Dealing with the Mobility Problem of Massive MIMO using Extended Prony’s Method

Haifan Yin†, Haiquan Wang*, Yingzhuang Liu†, David Gesbert§

†School of Electronic Information and Communications, Huazhong University of Science and Technology, 430074 Wuhan, China
*School of Communications Engineering, Hangzhou Dianzi University, 310018 Hangzhou, China
§Communications Systems Department, EURECOM, 06904 Biot, France
Email: yin@hust.edu.cn, tx_wang@hdu.edu.cn, liuyz@mail.hust.edu.cn, gesbert@eurecom.fr

Abstract—Massive MIMO is a key technology for 5th generation (5G) mobile communications. The large excess of base station (BS) antennas brings unprecedented spectral efficiency. However, during the initial phase of industrial testing, a practical challenge arises which undermines the actual deployment of massive MIMO and is related to mobility. In fact, testing teams reported that in moderate-mobility scenarios, e.g., 30 km/h of UE speed, the performance may drop 50% compared to the low-mobility scenario, a problem not foreseen by theoretical papers on the subject. In order to deal with this challenge, we propose a Prony-based angular-delay domain (PAD) prediction method, which is built on exploiting the angle-delay-Doppler structure of the multipath. Our theoretical analysis shows that when the number of base station antennas and the bandwidth are large, the prediction error of our PAD algorithm converges to zero for any UE velocity level, provided that only two accurate enough previous channel samples are available. Simulation results show that under the realistic channel model of 3GPP in rich scattering environment, our proposed method even approaches the performance of stationary scenarios where the channels do not vary at all.

Index Terms—massive MIMO, mobility, 5G, channel aging, channel prediction, Prony’s method

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) introduced in [1], is one of the key enablers of the 5G cellular systems. Compared to traditional MIMO with fewer BS antennas, massive MIMO can offer superior spectral efficiency and energy efficiency [2] at least in theory. The growing orthogonality between channel vectors of different UEs allows the base station to reject interference by inexpensive precoding methods, providing that Channel State Information (CSI) is known at base station. CSI acquisition is known to be a formidable problem in massive MIMO. An example of CSI acquisition issue is the pilot contamination problem. A rich body of literature has addressed this problem. The solutions vary from angular/amplitude domain discrimination [3] [4] [5], pilot coordination [3], multi-cell minimum mean square error (M-MMSE) [6] [7], etc.

Despite the technology hype and great expectations behind massive MIMO, some of the latest field trials have been more than disappointing when it comes to actual system performance (see [8] for example). In particular it appeared that CSI acquisition can be severely affected in mobility scenarios. This is related to the time-varying nature of wireless channel which itself limits its coherence time. In practical cellular networks, the processing delay at the base station is inevitable because of the highly sophisticated 5G protocol, scheduling, resource allocation, encoding/decoding, and channel training under UE power constraint. This implies that even in moderate-mobility scenarios, the processing delay can end up being larger than the coherence time, making it essentially unusable for multiuser beamforming. Channel aging effect is particularly harmful to massive MIMO, which relies on high precision CSI to achieve large multiplexing gains. It was for instance observed in industrial settings, that with a typical CSI delay of 4 milliseconds, the moderate-mobility scenario at 30 km/h leads to as much as 50% of the performance reduction versus in low-mobility scenario at 3 km/h. Some information theoretic efforts exploiting severely delayed CSI have been demonstrated to be possible but never tested out in practical 5G contexts [9], [10]. The effects of channel aging under a simple autoregressive model of the channel time variations were studied in [11]. A linear finite impulse reponse (FIR) Wiener predictor was proposed which could partially overcome the channel aging effect at the expense of high computational complexity.

In this paper, we revisit the problem of CSI acquisition by combining it with channel prediction algorithms. We propose a novel Prony-based angular-delay domain channel prediction algorithm by exploiting the structural information of the multi-path channel. The large number of BS antennas and the large bandwidth in 5G lead to higher resolution in both angular domain and delay domain. Our idea consists in exploiting this high resolution regime specifically. In practice the approach involves projecting the channel into an angular-delay grid, then capturing the channel variations in the angular-delay domain. Moreover, we propose to adopt here Prony’s method, traditionally used in the context of spectral analysis, for its ability to predict a uniformly sampled signal composed of damped complex exponentials. This feature turns out to be useful in the massive MIMO context because the training signal in 5G are normally periodic and the channel can be regarded as a sum of complex exponentials with each one corresponding to a path response having a Doppler component. We analyze the asymptotic performance of our PAD algorithm.
and prove that as the number of base station antennas and the bandwidth increase, the channel prediction error converges to zero, provided that only two accurate enough channel samples are available. Finally, since in practice, current channel estimates are noisy, we improve the performance of the PAD method by combining it with a denoising method using an adaptation of Tuft's-Kumaresan's method [12].

Simulations under the clustered delay line (CDL) channel model of 3GPP [13] shows that our proposed method at 60 km/h of UE speed is very close to the ideal case of a stationary setting. To the best of our knowledge, the study of channel prediction under such a realistic model of wideband massive MIMO has never been done before, and the high spatial-frequency resolution of 5G has never been fully exploited to solve the mobility challenge.

Notations: We use boldface to denote matrices and vectors. Specifically, $I$ denotes the identity matrix, $(X)^T$, $(X)^*$, and $(X)^H$ denote the transpose, conjugate, and conjugate transpose of a matrix $X$ respectively. $(X)^d$ is the Moore-Penrose pseudoinverse of $X$. $\|\cdot\|$ stands for the Frobenius norm. $\mathbb{E}\{\cdot\}$ denotes the expectation. $X \otimes Y$ is the Kronecker product of $X$ and $Y$. $\text{vec}(X)$ is the vectorization of the matrix $X$. $\text{diag}\{a_1, ..., a_N\}$ denotes a diagonal matrix or a block diagonal matrix with $a_1, ..., a_N$ at the main diagonal. $\hat{\cdot}$ is used for definition.

II. CHANNEL MODELS

For ease of exposition, we consider an arbitrary UE in a certain cell. We assume the antennas at the base station form a uniform planar array (UPA) with $N_r$ rows and $N_h$ columns as in commercial systems$^3$. Denote the number of antennas at the base station as $N_t$ and the number of antennas at the UE as $N_r$. It is clear that $N_t = N_rN_h$. The network operates in TDD mode and the UL and DL occupy the same bandwidth, which consists of $N_f$ subcarriers with spacing $\Delta f$. The channel is composed of $P$ multipath, with each path having a certain angle, delay, Doppler, and complex amplitude. We denote the elevation departure angle, azimuth departure angle, elevation arrival angle, and azimuth arrival angle of the $p$-th path as $\theta_{p,ZOD}$, $\phi_{p,ZOD}$, $\theta_{p,ZOA}$, and $\phi_{p,AOA}$ respectively. The DL channel at a certain time $t$ and a subcarrier with frequency $f$ is denoted as $h_{u,s}(f,t) \in \mathbb{C}^{N_r \times N_t}$. According to [13], the channel between the $s$-th base station antenna and the $u$-th UE antenna is modeled as

$$h_{u,s}(f,t) = \sum_{p=1}^{P} \beta_p e^{j2\pi f \tau_p} e^{-j2\pi f \varphi_p} e^{-j2\pi f \varphi_p} e^{-j2\pi f \varphi_p} e^{-j2\pi f \varphi_p},$$

where $\beta_p$ and $\tau_p$ are the complex amplitude and the delay of the $p$-th path respectively. $\lambda_0$ is the wavelength of center frequency. $\hat{r}_{tx,p}$ is the spherical unit vector with azimuth arrival angle $\phi_{p,AOA}$ and elevation arrival angle $\theta_{p,ZOA}$:

$$\hat{r}_{tx,p} \triangleq \begin{bmatrix} \sin \theta_{p,ZOA} \cos \phi_{p,AOA} \\ \sin \theta_{p,ZOA} \sin \phi_{p,AOA} \\ \cos \theta_{p,ZOA} \end{bmatrix}.$$  \hspace{1cm} (2)

Likewise, $\hat{r}_{tx,p}$ is the spherical unit vector defined as:

$$\hat{d}_{r,s} \triangleq \begin{bmatrix} \sin \theta_{p,ZOD} \cos \phi_{p,AOD} \\ \sin \theta_{p,ZOD} \sin \phi_{p,AOD} \\ \cos \theta_{p,ZOD} \end{bmatrix}.$$ \hspace{1cm} (3)

$\hat{d}_{r,s}$ is the $u$-th UE antenna’s location vector which contains the 3D cartesian coordinate. Similarly, $\hat{d}_{r,s}$ is the location vector of the $s$-th base station antenna. The last exponential term $e^{j2\pi f t}$ is the Doppler of the $p$-th path, where $t$ denotes time. $\omega_p$ is defined as $\omega_p \triangleq \hat{r}_{tx,p}^T \vec{v} / \lambda_0$, where $\vec{v}$ is the UE velocity vector with speed $v$, travel azimuth angle $\phi_v$, and travel elevation angle $\theta_v$:

$$\vec{v} \triangleq v \begin{bmatrix} \sin \theta_v \cos \phi_v \\ \sin \theta_v \sin \phi_v \\ \cos \theta_v \end{bmatrix}.$$ \hspace{1cm} (4)

An illustration of the coordinate system is shown in Fig. 1.

![Fig. 1. Definition of the coordinate system.](https://example.com/fig1.png)

Let $h_u(f,t) \in \mathbb{C}^{N_r \times N_t}$ denote the channel between all base station antennas and the $u$-th UE antenna at time $t$ and frequency $f$. We write the channels at all $N_f$ subcarriers in a matrix form:

$$H_u(t) \triangleq \begin{bmatrix} h_u^T(f_1,t) & h_u^T(f_2,t) & \cdots & h_u^T(f_{N_f},t) \end{bmatrix},$$ \hspace{1cm} (5)

where $f_i$ is the frequency of the $i$-th ($1 \leq i \leq N_f$) subcarrier.

III. DEALING WITH MOBILITY PROBLEM

A. The mobility problem of massive MIMO

As is well known, channel time variability can create inter-user interference induced by a precoder which is computed based on aging CSI. This impediment can be mitigated by anticipating the future channel variations. While predicting the future fading state of a wireless channel is a very challenging problem, the accounting of the specific space-time structure of the channel which arises in a broadband context (as in 5G) opens fresh perspectives for improvement.

B. A review of Prony’s method

Prony’s method proposed by Gaspard Riche de Prony in 1795 is a useful tool to analyze a uniformly sampled signal composed of a number of damped complex exponentials [14] and extract valuable information (e.g., the amplitudes and frequencies of the exponentials) which can be used for prediction.
A review of this method is given below. Suppose we have \( K \) samples of data \( y(k) \) which consist of \( N \) exponentially damped signals:

\[
y(k) = \sum_{n=1}^{N} \beta_n e^{(-\alpha_n + j2\pi f_n)k}, 0 \leq k \leq K - 1,
\]

where \( \alpha_n \) (positive) and \( f_n \) (\( 1 \leq n \leq N \)) are the pole damping factor and pole frequency respectively. \( \beta_n \) (\( 1 \leq n \leq N \)) is the complex amplitude. Note that in the special case of channel prediction, \( y(k) \) can be regarded as the uniformly sampled channel estimate. Define the following polynomial:

\[
P_0(z) = \prod_{n=1}^{N} (z-e^{s_n}) = \sum_{n=0}^{N} p_n z^n, z \in \mathbb{C},
\]

where \( s_n = -\alpha_n + j2\pi f_n \) for \( n = 1, \cdots, N \). It is clear that \( p_N = 1, e^{s_n}, (n = 1, \cdots, N) \) are zeros of \( P_0(z) \). For an arbitrary \( m \in \mathbb{N} \), one has

\[
N \sum_{n=0}^{N-1} p_n y(n+m) = -y(N+m), m \in \mathbb{N}.
\]

Thus, we may obtain the coefficients \( p_n \) with the \( 2N \) sampled data by solving

\[
\begin{bmatrix}
y(0) & y(1) & \cdots & y(N) \\
y(1) & y(2) & \cdots & y(N+1) \\
\vdots & \vdots & \ddots & \vdots \\
y(N-1) & y(N) & \cdots & y(2N-1)
\end{bmatrix}
\begin{bmatrix}
p_0 \\
p_1 \\
\vdots \\
p_N
\end{bmatrix}
= \mathbf{0}.
\]

Or equivalently, the following linear equations:

\[
\mathbf{Y} \mathbf{p} = -\mathbf{h},
\]

where \( \mathbf{Y} \) is a square Hankel matrix.

\[
\mathbf{Y} \triangleq \begin{bmatrix}
y(0) & y(1) & \cdots & y(N-1) \\
y(1) & y(2) & \cdots & y(N) \\
\vdots & \vdots & \ddots & \vdots \\
y(N-1) & y(N) & \cdots & y(2N-2)
\end{bmatrix},
\]

\[
\mathbf{p} \triangleq \begin{bmatrix}
p_0 \\
p_1 \\
\vdots \\
p_N
\end{bmatrix}^T, \quad \mathbf{h} \triangleq \begin{bmatrix}
y(N) \\
y(N+1) \\
\vdots \\
y(2N-1)
\end{bmatrix}^T.
\]

The least squares solution to Eq. (11) is given by

\[
\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \| \mathbf{Y} \mathbf{p} + \mathbf{h} \| ^2 = -\mathbf{Y}^\dagger \mathbf{h}.
\]

Note that we may need \( K \geq 2N \) samples to obtain all the coefficients \( p_n, \ n = 0, \cdots, N - 1 \).

C. Prony-based angular-delay domain channel prediction

As shown in Eq. (1), the channel is composed of \( P \) paths, each path has a Doppler term \( e^{j\omega \tau} \), \( \tau = 1, \cdots, P \). In practice, the number of paths can be large, e.g., several hundreds in rich scattering environment, which makes the prediction difficult. In order to cope with this problem, we propose a Prony-based angular-delay domain (PAD) channel prediction algorithm. The main idea is to convert the channel into another domain where the Doppler terms of different paths are less intertwined with each other. We choose this domain in such a way that it reflects the geometry of the antenna array and the wideband delay response structure of the channel. Denote a discrete Fourier transform (DFT) matrix of size \( N \times N \) by

\[
\mathbf{W}(N) \triangleq \frac{1}{\sqrt{N}} \begin{bmatrix}
\omega^{0.0} & \omega^{0.1} & \cdots & \omega^{(N-1).0} \\
\omega^{0.0} & \omega^{0.1} & \cdots & \omega^{(N-1).1} \\
\vdots & \vdots & \ddots & \vdots \\
\omega^{(N-1).0} & \omega^{(N-1).1} & \cdots & \omega^{(2N-1).(N-1)}
\end{bmatrix},
\]

where \( \omega \triangleq e^{-2\pi j/N} \). Since UPA antenna array is considered, we define a spatial orthogonal basis \( \mathbf{S} \) as

\[
\mathbf{S} \triangleq \mathbf{W}(N_h) \otimes \mathbf{W}(N_r),
\]

where \( N_h \) and \( N_r \) are the number of columns and the number of rows of antennas on the UPA respectively. Note that we assume the antenna numbering in \( \mathbf{h}_u(f, t) \) is first columns then rows. The frequency orthogonal basis is defined as \( \mathbf{F} \triangleq \mathbf{W}(N_f) \). We convert the channel Eq. (5) into angular-delay domain with the orthogonal basis of \( \mathbf{S} \) and \( \mathbf{F} \).

\[
\mathbf{G}_u(t) \triangleq \mathbf{S}^H \mathbf{H}_u(t) \mathbf{F}.
\]

\( \mathbf{G}_u(t) \) is in fact the orthogonal projection of the channel on the angular-delay grid of size \( N_s \times N_f \). Due to limited angular spread as well as the limited delay spread of the wireless propagating environment [3] [15], the channel in angular-delay domain is sparse. In other words, most of the elements in matrix \( \mathbf{G}_u(t) \) are very close to zero. As a result we may ignore the insignificant elements in \( \mathbf{G}_u(t) \) and focus on the predictions of the significant ones. Let \( \mathbf{g}_u(t) = \mathbf{vec}(\mathbf{G}_u(t)) \) and \( \mathbf{g}_u(t) \) be the re-arranged \( \mathbf{g}_u(t) \) with its absolute values in non-increasing order. The number of non-negligible angular-frequency positions \( N_s \) is defined as

\[
N_s = \arg \min_{N_s} \left\{ \sum_{u=1}^{N_r} \sum_{l=0}^{L} \left| \mathbf{g}_u(t_l, n) \right|^2 \geq \gamma \sum_{u=1}^{N_r} \sum_{l=0}^{L} \| \mathbf{G}_u(t_l) \|_F^2 \right\},
\]

where \( \mathbf{g}_u(t_l, n) \) is the \( n \)-th entry of \( \mathbf{g}_u(t_l) \), \( \gamma \) is a positive value that is close to 1. The physical meaning of \( \gamma \) is the ratio between the sum power of significant elements and the total power of the channel. An empirical value of \( \gamma \) can be 0.99. Note that \( N_s \) is normally much smaller than the size of the matrix \( \mathbf{G}_u(t) \): \( N_s \ll N_s \times N_f \). Thus by ignoring the insignificant elements, we may greatly reduce the computational complexity in channel prediction. We use \( \mathbf{g}_u(t), (n = 1, \cdots, N_s) \) to denote the \( n \)-th significant entry, which is located at the \( r(n) \)-th row and the \( c(n) \)-th column of \( \mathbf{G}_u(t) \). The channel can be approximated as

\[
\mathbf{\tilde{H}}_u(t) = \sum_{n=1}^{N_s} g_{u,n}(t) S_{r(n)} F_{c(n)},
\]
where \( s_t \) is the \( i \)-th column of \( S \) and \( f_i \) is the \( i \)-th column of \( F \). Assuming the CSI delay \( T_d = N_d \Delta T, N_d \in \mathbb{N} \), we seek to predict the channel \( H_u(t_{L+N_d}) \) using the known samples \( H_u(t_0), H_u(t_1), \ldots, H_u(t_L) \). Without loss of generality, we assume \( L \) is odd and let the order of the predictor \( N = (L + 1)/2 \). For a certain \( n, 1 \leq n \leq N_s \), we may obtain the Prony coefficients by solving

\[
\mathcal{G}(u,n)p(u,n) = -g(u,n),
\]

where \( \mathcal{G}(u,n) \triangleq \left[ \begin{array}{cccc}
g_{u,n}(t_0) & g_{u,n}(t_1) & \cdots & g_{u,n}(t_{N-1}) \\
g_{u,n}(t_1) & g_{u,n}(t_2) & \cdots & g_{u,n}(t_N) \\
\vdots & \vdots & \ddots & \vdots \\
g_{u,n}(t_{N-1}) & g_{u,n}(t_N) & \cdots & g_{u,n}(t_{2N-2})
\end{array} \right]
\]

1. Compute the angular-delay domain channel \( G_u(t_l) \) for \( l = 0, \ldots, L \) and \( u = 1, \ldots, N_u \) according to Eq. (17).
2. Find the significant values \( g_{u,n}(t_l) \) and their positions \( r(n) \) and \( c(n) \) for \( u = 1, \ldots, N_u, n = 1, \ldots, N_s, l = 0, \ldots, L \).
3. for \( u = 1, \ldots, N_u \):
4. for \( n = 1, \ldots, N_s \):
5. Compute the least squares estimate of the Prony coefficients as in Eq. (22);
6. Repeat Eq. (23) \( N_d \) times to compute the prediction \( \hat{g}_{u,n}(t_{L+N_d}) \);
7. end for
8. Reconstruct the channel prediction at \( t_{L+N_d} \) as in Eq. (18) with \( g_{u,n}(t) \) replaced by \( \hat{g}_{u,n}(t_{L+N_d}) \);
9. end for

**D. Performance analysis of the PAD algorithm**

The asymptotical performance of our PAD algorithm is now analyzed. In order to facilitate the analysis, we make a mild technical assumption that for any two paths \( p \neq q \), at least one of the three properties are non-identical: their elevation departure angles, their azimuth departure angles, and their delays. We denote the noisy observation of \( H_u(t) \) as \( \tilde{H}_u(t) \) and the predicted channel at \( t_{L+N_d} \) as \( \hat{H}_u(t_{L+N_d}) \). Our main theoretical result is shown in Theorem 1.

**Theorem 1** For an arbitrary delay \( N_d \in \mathbb{N}^+ \) and any UE velocity level, the asymptotic performance of the PAD algorithm yields:

\[
\lim_{N_v, N_h, N_f \to \infty} \frac{\| \hat{H}_u(t_{L+N_d}) - H_u(t_{L+N_d}) \|^2_F}{\| H_u(t_{L+N_d}) \|^2_F} = 0, \quad (24)
\]

under the condition that two channel samples are accurate enough, i.e., \( \forall k = L - 1, L, \ldots, N_d \),

\[
\lim_{N_v, N_h, N_f \to \infty} \frac{\| \hat{H}_u(t_k) - H_u(t_k) \|^2_F}{\| H_u(t_k) \|^2_F} = 0. \quad (25)
\]

**Proof:** See the full-length version [16].

Note that condition Eq. (25) may require some non-linear signal processing techniques. See [5] as an example of how this condition can be fulfilled for a multi-cell massive MIMO scenario in the presence of pilot contamination.

**E. Dealing with channel estimation error**

The channel estimate at the base station is always corrupted by noise, which undermines the performance of Prony’s method. Thus we propose to deal with noise with a supplementary method, which relies on the subspace structure of the channel sample matrix and the second-order long-term statistics of the noisy channel samples. It consists in the following two ingredients:

1) Tufts-Kumaresan’s method: The main idea of the Tufts-Kumaresan’s method [12] is to apply singular value decomposition (SVD) to the sample matrix, i.e., Eq. (12) or Eq. (20), of the linear prediction equations, and then remove the contributions of small singular values. Taking the estimate of \( \tilde{p}(u,n) \) for example, the SVD of \( \tilde{G}(u,n) \) can be written as:

\[
\tilde{G}(u,n) = U(u,n)\Sigma(u,n)V^H(u,n)
\]

where \( \Sigma_u(u,n) \) only contains the significant singular values of \( \tilde{G}(u,n) \). The Tufts-Kumaresan’s estimate of the Prony coefficients are given by

\[
\tilde{p}_k(u,n) = -V_s(u,n)\Sigma_s^{-1}(u,n)U^H_s(u,n)\tilde{g}(u,n). \quad (28)
\]

Note that \( \Sigma_s(u,n) \) can be obtained in a way that the minimum number of singular values satisfies

\[
\text{tr} \{ \Sigma_s(u,n) \} \geq \gamma_k \text{tr} \{ \Sigma(u,n) \}, \quad (29)
\]

where the threshold \( \gamma_k \) is no greater than 1, i.e., \( \gamma_k = 0.99 \).
2) Channel denoising with statistical information: The noisy channel samples between all base station antennas and the $u$-th UE antenna at time $t$ and frequency $f$ can be modeled as $\tilde{h}_u(f, t) = h_u(f, t) + n_u(f, t)$, \hfill (30)

where $h_u(f, t)$ is the accurate channel and $n_u(f, t)$ is the independent and identically distributed (i.i.d.) complex Gaussian noise with zero-mean and covariance $\sigma_n^2$. It is easy to obtain the covariance matrix of the noisy channel at the base station:

$$\tilde{R} = \mathbb{E} \left\{ \tilde{H}^H(f, t) \tilde{H}(f, t) \right\},$$ \hfill (31)

where the expectation is taken over time, frequency, or both. $\tilde{H}(f, t)$ is defined as

$$\tilde{H}(f, t) \triangleq [ \tilde{h}_1^H(f, t) \ \tilde{h}_2^H(f, t) \ \cdots \ \tilde{h}_{N_u}^H(f, t) ]^T.$$

From Eq. (30) we have

$$\tilde{R} = R + N_u \sigma_n^2 I,$$ \hfill (33)

where

$$R = \mathbb{E} \left\{ H^H(f, t) H(f, t) \right\},$$ \hfill (34)

with $H(f, t)$ being the accurate counterpart of $\tilde{H}(f, t)$. Due to the large number of BS antennas and the limited scattering environment, the channel covariance matrix $R$ has a low-rankness property \cite{3} \cite{15}, which means a fraction of the eigenvalues of $R$ are very close to zero. Thus we may exploit this property to have an estimate of the power of noise. The eigen-decomposition of $\tilde{R}$ is written as $\tilde{R} = \tilde{U} \Sigma \tilde{U}^H$ where $\Sigma = \text{diag}\{\sigma_1, ..., \sigma_{N_u}\}$. The estimate of the noise power $\sigma_n^2$ can be obtained by averaging the smallest eigenvalues of $\tilde{R}$. A linear filter $\mathbf{W}$ can be derived for channel denoising purpose:

$$\mathbf{W} = \arg \min_{\mathbf{W}} \mathbb{E} \left\{ \| \tilde{H}(f, t) \mathbf{W} - H(f, t) \|^2_F \right\}. $$ \hfill (35)

The solution is given by Proposition 1.

**Proposition 1** The linear solution to the optimization problem of Eq. (35) yields

$$\mathbf{W} = \tilde{U} \mathbf{D} \tilde{U}^H,$$ \hfill (36)

where $\mathbf{D}$ is a diagonal matrix with its $i$-th ($i = 1, \cdots, N_u$) diagonal entry being $\frac{\sigma_i - N_u \sigma_n^2}{\sigma_i}$.

**Proof:** The derivation is based on the linear minimum mean square error (LMMSE) criterion and is straightforward. \hfill $\square$

Note that the covariance matrix is computed based on samples of all $N_u$ UE antennas, since the scattering environments experienced by all co-located $N_u$ antennas is very similar. The denoising filter $\mathbf{W}$ can also be built for each UE antenna, however with higher complexity.

IV. NUMERICAL RESULTS

This section contains simulation results of our proposed channel prediction schemes. The simulation parameters are listed in Table I. Since we adopt the CDL-A channel model, the number of multipath is 460, i.e., for each UE, there are 23 clusters of multipath with each cluster containing 20 rays. Consider a typical setting of 5G at 3.5 GHz with 30 kHz of subcarrier spacing and normal cyclic prefix. Each slot, with a duration of 0.5 ms, contains 14 OFDM symbols \cite{17}. We assume that UE sends one SRS signal in each slot, which means one channel sample is available every 0.5 ms. The
structure (such as the one in [13]) do not necessarily conform with the simple AR(1) channel aging model.

Now the channel estimation error is taken into consideration, assuming the ratio between the channel power and the power of estimation noise is 20 dB, we plot the performances of our PAD algorithm with the denoising methods given by Sec. III-E in Fig. 3. Our proposed PAD algorithm with denoising methods in moderate mobility scenario of 30 km/h is very close to the low-mobility scenario of 3 km/h.

V. CONCLUSIONS

In this paper we addressed the practical challenge of massive MIMO - the mobility problem. We proposed Prony-based angular-delay domain channel prediction method which is based on the specific angle-delay-Doppler structure of the channel and relies on the high spatial-frequency resolution in 5G massive MIMO. Our theoretical analysis proves that the proposed PAD method is able to achieve asymptotically error-free prediction, provided that only two accurate channel samples are available. We also proposed a supplementary channel denoising method for PAD algorithm to deal with channel sample error. Simulation results show that in moderate mobility scenarios, our proposed methods achieve nearly ideal performance of stationary setting.

REFERENCES