

A Partial Reciprocity-based Channel Prediction Framework for FDD Massive MIMO with High Mobility

Ziao Qin, Haifan Yin, Yandi Cao, Weidong Li, and David Gesbert, *Fellow, IEEE*

Abstract—Massive multiple-input multiple-output (MIMO) is believed to deliver unrepresented spectral efficiency gains for 5G and beyond. However, a practical challenge arises during its commercial deployment, which is known as the “curse of mobility”. The performance of massive MIMO drops alarmingly when the velocity level of user increases. In this paper, we tackle the problem in frequency division duplex (FDD) massive MIMO with a novel Channel State Information (CSI) acquisition framework. A joint angle-delay-Doppler (JADD) wideband precoder is proposed for channel training. Our idea consists in the exploitation of the partial channel reciprocity of FDD and the angle-delay-Doppler channel structure. More precisely, the base station (BS) estimates the angle-delay-Doppler information of the UL channel based on UL pilots using Matrix Pencil (MP) method. It then computes the wideband JADD precoders according to the extracted parameters. Afterwards, the user estimates and feeds back some scalar coefficients for the BS to reconstruct the predicted DL channel. Asymptotic analysis shows that the CSI prediction error converges to zero when the number of BS antennas and the bandwidth increases. Numerical results with industrial channel model demonstrate that our framework can well adapt to high speed (350 km/h), large CSI delay (10 ms) and channel sample noise.

Index Terms—Massive MIMO, curse of mobility, channel prediction, FDD, angle-delay domain, partial reciprocity, Matrix Pencil, 5G.

I. INTRODUCTION

THE 5G wireless communication is being deployed in real life and is given great expectations on high throughput rate, low latency and high reliability. To achieve such intriguing merits of 5G, massive MIMO technology is indispensable. These benefits are brought by the large numbers of antennas at the BS side while eliminating uncorrelated noise and fast fading [1]. Massive MIMO system has shown great potential in improving spectral efficiency (SE) and energy efficiency (EE) [2]. Even though the pilot contamination problem limits massive MIMO system performance [3], this effect can be mitigated by exploiting the angular structure of channel [4] and differences of the channel power [5].

High SE performance depends on accurate CSI. Thanks to the channel reciprocity of TDD, CSI can be obtained by an acceptable pilot training overhead which scales with the

number of user equipments (UEs) instead of the number of BS antennas. Therefore, TDD mode may be the favorable choice for massive MIMO system. However, a large percent of current cellular communication system operates in FDD mode, thus massive MIMO operating in FDD mode has equal importance. The authors in [6] measured the performance at 2.6 GHz in the two modes and conclude that each enjoys its own advantages in different scenarios. Unfortunately, the CSI acquisition in FDD mode is more challenging due to the non-reciprocal UL and DL channel, and therefore the training and feedback overhead. Many research works have offered possible solutions to CSI acquisition in FDD massive MIMO system. The authors in [7] and [8] utilized a statistical channel information and user grouping based prebeamformer to reduce pilot training and feedback overhead, which is known as the “Joint spatial division and multi-plexing” (JSDM) method. The low-rankness property of channel correlation matrices was considered to design pilot training and feedback under a spatial correlation channel model in [9]. Spatial sparsity of massive MIMO channel can also be exploited through the compressed sensing (CS) method. In [10], the authors estimated the channel by extracting channel parameters through CS method in millimeter-wave massive MIMO. Another possible approach to address FDD massive MIMO channel estimation is based on channel parameter extraction. By exploiting angle information through a discrete Fourier transform (DFT) projection [11], the DL channel can be reconstructed through angular information and channel gain which are estimated separately. The authors in [12] introduced a Newtonized orthogonal matching pursuit (NOMP) method to detect angle, delay and gains and reconstructed the channel following a multipath channel model. Deep learning method was also utilized to reconstruct the DL channel [13]. However, the papers above mainly considered a block fading scenario where the channel was assumed to be constant for a period of time. This assumption is reasonable in a stationary or low-mobility scenario.

However, in practice, the system performance may degrade badly in mobility scenarios [14], [15] even in TDD mode. This effect is caused by the time-varying nature of channel. The outdated CSI severely corrupts the SE performance. Unlike in stationary settings, Doppler frequency shift becomes nontrivial in mobile environments. The authors in [16] proposed a data-aided channel prediction based on variational Bayesian inference (VBI) framework in high mobility scenario. A maximum-likelihood based method is introduced in [17] to estimate channel parameters in vehicle-to-vehicle (V2V) MIMO system.

Z. Qin, H. Yin, Y. Cao and W. Li are with Huazhong University of Science and Technology, 430074 Wuhan, China (e-mail: ziao_qin@hust.edu.cn, yin@hust.edu.cn, yandicao@hust.edu.cn, weidongli@hust.edu.cn).

D. Gesbert is with EURECOM, 06410 Biot, France (e-mail: gesbert@eurecom.fr).

This work was supported by the National Natural Science Foundation of China under Grant 62071191. The corresponding author is Haifan Yin.

Some works addressed the CSI delay influence on the channel in a theoretical view [18], [19]. In [15], the authors proposed a channel prediction method to solve the mobility problem utilizing Prony-based angle-delay domain channel prediction. The authors of [20] addressed the mobility problem in massive MIMO from a deep learning view. Nevertheless these papers mainly focused on TDD mode.

Different from TDD mode, closed-loop feedback of CSI from the UE to the BS is inevitable, which introduces CSI quantization error, in addition to even larger CSI delay. Especially in high-mobility scenario, the channel coherence time is much shorter than the low-mobility scenario and timely feedback is more challenging in FDD mode due to the different operating frequency bands between UL and DL. Worse still, the training and feedback overhead are much heavier than TDD, and thus has to be reduced. The state-of-the-art algorithms like CS [10], deep learning [20] and JSDM [8] method mainly focus on reducing the pilot training and feedback overhead. Some research works utilized maximum-likelihood method [17], deep learning [20] and machine learning method [21] to address the CSI aging problem in TDD. The authors in [22] utilized partial channel reciprocity in terms of the angular support to facilitate the CSI feedback in TDD for the case that the UE has unequal number of TX and RX antennas. In [23], a channel reconstruction method based on CSI-RS and SRS in TDD system was proposed. However, these methods did not consider the mobility problem in FDD and the given solutions were mostly NP-hard. As the high mobility demands timely CSI acquisition and high efficiency of channel prediction algorithm, these methods cannot directly apply in FDD massive MIMO with high-mobility. Recently, some works like [13], [24] proposed channel prediction methods for FDD massive MIMO. Unfortunately, the performances of these methods may not be guaranteed in a rich scattering environment with a large number of multipath, especially in high-mobility scenario. To the best of our knowledge, few works have addressed these real-world problems simultaneously in a practical multipath channel model.

In this paper, we aim to solve this problem with a novel CSI acquisition framework which is easy to deploy and has polynomial complexity. Even though the full channel reciprocity in FDD is not available like TDD, some frequency-unrelated channel parameters are reciprocal between the DL channel and the UL channel [25]–[27]. Through the channel measurement campaigns, the partial reciprocity in FDD was verified in [28]. The partial reciprocity allows us to extract some useful information from the UL channel estimation, e.g., the angle, delay, and Doppler frequency shift. We propose to extract such information through an efficient linear prediction method known as MP [29]. Once the information is obtained, we design a JADD spatial-frequency precoders for the wideband DL pilot transmission. The precoders capitalize on the channel sparsity in angle-delay domain, as well as the partial reciprocity. They will help reduce the training overhead and facilitate the DL channel reconstruction. Note that different from existing methods, our precoder are wideband and require joint operation from the BS and the UE. Afterwards, the UE estimates some complex scalar coefficients based on the

precoded DL training signal and feeds them back to the BS. Finally the BS reconstructs the DL CSI using the coefficients and the extracted UL channel parameters.

Different from previous channel reconstruction methods like [12], [13], [15], [23], [24], we devise a wideband precoder and JADD feedback framework. Our framework outperforms traditional methods which are typically based on the NP-hard solutions or failing to timely update CSI. Moreover, our approach is capable of predicting the channel in polynomial complexity. Simulation results under the 3rd Generation Partner Project (3GPP) channel model indicate that our proposed framework is robust to high mobility scenarios with even 350 km/h of UE speed and to large CSI delay. Moreover, we test our framework in different scattering environments, BS antenna configurations and noisy channel sample cases. The numerical results demonstrate the robustness of our framework.

Our main contributions are

- We address the mobility problem of FDD massive MIMO under an industrial multipath channel model, which was rarely considered in the literature. By exploiting the angle-delay-Doppler structure and the partial reciprocity of the channel, we propose a JADD CSI acquisition framework, which combats the outdated DL CSI and reduces the training overhead simultaneously.
- We propose to extract the Doppler frequency shifts using the MP method in angle-delay domain, where the channel shows more sparsity. This method requires less channel samples and achieves high accuracy for the Doppler estimation, due to the high spatial and frequency resolution of a wideband massive MIMO system.
- We propose a novel training and feedback framework for FDD massive MIMO. The key ingredients are a wideband precoder for DL pilots and the computation of the complex coefficients of the DL paths at the UE side. This precoding method requires a two-step joint operation of the BS and the UE. Only scalar coefficients need to be fed back to the BS. In this framework, the training and feedback overhead no longer depends on the number of the BS antennas and bandwidth, but on the angle-delay sparsity of the channel and the prediction order of the MP method.
- We derive the upper bound of the DL channel prediction error under limited BS antennas and bandwidth. Our asymptotic analysis shows the channel prediction error converges to zero when the number of antennas at the BS and the bandwidth increase while only two UL channel samples are needed. We also suggest the choice of the prediction order when applying our method.

The rest of the paper is organized as follows. Sec. II introduces our channel model. Sec. III demonstrates the UL channel parameters estimation method. Sec. IV discusses the DL pilot training, feedback, and DL channel reconstruction. Sec. V contains the performance analysis of our proposed framework. Sec. VI shows the numerical results of our framework. Sec. VII is the conclusion of our work.

Notations: The boldface front stands for vector and matrix. \otimes is Kronecker product symbol. $\text{diag}(\mathbf{X})$ means a diagonal matrix with \mathbf{X} as its diagonal elements and if \mathbf{X} is a block

matrix, $\text{diag}(\mathbf{X})$ denotes a block diagonal matrix. $\text{vec}(\mathbf{X})$ is the vectorization of \mathbf{X} . \mathbf{X}^\dagger , \mathbf{X}^T , and \mathbf{X}^H denote the Moore-Penrose inversion, transpose and conjugation of \mathbf{X} , respectively. $\mathbb{C}^{a \times b}$ is a matrix space with a rows and b columns. $|\mathbf{x}|$ denotes the absolute value of \mathbf{x} and $\|\mathbf{X}\|_2$ is the second-order induced norm of \mathbf{X} . $\Re(\alpha)$ denotes the real part of complex α . $\text{mod}(x)$ is the modular operation of x . \triangleq refers to the definition symbol. $\mathbb{E}\{x\}$ means calculating expectation of x . $\mathbf{R} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ means that \mathbf{R} satisfies zero-mean complex circular Gaussian distribution.

II. SYSTEM MODEL

This paper considers a wideband FDD massive MIMO system where the BS is equipped with a uniform planar array (UPA). The classical orthogonal frequency division multiplex (OFDM) modulation is adopted with N_f sub-carriers and a f_Δ subcarrier spacing. The number of BS antennas is $N_t = N_v N_h$, where N_v and N_h denote the number of antennas in a row and in a column, respectively. The center frequencies of UL and DL are f^u and f^d , respectively.

A multi-path channel model following [30] is adopted in our work. The number of paths of the channel is denoted by P . The corresponding parameters of each path p are the complex amplitude β_p , steering vector $\boldsymbol{\alpha}^u(\theta_p^u, \phi_p^u)$, Doppler frequency shift ω_p , and delay τ_p . Therefore, the UL channel between the BS and the UE k at a certain time t and frequency f is

$$\mathbf{h}_{k,r}^u(t, f) = \sum_{p=1}^P \beta_p^u \boldsymbol{\alpha}^u(\theta_p^u, \phi_p^u) e^{-j2\pi f \tau_p^u} e^{jw_p^u t}, \quad (1)$$

where the subscript k, r means the r -th antenna of the UE k and the superscript u denotes the UL channel. For simplicity, we drop the subscripts r and k here and afterwards. The UL Doppler frequency shift is defined as $w_p^u = v \cos \varphi_p^u f^u / c$, where v is the velocity of the UE and φ_p^u is the angle between path p and the 3D velocity vector of the UE \mathbf{v} . c is the speed of light. Denote the zenith angle and azimuth angle by θ_p^u, ϕ_p^u , respectively. Fig. 1 demonstrates the UPA antenna configuration in 3D-Cartesian coordinate system, the zenith angle θ , the azimuth angle ϕ , speed direction angle φ which is the angle between the path and the velocity vector of the UE. The transmit steering vector is $\boldsymbol{\alpha}^u(\theta_p^u, \phi_p^u) \in \mathbb{C}^{N_t \times 1}$ and is modeled as the Kronecker product of the vertical steering vector $\boldsymbol{\alpha}_v^u(\theta_p^u)$ and the horizontal steering vector $\boldsymbol{\alpha}_h^u(\theta_p^u, \phi_p^u)$

$$\boldsymbol{\alpha}^u(\theta_p^u, \phi_p^u) = \boldsymbol{\alpha}_h^u(\theta_p^u, \phi_p^u) \otimes \boldsymbol{\alpha}_v^u(\theta_p^u), \quad (2)$$

where

$$\boldsymbol{\alpha}_h^u(\theta_p^u, \phi_p^u) = \begin{bmatrix} 1 \\ e^{j2\pi \frac{l_h f^u}{c} \cos \theta_p^u \cos \phi_p^u} \\ \dots \\ e^{j2\pi \frac{l_h f^u}{c} (N_h - 1) \cos \theta_p^u \cos \phi_p^u} \end{bmatrix}, \quad (3)$$

$$\boldsymbol{\alpha}_v^u(\theta_p^u) = \begin{bmatrix} 1 \\ e^{j2\pi \frac{l_v f^u}{c} \cos \theta_p^u} \\ \dots \\ e^{j2\pi \frac{l_v f^u}{c} (N_v - 1) \cos \theta_p^u} \end{bmatrix}, \quad (4)$$

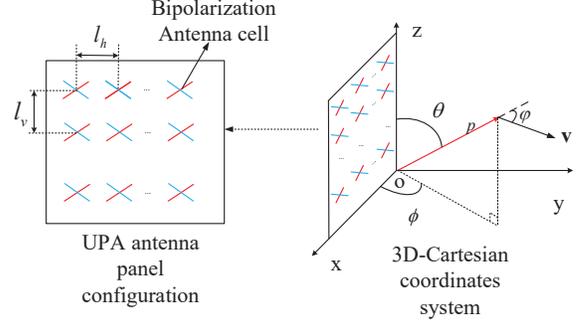


Fig. 1. UPA antenna configuration in 3D-Cartesian coordinate system.

and l_v, l_h are the spacing between the antennas in vertical direction and horizontal direction, respectively. Similarly, the DL channel is modeled as

$$\mathbf{h}^d(t, f) = \sum_{p=1}^P \beta_p^d \boldsymbol{\alpha}^d(\theta_p^d, \phi_p^d) e^{-j2\pi f \tau_p^d} e^{-j2\pi(f^d - f^u)\tau_p^d} e^{jw_p^d t}, \quad (5)$$

where d stands for the DL channel.

Unlike TDD, in FDD only some parameters of the UL and DL channels are reciprocal [26]

$$\tau_p^u = \tau_p^d, \theta_p^u = \theta_p^d, \phi_p^u = \phi_p^d, \frac{w_p^u}{w_p^d} = \frac{f^u}{f^d}. \quad (6)$$

The DL steering vector $\boldsymbol{\alpha}^d(\theta_p^d, \phi_p^d)$ is frequency-related and is calculated by the UL steering vector $\boldsymbol{\alpha}^u(\theta_p^u, \phi_p^u)$ with a rotation matrix as

$$\boldsymbol{\alpha}^d(\theta_p^d, \phi_p^d) = (\mathbf{R}_h(\theta_p^d, \phi_p^d) \otimes \mathbf{R}_v(\theta_p^d)) \cdot \boldsymbol{\alpha}^u(\theta_p^u, \phi_p^u), \quad (7)$$

where

$$\mathbf{R}_h(\theta_p^d, \phi_p^d) = \text{diag} \left(\begin{array}{c} 1 \\ e^{j2\pi \frac{l_h (f^d - f^u)}{c} \cos \theta_p^d \cos \phi_p^d} \\ \dots \\ e^{j2\pi (N_h - 1) \frac{l_h (f^d - f^u)}{c} \cos \theta_p^d \cos \phi_p^d} \end{array} \right), \quad (8)$$

$$\mathbf{R}_v(\theta_p^d) = \text{diag} \left(\begin{array}{c} 1 \\ e^{j2\pi \frac{l_v (f^d - f^u)}{c} \cos \theta_p^d} \\ \dots \\ e^{j2\pi (N_v - 1) \frac{l_v (f^d - f^u)}{c} \cos \theta_p^d} \end{array} \right), \quad (9)$$

are the horizontal rotation matrix and vertical rotation matrix, respectively.

In FDD mode, UL and DL symbols are transmitted successively in time domain. Fig. 2 demonstrates the flowchart of our proposed framework. The BS utilizes the SRS to extract the channel parameters, and based on the parameters, computes the wideband precoder for DL pilot. According to [31], the sounding reference signal (SRS) can be set as cyclical mode with a flexible periodicity T_{SRS} in units of slots. Considering a common configuration where the subcarrier-spacing is 30 kHz, the minimum SRS periodicity T_{SRS} can be as short as 0.5 ms. We denote the CSI delay by $T_d = N_d T_{\text{SRS}}$, where N_d is the

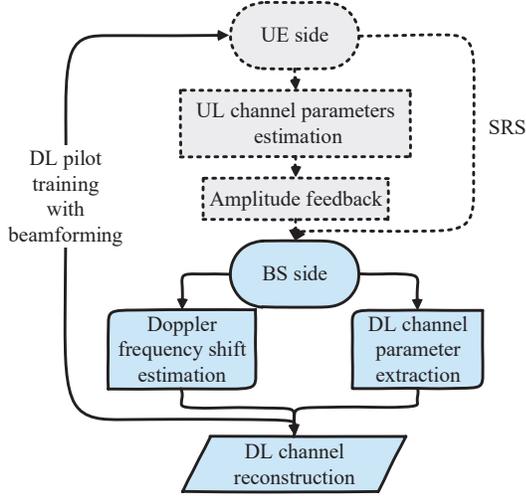


Fig. 2. Flowchart of our framework.

delay in unit of time slots. The UE computes the complex DL path coefficients upon receiving the DL pilot and feeds them back. The BS finally reconstructs the DL CSI based on the feedback and the extracted parameters.

III. UPLINK CHANNEL PARAMETERS EXTRACTION

Our proposed framework depends on the UL channel parameters extraction thanks to the merits of the partial channel reciprocity in FDD. The UL channel parameters is relatively easy to estimate through SRS at the BS side. In order to better exploit the sparsity of the multipath channel, we discuss the UL channel parameters extraction in angle-delay domain.

A. Angle-delay domain projection

In a wideband massive MIMO system, the UL channel in time domain can be written as

$$\mathbf{h}^u(t, f_n^u) = \sum_{p=1}^P \beta_p^u \boldsymbol{\alpha}^u(\theta_p^u, \phi_p^u) e^{-j2\pi f_n^u \tau_p^u} e^{jw_p^u t}, \quad (10)$$

where $f_n^u, n \in \{1, \dots, N_f\}$ is frequency of the n -th subcarrier. Then the UL channel in matrix form with all subcarriers is

$$\mathbf{H}^u(t) = \begin{bmatrix} \mathbf{h}^u(t, f_1^u), \dots, \mathbf{h}^u(t, f_{N_f}^u) \end{bmatrix}. \quad (11)$$

The vector form of (11) is

$$\mathbf{h}^u(t) = \text{vec}(\mathbf{H}^u(t)) = \sum_{p=1}^P \beta_p^u e^{-j2\pi f^u \tau_p^u} e^{jw_p^u t} \mathbf{r}_p^u, \quad (12)$$

where $\mathbf{r}_p^u = \mathbf{c}(\tau_p^u) \otimes \boldsymbol{\alpha}^u(\theta_p^u, \phi_p^u)$ is the angle-delay structure of path p and $\mathbf{c}(\tau_p^u)$ is the delay vector

$$\mathbf{c}(\tau_p^u) = e^{-j2\pi f^u \tau_p^u} \begin{bmatrix} 1 \dots e^{-j2\pi \tau_p^u (N_f-1) f_\Delta} \end{bmatrix}^T. \quad (13)$$

Similarly, the vectorized DL channel is

$$\mathbf{h}^d(t) = \sum_{p=1}^P \beta_p^d e^{-j2\pi f^d \tau_p^d} e^{jw_p^d t} \mathbf{r}_p^d. \quad (14)$$

A matrix $\mathbf{Q} \in \mathbb{C}^{N_t N_f \times N_t N_f}$ is used to project $\mathbf{h}^u(t)$ to the angle-delay domain [32], [33]

$$\mathbf{Q} = \mathbf{W}(N_f)^H \otimes \mathbf{W}(N_h) \otimes \mathbf{W}(N_v). \quad (15)$$

The DFT matrix $\mathbf{W}(X)$ is calculated by

$$\mathbf{W}(X) = \frac{1}{\sqrt{X}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & w^{1 \cdot 1} & \dots & w^{1 \cdot (X-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w^{(X-1) \cdot 1} & \dots & w^{(X-1) \cdot (X-1)} \end{bmatrix}, \quad (16)$$

where $w = e^{j\frac{2\pi}{X}}$. Then the channel in angle-delay domain $\hat{\mathbf{g}}^u(t) \in \mathbb{C}^{N_t N_f \times 1}$ is

$$\hat{\mathbf{g}}^u(t) = \mathbf{Q}^H \mathbf{h}^u(t). \quad (17)$$

By projecting $\mathbf{h}^u(t)$ to angle-delay domain, we can exploit the channel sparsity and obtain

$$\mathbf{h}^u(t) = \sum_{i=1}^{N_t N_f} \hat{g}_i^u(t) \mathbf{q}_i, \quad (18)$$

where \mathbf{q}_i is the i -th column of \mathbf{Q} and $\hat{g}_i^u(t) = \mathbf{q}_i^H \mathbf{h}^u(t)$ is the corresponding complex amplitude. Thanks to the channel sparsity in angle-delay domain, $\mathbf{h}^u(t)$ can be approximated with the linear combination of a relatively small number of selected columns of \mathbf{Q} which contain most power of the channel $\mathbf{h}^u(t)$. The set of column indices of \mathbf{Q} is found by

$$\mathcal{S} = \arg \min_{|\mathcal{S}|} \left\{ \sum_{l=1}^{N_L} \sum_{i \in \mathcal{S}} |\hat{g}_i^u(t_l)|^2 \geq \eta \sum_{l=1}^{N_L} |\hat{\mathbf{g}}^u(t_l)|^2 \right\}, \quad (19)$$

where $\hat{g}_i^u(t_l)$ is the i -th row of $\hat{\mathbf{g}}^u(t_l)$ and η denotes the power threshold. We use N_L channel samples in each UL channel parameter extraction. The size of \mathcal{S} is denoted by N_s which is referred to as the total number of selected columns in \mathbf{Q} .

In fact, the index set \mathcal{S} is time-varying and is updated in each UL channel parameter extraction. However, we drop the argument t for simplicity in the rest of the paper. Even though N_s is variant to channel sample and time, we tend to find a fixed N_s satisfying (19), which is more convenient to implement in practice. Note that N_s should be carefully chosen not only because it affects the estimation accuracy but also the computation complexity. Thus, there lies a trade-off of N_s between the performance and the complexity. The UL channel can be approximated with N_s angle-delay vectors

$$\tilde{\mathbf{h}}^u(t) = \sum_{i \in \mathcal{S}} \hat{g}_i^u(t) \mathbf{q}_i. \quad (20)$$

Comparing (20) and (12), the complex amplitude $\hat{g}_i^u(t)$ has an implicit physical meaning. Each vector \mathbf{q}_i maps the angle-delay structure \mathbf{r}_p^u and the corresponding $\hat{g}_i^u(t)$ maps the complex gain and Doppler frequency $\beta_p^u e^{-j2\pi f^u \tau_p^u} e^{jw_p^u t}$. An M -order superposition of exponentials is utilized to fit the complex amplitude $\hat{g}_i^u(t)$ as

$$\hat{g}_i^u(t) \triangleq \sum_{m=1}^M a_m^u(i) (z_m^u(i))^t, \quad (21)$$

where $z_m^u(i)$ denotes the Doppler frequency and $a_m^u(i)$ denotes the corresponding complex amplitude. Then the UL channel can be approximated by the following form

$$\tilde{\mathbf{h}}^u(t) = \sum_{i \in \mathcal{S}} \sum_{m=1}^M a_m^u(i) (z_m^u(i))^t \mathbf{q}_i. \quad (22)$$

For simplicity, we assume the same M for all selected angle-delay vectors. The value of M should be carefully chosen considering the complexity, Doppler frequency variety and the mismatch problem of DFT projection. More details will be discussed in Sec. V. In the following subsection, we aim to estimate the Doppler frequency $z_m^u(i)$ with MP method.

B. Matrix Pencil based Doppler Estimation

The problem of estimating $z_m^u(i)$ from channel samples in (21) has a form of a superposition of complex exponentials, where MP method is particularly applicable. MP method has the advantage of low computation complexity and noise-insensitive [29] over traditional polynomial methods like ESPRIT or Prony. Therefore, we apply this method in angle-delay domain in order to extract the Doppler. We should first introduce Assumption 1, which means the stationary time is larger than the CSI delay.

Assumption 1 During the period of CSI delay T_d , channel parameters such as angle and Doppler frequency shift are nearly unchanged.

This assumption often holds under a moderate mobility scenario [34]. Assume the CSI delay is 5 ms and the UE speed is 100 km/h for example, then the UE moves about 0.14 m during this CSI delay period. The position of the UE is approximately unchanged considering that the distance between the UE and the BS is much larger. Therefore, channel parameters such as angles and Doppler barely change during CSI delay period.

We first briefly introduce the principle of MP method. The Doppler $z_m^u(i)$ is referred to as the pole in MP. In this method, three parameters are crucial, i.e., sample quantity N_L , prediction order L and poles $z_m^u(i)$. Then, the prediction matrices $\mathbf{P}_1(i)$, $\mathbf{P}_0(i)$ are generated by the complex gain $\hat{g}_i^u(t)$ as

$$\mathbf{P}_1(i) = \begin{bmatrix} \hat{g}_i^u(t_{L+1}) & \hat{g}_i^u(t_L) & \cdots & \hat{g}_i^u(t_2) \\ \hat{g}_i^u(t_{L+2}) & \hat{g}_i^u(t_{L+1}) & \cdots & \hat{g}_i^u(t_3) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{g}_i^u(t_{N_L}) & \hat{g}_i^u(t_{N_L-1}) & \cdots & \hat{g}_i^u(t_{N_L-L+1}) \end{bmatrix},$$

$$\mathbf{P}_0(i) = \begin{bmatrix} \hat{g}_i^u(t_L) & \hat{g}_i^u(t_{L-1}) & \cdots & \hat{g}_i^u(t_1) \\ \hat{g}_i^u(t_{L+1}) & \hat{g}_i^u(t_L) & \cdots & \hat{g}_i^u(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{g}_i^u(t_{N_L-1}) & \hat{g}_i^u(t_{N_L-2}) & \cdots & \hat{g}_i^u(t_{N_L-L}) \end{bmatrix}.$$

Drop superscript u for simplicity and construct three matrices

$$\mathbf{Z}_0 = \text{diag}\{z_1(i), z_2(i), \cdots, z_M(i)\}, \quad (23)$$

$$\mathbf{Z}_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1(i) & z_2(i) & \cdots & z_M(i) \\ \vdots & \vdots & \ddots & \vdots \\ z_1(i)^{N_L-L-1} & z_2(i)^{N_L-L-1} & \cdots & z_M(i)^{N_L-L-1} \end{bmatrix},$$

$$\mathbf{Z}_2 = \begin{bmatrix} z_1(i)^{L-1} & z_1(i)^{L-1} & \cdots & 1 \\ z_1(i)^{L-1} & z_2(i)^{L-2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ z_M(i)^{L-1} & z_M(i)^{L-2} & \cdots & 1 \end{bmatrix}. \quad (24)$$

The complex amplitude $a_m^u(i)$ is given in the form of a diagonal matrix

$$\mathbf{A}_u(i) = \text{diag}\{a_1^u(i), a_2^u(i), \cdots, a_M^u(i)\}. \quad (25)$$

The following relationship holds according to [29]

$$\begin{cases} \mathbf{P}_0(i) = \mathbf{Z}_1(i) \mathbf{A}_u(i) \mathbf{Z}_2(i), \\ \mathbf{P}_1(i) = \mathbf{Z}_1(i) \mathbf{A}_u(i) \mathbf{Z}_0(i) \mathbf{Z}_2(i). \end{cases} \quad (26)$$

In order to describe the mechanism of how to obtain poles, Lemma 1 [29] is introduced

Lemma 1 If $M \leq L \leq N_L - M$, the solution to the singular generalized eigenvalue problem

$$\left(\mathbf{P}_0(i)^\dagger \mathbf{P}_1(i)\right) \mathbf{x} = z \mathbf{x}, \quad (27)$$

points the way to find poles $z_m^u(i)$. Each eigenvalue z equals to the pole $z_m^u(i)$. \mathbf{x} is the corresponding eigenvector.

After obtaining all poles, the Doppler frequency can be easily calculated. Algorithm 1 explains how to estimate the UL channel parameters like Doppler frequency shifts and the set of angle-delay indices.

Algorithm 1 Matrix Pencil based Doppler Estimation

- 1: Initialize N_L, L, M , start time t_s , end time t_e and obtain channel sample $\mathbf{h}^u(t)$
 - 2: Project $\mathbf{h}^u(t)$ to angle-delay domain as (18)
 - 3: Find a suitable N_s satisfying (19)
 - 4: **for** $t \in [t_s, t_e]$ **do**
 - 5: Obtain the index set \mathcal{S}
 - 6: **for** $n_i \in [1, N_s]$ **do**
 - 7: Generate prediction matrix $\mathbf{P}_1(n_i)$, $\mathbf{P}_0(n_i)$
 - 8: Using (27) to calculate the eigenvalue matrix $\mathbf{Z}_0(n_i)$
 - 9: Update $n_i = n_i + 1$
 - 10: **end for**
 - 11: Update $t = t + 1$
 - 12: **end for**
 - 13: Return the UL angle-delay index set and Doppler frequency shift.
-

C. Noisy channel sample analysis

The previous discussion is based on noise-free channel sample assumption. In such cases we can let the prediction order $L = M$ for simplicity. In realistic scenarios, only noisy channel samples are available. In this case, we propose to apply a minimum description length (MDL) criterion [35] to

detect the value M in (22) and cancel the noise by an M order subtraction Singular Value Decomposition (SVD) where the prediction order satisfies $L > M$. The value of M is minimized under an MDL criterion without prior decision or hypothesis as

$$M = \min_{x \in \{0, 1, \dots, L-1\}} \left\{ \log \left(\frac{\prod_{m=x+1}^L z_m(i)^{1/(L-x)}}{\frac{1}{L-x} \sum_{m=x+1}^L z_m(i)} \right)^{-N_L(L-x)} \right\},$$

where $z_m(i)$ is the singular value of

$$\mathbf{P}_{1,0} = [\mathbf{p}(t_{L+1}) \mid \mathbf{P}_0], \quad (28)$$

with $\mathbf{p}(t_{L+1})$ being the first column of \mathbf{P}_1 . After obtaining the value of M , the prediction matrix $\mathbf{P}_{1,0}$ is calculated after a rank- M truncated SVD

$$\mathbf{P}_{1,0} = \mathbf{U}_M \mathbf{\Lambda}_M \mathbf{V}_M^H, \quad (29)$$

where $\mathbf{U}_M, \mathbf{\Lambda}_M, \mathbf{V}_M$ are M -truncated left singular vector, singular value and right singular vector of $\mathbf{P}_{1,0}$, respectively. Then (27) in Lemma 1 becomes

$$(\mathbf{P}_{0,M}(i)^\dagger \mathbf{P}_{1,M}(i)) \mathbf{x} = z \mathbf{x}, \quad (30)$$

where

$$\begin{cases} \mathbf{P}_{0,M}(i) = \mathbf{U}_M \mathbf{\Lambda}_M \mathbf{V}_{M(1:M-1,:)}^H, \\ \mathbf{P}_{1,M}(i) = \mathbf{U}_M \mathbf{\Lambda}_M \mathbf{V}_{M(2:M,:)}^H, \end{cases} \quad (31)$$

and $\mathbf{V}_{M(1:M-1,:)}$ denotes the sub-matrix consists of the first row to the $(M-1)$ -th row and $\mathbf{V}_{M(2:M,:)}$ consists of the second row to the M -th row likewise. Finding the eigenvalue of (30) equals to obtaining the poles $z_m(i)$ in noisy channel sample case.

In this section, the UL channel parameters, such as Doppler frequency shift and angle-delay vector, are obtained at the BS and the UEs. These parameters will be used in the following section to facilitate the DL training.

IV. DOWNLINK TRAINING AND CHANNEL PREDICTION

Our CSI acquisition framework relies on channel parameters estimated from the UL channel samples. In Section III, we have obtained the UL angle-delay vectors \mathbf{q}_i and the Doppler frequency $z_m^u(i)$. In this section, we introduce the JADD pilot precoding scheme based on the extracted channel parameters and the DL channel reconstruction procedure.

A. Extract parameters from uplink channel parameters

Since the UL and DL are operating in different frequency bands, the angle-delay vectors and the Doppler shifts obtained from the UL channel samples have to adapt to the DL frequency band. Define the selected UL angle-delay vectors as

$$\mathbf{u}_j = \{\mathbf{q}_i | i = i_{s_j}, j \in \{1, 2, \dots, N_s\}\}, \quad (32)$$

where the index i_{s_j} denotes the j -th index in the UL angle-delay vector index set \mathcal{S} . In order to transform the UL angle-delay vector to the DL one, we introduce Proposition 1.

Proposition 1 *The DL angle-delay vector \mathbf{d}_j is obtained from the UL angle-delay vector \mathbf{u}_j by*

$$\mathbf{d}_j = (\mathbf{I}_{N_f} \otimes \mathbf{R}(\theta_j^d, \phi_j^d)) \mathbf{u}_j, j \in \{1, 2, \dots, N_s\}, \quad (33)$$

where $\mathbf{R}(\theta_j^d, \phi_j^d) = \mathbf{R}_h(\theta_j^d, \phi_j^d) \otimes \mathbf{R}_v(\theta_j^d)$.

Proof: Please refer to Appendix A. \square

Proposition 1 demonstrates how to acquire the DL angle-delay vector from the UL ones when the UE is equipped with single antenna. In practice, the UEs may have dual-polarized antennas. The generalization of our method is straightforward, as shown in Remark 1.

Remark 1 *If the UEs are equipped with dual-polarized antennas, the DFT matrix $\mathbf{W}(N_t)$ becomes*

$$\mathbf{W}(N_t) = \begin{bmatrix} \mathbf{W}(N_h) \otimes \mathbf{W}(N_v) & \\ & \mathbf{W}(N_h) \otimes \mathbf{W}(N_v) \end{bmatrix}.$$

Thus, the j -th DL angle-delay vector is now

$$\mathbf{d}_j = \left(\mathbf{I}_{N_f} \otimes \begin{bmatrix} \mathbf{R}(\theta_j^d, \phi_j^d) & \\ & \mathbf{R}(\theta_j^d, \phi_j^d) \end{bmatrix} \right) \mathbf{q}_i. \quad (34)$$

Then we calculate the DL Doppler frequency shift with the poles $z_m^u(i)$ obtained from the UL channel samples

$$e^{jw_m^d(j)} = e^{j \frac{\arccos\left(\Re\left\{\frac{z_m^u(i)}{|z_m^u(i)|}\right\}\right) f^d}{f^u}}. \quad (35)$$

B. DL pilot precoding and CSI reconstruction

With the angle-delay vectors and Doppler frequency shifts of the DL channel, we may reconstruct the DL CSI as

$$\tilde{\mathbf{h}}^d(t) = \sum_{j=1}^{N_s} \sum_{m=1}^M a_m^d(j) e^{jw_m^d(j)t} \mathbf{d}_j, \quad (36)$$

where $a_m^d(j)$ is the m -th complex amplitude corresponding to \mathbf{d}_j . In order to reconstruct the DL channel, $a_m^d(j)$ has to be estimated. We propose to do so with JADD precoded pilot signals. The proposed precoding matrix also helps to reduce the training overhead by exploiting the sparse structure of $\tilde{\mathbf{h}}^d(t)$.

The vectorized DL channel $\tilde{\mathbf{h}}^d(t)$ can be decomposed to three matrices as

$$\tilde{\mathbf{h}}^d(t) = \mathbf{D} \mathbf{E}(t) \mathbf{a}^d. \quad (37)$$

The DL angle-delay vector matrix $\mathbf{D} \in \mathbb{C}^{N_f N_t \times N_s}$ is

$$\mathbf{D} = [\mathbf{d}_1 \quad \mathbf{d}_2 \quad \dots \quad \mathbf{d}_{N_s}]. \quad (38)$$

The Doppler matrix $\mathbf{E}(t) \in \mathbb{C}^{N_s \times N_s M}$ is defined as

$$\mathbf{E}(t) = \begin{bmatrix} \mathbf{e}_1(t) & & & \\ & \mathbf{e}_2(t) & & \\ & & \ddots & \\ & & & \mathbf{e}_{N_s}(t) \end{bmatrix}, \quad (39)$$

where

$$\mathbf{e}_j(t) = [e^{jw_1^d(j)t} \quad e^{jw_2^d(j)t} \quad \dots \quad e^{jw_M^d(j)t}]. \quad (40)$$

The DL complex amplitude vector $\mathbf{a}^d \in \mathbb{C}^{N_s M \times 1}$ is

$$\mathbf{a}^d = [\mathbf{a}^d(1) \quad \mathbf{a}^d(2) \quad \cdots \quad \mathbf{a}^d(N_s)]^T, \quad (41)$$

where

$$\mathbf{a}^d(j) = [a_1^d(j) \quad a_2^d(j) \quad \cdots \quad a_M^d(j)]. \quad (42)$$

Using (37), we can design a precoding matrix to facilitate DL pilot training. This matrix is constructed based on the DL Doppler frequency shifts and angle-delay vectors. Traditionally, the idea of precoding is in spatial domain, where the signal is combined in the air from a receiver point of view. However in our scheme, the joint spatial-frequency precoding is a generalized wideband concept. Essentially, the training signal is combined in spatial domain at the BS side, and then combined in frequency domain at the UE side [36]. Denote the precoding matrix by $\mathbf{F}(t) \in \mathbb{C}^{N_t N_f \times N_s M}$:

$$\mathbf{F}(t) = [\mathbf{f}_1(t) \quad \mathbf{f}_2(t) \quad \cdots \quad \mathbf{f}_{N_s M}(t)]. \quad (43)$$

Each column of $\mathbf{F}(t)$, e.g., $\mathbf{f}_n(t) \in \mathbb{C}^{N_t N_f \times 1}$, is composed of the precoding vectors applied on all N_f subcarriers:

$$\mathbf{f}_n(t) = [\mathbf{f}_n(t, f_1)^T \quad \mathbf{f}_n(t, f_2)^T \quad \cdots \quad \mathbf{f}_n(t, f_{N_f})^T]^T,$$

where $\mathbf{f}_n(t, f_l) \in \mathbb{C}^{N_t \times 1}$ is the precoder for the l -th subcarrier in the wideband precoder $\mathbf{f}_n(t)$.

Denote the pilot matrix \mathbf{S} by

$$\mathbf{S} = [\mathbf{s}_1^T \quad \mathbf{s}_2^T \quad \cdots \quad \mathbf{s}_{N_s M}^T]^T, \mathbf{s}_n \in \mathbb{C}^{1 \times \tau}. \quad (44)$$

where τ is the length of pilot sequence. Then the transmitted pilot sequence at the l -th subcarrier by the BS is

$$\mathbf{g}^d(t, f_l) = \sum_{n=1}^{N_s M} \mathbf{f}_n(t, f_l) \mathbf{s}_n. \quad (45)$$

The received pilot signal by the UE at the l -th subcarrier is

$$\mathbf{x}^d(t, f_l) = \mathbf{h}^d(t, f_l)^T \mathbf{g}^d(t, f_l) + \mathbf{n}(t, f_l), \quad (46)$$

where $\mathbf{h}^d(t, f_l)$ denotes the DL channel at the l -th subcarrier and $\mathbf{n}(t, f_l)$ is the noise at the l -th subcarrier. The UE makes a summation over all subcarriers as

$$\mathbf{y}^d(t) = \sum_{l=1}^{N_f} \mathbf{x}^d(t, f_l) + \mathbf{n}(t). \quad (47)$$

The above-mentioned joint spatial-frequency precoding of the training signal can also be written in matrix form as

$$\mathbf{y}^d(t) = \mathbf{h}^d(t)^T \mathbf{F}(t) \mathbf{S} + \mathbf{n}(t). \quad (48)$$

In the following, we devise our precoding matrix $\mathbf{F}(t)$. In our framework, the DL channel is reconstructed as (37). Thus (48) is written as

$$\tilde{\mathbf{y}}^d(t) = \left((\mathbf{a}^d)^T \mathbf{E}(t)^T \mathbf{D}^T \mathbf{F}(t) \right) \mathbf{S} + \mathbf{n}(t). \quad (49)$$

The Gaussian noise vector $\mathbf{n}(t) \in \mathbb{C}^{1 \times \tau}$ has a distribution of $\mathbf{n}(t) \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$, where σ^2 is the noise power. Our purpose is to estimate the coefficient vector \mathbf{a}^d and feed it back to the BS. Notice that $\mathbf{E}(t)^T \mathbf{D}^T$ has a rank of N_s and has no right inverse matrix. Obviously, $\mathbf{E}(t)^T$ is of full column rank

and \mathbf{D}^T is of full row rank. Thus, there exists a right inverse matrix of \mathbf{D}^T , however, no right inverse matrix of $\mathbf{E}(t)^T$. The Moore-Penrose matrix of $\mathbf{E}(t)^T$ is introduced instead and the precoding matrix is designed as

$$\mathbf{F}(t) = (\mathbf{D}^T)^\dagger \left(\mathbf{E}(t)^T \right)^\dagger. \quad (50)$$

Substitute $\mathbf{F}(t)$ with (50) and (49) becomes

$$\tilde{\mathbf{y}}^d(t) = \left((\mathbf{a}^d)^T \mathbf{E}(t)^T \left(\mathbf{E}(t)^T \right)^\dagger \right) \mathbf{S} + \mathbf{n}(t). \quad (51)$$

After applying the precoding matrix, the dimension of \mathbf{S} in (49) reduces to $\mathbb{C}^{N_s M \times \tau}$. Due to the channel sparsity in angle-delay domain and a small M , the precoding matrix also reduces the training overhead, which does not scale with the number of BS antennas and the bandwidth. In principle we should guarantee $\tau \geq N_s M$. For simplicity, the length of the training sequence τ satisfies $\tau = N_s M$ and \mathbf{S} is designed as a unitary matrix. Based on (51), the unknown parameter \mathbf{a}^d can be obtained by least-square (LS) estimation

$$\hat{\mathbf{a}}^d = \left(\mathbf{S}^T \mathbf{E}(t)^\dagger \mathbf{E}(t) \right)^\dagger \tilde{\mathbf{y}}^d(t)^T. \quad (52)$$

The UEs should feed back the estimated complex coefficient vector $\hat{\mathbf{a}}^d$ to the BS. Therefore, the DL channel after a T_d CSI delay can be easily reconstructed at the BS as

$$\tilde{\mathbf{h}}^d(t + T_d) = \sum_{j=1}^{N_s} \sum_{m=1}^M \hat{a}_m^d(j) e^{j w_m^d(j)(t+T_d)} \mathbf{d}_j. \quad (53)$$

The reconstructed DL channel will be utilized in the downlink precoding for data transmission.

V. PERFORMANCE ANALYSIS

In our framework, the choice of N_s and L affect the channel prediction performance and the computation complexity. Hence, in this section, we focus on analyzing the impact of N_s and L on the prediction performance, the computational complexity, and the feedback overhead.

A. Channel prediction performance analysis

The DL channel prediction error is defined with the normalized mean square error (NMSE) metric as

$$\varepsilon \triangleq 10 \log \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t + T_d) - \tilde{\mathbf{h}}^d(t + T_d)}{\mathbf{h}^d(t + T_d)} \right\|_2^2 \right\}. \quad (54)$$

Define $t_p = t + T_d$ as the channel prediction offset. We revisit the DL channel reconstruction equation (37) and substitute $\mathbf{a}^d(t)$ with (52)

$$\tilde{\mathbf{h}}^d(t_p) = \tilde{\mathbf{h}}_1^d(t_p) + \tilde{\mathbf{h}}_2^d(t_p),$$

where

$$\tilde{\mathbf{h}}_1^d(t_p) = \mathbf{D} \mathbf{E}(t_p) (\mathbf{D} \mathbf{E}(t_p))^\dagger \mathbf{h}^d(t_p) = \mathbf{D} \mathbf{D}^\dagger \mathbf{h}^d(t_p), \quad (55)$$

$$\tilde{\mathbf{h}}_2^d(t_p) = \mathbf{D} \mathbf{E}(t_p) (\mathbf{S}^\dagger)^T \mathbf{n}(t_p)^T. \quad (56)$$

Then (54) becomes

$$\varepsilon \triangleq 10 \log \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p) - \tilde{\mathbf{h}}_2^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\}. \quad (57)$$

The following theorem gives the lower bound of the DL channel prediction error, which is derived by letting N_s take the maximum value, i.e., $N_s = N_f N_t$.

Theorem 1 *The lower bound of the DL channel prediction error of the proposed CSI acquisition framework is*

$$\varepsilon = 10 \log \left(\frac{\sigma^2 N_s M}{\mathbb{E} \|\mathbf{h}^d(t_p)\|_2^2} \right). \quad (58)$$

Proof: Please refer to Appendix B. \square

Theorem 1 gives the lower bound of the channel prediction error when all the angle-delay vectors in \mathbf{Q} are taken into account. This condition may not be easy to achieve due to the huge feedback overhead and high complexity. Fortunately, the sparsity of multipath angles and delays ensures a much smaller N_s in our framework. Another important parameter is the prediction order L , which may remain small in wideband massive MIMO regime, as shown in the following theorem. First define N_P as the number of non-identical angle-delay structures of all the DL paths.

Theorem 2 *When $L = 1, N_L = 2, N_s = N_P$, the DL channel prediction error converges to zero as the number of BS antennas and bandwidth increase*

$$\lim_{N_t, N_f \rightarrow \infty} \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} = 0. \quad (59)$$

Proof: Please refer to Appendix C. \square

Theorem 2 gives an asymptotic channel prediction performance of our framework. When the number of antennas and bandwidth are finite, we introduce Remark 2 for choosing a proper prediction order L .

Remark 2 *Given any L satisfying $M \leq L \leq N_L - L$, the channel prediction error yields*

$$\begin{aligned} & \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} - \mathbb{E} \left\{ \frac{N_s M \sigma^2}{\|\mathbf{h}^d(t_p)\|_2^2} \right\} \leq \varepsilon \\ & \leq \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} + \mathbb{E} \left\{ \frac{N_s M \sigma^2}{\|\mathbf{h}^d(t_p)\|_2^2} \right\} \end{aligned} \quad (60)$$

Proof: Please refer to Appendix D. \square

We notice that the difference between the upper bound and lower bound of the channel prediction error is $\mathbb{E} \left\{ \frac{2N_s M \sigma^2}{\|\mathbf{h}^d(t_p)\|_2^2} \right\}$, which is scaling with $1/\text{SNR}$ and is very small when the number of antennas and the bandwidth are large. Normally M cannot be known in advance and we assume $M = L$ on noise-free channel sample condition and $M < L$ on noisy channel sample condition, respectively. Therefore, greater L cannot bring significant performance improvement. Remark 2

indicates that we should choose as small L as possible for a given M and N_L . This observation is also confirmed in simulation of Sec. VI.

However, the limited number of antennas causes DFT mismatch problem. Thus, each \mathbf{d}_j cannot accurately map the exact angle-delay structure of the DL channel. Therefore, each angle-delay vector may correspond to multiple Doppler frequency shifts. Bigger L may better fit the corresponding Doppler frequency of each angle-delay vector. Thus, there lies a trade-off in the choice of L . Since the diversity of Doppler frequency shift cannot be known apriori in realistic applications, the optimal L is difficult to obtain. Thus a relatively small L satisfying $M \leq L \leq N_L - L$ is recommended.

B. Complexity and feedback overhead analysis

Our DL channel reconstruction framework consists of five parts, i.e., the DFT projection, the MP based Doppler estimation, the UL to DL transformation of angle-delay vectors and Dopplers, the DL training, and channel reconstruction. The DFT projection can be realized with fast Fourier transform (FFT), which has a complexity of $\mathcal{O}(N_f N_t \log_2(N_f N_t))$. The complexity of MP method is mainly the SVD, i.e., $\mathcal{O}((N_L - L)^2 L + (N_L - L) L^2)$. The complexity of parameter transformation procedure is $\mathcal{O}(N_s N_v + N_s N_h) + \mathcal{O}(N_f^2 N_t^2) + \mathcal{O}(N_s M \log^2(N_s M)) + \mathcal{O}(N_f N_t N_s^2 M^2)$. The DL training contains a matrix inversion and the complexity is $\mathcal{O}(N_f N_t N_s M)$. The channel reconstruction entails a matrix inversion and SVD which have the complexity of $\mathcal{O}(N_s^3 M + N_s^3 M^2) + \mathcal{O}(N_s^3 M^2) + \mathcal{O}(N_s^2 M^2) + \mathcal{O}((N_s M)^{2.37})$. The overall complexity of the channel reconstruction procedure is thus $\mathcal{O}(N_f N_t N_s^2 M^2)$. Obviously, our framework is of polynomial complexity and requires no iterative computing like CS methods or machine learning methods.

The feedback overhead is now analyzed for a given channel coherence time T_c . In Enhanced Type II codebook [37], the feedback overhead scales with $L_1 L_2$, where L_1, L_2 are smaller than N_t, N_f . In classical CS methods like [38], the feedback overhead depends on the reduced dimension N_r of the channel and scales with $N_r N_f$. In other methods like NOMP [12] and deep learning [13], the feedback overhead depends on the number of paths L_p , which is large in rich scattering environments. And for a wideband system, the feedback overhead of these methods scales with $L_p N_f$. Thanks to the channel prediction capability, the feedback overhead of our framework is $N_s M / N_c$ scalars for one channel coherence time T_c , where $N_c \geq 1$ means only one set of $N_s M$ feedback coefficients is required for a time interval of $N_c T_c$. Hence, our framework has the advantage of reduced feedback over these traditional methods.

VI. NUMERICAL RESULTS

In this section, we validate the proposed JADD framework with the industrial channel model of the cluster-delay-line-A (CDL-A) defined by 3GPP [30] in a rich scattering scenario. Unless particularly specified, CDL-A channel model contains

TABLE I
SYSTEM PARAMETERS IN SIMULATIONS

| Physical meaning | Default value |
|--------------------------------|---|
| Channel model | CDL-A |
| Bandwidth | 20 MHz |
| UL carrier frequency | 1.92 GHz |
| DL carrier frequency | 2.11 GHz |
| Subcarrier spacing | 30 kHz |
| Resource block | 51 |
| Angle spread RMS | (87.1°, 33.6°, 102.1°, 24.7°) |
| Delay spread | 300 ns |
| Number of paths | 460 |
| Transmit antenna configuration | $(N_v, N_h, P_t) = (2, 8, 2)$, polarization direction are 0°, 90° |
| Receive antenna configuration | $(N_v, N_h, P_t) = (1, 1, 2)$, polarization direction are ±45° |
| Slot duration | 0.5 ms |
| Number of UEs | 8 |

a total of 23 clusters with 20 paths inside each cluster. Following the n65 new radio (NR) band in [39], the UL center frequency is 1.92 GHz and the DL center frequency is 2.11 GHz. The bandwidth of UL and DL are both 20 MHz with a 30 kHz subcarrier spacing, implying that 51 resource blocks (RBs) are available per time slot. In this configuration, each time slot contains 14 OFDM symbols and is as short as 0.5 ms, which denotes the SRS signal cycle length. The BS antenna configuration is $(N_v, N_h, P_t) = (2, 8, 2)$, where P_t is the number of polarizations for each antenna element. The spacing between the antennas in vertical direction and horizontal direction are both $0.5c/f^d$. Table I gives all the other parameters used in our simulation unless otherwise specified. The DL precoding process is the Eigen Zero Forcing (EZF) [40] and the UEs apply Minimum Mean Square Error-Interference Rejection Combining (MMSE-IRC) receiver. The performance of our framework is shown in two metrics, the SE and the prediction error (PE). The PE is defined by (54). The spectral efficiency R_s is calculated over a period of time and all subcarriers by

$$R_s = \mathbb{E} \left\{ \sum_{k=1}^K \log \left(1 + \frac{\|\bar{\mathbf{h}}_k^d(t, f) \mathbf{G}_k(t, f)\|_2^2}{\sigma_k^2 + \sum_{j \neq k} \|\bar{\mathbf{h}}_j^d(t, f) \mathbf{G}_j(t, f)\|_2^2} \right) \right\},$$

where $\mathbf{G}_k(t, f)$ is the precoding matrix and $\bar{\mathbf{h}}_k^d(t, f)$ is the estimated channel. σ_k^2 is the noise power at UE k .

Three baseline schemes are introduced as the benchmarks. All baselines follow the same channel parameters in Table I. The curves labeled with ‘‘Enhanced Type II with perfect CSI’’ are the performances of Enhanced Type II codebook where perfect CSI is known by the UEs and there is no CSI delay. The first baseline is Enhanced Type II codebook, yet only delayed CSI is known by the UEs. The second baseline is utilizing an adaptive and parameter free recurrent neural structure (APF-RNS) based on deep learning [41] for real-time prediction. In this method, the DL channel is predicted according to the temporal correlation with the recent history DL channel data with CSI delay, however, without any CSI compression or quantization. The APF-RNS network structure

follows the configuration in [41], and adopts 32 long-short term memory (LSTM) units as the hidden layer. In addition, the history channel data for training and testing is generated by the same channel model with parameters of Table I. We perform the online training of APF-RNS with the known length of 20 and the prediction length of 1. The last baseline is a traditional CS method, called TVAL3 [42]. We apply this method to compress the dimension of the DL channel and recover it through a TVAL3 solver. The basic parameter setting is following [43] and the compression ratio is $\rho = 1/4$. Our proposed scheme is referred to as JADD scheme in our simulations.

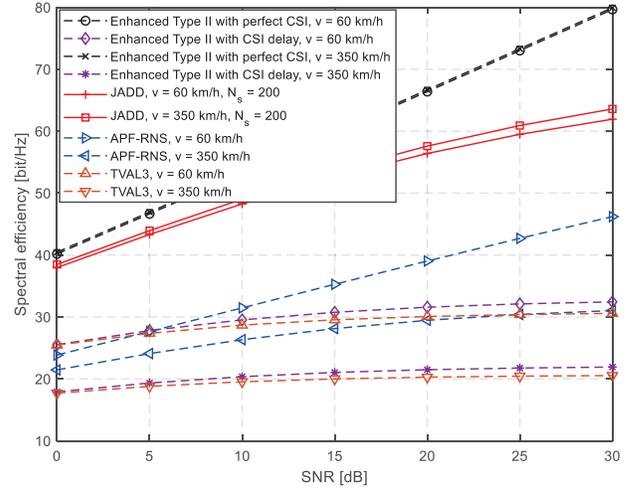


Fig. 3. SE performance vs. SNR, different UE speeds, noise-free channel samples, $L = 2$, $T_d = 5$ ms.

In Fig. 3, our framework is evaluated with the SE metric. In both high speed (350 km/h) and low speed (60 km/h), our JADD outperforms Enhanced Type II, APF-RNS and TVAL3, which demonstrate the superiority of our scheme in different mobility scenarios. We also conclude that CSI delay causes severe performance dropping in all baselines, especially in high mobility scenario. Note that there exists a small SE difference between the cases when the speed of UE is 60 km/h and 350 km/h. This phenomenon is caused by the different angle-delay sparsity and the corresponding Doppler frequency of the channel under different mobility scenarios.

Fig. 4 shows the PE performance under different CSI delay and different mobility levels. Following the coefficient quantization method in 5G [37], we use a C_a -bit geometric sequence codebook and a C_p -bit geometric sequence codebook to quantize the amplitude and the phase of the feedback coefficients, respectively.

Because of the CSI delay, the PE of Enhanced Type II is unsatisfactory, especially in high mobility scenarios. Compared with Enhanced Type II, our JADD can overcome the performance degradation brought by the CSI delay since our novel channel reconstruction framework (53) can well predict the DL channel. We may conclude that our scheme can well adapt to different CSI delay even with quantization errors of feedback coefficients.

Fig. 5 demonstrates the SE performance under different N_s values while the performance of our framework always

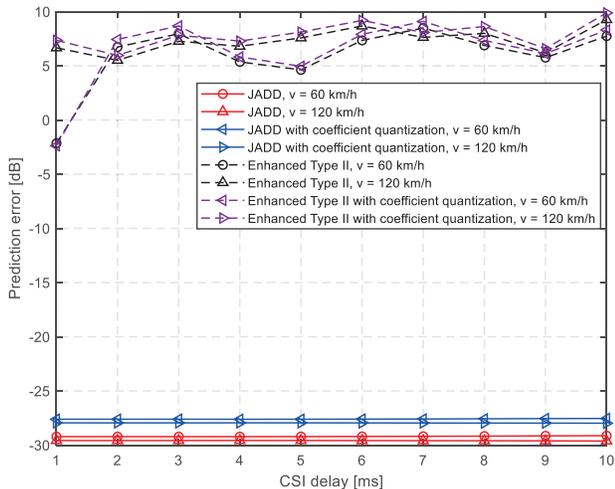


Fig. 4. PE performance vs. CSI delay, noise-free channel samples, feedback coefficients quantized, $N_s = 200$, $L = 2$, $C_a = 4$, $C_p = 6$.

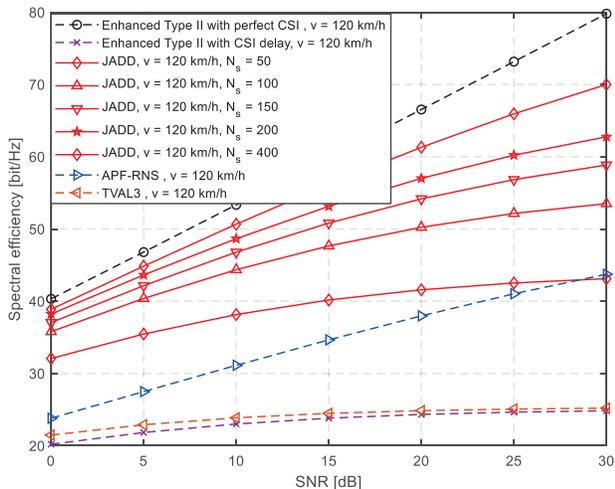


Fig. 5. SE performance vs. SNR, different N_s , noise-free channel samples, $L = 2$, $T_d = 5$ ms.

surpasses the three baselines even with a small N_s . The SE of JADD quickly increases with N_s .

Fig. 6 shows the PE performance of our method under different prediction order L . The results show that the value of L has little impact on the PE performance, which is aligned with Remark 2. We notice that the PE with coefficient quantizations varies a little with the value of L . This phenomenon is reasonable. The quantization error results in the failure of Remark 2 and bigger L leads to higher Doppler frequency resolution, hence better PE performance. However, in Fig. 6, bigger L leads to only a small improvement of the PE with quantization but much heavier computation complexity. Therefore, in realistic application, a small L is still preferable considering the trade-off between the complexity and prediction error.

In the follow, we focus on evaluating the robustness of our method in the cases of different antenna configurations, different channel models and inaccurate CSI samples.

The previous numerical results are based on the same antenna configuration and Fig. 7 shows the SE performance

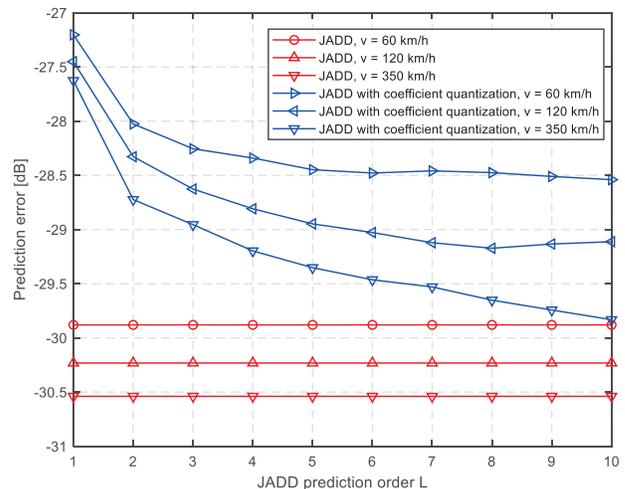


Fig. 6. PE performance vs. prediction order L , noise-free channel samples, feedback coefficients quantized, $N_s = 200$, $T_d = 5$ ms, $C_a = 4$, $C_p = 6$.

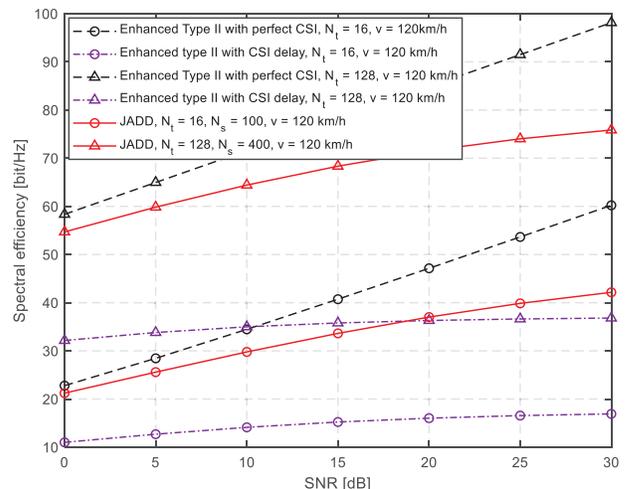


Fig. 7. SE performance vs. SNR, different BS antenna configurations, noise-free channel samples, $L = 2$, $\eta = 0.99$, $T_d = 5$ ms.

of JADD under different BS antenna configurations. Note that the value of N_s should be carefully chosen to assure (19) under different N_t configuration given the threshold η . The results show that the SE performance of our framework always outperforms the Enhanced Type II codebook with CSI delay.

In a rich scattering environment, our framework performs well as discussed above. In fact, different scattering environments render different physical features of the channel. The channel model of CDL-D [30] which contains a line of sight (LOS) path is also considered. The numerical result is demonstrated in Fig. 8. Our framework still performs well in this case. We notice that the same N_s in CDL-D achieves better SE performance as in CDL-A. This phenomenon is reasonable because CDL-D channel has higher angle-delay sparsity.

So far, the previous numerical results are achieved under the noise-free channel sample condition. Now we show the performance of JADD under noisy channel sample case. The channel sample noise is modeled as i.i.d. Gaussian and the noise power is characterized by channel sample SNR.

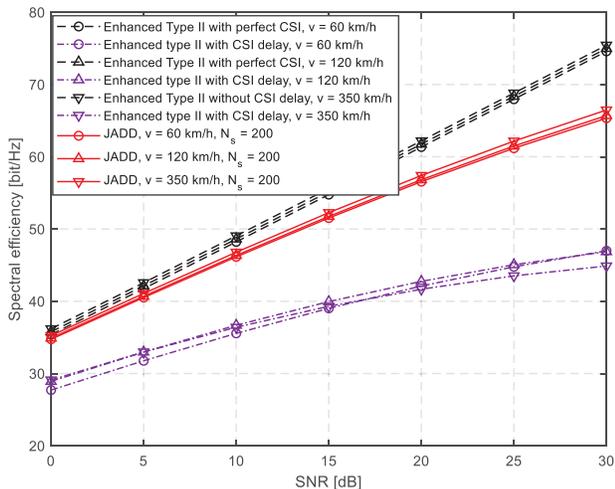


Fig. 8. SE performance under CDL-D channel model, noise-free channel samples, $L = 2$, $T_d = 5$ ms.

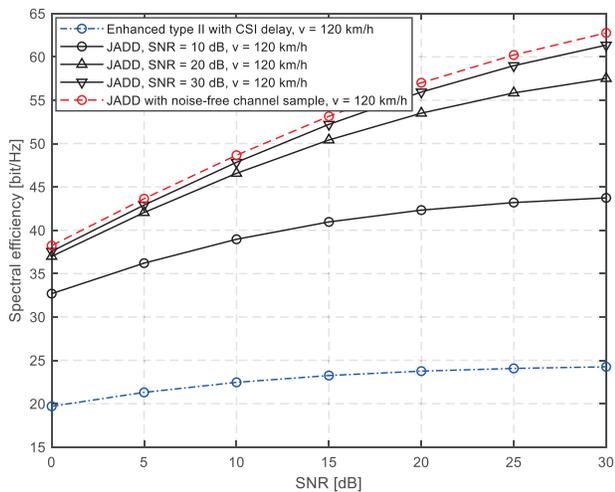


Fig. 9. SE performance vs. SNR, noisy channel samples with sampling SNR = {10 dB, 20 dB, 30 dB}, $N_s = 200$, $L = 2$, $T_d = 5$ ms.

Fig. 9 shows the SE performance of our framework under different channel sample SNRs. Our framework outperforms the Enhanced Type II codebook with CSI delay. Therefore, our scheme is robust to noisy channel samples.

VII. CONCLUSION

In this paper we proposed a novel channel prediction framework to address the curse of mobility in FDD massive MIMO, which suffers from the problems of both the CSI aging and large training overhead. Our framework combined the merits of partial channel reciprocity in FDD and the angle-delay-Doppler structure of the multipath channel. The DL channel was reconstructed with parameters extracted from the UL channel parameters and some coefficients fed back from the UEs. In particular, the BS calculates the angle-delay vectors of the UL channel and estimates the Doppler frequency shifts using the MP method. A wideband JADD precoding matrix was then proposed to facilitate the acquisition of the desired coefficients, and meanwhile, reduce the training overhead. At the UE side, only some scalar coefficients were computed

and fed back to the BS. Our asymptotic analysis showed the prediction error converges to zero as the number of BS antennas increases while only two UL channel samples are available. A scheme to choose a proper prediction order was also discussed. The numerical results demonstrated that our framework works well even in high-mobility scenarios with large CSI delays.

APPENDIX A PROOF OF PROPOSITION 1

Proof: First we need to derive the relationship between i and j . The i -th column of DFT matrix is

$$\mathbf{q}_i = \left[\mathbf{w}_i \quad e^{-j\frac{2\pi}{N_f}i} \mathbf{w}_i \quad \dots \quad e^{-j\frac{2\pi}{N_f}(N_f-1)i} \mathbf{w}_i \right]^T, \quad (61)$$

where $\mathbf{w}_i = \left[1 \quad e^{-j\frac{2\pi}{N_t}i} \quad \dots \quad e^{-j\frac{2\pi}{N_t}i(N_t-1)} \right]^T$. Comparing \mathbf{q}_i in (61) and \mathbf{r}_p in (12), angle information θ_p^u, ϕ_p^u and delay information τ_p^u are closely related to elements in \mathbf{q}_i . Each angle-delay vector index i corresponds to an index combination (i^a, i^h, i^v) , where i^a denotes the angle index, i^h denotes the horizontal angle index and i^v denotes the vertical angle index. The relationship between i^a and i is

$$i^a = \begin{cases} \text{mod } \frac{i}{N_t}, \text{mod } \frac{i}{N_t} \neq 0, \\ N_t, \text{mod } \frac{i}{N_t} = 0. \end{cases} \quad (62)$$

Then, the horizontal and the vertical angle index are calculated as

$$i^h = \begin{cases} \frac{(i^a - i^v)}{N_v} + 1, i^v \neq 0, \\ \frac{i^a}{N_v}, i^v = 0, \end{cases} \quad (63)$$

$$i^v = \begin{cases} \text{mod } \frac{i^a}{N_v}, \text{mod } \frac{i^a}{N_v} \neq 0, \\ N_v, \text{mod } \frac{i^a}{N_v} = 0. \end{cases} \quad (64)$$

Thus, each UL angle-delay i map a DL channel angular index j , $j \in \{1, 2, \dots, N_s\}$ as

$$\sin \theta_j \sin \phi_j = \frac{i^h c}{l_h f^u N_h}, \quad \cos \theta_j = \frac{c i^v}{l_v f^u N_v}. \quad (65)$$

According to (7), the DL angle-delay structure can be calculated as

$$\begin{aligned} \mathbf{r}^d(\theta_j, \phi_j) &= (\mathbf{I}_{N_f} \mathbf{c}(\tau_i^u)) \otimes ((\mathbf{R}_h(\theta_j, \phi_j) \otimes \mathbf{R}_v(\theta_j)) \boldsymbol{\alpha}^u(\theta_i, \phi_i)) \\ &= (\mathbf{I}_{N_f} \otimes (\mathbf{R}_h(\theta_j, \phi_j) \otimes \mathbf{R}_v(\theta_j))) \cdot (\mathbf{c}(\tau_i^u) \otimes \boldsymbol{\alpha}^u(\theta_i, \phi_i)). \end{aligned}$$

We use the angle-delay vector to approximate the DL angle-delay structure

$$\mathbf{d}_j \triangleq (\mathbf{I}_{N_f} \otimes (\mathbf{R}_h(\theta_j, \phi_j) \otimes \mathbf{R}_v(\theta_j))) \mathbf{u}_j. \quad (66)$$

Then the Proposition 1 is proved. \square

APPENDIX B PROOF OF THEOREM 1

Proof: The vectorized DL channel can be estimated by a series of angle-delay vectors superposition mapped by the columns of DFT matrix \mathbf{Q} . As N_s increases, angle-delay resolution improves. The power leakage problem alleviates, thus, ε decreases.

When $N_s = N_f N_t$, $\mathbf{D}\mathbf{D}^\dagger = \mathbf{I}_{N_f N_t}$ holds. The prediction error (54) becomes

$$\begin{aligned} \varepsilon &= 10 \log \left(\frac{\mathbb{E} \left\| \mathbf{h}^d(t_p) - \mathbf{h}^d(t_p) \mathbf{I} - \tilde{\mathbf{h}}_2^d(t_p) \right\|_2^2}{\mathbb{E} \left\| \mathbf{h}^d(t_p) \right\|_2^2} \right) \\ &= 10 \log \left(\frac{\mathbb{E} \left\{ \tilde{\mathbf{h}}_2^d(t_p) \tilde{\mathbf{h}}_2^d(t_p)^H \right\}}{\mathbb{E} \left\| \mathbf{h}^d(t_p) \right\|_2^2} \right) \\ &= 10 \log \left(\frac{\mathbb{E} \left\| \mathbf{D}\mathbf{E}(t_p) (\mathbf{S}^\dagger)^T \right\|_2^2 \mathbb{E} \left\| \mathbf{n}(t_p)^T \right\|_2^2}{\mathbb{E} \left\| \mathbf{h}^d(t_p) \right\|_2^2} \right). \end{aligned} \quad (67)$$

The training sequence matrix is denoted in column vectors and row vectors form

$$\begin{aligned} (\mathbf{S}^\dagger)^T &= \begin{bmatrix} \mathbf{s}_{1,1} & \mathbf{s}_{1,2} & \cdots & \mathbf{s}_{N_s, M} \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{s}(1) & \mathbf{s}(2) & \cdots & \mathbf{s}(N_s M) \end{bmatrix}, \end{aligned} \quad (68)$$

where $\mathbf{s}_{j,m} = [s_{j,m}(1) \ s_{j,m}(2) \ \cdots \ s_{j,m}(N_s M)]$ and $\mathbf{s}(n) = [s_{1,1}(n) \ s_{1,2}(n) \ \cdots \ s_{N_s, M}(n)]^T$. Calculate the matrix

$$\mathbf{D}\mathbf{E}(t_p) (\mathbf{S}^\dagger)^T = \begin{bmatrix} \sum_{j=1}^{N_s} \mathbf{d}_j \sum_{m=1}^M e^{jw_{j,m}^d(t_p)} s_{j,m}(1) \\ \sum_{j=1}^{N_s} \mathbf{d}_j \sum_{m=1}^M e^{jw_{j,m}^d(t_p)} s_{j,m}(2) \\ \cdots \\ \sum_{j=1}^{N_s} \mathbf{d}_j \sum_{m=1}^M e^{jw_{j,m}^d(t_p)} s_{j,m}(N_s M) \end{bmatrix}^T. \quad (69)$$

Let $\sum_{m=1}^M e^{jw_{j,m}^d(t_p)} s_{j,m}(n) = \lambda_{j,n}(t_p)$ and obtain

$$\begin{aligned} &\left\| \mathbf{D}\mathbf{E}(t_p) (\mathbf{S}^\dagger)^T \right\|_2^2 \\ &= \sum_{n=1}^{N_s M} \left(\left(\sum_{j=1}^{N_s} \lambda_{j,n}(t_p)^H \mathbf{d}_j^H \right) \left(\sum_{j=1}^{N_s} \lambda_{j,n}(t_p) \mathbf{d}_j \right) \right). \end{aligned} \quad (70)$$

The following lemma shows that \mathbf{d}_j has unit norm and mutual orthogonality.

Lemma 2 For any $i, j = 1, \dots, N_s$, we have

$$\mathbf{d}_i^H \mathbf{d}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}. \quad (71)$$

Proof: Eq. (71) is derived as

$$\begin{aligned} \mathbf{d}_i^H \mathbf{d}_j &= \mathbf{q}_i^H \left((\mathbf{I}_{N_f} \otimes (\mathbf{R}_h(\theta_i, \phi_i) \otimes \mathbf{R}_v(\theta_i))) \right)^H \\ &\cdot (\mathbf{I}_{N_f} \otimes (\mathbf{R}_h(\theta_j, \phi_j) \otimes \mathbf{R}_v(\theta_j))) \mathbf{q}_j \\ &= \mathbf{q}_i^H (\mathbf{I}_{N_f} \otimes \mathbf{R}_{i,j}) \mathbf{q}_j, \end{aligned} \quad (72)$$

where

$$\mathbf{R}_{i,j} = \left(\mathbf{R}_h(\theta_i, \phi_i)^H \mathbf{R}_h(\theta_j, \phi_j) \right) \otimes \left(\mathbf{R}_v(\theta_i)^H \mathbf{R}_v(\theta_j) \right).$$

When $i = j$, as \mathbf{q}_i is a column of the DFT matrix \mathbf{Q} , (72) becomes

$$\mathbf{d}_j^H \mathbf{d}_j = \mathbf{q}_i^H (\mathbf{I}_{N_f} \otimes (\mathbf{I}_{N_h} \otimes \mathbf{I}_{N_v})) \mathbf{q}_i = 1. \quad (73)$$

Since $\mathbf{R}_h(\theta, \phi)$ in (8), $\mathbf{R}_v(\theta)$ in (9) are both unitary matrices and $\mathbf{q}_i^H \mathbf{q}_j = 0, \forall i \neq j$, obviously $\mathbf{d}_i^H \mathbf{d}_j = 0, \forall i \neq j$ holds. \square

Using Lemma 2, the following expectation can be easily calculated as

$$\begin{aligned} &\mathbb{E} \left\| \mathbf{D}\mathbf{E}(t_p) (\mathbf{S}(t_p)^\dagger)^T \right\|_2^2 \\ &= \sum_{n=1}^{N_s M} \left(\left(\sum_{j=1}^{N_s} \sum_{m=1}^M e^{-jw_{j,m}^d(t_p)} s_{j,m}(n)^H \right) \right. \\ &\quad \left. \left(\sum_{j=1}^{N_s} \sum_{m=1}^M e^{jw_{j,m}^d(t_p)} s_{j,m}(n) \right) \right) \\ &= \sum_{n=1}^{N_s M} \mathbf{s}(n)^H \mathbf{s}(n) = N_s M. \end{aligned} \quad (74)$$

The equation (67) can be easily calculated as Theorem 1. \square

APPENDIX C PROOF OF THEOREM 2

Proof: When $N_t, N_f \rightarrow \infty$, the following relationship is implicitly holds

$$\tilde{\mathbf{h}}^u(t_p) = \sum_{i=1}^{N_f N_t} \hat{g}_i^u(t_p) \mathbf{q}_i = \sum_{i \in \mathcal{S}} \hat{g}_i^u(t_p) \mathbf{q}_i, |\mathcal{S}| = \mathcal{N}_P, \quad (75)$$

where \mathbf{q}_i is i -th column of the DFT matrix \mathbf{Q} , $\hat{g}_i^u(t_p) = \mathbf{q}_i^H \mathbf{h}^u(t_p)$ and \mathcal{S} is set of indices of which $\hat{g}_i^u(t_p) \neq 0$. Define the space generated by $\mathbf{q}_i \in \mathcal{S}$ as

$$\mathcal{U}_P = \text{span} \{ \mathbf{q}_i : i \in \mathcal{S} \}. \quad (76)$$

Define N_P as the number of non-identical angle-delay structure of the UL channel. The angle-delay structure of two non-intertwined paths p, q are asymptotically orthogonal [15]

$$\lim_{N_t, N_f \rightarrow \infty} \frac{(\mathbf{r}_p^u)^H \mathbf{r}_q^u}{\sqrt{N_t N_f}} = 0. \quad (77)$$

According to [15], the angle-delay structure of two paths p, q holds the orthogonality after projecting to an orthogonal space \mathbf{S} which is a DFT matrix. In fact, the angle-delay structure \mathbf{r}_p^u lies in the space \mathcal{U}_P when $N_t, N_f \rightarrow \infty$. Therefore, the following relationship holds

$$\lim_{N_t, N_f \rightarrow \infty} \frac{\left\| \mathbf{U}_p^H \mathbf{r}_p^u \right\|_2^2}{N_t N_f} = 1, \quad \lim_{N_t, N_f \rightarrow \infty} \frac{\left\| \mathbf{U}_q^H \mathbf{r}_p^u \right\|_2^2}{N_t N_f} = 0, \quad (78)$$

where \mathbf{U}_p is the sub-matrix formed by $\mathbf{q}_i, i \in \mathcal{S}_p$. The index set \mathcal{S}_p is defined by

$$\lim_{N_t, N_f \rightarrow \infty} \left| \frac{q_i^H \mathbf{r}_p^u}{N_t N_f} \right| > 0, i \in \mathcal{S}_p, \quad \lim_{N_t, N_f \rightarrow \infty} \left| \frac{q_i^H \mathbf{r}_p^u}{N_t N_f} \right| = 0, i \notin \mathcal{S}_p.$$

The sub-matrix \mathbf{U}_q and set \mathcal{S}_q are defined likewise. Furthermore, $N_P = \mathcal{N}_P$ due to the orthogonality between \mathcal{S}_p and \mathcal{S}_q .

The condition $L = 1, N_L = 2$ gives the lower bound of matrix pencil parameter configuration, which means that only

one pole needs to be estimated, i.e., $M = 1$. Then, the UL channel (20) becomes

$$\tilde{\mathbf{h}}^u(t_p) = \sum_{i \in \mathcal{S}} a^u(i) z^u(i)^{t_p} \mathbf{q}_i. \quad (79)$$

Comparing to (22), (79) indicates that each angle-delay vector \mathbf{q}_i corresponds to only one Doppler frequency shift $z^u(i)$. Given any N_d the pole $z^u(i)$ is calculated through Algorithm 1, which means the Doppler frequency shift corresponding to \mathbf{q}_i is obtained.

Then we focus on the DL channel analysis. In Lemma 2, we have proved that the DL angle-delay vector \mathbf{d}_j shares the same orthogonality like \mathbf{q}_i , hence, the asymptotic properties (77)-(79) hold for \mathbf{r}_p^d and \mathbf{d}_j . Each \mathbf{d}_j is calculated by (33). Denote the collection of all \mathbf{d}_j as $\mathbf{D}_a \in \mathbb{C}^{N_t N_f \times N_t N_f}$ which is a unitary matrix. Similarly we define \mathcal{D} like the \mathcal{S} in (75) and \mathcal{D}_p like the \mathcal{S}_p in (79), respectively. The condition $N_s = N_P$ makes sure the following relationship holds

$$\|\mathbf{D}_a^H \mathbf{h}^d(t_p)\|_2^2 = \left\| \sum_{j \in \mathcal{D}, p \in \mathcal{D}_p} \beta_p^d e^{-j2\pi f^d \tau_p} e^{jw_p^d t_p} \mathbf{d}_j^H \mathbf{r}_p^d \right\|_2^2.$$

Using (35), the DL Doppler frequency shift satisfies

$$e^{jw^d(j)} = e^{jw_p^d}, j \in \mathcal{D}, p \in \mathcal{D}_p. \quad (80)$$

The asymptotic performance of the DL channel prediction is

$$\begin{aligned} \lim_{N_t, N_f \rightarrow \infty} \varepsilon &= \lim_{N_t, N_f \rightarrow \infty} \frac{\|\mathbf{D}_a (\mathbf{D}_a^H \mathbf{h}^d(t_p) - \mathbf{D}_a^H \tilde{\mathbf{h}}^d(t_p))\|_2^2}{N_t N_f \left(\sum_P |\beta_p^d| \right)^2} \\ &= \lim_{N_t, N_f \rightarrow \infty} \frac{\|\mathbf{D}_a^H \mathbf{h}^d(t_p) - \mathbf{D}_a^H \tilde{\mathbf{h}}^d(t_p)\|_2^2}{N_t N_f \left(\sum_P |\beta_p^d| \right)^2}. \end{aligned} \quad (81)$$

Using the property of norm, we can relax (81)

$$\begin{aligned} \lim_{N_t, N_f \rightarrow \infty} \varepsilon &\leq \lim_{N_t, N_f \rightarrow \infty} \frac{\left\{ \left\| \sum_{j \in \mathcal{D}, p \in \mathcal{D}_p} \beta_p^d e^{-j2\pi f^d \tau_p} e^{jw_p^d t_p} \mathbf{d}_j^H \mathbf{r}_p^d \right\|_2^2 + \left\| \sum_{j \in \mathcal{D}} a^d(j) e^{jw^d(j) t_p} \right\|_2^2 \right\}}{N_t N_f \left(\sum_P |\beta_p^d| \right)^2} \\ &\leq \lim_{N_t, N_f \rightarrow \infty} \frac{\left(\sum_{j \in \mathcal{D}, p \in \mathcal{D}_p} |\beta_p^d| \right)^2 + \left(\sum_{j \in \mathcal{D}} |a^d(j)| \right)^2}{N_t N_f \left(\sum_P |\beta_p^d| \right)^2}. \end{aligned} \quad (82)$$

Notice that $\frac{\left(\sum_{j \in \mathcal{D}, p \in \mathcal{D}_p} |\beta_p^d| \right)^2 + \left(\sum_{j \in \mathcal{D}} |a^d(j)| \right)^2}{\left(\sum_P |\beta_p^d| \right)^2}$ is a constant

which is independent with $N_t N_f$. Therefore, (82) becomes

$$\begin{aligned} \lim_{N_t, N_f \rightarrow \infty} \frac{\|\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}^d(t_p)\|_2^2}{\|\mathbf{h}^d(t_p)\|_2^2} \\ \leq \lim_{N_t, N_f \rightarrow \infty} \frac{\left(\sum_{j \in \mathcal{D}, p \in \mathcal{D}_p} |\beta_p^d| \right)^2 + \left(\sum_{j \in \mathcal{D}} |a^d(j)| \right)^2}{N_t N_f \left(\sum_P |\beta_p^d| \right)^2} = 0, \end{aligned} \quad (83)$$

Then Theorem 2 is proved. \square

APPENDIX D

PROOF OF REMARK 2

Proof: Obviously when $N_s = N_f N_t$, the value of L will not affect ε . We only need to prove the case when N_s satisfies $N_s < N_f N_t$. Clearly $\tilde{\mathbf{h}}_1^d(t_p)$ is independent with M , only $\tilde{\mathbf{h}}_2^d(t_p)$ need to be analyzed. We have proved that $\|\mathbf{D}\mathbf{E}(t_p) (\mathbf{S}^\dagger)^T\|_2^2 = N_s M$ in Appendix B. The expectation part in (57) is calculated as

$$\begin{aligned} &\mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p) - \tilde{\mathbf{h}}_2^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} \\ &\leq \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} + \mathbb{E} \left\{ \left\| \frac{\tilde{\mathbf{h}}_2^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} \\ &\leq \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} + \mathbb{E} \left\{ \left\| \frac{\mathbf{D}\mathbf{E}(t_p) (\mathbf{S}^\dagger)^T \mathbf{n}(t_p)^T}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} \\ &\leq \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} \\ &+ \mathbb{E} \left\{ \left\| \mathbf{D}\mathbf{E}(t_p) (\mathbf{S}^\dagger)^T \right\|_2^2 \left\| \frac{\mathbf{n}(t_p)^T}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} \\ &\leq \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} + \mathbb{E} \left\{ \frac{N_s M \sigma^2}{\|\mathbf{h}^d(t_p)\|_2^2} \right\}. \end{aligned} \quad (84)$$

Similarly, we may derive

$$\begin{aligned} &\mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p) - \tilde{\mathbf{h}}_2^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} \\ &\geq \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} - \mathbb{E} \left\{ \left\| \frac{\mathbf{D}\mathbf{E}(t_p) (\mathbf{S}^\dagger)^T \mathbf{n}(t_p)^T}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} \\ &\geq \mathbb{E} \left\{ \left\| \frac{\mathbf{h}^d(t_p) - \tilde{\mathbf{h}}_1^d(t_p)}{\mathbf{h}^d(t_p)} \right\|_2^2 \right\} - \mathbb{E} \left\{ \frac{N_s M \sigma^2}{\|\mathbf{h}^d(t_p)\|_2^2} \right\}. \end{aligned} \quad (85)$$

Thus Remark 2 is proved. \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, 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, 2013.

- [3] J. Jose, A. Ashikhmin, T. L. Marzetta *et al.*, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.
- [4] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.
- [5] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, 2014.
- [6] J. Flordelis, F. Rusek, F. Tufvesson *et al.*, "Massive MIMO performance—TDD versus FDD: What do measurements say?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2247–2261, 2018.
- [7] A. Adhikary, J. Nam, J. Ahn *et al.*, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.
- [8] Y. Song, C. Liu, Y. Liu *et al.*, "Joint spatial division and multiplexing in massive MIMO: A neighbor-based approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7392–7406, 2020.
- [9] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, 2015.
- [10] B. Wang, M. Jian, F. Gao *et al.*, "Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 5893–5908, 2019.
- [11] H. Lin, F. Gao, S. Jin *et al.*, "A new view of multi-user hybrid massive MIMO: Non-orthogonal angle division multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2268–2280, 2017.
- [12] Y. Han, T.-H. Hsu, C.-K. Wen *et al.*, "Efficient downlink channel reconstruction for FDD multi-antenna systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3161–3176, 2019.
- [13] Y. Han, M. Li, S. Jin *et al.*, "Deep learning-based FDD non-stationary massive MIMO downlink channel reconstruction," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1980–1993, 2020.
- [14] Fraunhofer IIS and Fraunhofer HHI, "RP-191951: Mobility enhancements for MIMO," in *3GPP TSG RAN WG#85*, January 2019, Newport Beach California, USA.
- [15] H. Yin, H. Wang, Y. Liu, and D. Gesbert, "Addressing the curse of mobility in massive MIMO with prony-based angular-delay domain channel predictions," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2903–2917, 2020.
- [16] X. Xia, K. Xu, S. Zhao, and Y. Wang, "Learning the time-varying massive MIMO channels: Robust estimation and data-aided prediction," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8080–8096, 2020.
- [17] R. Wang, O. Renaudin, C. U. Bas *et al.*, "High-resolution parameter estimation for time-varying double directional V2V channel," *IEEE Trans. Veh. Technol.*, vol. 16, no. 11, pp. 7264–7275, 2017.
- [18] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4418–4431, 2012.
- [19] X. Yi, S. Yang, D. Gesbert, and M. Kobayashi, "The degrees of freedom region of temporally correlated MIMO networks with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 494–514, 2014.
- [20] C. Wu, X. Yi, Y. Zhu, W. Wang, L. You, and X. Gao, "Channel prediction in high-mobility massive MIMO: From spatio-temporal autoregression to deep learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1915–1930, 2021.
- [21] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO channel prediction: Kalman filtering vs. machine learning," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 518–528, 2021.
- [22] J. Tan and L. Dai, "Channel feedback in TDD massive MIMO systems with partial reciprocity," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 12 960–12 974, 2021.
- [23] H. Lee, H. Choi, H. Kim, S. Kim, C. Jang, Y. Choi, and J. Choi, "Downlink channel reconstruction for spatial multiplexing in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6154–6166, 2021.
- [24] W. Peng, W. Li, W. Wang, X. Wei, and T. Jiang, "Downlink channel prediction for time-varying FDD massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1090–1102, 2019.
- [25] K. Hugl, K. Kalliolu, J. Laurila *et al.*, "Spatial reciprocity of uplink and downlink radio channels in FDD systems," in *Proc. COST*, vol. 273, no. 2. Citeseer, 2002, p. 066.
- [26] 3GPP, *Study on elevation beamforming / Full-Dimension (FD) Multiple Input Multiple Output (MIMO) for LTE (Release 13)*. Technical Report TR 36.897, available: <http://www.3gpp.org>, 2015.
- [27] D. Fan, F. Gao, G. Wang *et al.*, "Angle domain signal processing-aided channel estimation for indoor 60-GHz TDD/FDD massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1948–1961, 2017.
- [28] Z. Zhong, L. Fan, and S. Ge, "FDD massive MIMO uplink and downlink channel reciprocity properties: Full or partial reciprocity?" in *IEEE Glob. Commun. Conf., GLOBECOM - Proc.*, 2020, pp. 1–5.
- [29] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, 1990.
- [30] 3GPP, *Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)*. Technical Report TR 38.901, available: <http://www.3gpp.org>, 2020.
- [31] —, *NR; Physical channels and modulation (Release 16)*. Technical Report TR 38.211, available: <http://www.3gpp.org>, 2021.
- [32] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, 2002.
- [33] 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, 2013.
- [34] Fraunhofer IIS, Fraunhofer HHI, and Deutsche Telekom, "RP-193072: Measurement results on Doppler spectrum for various UE mobility environments and related CSI enhancements," in *3GPP TSG RAN WG#86*, December 2019, Sitges, Spain.
- [35] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, 1985.
- [36] H. Yin and D. Gesbert, "A partial channel reciprocity-based codebook for wideband FDD massive MIMO," *IEEE Trans. Wireless Commun.*, 2022.
- [37] 3GPP, *NR; Physical layer procedures for data (Release 16)*. Technical Report TR 38.214, available: <http://www.3gpp.org>, 2021.
- [38] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, 2015.
- [39] 3GPP, *NR; Physical layer procedures for data (Release 17)*. Technical Report TR 38.104, available: <http://www.3gpp.org>, 2021.
- [40] L. Sun and M. R. McKay, "Eigen-based transceivers for the MIMO broadcast channel with semi-orthogonal user selection," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5246–5261, 2010.
- [41] Y. Zhu, X. Dong, and T. Lu, "An adaptive and parameter-free recurrent neural structure for wireless channel prediction," *IEEE Trans. Wireless Commun.*, vol. 67, no. 11, pp. 8086–8096, 2019.
- [42] C. Li, W. Yin, H. Jiang, and Y. Zhang, "An efficient augmented lagrangian method with applications to total variation minimization," *Computational Optimization and Applications*, vol. 56, no. 3, pp. 507–530, 2013.
- [43] C. Li, W. Yin, and Y. Zhang, "User's guide for TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms," *CAAM report*, vol. 20, no. 46-47, p. 4, 2009.



Ziao Qin received the B.Sc. degree in Information Engineering from Beijing Institute of Technology, Beijing, China, in 2014. From 2014 to 2017, he works in industry in Beijing. Since 2018, he has been a graduate student at Huazhong University of Science and Technology, Wuhan, China. He is currently pursuing the Ph.D. degree in Information and Communications Engineering. His research interests include channel estimation, signal processing, codebook design, and beamforming for massive MIMO systems.



Haifan Yin received the Ph.D. degree from Télécom ParisTech in 2015. He received the B.Sc. degree in Electrical and Electronic Engineering and the M.Sc. degree in Electronics and Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2012 respectively. From 2009 to 2011, he has been with Wuhan National Laboratory for Optoelectronics, China, working on the implementation of TD-LTE systems as an R&D engineer. From 2016 to 2017, he has been a DSP engineer in Sequans Communications - an

IoT chipmaker based in Paris, France. From 2017 to 2019, he has been a senior research engineer working on 5G standardization in Shanghai Huawei Technologies Co., Ltd., where he made substantial contributions to 5G standards, particularly the 5G codebooks. Since May 2019, he has joined the School of Electronic Information and Communications at Huazhong University of Science and Technology as a full professor. His current research interests include 5G and 6G networks, signal processing, machine learning, and massive MIMO systems. H. Yin was the national champion of 2021 High Potential Innovation Prize awarded by Chinese Academy of Engineering, a winner of 2020 Academic Advances of HUST, and a recipient of the 2015 Chinese Government Award for Outstanding Self-financed Students Abroad.



Weidong Li received the B.Sc. degree in Electronic Information Science and Technology from Nanjing Agricultural University, Nanjing, China, in 2017, and the M.Sc. degree in Electronic Engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2020. He is currently pursuing the Ph.D. degree with the School of Electronic Information and Communications at Huazhong University of Science and Technology, Wuhan, China. His research interests include massive MIMO and signal processing.



Yandi Cao received the B.Sc. degree in Communication Engineering from Chongqing University, Chongqing, China, in 2020. She is currently pursuing the Ph.D. degree with the School of Electronic Information and Communications at Huazhong University of Science and Technology, Wuhan, China. Her research interests include massive MIMO and machine learning.



David Gesbert (Fellow, IEEE) is Director of EURECOM, Sophia Antipolis, France (www.eurecom.fr). He received the Ph.D. degree from TelecomParis, France, in 1997. From 1997 to 1999, he was with the Information Systems Laboratory, Stanford University. He was the Founding Engineer of Iospan Wireless Inc., a Stanford spin off pioneering MIMO-OFDM (currently Intel). Before joining EURECOM in 2004, he was with the Department of Informatics, University of Oslo, as an Adjunct Professor. He has published about 350 articles and 25 patents,

7 of them winning IEEE Best paper awards. He has been the Technical Program Co-Chair for ICC2017 and has been named a Thomson-Reuters Highly Cited Researchers in computer science. He is a Board Member for the OpenAirInterface (OAI) Software Alliance. In 2015, he has been awarded an ERC Advanced Grant. In 2020, he was awarded funding by the French Interdisciplinary Institute on Artificial Intelligence for a Chair in the area of AI for the future IoT. In 2021, he received the Grand Prix in Research from IMT-French Academy of Sciences.