Channel Prediction Based on Non-Uniform Pilot Pattern for Mobile Massive MIMO Scenarios

Yi Shi, Xianling Wang, and Zhiyuan Jiang

Abstract—Massive multiple input multiple output (MIMO) is a broadly used technique that can provide numerous gains in spectral efficiency. However, the degradation of beamforming performance due to outdated channel state information at the transmitter side (CSIT) induced by the mobility of users has been a significant problem waiting to be solved. It is reported that system performance will decrease 50 percent even in a moderate 30 km/h speed scenario. However, the CSI cannot be simply reconstructed through interpolation in high mobility scenarios due to the limitation of pilot density — the phenomenon is known as "Doppler aliasing". To address this, we propose a novel non-uniform pilot pattern that can provide more spectrum resolution compared with the uniform pilot currently used in most communication protocols. Meanwhile, we maintain the density of pilots in order not to sacrifice the payload resources. Based on the novel pilot setting, we propose two-channel prediction schemes with compressive sensing and matrix completion methods. Simulation results show our scheme can outperform deep learning-based and auto-regressive-based methods for about 15 percent in terms of average throughput in the simulated channel generated from the COST2100 channel model. To further verify the applicability, we apply our schemes in real channels measured from a channel sounding campaign, the proposed methods also achieve 5 percent gain which validates their superiority over conventional methods.

Index Terms—Channel aging, channel prediction, compressive sensing, massive MIMO, matrix completion.

I. INTRODUCTION

MASSIVE multiple input and multiple output (MIMO) is a developed technology that has been proven [1] to greatly promote spectral efficiency. It has attracted a lot of attention both academically and industrially and has been applied in many aspects of communication systems, including interference cancellation, multiplexing and etc. The increasing degree of freedom brought by the unprecedented number of the antenna can drastically promote the performance of beamforming. Most of the beamforming techniques [2], [3] assume perfect channel state information at the transmitter side (CSIT) to guarantee the system performance. However, as a matter of fact, the idealism of CSIT can be jeopardized by many factors, the mobility of user ends is one of them,

both in TDD and FDD systems, because the channel will be outdated when the time it is used in downlink beamforming. It is reported that there exists a 50% performance degradation even in 30 km/h mobility scenario [4]. Meanwhile, with the overwhelming need for communication in moderate, even high mobility scenarios, such as Internet of vehicles and high-speed train. This problem should be urgently tackled.

Generally speaking, the channel prediction task of a pilot-assisted system can be logically divided into two parts: prediction of CSI on the future pilot and non-pilot symbols. To the best of our knowledge, most of the proposed channel prediction schemes are focused on addressing the first part and leverage conventional interpolation methods to reconstruct the non-pilot symbols. These works can be generally split into two categories: methods with or without knowing channel temporal correlation function in advance.

A. Methods with Known Temporal Correlation

In the first category, the temporal correlation function is supposed to be known in advance according to the channel model concerned. Among all the considered channel models, the Jakes model is the most frequently used one. It assumes N equal-strength rays arrive at a moving receiver with uniformly distributed angles. In [5], the authors propose a channel predictor based on the Wiener filter with the temporal correlation function derived from the Jakes channel model. Similarly, ref [6], [7] analyze the performance of the Kalman filter on predicting the channel in the massive MIMO case. Ref [8] considers a scenario where the mobility of users are different, resulting in different temporal correlation function and it proposes an adaptive Kalman filter to address this problem. In addition, the influence of pilot contamination is also considered. However, knowing the temporal correlation function in advance is a strict assumption because the Jakes model differs a lot from the real-world channel which significantly degrades the performance when applied in a real channel.

B. Methods with Unknown Temporal Correlation

In the second category, the temporal correlation function is extrapolated from the observed channel instead. Ref [4] proposes an angular-delay domain prediction scheme based on the Prony method, an auto-regressive algorithm, to predict the CSI on dominant angular delay grids. It shows a good performance in terms of pilot symbol prediction both in low and high-mobility scenarios. Similarly, ref [9] utilize the burg method and modified covariance method to calculate the

temporal correlation coefficient. Besides, some other works leverage super-resolution methods to extract temporal-related parameters like the Doppler frequency of each multipath component (MPC) and then reconstruct the future channel. It can predict arbitrary future CSI as long as the parameters do not change. Ref [10] proposes a two-step estimation of signal parameters via rotational invariance technique (ESPRIT) algorithm to extract the delay, Doppler shift, and complex amplitude of each MPC in an orthogonal frequency division multiplexing (OFDM) MIMO system. Ref [11] considers a doubly diverting channel and uses the EMVD method to solve the parameter mismatching problem of ESPRIT. In addition, deep learning methods have also been applied in channel prediction. Ref [12] exploits a combination of convolutional neural network (CNN) and long short term memory RNN (LSTM) to predict future CSI in a 5 minute interval. The user mobility is not considered in this work. Ref [13] also uses LSTM to predict future CSI for vehicular beamtracking, but the acquisition of the CSI on future non-pilot symbols is not mentioned.

Although the methods in the second category do not rely on the prior assumption of the channel temporal correlation, they only evaluate the prediction performance on future pilot symbols and the observed CSI is evenly spaced in the time domain which can be acquired with a uniform pilot pattern. However, if we interpolate the future non-pilot symbols and take the prediction error on them into consideration, the performance will drastically degrade in moderate to high user mobility due to Doppler aliasing caused by the inadequate density of evenly spaced pilot, which will be further elaborated in Section II.

C. Contribution of This Paper

- A novel non-uniform pilot design is proposed to enable the Doppler spectrum sensing in high mobility scenarios, without sacrificing the payload resources compared with a uniform pilot case.
- A model-based prediction scheme based on compressive sensing is proposed. It estimates the sparse Doppler spectrum on each angular-delay domain grid with the help of high resolution enabled by massive MIMO. To address the cyclic predicting problem caused by discrete Fourier transform (DFT) sparsifying matrix, we introduce a redundant dictionary and band-exclusion method to find dominant Doppler peaks. The necessary recovery criteria is then analyzed. Furthermore, to promote stability of recovery, we combine with multiple measurement vector (MMV) and propose Rank-Aware based orthogonal matching pursuit (OMP) with band-excluded redundant dictionary (RAMBLE) algorithm.
- To further mitigate the additional estimation error introduced by parameter extraction, a model-free prediction scheme is then proposed which transfers the channel prediction task into a matrix completion problem. It can directly infer future CSI based on past observations. We first give proof that the number of MPC equals the rank of a Hankel matrix generated by the observed CSI then
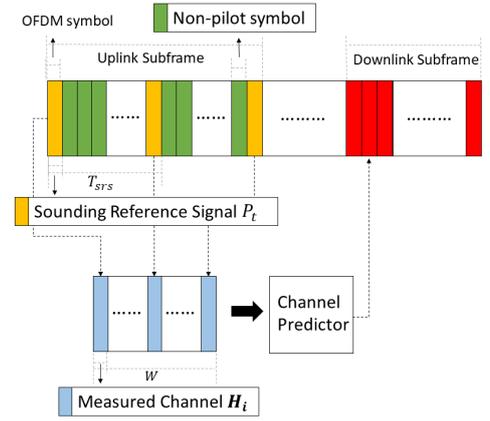
Fig. 1. In the uplink period, the yellow strip represents the pilot slot, the green strip represents the non-pilot slot and CSI can be measured on pilot slots and forms an observation window whose size is W, the channel predictor can predict future CSI for precoding in downlink subframe which is represented by red strips.

use iterated soft threshold (IST) method to reconstruct the future CSI in a rank minimization constraint.
- The performance of the proposed algorithms is tested both in a simulated channel generated by the COST2100 channel model and a real channel measured through a channel-sounding campaign. The results show that both algorithms can achieve a good channel capacity compared with the state-ofthe-art methods.

Notation: We use the following notation throughout the paper. Bold letter $\boldsymbol{X}$ is used for matrices or vectors. Non-bold letter x is used for scalars. $(\cdot)^H$ and $(\cdot)^T$ represent the conjugate transpose and transpose. We use $\|\cdot\|_p$ to denote p norm. $|\cdot|$ is the absolute value of its argument or the cardinality of a set. $\langle\cdot\rangle$ denotes the inner product of two vectors. $\boldsymbol{X} \otimes \boldsymbol{Y}$ is the Kronecker product of $\boldsymbol{X}$ and $\boldsymbol{Y}$. $\mathrm{vec}(\cdot)$ yields a vector for a matrix argument and $\mathcal{CN}(\mu, \sigma^2)$ denotes a complex circularly symmetric Gaussian random vector with mean μ and correlation σ^2. $mod(\cdot)$ and $round(\cdot)$ are modulo and round operation, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a TDD massive MIMO OFDM system, the base station, hereafter denoted by BS, is equipped with a planar antenna array which has $N_{\mathrm{t}} = mn$ antennas, where m is the number of columns and n is the number of rows. The user equipment, hereafter denoted by UE, is equipped with a linear array that has N_{r} antenna elements. In the uplink subframe, the UE will transmit the sounding reference signal (SRS) [14] for measuring CSI at the BS side and BS will precode the signal at the downlink subframe with the predicted channel.

Because of the intermittency of the pilot, BS could not acquire consecutive observations of CSI at the uplink subframe, We let $\mathbb{M} = \{\boldsymbol{H}_1, \cdots, \boldsymbol{H}_{\hat{i}}, \hat{i} \leq W\}$ as the measured CSI set, where $\boldsymbol{H}_{\hat{i}} \in \mathbb{C}^{N_c \times N_r \times N_t}$ is the measured CSI at the ith pilot symbol in a W-length observation window, N_c is the number of subcarriers. In the downlink subframe, BS will use a CSI predictor, denoted as $\mathbb{G}$, to predict downlink CSI

based on past observations. The frame structure and channel prediction process are depicted in Fig. 1. It is a uniform pilot design where the interval of the pilot is denoted by T_{srs}. In the uplink period, the yellow strip represents the pilot symbol and $\boldsymbol{H}_{\hat{i}}$ can be measured on them and forms an observation window whose size is W. The green strip represents the non-pilot symbol. The channel predictor forecasts the future CSI for precoding in the downlink subframe which is represented by red strips. A straightforward prediction method is to use the CSI measured from the last pilot symbol to do beamforming. However, because of the mobility of UE, the channel will rapidly change temporally, so maintaining the last measured CSI will significantly degrade the system performance. Hence, a more dedicated channel predictor should be designed.

Define $\mathbb{G}_t(\mathbb{M})$ as the predicted channel of the tth OFDM symbol following the observation window. Channel prediction aims to design an optimal channel predictor, denoted as $\hat{\mathbb{G}}$, to minimize the time average normalized mean squared error (TNMSE) between predicted CSI and its ground-truth value, which can be formulated as follows,

$$\hat{\mathbb{G}} = \arg\min_{\mathbb{G}} \lim_{T \to +\infty} \frac{1}{T} \sum_{\hat{i}=1}^{T} \frac{\|\mathbb{G}_{\hat{i}}(\mathbb{M}) - \boldsymbol{H}_{\hat{i}}\|_2}{\|\boldsymbol{H}_{\hat{i}}\|_2}. \quad (1)$$

A. Channel Model

The $N_{\mathrm{r}} \times N_{\mathrm{t}}$ spatial channel between the BS and UE on each time-frequency grid can be represented as

$$\boldsymbol{H}_{q,k} = \sum_{\hat{n} \in \Gamma} \sum_{\hat{j}=1}^{P} \alpha_{\hat{n},\hat{j}} \boldsymbol{a}_{\mathrm{t}}(\theta_{\hat{n},\hat{j}}, \psi_{\hat{n},\hat{j}}) \boldsymbol{a}_{\mathrm{r}}(\phi_{\hat{n},\hat{j}}) \quad (2)$$
$$\times e^{j2\pi q \nu_{\hat{n},\hat{j}} \Delta t} e^{-j2\pi k \tau_{\hat{n},\hat{j}} \Delta f}.$$

We denote Γ as cluster set of the scattering environment, $|\Gamma|$ is the number of clusters, each cluster contain P sub-paths. q is OFDM symbol index, k is subcarrier index. $\alpha_{\hat{n},\hat{j}}, \theta_{\hat{n},\hat{j}}, \psi_{\hat{n},\hat{j}}, \phi_{\hat{n},\hat{j}}, \nu_{\hat{n},\hat{j}}, \tau_{\hat{n},\hat{j}}$ represent fading coefficient, elevation of departure (EoD), azimuth of departure (AoD), azimuth of arrival (AoA), Doppler frequency and delay of each path respectively. Δt and Δf is symbol duration time and subcarrier spacing. $\boldsymbol{a}_{\mathrm{r}}(\cdot), \boldsymbol{a}_{\mathrm{t}}(\cdot)$ denote the steering vector at transmitter and receiver end which have the following form,

$$\boldsymbol{a}_{\mathrm{t}}(\theta_{\hat{n},\hat{j}}, \psi_{\hat{n},\hat{j}}) = \boldsymbol{V}_{\mathrm{x}}(\theta_{\hat{n},\hat{j}}, \psi_{\hat{n},\hat{j}}) \otimes \boldsymbol{V}_{\mathrm{y}}(\theta_{\hat{n},\hat{j}}, \psi_{\hat{n},\hat{j}}), \quad (3)$$

$$\boldsymbol{a}_{\mathrm{r}}(\phi_{\hat{n},\hat{j}}) = [1, e^{jkd\sin\phi_{\hat{n},\hat{j}}}, \cdots, e^{jk(N_{\mathrm{r}}-1)d\sin\phi_{\hat{n},\hat{j}}}]^T, \quad (4)$$

where $\boldsymbol{V}_{\mathrm{x}}(\cdot)$ and $\boldsymbol{V}_{\mathrm{y}}(\cdot)$ can be viewed as the steering vectors on the horizontal and vertical direction, respectively, with

$$\boldsymbol{v}_{\mathrm{x}}(\theta_{\hat{n},\hat{j}}, \psi_{\hat{n},\hat{j}}) = [1, e^{jkd_{\mathrm{x}}\cos\theta_{\hat{n},\hat{j}}\cos\psi_{\hat{n},\hat{j}}}, \quad (5)$$
$$\cdots, e^{jk(n-1)d_{\mathrm{x}}\cos\theta_{\hat{n},\hat{j}}\cos\psi_{\hat{n},\hat{j}}}]^T$$

and

$$\boldsymbol{v}_{\mathrm{y}}(\theta_{\hat{n},\hat{j}}, \psi_{\hat{n},\hat{j}}) = [1, e^{jkd_{\mathrm{y}}\cos\theta_{\hat{n},\hat{j}}\sin\psi_{\hat{n},\hat{j}}}, \quad (6)$$
$$\cdots, e^{jk(m-1)d_{\mathrm{y}}\cos\theta_{\hat{n},\hat{j}}\sin\psi_{\hat{n},\hat{j}}}]^T,$$

where $k = 2\pi/\lambda$ and λ is the carrier wavelength, $d, d_{\mathrm{x}}, d_{\mathrm{y}}$ is the inter-element spacing.

We should notice here that the parameters of each path in the same cluster will be slightly different, it is practically reasonable because, for a certain kind of reflector, the incident angle of all rays will not be exactly the same.

B. Doppler Aliasing

Most of the related channel prediction works focus on the prediction of future pilot symbols, and then reconstruct the non-pilot symbol with conventional interpolation methods, such as the DFT-based method. However, the conventional interpolation method will incur aliasing when the uniform pilot is not dense enough, especially in high-mobility cases. We name this kind of aliasing "Doppler aliasing" and would illustrate it in the angular-delay channel domain.

Define $\boldsymbol{h}_{q,k,r:}$ as the rth column of $\boldsymbol{H}_{q,k}$, which is the antenna pair response related to the rth receive antenna, we first construct $\bar{\boldsymbol{H}}_{q,r} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{c}}}$ as,

$$\bar{\boldsymbol{H}}_{q,r} = [\boldsymbol{h}_{q,1,r:}, \boldsymbol{h}_{q,2,r:}, \cdots, \boldsymbol{h}_{q,N_{\mathrm{c}},r:}], \quad (7)$$

where the subscript denotes the qth symbol and the rth receive antenna. The angular-delay domain channel can then be expressed as,

$$\hat{\boldsymbol{H}}_{q,r} = (\boldsymbol{D}_m \otimes \boldsymbol{D}_n)\bar{\boldsymbol{H}}_{q,r}\boldsymbol{D}_{N_{\mathrm{c}}}^H, \quad (8)$$

where $\boldsymbol{D}_{\hat{k}} \in \mathbb{C}^{\hat{k} \times \hat{k}}$ is a normalized Discrete Fourier Transform matrix of $\hat{k}$th order. We denote the channel on the (i,j)th angular-delay grid as $\hat{H}_{q,r}(i,j)$,

Next we define

$$F_{q,r}(i,j) = \sum_{q=1}^{W-1} \hat{H}_{q,r}(i,j)e^{-j\frac{2\pi}{W}kq}, \quad q = 0, 1, \cdots, W-1,$$
$$(9)$$

as the Doppler spectrum component of $\hat{\boldsymbol{h}}_r(i,j) = [\hat{H}_{1,r}(i,j), \cdots, \hat{H}_{W,r}(i,j)]^T \in \mathbb{C}^{W \times 1}$ and we let

$$\boldsymbol{f}_r(i,j) = [F_{1,r}(i,j), \cdots, F_{W,r}(i,j)], \quad (10)$$

as the Doppler spectrum of the rth antenna on the (i,j)th angular-delay domain grid in the observation window.

According to the Nyquist sampling theorem, for a lowpass like $\boldsymbol{f}_r(i,j)$, with its upper frequency denoted as $f_{\mathrm{H}}(i,j)$, if we want to reconstruct the Doppler spectrum from evenly spaced channel measurements, the minimum sampling rate should satisfy,

$$f_{\mathrm{s}} = \frac{1}{T_{\mathrm{srs}}} > 2f_{\mathrm{H}}(i,j). \quad (11)$$

If T_{srs} violates (11), the Doppler spectrum will be aliasing and can not be reconstructed through a lowpass filter. According to $\nu_{\hat{n},\hat{j}} = v\cos\beta_{\hat{n},\hat{j}}/\lambda$, $v, \beta_{\hat{n},\hat{j}}, \lambda$ is the velocity, intersection angle and wavelength of each path. When UE mobility increases, the Doppler frequency of each path will increase, and the sampling rate should be correspondingly greater to satisfy the sampling theorem, it means that pilots should be denser, but in reality, time-frequency resources are limited, denser pilots also mean sacrifice of payload resources.

C. Bandpass Sinc Interpolation

To address the problem proposed above, we first consider a special case that makes the problem easier to be tackled. In high mobility scenario, the Doppler frequency of each cluster will become larger and if all the Doppler frequencies are gathered in a certain range, the Doppler spectrum can then be viewed as a bandpass signal, then bandpass interpolation can be applied to reconstruct if it satisfies the following requisition.

For a bandpass signal which upper and lower frequency are f_H and f_L, the sampling rate f_s should satisfy

$$\frac{2f_H}{\hat{m}+1} \leq f_s \leq \frac{2f_L}{\hat{m}}, \quad (12)$$

where $\hat{m} > 0$ is an integer and satisfy

$$\hat{m} \leq \frac{f_L}{B}, \quad (13)$$

where B is the bandwidth of the signal. Then we can apply the following method to interpolate the CSI of non-SRS symbols.

$$\hat{H}_{t,r}(i,j) = \frac{2\omega_c \Delta T}{\pi} \sum_{\tilde{n}=-\infty}^{+\infty} \hat{H}_{\tilde{n}\Delta t,r}(i,j) \sin\left(\frac{\omega_c(t-\tilde{n}\Delta T)}{\omega_c(t-\tilde{n}\Delta T)}\right)$$
$$\times \cos(\omega_c(t-\tilde{n}\Delta T)), \quad (14)$$

where $\omega_c = 2\pi f_c$, f_c is the center frequency of a bandpass signal.

Under the bandpass signal assumption, we can reconstruct the CSI without increasing the density of pilots through bandpass interpolation. However, this method will face two main problems: firstly, the bandwidth and upper/lower frequency of the Doppler spectrum are hard to estimate in practice. Secondly, it's an ideal hypothesis that the Doppler spectrum is a bandpass signal, when there are lowpass frequency components on the Doppler spectrum that violates the assumption, bandpass interpolation will no longer work because it will still involve aliasing.

To overcome the problems mentioned above, we introduce two methods that combine the prediction on pilot and non-pilot symbols as an integral. The first one is a model-based algorithm called RAMBLE, which uses compressive sensing to extract each non-zero component on the Doppler spectrum and infers future CSI after reconstructing the channel. The second one is model-free, which leverages the matrix interpolation method to directly infer CSI based on past observations without extracting channel parameters.

III. COMPRESSIVE SENSING BASED PREDICTION SCHEME

A. Compressive Sensing (CS) and Non-Uniform Pilot Design

Let $\boldsymbol{y} \in \mathbb{C}^{\overline{n}}$ denote a vector signal, CS considers a compressed measurement $\boldsymbol{y}_c = \boldsymbol{\Phi}\boldsymbol{y}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{\overline{m} \times \overline{n}}$ is a sensing matrix, with $\overline{m} < \overline{n}$, therefore CS is solving an underdetermined equation, it is only applicable when $\boldsymbol{y}$ is sparse, since $\boldsymbol{y}$ is normally not sparse in its shape, its sparse representation can be obtained through a sparsifying transform $\boldsymbol{\Psi}$, in order to reconstruct the signal in sub-Nyquist
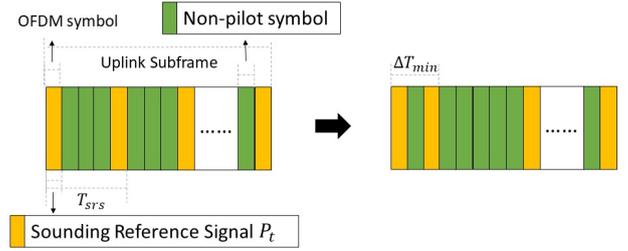
Fig. 2. This figure shows the non-uniform pilot pattern design, the minimum interval of pilot is denoted by ΔT_{min} and the density of pilot is remain the same as in uniform case.

rate, the sensing matrix should be well-designed to satisfy the restricted isometry property (RIP) [15]

$$(1-\delta)\|\boldsymbol{y}\|_2 \leq \|\boldsymbol{\Phi}\boldsymbol{y}\|_2 \leq (1+\delta)\|\boldsymbol{y}\|_2, \quad (15)$$

where $0 \leq \delta \leq 1$ is the RIP parameter. Typically, a random matrix with entries independently sampled from a subgaussian distribution can satisfy this property.

In our scenario, the sensing matrix reflects the sampling pattern of pilots, the Gaussian random matrix is not available here because BS can not acquire the CSI on every OFDM symbol, as a substitution, we choose a random binary matrix denoted as $\boldsymbol{S}$. If a symbol is chosen to insert a pilot, the corresponding entry is set to one, otherwise, it will be zero. Therefore, the pilot is no longer uniform because of the randomness.

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \Big\} \rho W \quad (16)$$

The sampling matrix is characterized by two parameters, ΔT_{min} and ρ, where ΔT_{min} denotes the minimum spacing of non-uniform SRS and it determines the Doppler frequency resolution of the non-uniform pilot. In order to have enough granularity to resolve the maximum Doppler frequency, ΔT_{min} should satisfy $1/\Delta T_{min} > f_H(i,j)$. ρ denotes the density of pilots in the observation window. To guarantee fairness, we should not increase the density of the non-uniform pilot compared with the traditional scheme, so we keep the non-uniform pilot density the same as in the uniform case, without compromising the available data resources. Therefore, the total number of the non-uniform pilot in the observation window can be expressed $\rho W = \lfloor W\Delta t/T_{srs} \rfloor$. The non-uniform pilot pattern design is depicted in Fig. 2.

To reconstruct the Doppler spectrum from measurements, we can formulate the optimization problem on the (i,j)th angular-delay domain grid (ADG) of each receive antenna.

$$\hat{F} = \arg\min_{\boldsymbol{f}} \|\boldsymbol{f}_r(i,j)\|_0$$
$$\text{s.t., } \boldsymbol{y}_r(i,j) = \boldsymbol{S}\boldsymbol{D}_W^H \boldsymbol{f}_r(i,j), \quad (17)$$

where $\boldsymbol{y}_r(i,j)$ is the available observations on the selected ADG. The above problem can be solved if the Doppler spectrum of ADG is sparse enough, thanks to the sparse property of the angular delay domain, each ADG is influenced by a few clusters in which angular and delay are nearby. Hence, the number of dominant clusters in each ADG is approximately the sparsity of its Doppler spectrum. This is partially applicable if each cluster contains only one path. Each cluster has its Doppler frequency distribution because there is more than one path in each cluster, the influence of this will be further discussed in the next subsection. Problem (17) is both numerically unstable and an NP-complete problem that requires an exhaustive enumeration of all possible combinations for the locations of the nonzero entries in $\boldsymbol{f}_r(i,j)$ [16]. A classical method is minimize the L-1 norm instead, so (17) is transformed to

$$\hat{F} = \arg\min_{\boldsymbol{f}} \|\boldsymbol{f}_r(i,j)\|_1$$
$$\text{s.t., } \boldsymbol{y}_r(i,j) = \boldsymbol{S}\boldsymbol{D}_W^H\boldsymbol{f}_r(i,j). \tag{18}$$

There are many well-known algorithms to solve (18), such as basis pursuit [17], orthogonal basis pursuit [18] and iterative thresholding algorithm [19], we choose OMP for its simplicity. OMP is an iterative method, it finds the atom that has the highest inner product with the signal and subtracts from the signal during each iteration.

After extracting all the non-zero Doppler components from the observations, we use regularized least squared method [20] to fit the amplitude of each Doppler frequency component to mitigate the influence of processing noise during channel estimation. Then the CSI can be reconstructed with a summation of all the frequency components and the future CSI can be deduced by increasing the time index.

However, if we use an orthonormal DFT basis as the sparsifying matrix, the predicted CSI will be the duplication of the observations because of the periodic property of DFT which can be expressed as follows.

$$\omega_N^k = \omega_N^{k+mN}, \tag{19}$$

where $\omega = e^{-2\pi i/N}$. This phenomenon is depicted in Fig. 3, the left and right figures show the amplitude and angle temporal variation of CSI on a certain ADG, the yellow frame represents the observation window and the green frame represents the prediction, which is the duplication of the observation.

In the real world, it is impossible that the actual Doppler frequency locates exactly based on the DFT, so the real channel will not be periodic. To address the problems, we resort to a Redundant dictionary to supply extra resolution capability.

B. Redundant Dictionary Based Channel Prediction Method

Normally, the signal will be sparse on an orthonormal basis. However, more often than not, sparsity is expressed not in terms of an orthonormal basis but in terms of an overcomplete dictionary. In this setting, the dictionary need not be orthonormal or even incoherent and often it will be overcomplete, meaning it has far more columns than rows. In our scenario,

(a)

(b)
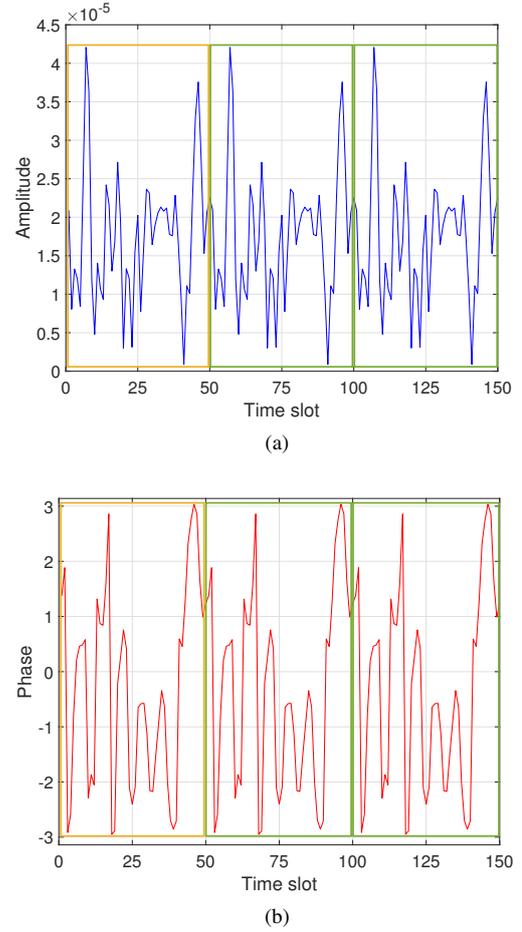
Fig. 3. The left and right figures show the variation of amplitude and phase of observed CSI and predicted CSI respectively, the yellow frame is the observation part and the red one is the predicted part. It can be seen from the graph that the prediction is the duplication of the observation because of the cyclic property of DFT.

to decrease the gap between real Doppler frequency and DFT basis, we use oversampled DFT matrix as a dictionary, so $\boldsymbol{y}_r(i,j)$ can be expressed by a redundant dictionary which is the first ρW rows of the Hermitian transpose of a FW order DFT matrix

$$\boldsymbol{y}_r(i,j) = \boldsymbol{S}\hat{\boldsymbol{D}}_{FW}^H\hat{\boldsymbol{f}}_r(i,j), \tag{20}$$

where $\hat{\boldsymbol{D}}_{FW}$ is shown in (21), and F is the refinement factor, which reflects the frequency resolution of the dictionary. The gap between real frequency and DFT basis will be smaller when F increases. However, a large F will induce the neighboring columns of $\hat{\boldsymbol{D}}_{FW}$ strongly coherent, which will cause OMP to choose two adjacent columns consecutively, it often happens because there is more than one path in a cluster [21] and the parameters of them will not be identified which will cause a small range of Doppler spread. Hence, the contribution of the selected spike could not be eliminated from the residual. There are two negative effects brought by this problem, the first one is the projection computation in OMP will become ill-conditioned [22], and the second one is not all the dominant Doppler spikes can be found because some estimated Doppler indexes will assemble on the same spike. To circumvent this,

$$\hat{\boldsymbol{D}}_{FW} = \frac{1}{\sqrt{FW}} \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{W-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{FW-1} & \omega^{2(FW-1)} & \omega^{3(FW-1)} & \cdots & \omega^{(W-1)(FW-1)} \end{bmatrix} \Big\} FW \qquad (21)$$

where the brace above spans W.

$$\begin{aligned} |\boldsymbol{R}_1^H \boldsymbol{\Phi}_{S_{\max}}| &= |F_{\max} + \sum_{j \neq S_{\max} \cap j \in \Upsilon_{S_{\max}}} F_j \boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_{S_{\max}} + \cdots + \sum_{j \in \Upsilon_K} F_j \boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_{S_{\max}} + \boldsymbol{e}^H \boldsymbol{\Phi}_{S_{\max}}| \\ &\geq F_{\max} - F_{\max}(\delta - 1)\hat{\eta} - F_{\max}(K - 1)\delta\eta - \|\boldsymbol{e}\|_2 \end{aligned} \qquad (22)$$

we use the band exclusion method introduced in [23]. First, define the coherent band of the dictionary as follows.

Let $\eta > 0$. Define the η-coherence band of column k to be the set

$$B_\eta(\overline{k}) = \{\overline{i} | \mu(\overline{i}, \overline{k}) > \eta\}, \qquad (23)$$

where $\mu(\overline{i}, \overline{k}) = \frac{|\boldsymbol{\Phi}_{\overline{i}}^H \boldsymbol{\Phi}_{\overline{k}}|}{\|\boldsymbol{\Phi}_{\overline{i}}\|_2 \|\boldsymbol{\Phi}_{\overline{k}}\|_2}$ the η-coherence band of the index set $\mathbb{S}$ is

$$B_\eta(\mathbb{S}) = \cup_{\overline{k} \in \mathbb{S}} B_\eta(\overline{k}), \qquad (24)$$

and denote $B_\eta^{(2)}(\overline{k}) = B_\eta(B_\eta(\overline{k}))$. Then we can modify the matching step in OMP as follows:

$$\overline{i}_{\max} = \arg\min_{\overline{i}} \left| \langle \boldsymbol{\Phi}_{\overline{i}}^H, \boldsymbol{R} \rangle \right|, \overline{i} \notin B_\eta^{(2)}(\boldsymbol{\Omega}), \qquad (25)$$

where $\boldsymbol{\Omega}$ is the estimated index set. Equation (25) skips the neighboring zone of each estimated Doppler component under the assumption that the Doppler frequencies of every two clusters should be separated at least $B_\eta^{(2)}(\overline{k})$ apart. This is acceptable because clusters are normally spatially uncorrelated. Besides, we uniformly add a few adjacent frequencies near each estimated non-zero Doppler to simulate the Doppler frequency distribution in a single cluster after iteration reaches its end. The algorithm is illustrated in Algorithm 1. The modifications compared with traditional OMP are highlighted with boxes.

Next, we will analyze the recovery stability of Algorithm 1, first define $\Upsilon_{\overline{i}}$ as the Doppler index set spanned by the $\overline{i}$th Doppler spike in one ADG, which can be expressed as $\Upsilon_{\overline{i}} = \{S_{\overline{i}} - \delta/2, \cdots, S_{\overline{i}} + \delta/2\}$, where $S_{\overline{i}}$ is the center Doppler frequency and δ is the Doppler index spread. In a practical case, the non-Doppler term in $\boldsymbol{f}_r$ will not be strictly zero, so the range of a Doppler spike can be defined by a manually set threshold. Next, define $\Pi[\hat{\boldsymbol{f}}_r(i, j)] = \{S_1, \cdots, S_K\}$ where K is the number of spikes. An ideal recovery is defined as each estimated Doppler index being within the η-coherence band of each element in $\Pi[\hat{\boldsymbol{f}}_r(i, j)]$.

Theorem 1 (stable recovery criteria for band-excluded OMP):

Let $\hat{\boldsymbol{f}}_r(\hat{m}, \hat{n})$ has K non-zero spikes and the length of each spike is δ. Let $\eta > 0$ and $\hat{\eta} = \max \mu(i, j) \quad \forall i, j \in [S_{\overline{k} - \delta/2}, \cdots, S_{\overline{k} + \delta/2}]$. Suppose that

$$B_\eta(a) \cap B_\eta^{(2)}(b) = \emptyset \quad \forall a, b \in \Pi[\hat{\boldsymbol{f}}_r(\hat{m}, \hat{n})], \qquad (26)$$

if

$$(2K - 1)\delta\eta + (\delta - 1)\hat{\eta} + 2\frac{\|\boldsymbol{e}\|_2}{F_{\max}} < 1, \qquad (27)$$

and

$$\frac{F_{\max}(\eta\delta K - 1) + 2\eta(K - 1)A_{\max} + 2\|\boldsymbol{e}\|_2}{F_{\min}[1 - \hat{\eta}(\delta - 1) - \eta\delta K]} \leq 1, \qquad (28)$$

are satisfied where

$$\begin{aligned} A_{\max} = \frac{1}{1 - \eta(K - 2)}&[F_{\max} + F_{\max}(\delta - 1)\hat{\eta} \qquad (29) \\ &+ F_{\max}(K - 1)\delta\eta + \|\boldsymbol{e}\|_2], \end{aligned}$$

and $F_{\max} = \max \hat{\boldsymbol{f}}_r(i, j)$ and $F_{\min} = \min \hat{\boldsymbol{f}}_r(i, j)$, then the reconstructed Doppler index will fall in $B_\eta(\Pi[\hat{\boldsymbol{f}}_r(i, j)])$.

Proof 1: Firstly, we will prove that the first chosen index would fall in $\Pi[(\hat{\boldsymbol{f}}_r(i, j)]$, Let $S_{\max}$ be the index of the largest component in the absolute value of $\hat{\boldsymbol{f}}_r(i, j)$. In the first iteration, $\boldsymbol{R}_1 = \sum_{i=1}^K \sum_{j \in \Upsilon_1} F_j \boldsymbol{\Phi}_j$, the subscript of $\boldsymbol{R}$ denotes the iteration index and the subscript of $\boldsymbol{\Phi}$ denotes the column index. In addition, we ignore the superscript of F here for simplicity, as shown in (22).

For $\forall l \notin B_\eta(\Pi[\hat{\boldsymbol{f}}_r(i, j)])$

$$\begin{aligned} |\boldsymbol{R}_1^H \boldsymbol{\Phi}_l| &= |\sum_{j \in \Upsilon_1} F_j \boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_l + \cdots + \sum_{j \in \Upsilon_K} F_j \boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_l + \boldsymbol{e}^H \boldsymbol{\Phi}_l| \\ &\leq F_{\max} K \delta\eta + \|\boldsymbol{e}\|_2, \end{aligned} \qquad (30)$$

if the right-hand side of (22) is greater than the right-hand side of (30), we get

$$(2K - 1)\delta\eta + (\delta - 1)\hat{\eta} + 2\frac{\|\boldsymbol{e}\|_2}{F_{\max}} < 1. \qquad (31)$$

Then the first chosen index will be $S_{\max}$. Next we suppose the first $k - 1$ selected indices are in $B_\eta(S_i), S_i \in \Pi(\hat{\boldsymbol{f}}_r(i, j))$, the $k - 1$th residual can be expressed as

$$\boldsymbol{R}_{k-1} = \hat{\boldsymbol{f}}_r(i, j) - A_{I_1}\boldsymbol{\Phi}_{I_1} - A_{I_2}\boldsymbol{\Phi}_{I_2} - \cdots - A_{I_{k-1}}\boldsymbol{\Phi}_{I_{k-1}}, \qquad (32)$$

where $\{A_{I_1}, A_{I_2}, \cdots, A_{I_{k-1}}\}$ is the estimated coefficients, we can derive each coefficient as

$$
\begin{aligned}
A_{I_i} = &\sum_{j \in \Upsilon_1} F_j \boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_{I_i} + \sum_{j \in \Upsilon_2} F_j \boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_{I_i} + \cdots \\
&+ \sum_{j \in \Upsilon_K} F_j \boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_{I_i} + \boldsymbol{e}^H \boldsymbol{\Phi}_{I_i} \\
&- \sum_{n \neq i} A_{I_n} \boldsymbol{\Phi}_{I_n}^H \boldsymbol{\Phi}_{I_i}, i = 1, \cdots, k-1,
\end{aligned}
\tag{33}
$$

which implies

$$
\begin{aligned}
|A_{I_i}| \leq &F_{\max} + F_{\max}(\delta - 1)\hat{\eta} + F_{\max}(K-1)\delta\eta + \|\boldsymbol{e}\|_2 \\
&+ \eta \sum_{n \neq i} |A_{I_n}|, i = 1, \cdots, k-1
\end{aligned}
\tag{34}
$$

Let $A_{\max} = \max_{j=1, \cdots, k-1} |A_{I_j}|$, (34) can then be

$$
\begin{aligned}
A_{\max} \leq &F_{\max} + F_{\max}(\delta - 1)\hat{\eta} + F_{\max}(K-1)\delta\eta + \|\boldsymbol{e}\|_2 \\
&+ \eta(k-2)A_{\max}.
\end{aligned}
\tag{35}
$$

hence,

$$
\begin{aligned}
A_{\max} \leq \frac{1}{1 - \eta(k-2)} [&F_{\max} + F_{\max}(\delta - 1)\hat{\eta} \\
&+ F_{\max}(K-1)\delta\eta + \|\boldsymbol{e}\|_2],
\end{aligned}
\tag{36}
$$

when $k = K$, the right-hand side of (36) will get its maximum value.

Now we show that the kth selected index will fall in $B_\eta(\Pi(\hat{\boldsymbol{f}}_r(i,j)))$. For the kth residual, we have (37). For $\forall l \notin B_\eta^{(2)}(S_1, \cdots, S_{k-1}) \cup B_\eta(S_k, \cdots, S_K)$

$$
\begin{aligned}
\mathbf{R}_{k-1}^H \boldsymbol{\Phi} = |&\sum_{j \in \Upsilon_1} F_j \boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_l + \cdots + \sum_{j \in \Upsilon_K} F_j \boldsymbol{\Phi}_j^H \boldsymbol{\Phi}_l + \mathbf{e}^H \boldsymbol{\Phi}_l \\
&- A_{I_1} \boldsymbol{\Phi}_{I_1}^H \boldsymbol{\Phi}_l - \cdots - A_{I_{k-1}} \boldsymbol{\Phi}_{I_{k-1}}^H \boldsymbol{\Phi}_l| \\
\leq &\eta F_{\min} K \delta + \|\mathbf{e}\|_2 + \eta(k-1)A_{\max}
\end{aligned}
\tag{39}
$$

If the right-hand side of (37) is gerater than the right-hand side of (39), we get

$$
\frac{F_{\max}(\eta \delta K - 1) + 2\eta(K-1)A_{\max} + 2\|\boldsymbol{e}\|_2}{F_{\min}[1 - \hat{\eta}(\delta - 1) - \eta K \delta]} \leq 1,
\tag{40}
$$

which end our proof.

Remark 1: From the above stable recovery criteria, there exists a compromise among η, δ, K, $F_{\max}/F_{\min}$. If we wants to recover the Doppler spectrum with more spikes with a fixed refinement factor, η should be smaller which is equivalent to more time domain samples. The recovery probability will decrease if there are more distinct Doppler components in one spike or the dynamic range $F_{\max}/F_{\min}$ is large. In addition, (26) is not a strict requirement because clusters will normally be separately located.

C. Multiple Measurement Vector

To further improve the recovery stability in the presence of noise and the increasing number of non-zero Doppler spikes, MMV method is leveraged to further excavate the potential of multiple antenna at the receiver end.

Algorithm 1: Redundant dictionary OMP-based angular delay domain channel prediction

1 **Input**: Observation $\mathbb{M}$,
    $\boxed{\text{redundant dictionary } \boldsymbol{\Phi} = \boldsymbol{S}\hat{\boldsymbol{D}}_{FW}^H}$, threshold γ, regularization factor ξ, $f_s = \frac{1}{\Delta T_{\min}}$, band exclusion factor η

2 **Stage1: Transform channel to angular delay domain**

3     $\hat{\boldsymbol{H}}_{q,r} = (\boldsymbol{D}_m \otimes \boldsymbol{D}_n)\bar{\boldsymbol{H}}_{q,r}\boldsymbol{D}_{N_c}^H \quad r = 1 \cdots N_r$

4 **Stage2: Find non-zero position of Doppler spectrum on each ADG**

5 **for** $r = 1 : N_r$ **do**

6   **for** *each ADG* **do**

7        $\boldsymbol{y}_r(i,j) = [\hat{H}_{1,r}(i,j) \cdots \hat{H}_{|\mathbb{M}|,r}(i,j)]$

8   **Initialization**: $\boldsymbol{R} = \boldsymbol{y}_r(i,j)$, $\boldsymbol{L} = \varnothing$, $k = 0$

9   **while** $\frac{\|\boldsymbol{R}\|_2}{\|\boldsymbol{y}_r(i,j)\|_2} > \gamma$ **do**

10          $pos = \arg\max_i(|\langle \boldsymbol{\Phi}_i^H, \boldsymbol{R} \rangle|)$, $\boxed{i \notin B_\eta^{(2)}(\boldsymbol{\Omega})}$

11   Add $\boldsymbol{\Phi}_{pos}$ to $\boldsymbol{L}$

12   Set $\boldsymbol{\Phi}_{pos}$ as zero column

13          $\boldsymbol{Z} = (\boldsymbol{L}^H \boldsymbol{L})^{-1}\boldsymbol{L}\boldsymbol{y}_r(i,j)$

14          $\boldsymbol{R} = \boldsymbol{y}_r(i,j) - \boldsymbol{L}\boldsymbol{Z}$

15          $\boldsymbol{\Omega}(k) = pos$

16          $k = k + 1$

17 **Stage3: Fitting the complex amplitude with RLS**

18     $\boxed{\text{Uniformly padding frequencies in} \boldsymbol{\Omega}}$

19   Transform $\boldsymbol{\Omega}$ to real Doppler frequency $\hat{\boldsymbol{f}}$

20     $A_{i,j} = e^{j2\pi \hat{f}(j)(i-1)/f_s}$

21     $\hat{\boldsymbol{Z}} = (\boldsymbol{A}^H \boldsymbol{A} + \xi \boldsymbol{I})^{-1}\boldsymbol{A}^H \boldsymbol{y}_r(i,j)$

22 **Stage4: Predicting future CSI**

23   **Given** $\forall t > W$

24     $\tilde{H}_{t,r}(i,j) = \sum_{l=1}^{|\hat{\boldsymbol{f}}|} \hat{z}(l)e^{j2\pi \hat{f}(l)\frac{t-1}{f_s}}$

25 **Output**: $\tilde{\boldsymbol{H}}_r, r = 1 \cdots N_r$

First define the support of a collection of vectors $\boldsymbol{X} = [\boldsymbol{x_1}, \cdots, \boldsymbol{x_l}]$ as the union over all individual supports,

$$
\mathrm{supp}(\boldsymbol{X}) = \bigcup_i \mathrm{supp}(\boldsymbol{x}_i).
\tag{41}
$$

In other words, a joint sparse matrix defined above contains several sparse vectors where the non-zero component in each of them is in the same position. Given the definition of $\boldsymbol{X}$, a noiseless MMV problem can be defined as follows,

Given $\boldsymbol{Y} \in \mathbb{C}^{m \times l}$ and $\hat{\boldsymbol{\Phi}} \in \mathbb{C}^{m \times n}$ with $m < n$

$$
\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} |\mathrm{supp}(\boldsymbol{X})| \quad s.t. \boldsymbol{Y} = \hat{\boldsymbol{\Phi}}\boldsymbol{X}.
\tag{42}
$$

It is a special case when $l = 1$ then $\boldsymbol{X}$ shrinks into a vector form which is named by single measurement vector (SMV) problem. It can be solved previous reconstruction method because algorithm 1 treats each receive antenna as an individual.

For the SMV problem, it is well known that a necessary and sufficient condition for the measurements $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}$ to uniquely

$$|\boldsymbol{R}_{k-1}^{H}\boldsymbol{\Phi}_{S_k}| = |\sum_{j\in\Upsilon_1}\mathbb{F}_j\boldsymbol{\Phi}_j^H\boldsymbol{\Phi}_{S_k} + \cdots + \sum_{j\in\Upsilon_K}\mathbb{F}_j\boldsymbol{\Phi}_j^H\boldsymbol{\Phi}_{S_k} + \boldsymbol{e}^*\boldsymbol{\Phi}_{S_k} - A_{I_1}\boldsymbol{\Phi}_{I_1}^H\boldsymbol{\Phi}_{S_k} - \cdots - A_{I_{k-1}}\boldsymbol{\Phi}_{I_{k-1}}^H\boldsymbol{\Phi}_{S_k}| \tag{37}$$
$$\geq F_{\min} - \hat{\eta}(\delta-1)F_{\min} - \eta(K-1)\delta F_{\max} - \|\boldsymbol{e}\|_2 - \eta(k-1)A_{\max}$$

$$[\text{vec}(\hat{\boldsymbol{H}}_{q,r}+\boldsymbol{n})]_i = [(\boldsymbol{D}_{N_c}\otimes\boldsymbol{D}_m\otimes\boldsymbol{D}_n)\text{vec}(\bar{\boldsymbol{H}}_{q,r}+\boldsymbol{n})]_i$$
$$= \sum_{n\in\Gamma}\sum_{j=1}^{P}\alpha_{k_1,k_2,k_3}mnN_c\text{Sa}(2\pi(\hat{\theta}_{k_1}-\theta_{n,j})\frac{m}{2})\text{Sa}(2\pi(\hat{\psi}_{k_2}-\psi_{n,j})\frac{n}{2})\text{Sa}(2\pi(\hat{\tau}_{k_3}-\tau_{n,j})\frac{N_c}{2}) + \frac{\sigma^2}{mnN_c} \tag{38}$$

---

determine each k-sparse vector $\boldsymbol{x}$ is given by

$$k < \frac{\text{spark}(\boldsymbol{\Phi})}{2}, \tag{43}$$

where spark of $\boldsymbol{\Phi}$ is defined as the smallest number of columns of $\boldsymbol{\Phi}$ that are linearly independent.

Similarly, a sufficient condition from the measurements $\boldsymbol{Y} = \hat{\boldsymbol{\Phi}}\boldsymbol{X}$, to uniquely determine the jointly sparse matrix $\boldsymbol{X}$ has been proposed in [24],

$$|\text{supp}(\boldsymbol{X})| < \frac{\text{spark}(\boldsymbol{\Phi}) - 1 + \text{rank}(\boldsymbol{Y})}{2}, \tag{44}$$

which implies that the measurements can be reduced if the rank of $\boldsymbol{Y}$ is greater than 1.

In our scenario, the receive antenna can be considered as multiple measurements, We can thus combine $\hat{\boldsymbol{f}}_r(i,j), r = 1,\cdots,N_r$ as a matrix $\boldsymbol{F}(i,j)$

$$\boldsymbol{F}(i,j) = [\hat{\boldsymbol{f}}_1(i,j), \hat{\boldsymbol{f}}_2(i,j), \cdots, \hat{\boldsymbol{f}}_{N_r}(i,j)]. \tag{45}$$

To implement the MMV algorithm on $\boldsymbol{F}(i,j)$, one should first guarantee that $\boldsymbol{F}(i,j)$ satisfies the joint sparse property.

Theorem 2 (asymptotic joint sparse property):
For $\forall i,j = 1,\cdots,N_r, i \neq j$

$$\lim_{N_c,N_t\to+\infty}\text{supp}(\hat{\boldsymbol{f}}_i(k,q)) = \text{supp}(\hat{\boldsymbol{f}}_j(k,q)). \tag{46}$$

Proof 2:
Without loss of generality, consider an additive noise induced by channel measurement, denoted by $\boldsymbol{n} \sim \mathcal{CN}(0,1)$. For the rth receive antenna, the observed channel in each AGD can then be expressed as follows, $[\cdot]_i$ here denotes the ith entry of a vector as shown in (38):

where $k_1 = mod(i,n)$, $k_2 = round(i/n)$, $k_3 = round(i/mn)$. $\hat{\theta}_{k_1}, \hat{\psi}_{k_2}, \hat{\tau}_{k_3}$ are equivalent estimation value on each grid. When m, n, N_c approximate to infinity, the gap between estimation value and real value becomes smaller due to the higher resolution, then $Sa(\cdot)$ will change into Dirac function, which implies that the Doppler spectrum of each receiving antenna on an ADG will contain the same single path and will not be influenced by the power of other ADGs, which end our proof.

Remark 2: The above definition implies that when the number of antenna and subcarrier goes to infinity, the observation will be a rank-deficient matrix, which causes the MMV to compromise its ability to decrease the number of measurements but to mitigate the influence of channel noise.

To implement MMV into our scheme, we adopt the rank-aware method proposed in [25], it decomposes the measurement matrix to orthogonal basis and computes the projection power of each column on it, we still adopt band exclusion strategy while selecting the columns. The details of the algorithm are illustrated in Algorithm 2. We name the rank-aware based omp with band-excluded redundant dictionary as RAMBLE algorithm.

IV. MATRIX COMPLETION PREDICTION SCHEME

The previously proposed algorithm relies on the channel modeling of the real-world channel, it estimates the Doppler parameters from the observed channel and reconstructs the future channel based on known channel function. In this section, a model-free channel prediction scheme is proposed which does not rely on the modeling of channel. It directly deduces the future channel solely based on past observations.

The problem is still considered in the angular-delay domain, for the (i,j)th grid of the rth receive antenna, construct a Hankel matrix with the past observations as follows, the subscript of receive antenna is omitted for the simplicity of expression:

$$\boldsymbol{B} = \begin{bmatrix} \hat{H}_{1,r}(i,j) & \hat{H}_{2,r}(i,j) & \cdots & \hat{H}_{\frac{W+1}{2},r}(i,j) \\ \hat{H}_{2,r}(i,j) & \hat{H}_{3,r}(i,j) & \cdots & \hat{H}_{\frac{W+3}{2},r}(i,j) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{H}_{\frac{W+1}{2},r}(i,j) & \hat{H}_{\frac{W+3}{2},r}(i,j) & \cdots & \hat{H}_{W,r}(i,j) \end{bmatrix}. \tag{47}$$

As one can not get all the CSI in the observation window, the actual $\boldsymbol{B}$ will be a sparse matrix with many entries left zero. Considering the predicted CSI, $\boldsymbol{B}$ will be padded with a lower triangle matrix of zero which represents the future CSI, denote the padded matrix as $\hat{\boldsymbol{B}}$

$$\hat{\boldsymbol{B}} = \begin{bmatrix} \hat{H}_{1,r}(i,j) & 0 & \cdots & \hat{H}_{\frac{W+N+1}{2},r}(i,j) \\ 0 & \hat{H}_{3,r}(i,j) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{H}_{\frac{W+N+1}{2},r}(i,j) & 0 & \cdots & \boldsymbol{0}_{N\times N} \end{bmatrix}, \tag{48}$$

where $\boldsymbol{0}_{N\times N}$ is a lower triangle matrix of zero.

Algorithm 2: Rank-aware based OMP channel prediction with band-excluded redundant dictionary

---

1 Input: Observation $\mathbb{M}$, $\mathbf{\Phi} = \mathbf{S}\hat{\mathbf{D}}_{FW}^{H}$, threshold γ, regularization factor ξ, $f_s = \frac{1}{\Delta T_{\min}}$, band exclusion factor η

2 Stage1: Transform channel to angular delay domain

3 $\hat{\mathbf{H}}_{q,r} = (\mathbf{D}_m \otimes \mathbf{D}_n)\bar{\mathbf{H}}_{q,r}\mathbf{D}_{N_c}^{H}$ $\mathbf{H}_j \in \mathbb{M}$ $r = 1, \cdots, N_r$

4 Stage2: Generate measurement matrix from different antennas

5 $\breve{\mathbf{H}}_j = [\hat{\mathbf{H}}_{j,1}, \hat{\mathbf{H}}_{j,2}, \cdots, \hat{\mathbf{H}}_{j,N_r}]$

6   **for** *each ADG* **do**

7 $\mathbf{Y}(i,j) = [\breve{\mathbf{H}}_1(i,j)\cdots\breve{\mathbf{H}}_{|\mathbb{M}|}(i,j)] \in C^{\rho W \times N_r}$

8 Stage3: Find non-zero position of Doppler spectrum on each ADG

9 for *each ADG* **do**

10   **Initialization**: $\mathbf{R} = \mathbf{Y}(i,j)$, $\mathbf{L} = \varnothing$, $k = 0$, $\mathbf{\Phi}^{'} = \mathbf{\Phi}$

11   **while** $\frac{\|\mathbf{R}\|}{\|\mathbf{Y}(i,j)\|} > \gamma$ **do**

12   Calculate orthonomal basis for residual: $\mathbf{U}^{(n-1)} = \text{Orth}(\mathbf{R}^{(n-1)})$

13 $pos = \arg\max_i(\|\langle\mathbf{\Phi}_i^H, \mathbf{U}^{(n-1)}\rangle\|_2)$ $i \notin B_\eta^{(2)}(\mathbf{\Omega})$

14 $\mathbf{\Omega}(k) = pos$

15   Add $\mathbf{\Phi}_{pos}$ to $\mathbf{L}$

16   Set $\mathbf{\Phi}_{pos}$ as zero column

17 $\mathbf{Z} = (\mathbf{L}^H\mathbf{L})^{-1}\mathbf{L}^H\mathbf{Y}(i,j)$

18   Calculate orthogonal projector $\mathbf{P}_{\mathbf{\Omega}}^{\perp} = \mathbf{I} - \mathbf{\Phi}_{\mathbf{\Omega}}\mathbf{\Phi}_{\mathbf{\Omega}}^H$

19 $\mathbf{R}^{(n)} = \mathbf{Y}(i,j) - \mathbf{L}\mathbf{Z}$

20 $\mathbf{\Phi}^{'} = \mathbf{P}_{\mathbf{\Omega}}^{\perp}\mathbf{\Phi}$

21   Normalize each column of $\mathbf{\Phi}^{'}$

22 $k = k + 1$

23 Stage3: Fitting the complex amplitude with RLS

24 Stage4: Predicting subsequent CSI

25 Output: $\tilde{\mathbf{H}}_{r,t}, r = 1, \cdots, N_r$

Therefore, the prediction of future CSI can now be transformed to complete $\hat{\mathbf{B}}$ of those zeros entries. Meanwhile, a criteria is needed to help supervise the reconstruction process. First, we will prove that if $\hat{H}_{q,r}(i,j)$ can be expressed as a sum of exponential, the rank of $\hat{\mathbf{B}}$ constructed from $\hat{H}_{q,r}(i,j)$ is equal to the number of components in $\hat{H}_{q,r}(i,j)$.

Proof 3: Suppose a signal is the sum of exponential which can be expressed as

$$S_n = \sum_{k=1}^{K} a_k e^{j2\pi f_k n} \quad n = 0, \cdots, L, \quad (49)$$

where L is the total length of the signal, then one can construct a Hankel matrix from S_n

$$\mathbf{M}_s = \begin{bmatrix} S_0 & S_1 & \cdots & S_{N-1} \\ S_1 & S_2 & \cdots & S_N \\ \vdots & \vdots & \ddots & \vdots \\ S_M & S_{M+1} & \cdots & S_{M+N-1} \end{bmatrix}, \quad (50)$$

where both M and N are integer and satisfy $L = M + N - 1$, M and N should be equal to guarantee $\mathbf{M}_s$ be a square matrix. According to [26], $\mathbf{M}_s$ can then be decomposed as

$$\mathbf{M}_s = \mathbf{\Omega}_M \mathbf{A} \mathbf{\Omega}_N^H. \quad (51)$$

$\mathbf{A}$ is a diagonal matrix with its entries be the amplitude of each exponential.

$$\mathbf{A} = diag(a_1, a_2, \cdots, a_K) \quad (52)$$

$\mathbf{\Omega}_x$ is a $x \times K$ Vandermond matrix defined as follows.

$$\mathbf{\Omega}_x = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{jf_1} & e^{jf_2} & \cdots & e^{jf_K} \\ \vdots & \vdots & \ddots & \vdots \\ e^{jf_1(x-1)} & e^{jf_2(x-1)} & \cdots & e^{jf_K(x-1)} \end{bmatrix}. \quad (53)$$

Equation (51) implies that for $\forall i, j = 1, \cdots, K$, and $i \neq j$, if $f_i \neq f_j$, then the rank of M is equal to the number of exponential components in S_n. based on our channel model, the entries in $\hat{\mathbf{B}}$ satisfies (49), therefore, the rank of $\hat{\mathbf{B}}$ is approximately equal to the number of Doppler spike in the corresponding ADG. To satisfy the sparse property of the scattering environment, the aim of reconstruction is to use the least number of the component to fit observations, which is equivalent to minimizing the rank of $\hat{\mathbf{B}}$.

One can now reconstruct the matrix with a rank minimization constraint, the optimization problem can be formulated as follows.

$$\bar{\mathbf{B}} = \arg\min \text{rank}(\hat{\mathbf{B}})$$
$$\text{s.t.,} \ \bar{\mathbf{B}}_{i,j} = \hat{\mathbf{B}}_{i,j} \quad \hat{\mathbf{B}}_{i,j} \neq 0, \quad (54)$$

where the subscript represents the non-zero entry in the ith column and the jth row, denotes the set of non-zero entry index tuple as $\boldsymbol{\kappa}$. However, (54) is an NP-hard problem, an alternative way is to minimize the nuclear norm of $\hat{\mathbf{B}}$. Then (54) becomes

$$\bar{\mathbf{B}} = \arg\min \|\hat{\mathbf{B}}\|_*$$
$$\text{s.t.,} \ \bar{\mathbf{B}}_{i,j} = \hat{\mathbf{B}}_{i,j} \quad \hat{\mathbf{B}}_{i,j} \neq 0, \quad (55)$$

where $\|\hat{\mathbf{B}}\|_*$ is the sum of singular values of $\hat{\mathbf{B}}$.

To solve the above problem, soft iterative thresholding (SIT) proposed in [27] is leveraged. It is an iterative algorithm, in the kth iteration, it decomposes $\hat{\mathbf{B}}$ through singular value decomposition (SVD) as $\hat{\mathbf{B}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrix and $\mathbf{\Sigma}$ is a positive diagonal matrix. Then for each $\tau > 0$, define singular value shrinkage operator $\mathbb{D}_\tau$ as

$$\mathbb{D}_\tau(\mathbf{\Sigma}) = diag(\{\sigma_i - \tau\}_+), \quad (56)$$

Algorithm 3: Matrix completion based channel pre-diction method

---

1 **Input**: Observation $\mathbb{M}$
2 $\quad \hat{\boldsymbol{H}}_{q,r} = (\boldsymbol{D}_m \otimes \boldsymbol{D}_n) \bar{\boldsymbol{H}}_{q,r} \boldsymbol{D}_{N_c}^H \quad \boldsymbol{H}_j \in \mathbb{M} \quad r = 1 \cdots N_r$
3 **Stage2: Predicting future CSI**
4 **for** $r = 1 : N_r$ **do**
5 $\quad$ **for** *each ADG* **do**
6 $\quad\quad$ **Construct Hankel matrix $\hat{B}$ from observations $\hat{H}_{q,r}$**
7 $\quad\quad$ **Complete $\hat{B}$ with SIT**

8 **Stage3: Lowpass interpolation of future CSI**
9 **Output**: $\tilde{\boldsymbol{H}}_r, r = 1 \cdots N_r$

---

where t_+ is the positive part of t, namely $t_+ = \max(0, t)$, this operator effectively shrinking singular values to zero. Then, if we apply this operator to $\hat{\boldsymbol{B}}$, we get

$$\mathbb{D}_\tau(\hat{\boldsymbol{B}}) = \boldsymbol{U}\mathbb{D}_\tau(\boldsymbol{\Sigma})\boldsymbol{V}^*. \tag{57}$$

To maintain the non-zero entries equal to the initial value in each iteration, it takes following step for a fixed $\tau > 0$ and a sequence $\{\delta_k\}$ of positive step sizes.

$$\tilde{\boldsymbol{B}}^{k-1} = \mathbb{D}_\tau(\hat{\boldsymbol{B}}^{k-1}), \tag{58}$$

$$\hat{\boldsymbol{B}}^k = \hat{\boldsymbol{B}}^{k-1} + \delta_k(\hat{\boldsymbol{B}}^0 - \tilde{\boldsymbol{B}}_\kappa^{k-1}), \tag{59}$$

where $\tilde{\boldsymbol{B}}_\kappa^{k-1}$ is a matrix with its non-zero entry index the same as $\hat{\boldsymbol{B}}^0$ and its non-zero entry value the same as $\tilde{\boldsymbol{B}}^{k-1}$.

Because the neighboring entry in the Hankel matrix is equally spaced depending on the minimum interval of the non-uniform pilot, if the minimum interval exceeds the interval between two OFDM symbols, after iteration reaches its maximum value, we leverage the lowpass DFT interpolation method mentioned before to reconstruct all the future CSI based on the predicted output of MC. The details are listed in Algorithm 3.

V. SIMULATION RESULTS

A symbol level simulation is conducted where BS will operate beamforming to the transmitted signal according to the predicted channel. Eigen-based method [28] is used as the precoding method and MMSE is used at the user end as the equalization method.

The proposed algorithms are applied in both simulated channels and real channels acquired from channel sounding. Firstly, the COST2100 channel model [21] is used to generate a simulated channel with its setting listed in Table I. We choose semi-urban-300M as our channel scenario and set the carrier frequency and bandwidth to 3.5 GHz and 20 MHz respectively which is a default setting in the long term evolution (LTE) system. The scattering environment consists of 25 clusters and each of them contains 5 MPC. BS is equipped with a 4×8 planar array and each antenna is unipolar and omni-directional. Both single and multi-user cases are considered and we set the

TABLE I
COST2100 SETTINGS.

Parameter	Value
Channel type	Semi-urban-300M
Center frequency	3.5 GHz
$|\Gamma|$	25
P	5
Bandwidth	20 MHz
N_t	32 (4×8 UPA)
N_r	4 ULA (single-user) or 2 ULA (multi-user)
BS position	$[0, 0, 10]$ m
User velocity	60 km/h or 120 km/h
User number	1 or 4
Single-user position	$[200, 400, 0]$ m
Multi-user position	$[200, -400, 0]$ m $[100, -500, 0]$ m $[300, -300, 0]$ m $[400, -200, 0]$ m

TABLE II
COMPARISON OF TNMSE BETWEEN UNIFORM AND NON-UNIFORM PILOTS AT DIFFERENT VELOCITIES.

	60 km/h	120 km/h
Even pilot	0.8234	1.0082
Non-uniform pilot	0.5298	0.6313
Random pilot	0.5844	0.6983

number of user antenna as 4 and 2 in single and multi-user cases, respectively. Two velocities, 60 km/h and 120 km/h, are simulated covering moderate and high-speed scenarios.

The relationship between channel prediction error versus user mobility is first shown in uniform pilot and non-uniform pilot cases with the proposed RAMBLE method. It predicts the CSI on the adjacent future symbol after the observation window, then the window will slide forward which represents the acquirement of newly updated CSI from channel estimation and the prediction will be operated again. The window size is set to 400 and the interval between symbols is 0.5 ms. In addition, the pilot symbol interval in the uniform pilot case is set to 5 ms which is a default SRS interval in the LTE protocol. For non-uniform pilot design, two strategies are tested: non-uniform pilot which meets the minimum interval criteria, and random pilot which ignores the previous criteria so the minimum spacing will be arbitrary. Both non-uniform patterns guarantee the same pilot density as in the uniform case so the density ρ will equal 0.025. It can be seen from Table II that non-uniform pilot can outperform uniform pilot in both 60 km/h and 120 km/h velocity, which implies that non-uniform pilot can provide more resolution to distinguish higher Doppler frequency. Meanwhile, it shows a better performance compared with the random pilot because the latter provides more resolution than it needs which will cause the pilot distribution in the observation window to become imbalanced.

Secondly, the advantage of MMV over SMV is depicted in Fig. 4 which shows TNMSE versus channel measurement noise in 60 km/h and 120 km/h user velocity. One can observe from Fig. 4 that MMV can outperform SMV in low and high SNR regimes, which implies that it can not only combat channel measurement noise but also offer more extra measurements because SMV deems each receive antenna as an individual and MMV leverages the extra information brought
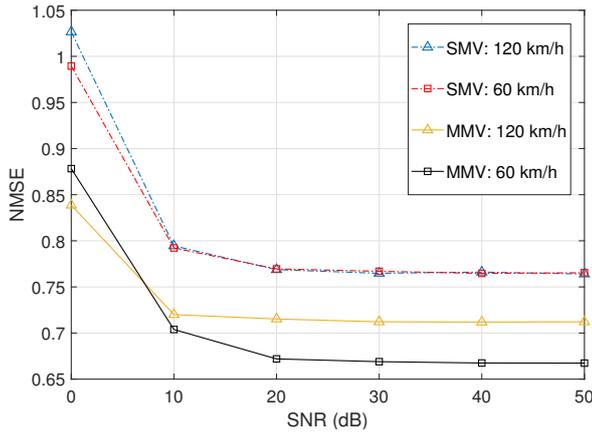
Fig. 4.   TNMSE vs. channel measurement noise level in 60 km/h and 120 km/h user speed.

Fig. 5.   Average throughput loss vs. SNR, noise-free channel measurements, $N_t = 32$.

by multiple measurements.

Next, to manifest the superiority of our algorithms, we take the algorithm proposed in [4] and bayesian regularized neural network (BRNN) [29] as our baselines, both of them use uniform pilot patterns. The former uses the vector Prony method to predict the channel on the future pilot symbol in the angular-delay domain, the order of the Prony method is set to 3. The BRNN is a neural network using Bayesian regularization, the complex CSI is separated into real and imaginary parts and stacked together as the network input. The number of hidden layer neurons is set to 50 and the tangent sigmoid is used as an activation function. We then use the lowpass/bandpass DFT interpolation method to predict the CSI on non-pilot symbols based on the profile of the Doppler spectrum. We assume here that the profile, center frequency, and bandwidth of the Doppler spectrum can be served as prior although they are not acquirable in practice. The window size of both algorithms is set the same as RAMBLE and MC as 400. The average throughput loss in single user case is depicted in Fig. 5, the performance of the Prony + interpolation method and BRNN + interpolation method is quite limited due to the aliasing of the Doppler spectrum. With the help of increased resolution brought by the non-uniform pilot, RAMBLE and MC achieve a good performance both in low and high SNR regimes.

In addition, the non-ideal characteristic of the actual system is also considered which has a channel estimation delay, meaning that RAMBLE and MC can not get the instantaneous channel measurement when the observation window is sliding forward. Therefore, the updated CSI is excluded from the observation window and the result is shown in Fig. 6. The performance of both proposed algorithms can still outperform the traditional methods to a great extent.

The curve of average throughput loss versus SNR in the multiuser case is depicted in Fig. 8, and the number of users is set to 4. It can be shown from the graph that the proposed algorithms can reduce the average throughput loss by about 10 percent and MC shows a better performance compared with RAMBLE.
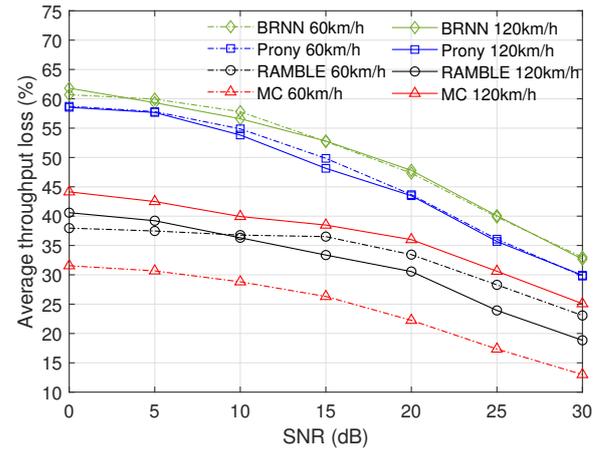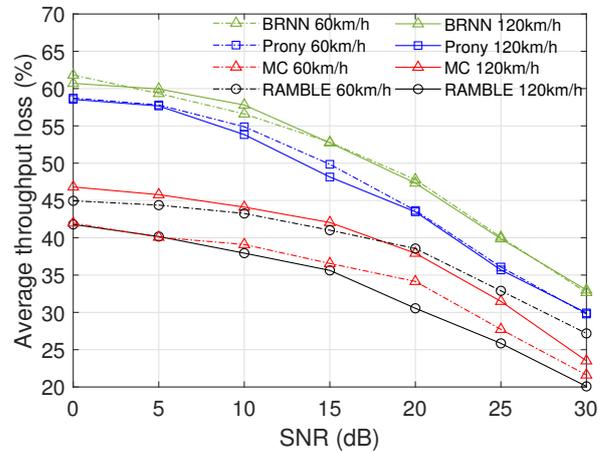
Fig. 6.   Average throughput loss vs. SNR, noise-free channel measurements with channel processing delay, $N_t = 32$.

To further showcase the proposed algorithms can be applied to the real measured channel. A channel-sounding campaign is conducted on campus. A 3.5 GHz channel sounder is adopted and it is equipped with a 3×8 bipolar antenna panel at the transmitter end and a two-element array at the receiver end. The channel is measured fixedly with 3 m interval and the route is depicted in Fig. 7 where the red star is the location of the transmitter and the red line represents the channel sounding route. To generate mobility from stationary channel data, we uniformly extract points from the route to simulate a mobile scenario, the equivalent speed will get greater if the spacing between the extracted points increases. The simulation result is shown in Fig. 9, both MC and RAMBLE outperform the Prony method for about 5 percent, which verify the applicability of the proposed algorithm. In addition, MC performs better than RAMBLE in real channel cases, which benefits from its direct usage of channel observations, avoiding incurring extra noise while estimating the intermediate parameters based on the channel model.

Fig. 7.   Channel sounding route and transmitter location.
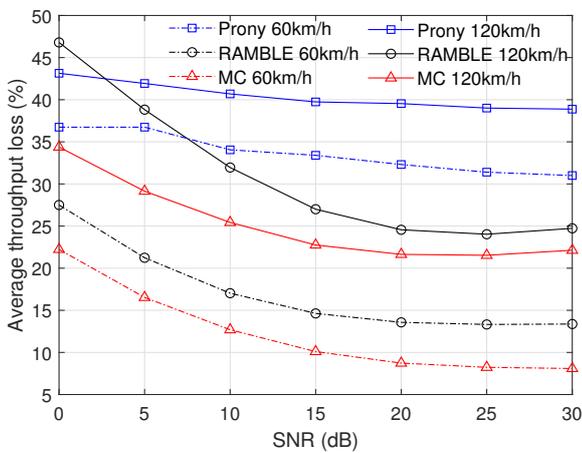
Fig. 8.   Average throughput loss vs. SNR in different velocities, noise-free channel measurements, 4 UEs, $N_t = 32$.
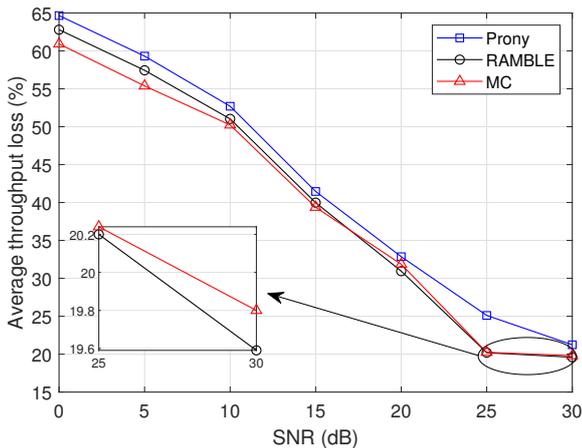
Fig. 9.   Average throughput loss versus SNR in real measured channel case.

VI. CONCLUSION AND FUTURE WORK

In this work, the existing problem of current channel prediction schemes is first pointed out, which is the Doppler aliasing problem in moderate to high mobility scenarios because of the density limitation of the uniform pilot pattern. Hence the reconstruction performance of non-pilot symbols will drastically degrade. To solve this, a non-uniform pilot pattern is proposed that differs from the uniform pilot pattern used in most communication protocols. Then two channel prediction methods are proposed, the first one uses compressive sensing to reconstruct the Doppler spectrum of each angular-delay domain grid. To solve the repetition problem induced by the orthonormal DFT transformation matrix, redundant dictionary, and band exclusion are introduced to mitigate the basis mismatching problem. To further combat the channel-measuring noise and increase recovery stability, multiple measurements from different receiving antennas are leveraged to promote the performance. To bridge the gap between the real channel and channel model, the second method is data-driven which considers the channel prediction task as a matrix completion problem constrained by the rank of the Hankel matrix constructed from measurements.

Simulation results show that the proposed methods can outperform the state-ofthe-art algorithms both in the simulated channel and real measured channel in terms of throughput. In the simulated multiuser channel setting, the proposed methods can reduce the average channel throughput loss by about 15 percent both in 60 and 120 km/h user speed. The applicability of the proposed algorithms, in reality, is also verified with a measured channel where the proposed algorithm can get about 5 percent channel throughput-loss gain compared with conventional algorithms.

In this paper, there is a latent assumption that the support of the Doppler spectrum in the observation window is time-invariant, which implies that the Doppler frequency of each cluster should not change greatly in a certain period of time. Therefore, a smaller window size is chosen when the mobility becomes larger, which will lessen the available channel measurements and degrade the recovery performance. However, if one extends the "Doppler stationary" to a broader sense, which means the variation of Doppler support is a fixed function of time, this kind of long-term stationary could enlarge the window size and further promote the performance. Secondly, the discussion of the non-uniform pilot pattern in this paper only considers the minimum spacing, the optimum pattern of the non-uniform pilot is still an open problem [30].

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge university press, 2005.

[2] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 86–105, 2007.

[3] M.-C. Lee, W.-H. Chung, and T.-S. Lee, "Generalized precoder design formulation and iterative algorithm for spatial modulation in MIMO systems with CSIT," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1230–1244, 2015.

[4] 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.

[5] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, 2013.

[6] S. Kashyap, C. Mollén, E. Björnson, and E. G. Larsson, "Performance analysis of (TDD) massive MIMO with kalman channel prediction," in *Proc. IEEE ICASSP*, 2017.

[7] A. Khrwat, B. Sharif, C. Tsimenidis, and S. Boussakta, "Channel prediction for precoded spatial multiplexing multiple-input multiple-output systems in time-varying fading channels," *IET Signal Process.*, vol. 3, no. 6, pp. 459–466, 2009.

[8] V. Arya and K. Appaiah, "Kalman filter based tracking for channel aging in massive MIMO systems," in *Proc. IEEE SPCOM*, 2018, pp. 362–366.

[9] K. E. Baddour and N. C. Beaulieu, "Improved pilot-assisted prediction of unknown time-selective rayleigh channels," in *Proc. IEEE ICC*, 2006.

[10] I. C. Wong and B. L. Evans, "Wlc43-5: Low-complexity adaptive high-resolution channel prediction for OFDM systems," in *Proc. IEEE GLOBECOM*, 2006.

[11] R. Adeogun, P. Teal, and P. Dmochowski, "Parametric schemes for prediction of wideband MIMO wireless channels," *arXiv preprint arXiv:1408.0581*, 2014.

[12] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 227–236, 2018.

[13] S. Moon, H. Kim, and I. Hwang, "Deep learning-based channel estimation and tracking for millimeter-wave vehicular communications," *J. Commun. Netw.*, vol. 22, no. 3, pp. 177–184, 2020.

[14] E. U. T. R. Access, "Physical channels and modulation, 3GPP TS 36.211," *V10*, vol. 2, 2011.

[15] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[16] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[17] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[18] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[19] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[20] D. Calvetti and L. Reichel, "Tikhonov regularization of large linear problems," *BIT Numer. Math.*, vol. 43, no. 2, pp. 263–283, 2003.

[21] L. Liu *et al.*, "The COST 2100 MIMO channel model," *IEEE Wireless Commun.*, vol. 19, no. 6, pp. 92–99, 2012.

[22] V. Michel and R. Telschow, "The regularized orthogonal functional matching pursuit for ill-posed inverse problems," *SIAM J. Numer. Analysis*, vol. 54, no. 1, pp. 262–287, 2016.

[23] A. Fannjiang and W. Liao, "Coherence pattern–guided compressive sensing with unresolved grids," *SIAM J. Imaging Sci.*, vol. 5, no. 1, pp. 179–202, 2012.

[24] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, 2006.

[25] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135–1146, 2012.

[26] J. Ying *et al.*, "Vandermonde factorization of hankel matrix for complex exponential signal recovery—application in fast NMR spectroscopy," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5520–5533, 2018.

[27] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[28] T. Chunchang, Y. Wenqian, F. Liang, W. Zhijie, and W. Yafeng, "On the performance of eigen based beamforming in LTE-advanced," in *Proc. IEEE PRIMRC*, 2009.

[29] D. T. Mirikitani and N. Nikolaev, "Recursive bayesian recurrent neural networks for time-series modeling," *IEEE Trans. Neural Netw.*, vol. 21, no. 2, pp. 262–274, 2009.

[30] R. Venkataramani and Y. Bresler, "Optimal sub-nyquist nonuniform sampling and reconstruction for multiband signals," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2301–2313, 2001.
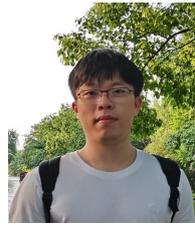
Yi Shi [S'21] (yishi1996@shu.edu.cn) received the B.E. degree from the School of Communication and Information Engineering, Shanghai University, in 2019, where he is currently pursuing the Ph.D. degree with the School of Communication and Information Engineering. His main research interests include channel prediction and application specific integrated circuit.

Xianling Wang [S'22] (lance_wang@shu.edu.cn) received the B.E. degree from the School of Communication and Information Engineering, Shanghai University, in 2020, where he is currently pursuing the Ph.D. degree with the School of Communication and Information Engineering. His main research interests include channel measurement, estimation and prediction.

Zhiyuan Jang [S'12-M'15] (jiangzhiyuan@shu.edu.cn) received the B.S. and Ph.D. degrees from the Electronic Engineering Department, Tsinghua University, China, in 2010 and 2015, respectively. He is currently a Professor at the School of Communication and Information Engineering, Shanghai University, Shanghai, China. He visited the WiDeS Group, University of Southern California, Los Angeles, CA, USA, from 2013 to 2014. He worked as an experienced researcher at Ericsson from 2015 to 2016. He visited ARNG at the University of Southern California, Los Angeles, CA, USA, from 2017 to 2018. He worked as a Wireless Signal Processing Scientist at Intel Labs, Hillsboro, OR, USA, in 2018. His current research interests include URLLC in wireless networked control systems and signal processing in MIMO systems. He serves as a TPC member for IEEE INFOCOM, ICC, GLOBECOM, and WCNC. He received the ITC Rising Scholar Award in 2020, the Best Paper Award at the IEEE ICC 2020, the best In-Session Presentation Award of IEEE INFOCOM 2019, and the Exemplary Reviewer Award of IEEE WCL in 2019. He serves as an Associate Editor for the IEEE/KICS Journal of Communications and Networks, and a Guest Editor for the IEEE IoT Journal.