Adaptive Data-Aided Time-Varying Channel Tracking for Massive MIMO Systems

Ribhu Chopra, *Member, IEEE*, Chandra R. Murthy, *Fellow, IEEE*, Kumar Appaiah, *Senior Member, IEEE*

***Abstract*—The time varying nature of the wireless propagation channel causes a mismatch between the true channel at the time of data transmission and its available estimate based on previously received pilot symbols, and is known to impair the performance of the massive multiple input multiple output (MIMO) systems. In this paper, we develop and evaluate adaptive data aided channel tracking and data detection algorithms to counter the effects of channel aging for uplink and downlink massive MIMO systems. We first present a recursive least squares (RLS) algorithm for tracking the matrix uplink channel at the base station (BS), and derive bounds on its MSE performance. We also derive a linear complexity stochastic gradient descent (SGD) algorithm for tracking the uplink channel, along with its performance bounds. Following this, we develop RLS and SGD based algorithms for tracking the scalar effective downlink channel at each UE, and derive their performance guarantees. Finally, via Monte Carlo simulations, we validate the efficacy of the algorithms in terms of their mean squared error performance, and demonstrate the gains achievable by channel tracking in the form of the improvement in the symbol error rates.**

***Index Terms*—Massive MIMO, channel aging, recursive least squares, least mean squares**

I. Introduction

Massive multiple input multiple output (mMIMO) is the name given to a cellular communication system architecture with a large number of user equipments (UEs) being served over the same time-frequency resource by an even larger number of centrally located base station (BS) antennas [1]. Early results on mMIMO [2]–[5] have established that the availability of a large number of degrees of freedom in such a system leads to quasi orthogonality among the channels to the different users. Therefore, under the assumption of availability of accurate channel state information (CSI) at the BS, conjugate beamforming and combining potentially produce orthogonal effective channels to/from the UEs [1]. Due to this, mMIMO systems can potentially offer high spectral, energy, and computational efficiencies, at the cost of a large number of BS antennas. However, this seemingly simple trade off is only valid under the assumption that accurate CSI is available at the BS [1], [6]. In practical systems, this assumption is violated due to multiple causes, including additive noise in channel estimation [4], pilot contamination [7], reciprocity calibration imperfections [8]–[10] and channel aging [11].

R. Chopra is with the Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Assam, India. (email: ribhufec@iitg.ac.in). C. R. Murthy is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. (email: cmurthy@iisc.ac.in). K. Appaiah is with the Indian Institute of Technology Bombay, Mumbai, Maharashtra, India. (email: akumar@ee.iitb.ac.in)

Conventionally, the temporal variations of communication channels are modeled using the block fading channel model. Under this model, the channel coefficients remain unchanged for one coherence interval, and take independent values from the same distribution during the next coherence interval. However, this model fails to capture the continuous variation in the propagation channel due to UE movement [11] and phase noise [12]. It is therefore more appropriate to model the time variations in the channel as a slow temporal drift rather than a jump from one set of channel coefficients to another. It can be observed that under such a model, the channel estimate available at any node ages, or becomes obsolete, over time: this phenomenon is termed as *channel aging*. In an mMIMO system, channel aging limits the rates achievable by the UEs, especially in high mobility scenarios [11]–[14]. The effect of channel aging is critically dependent on several factors, including user mobility, system dimensions [15], and the underlying communication protocol [16].

As mentioned above, when the initial channel estimates at the BS obtained from pilot signals at the start of the frame are fairly accurate, accurate data detection is possible via simple linear processing in mMIMO systems. Therefore, the first few data symbols can be detected with high fidelity in both uplink and downlink directions. These symbols can then be potentially used as virtual pilots to track the variations in the corresponding uplink/downlink channels over time [17], thus mitigating the effects of channel aging. In this paper, we develop and evaluate different adaptive filtering algorithms for tracking aging channels in mMIMO systems.

A. Related Work

The past works that study the effects of channel aging in mMIMO systems mainly consider time division duplex (TDD) systems, as they allow one to exploit channel reciprocity to reduce the training overheads [11]–[14], [18], [19]. Also, the focus has been on the spectral efficiency (SE) achievable via linear signal processing techniques for precoding and data detection. Both the sum SE [12]–[14] and the average per user SE [15], [16] under aging channels have been analyzed. The large system dimensions of mMIMO enables the use of asymptotic techniques such as deterministic equivalent (DE) analysis [13], [15] and the use-and-then-forget bounds [16], to accurately characterize the performance of mMIMO systems under aging. The authors in [12] also discuss the non-asymptotic achievable rates for mMIMO systems under channel aging, incorporating the effect of phase noise. The key message of these papers is that, although the propounded

benefits of mMIMO such as power scaling [20] still hold in the presence of channel aging, increased user mobility and the consequent expedited aging of channels [21] can lead to severe performance losses. In the context of orthogonal frequency division multiplexing (OFDM)-based communications, the impact of user mobility on the uplink performance with frequency selective channels was analyzed in [22]. However, none of the above-mentioned papers deal with *tracking* the temporal variations in a massive MIMO channel, or analyze the performance of channel tracking algorithms.

Recently, the idea of using Kalman filtering based channel estimators for tracking aging channels was introduced in [23], [24], where the channel estimates from each frame were preserved, and updated using the pilot symbols received in the next frame. The idea of pilot-aided channel tracking has also been explored using machine learning based techniques in [25]–[27]. For slowly varying channels, these schemes improve the quality of the channel estimates, but the performance degrades when the channel is fast-varying. In [17], we developed Kalman filtering based data aided channel tracking algorithm for mMIMO systems with aging channels. We showed that this approach mitigates the losses due to channel aging, resulting in better SEs both in the uplink and the downlink. However, the Kalman filtering approach, although mean squared error (MSE)-optimal, requires exact knowledge of the second order channel statistics at the BS as well as at the UEs, and is also computationally intensive due to the matrix inverses involved. These factors limit the applicability of this approach and reduce the benefits obtained. In this paper, we overcome these issues by developing and analyzing alternative low complexity and adaptive channel tracking algorithms.

B. Contributions

We develop adaptive CSI tracking algorithms for an mMIMO system under aging channels. The problem of tracking a time-varying mMIMO channel is challenging due to the large system dimensions involved. We therefore rely on results from random matrix theory [11], [28] to obtain low complexity data detection procedures. The detected data symbols are then used to update the channel estimates via recursive least squares (RLS) and stochastic gradient descent (SGD) based methods.

We first discuss the problem of adaptive channel tracking for *uplink* mMIMO channels. We note that under the assumption of a static channel and known transmitted data symbols, channel estimation can be posed as a least squares regression problem without requiring the knowledge of the second order channel statistics. Furthermore, in case of sequentially arriving symbols, this least squares problem can be solved via the recursive least squares (RLS) algorithm, involving a rank-1 update of the available channel estimate based on the detected data symbols (See Theorem 1), which makes the approach computationally inexpensive. We then invoke the direct averaging approximation [29] to derive the MSE performance of this RLS based channel tracking algorithm under aging channels (See Theorem 2).

Noting that the computational complexity of the RLS algorithm is quadratic in the the number of users, we develop an even lower complexity solution for the channel tracking problem. To elaborate, we revisit the MMSE channel estimation problem and develop a data-aided stochastic gradient descent (SGD) based solution (See Theorem 3). We show that this results in a linear computational complexity in both the number of BS antennas as well as the number of users. We analyze the tracking performance of SGD based channel tracking in the uplink in Theorem 4. In both cases, our analysis accounts for the effect of error propagation due to incorrectly detected data symbols.

We discuss the symbol estimation procedure and evaluate the underlying probabilities of error in Sec. III-C. Finally, we evaluate the performance of the proposed channel tracking algorithms in Section III-D and prescribe optimal values for the tracking parameters. We also empirically illustrate the benefit of channel tracking in terms of the overall system performance by evaluating the symbol error rates (SERs) with and without adaptive channel tracking.

Next, we discuss *downlink* time-varying channel estimation, and derive update equations for data-aided tracking of the effective scalar downlink channel, using both the RLS and SGD formulations (See Theorems 5 and 7 respectively), and analyze their tracking performances (See Theorems 6 and 8.). Again, we evaluate the probabilities of error in Sec. IV-C and use Monte Carlo simulations to illustrate the benefits of data-aided channel tracking in terms of the MSE and SER performance of these algorithms in Sec. IV-E.

The key takeaway of this work is that, in the context of aging mMIMO channels, adaptive tracking results in a significantly better performance compared to no channel tracking at virtually no additional computational complexity at the BS and the UEs. Furthermore, and in contrast with Kalman tracking [17], it is computationally simpler by 1-2 orders of magnitude (see Fig. 7) and does not require the knowledge of the second order channel statistics. Explicitly, the advancement in this paper over our previous work [17] are the development of low-complexity adaptive tracking algorithms, their careful theoretical performance analysis in terms of the MSE in the channel estimates, both for uplink and downlink channel estimation, and both for the RLS and SGD algorithms. Further, we present simple and interpretable expressions for the steady-state MSE that uncover the dependence of the tracking performance on the key system parameters. In the next section, we present our system model, which is the point of departure of this work.

II. System Model

We consider a canonical time division duplexed (TDD) single cell mMIMO system consisting of an N antenna base station (BS) serving K single antenna user equipments (UEs) over rich scattering Rayleigh fading channels. We represent the path loss coefficient and the velocity of the kth UE by β_k and v_k, respectively. As illustrated in the frame structure shown in [17, Fig. 1], each frame consists of a total of T channel uses, and is subdivided into an uplink subframe of T_u channel uses, and a downlink subframe of T_d channel uses. Each of these sub-frames is further divided into a training phase of duration τ_x and a transmission phase of duration

$T_x - \tau_x$, $x \in \{u, d\}$. In the TDD mode of operation, it is conventional to assume that $\tau_u \geq K$ to ensure that orthogonal pilot sequences can be assigned to the UEs, and that $\tau_d \geq 0$. For convenience, we assume the time index of the symbols, n, to take values between $n = -\tau_x + 1$ and $n = T_x - \tau_x$, within each subframe. With this convention, in both the uplink and downlink subframes, the first τ_x channel uses correspond to the training phase, and data transmission starts at $n = 1$. We assume knowledge of the second-order channel statistics and the AR coefficients at the BS, as these typically vary slowly with time. The estimation of these parameters is a separate problem and is beyond the scope of this paper.

Denoting the fast fading coefficient of the channel between the ith BS antenna and the kth UE at the nth instant as $h_{ik}[n]$, we can write the effective channel between the ith BS antenna and the kth UE as $\sqrt{\beta_k} h_{ik}[n]$. We denote the vector channel to the kth UE by $\mathbf{h}_k[n] \triangleq [h_{1k}[n], \dots h_{Nk}[n]]^T$. The temporal evolution of the channel is modeled by a first order auto-regressive (AR-1) process [30], [31]:

$$\mathbf{h}_k[n] = \rho_k[n]\mathbf{h}_k[0] + \bar{\rho}_k[n]\mathbf{z}_{h,k}[n], \tag{1}$$

with $\mathbf{z}_{h,k}[n] \in \mathbb{C}^K$ being a temporally and spatially white zero-mean circularly symmetric complex Gaussian (ZMCSCG) innovation process with unit variance entries,[1] $\rho_k[n] = (\rho_k[1])^n \triangleq (J_0(2\pi f_{d,k} T_s))^n$ where $J_0(.)$ is the Bessel function of the first kind and zeroth order [32], T_s is the sampling period at the BS, and $f_{d,k} = v_k f_c / c$ is the Doppler frequency corresponding to the kth UE, f_c is the carrier frequency, and c is the speed of light. Also, $\mathbf{z}_{h,k}[n]$ has the same distribution as $\mathbf{h}_k[0]$ and the two are independent of each other.

During uplink training, from $n = -\tau_u + 1$ to $n = 0$, the UEs transmit τ_u mutually orthogonal pilot symbols, with the kth UE's pilot being transmitted with the energy $\mathcal{E}_{p,k}$. These pilots are used by the BS to obtain MMSE estimates $\hat{\mathbf{h}}_k[0] \in \mathbb{C}^{K \times 1}$ of $\mathbf{h}_k[0]$ related to the latter as

$$\mathbf{h}_k[0] = \hat{\mathbf{h}}_k[0] + \tilde{\mathbf{h}}_k[0], \tag{2}$$

with $\tilde{\mathbf{h}}_k[0] \in \mathbb{C}^{K \times 1}$ being the channel estimation error, such that $E[\hat{\mathbf{h}}_k[0]\tilde{\mathbf{h}}_k^H[0]] = \mathbf{O}_N$. Here, $\mathbf{O}_N$ denotes the $N \times N$ all zero matrix. Since the entries of $\mathbf{h}_k[0]$ are i.i.d. ZMCSCG for all k, $\hat{\mathbf{h}}_k[0]$ also has i.i.d. entries. Let $\bar{b}_k^2[0] = E[|\tilde{h}_{ki}[0]|^2]$ denote the mean squared channel estimation error at $n = 0$, and by extension $b_k^2[0] = 1 - \bar{b}_k^2[0] = E[|\hat{h}_{ki}[0]|^2]$. It can be shown that, if the pilot signal from the kth UE is of the form $\delta[n - K + k]$, then $b_k[0] = \rho_k[K - k]\sqrt{\frac{\beta_k \mathcal{E}_{p,k}}{\beta_k \mathcal{E}_{p,k} + N_0}}$, and in general, $b_k[0]$ can be bounded as [15]

$$b_k[0] \geq \rho_k[K-1]\sqrt{\frac{\beta_k \mathcal{E}_{p,k}}{\beta_k \mathcal{E}_{p,k} + N_0}}. \tag{3}$$

We will use the above bound later for deriving the MSE performance of the proposed channel tracking algorithms.

Following pilot transmission, all the UEs simultaneously transmit data symbols during the next $T_u - \tau_u$ instants. The symbol transmitted by the kth UE at the nth instant is denoted by $s_{u,k}[n]$, and is transmitted with an energy $\mathcal{E}_{u,s,k}$. Therefore, the symbol received at the BS during $(1 \leq n \leq T_u - \tau_u)$ is

$$\mathbf{y}_u[n] = \sum_{k=1}^{K} \sqrt{\beta_k \mathcal{E}_{u,s,k}} \mathbf{h}_k[n] s_{u,k}[n] + \sqrt{N_0}\mathbf{w}_u[n] \in \mathbb{C}^N, \tag{4}$$

with $\mathbf{w}_u[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ being the additive white Gaussian noise (AWGN) vector at the BS, and N_0 being the noise power spectral density (PSD). This can be equivalently expressed as

$$\mathbf{y}_u[n] = \mathbf{H}[n](\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}}) \odot \mathbf{s}_u[n] + \sqrt{N_0}\mathbf{w}_u[n], \tag{5}$$

with $\mathbf{H}[n] \triangleq [\mathbf{h}_1[n], \mathbf{h}_2[n], \dots, \mathbf{h}_K[n]]$, $\boldsymbol{\mathcal{E}}_{u,s} \triangleq [\mathcal{E}_{u,s,1}, \mathcal{E}_{u,s,2}, \dots, \mathcal{E}_{u,s,K}]^T$, $\boldsymbol{\beta} \triangleq [\beta_1, \beta_2, \dots, \beta_K]^T$, and $\mathbf{s}_u[n] \triangleq [s_{u,1}[n], s_{u,2}[n], \dots, s_{u,K}[n]]^T$. Here, $\mathbf{A} \odot \mathbf{B}$ represents the Hadamard (element wise) product between two identically sized matrices $\mathbf{A}$ and $\mathbf{B}$.

The BS uses linear combining vectors $\mathbf{v}_k[n]$ $(1 \leq k \leq K)$ generated based on the available channel estimates to recover the symbols transmitted by the kth UE as $\hat{s}_{u,k}[n]$. We denote the probability of symbol error of the kth UE's symbol at the nth instant by $\mathtt{P}_{u,k}[n]$.

Similarly, in the downlink sub-frame, the BS uses the available channel estimates to compute the precoding matrix $\mathbf{P} \in \mathbb{C}^{N \times K}$ for downlink data transmission during the next T_d channel uses. Let $s_{d,k}[n]$ be the data symbol to be transmitted to the kth UE at the nth instant, such that the symbol vector being transmitted to all the UEs can be written as $\mathbf{s}_d[n] = [s_{d,1}[n], \dots, s_{d,K}[n]]^T$. Now, letting $\mathcal{E}_{d,s,k}$ be the downlink energy allocated to the kth UE, we can write the signal transmitted in the downlink as

$$\mathbf{x}[n] = \sum_{k=1}^{K} \sqrt{\mathcal{E}_{d,s,k}} \mathbf{p}_k s_{d,k}[n], \tag{6}$$

such that $E[\|\mathbf{x}[n]\|_2^2] = \mathcal{E}_{d,s}$, with $\mathbf{p}_k$ being the kth column of $\mathbf{P}$. Consequently, the symbol received at the kth UE can be written as

$$y_{d,k}[n] = \sum_{l=1}^{K} \sqrt{\beta_k \mathcal{E}_{d,s,l}} \mathbf{h}_k^T[n] \mathbf{p}_l s_{d,l}[n] + \sqrt{N_0} w_{d,k}[n], \tag{7}$$

with $w_{d,k}[n] \sim \mathcal{CN}(0, 1)$ being the zero mean circularly symmetric complex AWGN. Let $g_{kl}[n] \triangleq \frac{1}{\sqrt{N}} \sqrt{\beta_k \mathcal{E}_{d,s,l}} \mathbf{h}_k^T[n] \mathbf{p}_l$ denote the effective downlink channel for the lth UE's data stream at the kth UE. Then, we can write

$$y_{d,k}[n] = \sqrt{N} g_{kk}[n] s_{d,k}[n] + \sqrt{N} \sum_{\substack{l=1 \\ l \neq k}}^{K} g_{kl}[n] s_{d,l}[n] + \sqrt{N_0} w_{d,k}[n]. \tag{8}$$

We can also write the time evolution of the effective downlink channel to the kth UE as

$$g_{kk}[n] = \rho_k[1] g_{kk}[n-1] + \bar{\rho}_k[1] \zeta_{kk}[n], \tag{9}$$

where $\zeta_{kk}[n] \triangleq \sqrt{\beta_k \mathcal{E}_{d,s,k}/N} \mathbf{z}_{h,k}^T[n] \mathbf{p}_k$ denotes the innovation component in the effective channel to the kth UE. It can be shown that for a wide sense stationary innovation process,

[1] Throughout the paper, we use the notation $\bar{\rho} \triangleq \sqrt{1 - \rho^2}$.

$\zeta_{kk}[n]$ can also be assumed to be wide sense stationary, such that $\text{var}(\zeta_{kk}[n]) = \text{var}(g_{kk}[n]) = \beta_k \mathcal{E}_{d,s,k}$.

The UE employs the *use-and-then-forget* [1] technique to detect the received symbol by replacing the effective channel coefficient with its expected value, and treating interference as noise [1], with the probability of symbol error at the nth instant being denoted by $\mathtt{P}_{d,k}[n]$.

III. The Uplink Case

In this section, we develop and analyze algorithms to track the matrix channel during the uplink data transmission phase. We note that both the RLS and the SGD algorithms were originally proposed for estimating static systems, but are able to track time varying systems due to their adaptive nature. Therefore, in this section, we develop these algorithms under the assumption of static channels, and then study their tracking behavior under time varying channels. Both these algorithms are initialized using the channel estimate at time 0, $\hat{\mathbf{h}}_k[0]$.

A. RLS Based Uplink Channel Tracking

At the outset, if we assume the channel matrix $\mathbf{H} \in \mathbb{C}^{K \times N}$ to be static, using (5), we can write the signal vector received at the BS in the nth instant compactly as

$$\mathbf{y}_u[n] = \mathbf{H}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}}) \odot \mathbf{s}_u[n] + \sqrt{N_0}\mathbf{w}_u[n], \quad (10)$$

Letting $\hat{\mathbf{s}}_u[n]$ be the estimate of the received symbols at the nth instant, and $\tilde{\mathbf{s}}_u[n]$ be the corresponding symbol error vector, such that $\mathbf{s}_u[n] = \hat{\mathbf{s}}_u[n] + \tilde{\mathbf{s}}_u[n]$, we can rewrite (10) as

$$\begin{aligned}\mathbf{y}_u[n] = \mathbf{H}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}}) \odot \hat{\mathbf{s}}_u[n] \\ + \mathbf{H}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}}) \odot \tilde{\mathbf{s}}_u[n] + \sqrt{N_0}\mathbf{w}_u[n]. \quad (11)\end{aligned}$$

This received uplink signal can then be used to update the channel estimate at the BS according to Theorem 1.

Theorem 1. *The estimate of the channel matrix at the nth $(0 \leq n \leq T_u - \tau_u)$ instant is updated as*

$$\hat{\mathbf{H}}[n] = \hat{\mathbf{H}}[n-1] + \tilde{\mathbf{y}}_u[n]\mathbf{g}^H[n] diag(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})^{-1}, \quad (12)$$

with

$$\tilde{\mathbf{y}}_u[n] \triangleq \mathbf{y}_u[n] - \hat{\mathbf{H}}[n-1] diag(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n],$$

being the a-posteriori error,

$$\mathbf{g}^H[n] \triangleq \frac{\mu_r^{-1}\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n-1]}{1 + \mu_r^{-1}\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n-1]\hat{\mathbf{s}}_u[n]} \quad (13)$$

and

$$\boldsymbol{\Phi}^{-1}[n] = \mu_r^{-1}\boldsymbol{\Phi}^{-1}[n-1] - \mu_r^{-1}\mathbf{g}[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n-1]. \quad (14)$$

being the RLS gain, $\boldsymbol{\Phi}[n] \triangleq \sum_{l=1}^{n} \mu_r^{n-l}\hat{\mathbf{s}}_u[l]\hat{\mathbf{s}}_u^H[l]$, and $0 < \mu_r \leq 1$ is the forgetting factor.

Proof. See Appendix-A. ∎

We note that this algorithm requires only the decoded data symbols, unlike the Kalman filtering-based approach [17] which requires the second order channel statistics.

As mentioned earlier, the RLS based tracking algorithm is developed for a static channel $\mathbf{H}$. Therefore, in order to extend the approach to the case of aging channels, and to prescribe optimal values for the forgetting factor μ_r, we next assess its tracking performance under time varying channels in the following theorem.

Theorem 2. *The mean squared channel estimation error for the kth UE's channel at the nth $(0 \leq n \leq T_u - \tau_u)$ instant is given as*

$$\begin{aligned}\bar{b}_k^2[n] = \mu_r^2\bar{b}_k^2[n-1] + \mu_r^2\bar{\rho}_k^2[1] \\ + (1-\mu_r)^2\left(\frac{N_0}{\beta_k \mathcal{E}_{u,s,k}} + 4\mathtt{P}_{u,k}[n]\right). \quad (15)\end{aligned}$$

Proof. See Appendix-B. ∎

Remark: At steady state, if $\bar{b}_k^2[n] \approx \bar{b}_k^2[n-1]$, the above reduces to $\bar{b}_k^2 \approx \frac{\mu_r^2\bar{\rho}_k^2[1]}{(1-\mu_r^2)} + \frac{(1-\mu_r)^2}{(1-\mu_r^2)}(N_0/(\beta_k\mathcal{E}_{u,s,k}) + 4\mathtt{P}_{u,k}[n])$. Thus, the steady-state error is the sum of two terms. The first term represents the error due to the temporal variation in the channel. The second term captures the errors due to channel estimation at finite SNR and error propagation. It is easy to optimize the steady state error with respect to μ_r and obtain a good initialization for the forgetting factor.

It is easy to show that the principle of orthogonality, that is, $E[\mathbf{h}[n]\tilde{\mathbf{y}}_u^H[n]] = E[\hat{\mathbf{h}}[n]\tilde{\mathbf{y}}_u^H[n]]$, does not hold here, resulting in the channel estimate not being independent of the channel estimation error. Hence, the desired signal is no longer independent of the interference due to channel estimation error, and the worst case noise theorem from [33] cannot be used to quantify the achievable rate. Therefore, in order to quantify the performance of this algorithm, we have to rely on simulation based metrics such as the bit error rate. These are discussed in detail in Sec. III-D.

B. SGD based Uplink Channel Tracking

We now present a stochastic gradient descent based technique for tracking the channel state under aging. Similar to the RLS based approach, we first develop the algorithm for a static channel, and then analyze its tracking performance under an aging channel.

Theorem 3. *The SGD based algorithm for updating the estimate of the channel at the BS at the nth $(0 \leq n \leq T_u - \tau_u)$ instant is given by*

$$\hat{\mathbf{H}}[n+1] = \hat{\mathbf{H}}[n] + \mu_a\tilde{\mathbf{y}}_u[n](diag(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n])^H, \quad (16)$$

with μ_a being the adaptation step size, and the estimate of the received signal at the nth instant defined as $\hat{\mathbf{y}}_u[n] = \hat{\mathbf{H}}[n] diag(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n]$, such that $\tilde{\mathbf{y}}_u[n] = \mathbf{y}_u[n] - \hat{\mathbf{y}}_u[n]$.

Proof. See Appendix-C. ∎

In the following Theorem, we evaluate the mean squared estimation error of the SGD based update.

Theorem 4. *The mean squared estimation error $(\bar{b}_{u,k}^2[n+1])$ of the kth channel vector at the $(n+1)$th $(0 \leq n \leq T_u - \tau_u)$instant is given as*

$$\bar{b}_{u,k}^2[n+1] = (1 - \mu_a\beta_k\mathcal{E}_{u,s,k})^2\,\rho_k^2[1]\bar{b}_{u,k}^2[n]$$

$$+\bar{\rho}_k^2[1]+(\mu_a\beta_k\mathcal{E}_{u,s,k})^2(1-\rho_k[1])^2\bar{b}_{u,k}^2[n]$$
$$+\mu_a^2(N_0+4\mathtt{P}_{u,k}[n]\beta_k\mathcal{E}_{u,s,k})\sum_{l=1}^{K}\beta_l\mathcal{E}_{u,s,l}. \quad (17)$$

Proof. See Appendix-D. ■

Remark: It can be shown that, for small μ_a, the steady-state MSE of the SGD-based algorithm is $\bar{b}_{u,k}^2 \approx \frac{\bar{\rho}_k^2[1]}{\bar{\rho}_k^2[1]+2\mu_a\beta_k\mathcal{E}_{u,s,k}\rho_k^2[1]} + \frac{\mu_a^2(N_0+4\mathtt{P}_{u,k}[n]\beta_k\mathcal{E}_{u,s,k})\sum_{l=1}^{K}(\beta_l\mathcal{E}_{u,s,l})}{\bar{\rho}_k^2[1]+2\mu_a\beta_k\mathcal{E}_{u,s,k}\rho_k^2[1]}$. The first term represents the error due to the temporal variation in the channel, and the second term to the error due to finite SNR based channel estimation and error propagation due to symbol detection errors. It can be observed that too small a value of the learning parameter μ_a will result in the first term dominating $\bar{b}_{u,k}^2$, and too large a value of μ_a will lead to the dominance of the second term. Minimizing the steady state error with respect to μ_a is straightforward.

Here again, due to the replacement of the deterministic gradient by the stochastic gradient, the channel estimation error is no longer orthogonal to the channel estimate, and the worst case noise theorem cannot be used to bound the achievable rates. Hence, the performance of the SGD based channel tracker also needs to be quantified via simulations.

C. Symbol Estimation

The channel tracking algorithm described above provides the BS receiver with updated channel estimates at each symbol instant. These channel estimates need to be used to decode the data symbols, which are then used to update the channel estimates for the subsequent received symbol. We now analyze the symbol decoding performance of the above estimator. Let the true channel at the nth instant be expressed in terms of its most recent estimate, obtained at the previous time instant, as

$$\mathbf{H}[n]=\hat{\mathbf{H}}[n-1]\mathrm{diag}(\boldsymbol{\rho}[1])+\tilde{\mathbf{H}}[n-1]\mathrm{diag}(\boldsymbol{\rho}[1])+\mathbf{Z}_h[n]\mathrm{diag}(\bar{\boldsymbol{\rho}}[1]), \quad (18)$$

where $\mathbf{Z}_h[n] \in \mathbb{C}^{N\times K}$ is the channel innovation matrix defined as $\mathbf{Z}_h[n]=[\mathbf{z}_{h,1}[n],\ldots,\mathbf{z}_{h,K}[n]]$. Consequently, the uplink signal received at the nth instant can be expanded as

$$\mathbf{y}_u[n]=\hat{\mathbf{H}}[n-1](\boldsymbol{\rho}[1]\odot\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\odot\mathbf{s}_u[n]+\tilde{\mathbf{H}}[n-1](\boldsymbol{\rho}[1]\odot\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\odot\mathbf{s}_u[n]+\mathbf{Z}_h[n](\bar{\boldsymbol{\rho}}[1]\odot\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\odot\mathbf{s}_u[n]+\sqrt{N_0}\mathbf{w}_u[n]. \quad (19)$$

Assuming that all the UEs transmit from the constellation $\mathcal{S}$, the maximum likelihood estimate of the symbol vector can be obtained at the BS by solving

$$\hat{\mathbf{s}}_u[n]=\arg\min_{\mathbf{s}\in\mathcal{S}^K}\|\mathbf{y}_u[n]-\hat{\mathbf{H}}[n-1](\boldsymbol{\rho}[1]\odot\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\odot\mathbf{s}\|_2. \quad (20)$$

The solution of (20) has an exponential computational complexity in the number of UEs. Hence, we can use a combining matrix $\mathbf{V}[n]$ to post-process the received signal to obtain $\mathbf{r}_u[n]=\mathbf{V}^H[n]\mathbf{y}_u[n]$. Letting $r_{u,k}[n]$ denote the kth element of $\mathbf{r}_u[n]$ and $\mathbf{v}_k[n]$ denote the kth column of $\mathbf{V}[n]$, we get

$$r_{u,k}[n]=\rho_k[1]\beta_k\mathcal{E}_{u,s,k}\mathbf{v}_k^H[n]\hat{\mathbf{h}}_k[n-1]s_{u,k}[n]+\rho_k[1]\beta_k\mathcal{E}_{u,s,k}\mathbf{v}_k^H[n]\tilde{\mathbf{h}}_k[n-1]s_{u,k}[n]+\rho_k[1]\beta_k\mathcal{E}_{u,s,k}\mathbf{v}_k^H[n]\mathbf{z}_{h,k}[n]s_{u,k}[n]+\sum_{\substack{l=1\\l\neq k}}^{K}\beta_l\mathcal{E}_{u,s,l}\mathbf{v}_k^H[n]\mathbf{h}_l[n]s_{u,l}[n]+\sqrt{N_0}\mathbf{v}_k^H[n]\mathbf{w}_u[n]. \quad (21)$$

Here, the first term corresponds to the desired signal, the second to the interference due to the channel estimation error, the third to the interference due to aging, the fourth to the inter stream interference from different UEs, and the last to the additive noise. Now, the symbols from different UEs can be detected individually from (21), with the symbol transmitted by the kth UE estimated as

$$\hat{s}_{u,k}=\arg\min_{s\in\mathcal{S}}|r_{u,k}[n]-\rho_k[1]\beta_k\mathcal{E}_{u,s,k}\mathbf{v}_k^H[n]\hat{\mathbf{h}}_k[n-1]s|. \quad (22)$$

Note that in case the BS employs conjugate beamforming, $\mathbf{V}[n]=\hat{\mathbf{H}}[n-1]$ and if the BS uses generalized MMSE beamforming, then $\mathbf{V}[n]=\hat{\mathbf{H}}[n-1](\hat{\mathbf{H}}^H[n-1]\hat{\mathbf{H}}[n-1]+\xi\mathbf{I}_N)^{-1}$, with ξ being the regularizing parameter. Now, we will declare an error event to have occurred in the data stream of the kth UE if $\hat{s}_{u,k}[n]\neq s_{u,k}[n]$. The probability of such an event is denoted as $\mathtt{P}_{u,k}[n]$. However, due to the adaptive nature of the channel estimation algorithms, and the consequent non-orthogonality of the channel estimates and estimation errors, this cannot be expressed in a closed form, and is evaluated via simulations in the next subsection.

subsectionComputational Complexity

1) RLS Based Tracking: We note that the channel update in (12) and the computation of the a-posteriori error vector require $\mathcal{O}(KN)$ floating point operations each, and the computation of the RLS gain in (13) and rank 1 update of $\boldsymbol{\Phi}^{-1}[n]$ in (14) require $\mathcal{O}(K^2)$ floating point operations. Consequently, the overall computational complexity of RLS based channel tracking is $\mathcal{O}(K(K+N))$.

2) SGD Based Tracking: Here, the computation of the symbol error vector $\tilde{\mathbf{y}}[\mathbf{n}]$, the calculation of the gradient, and the update of the available estimate, all require $\mathcal{O}(KN)$ floating point operations, making the overall computational complexity of SGD based tracking as $\mathcal{O}(KN)$.

3) Optimal Tracking: As described in [17], the computational complexity of this algorithm is $\mathcal{O}((KN)^2+N^3)$.

We note that the computational complexity of MRC-based data detection, as described in Section III-C, is $\mathcal{O}(KN|\mathcal{S}|)$ for a constellation $\mathcal{S}$, whereas the computational complexity for generalized MMSE combining based data detection [17] is $\mathcal{O}(KN^3|\mathcal{S}|)$; these computations need to be performed with pilot-only based channel estimation followed by data detection also. Hence, we see that channel tracking can be included at the BS at little additional computational cost compared to the no-tracking case.

D. Performance Evaluation

In this subsection, we use Monte Carlo simulations to quantify the performance of the proposed channel tracking schemes, and to contrast them against a conventional mMIMO system with no tracking. To simplify things, we consider a single cell, single carrier system with $K=16$ UEs and a 256

TABLE I: Simulation Parameters

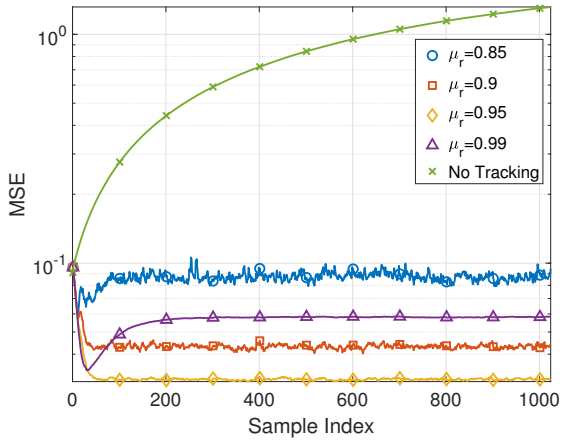
N	The number of antennas on the BS.	256
K	The number of users.	16
f_c	Carrier frequency.	3 GHz
T_s	Sampling Interval	10 μs
T	Frame Duration	1024 Samples
	Pilot and data SNR	10 dB

antenna BS. The system operates at a carrier frequency (f_c) of 3 GHz, such that the signal (single carrier) bandwidth is 100 kHz. Unless specified otherwise, the data and pilot SNRs are assumed to be 10 dB. The BS is assumed to sample at the Nyquist rate of the complex baseband signal, i.e., at 100 kHz, and the frame duration (T) is assumed to be 1024 symbols [15]. Also, we assume the channel to age according to the AR-1 model, i.e., $\rho[n] = \rho^n$, with the correlation coefficient ρ taking values in the range $0.999 \leq \rho \leq 0.99999$ [31]. For the purpose of these simulations, we consider $\rho = 0.99999$, $\rho = 0.9999$ and $\rho = 0.999$, corresponding to UE velocities of 10, 70, and 270 km/h, respectively [31]. We note that this model coincides with the 3GPP SCM channel model [34], extended to a multiuser massive MIMO setup. We summarize these parameters in Table I. Also, for simplicity, we assume the UEs to be equidistant from the BS, and average the performance over $1,000$ independent Monte Carlo runs. We note that, in our previous work [17], we had compared the performance against the state of the art [23], [24] and showed that the technique in [17] offers the best performance. Hence, in this paper, we consider the adaptive MMSE receiver developed in [17] as the baseline for our comparisons.

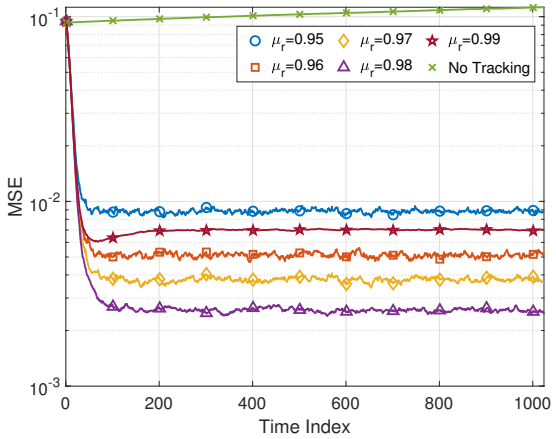
In Figs. 1 (a) and (b), we plot the tracking performance of the RLS based tracking algorithm for different values of the forgetting factor μ_r under different channel aging conditions. We see that using an appropriate forgetting factor improves the MSE for both fast ($\rho = 0.999$) and slow ($\rho = 0.99999$) moving UEs. Also, while $\mu_r = 0.95$ works best for $\rho = 0.999$ and $\mu_r = 0.98$ works best for $\rho = 0.99999$, using an intermediate value like $\mu_r = 0.97$ offers reasonably good performance across all UE velocities.

In Fig 2, we plot the BER with RLS based channel tracking for uncoded BPSK transmission as a function of the data SNR at different sample indices for $\rho = 0.999$. The value of the forgetting factor was fixed at $\mu_r = 0.9$, based on the results in Fig. 1(a). We observe that in the presence of channel tracking, at sufficiently large data SNRs, the availability of better channel estimates due to tracking results in significant improvement in the BER at higher time indices. RLS based tracking can boost the performance at a BER of 10^{-3} by as much as 1 dB. For better visualization, in Fig 3, we plot the BER at different SNRs as a function of time across four data frames. We observe that the BER at lower SNRs continues to worsen due to poor channel quality and error propagation, whereas, at sufficiently high SNRs (close to 3 dB), it continues to improve with the available channel estimates. Note that the slight variation in the BER at lower data SNRs is due to the finite number of samples being used for averaging.

In Fig 4, we plot the theoretically computed and the simulated MSE performances of the RLS based tracking algorithm as a function of the data and pilot SNRs. We observe that the simulated MSEs agree with the theoretically derived results,



(a)



(b)

Fig. 1: Channel MSE with and without RLS based tracking, as a function of the time index for (a) $\rho = 0.999$ (b) $\rho = 0.99999$ and different forgetting factors μ_r.

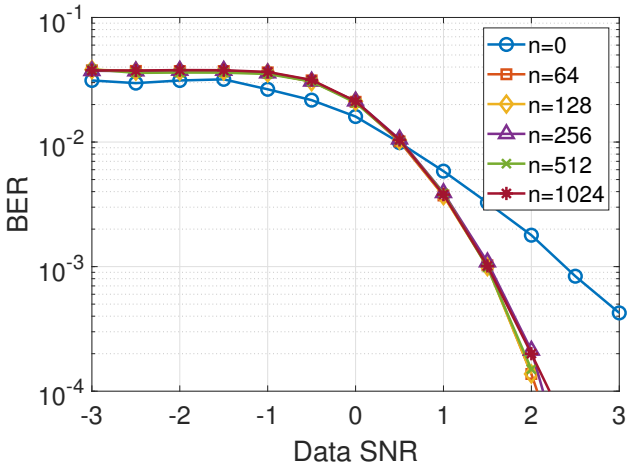


Fig. 2: BER with RLS based tracking, as a function the data SNR for different time indices.

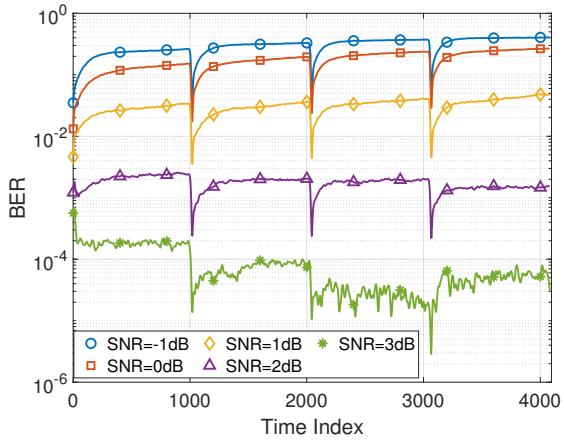


Fig. 3: BER with RLS based tracking, as a function time for different data SNRs.

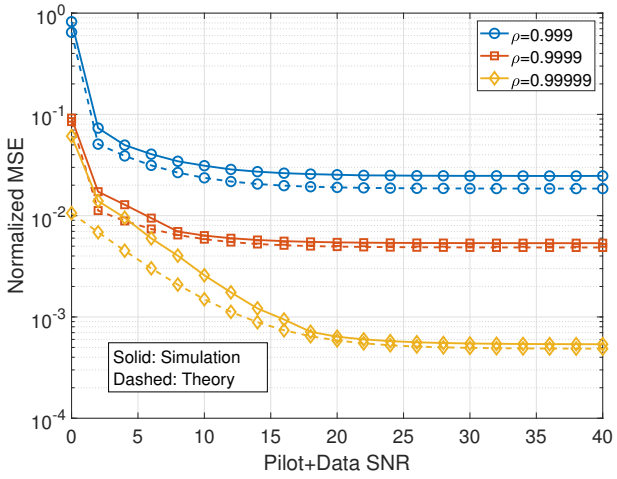
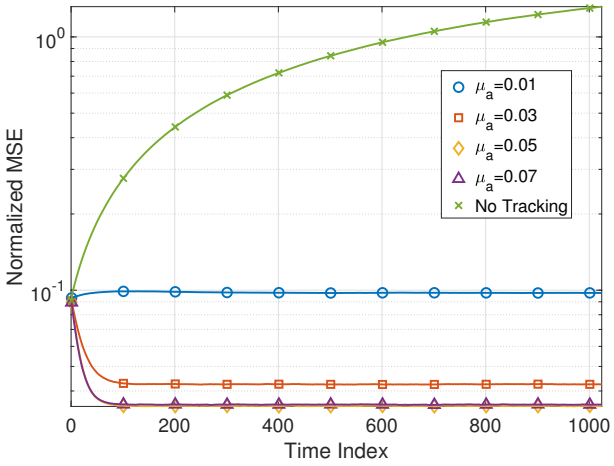
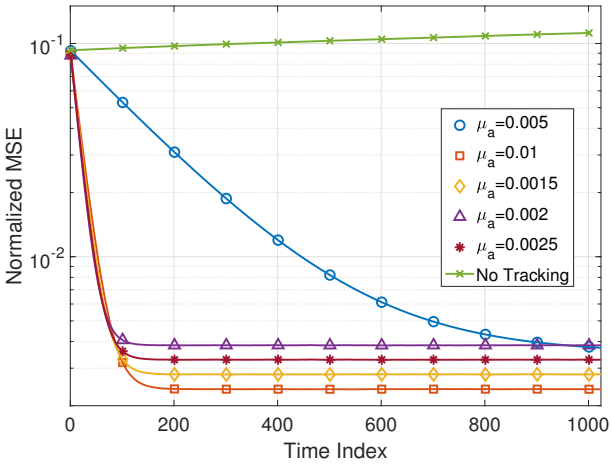


Fig. 4: Normalized MSE with RLS based tracking *vs.* data and pilot SNRs for different rates of channel aging.



(a)



(b)

Fig. 5: Channel MSE with and without SGD based tracking, as a function of the time index for (a) $\rho = 0.999$ (b) $\rho = 0.99999$ and different values of the step size parameter μ_a.

and converge close to $\mu_r^2 \bar{\rho}_k^2$[1], as predicted by Theorem 2.

We now discuss the performance of the SGD based channel tracking algorithm. In Figs. 5 (a) and (b), we plot the MSE as a function of the sample index for different values of the adaptation parameter μ_a, for $\rho = 0.999$ and $\rho = 0.99999$, respectively. Similar to RLS based tracking, a suitable choice of the adaptation parameter can reduce the MSE by more than one order of magnitude compared to the no tracking case. However, in this case, since μ_a affects both the speed of convergence and the residual error after convergence, it should be judiciously chosen.

In Fig. 6, we compare the MSE achieved for the RLS, SGD, and the optimal channel tracking algorithm discussed in [17], for different values of ρ as a function of time. The MSE-optimal channel tracking algorithm from [17], which uses the knowledge of the second-order channel statistics, performs the best, but it is closely followed by the RLS and SGD algorithms, which is as expected. To illustrate the relative computational complexity of the algorithms, we plot the time (in seconds) required for 1000 iterations of the three algorithms as a function of the number of BS antennas in Fig. 7. The MSE-optimal channel tracking algorithm is significantly more computationally intensive than either RLS or SGD based tracking. In fact, the algorithms in this paper are 1-2 orders of magnitude faster than the approach in [17]. We note that, due to the efficient matrix inversion algorithms used by MATLAB, the complexity of RLS based tracking turns out to be smaller than $\mathcal{O}(N^2)$, making it preferable over SGD. Also, as discussed in [17], the computational complexity of optimal tracking is of comparable order as that of MMSE data detection ($\mathcal{O}(N^3)$) at the BS. Therefore, we conclude that RLS and SGD-based channel tracking can boost the performance of a massive MIMO system with or without channel aging with a negligible increase in computational cost compared to the no-tracking case.

As the final experiment in this section, in Fig. 8, we plot the uplink symbol error rate (SER) as a function of the SNR with $K = 8$ users transmitting data to the BS using the 8-PSK case

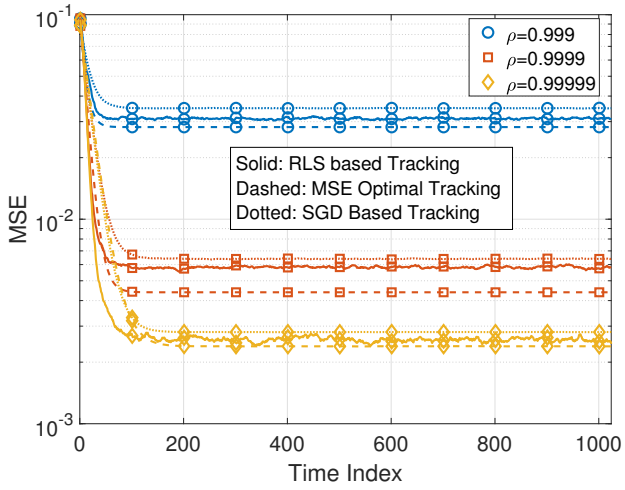


Fig. 6: Channel MSE with optimal, RLS, and SGD based tracking as a function of time for different values of ρ.

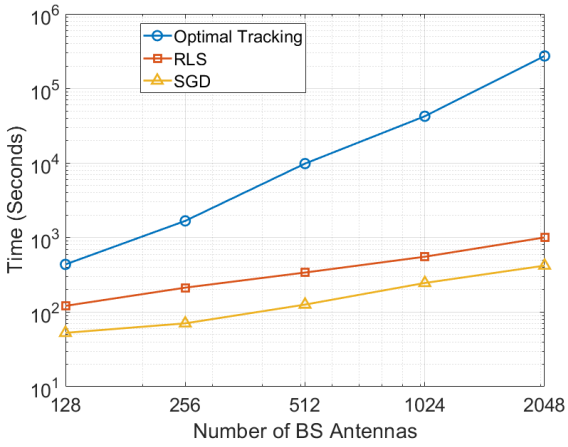


Fig. 7: Run time (seconds) for 100 iterations of the channel tracking algorithms *vs.* the number of BS antennas.

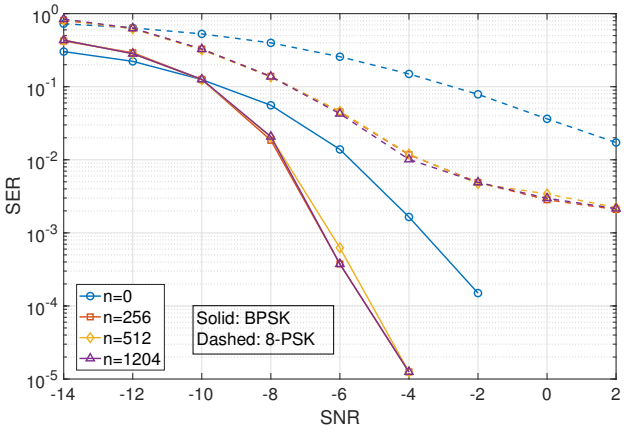


Fig. 8: Uplink SER performance with $K = 8$ users with data transmission from the 8-PSK constellation, as a function of SNR with SGD-based tracking for different time indices, with $\mu_a = 0.05$ and $\rho = 0.999$.

constellation (dashed lines). We also plot the SER obtained with BPSK transmission using solid lines in the same figure. We see that, at low SNR, when the SER is of the order 0.1, there is a gap of about 3 dB between BPSK and 8-PSK. As the SNR increases, the SER with 8-PSK saturates due to the interference caused by the residual channel error, especially in the multi-user setting. We also see that, compared to the first data symbol ($n = 0$), the SER for later symbols $n \geq 256$ is lower; this is because the detected data symbols help to improve the accuracy of the channel estimates. Moreover, the channel estimation error saturates within about 100 symbols (see Fig. 5), so the performances coincide for all $n \geq 256$, as expected. Finally, for BPSK transmission, comparing the SER obtained from the SGD-based algorithm with that from the RLS-based algorithm (Fig. 2), we see that SGD achieves an SER of 10^{-2} at about 7.5 dB lower SNR than RLS-based tracking. This is partly because Fig. 3 is for the $K = 16$ users case while Fig. 8 of this document is for $K = 8$. It is also because, although SGD and RLS have similar channel tracking performance at high SNR (see Fig. 6, which shows the MSE behavior at an SNR of 10 dB), at lower SNR, SGD significantly outperforms RLS in terms of its MSE, and this allows the former to achieve lower SER than the latter.

IV. The Downlink Case

We now present algorithms for channel tracking at the UEs during the downlink data transmission phase. The key difference between channel tracking in the uplink and the downlink is that, in the latter case, one only needs to track the overall effective scalar channel seen by the UEs. As a consequence, the downlink channel tracking algorithms are computationally simpler than the corresponding algorithms for uplink channel tracking. Both the algorithms presented here are initialized with the deterministic equivalent of the effective downlink channel, $\sqrt{\beta_k \mathcal{E}_{d,s,k}}$.

A. RLS Based Tracking

Assuming the effective downlink channel coefficient g_{kk} to be constant across time, we can write the downlink signal received at the kth UE at the nth instant as

$$y_{d,k}[n] = \sqrt{N} g_{kk} s_{d,k}[n] + \sqrt{N} \sum_{\substack{l=1 \\ l \neq k}}^{K} g_{kl} s_{d,l}[n] + \sqrt{N_0} w_k[n]. \tag{23}$$

Also, letting $\hat{s}_{d,k}[n]$ be the estimate of the received symbol at the nth instant, with the corresponding symbol error being $\tilde{s}_{d,k}[n]$, (23) takes the form,

$$y_{d,k}[n] = \sqrt{N} g_{kk} \hat{s}_{d,k}[n] + \sqrt{N} g_{kk} \tilde{s}_{d,k}[n] + \sqrt{N} \sum_{\substack{l=1 \\ l \neq k}}^{K} g_{kl} s_{d,l}[n] + \sqrt{N_0} w_k[n]. \tag{24}$$

Theorem 5. *The estimate of the effective downlink channel at the nth ($T_u \leq n \leq T_u + T_d - \tau_d$) instant is given as*

$$\hat{g}_{kk}[n] = \hat{g}_{kk}[n-1] + c_k[n]\tilde{y}_{d,k}[n], \tag{25}$$

where $\tilde{y}_{d,k}[n] = y_{d,k}[n] - \sqrt{N}\hat{g}_{kk}[n-1]\hat{s}_k[n]$, and

$$c_k[n] \triangleq \frac{1}{\sqrt{N}} \frac{\frac{\hat{s}^*_{d,k}[n]}{\sum_{m=1}^{n-1}|\hat{s}_{d,k}[m]|^2}}{\mu_r + \frac{|\hat{s}_{d,k}[n]|^2}{\sum_{m=1}^{n-1}|\hat{s}_{d,k}[m]|^2}}, \tag{26}$$

is the RLS gain, with μ_r being the forgetting factor.

Proof. See Appendix-E. ∎

We next assess the tracking performance of this algorithm for aging channels. For this, we let $\bar{a}_k^2[n] \triangleq E[|\tilde{g}_{kk}[n]|^2]$.

Theorem 6. *The estimation error in the effective downlink channel at the nth ($T_u \leq n \leq T_u + T_d - \tau_d$) instant can be iteratively computed as*

$$\bar{a}_k^2[n] = \mu_r^2\bar{a}_k^2[n-1] + \mu_r^2\bar{\rho}_k^2[1]\beta_k\mathcal{E}_{d,s,k} + (1-\mu_r)^2\left(\beta_k\sum_{\substack{l=1\\l\neq k}}^{K}\mathcal{E}_{d,s,l} + N_0 + 4\beta_k\mathcal{E}_{d,s,k}\mathtt{P}_{d,k}[n]\right). \tag{27}$$

Proof. See Appendix-F. ∎

Remark: Similar to the uplink case, we can show that the steady state mean squared error takes the from $\bar{a}_k^2 \approx \frac{\mu_r^2\bar{\rho}_k^2[1]}{(1-\mu_r^2)}\beta_k\mathcal{E}_{d,s,k} + \frac{(1-\mu_r)^2}{(1-\mu_r^2)}(N_0 + \beta_k\sum_{l\neq k}\mathcal{E}_{d,s,l} + 4\mathtt{P}_{u,k}[n]\beta_k\mathcal{E}_{u,s,k})$. Again, it is straightforward to optimize this expression to obtain a good initial value for the forgetting factor μ_r. Also, as before, in order to quantify the data transmission performance of this algorithm, we have to rely on simulation based metrics such as the bit error rate.

B. SGD based Tracking

We now discuss a stochastic gradient descent based technique for tracking the effective downlink channel.

Theorem 7. *The stochastic gradient based update for the channel estimate $\hat{g}_{kk}$ at the $(n+1)$th ($T_u \leq n \leq T_u + T_d - \tau_d$) instant is given by*

$$\hat{g}_{kk}[n+1] = \hat{g}_{kk}[n] + \mu_a\tilde{y}_{d,k}[n]\hat{s}^*_{d,k}[n], \tag{28}$$

with μ_a being the adaptation step size, and the estimation error in the received signal at the nth instant being defined as,

$$\tilde{y}_{d,k}[n] = y_{d,k}[n] - \sqrt{N}\hat{g}_{kk}[n]\hat{s}_{d,k}[n]. \tag{29}$$

Proof. See Appendix-G ∎

We now discuss the tracking performance of the above algorithm for a time varying channel.

Theorem 8. *The mean squared estimation error ($\bar{a}_k^2[n+1]$) of the kth UE's desired signal channel at the $(n+1)$th ($T_u \leq n \leq T_u + T_d - \tau_d$) instant is given as,*

$$\bar{a}_k^2[n+1] = (1-\mu_a)^2\rho_k^2[1]\bar{a}_k^2[n] + \bar{\rho}_k^2[1] + \mu_a^2\left(N_0 + \beta_k\sum_{\substack{l=1\\l\neq k}}^{K}\mathcal{E}_{d,s,l} + 4\mathtt{P}_{d,k}[n]\beta_k\mathcal{E}_{d,s,k}\right). \tag{30}$$

Proof. See Appendix-H. ∎

Remark: We can again approximate the steady state mean squared estimation error as the sum of two terms: $\bar{a}_k^2 \approx \frac{\bar{\rho}_k^2[1]}{\mu_a^2+2\mu_a\rho_k^2[1]} + \frac{\mu_a^2}{\mu_a^2+2\mu_a\rho_k^2[1]}\left(N_0 + \beta_k\sum_{\substack{l=1\\l\neq k}}^{K}\mathcal{E}_{d,s,l} + 4\mathtt{P}_{d,k}[n]\beta_k\mathcal{E}_{d,s,k}\right)$, with the first term corresponding to the error due to aging, and the second term corresponding to the error due to the finite data SNR and error propagation.

C. Symbol Estimation

Letting the channel estimate available at the kth UE at the nth instant be $\hat{g}_{kk}[n]$, and the corresponding estimation error be $\tilde{g}_{kk}[n]$, we can write the received signal at the kth UE as

$$y_{d,k}[n] = \sqrt{N}\hat{g}_{kk}[n]s_{d,k}[n] + \sqrt{N}\tilde{g}_{kk}[n]s_{d,k}[n] + \sqrt{N}\sum_{\substack{l=1\\l\neq k}}^{K}g_{kl}[n]s_{d,l}[n] + \sqrt{N_0}w_{d,k}[n]. \tag{31}$$

Now assuming that the symbols transmitted to the kth UE are contained in the constellation $\mathcal{S}$, we can write an estimate $\hat{s}_{d,k}[n]$ of $s_{d,k}[n]$ as

$$\hat{s}_{d,k}[n] = \arg\min_{s\in\mathcal{S}}|y_{d,k}[n] - \sqrt{N}\hat{g}_{kk}[n]s|. \tag{32}$$

This operation entails $\mathcal{O}(|\mathcal{S}|)$ complex floating point operations. Also, we declare the presence of an error in the data detected by the kth UE if $\hat{s}_{d,k}[n] \neq s_{d,k}[n]$. The probability error in the kth UE's detected data is denoted by $\mathtt{P}_{d,k}[n]$. As argued in the uplink case, due to the adaptive nature of the channel estimation algorithms $\mathtt{P}_{d,k}[n]$ cannot be expressed in a closed form, and is evaluated via simulations.

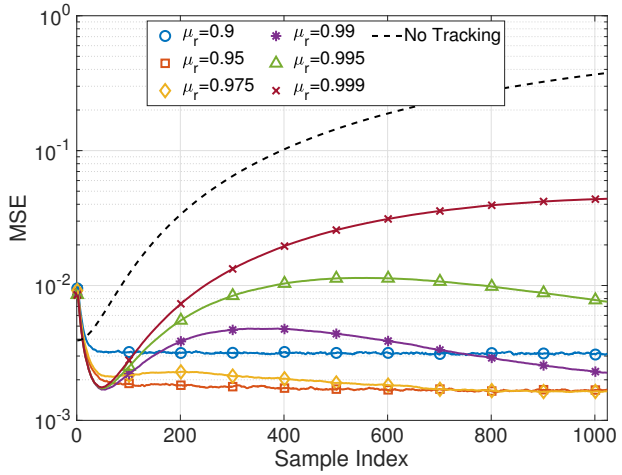
D. Computational Complexity

1) RLS Based Tracking: We note that the channel update step at the kth user requires the the computation of $c_k[n]$ and $\tilde{y}_{d,k}[n]$ followed by the update in (25). These steps entail $\mathcal{O}(1)$ floating point operations, making the overall computational complexity of RLS-based update $\mathcal{O}(1)$ at each user.

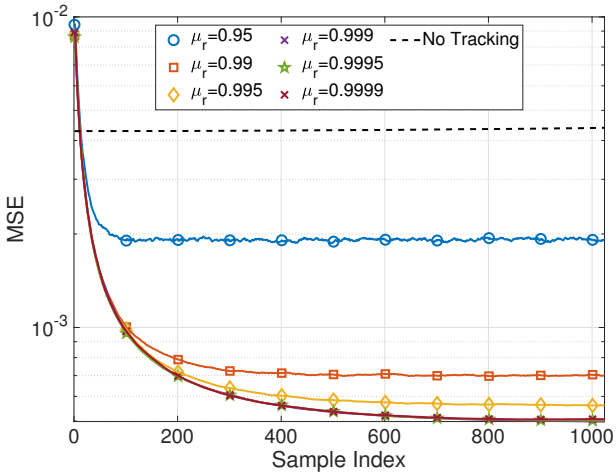
2) SGD Based Tracking: The computation of the symbol error $\tilde{\mathbf{y}}_{d,k}[n]$, and the channel estimate update in (28) both require $\mathcal{O}(1)$ floating point operations, making the overall computational complexity of SGD based tracking also $\mathcal{O}(1)$.

3) Optimal Tracking: As discussed in [17], the computational complexity of optimal tracking is also $\mathcal{O}(1)$.

We thus conclude that downlink channel tracking is possible at the UEs at no increase in the order of computational cost compared to the no-tracking case.



(a)



(b)

Fig. 9: Channel MSE with and without RLS based tracking, as a function of the time index for (a) $\rho = 0.999$ (b) $\rho = 0.99999$ and different forgetting factors μ_r.

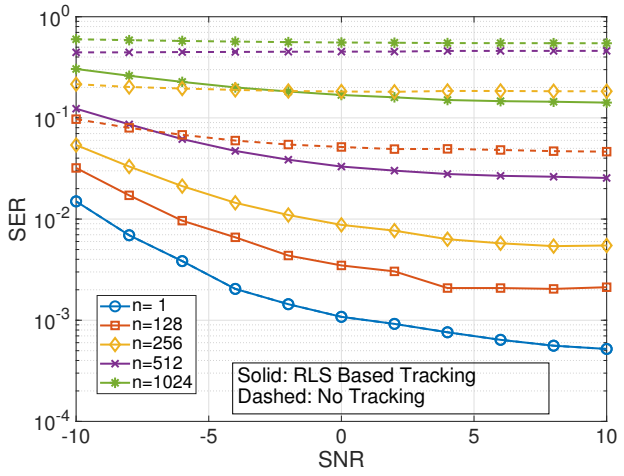
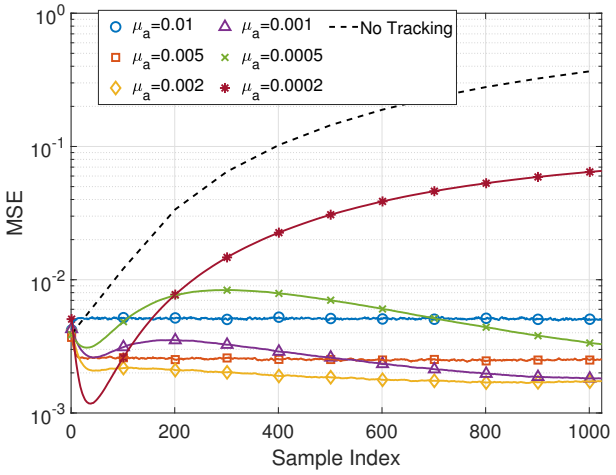
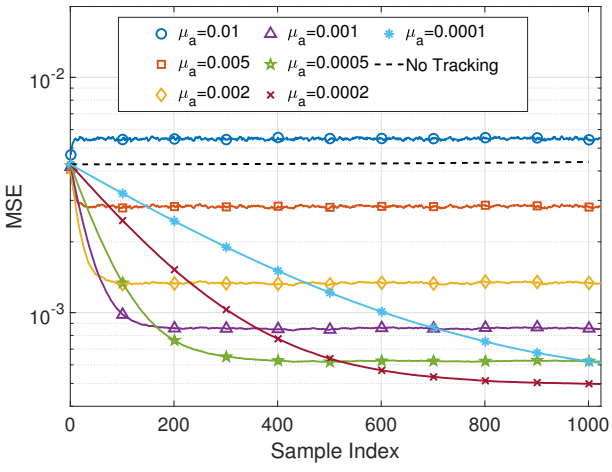


Fig. 10: SER with and without RLS filtering based tracking *vs.* the data SNR, for $\rho = 0.999$ at different time indices.



(a)



(b)

Fig. 11: Channel MSE with and without SGD based tracking, as a function of the time index for (a) $\rho = 0.999$ (b) $\rho = 0.99999$ and different adaptation step sizes μ_a.

E. Performance Evaluation

In Figs. 9(a) and 9(b), we plot the performance of the RLS based tracking algorithm for different values of the forgetting factor μ_r and under different channel aging conditions. We see that using an appropriate choice of the forgetting factor significantly reduces the MSE in both the cases.

In Fig. 10 we plot the symbol error rate performance of the RLS based tracking algorithm for different time indices for a 4-PAM system for $\rho = 0.999$. The forgetting factor μ_r in all these cases is fixed at $\mu_r = 0.95$.

In Figs. 11 (a) and (b), we plot the performance of the SGD based tracking algorithm for different values of the learning rate μ_a under different channel aging conditions. We see that using an appropriate choice of the adaptation step size can lead to a significant improvement in the MSE of the effective downlink channel at the UEs.

In Fig. 12, we plot the SER of SGD based channel tracking at different time indices for 4-PAM data transmission over a

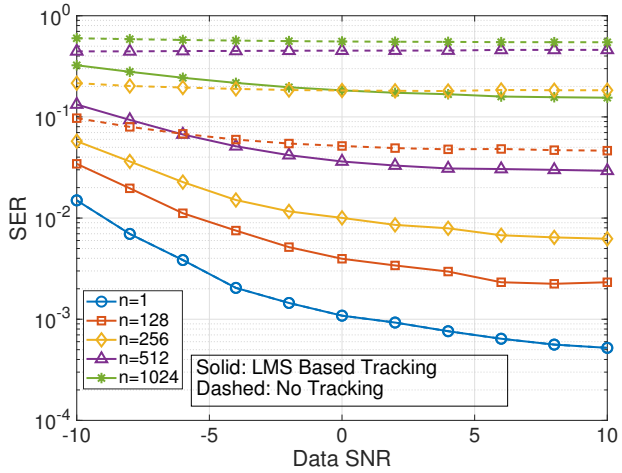


Fig. 12: SER with and without SGD filtering based tracking *vs.* the data SNR, for $\rho = 0.999$ at different time indices.

fast-varying channel (with $\rho = 0.999$.) The adaptation step size is fixed at $\mu_l = 0.002$. We see that channel tracking can reduce the SER by more than one order of magnitude under fast-varying channels.

V. Conclusions

In this work, we developed and evaluated data aided adaptive channel tracking algorithms for mMIMO systems under channel aging for Rayleigh fading time-varying channels whose temporal evolution follows the AR(1) model. Specifically, we theoretically derived the tracking-MSE performance of two approaches, an RLS-based approach and an SGD-based approach. Through extensive simulations, we showed that the low complexity RLS and SGD based channel tracking algorithms perform at par with MMSE-optimal Kalman filter based channel tracking, but with orders of magnitude computational complexity. In fact, using the algorithms developed here, one can obtain the benefits of channel tracking with only a marginal increase in computational complexity compared to the no-tracking case, making the algorithm attractive for implementation. Future work can consider the effects of other imperfections such as synchronization errors and pilot contamination on the tracking performance of the system.

Appendices

A. Proof of Theorem 1

Letting $\mathbf{Y}_{u,n} = [\mathbf{y}_u[1], \ldots, \mathbf{y}_u[n]]$, $\hat{\mathbf{S}}_{u,n} = [\hat{\mathbf{s}}_u[1], \ldots, \hat{\mathbf{s}}_u[n]]$, $\tilde{\mathbf{S}}_{u,n} = [\tilde{\mathbf{s}}_u[1], \ldots, \tilde{\mathbf{s}}_u[n]]$, and $\mathbf{W}_{u,n} = [\mathbf{w}_{u,n}[1], \ldots, \mathbf{w}_{u,n}[n]]$ we can write (11) as

$$\begin{aligned}\mathbf{Y}_n = \mathbf{H}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}}) \odot \hat{\mathbf{S}}_{u,n} \\ + \mathbf{H}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}}) \odot \tilde{\mathbf{S}}_{u,n} + \sqrt{N_0}\mathbf{W}_{u,n}. \quad (33)\end{aligned}$$

Since $\hat{\mathbf{S}}_{u,n}$ is known at the BS, we can use it to obtain the least squares estimate $\hat{\mathbf{H}}[n]$ of $\mathbf{H}$ as $\hat{\mathbf{H}}[n] = \mathbf{Y}_n\hat{\mathbf{S}}_{u,n}^{\dagger}\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})^{-1}$, with $\mathbf{A}^{\dagger}$ representing the Moore-Penrose inverse of the matrix $\mathbf{A}$. We define the empirical autocorrelation matrix of the estimated symbols and empirical cross correlation matrix between the received samples, respectively, as

$$\boldsymbol{\Phi}[n] = \hat{\mathbf{S}}_{u,n}\hat{\mathbf{S}}_{u,n}^H = \boldsymbol{\Phi}[n-1] + \hat{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n], \quad (34)$$

$$\boldsymbol{\Gamma}[n] = \mathbf{Y}_{u,n}\hat{\mathbf{S}}_{u,n}^H = \boldsymbol{\Gamma}[n-1] + \mathbf{y}_u[n]\hat{\mathbf{s}}_u^H[n]. \quad (35)$$

Letting $\hat{\mathbf{y}}_u[n] = \hat{\mathbf{H}}[n-1]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n]$, it is easy to show that

$$\begin{aligned}\hat{\mathbf{H}}[n]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}}) = \boldsymbol{\Gamma}[n]\boldsymbol{\Phi}^{-1}[n] \\ = (\boldsymbol{\Gamma}[n-1] + \mathbf{y}[n]\hat{\mathbf{s}}_u^H[n])\boldsymbol{\Phi}^{-1}[n] = \Bigg(\boldsymbol{\Gamma}[n-1]\Bigg(\boldsymbol{\Phi}^{-1}[n-1] \\ - \frac{\boldsymbol{\Phi}^{-1}[n-1]\hat{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n-1]}{1 + \hat{\mathbf{s}}^H[n]\boldsymbol{\Phi}^{-1}[n-1]\hat{\mathbf{s}}_u[n]}\Bigg) + \mathbf{y}[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]\Bigg) \\ = \hat{\mathbf{H}}[n-1]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}}) \\ + (\mathbf{y}[n] - \hat{\mathbf{y}}[n])\frac{\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n-1]}{1 + \hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n-1]\hat{\mathbf{s}}_u[n]}. \quad (36)\end{aligned}$$

Defining $\mathbf{g}^H[n] \triangleq \frac{\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n-1]}{1+\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n-1]\hat{\mathbf{s}}_u[n]}$ as the RLS gain, the estimate of the channel matrix at the nth instant is given by (12) with $\tilde{\mathbf{y}}_u[n] = \mathbf{y}_u[n] - \hat{\mathbf{y}}_u[n]$. Introducing a forgetting factor $0 \leq \mu_r \leq 1$ to provide higher weightage to more recent samples, can modify (34) and (35) as $\boldsymbol{\Phi}[n] = \mu_r\boldsymbol{\Phi}[n-1] + \hat{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]$, $\boldsymbol{\Gamma}[n] = \mu_r\boldsymbol{\Gamma}[n-1] + \hat{\mathbf{y}}_u[n]\hat{\mathbf{s}}_u^H[n]$. Then, (12) can be used to update the channel estimates, with the gain $\mathbf{g}[n]$ taking the value given by (13).

B. Proof of Theorem 2

Defining the channel estimation error at the nth instant as $\tilde{\mathbf{H}}[n] = \mathbf{H}[n] - \hat{\mathbf{H}}[n]$, and assuming that the channel ages according to (1), we can write

$$\begin{aligned}\tilde{\mathbf{H}}[n] = \mathbf{H}[n-1]\text{diag}(\boldsymbol{\rho}[1]) + \mathbf{Z}_h[n]\text{diag}(\bar{\boldsymbol{\rho}}[1]) - \hat{\mathbf{H}}[n-1] \\ - (\tilde{\mathbf{y}}[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n])\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})^{-1}. \quad (37)\end{aligned}$$

Also, as discussed in [31], for velocities of up to 250 km/hr, and in the sub-6 GHz range for the carrier frequency, $0.999 \leq \rho[1] \leq 1$. Therefore, we can approximate $\rho[1] \approx 1$ [29] in the first term above, and consequently,

$$\begin{aligned}\tilde{\mathbf{H}}[n] = \tilde{\mathbf{H}}[n-1] + \mathbf{Z}_h[n]\text{diag}(\bar{\boldsymbol{\rho}}[1]) \\ - (\tilde{\mathbf{y}}[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n])\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})^{-1}. \quad (38)\end{aligned}$$

But

$$\begin{aligned}\tilde{\mathbf{y}}[n] =& \mathbf{H}[n]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n] \\ & - \hat{\mathbf{H}}[n-1]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n] \\ & + \mathbf{H}[n]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\tilde{\mathbf{s}}_u[n] + \sqrt{N_0}\mathbf{w}_u[n] \\ =& \tilde{\mathbf{H}}[n-1]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n] \\ & + \mathbf{Z}_h[n]\text{diag}(\bar{\boldsymbol{\rho}} \odot \sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}[n] \\ & + \mathbf{H}[n]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\tilde{\mathbf{s}}_u[n] + \sqrt{N_0}\mathbf{w}_u[n], \quad (39)\end{aligned}$$

and thus,

$$\tilde{\mathbf{H}}[n] = \tilde{\mathbf{H}}[n-1] - \tilde{\mathbf{H}}[n-1]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})$$

$$\times \hat{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}$$
$$+\mathbf{Z}_h[n]\text{diag}(\bar{\boldsymbol{\rho}}[1]) - \mathbf{Z}_h[n]\text{diag}(\bar{\boldsymbol{\rho}}\odot\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n]$$
$$\times \hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}$$
$$+\mathbf{H}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})$$
$$\times \tilde{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}$$
$$+\sqrt{N_0}\mathbf{w}_u[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}. \quad (40)$$

Defining $\tilde{\mathbf{h}}[n] = \text{vec}(\tilde{\mathbf{H}}[n])$ and $\mathbf{z}_h[n] = \text{vec}(\mathbf{Z}_h[n])$, we get

$$\tilde{\mathbf{h}}[n] = \tilde{\mathbf{h}}[n-1] - (\text{diag}\left(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}}\right)\hat{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]\times$$
$$\boldsymbol{\Phi}^{-1}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}\otimes\mathbf{I}_N)\tilde{\mathbf{h}}[n-1]$$
$$+(\text{diag}(\bar{\boldsymbol{\rho}}[1])\otimes\mathbf{I}_N)\mathbf{z}_h[n] - (\text{diag}(\bar{\boldsymbol{\rho}}\odot\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]$$
$$\times\boldsymbol{\Phi}^{-1}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}\otimes\mathbf{I}_N)\mathbf{z}_h[n]$$
$$+(\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\tilde{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}$$
$$\otimes\mathbf{I}_N)\mathbf{h}[n]+\sqrt{N_0}\text{vec}(\mathbf{w}_u[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}) \quad (41)$$

This can be rewritten as

$$\tilde{\mathbf{h}}[n] = ((\mathbf{I}_K - \text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]$$
$$\times\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1})\otimes\mathbf{I}_N)(\tilde{\mathbf{h}}[n-1]+(\text{diag}(\bar{\boldsymbol{\rho}}[1])\otimes\mathbf{I}_N)\mathbf{z}_h[n])$$
$$+(\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\tilde{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]$$
$$\times\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}\otimes\mathbf{I}_N)\mathbf{h}[n]$$
$$+\sqrt{N_0}\text{vec}(\mathbf{w}_u[n]\hat{\mathbf{s}}_u^H[n]\boldsymbol{\Phi}^{-1}[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}). \quad (42)$$

Now, since $E[\mathbf{s}_u[n]\mathbf{s}_u^H[n]] = \mathbf{I}_K$, for large n, $\boldsymbol{\Phi}[n] \approx \frac{1}{(1-\mu_r)}\mathbf{I}_K$ and consequently, $\boldsymbol{\Phi}^{-1}[n] \approx (1-\mu_r)\mathbf{I}_K$. Thus

$$\tilde{\mathbf{h}}[n] = ((\mathbf{I}_K - (1-\mu_r)\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n]$$
$$\times\hat{\mathbf{s}}_u^H[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}))\otimes\mathbf{I}_N)$$
$$\times(\tilde{\mathbf{h}}[n-1]+(\text{diag}(\bar{\boldsymbol{\rho}}[1])\otimes\mathbf{I}_N)\mathbf{z}_h[n])$$
$$+(1-\mu_r)(\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\tilde{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]$$
$$\times\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}\otimes\mathbf{I}_N)\mathbf{h}[n]$$
$$+\sqrt{N_0}(1-\mu_r)\text{vec}(\mathbf{w}_u[n]\hat{\mathbf{s}}_u^H[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}). \quad (43)$$

Also, since $\mu_r \approx 1$, and $1-\mu_r \ll 1$, we can use the direct averaging method [29] to approximate the above as

$$\tilde{\mathbf{h}}[n] = \mu_r(\tilde{\mathbf{h}}[n-1]+(\text{diag}(\bar{\boldsymbol{\rho}}[1])\otimes\mathbf{I}_N)\mathbf{z}_h[n])$$
$$+(1-\mu_r)(\text{diag}(\tilde{\mathbf{s}}_u[n]\hat{\mathbf{s}}_u^H[n]\otimes\mathbf{I}_N)\mathbf{h}[n]$$
$$+\sqrt{N_0}(1-\mu_r)\text{vec}(\mathbf{w}_u[n]\hat{\mathbf{s}}_u^H[n]\text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})^{-1}). \quad (44)$$

We assume that the symbol errors exist in the kth UE's data stream with a probability $\mathtt{P}_{u,k}[n]$, i.e., $E[|\tilde{s}_{u,k}[n]|^2] \leq 4\mathtt{P}_{u,k}[n]$ [35], such that $E[\tilde{\mathbf{s}}_u[n]\tilde{\mathbf{s}}_u^H[n]] = \underline{\mathtt{P}_u}[n]$, with $\underline{\mathtt{P}_u}[n] = \text{diag}(\mathtt{P}_{u,1},\ldots,\mathtt{P}_{u,K})$. We also assume that the symbol estimates are independent of the corresponding estimation errors. Also, letting $\bar{\mathbf{B}}_k[n] = E[\tilde{\mathbf{h}}_k[n]\tilde{\mathbf{h}}_k^H[n]]$, and $\bar{\mathbf{B}}[n] = E[\tilde{\mathbf{h}}[n]\tilde{\mathbf{h}}^H[n]]$, we can write

$$\bar{\mathbf{B}}^2[n] = \mu_r^2\bar{\mathbf{B}}^{\mathbf{2}}[n-1] + \mu_r^2\text{diag}(\bar{\boldsymbol{\rho}}^2[1])\otimes\mathbf{I}_N$$
$$+(1-\mu_r)^2(N_0\text{diag}(\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s})^{-1} + 4\text{diag}(\underline{\mathtt{P}_u}[n]))\otimes\mathbf{I}_N. \quad (45)$$

We note that all the terms in the above equation are diagonal matrices with the first being a function of $\bar{\mathbf{B}}^2[n-1]$. We also note from the system model that $\bar{\mathbf{B}}^2[0]$ is diagonal, and consequently $\bar{\mathbf{B}}^2[n]$ is diagonal for all n. Also defining $\bar{\mathbf{b}}[n] = [\bar{b}_1[n],\bar{b}_2[n],\ldots,\bar{b}_K[n]]^T$, we can write, $\bar{\mathbf{B}}[n] = \text{diag}(\bar{\mathbf{b}}[n])\otimes\mathbf{I}_N$, and consequently,

$$\bar{\mathbf{b}}^2[n] = \mu_r^2\bar{\mathbf{b}}^2[n-1] + \mu_r^2(\bar{\boldsymbol{\rho}}^2[1])$$
$$+(1-\mu_r)^2(N_0(\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s})^{-1} + 4(\underline{\mathtt{P}_u}[n])). \quad (46)$$

Considering the kth entry of (46) completes the proof.

C. Proof of Theorem 3

Defining $\hat{\mathbf{u}}[n] = \text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{s}}_u[n]$ and $\tilde{\mathbf{u}}[n] = \text{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\tilde{\mathbf{s}}_u[n]$ for ease of notation, and using (11), with the knowledge of $\hat{\mathbf{u}}[n]$ and $\mathbf{y}_u[n]$, $\hat{\mathbf{H}}_o$, the optimal MMSE estimate of $\mathbf{H}$, satisfies

$$\hat{\mathbf{H}}_o = \arg\min_{\mathbf{H}} E\left[\|\mathbf{y}_u[n] - \mathbf{H}\hat{\mathbf{u}}[n]\|_2^2\right] = \arg\min_{\mathbf{H}} J(\mathbf{H}), \quad (47)$$

where $J(\mathbf{H})$, defined implicitly, is the cost function. Letting $\sigma_y^2 = E[\|\mathbf{y}_u[n]\|_2^2]$, $\mathbf{R}_{\hat{u}\hat{u}} = E[\hat{\mathbf{u}}[n]\hat{\mathbf{u}}^H[n]]$, and $\mathbf{R}_{y\hat{u}} = E[\mathbf{y}_u[n]\hat{\mathbf{u}}^H[n]]$, we can write,

$$J(\mathbf{H}) = \sigma_y^2 - \text{Tr}\{\mathbf{H}\mathbf{R}_{y\hat{u}}^H\} - \text{Tr}\{\mathbf{R}_{y\hat{u}}\mathbf{H}^H\} + \text{Tr}\{\mathbf{R}_{\hat{u}\hat{u}}\mathbf{H}^H\mathbf{H}\}. \quad (48)$$

We can now write the gradient of $J(\mathbf{H})$ w.r.t. $\mathbf{H}$ as,

$$\nabla_{\mathbf{H}}J = \mathbf{H}\mathbf{R}_{\hat{u}\hat{u}} - \mathbf{R}_{y_u\hat{u}} = -E[(\mathbf{y}_u[n] - \mathbf{H}\hat{\mathbf{u}}[n])\hat{\mathbf{u}}^H[n]]$$
$$= -E[\tilde{\mathbf{y}}_u[n]\hat{\mathbf{u}}^H[n]]. \quad (49)$$

Therefore, using the method of steepest descent, and replacing the statistical average with its instantaneous value, we can iteratively update the channel estimate at the BS as (16).

D. Proof of Theorem 4

We can write $\hat{\mathbf{H}}[n+1]$ as

$$\hat{\mathbf{H}}[n+1] = \hat{\mathbf{H}}[n] + \mu_a\tilde{\mathbf{y}}_u[n]\hat{\mathbf{u}}^H[n]$$
$$= \hat{\mathbf{H}}[n] + \mu_a\mathbf{H}[n]\hat{\mathbf{u}}[n]\hat{\mathbf{u}}^H[n] - \mu_a\hat{\mathbf{H}}[n]\hat{\mathbf{u}}[n]\hat{\mathbf{u}}^H[n]$$
$$+\mu_a\mathbf{H}[n]\tilde{\mathbf{u}}[n]\hat{\mathbf{u}}^H[n] + \mu_a\mathbf{w}_u[n]\hat{\mathbf{u}}^H[n]. \quad (50)$$

Considering the above, we can write the estimation error at the nth instant as

$$\tilde{\mathbf{H}}[n+1] = \tilde{\mathbf{H}}[n]\left(\mathbf{I}_K - \mu_a\hat{\mathbf{u}}[n]\hat{\mathbf{u}}^H[n]\right)\text{diag}(\boldsymbol{\rho}[1])$$
$$+\mathbf{Z}_h[n]\text{diag}(\bar{\boldsymbol{\rho}}[1]) - \mu_a\tilde{\mathbf{H}}[n](\hat{\mathbf{u}}[n]\hat{\mathbf{u}}^H[n])(\mathbf{I}_K - \text{diag}(\boldsymbol{\rho}[1]))$$
$$-(\mathbf{H}[n]\tilde{\mathbf{u}}[n] + \mathbf{w}_u[n])\hat{\mathbf{u}}^H[n]. \quad (51)$$

Letting $\tilde{\mathbf{h}}[n] = \text{vec}(\tilde{\mathbf{H}}[n])$, $\mathbf{h}[n] = \text{vec}(\mathbf{H}[n])$, $\mathbf{z}_h[n] = \text{vec}(\mathbf{Z}_h[n])$, and using the small step size approximation [29] to replace $\mu_a\hat{\mathbf{u}}[n]\hat{\mathbf{u}}^H[n]$ with its mean, we get

$$\tilde{\mathbf{h}}[n+1] = (\mathbf{I}_{NK} - \mu_a(\mathbf{R}_{\hat{u}\hat{u}}\otimes\mathbf{I}_N))\,(\text{diag}(\boldsymbol{\rho}[1])\otimes\mathbf{I}_N)\tilde{\mathbf{h}}[n]$$
$$+(\text{diag}(\bar{\boldsymbol{\rho}})\otimes\mathbf{I}_N)\mathbf{z}_h[n] - \mu_a(\hat{\mathbf{u}}^*[n]\tilde{\mathbf{u}}^T[n]\otimes\mathbf{I}_N)\mathbf{h}[n]$$

$-\mu_a(\mathbf{R}_{\hat{u}\hat{u}}(\mathbf{I}_K-\text{diag}(\boldsymbol{\rho}[1]))\otimes\mathbf{I}_N)\tilde{\mathbf{h}}[n]-\mu_a(\hat{\mathbf{u}}^*[n]\otimes\mathbf{I}_N)\mathbf{w}_u[n]$.

Now, since the symbols transmitted by the different UEs are uncorrelated, $\mathbf{R}_{\hat{u}\hat{u}} = \text{diag}(\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s})$, and therefore, the estimation error of the (ik)th channel coefficient evolves as

$$\begin{aligned}\tilde{h}_{ik}[n+1] &= (1-\mu_a\beta_k\mathcal{E}_{u,s,k})\,\rho_k[1]\tilde{h}_{ik}[n]\\ &+\bar{\rho}_k[1]z_{h,ik}[n]-\mu_a\beta_k\mathcal{E}_{u,s,k}(1-\rho_k[1])\tilde{h}_{ik}[n]+\omega_{ik}[n].\end{aligned} \tag{52}$$

with $\omega_{ik}[n]$ being the (ik)th entry of $-\mu_a(\mathbf{H}[n]\tilde{\mathbf{u}}[n]+\mathbf{w}_u[n])\hat{\mathbf{u}}^H[n]$, which has zero mean and a variance $\mu_a^2(N_0+4\mathtt{P}_{u,k}[n]\beta_k\mathcal{E}_{u,s,k})\sum_{l=1}^K\beta_k\mathcal{E}_{u,s,k}$. Note that the variance does not depend on i, the antenna index at the base station, as the channel statistics are identical across BS antennas and only differ from UE to UE. This completes the proof of (17).

E. Proof of Theorem 5

Defining $\mathbf{y}_{d,k,n} = [y_{d,k}[1],\ldots,y_{d,k}[n]]^T$, $\hat{\mathbf{s}}_{d,k,n} = [\hat{s}_{d,k}[1],\ldots,\hat{s}_{d,k}[n]]^T$, $\tilde{\mathbf{s}}_{d,k,n} = [\tilde{s}_{d,k}[1],\ldots,\tilde{s}_{d,k}[n]]^T$, and $\mathbf{w}_{k,n} = [w_k[1],\ldots,w_k[n]]^T$ we can write

$$\mathbf{y}_{d,k,n} = \sqrt{N}(g_{kk}\hat{\mathbf{s}}_{d,k,n}+g_{kk}\tilde{\mathbf{s}}_{d,k,n}+\sum_{\substack{l=1\\l\neq k}}^K g_{kl}\mathbf{s}_{d,l,n})+\sqrt{N_0}\mathbf{w}_{k,n}. \tag{53}$$

Since $\hat{\mathbf{s}}_{d,k,n}$ is available at the kth UE, we can use it to obtain the least squares estimate $\hat{g}_{kk}[n]$ of $g_{kk}[n]$ as $\hat{g}_{kk}[n] = \frac{1}{\sqrt{N}}\frac{\hat{\mathbf{s}}_{d,k,n}^H\mathbf{y}_{d,k,n}}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}$, with $\|\hat{\mathbf{s}}_{d,k,n}\|_2$ denoting the ℓ_2 norm of $\hat{\mathbf{s}}_{d,k,n}$ such that $\|\hat{\mathbf{s}}_{d,k,n}\|_2^2 = \|\hat{\mathbf{s}}_{d,k,n-1}\|_2^2+|\hat{s}_{d,k}[n]|^2$. We note that $\frac{1}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2} = \frac{1}{\|\hat{\mathbf{s}}_{d,k,n-1}\|_2^2}\left(\frac{1}{1+\frac{|\hat{s}_{d,k}[n]|^2}{\|\hat{\mathbf{s}}_{d,k,n-1}\|_2^2}}\right)$, and $\hat{\mathbf{s}}_{d,k,n}^H\mathbf{y}_{d,k,n} = \hat{\mathbf{s}}_{d,k,n-1}^H\mathbf{y}_{d,k,n-1}+y_{d,k}[n]\hat{s}_{d,k}^*[n]$. Letting $\hat{y}_{d,k}[n] = \hat{g}_{kk}[n-1]\hat{s}_{d,k}[n]$, it is easy to show that

$$\begin{aligned}\hat{g}_{kk}[n] &= \frac{1}{\sqrt{N}}\frac{(\hat{\mathbf{s}}_{d,k,n-1}^H\mathbf{y}_{d,k,n-1}+y_{d,k}[n]\hat{s}_{d,k}^*[n])}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}\\ &= \frac{1}{\sqrt{N}}\frac{\hat{\mathbf{s}}_{d,k,n-1}^H\mathbf{y}_{d,k,n-1}}{\|\mathbf{s}_{d,k,n-1}\|_2^2}\left(1-\frac{\hat{s}_{d,k}[n]\hat{s}_{d,k}^*[n]}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}\right)\\ &\quad+\frac{1}{\sqrt{N}}y_{d,k}[n]\hat{s}_{d,k}^*[n]\left(\frac{1}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}\right)\\ &= \hat{g}_{kk}[n-1]+\frac{1}{\sqrt{N}}(y_{d,k}[n]-\hat{y}_{d,k}[n])\hat{s}_{d,k}^*[n]\left(\frac{1}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}\right).\end{aligned} \tag{54}$$

Defining $c_k[n] \triangleq \frac{1}{\sqrt{N}}\frac{\frac{\hat{s}_{d,k}^*[n]}{\|\mathbf{s}_{d,k,n-1}\|_2^2}}{1+\frac{|\hat{s}_{d,k}[n]|^2}{\|\hat{\mathbf{s}}_{d,k,n-1}\|_2^2}}$ as the RLS gain, we can write the estimate of the effective downlink channel at the nth instant as (25).

Introducing a forgetting factor μ_r ($0\leq\mu_r\leq 1$), to allow higher weightage to the recent observations, we obtain $\|\hat{\mathbf{s}}_{d,k,n}\|_2^2 = \mu_r\|\hat{\mathbf{s}}_{d,k,n-1}\|_2^2+|\hat{s}_{d,k}[n]|^2$, $\frac{1}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2} = \frac{1}{\|\hat{\mathbf{s}}_{d,k,n-1}\|_2^2}\left(\frac{1}{\mu_r+\frac{|\hat{s}_{d,k}[n]|^2}{\|\hat{\mathbf{s}}_{d,k,n-1}\|_2^2}}\right)$, and $\hat{\mathbf{s}}_{d,k,n}^H\mathbf{y}_{d,k,n} = \mu_r\hat{\mathbf{s}}_{d,k,n-1}^H\mathbf{y}_{d,k,n-1}+y_{d,k}[n]\hat{s}_{d,k}^*[n]$. Thus, (25) can be used to obtain updated estimates of the channel, with the gain $c_k[n]$ taking the value given by (26).

F. Proof of Theorem 6

Defining the channel estimation error at the nth instant as $\tilde{g}_{kk}[n] = g_{kk}[n]-\hat{g}_{kk}[n]$, and assuming the channel to age according to (9) and approximating $\rho[1]\approx 1$ [29], we have

$$\begin{aligned}\tilde{g}_{kk}[n] &= \tilde{g}_{kk}[n-1]+\bar{\rho}_k[1]\zeta_{kk}[n]\\ &\quad-\frac{1}{\sqrt{N}}\left(\tilde{y}_{d,k}[n]\frac{\hat{s}_{d,k}^*[n]}{\|\mathbf{s}_{d,k,n}\|_2^2}\right).\end{aligned} \tag{55}$$

But

$$\begin{aligned}\tilde{y}_{d,k}[n] &= \sqrt{N}\tilde{g}_{kk}[n]\hat{s}_{d,k}[n]+\sqrt{N}g_{kk}\tilde{s}_{d,k}[n]\\ &\quad+\sqrt{N}\sum_{\substack{l=1\\l\neq k}}^K g_{kl}[n]s_{d,l}[n]+\sqrt{N_0}w_{d,k}[n]\\ &= \sqrt{N}\tilde{g}_{kk}[n-1]\hat{s}_{d,k}[n]+\sqrt{N}\bar{\rho}_k[1]\zeta_{kk}[n]\hat{s}_{d,k}[n]\\ &+\sqrt{N}g_{kk}[n]\tilde{s}_{d,k}[n]+\sqrt{N}\sum_{\substack{l=1\\l\neq k}}^K g_{kl}[n]s_{d,l}[n]+\sqrt{N_0}w_{d,k}[n],\end{aligned}$$

and consequently,

$$\begin{aligned}\tilde{g}_{kk}[n] &= \tilde{g}_{kk}[n-1]-\tilde{g}_{kk}[n-1]\frac{|\hat{s}_{d,k}[n]|^2}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}\\ &\quad+\bar{\rho}_k[1]\zeta_{kk}[n]-\bar{\rho}_k[1]\zeta_{kk}[n]\frac{|\hat{s}_{d,k}[n]|^2}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}\\ &\quad+g_{kk}[n]\frac{\tilde{s}_{d,k}[n]\hat{s}_{d,k}^*[n]}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}+\sum_{\substack{l=1\\l\neq k}}^K g_{kl}[n]\frac{s_{d,l}[n]\hat{s}_{d,k}^*[n]}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}\\ &\quad+\sqrt{N_0}\frac{w[n]\hat{s}_{d,k}^*[n]}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}.\end{aligned} \tag{56}$$

Since $E[|s_{d,k}[n]|^2] = 1$, for large n, we can approximate $\frac{1}{\|\hat{\mathbf{s}}_{d,k,n}\|_2^2}\approx 1-\mu_r$. Now,

$$\begin{aligned}\tilde{g}_{kk}[n] &= (1-(1-\mu_r))(\tilde{g}_{kk}[n-1]+\bar{\rho}_k[1]\zeta_{kk}[n])|\hat{s}_{d,k}[n]|^2\\ &+(1-\mu_r)\sum_{\substack{l=1\\l\neq k}}^K g_{kl}[n]s_{d,l}[n]\hat{s}_{d,k}^*[n]+(1-\mu_r)\hat{s}_{d,k}^*[n]\tilde{s}_{d,k}[n]g_{kk}[n]\\ &\quad+(1-\mu_r)\sqrt{N_0}w_{d,k}[n]\hat{s}_{d,k}^*[n].\end{aligned} \tag{57}$$

Also, since $\mu_r\approx 1$, and $1-\mu_r\ll 1$, we can use the direct averaging method [29] to approximate the above as

$$\begin{aligned}\tilde{g}_{kk}[n] &= \mu_r(\tilde{g}_{kk}[n-1]+\bar{\rho}[1]\zeta_{kk}[n])\\ &+(1-\mu_r)\sum_{\substack{l=1\\l\neq k}}^K g_{kl}[n]s_{d,l}[n]\hat{s}_{d,k}^*[n]+(1-\mu_r)\hat{s}_{d,k}^*[n]\tilde{s}_{d,k}[n]g_{kk}[n]\\ &\quad+(1-\mu_r)\sqrt{N_0}w_{d,k}[n]s_{d,k}^*[n].\end{aligned} \tag{58}$$

We note that all the terms being summed on the RHS of (58) are independent, therefore, the variance of their sum is the same as the sum of their

variances. We also know that $E[|\tilde{g}_{kk}[n]|^2] = \bar{a}_k^2[n]$, $E[|\zeta_{kk}[n]|^2] = \beta_k \mathcal{E}_{d,s,k}$, $E[|g_{kl}[n]s_{d,l}[n]\hat{s}^*_{d,k}[n]|^2] = E[|g_{kl}[n]|^2]E[|s_{d,l}[n]|^2]E[|\hat{s}^*_{d,k}[n]|^2] = \beta_k \mathcal{E}_{d,s,l}$, and $E[|\hat{s}^*_{d,k}[n]\tilde{s}_{d,k}[n]g_{kk}[n]|^2] \leq 4\mathrm{P}_{e,k}\beta_k\mathcal{E}_{d,s,k}$. Substituting these into the variance of $\tilde{g}_{kk}[n]$ completes the proof.

G. Proof of Theorem 7

Using (24), with the knowledge of $\hat{s}_{d,k}[n]$ and $y_{d,k}[n]$, we can write $\hat{g}_{kk,o}$, the MMSE estimate of g_{kk}, as

$$\hat{g}_{kk,o} = \arg\min_{g} E\left[|y_{d,k}[n] - \sqrt{N}g\hat{s}_{d,k}[n]|_2^2\right] = \arg\min_{g} J(g). \tag{59}$$

Letting $\sigma_y^2 = E[|y_{d,k}[n]|_2^2]$, $\sigma_{\hat{s}}^2 = E[|\hat{s}_{d,k}|^2]$, and $\sigma_{y\hat{s}} = E[y_{d,k}[n]\hat{s}^*_{d,k}[n]]$, we can write,

$$J(g) = \sigma_y^2 - 2\Re\{\sqrt{N}g\sigma^*_{y\hat{s}}\} + N|g|^2\sigma_{\hat{s}}^2. \tag{60}$$

We can now write the gradient of $J(g)$ at g^* as,

$$\begin{aligned}\nabla_{g^*}J &= \sqrt{N}g\sigma_{\hat{s}}^2 - N\sigma_{y\hat{s}} \\ &= -\sqrt{N}E[(y_{d,k}[n] - \sqrt{N}g\hat{s}_{d,k}[n])\hat{s}^*_{d,k}[n]] \\ &= -E[\tilde{y}_{d,k}[n]\hat{s}^*_{d,k}[n]],\end{aligned} \tag{61}$$

where the estimate of the received signal at the nth instant is defined as $\hat{y}_{d,k}[n] = \sqrt{N}\hat{g}_{kk}[n]\hat{s}_{d,k}[n]$, and $\tilde{y}_{d,k}[n] = y_{d,k}[n] - \hat{y}_{d,k}[n]$. Approximating the expectation term in (61) with its instantaneous stochastic value, and introducing an adaptation step size μ_a, we can iteratively update the channel estimate as (28).

H. Proof of Theorem 8

We can write $\hat{g}_{kk}[n+1]$ as,

$$\begin{aligned}\hat{g}_{kk}[n+1] &= \hat{g}_{kk}[n] + \mu_a\tilde{y}_{d,k}[n]\hat{s}^*_{d,k}[n] \\ &= \hat{g}_{kk}[n] + \mu_a g_{kk}[n]|\hat{s}_{d,k}[n]|^2 - \mu_a\hat{g}_{kk}[n]|\hat{s}_{d,k}[n]|^2 \\ &\quad + \mu_a\hat{g}_{kk}[n]\tilde{s}_{d,k}[n]\hat{s}^*_{d,k}[n] \\ &\quad + \mu_a\sum_{\substack{l=1\\l\neq k}} g_{kl}[n]s_{d,l}[n]\hat{s}^*_{d,k}[n] + \mu_a w_{d,k}[n]\hat{s}^*_{d,k}[n].\end{aligned} \tag{62}$$

We can now write the channel estimation error at the nth instant as

$$\begin{aligned}\tilde{g}_{kk}[n+1] &= \rho_k[1]\tilde{g}_{kk}[n]\left(1 - \mu_a|\hat{s}_{d,k}[n]|^2\right) + \bar{\rho}_k[1]\zeta_{kk}[n] \\ &\quad - \mu_a\left(\sqrt{N}\hat{g}_{kk}[n]\tilde{s}_{d,k}[n] + \sqrt{N}\sum_{\substack{l=1\\l\neq k}} g_{kl}[n]s_{d,l}[n] + w_{d,k}[n]\right)\hat{s}^*_{d,k}[n].\end{aligned}$$

Using the small step size approximation [29] to replace $\mu_a|\hat{s}[n]|^2$ with its expected value, we can write

$$\begin{aligned}\tilde{g}_{kk}[n+1] &= \rho_k[1]\left(1 - \mu_a\right)\tilde{g}_{kk}[n] + \bar{\rho}_k[1]\zeta_{kk}[n] \\ &\quad - \sqrt{N}\mu_a(\hat{s}^*_{d,k}[n]\tilde{s}_{d,k}[n])g_{kk}[n] - \sqrt{N}\mu_a\sum_{\substack{l=1\\l\neq k}} g_{kl}[n]s_{d,l}[n]\hat{s}^*_{d,k}[n] \\ &\quad - \mu_a w_{d,k}[n]\hat{s}^*_{d,k}[n].\end{aligned} \tag{63}$$

Letting $\omega_k = \mu_a\sqrt{N}(\hat{s}^*_{d,k}[n]\tilde{s}_{d,k}[n])g_{kk}[n] - \mu_a\sqrt{N}\sum_{\substack{l=1\\l\neq k}} g_{kl}[n]s_{d,l}[n]\hat{s}^*_{d,k}[n] - \mu_a w_k[n]\hat{s}^*_{d,k}[n]$, it can be shown to have a zero mean and a variance $\mu_a^2(N_0 + \beta_k\sum_{\substack{l=1\\l\neq k}}\mathcal{E}_{d,s,l} + 4\mathrm{P}_{d,k}[n]\beta_k\mathcal{E}_{d,s,k})$. Substituting this into the mean squared value of (63) results in (30).

REFERENCES

[1] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, Cambridge, UK, 2016.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[3] T. Narasimhan and A. Chockalingam, "Channel hardening-exploiting message passing (CHEMP) receiver in large-scale MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 847–860, Oct. 2014.

[4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[6] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. New York, NY, USA: Cambridge University Press, 2014.

[7] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[8] J. Vieira, F. Rusek, O. Edfors, S. Malkowsky, L. Liu, and F. Tufvesson, "Reciprocity calibration for massive MIMO: Proposal, modeling, and validation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3042–3056, May 2017.

[9] X. Jiang, A. Decurninge, K. Gopala, F. Kaltenberger, M. Guillaud, D. Slock, and L. Deneire, "A framework for over-the-air reciprocity calibration for TDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 5975–5990, Sep. 2018.

[10] R. Chopra, C. R. Murthy, H. A. Suraweera, and E. G. Larsson, "Blind channel estimation for downlink massive MIMO systems with imperfect channel reciprocity," *IEEE Trans. Signal Process.*, vol. 68, pp. 3132–3145, 2020. [Online]. Available: https://doi.org/10.1109/TSP.2020.2988570

[11] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 338–351, Aug. 2013.

[12] A. K. Papazafeiropoulos, "Impact of general channel aging conditions on the downlink performance of massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1428–1444, Feb. 2016.

[13] A. K. Papazafeiropoulos and T. Ratnarajah, "Deterministic equivalent performance analysis of time-varying massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5795–5809, Oct. 2015.

[14] C. Kong, C. Zhong, A. K. Papazafeiropoulos, M. Matthaiou, and Z. Zhang, "Sum-rate and power scaling of massive MIMO systems with channel aging," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4879–4893, Dec. 2015.

[15] R. Chopra, C. Murthy, H. Suraweera, and E. Larsson, "Performance analysis of FDD massive MIMO systems under channel aging," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1094–1108, Feb. 2018.

[16] R. Chopra, C. R. Murthy, H. A. Suraweera, and E. G. Larsson, "Analysis of nonorthogonal training in massive MIMO under channel aging with SIC receivers," *IEEE Signal Process. Lett.*, vol. 26, no. 2, pp. 282–286, Feb. 2019.

[17] R. Chopra and C. R. Murthy, "Data aided MSE-optimal time varying channel tracking in massive MIMO systems," *IEEE Trans. Signal Process.*, vol. pp, p. pp, Jun. 2021.

[18] G. Amarasuriya and H. V. Poor, "Impact of channel aging in multi-way relay networks with massive MIMO," in *Proc. IEEE Intl. Conf. Commun. (ICC 2015)*, London, UK, Jun. 2015, pp. 1951–1957.

[19] Q. Bao, H. Wang, Y. Chen, and C. Liu, "Downlink sum-rate and energy efficiency of massive MIMO systems with channel aging," in *Proc. 8th Int. Conf. on Wireless Commun. and Signal Process. (WCSP)*, Yangzhou, China, Oct. 2016, pp. 1–5.

[20] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[21] H. Rydén, "Massive MIMO in LTE with MRT precoder: Channel aging and throughput analysis in a single-cell deployment," Master's thesis, Linköping University, Sweden, 2014.

[22] A. Anand and C. R. Murthy, "Impact of subcarrier allocation and user mobility on the uplink performance of multiuser massive mimo-ofdm systems," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5285–5299, 2022.

[23] S. Kashyap, C. Mollén, E. Björnson, and E. G. Larsson, "Performance analysis of TDD massive MIMO with Kalman channel prediction," in *Proc. Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2017)*, New Orleans, LA, Mar. 2017, pp. 3554–3558.

[24] V. Arya and K. Appaiah, "Kalman filter based tracking for channel aging in massive MIMO systems," in *2018 International Conference on Signal Processing and Communications (SPCOM) (SPCOM 2018)*, Indian Institute of Science, Bangalore, India, Jul. 2018.

[25] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive mimo channel prediction: Kalman filtering vs. machine learning," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 518–528, 2021.

[26] C. Wu, X. Yi, Y. Zhu, W. Wang, L. You, and X. Gao, "Channel prediction in high-mobility massive mimo: From spatio-temporal autoregression to deep learning," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 1915–1930, 2021.

[27] H. Jiang, M. Cui, D. W. K. Ng, and L. Dai, "Accurate channel prediction based on transformer: Making mobility negligible," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2717–2732, 2022.

[28] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. Cambridge University Press, Cambridge, UK, 2011.

[29] S. Haykin, *Adaptive Filter Theory*. Pearson Education, 2002.

[30] W. C. Jakes and D. C. Cox, Eds., *Microwave Mobile Communications*, 2nd ed. IEEE Press, New York: IEEE Press, 1994.

[31] R. Chopra, C. R. Murthy, and H. A. Suraweera, "On the throughput of large MIMO beamforming systems with channel aging," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1523–1527, Nov. 2016.

[32] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York: Dover, 1964.

[33] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[34] I. Xirouchakis. (1999) Spatial channel model for MIMO simulations. a ray based simulator based on 3GPP TR 25.996 v.6.1.0 [online].

[35] R. Chopra, C. R. Murthy, and R. Annavajjala, "Multistream distributed cophasing," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 1042–1057, Feb. 2017.

Ribhu Chopra (S'11–M'17) received the B.E. degree in Electronics and Communication Engineering from Panjab University, Chandigarh, India in 2009, and the M. Tech. and Ph. D. Degrees in Electronics and Communication Engineering from the Indian Institute of Technology Roorkee, India in 2011 and 2016 respectively. He worked as a project associate at Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore from Aug. 2015, till May 2016. From May 2016 to March 2017 he worked as an institute research associate at the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. In April 2017, he joined the department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Assam, India. His research interests include statistical and adaptive signal processing, massive MIMO communications, and cognitive communications.

Chandra R. Murthy (S'03–M'06–SM'11–F'23) received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology Madras, India, in 1998, the M. S. and Ph. D. degrees in Electrical and Computer Engineering from Purdue University and the University of California, San Diego, USA, in 2000 and 2006, respectively. From 2000 to 2002, he worked as an engineer for Qualcomm Inc., where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. From Aug. 2006 to Aug. 2007, he worked as a staff engineer at Beceem Communications Inc. on advanced receiver architectures for the 802.16e Mobile WiMAX standard. In Sept. 2007, he joined the Department of Electrical Communication Engineering at the Indian Institute of Science, Bangalore, India, where he is currently working as a Professor.

His research interests are in the areas of energy harvesting communications, multiuser MIMO systems, and sparse signal recovery techniques applied to wireless communications. Papers coauthored by him have received Student/Best Paper Awards at the NCC 2014, IEEE ICASSP 2018, IEEE ISIT 2021, IEEE SPAWC 2022, and NCC 2023. He has served on the editorial board of the IEEE SIGNAL PROCESSING LETTERS, SADHANA JOURNAL, and the IEEE TRANSACTIONS ON COMMUNICATIONS. Currently, he is an elected member of the IEEE SAM Technical Committee and is a senior area editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE TRANSACTIONS ON INFORMATION THEORY.

Kumar Appaiah (S'08–M'13–SM'21) received the B.Tech. and M.Tech. degrees in electrical engineering from IIT Madras, Chennai, India in 2008, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin, TX, USA, in 2013. From 2013 to 2014, he was a Senior Engineer with Qualcomm Flarion Technologies, Bridgewater, NJ, USA. Since 2014, he has been working in the Department of Electrical Engineering at IIT Bombay, Mumbai, India, where he is currently an Associate Professor. His research interests include signal processing for optical communication, and multiplexing in wireless and fiber-optic communication systems.