. THEORETICAL PERFORMANCE ANALYSIS

In this section, we analyze the accuracy of our rank-1 subspace channel estimator for massive MIMO uplink.

A. Error of Estimated Channel AoAs

We first investigate the estimation accuracy of the AoA information of each user. We easily note that the norm of the Gaussian noise matrix in (10) obeys the following rule:

$$\|\bar{\mathbf{N}}\|_2 = \mathcal{O}\left(\sqrt{M \log_2(M)}\right). \quad (20)$$

On this basis, if the stack length is chosen to be $L \sim \mathcal{O}(M)$ [35], e.g., $L = M/2$, we can acquire the accurate pseudo-spectrum estimation. When $M \rightarrow \infty$ and the Gaussian noise matrix $\bar{\mathbf{N}}$ meets the condition, $\|\bar{\mathbf{N}}\|_2^2 \sim M \log_2(M)$, the error of the estimated pseudo-spectrum is bounded [35]:

$$|P(\theta) - P_0(\theta)| = \mathcal{O}\left(\frac{\sqrt{\log_2(M)}}{\sqrt{M}}\right), \quad (21)$$

where $P_0(\theta)$ is the exact spatial pseudo-spectrum in the absence of additive noise. As such, for massive MIMO uplink, we are able to obtain the highly accurate spatial pseudo-spectrum, even in the presence of additive noise.

Furthermore, the error of the estimated AoA is also bounded, as specified by the following Lemma 1 [35].

Lemma 1. Assume that $L \geq P$ and $M - L \geq P$. Define

$$f(L) \triangleq \frac{1}{L} \sqrt{\frac{2}{\pi}} \left(\frac{2}{\pi} - \frac{4}{L}\right)^{-\frac{1}{2}}. \quad (22)$$

Denote the minimal gap between two adjacent AoAs as $d = \min_{p \neq p'} |\theta_{p,k} - \theta_{p',k}|$. If

$$d \geq \max\{f(L), f(M-L)\}, \quad (23)$$

then the AoA estimation error is bounded for large M :

$$\delta_{p,k} \triangleq |\hat{\theta}_{p,k} - \theta_{p,k}| = \mathcal{O}\left(\sqrt{\log_2(M)}/M^{\frac{3}{2}}\right), \quad \forall p. \quad (24)$$

Based on Lemma 1, it is seen that when M is sufficiently large, as is the case for massive MIMO, the estimated AoA information is highly accurate. Also observe that the AoA estimation error meets the necessary condition for the unbiased channel gain estimate stated in Theorem 2. More importantly, this high resolution AoA information enables the near-ML estimation of unknown CSI, as seen late in Section IV-C.

B. Error of Estimated Channel Gains

We now analyze the estimation accuracy of the channel gain estimates, obtained by the post-reception beamforming using our highly accurate AoA estimates. Denote the error of the p th AoA of the k th UE as $\delta_{p,k} = |\hat{\theta}_{p,k} - \theta_{p,k}|$ and the variance of the pilot signal as σ_x^2 . After the post-reception beamforming, the noise term n_M in (46) of Appendix A has the variance

$$\bar{\sigma}_M^2 = \frac{1}{M} \exp\left(-2 \ln(2) M^2 \delta_{p,k}^2\right) \frac{\sigma_n^2}{B \sigma_x^2}. \quad (25)$$

Here, the approximate Gaussian beamformer is considered; see (49) in Appendix B. From (24), for large M , we have $\delta_{p,k} \rightarrow \mathcal{O}(\sqrt{\log_2(M)}/M^{1+\epsilon})$, with $\epsilon \rightarrow 1/2$. By applying the Taylor series expansion $\exp(x) = 1 + x + o(x^2)$, we have

$$\bar{\sigma}_M^2 = \frac{1}{M} \left(1 - 2 \ln(2) \mathcal{O}\left(\frac{\log_2(M)}{M^{2\epsilon}}\right)\right) \frac{\sigma_n^2}{B \sigma_x^2} \xrightarrow{M \text{ is large}} \frac{\sigma_n^2}{M B \sigma_x^2}. \quad (26)$$

Moreover, the channel gain estimation is the classical ML estimation (14), whose likelihood density is a complex Gaussian distribution with mean $\alpha_{p,k}$ and variance $\bar{\sigma}_M^2$, namely,

$$p\left(\alpha_{p,k} - \frac{1}{M} \mathbf{a}_M^H(\hat{\theta}_{p,k}) \mathbf{y}_k\right) \sim \mathcal{CN}(\alpha_{p,k}, \bar{\sigma}_M^2) \quad (27)$$

The following theorem shows that the estimates $\{\hat{\alpha}_{p,k}\}_{p=0}^{P-1}$ are unbiased with the estimation MSE bound $\sigma_n^2/(MB\sigma_x^2)$.

Theorem 3. *When $L = O(M)$ (e.g., $L = M/2$) and the minimal gap between AoAs satisfies (23), the channel gain estimated via the rank-1 subspace method is asymptotically unbiased for a large M , namely,*

$$\mathbb{E}[\hat{\alpha}_{p,k}] = \alpha_{p,k}, \text{ for large } M, \quad (28)$$

and the estimation MSE is bounded by $\sigma_n^2/(MB\sigma_x^2)$.

Proof. See Appendix C. $\square$

Note that, unlike the classical FFT method whose channel gain estimate is biased, our rank-1 subspace method now attains the unbiased channel gain estimate.

C. Error of Estimated Channel Vector

Based on the above error analysis on the estimates of the AoAs and the channel gains, in the angular domain, the residual error of the k th channel vector can be written as:

$$\begin{aligned} \mathbb{E}[\|\hat{\mathbf{h}}_k - \mathbf{h}_k\|_2^2] &= \mathbb{E}\left[\sum_{p=0}^{P-1} |\hat{h}_{p,k} - h_{p,k}|^2\right] \\ &\stackrel{\delta_{p,k} \rightarrow 0}{\approx} \mathbb{E}\left[\sum_{p=0}^{P-1} |\hat{\alpha}_{p,k} \cos(\delta_{p,k}) - \alpha_{p,k}|^2\right]. \end{aligned} \quad (29)$$

Recall from (24) that the MSE of AoA estimates of our new method is $O(\frac{\log_2 M}{M^3})$, while that of the ML method is $O(\frac{1}{M^3})$ [53]. We argue that, although the MSE gap between our method and ML is $O(\log_2 M)$, the near-ML CSI estimation will be attained in the high SNR region, given the shared dominate decaying rate of $O(\frac{1}{M^3})$.

To be specific, when the number of antennas is sufficiently large, e.g., $M \geq 100$, we are able to omit the error of AoA estimates in the high SNR region. That is to say, we will asymptotically have $\cos(\delta_{p,k}) \xrightarrow{M \text{ is large}} 1$. Further consider that the channel gains were estimated via the ML criterion [54], the CRLB of the channel gain estimation is achievable. As a result, the estimation error of the p th component in the k th channel vector is:

$$\mathbb{E}[\|\hat{h}_{p,k} - h_{p,k}\|_2^2] \stackrel{M \text{ is large}}{\approx} \mathbb{E}[|\hat{\alpha}_{p,k} - \alpha_{p,k}|^2] \geq \frac{\sigma_n^2}{MB\sigma_x^2}. \quad (30)$$

D. Gain Over Linear MMSE Estimator

Based on the signal model (17), the classical linear MMSE estimator would also obtain an accurate channel estimation. But it requires the MIMO channel covariance matrix $\mathbf{R}_H \in \mathbb{C}^{MK \times MK}$ and the channel AWGN covariance matrix $\mathbf{R}_N \in \mathbb{C}^{KB \times KB}$ to compute the channel estimate [54]

$$\text{vec}(\hat{\mathbf{H}}^{\text{mmse}}) = \mathbf{R}_H \tilde{\mathbf{X}}^H (\tilde{\mathbf{X}} \mathbf{R}_H \tilde{\mathbf{X}}^H + \mathbf{R}_N)^{-1} \text{vec}(\mathbf{Y}), \quad (31)$$

where $\tilde{\mathbf{X}} = \mathbf{X}^T \otimes \mathbf{I}$, and $\otimes$ is the Kronecker product. Note that, the acquisition of the MIMO channel covariance matrix and its inverse are very costly. So, another simpler LS channel estimator computes the MIMO channel matrix estimate as

$$\hat{\mathbf{H}}^{\text{ls}} = \mathbf{Y} \mathbf{X}^\dagger. \quad (32)$$

Yet it has a larger estimation error than the linear MMSE estimator.

Note that the linear MMSE estimator should not outperform the ML method whose CRLB of the estimation error is given in [55], [56]. Therefore, using this CRLB, we have

$$\mathbb{E}\left[\|\hat{\mathbf{h}}_{p,k}^{\text{mmse}} - \mathbf{h}_{p,k}\|_2^2\right] \geq \frac{\sigma_n^2}{B\sigma_x^2 + \rho_h^2 \sigma_n^2}, \quad (33)$$

where $\mathbb{E}\left[\frac{\partial \ln p_h(\mathbf{h})}{\partial \mathbf{h}^*} \frac{\partial^H \ln p_h(\mathbf{h})}{\partial \mathbf{h}^*}\right] \triangleq \rho_h^2 \mathbf{I}_M$; $\rho_h^2 = P + 1$ and $p_h(\mathbf{h})$ is the prior distribution of the channel vector $\mathbf{h}_k$.

Comparing (30) with (33), it can be seen that our rank-1 subspace estimator is capable of attaining an higher estimation accuracy than the linear MMSE estimator.

Theorem 4. *Define the normalized MSE of the channel estimation as $\gamma = \mathbb{E}[\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2 / \|\mathbf{H}\|_F^2]$. In order to achieve a target normalized MSE γ (e.g., $\gamma < 10^{-2}$), the required SNR for our rank-1 subspace method is $\text{snr}_1 \triangleq 10 \log_{10}(E_1/\sigma_n^2)$, while the required SNR for the linear MMSE estimator is $\text{snr}_2 \triangleq 10 \log_{10}(E_2/\sigma_n^2)$, where E_1 and E_2 are the respective transmission powers of the two schemes. When M is large and the stack length is $L = O(M)$, we have:*

$$\text{snr}_2 - \text{snr}_1 = O(\log_{10}(M)). \quad (34)$$

The proof of Theorem 4 is straightforward. From (31), we have $\text{snr}_1 \triangleq 10 \log_{10}(E_1/\sigma_n^2) = 10 \log_{10}(\frac{\sigma_x^2}{MB\sigma_x^2})$. While for the linear MMSE method, we have $\text{snr}_2 \triangleq 10 \log_{10}(E_2/\sigma_n^2) = 10 \log_{10}(\frac{\sigma_n^2}{B\sigma_x^2 + \rho_h^2 \sigma_n^2})$. Thus, we obtain $\text{snr}_2 - \text{snr}_1 = 10 \log_{10}\left[\frac{M}{1 + \rho_h^2/B \times \sigma_n^2/\sigma_x^2}\right] = O(\log_{10}(M))$, when σ_n^2/σ_x^2 is small.

Remark 1. *Theoretical analysis shows that our new estimator is capable of acquiring the high-resolution AoA for large M and $L \sim O(M)$. When M is small, this result may be invalid. Later in the numerical simulation, it will be shown that the above theoretical result holds when $M > M_0$ with $M_0 = 64$. Although our rank-1 subspace estimator achieves excellent estimation performance for massive MIMO systems, its computational complexity on the order of $O(KM^3)$ still makes it impractical. Note that the same problem is also applied to the linear MMSE estimator², which has the complexity of $O(M^3)$. It is critical to achieve a low-complexity deployment for massive MIMO system, while maintaining the high performance.*

V. FAST RANK-1 SUBSPACE ESTIMATOR

In this section, we design a fast implementation of our rank-1 subspace method. From Subsection III-C, the complexity of this rank-1 subspace estimator mainly comes from the SVD of the Hankel matrix $\tilde{\mathbf{Y}}$. Without loss of generality, in the sequel we assume $L = M/2$. Therefore, from (10), the space-embedding Hankel matrix $\tilde{\mathbf{Y}} \in \mathbb{C}^{L \times L}$ becomes a symmetric positive semi-definite (SPSD) matrix, i.e., $\tilde{\mathbf{Y}} = \tilde{\mathbf{Y}}^T$. As discussed previously, the rank of this SPSP matrix $\tilde{\mathbf{Y}}$ is

²For the correlated MIMO channel matrix, the computational complexity of the linear MMSE estimator becomes $O((MB)^3)$ [17], [18].

restricted. By exploiting this low-rank property, we resort to a randomized low-rank approximation technique that is widely applied in linear algebra and scientific computing [18], [50]–[52], [57], [58], to construct a fast rank-1 subspace CSI estimator without impeding the estimation accuracy. It is noteworthy that, although this fast method may substantially reduce the computational complexity in the case of low-rank channel matrix, it is not an essential component of our two-stage CSI estimator. For example, for the full-rank channel matrix, there is no need to apply it.

A. Randomized Low-rank Approximation

In order to minimize the overall computational complexity, we focus on *approximate* rather than *exact* computation of the SVD of $\bar{\mathbf{Y}} \in \mathbb{C}^{L \times L}$. To achieve this goal, we first approximate $\bar{\mathbf{Y}}$ with a special structure [32]:

$$\bar{\mathbf{Y}} \approx \mathbf{C}\mathbf{W}\mathbf{C}^H, \quad (35)$$

where the small matrix sketch $\mathbf{C} \in \mathbb{C}^{L \times s}$ is abstracted from $\bar{\mathbf{Y}}$ via random sampling³, while the small matrix $\mathbf{W} \in \mathbb{C}^{s \times s}$ is computed based on the chosen sketch $\mathbf{C}$.

To abstract $\mathbf{C}$ from $\bar{\mathbf{Y}}$ via random sampling, we specify the sampling length s that satisfy $\text{rank}(\bar{\mathbf{Y}}) \leq s \ll M$, and randomly pick up the s columns of $\bar{\mathbf{Y}}$ according to the uniform distribution $\mathcal{U}\{1, 2, \dots, L\}$. Denoting the sampled column indexes as the set $\mathcal{I} \triangleq \{i_1, i_2, \dots, i_s\}$ with $|\mathcal{I}| = s$, we have

$$\mathbf{C} = \bar{\mathbf{Y}}(:, \mathcal{I}) = \bar{\mathbf{Y}}\mathbf{\Pi}, \quad (36)$$

where $\mathbf{\Pi} \in \mathbb{R}^{L \times s}$ is the uniform sampling matrix, which selects the s columns of $\frac{\sqrt{L}}{\sqrt{s}}\mathbf{I}_L$ with the column index set $\mathcal{I}$, namely, $\mathbf{\Pi} = \frac{\sqrt{L}}{\sqrt{s}}\mathbf{I}_L(:, \mathcal{I})$.

Given the small sketch $\mathbf{C} \in \mathbb{C}^{L \times s}$, the weight matrix $\mathbf{W} \in \mathbb{C}^{s \times s}$ is determined, by minimizing the approximation error

$$\mathbf{W} = \arg \min_{\mathbf{W}' \in \mathbb{C}^{s \times s}} \|\bar{\mathbf{Y}} - \mathbf{C}\mathbf{W}'\mathbf{C}^H\|_F^2 = \mathbf{C}^\dagger \bar{\mathbf{Y}} (\mathbf{C}^\dagger)^H. \quad (37)$$

Considering that the above exact solution requires the computational complexity of $\mathcal{O}(2s^2M + sM^2)$, we choose to solve it approximately. Specifically, since the optimization problem (37) is *over-determined*, $\mathbf{W}$ can be derived in a more efficient way, by further sketching both $\bar{\mathbf{Y}}$ and $\mathbf{C}\mathbf{W}\mathbf{C}^H$, as is done by the Nystrom method [32], [52], [59]. Hence, we have

$$\begin{aligned} \mathbf{W} &= \arg \min_{\mathbf{W}' \in \mathbb{C}^{s \times s}} \|\mathbf{\Pi}^T (\bar{\mathbf{Y}} - \mathbf{C}\mathbf{W}'\mathbf{C}^H) \mathbf{\Pi}\|_F^2, \\ &= (\mathbf{\Pi}^T \bar{\mathbf{Y}} \mathbf{\Pi})^\dagger = (\bar{\mathbf{Y}}(\mathcal{I}, \mathcal{I}))^\dagger. \end{aligned} \quad (38)$$

With $\mathbf{C}$ and $\mathbf{W}$, we obtain the randomized low-rank approximation $\bar{\mathbf{Y}}$ as in (35). To compute its SVD, we first obtain the SVD of $\mathbf{C}$ as $\mathbf{C} = \mathbf{U}_c \Sigma_c \mathbf{V}_c^H$, and then approximate $\bar{\mathbf{Y}}$ by:

$$\bar{\mathbf{Y}} \approx \mathbf{U}_c \underbrace{\Sigma_c \mathbf{V}_c^H \mathbf{W} \mathbf{V}_c \Sigma_c^H}_{\triangleq \mathbf{B}} \mathbf{U}_c^H = \mathbf{U}_c \mathbf{B} \mathbf{U}_c^H. \quad (39)$$

³Note that, in the general case ($L \neq M/2$), the Hankel matrix $\bar{\mathbf{Y}}$, constructed from the received signal vector $\mathbf{y}$, is not symmetric. For simplicity, here we assume $L = M/2$.

TABLE I
COMPUTATIONAL COMPLEXITY OF VARIOUS CSI ESTIMATORS.

Estimator	Computational complexity
LS [1]	$\mathcal{O}(KBM + KB^2)$
SVP [20] ¹	$\mathcal{O}(t_{\text{svp}} r_{\text{est}}(K^2 + K)M + 2K^2)$
FFT [34]	$\mathcal{O}(K(B + \log_2(M))M + KP^2)$
WPEACH [17] ²	$\mathcal{O}(N_l(BK^2 + B^2K)M^3 + N_l^3)$
Linear MMSE [11] ³	$\mathcal{O}(B(2K^2 + 2BK + B^2)M^3)$
LSE-SMP [29] ⁴	$\mathcal{O}(t_{\text{lse-smp}}K(B^2 + K^2)M^3)$
Random-MMSE [18] ⁵	$\mathcal{O}(B_1^3M^3 + s_r^3B_2^2 + s_r^2s_c)$
Proposed ⁶	$\mathcal{O}(KM^3 + KNPM)$
Proposed fast ⁶	$\mathcal{O}(K(P^2 + NP)M)$

1. t_{svp} : the number of iterations, and r_{est} : the order of low rank.

2. N_l : truncation order of Taylor series expansion.

3. $\mathcal{O}(M^3)$ for uncorrelated MIMO channels.

4. $t_{\text{lse-smp}}$: the number of iterations.

5. B_1/B_2 : the pilot lengths, and s_r/s_c : the sampling lengths.

6. Complexity $\mathcal{O}(KNPM)$ for matrix multiplications has fast parallel acceleration implementation [49].

As $\mathbf{B}$ is a Hermitian matrix, i.e., $\mathbf{B} = \mathbf{B}^H$, it has the SVD of $\mathbf{B} = \mathbf{U}_B \Sigma_B \mathbf{U}_B^H$. Substituting the SVD of $\mathbf{B}$ into (39) leads to the approximate rank-restricted SVD of $\bar{\mathbf{Y}}$ as

$$\bar{\mathbf{Y}} \approx \mathbf{U}_c \mathbf{U}_B \Sigma_B \mathbf{U}_B^T \mathbf{U}_c^H = \tilde{\mathbf{U}}_P \Sigma_B \tilde{\mathbf{U}}_P^H, \quad (40)$$

where the unitary matrix $\tilde{\mathbf{U}}_P \in \mathbb{C}^{L \times s}$, given by

$$\tilde{\mathbf{U}}_P \triangleq \mathbf{U}_c \mathbf{U}_B, \quad (41)$$

defines approximately the signal subspace. On this basis, we can approximate the pseudo-spectrum (12) by

$$\tilde{P}(\theta) = \frac{1}{\mathbf{a}_L^H(\theta)(\mathbf{I}_L - \tilde{\mathbf{U}}_P \tilde{\mathbf{U}}_P^H) \mathbf{a}_L(\theta)}, \quad (42)$$

and search for its P highest peaks on the N gridded angles $\{-\pi + n\Delta\theta, n = 0, 1, \dots, N-1\}$. In this stage, other subspace-based methods, e.g., ESPRIT, can also be applied to compute the above pseudo-spectrum and thus acquire unknown AoAs. When the exact number of P is known as *a priori*, highly accurate AoAs can be attained with ESPRIT. While this prior information is not known, the above MUSIC method can still acquire the accurate result.

B. Complexity Analysis

The computational complexity of our fast rank-1 subspace estimator for the single UE involves three parts: 1) compute the SVD of $\bar{\mathbf{Y}}$ approximately, which has the complexity of $\mathcal{O}(s^2M + s^3)$, 2) evaluate the spatial pseudo-spectrum, which has the complexity of $\mathcal{O}(NLP)$, and 3) estimate the channel gains, with the complexity of $\mathcal{O}(M)$. Since both the number of users K and the number of paths P are very small ($\ll M$) and further considering that $L = \frac{M}{2}$ and in practice the number of sampled columns $s = \mathcal{O}(P)$, the overall computational complexity of our fast method for the K -user case is

$$C_{\text{fast rank-1}} = \mathcal{O}(K(P^2 + NP)M). \quad (43)$$

Hence, the complexity of our fast method scales linearly with M . This is two orders of magnitude lower than that of the linear MMSE method, which has the complexity of $\mathcal{O}(M^3)$. This is because in practice we usually have $P < 6, M > 64, K \ll M$ in the multi-users massive MIMO scenario [43], and hence $K(P^2 + NP)M \ll M^3$.

Table I compares the computational complexity of the following massive MIMO channel estimators: the linear LS [1], the SVP [20], the FFT [34], the WPEACH [17], the linear MMSE [11], the LSE-SMP [29], the random-MMSE [18], our rank-1 subspace estimator, and our fast rank-1 subspace estimator. We will show that our low-complexity fast rank-1 subspace estimator is capable of outperforming the linear MMSE estimator through both theoretical analysis and numerical simulation.

C. Theoretical Analysis on Approximated Pseudo-spectrum

We first give the error bound on the approximate pseudo-spectrum $\tilde{P}(\theta)$ (42) computed based on the subspace $\tilde{U}_P$. To be self-contained, here we provide the following Theorem 5 to demonstrate the accuracy of the fast method, while the proof of Theorem 5 can be found in our previous work [32].

Theorem 5. Give $C = \bar{Y}\Pi$ and $W = (\bar{Y}(\mathcal{I}, \mathcal{I}))^\dagger$, where $\Pi \in \mathbb{R}^{L \times s}$ is a uniform sampling matrix. Let the signal subspace $\tilde{U}_P$ be computed by (41) and the pseudo-spectrum be approximated by (42). Define the coherence of $\tilde{U}_P$ as $\mu(\tilde{U}_P) \triangleq \frac{L}{P} \max_{1 \leq m \leq L} \|\tilde{U}_P(m, :)\|_2^2 \in [1, \frac{L}{P}]$ and give a constant $\delta \in (0, 1)$. When the sampling length s satisfies

$$s \geq 4.5P \log_2 \left(\frac{P}{\delta} \right) \mu(\tilde{U}_P), \quad (44)$$

the following relation holds with probability at least $1 - \delta$

$$\sqrt{P(\theta)} \leq \left(1 + \frac{L}{\sqrt{s}} \frac{\sigma_{P+1}(\bar{Y})}{\sigma_P(\bar{Y})} \right) \sqrt{\tilde{P}(\theta)}. \quad (45)$$

Theorem 5 gives the lower error bound on the approximate pseudo-spectrum. When the SNR is high, the spectrum gap $\frac{\sigma_{P+1}(\bar{Y})}{\sigma_P(\bar{Y})}$ is very small, i.e., $\frac{\sigma_{P+1}(\bar{Y})}{\sigma_P(\bar{Y})} \rightarrow 0$, and the approximation error becomes ignorable.

Remark 2. Based on the above lower bound, we can show that the approximated pseudo-spectrum should not miss the true AoAs. On one hand, the exact pseudo-spectrum peaks at a true AoA and is significantly larger than the noise baseline at the non-AoA region. Specifically, $P(\theta_p)/P(\theta_{null}) \sim L$, where $\theta_{null} \in [0, \pi]/\{\theta_p\}_{p=1}^P$. If a single AoA was missed in the approximated pseudo-spectrum, we expect to have $\tilde{P}(\theta_p) \sim P(\theta_{null})$. On the other hand, Theorem 5 suggests that $P(\theta_p)/\tilde{P}(\theta_p) \leq (1 + \epsilon)^2$, where $\epsilon \triangleq \frac{L}{\sqrt{s}} \frac{\sigma_{P+1}(\bar{Y})}{\sigma_P(\bar{Y})}$. Due to the low-rank property of Hankel matrix $\bar{Y}$, in the high SNR region, we should have $\epsilon < 1$. Combining the both arguments results in the contradictory result of $(1 + \epsilon)^2 \geq L$. This indicates that $\tilde{P}(\theta_p) \sim P(\theta_{null})$ should be untrue, in the case of large M . Thus, the approximated pseudo-spectrum should accurately capture the real AoAs.

Hence, like the exact pseudo-spectrum (12), the approximated pseudo-spectrum (42) can be also used to acquire the accurately AoA information. Furthermore, the rest of the fast rank-1 subspace estimator are identical to those of the rank-1 subspace estimator. Therefore, we may conclude that the fast rank-1 subspace estimator should not degrade the achievable performance of the rank-1 subspace estimator.

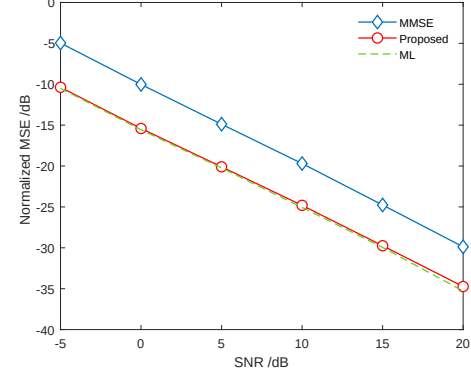


Fig. 1. Normalized MSE performance of our rank-1 subspace estimator and the linear MMSE estimator as functions of SNR for the full-rank, sparse and uncorrelated channel, given $M = 128$, $K = 40$, $B = 2K$ and $L = M/2$.

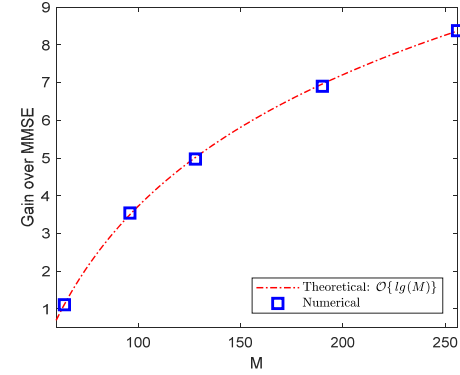


Fig. 2. SNR gain of our rank-1 subspace method over the classical linear MMSE estimator as function of M for the full-rank sparse channel of Fig. 1, with $K = 40$, $B = 2K$ and $L = M/2$.

VI. NUMERICAL ANALYSIS

In this section, we compare the performance and complexity of various massive MIMO channel estimators based on numerical simulations using several evaluation metrics, including the normalized mean square error (NMSE), the averaged CPU runtime, and the channel capacity.

A. Full-rank and Sparse Channel

We first compare our rank-1 subspace estimator with the classical linear MMSE method, in the case of full-rank and sparse channel matrix, i.e., $\text{rank}(\mathbf{H}) = \min\{M, K\} = K$ and $P \ll M$. That means, the channels are uncorrelated. In the simulation, we first randomly generate P independent AoAs of each channel vector and P i.i.d. Rayleigh fading gains. With the assumption of ULA, these spatial channel parameters can be transformed to the temporal domain, and then the channel matrix is obtained. As discussed, the linear MMSE method requires the channel covariance $\mathbf{R}_H$. We assume that such *a priori* statistical information is available, thus leading to the *genie-aided* linear MMSE estimator [60]. This may be impractical as acquiring an accurate estimation of the channel covariance requires considerable time resources. By contrast, our method excludes such an impractical prior knowledge.

Fig. 1 compares the NMSE performance of our rank-1 subspace estimator with that of the classical linear MMSE estimator and the ML estimator, for the massive MIMO system with the number of antennas $M = 128$, the number of users $K = 40$ and the length of pilot signal $B = 2K$. The number of paths is $P = 5 \sim 7$ for each channel vector, and we

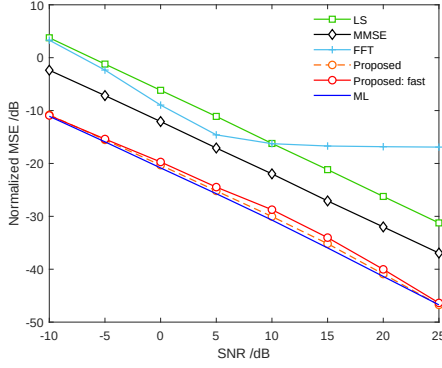


Fig. 3. Normalized MSE performance of five channel estimators as functions of SNR for the full-rank, sparse and uncorrelated channel, given $M = 256$, $K = 40$, $B = 2K$ and $L = M/2$.

set the stack length to $L = M/2$ for our estimator. Observe from Fig. 1 that our new method indeed approximates the ML estimator, and dramatically outperforms the linear MMSE estimator. For example, at $\text{SNR} = 20$ dB, the NMSE of our rank-1 subspace estimator is more than 6 dB lower than that of the linear MMSE method. This significant gain is mainly attributed to two key factors: 1) the highly accurate AoA information acquired by the rank-1 subspace method, and 2) the largely improved SNR by the *post-reception* beamforming.

Fig. 2 depicts the SNR gain achieved by our new method over the linear MMSE estimator under different M . First, this SNR gain increases with M , since the SNR after the post-reception beamforming increases with M . Second, the numerical results show that the attained gain over the linear MMSE method indeed shows a *logarithmic* increasing trend with M , which validates the theoretical result of Theorem 4. Therefore, by applying our new estimator, the UEs far away from the BS can be admitted to establish reliable communication links, and the coverage area of massive MIMO system can be effectively extended.

Next we use the same full-rank channel model with $M = 256$, $L = 128$, $K = 40$, $B = 2K$ and $P = 7$, to compare the NMSE performance of several state-of-the-art channel estimators in Fig. 3. Again, the propagation channels of K UEs are uncorrelated. In this case, many low-rank based channel estimators, including the SVP and WPEACH methods, are inapplicable. As expected, the LS estimator and the FFT-based estimator acquire less accurate CSI than the linear MMSE method. As a classical angular-domain estimator [34], the FFT-based method also acquire AoAs first. However, due to its limited resolution and biased channel gain estimation, it exhibits an error floor at the high SNR region, as can be clearly seen in Fig. 3. Observe that our fast rank-1 subspace estimator achieves the same NMSE performance as the rank-1 subspace estimator. The numerical result of Fig. 3 thus validates the theoretical analysis of Subsection V-C, namely, the approximation of the fast rank-1 subspace implementation does not degrade the achievable performance. Hence, the fast rank-1 subspace estimator significantly outperforms the linear MMSE estimator, in terms of channel estimation accuracy. Note that for the fast rank-1 subspace method, we set the sampling length to $s = \lceil 1.5 \times P \rceil$, i.e., $s \sim \mathcal{O}(P)$, which is

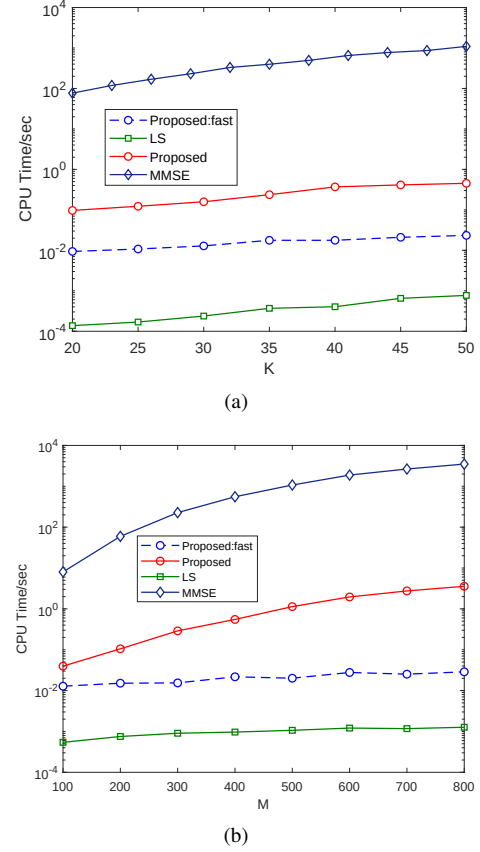


Fig. 4. Time complexity of four channel estimators for the full-rank sparse channel of Fig. 3: (a) as functions of K , given $M = 256$, $B = 2K$ and $L = M/2$, and (b) as functions of M , given $K = 40$, $B = 2K$ and $L = M/2$.

less strict than the requirement of (44); recall that (44) is a sufficient condition. With this setting on s , it is sufficient to attain a highly accurate AoA information.

To evaluate the computational complexity of various channel estimators, except for the comparative results in Table 1, we also provide the CPU run-time to measure time complexity or processing delay. The operating system runs at 2.7 GHz basic frequency with RAM 32 GB. For the same full-rank system of Fig. 3, Fig. 4(a) compares the time complexity of four channel estimators by varying the number of users K . As expected, the LS estimator imposes the lowest computational complexity, which is its major advantage in practical applications. Observe that our fast rank-1 subspace method incurs a lower complexity compared to the linear MMSE estimator. This is remarkable, considering the fact that it significantly outperforms the linear MMSE method, in terms of channel estimation accuracy. Fig. 4(b) shows the time complexity of the four channel estimators under different M . As shown, the complexity of our fast method increases linearly with M . This verifies the complexity analysis of Subsection IV-B. It can also be seen that for $M > 200$, the complexity of our fast estimator is dramatically lower than the linear MMSE estimator. In particular, for $M = 800$, its complexity is around 10^6 times lower than the linear MMSE method. This agrees with the analysis of Subsection IV-B, which indicates that for massive MIMO with large M , the complexity of our fast method is on

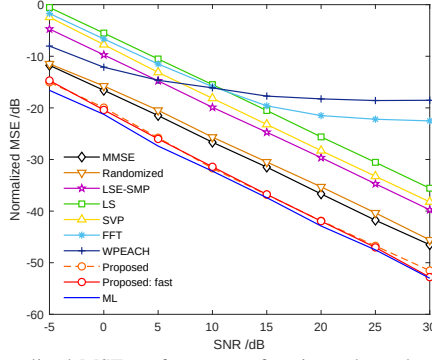


Fig. 5. Normalized MSE performance of various channel estimator for the low-rank, sparse and correlated channel, given $\text{rank}(\mathbf{H}) = 20$, $M = 256$, $K = 40$, $L = M/2$, and $B = K$.

two orders of magnitude lower than that of the linear MMSE method.

B. Low-rank and Sparse Channel

We further consider the case of low-rank and sparse massive MIMO, i.e., $\text{rank}(\mathbf{H}) < K$ and $P \ll M$. As reported in [19], multiple users may experience the similar propagation environments, and hence the channel involves correlated paths. In this case, the channel matrix is characterized by both low rank and sparsity. Many channel estimators were developed to exploit these two features [20], [34]. In the simulation, we firstly generate a set of random AoAs and fading gains, while assuming multiple channel vectors of clustered users share the same set of spatial parameters. After the spatial channel parameters are transformed to the temporal domain, the low-rank and sparse channel matrix can be obtained. We assume $P = 5 \sim 7$, $M = 256$, $K = 40$, $\text{rank}(\mathbf{H}) = 20$, $L = M/2$, and $B = 2K$. For the fast method, the sampling length is $s = 20$.

Fig. 5 compares the normalized MSE performance of various channel estimators. Similar to the FFT-based method, the WPEACH also exhibits a high error floor at the high SNR region. This is due to the approximation based on a limited expansion order (e.g., $3 \sim 5$). The SVP achieves 3dB gain over the LS estimator. The sparsity-based estimator, LSE-SMP, improves the NMSE by 2dB over the SVP, but its NMSE is still 7dB higher than the linear MMSE estimator. The randomized-MMSE [18], which relies on randomized matrix approximation to reduce the complexity, is capable of achieving a comparable accuracy as the linear MMSE estimator. Observe that our rank-1 subspace method and its fast version attain the similar performance, both of which approximate the ML estimator and again outperform the linear MMSE estimator considerably.

Fig. 6 depicts the complexity of these channel estimators as the functions of M . The LSE-SMP requires the highest computational complexity, as it involves iterative computation of large matrices. The high complexity of the WPEACH can be attributed to the complex computation of weight coefficients required in approximating inverse via matrix series expansion. Observe that in correlated MIMO, the complexity of the linear MMSE method also increases considerably. By separately acquiring two small sub-matrices [18], the randomized-MMSE

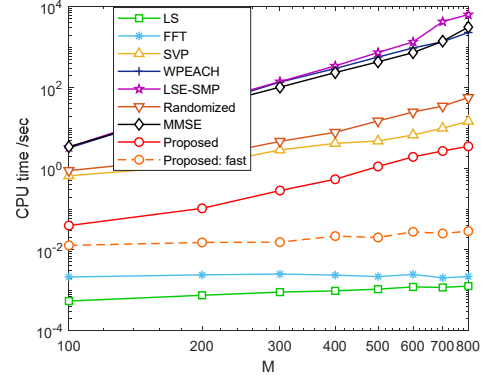


Fig. 6. Time complexity of various channel estimators as functions of M for the low-rank sparse channel of Fig. 5, given $M = 256$, $K = 40$, $L = M/2$, and $B = K$.

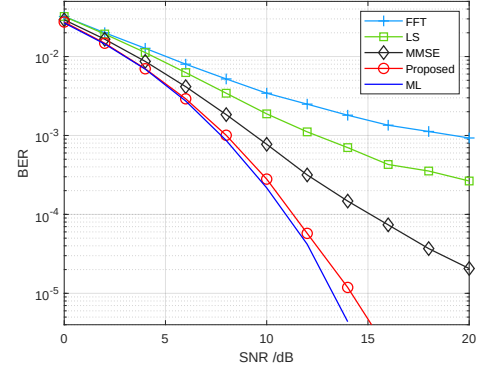


Fig. 7. BER performance of various channel estimators in the 5G NR scenario; $M = 256$, $K = 40$ and $P = 7$; $\text{rank}(\mathbf{H}) = K$; $B = 80$; QPSK.

has significantly lower complexity than the linear MMSE method, but its computational complexity is still considerably higher than our rank-1 subspace method. The LS and FFT methods offer the lowest complexity, but their estimation performance are much poorer compared with the linear MMSE method. Most noticeably, our fast rank-1 subspace method also requires a very low computational complexity. Specifically, at $M = 300$, its time complexity is $10000\times$ faster than the linear MMSE method. Among all the existing massive MIMO channel estimators, our new method is the first one that not only significantly outperforms the popular linear MMSE estimator, but also imposes a dramatically lower complexity than the latter.

C. BER Performance

We further evaluate the bit error rate (BER) performance of various CSI estimators, in the context of 5G NR configurations. To be specific, we consider the sparse uncorrelated channel matrix. The detailed simulation parameters are given as follows: $M = 256$, $K = 40$, $B = 2K$ and $L = M/2$. The QPSK modulation is assumed, and the deterministic pilot sequence (i.e., DM-RS) of length $B = 80$ is used. In practice, our new estimator can be applied to both random and deterministic pilot sequences. In the simulation, totally 1.6×10^6 data symbols are used ($N_{\text{frame}} = 10$, $N_{\text{packet}} = 2000$); the BER curve is then plotted based on the Monte-Carlo method. From Fig. 7, we can see that our new CSI estimator

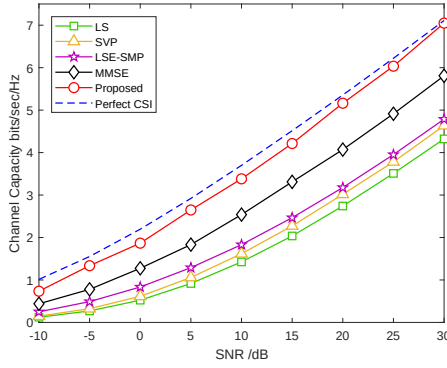


Fig. 8. Channel capacity achieved by various channel estimators for the MIMO system with $M = 256$, $K = 40$ and $P = 4$.

attains the near-ML BER performance, while outperforming the classical CSI estimators (e.g., LS and linear MMSE).

D. Channel Capacity and Spectral Efficiency

We then evaluate the channel capacity of the massive MIMO system, assuming $M = 128$, $K = 40$ and $P = 7$. For simplicity, we directly present the calculated results of the sum rate. One can refer to [61] for the detailed equation. Fig. 8 depicts the channel capacity achieved by various channel estimators, in comparison with the channel capacity of the perfect CSI. It can be seen from Fig. 8 that the channel capacity is significantly enhanced by our rank-1 subspace estimator, compared to the other methods. For example, at $\text{SNR} = 30$ dB, the achieved channel capacity of the linear MMSE estimator is 5.8 bits/sec/Hz, while our new method improves the channel capacity to 7.1 bits/sec/Hz, which asymptotically closes the *capacity gap* between the estimated CSI and the perfect CSI (where the estimation MSE is zero).

Furthermore, we evaluate the spectral efficiency of various CSI estimators in massive MIMO communications. In the simulation, we set $M = 256$, $K = 40$, $P = 5$, $B = 2K$ and $\text{rank}\{\mathbf{H}\} = K$. Note that, with the finite number of antennas ($M = 256$), we may assume that the inter-cell interference (ICI) from nearby cells can be reasonably ignored, as the pilot contamination in this case can be negligible. Thus, we focus on the popular regularized zero-forcing (RZF) precoding scheme [62]. On this basis, the signal-to-noise-plus-interference ratio (SINR) can be calculated. Finally, we obtain the instantaneous spectral efficiency of various CSI estimators. From Fig. 9, we observe that our proposed method achieves comparable spectral efficiency as the ML estimator, while the others fail to do so.

E. Complexity versus Accuracy Region

Next we jointly evaluate the two aspects of channel estimation in massive MIMO communications, specifically, complexity vs accuracy. The first performance metric is related to the implementation cost, such as time or energy; while the second metric corresponds to the benefit, i.e., estimation accuracy or channel capacity. Traditionally, these two conflicting aspects have to be balanced in massive MIMO systems, as done by different channel estimators. To achieve the high accuracy in

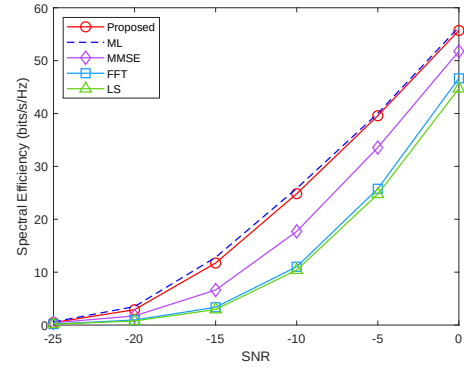


Fig. 9. Spectral efficiency of various channel estimators as functions of SNR; $M = 256$, $K = 40$, $L = M/2$, and $B = 2K$. The regularized zero-forcing (RZF) precoding scheme is adopted.

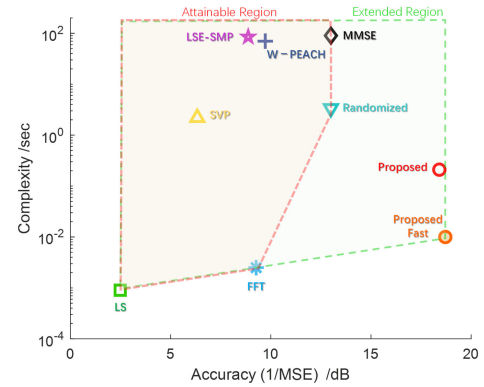


Fig. 10. Complexity vs accuracy in the CSI acquisition of massive MIMO communications, given $M = 256$, $K = 40$ and $\text{SNR} = 20$ dB.

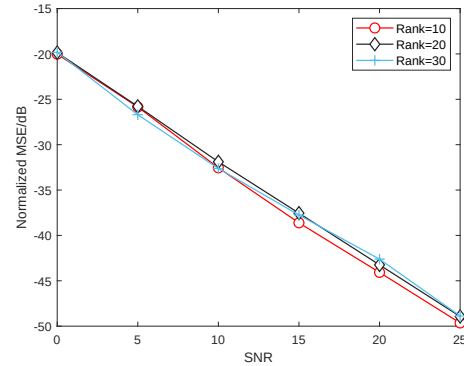


Fig. 11. Performance of the proposed CSI estimator under different rank values of channel matrices; $M = 256$, $K = 40$ and $B = 2K$.

the classical estimation sense, the complexity may become impractical, as in the case of the linear MMSE estimator. To reduce the complexity to a minimum, the estimation accuracy and hence the channel capacity has to be seriously sacrificed, as in the case of the LS estimator. Within the classical paradigm of trading such two aspects [17], [20], [29], [34], the attainable complexity-accuracy region in the CSI acquisition is largely limited, as illustrated in the attainable region of Fig. 10.

As a key breakthrough, our rank-1 subspace estimator surprisingly extends this complexity-accuracy region considerably, as seen in Fig. 10. Unlike the CRLB of the classical

linear MMSE method whose MSE is proportional to the noise variance, our theoretical analysis shows that the attainable MSE of our new estimator is further scaled by $\frac{1}{M}$ when M is large (e.g. $M > 64$). Hence, the estimation accuracy of our rank-1 subspace estimator is far beyond the limit of the classical linear MMSE estimator, and it substantially enhances the channel capacity and the coverage area. Moreover, its fast version, which does not degrade the estimation accuracy, has the complexity scaled linearly with M . This new massive MIMO channel estimator therefore offers the great promise to the emerging 5G-Advanced and 6G communications.

F. Practical Considerations

1) *Effects of Channel Correlation Factor*: In the above, we have evaluated the performance of our proposed method in both full-rank and low-rank channels. As seen, our method is independent of the rank value of a channel matrix. It was shown in [63] that the channel correlation factor, or the spatial correlation matrices at the BS, affects the low rank property of massive MIMO channels. In principle, the higher the channel correlation factor, the lower the channel rank value. Although a unified relationship between the channel correlation factor and the channel rank value is hard to describe, we may alternatively evaluate the estimators' performance under different rank values, corresponding to various channel correlation factors. In Fig. 11, we present the NMSE performance with different rank values. From the averaged NMSE performance obtained with a large number of random channel realizations, we conclude that the proposed method accurately acquires unknown channel matrix.

2) *Effects of Sector-Level Beamforming*: In practice, the BS may adopt the sector-level beamforming to enhance the system's performance. Combined with classical CSI estimators (e.g., LS), the sector-level beamforming may further improve the accuracy of channel estimation. Our proposed method can also be applied in this scenario. Nevertheless, the improvement on the estimation accuracy may be marginal. This is mainly due to the high-accuracy estimation of AoAs, as well as the subsequent post-reception beamforming in acquiring unknown fading gains based on the ML criterion. Hence, for our two-stage CSI estimator, the MSE performance may be comparable with the case of using omnidirectional antenna.

3) *2D Uniform Rectangular Array (URA) Scenario*: Furthermore, we demonstrate that the proposed channel estimator can be directly applied to the 2D URA based massive MIMO to accomplish a highly accurate channel estimation. Here, the only difference is that the received signal becomes a two-dimensional matrix. From Fig. 12, it is seen that our new estimator attains significant performance gain over the classical LS and FFT estimators.

4) *Another Potential Solution*: Lastly, we note that another popular scheme, i.e., approximate message passing (AMP), shares the similar linear inverse problem formulation. In this respect, the orthogonal AMP (OAMP) [64], or recently developed memory AMP (MAMP) [65], may also have the potential to accomplish the Bayesian optimal estimation. It is interesting to consider the application of such AMP methods

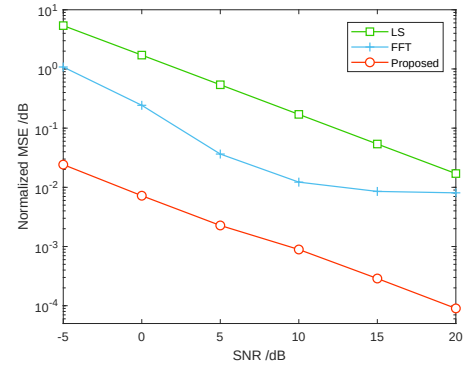


Fig. 12. Normalized MSE performance of three channel estimators for 2D URA based MIMO, given $M = 32 \times 32$, $K = 1$, $B = 24$ and $P = 5$.

to realize the high-accuracy CSI estimation in future research.

VII. CONCLUSIONS

For massive MIMO systems, the existing channel estimators do not trade off estimation accuracy and computational complexity well, and the attainable accuracy-complexity region is fundamentally limited. In this work, we have presented a novel rank-1 subspace channel estimator, which could approximate the ML method while avoiding the computationally exhaustive search of high-dimensional space. Our new estimator first acquires the high-resolution AoA estimate, and then attains the unbiased and highly accurate channel gains by the post-reception beamforming. Given the number of antennas M , our theoretical analysis has shown that this new method substantially outperforms the classical linear MMSE estimator, attaining the extra gain that increases with $\log_{10}(M)$. We have also derived a fast implementation of this rank-1 subspace estimator by leveraging the low-rank property of the constructed Hankel matrix. This fast version does not degrade the achievable performance, and yet it reduces the computational complexity to the order of $\mathcal{O}(M)$, which is orders of magnitude lower than the linear MMSE estimator. Numerical simulations have validated our theoretical analysis. Our new estimator has thus resolved a long-standing dilemma in massive MIMO signal processing, namely, accuracy vs complexity. It is the first estimator that goes far beyond the estimation accuracy limit set by the linear MMSE estimator, and yet only imposes the complexity in the order of the LS estimator. We believe it has great potential in the 5G-Advanced and 6G communications.

APPENDIX

A. Proof of Theorem 1

Proof. Substituting $\hat{\theta}_p$ by θ_p in (5) as well as noting the facts of $\mathbf{a}_M^H(\theta_p)\mathbf{a}_M(\theta_p) = M$ and $\mathbf{a}_M^H(\theta_p)\mathbf{a}_M(\theta_{p'}) \rightarrow 0$ as $M \rightarrow \infty$ for $p \neq p'$, we immediately have

$$\begin{aligned} \hat{\alpha}_p &= \frac{1}{M} \mathbf{a}_M^H(\theta_p) \left(\sum_{p'=0}^{P-1} \alpha_{p'} \mathbf{a}_M(\theta_{p'}) \right) + \frac{1}{M} \mathbf{a}_M^H(\theta_p) \mathbf{n}, \\ &= \alpha_p + n_M, \end{aligned} \quad (46)$$

where $\mathbf{n} \triangleq \mathbf{N}\mathbf{x} \in \mathbb{C}^{M \times 1}$ with elements having the variance $\sigma_n^2/B\sigma_x^2$, and $n_M \triangleq \mathbf{a}_M^H(\theta_p)\mathbf{n}/M$ which is a zero-mean AWGN with the variance $\text{var}[n_M] = \sigma_n^2/MB\sigma_x^2$. Therefore, the estimated channel gains using (5) are unbiased, i.e.,

$$\mathbb{E}[\hat{\alpha}_p] = \alpha_p, \quad 0 \leq p \leq P-1. \quad (47)$$

With the linear Gaussian model of (46), the CRLB of the above channel estimator satisfies [54]:

$$\text{var}[\hat{\alpha}_p] \geq \text{var}[n_M] = \frac{\sigma_n^2}{MB\sigma_x^2}, \quad 0 \leq p \leq P-1. \quad (48)$$

This completes the proof of Theorem 1. $\square$

B. Proof of Theorem 2

Proof. For the ULA based massive MIMO system, the main-lobe gain of a post-reception beamformer is modeled as a circularly symmetric Gaussian function [66]–[68], i.e.,

$$G(\theta) = M \exp\left(-\beta(\theta - \theta_p)^2\right), \quad (49)$$

where $\beta \triangleq \ln(2)/\Delta\theta_{-3\text{dB}}^2$ with the 3dB beamwidth given by $\Delta\theta_{-3\text{dB}} = 2/M$. Upon having the estimated AoAs $\{\hat{\theta}_p\}_{p=0}^{P-1}$, the averaged channel gain estimation becomes:

$$\mathbb{E}[\hat{\alpha}_p] = \exp\left(-2\ln(2)M^2(\hat{\theta}_p - \theta_p)^2\right) \alpha_p. \quad (50)$$

When the number of antennas M is large, in order to attain an unbiased channel gain estimate, the AoA estimation error must satisfy certain condition. More specifically, in order to obtain $\mathbb{E}[\hat{\alpha}_p] = \alpha_p$, we must have

$$M^2(\hat{\theta}_p - \theta_p)^2 \xrightarrow{M \text{ is large}} 0. \quad (51)$$

It can be readily seen that when $\exists \varepsilon > 0$ such that the AoA estimation error satisfies $|\hat{\theta}_p - \theta_p| \sim \mathcal{O}(1/M^{1+\varepsilon})$, the unbiased channel gain estimate can be attained by (5). $\square$

C. Proof of Theorem 3

Proof. From (50), the expectation on the estimated channel gain is given by

$$\begin{aligned} \mathbb{E}[\hat{\alpha}_{p,k}] &= \exp\left(-2\ln(2)M^2\delta_{p,k}^2\right) \alpha_{p,k} \\ &= \left(1 - 2\ln(2)\mathcal{O}\left(\frac{\log_2(M)}{M^{2\varepsilon}}\right)\right) \alpha_{p,k} \xrightarrow{M \text{ is large}} \alpha_{p,k}. \end{aligned} \quad (52)$$

The CRLB of the estimation error is defined by [54]

$$\mathbb{E}\left[|\hat{\alpha}_{p,k} - \alpha_{p,k}|^2\right] \geq -\left(\frac{\partial \ln p\left(\alpha_{p,k} - \frac{1}{M}\mathbf{a}_M^H(\hat{\theta}_{p,k})\mathbf{y}_k\right)}{\partial \alpha_{p,k}^2}\right)^{-1}. \quad (53)$$

From (27) and (26), we conclude that the MSE of the channel gain estimate meets:

$$\mathbb{E}\left[|\hat{\alpha}_{p,k} - \alpha_{p,k}|^2\right] \geq \bar{\sigma}_M^2 \xrightarrow{M \text{ is large}} \frac{\sigma_n^2}{MB\sigma_x^2}. \quad (54)$$

This completes the proof. $\square$

REFERENCES

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [2] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [5] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
- [6] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, Fourth Quarter 2015.
- [7] N. Shlezinger, G. C. Alexandropoulos, M. F. Imani, Y. C. Eldar, and D. R. Smith, "Dynamic metasurface antennas for 6G extreme massive MIMO communications," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 106–113, Apr. 2021.
- [8] J. Wang, C.-X. Wang, J. Huang, H. Wang, and X. Gao, "A general 3D space-time-frequency non-stationary THz channel model for 6G ultra-massive MIMO wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1576–1589, Jun. 2021.
- [9] M. Cheng, S. Yang, and X. Fang, "Adaptive antenna-activation based beamforming for large-scale MIMO communication systems of high speed railway," *China Commun.*, vol. 13, no. 9, pp. 12–23, Sep. 2016.
- [10] S. Yang, C. Zhou, T. Lv, and L. Hanzo, "Large-scale MIMO is capable of eliminating power-thirsty channel coding for wireless transmission of HEVC/H.265 video," *IEEE Wireless Commun.*, vol. 23, no. 3, pp. 57–63, Jun. 2016.
- [11] N. Shariati, J. Wang, and M. Bengtsson, "Robust training sequence design for correlated MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 107–120, Jan. 2014.
- [12] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, "An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Oct. 2014.
- [13] T. Lv, F. Tan, H. Gao, and S. Yang, "A beamspace approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *Signal Processing*, vol. 121, pp. 30–45, Apr. 2016.
- [14] Y. Zhou, Z. Fei, S. Yang, J. Kuang, S. Chen, and L. Hanzo, "Joint angle estimation and signal reconstruction for coherently distributed sources in massive MIMO systems based on 2-D unitary esprit," *IEEE Access*, vol. 5, pp. 9632–9646, 2017.
- [15] T. Lv, S. Yang, and H. Gao, "Semi-blind channel estimation relying on optimum pilots designed for multi-cell large-scale MIMO systems," *IEEE Access*, vol. 4, pp. 1190–1204, 2016.
- [16] H. Wang, G. Li, S. Zheng, S. Yang, and P. Pan, "An approach to reduce the overhead of training sequences in FDD massive MIMO downlink systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1301–1305, Aug. 2019.
- [17] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815–830, Oct. 2014.
- [18] B. Li, S. Wang, J. Zhang, X. Cao, and C. Zhao, "Randomized approximate channel estimator in massive-MIMO communication," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2314 – 2318, Oct. 2020.
- [19] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.
- [20] W. Shen, L. Dai, B. Shim, S. Mumtaz, and Z. Wang, "Joint CSIT acquisition based on low-rank matrix completion for FDD massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2178–2181, Dec. 2015.
- [21] P. A. Eliasi, S. Rangan, and T. S. Rappaport, "Low-rank spatial channel estimation for millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2748–2759, May 2017.
- [22] S. Moshavi, E. Kanterakis, and D. L. Schilling, "Multistage linear receivers for DS-CDMA systems," *International Journal of Wireless Information Networks*, vol. 3, pp. 1–17, Jan. 1996.

- [23] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
- [24] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [25] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, Australia, Apr. 2015, pp. 2909–2913.
- [26] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
- [27] J. Rodriguezfernandez, N. Gonzalezprelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [28] C. Huang, L. Liu, and C. Yuen, "Asymptotically optimal estimation algorithm for the sparse signal with arbitrary distributions," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 10070–10075, Oct. 2018.
- [29] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mm-Wave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Jan. 2019.
- [30] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, Jul. 1996.
- [31] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [32] B. Li, S. Wang, J. Zhang, X. Cao, and C. Zhao, "Ultra-fast accurate AoA estimation via automotive massive-MIMO radar," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1172–1186, Feb. 2022.
- [33] Z. Guo, X. Wang, and W. Heng, "Millimeter-wave channel estimation based on 2-D beamspace MUSIC method," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5384–5394, Aug. 2017.
- [34] D. Fan, F. Gao, G. Wang, Z. Zhong, and A. Nallanathan, "Angle domain signal processing-aided channel estimation for indoor 60GHz TDD/FDD massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1948–1961, Sep. 2017.
- [35] W. Liao and A. Fannjiang, "MUSIC for single-snapshot spectral estimation: Stability and super-resolution," *Applied and Computational Harmonic Analysis*, vol. 40, no. 1, pp. 33–67, Jan. 2016.
- [36] S. Fortunati, R. Grasso, F. Gini, M. S. Greco, and K. LePage, "Single-snapshot DOA estimation by using compressed sensing," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 120, pp. 1–17, Jul. 2014.
- [37] M. A. Maisto, A. Dell'Aversano, I. Russo, A. Brancaccio, and R. Solimene, "A single-snapshot MUSIC algorithm for ADAS radar processing," in *Proc. 2022 Microwave Mediterranean Symposium (MMS)*, Pizzo Calabro, Italy, May 2022, pp. 1–6.
- [38] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.
- [39] L. Dai, X. Gao, S. Han, I. Chih-Lin, and X. Wang, "Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," in *Proc. 2016 IEEE/CIC International Conference on Communications in China (ICCC)*, Chengdu, China, Jul. 2016, pp. 1–6.
- [40] B. Li, Z. Zhou, W. Zou, X. Sun, and G. Du, "On the efficient beamforming training for 60GHz wireless personal area networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 504–515, Feb. 2013.
- [41] B. Li, C. Zhao, M. Sun, H. Zhang, Z. Zhou, and A. Nallanathan, "A Bayesian approach for nonlinear equalization and signal detection in millimeter-wave communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3794–3809, Jul. 2015.
- [42] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-Wave multipath clustering and channel modeling," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1445–1455, Mar. 2014.
- [43] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
- [44] D. Lai and G. Chen, "Dynamical systems identification from time-series data: A hankel matrix approach," *Mathematical and Computer Modelling*, vol. 24, no. 3, pp. 1–10, Aug. 1996.
- [45] L. Lo Presti, M. La Cascia, S. Sclaroff, and O. Camps, "Hankel-based dynamical systems modeling for 3D action recognition," *Image and Vision Computing*, vol. 44, pp. 29–43, Dec. 2015.
- [46] C. Georgakis, Y. Panagakis, and M. Pantic, "Dynamic behavior analysis via structured rank minimization," *International Journal of Computer Vision*, vol. 126, pp. 333–357, Apr. 2018.
- [47] P. Hacker and B. Yang, "Single snapshot DOA estimation," *Advances in Radio Science*, vol. 8, pp. 251–256, Oct. 2010.
- [48] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [49] B. Li, S. Wang, Z. Feng, J. Zhang, X. Cao, and C. Zhao, "Fast pseudospectrum estimation for automotive massive MIMO radar," *IEEE Internet of Things Journal*, vol. 8, no. 20, pp. 15 303–15 316, Oct. 2021.
- [50] D. P. Woodruff, "Sketching as a tool for numerical linear algebra," *Foundations and Trends® in Theoretical Computer Science*, vol. 10, no. 1–2, pp. 1–157, 2014.
- [51] S. Wang, L. Luo, and Z. Zhang, "SPSD matrix approximation via column selection: Theories, algorithms, and extensions," *Journal of Machine Learning Research*, vol. 17, no. 49, pp. 1697–1745, May 2016.
- [52] P. Drineas and M. W. Mahoney, "On the Nyström method for approximating a Gram matrix for improved kernel-based learning," *Journal of Machine Learning Research*, vol. 6, no. 72, pp. 2153–2175, Dec. 2005.
- [53] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
- [54] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. USA: Prentice-Hall, 1993.
- [55] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.
- [56] L. Berriche, K. Abed-Meraim, and J.-C. Belfiore, "Cramer-Rao bounds for MIMO channel estimation," in *Proc. 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. IV, Montreal, Canada, May 2004, pp. 397–400.
- [57] B. Li, P. Chen, H. Liu, W. Guo, X. B. Cao, J. Z. Du, C. L. Zhao, and J. Zhang, "Random sketch learning for deep neural networks in edge computing," *Nature Computational Science*, vol. 1, pp. 221–228, Mar. 2021.
- [58] B. Li, Z. P. Wei, J. T. Wu, S. Yu, T. Zhang, C. L. Zhu, D. Z. Zheng, W. S. Guo, C. L. Zhao, and J. Zhang, "Machine learning-enabled globally guaranteed evolutionary computation," *Nature Machine Intelligence*, vol. 5, no. 4, p. 457–467, 2023.
- [59] A. Gittens and M. W. Mahoney, "Revisiting the Nyström method for improved large-scale machine learning," *Journal of Machine Learning Research*, vol. 17, no. 117, pp. 3977–4041, Apr. 2016.
- [60] E. Balevi, A. Doshi, and J. G. Andrews, "Massive MIMO channel estimation with an untrained deep neural network," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2079–2090, Mar. 2020.
- [61] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
- [62] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Nattagunathan, "Massive mimo linear precoding: A survey," *IEEE Systems Journal*, pp. 1–12, 2018.
- [63] A. Forenza, D. J. Love, and R. W. Heath, "Simplified spatial correlation models for clustered mimo channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 1924–1934, 2007.
- [64] J. J. Ma and P. Li, "Orthogonal AMP," *IEEE Access*, pp. 2020–2033, 2017.
- [65] L. Liu, S. Huang, and B. M. Kurkoski, "Memory AMP," *IEEE Transactions on Information Theory*, pp. 8015–8039, 2022.
- [66] I. Toyoda, *Reference antenna model with side lobe for TG3c evaluation*, IEEE Std. IEEE 802.15-06-0474-00-003c, 2006.
- [67] A. Maltsev, *Channel models for 60GHz WLAN systems*, IEEE Std. IEEE 802.11-09/0334r8, 2010.
- [68] C. Fan, B. Li, C. Zhao, W. Guo, and Y. Liang, "Learning-based spectrum sharing and spatial reuse in mm-Wave ultradense networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4954–4968, Jun. 2018.



Bin Li received the B.Eng. degree in electrical information engineering from Beijing University of Chemical Technology in 2007, and the Ph.D. degree in communication and information engineering from Beijing University of Posts and Telecommunications (BUPT) in 2013. In the same year, he joined BUPT, where he is currently an Associate Professor with the School of Information and Communication Engineering. His current research interests include signal processing for wireless communications and machine learning, such as millimeter-wave communications, UAV communications, MIMO communication/radar systems. He received the 2011 ChinaCom Best Paper Award, the 2015 IEEE WCSP Best Paper Award. He is an Associate Editor of *China Communications*.



Ziping Wei received the B.Eng. degree in communication engineering from Civil Aviation University of China (CAUC) in 2018. He is currently pursuing the Ph.D. degree with the Key Laboratory of Universal Wireless Communications, Ministry of Education, BUPT, China. His research interests include MIMO signal processing, channel estimation, and radar signal processing.



Shaoshi Yang (Senior Member, IEEE) received the B.Eng. degree in information engineering from Beijing University of Posts and Telecommunications (BUPT), China, in 2006, and the Ph.D. degree in electronics and electrical engineering from University of Southampton, UK, in 2013. From 2008 to 2009, he was a Researcher with Intel Labs China. From 2013 to 2016, he was a Research Fellow with the School of Electronics and Computer Science, University of Southampton. From 2016 to 2018, he was a Principal Engineer with Huawei Technologies Co., Ltd., where he made significant contributions to Huawei's products and solutions on 5G base stations, wideband IoT, and cloud gaming/VR. He is currently a Full Professor with BUPT. His research interests include 5G/5G-A/6G, massive MIMO, mobile ad hoc networks, distributed artificial intelligence, and cloud gaming/VR. He received numerous research awards from University of Southampton, Huawei, IEEE ComSoc, Xiaomi Foundation, China Association of Inventions, and China Industry-University-Research Institute Collaboration Association. He was/is an Editor for *IEEE Systems Journal*, *IEEE Wireless Communications Letters*, and *Signal Processing* (Elsevier). He is a standing committee member of the China Computer Federation (CCF) Technical Committee on Distributed Computing and Systems. He was also a Guest Researcher with the Isaac Newton Institute for Mathematical Sciences, Cambridge University. For more details of his research progress, please refer to <https://shaoshiyang.weebly.com/>



Yang Zhang received the B.Eng. degree in communication engineering in 2007, and the M.Eng. degree in computer applied technology in 2010, both from China University of Petroleum (CUP), and the Ph.D. degree in communication and information engineering from Beijing University of Posts and Telecommunications (BUPT) in 2014. He is currently with China Satellite Network Innovation Research Institute Co., Ltd. His research interests include MIMO signal processing, satellite communications and networking.



Jun Zhang (Senior Member, IEEE) received the M.S. degree in statistics with the Department of Mathematics from Southeast University, Nanjing, China, in 2009, and the Ph.D. degree in communication and information systems with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in 2013. From 2013 to 2015, he was a Postdoctoral Research Fellow with Singapore University of Technology and Design, Singapore. Since 2015, he is on the faculty of the Jiangsu Key Laboratory of Wireless Communications, School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, where he is currently a Professor. His research interests include massive MIMO communications, RIS-assisted wireless communications, UAV-assisted wireless communications, physical layer security, and large dimensional random matrix theory. Dr. Zhang was a recipient of the IEEE GLOBECOM Best Paper Award in 2016, the IEEE APCC Best Paper Award in 2017, the IEEE JC&S Symposium Best Paper Award in 2022, the IEEE/CIC ICCS Best Paper Award in 2023, and the WCSP Best Paper Award in 2023. He was an Associate Editor for *IEEE Communications Letters*.



Chenglin Zhao received the B.Eng. degree in radio technology from Tianjin University in 1986, and the M.Eng. degree in circuits and systems from Beijing University of Posts and Telecommunications (BUPT) in 1993, and the Ph.D. degree in communication and information systems from BUPT, in 1997. At present, he serves as a Professor in BUPT. His research is focused on emerging technologies of short-range wireless communications, cognitive radio, 60GHz millimeter-wave communications.



Sheng Chen (Life Fellow, IEEE) received his B.Eng. degree from East China Petroleum Institute, Dongying, China, in 1982, and his Ph.D. degree from City, University of London, in 1986, both in control engineering. In 2005, he was awarded the higher doctoral degree, Doctor of Sciences (DSc), from University of Southampton, UK. From 1986 to 1999, He held research and academic appointments at University of Sheffield, University of Edinburgh and University of Portsmouth, all in UK. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he holds the post of Professor in Intelligent Systems and Signal Processing. Dr. Chen's research interests include machine learning, neural networks and wireless communications. He has published over 700 research papers. He has 19,000+ Web of Science citations with h-index 61 and 38,000+ Google Scholar citations with h-index 83. He is also a Fellow of the Royal Academy of Engineering, UK, a Fellow of Asia-Pacific Artificial Intelligence Association (AAIA), and a Fellow of IET. He is one of the original ISI highly cited researcher in engineering (March 2004).