Data Aided MSE-Optimal Time Varying Channel Tracking in Massive MIMO Systems

Ribhu Chopra, *Member, IEEE*, and Chandra R. Murthy, *Senior Member, IEEE*

***Abstract*—The temporal variations in the wireless propagation channel, referred to as *channel aging*, cause a mismatch between the estimated channel and the channel state at the time of data detection. This mismatch has been shown to severely impair the performance of massive MIMO systems. In this paper, we present data aided MSE-optimal channel tracking algorithms to decode the received symbols and update the channel estimates at the base station (BS) and the UEs. These algorithms combine ideas from Kalman filtering for channel tracking and deterministic equivalent analysis for symbol estimation. In the uplink case, we first develop a minimum mean squared error (MMSE) data estimator and the MSE-optimal channel predictor based on the Kalman filtering algorithm. We analytically show that the updated channel estimate obtained from the estimates of the data symbols leads to significantly larger signal to interference-plus-noise ratio (SINR), and hence achievable rate, compared to that obtained from the channel estimate based on pilot symbols. Following this, in the downlink case, we develop an algorithm to track the effective channel at the UEs and analyze its MSE, SINR and achievable rate performance. We show that tracking the effective downlink channel mitigates the effects of channel aging and leads to improved performance. However, since the beamforming matrices at the BS are not updated, downlink channel tracking is not as effective as uplink channel tracking. Finally, via Monte Carlo simulations, we validate our derived results, and demonstrate the gains achievable by tracking the time-varying channels in massive MIMO systems.**

***Index Terms*—Massive MIMO, channel aging, Kalman filtering**

I. Introduction

A. Motivation

Cellular communication systems with a large number of base station (BS) antennas serving a large number of user equipments (UEs), referred to as massive multiple input multiple output (MIMO) systems, have seen much research interest in recent years [1]. The large number of BS antennas results in quasi-orthogonality among the channels to the different users [2]–[5]. Also, the effective channels to the UEs reduce to their deterministic equivalents (DEs) [6]. Further, if accurate channel state information (CSI) is available, simple linear precoding and combining techniques at the BS can lead to large array gains and minimal inter user interference [2], [7]. However, in practice, the accuracy of the CSI available at the BS is impaired due to various factors such as additive noise in channel estimation [4], pilot contamination [8], reciprocity calibration imperfections [9], [10] and channel aging [11]. The latter is the main focus of this work.

Channel aging is caused by the time varying nature of the channels between the BS and the mobile UEs, and manifests as a slow drift of the actual propagation channel from the estimate available at the BS [11]. This evolutionary nature of an aging channel is not amenable to the conventional block fading model of communication channels. In the block fading model, the channel coefficients are assumed to remain unchanged for one coherence interval, and take independent values from the same distribution during the next coherence interval [12]. It has been argued that the notion of coherence time should be replaced with that of a usable time/outage time in case of aging channels [12], [13]. The mismatch between the actual and the estimated channels due to aging results in the SINR (and consequently, the achievable rate) at the UEs becoming a decreasing function of time [11], [14]–[16]. Due to this, depending on the choice of the underlying communication protocol, fast channel aging may limit the maximum system dimensions [12], [17], [18].

Now, although the effective channel quality degrades as the lag between the training and data transmission increases, the initial performance (i.e., at low lag values) is still close to that of an ideal massive MIMO system without aging. This can be exploited to accurately estimate the transmitted data symbols with low complexity. Then, the estimated data symbols, along with the received signal, can be fed into a Kalman filter to update the channel estimates, and thereby track the evolution of the channel coefficients. In this paper, we derive and analyze the performance of data aided channel tracking algorithms in both uplink and downlink of a massive MIMO system.

B. Related Work

Classically, time varying channels have been modeled using either the auto-regressive (AR) model [11], [19] or the basis expansion model (BEM) [20], [21]. The BEM model entails tracking channel variations by estimating parameters of a parsimonious representation of the channel. This approach has been widely studied in the context of the estimating doubly selective OFDM channels [22], [23], and has also recently been applied to the modeling and tracking of sparse millimeter wave (mmWave) massive MIMO channels [24], [25]. While BEM remains an interesting direction for modeling and tracking time varying channels in massive MIMO systems, starting from [11], most of the literature discussing channel aging in massive MIMO systems is based on the AR model. Therefore, in this paper, we focus on tracking time-varying multi-user massive MIMO channels under the AR model.

R. Chopra is with the Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Assam, India, 781039. (email: ribhufec@iitg.ac.in). C. R. Murthy is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, Karnataka, India, 560012. (email: cmurthy@iisc.ac.in).

The canonical massive MIMO system [2] assumes a time division duplexed (TDD) operation due to the prohibitive cost of downlink training. Here, training is performed in the uplink direction, and the downlink channels are estimated by exploiting channel reciprocity [17], [26]–[30]. Consequently, a majority of the existing work on channel aging [11], [14]–[16], [31], [32] has also focussed on TDD systems. These studies exclusively consider linear signal processing techniques such as the maximal ratio combining (MRC) [11] and the minimum mean squared error (MMSE) combining [14] at the receiver, and the matched filter precoding (MFP) [11] and the regularized zero forcing (RZF) [15] precoding at the transmitter.

The most commonly used metric for quantifying the system performance and the loss due to channel aging is the achievable rate of the system. In this respect, both the sum rate [14]–[16] and the average per user rate [17], [18] of a massive MIMO system with an aging channel have been examined. Since an aging channel results in a time varying SINR, and consequently an achievable rate that varies as a function of the sample index throughout a frame, both the average [17] and the worst case [12] rates have been used as metrics for the effects of channel aging. The achievable rates for these systems are generally derived using the DE analysis [14], [17], or the use and forget bounds [18]. The authors in [15] have also discussed the non-asymptotic achievable rates for massive MIMO systems under channel aging, incorporating the effect of phase noise. A key takeaway from these papers is that the expedited aging of channels in high mobility scenarios leads to severe performance loss [33].

The early works on channel aging [11], [14]–[16] ignored the effect of aging during the training and channel estimation phase. This leads to simple yet inaccurate expressions of the channel estimation errors, and consequently of the achievable rates. Kalman filter based estimators for aging channels were presented in [12], [34], [35], where the channel is tracked one frame to the next using pilot symbols received at the start of each frame. The notion of *outage time*, defined as the time for which the channel can be used for data transmission after CSI acquisition, was introduced in [13].

All the works discussed above focus exclusively on pilot based training, and do not explore the potential of using data symbols for data aided channel estimation and tracking. In this work, we address this gap by developing data aided channel tracking algorithms for massive MIMO systems.

C. Contributions

In this paper, we develop MSE-optimal data aided channel tracking schemes for a single cell massive MIMO system under aging channels. We first estimate the data symbols using a low complexity procedure based on results from random matrix theory [6], [11]. Then, we use the recovered data symbols to update the channel estimates using the Kalman filtering principle (See Theorems 1 and 2 in the uplink case, and Theorems 5 and 6 in the downlink case.) We perform these two steps successively upon reception of each data symbol, allowing us to track the channel evolution with time (See Algorithms 1 and 2 for the uplink and downlink cases, respectively). In addition, using tools from random matrix theory, we derive expressions for the MSE (See Theorems 3 and 4 in the uplink case, and Theorems 7 and 8 in the downlink case), SINR, and achievable rate performance (See Sections IV and VI) of the time-varying channel tracking algorithm. In particular, we make use of the expression for the MSE to obtain the Kalman filtering gain employed in the tracking algorithm.

We consider the tracking of both uplink and downlink channels. In the uplink case, pilot symbols transmitted by the UEs are used to obtain the initial channel estimate. We employ a data aided Kalman filter to track the channel using the estimated uplink data symbols (See Algorithm 1). Using results from random matrix theory, we analyze the performance of the algorithm in terms of the MSE in the channel estimate and the achievable rate of the system. In the downlink case, the BS fixes its precoding vectors based on the latest channel estimate obtained from the uplink phase. Therefore, the UEs only need to track the scalar effective downlink channel, i.e., the inner product between the precoding vector and the current channel state (See Section V). We develop an algorithm for tracking the effective downlink channels at the UEs using the downlink data symbols (See Algorithm 2). In particular, we do not require dedicated downlink training. We analyze the MSE and achievable rate performance of the data aided Kalman channel tracker. For comparison purposes, we also analyze the performance of the system without channel tracking.

Our analysis accounts for three sources of errors: due to noise in the pilot symbols, due to the temporal variation of the channel, and due to incorrect symbol recovery affecting the estimation of future channel states. In particular, we elucidate the inter-dependence between the symbol errors and channel tracking errors. For example, in the uplink case, Theorem 3 derives the MSE in the channel estimate, and shows how it depends on the MSE in the data symbols. The MSE in the data symbols, derived in Theorem 4, in turn depends on MSE in the previous channel estimate, which relates it back to Theorem 3. Accounting for these effects requires careful bookkeeping, and is a new aspect that is developed in this paper.

Our key finding is that Kalman filter based tracking helps to counter the effects of channel aging, and significantly increases the usable time of the time-varying channel in massive MIMO systems. We illustrate this via detailed Monte Carlo simulations.

Organization: In Section II, we discuss the system model considered in this work. In Section III, we develop a Kalman filter based data aided channel tracking algorithm for estimating the channel matrix from all the users in the uplink, and derive the MMSE estimates of the received symbols. In Section IV, we present a deterministic equivalent analysis of the achievable rate, both with and without Kalman filter based tracking. Next, in the downlink case, develop an algorithm for tracking the effective scalar channel at each user in Section V. We present a DE analysis of achievable rate by the blind downlink tracking algorithm in Section VI. In Section VII, we validate our results via Monte Carlo simulations and illustrate the effect of system parameters such as the user velocities and the frame duration on the performance. Finally, we present our concluding remarks in Section VIII.

TABLE I: A list of notation followed in the paper.

N	The number of antennas on the BS.
K	The number of users.
$\rho_k[1]$	The temporal correlation coefficient of the channel.
$\mathbf{h}_k[n]$	The channel vector between the BS and the kth user at the nth instant.
$\mathbf{z}_{h,k}[n]$	The innovation component for the channel vector between the BS and the kth user at the nth instant.
$\mathbf{h}_k[n\|\mathcal{Y}_{n-1}]$	The estimate of the channel vector between the BS and the kth user at the nth instant based on the information available till the $(n-1)$th instant.
$b^2_{u,k}[n,l]$	The mean squared channel estimation error for the channel at the nth instant, based on the information till the $(n-1)$th instant.
λ	The downlink power control coefficient
$s_{u,k}[n]$	The uplink symbol transmitted by the kth user
$s_{d,k}[n]$	The downlink symbol transmitted to the kth user
$\mathbf{G}[n]$	The Kalman filtering gain for the uplink channel tracker
$\mathbf{R}_{\tilde{y}\tilde{y}}[n]$	The covariance matrix of the innovation component in the uplink.
$\bar{a}^2_{u,k}[n]$	The MSE of in the estimation of the uplink symbol sent by the kth user.
$g_{kl}[n]$	The effective downlink channel to the kth user for the lth user's data stream.
$\sigma^2_{\tilde{y}\tilde{y},k}[n]$	The variance of the innovation component in the downlink.
$\Gamma_k[n]$	The Kalman filtering gain for the kth downlink user at the nth instant

Notation: Boldface lowercase and uppercase letters represent vectors and matrices, respectively. The kth column of the matrix $\mathbf{A}_u$ is denoted by $\mathbf{a}_{u,k}$. $(\cdot)^H$ represents the Hermitian operation on a vector or a matrix, and $(\cdot)^\dagger$ represents its Moore-Penrose pseudoinverse. $\mathbf{I}_K, \mathbf{0}_K$ and $\mathbf{O}_K$ represent the $K \times K$ identity matrix, the $K \times 1$ all-zero vector and the $K \times K$ all-zero matrix, respectively. The ℓ_2 norm of a vector and the Frobenius norm of a matrix are denoted by $\|\cdot\|_2$ and $\|\cdot\|_F$, respectively. $\mathcal{CN}(\mu, \sigma^2)$ represents a circularly symmetric complex Gaussian random variable with mean μ and variance σ^2. $E[\cdot]$ and var$(\cdot)$ represent the mean and variance of a random variable. In general, $\hat{x}$ and $\tilde{x}$ denote the estimate and the corresponding estimation error of a random variable x. Also, throughout this paper, *for any* $-1 \leq x \leq 1$, $\bar{x}$ *is used to denote* $\bar{x} = \sqrt{1-|x|^2}$. The other notations used in this paper are listed in Table I.

II. System Model

We consider a single cell massive MIMO system operating in the time division duplex (TDD) mode with an N antenna base station serving K single antenna users. We let β_k and v_k respectively denote the path loss coefficient and the velocity of the kth user [17]. Letting $h_{ik}[n]$ denote the fast fading component of the channel between the kth UE and the ith BS antenna at the nth instant, the effective channel between the ith BS antenna and the kth UE is given as $\sqrt{\beta_k}h_{ik}[n]$. Also, we let $\mathbf{h}_k[n] = [h_{1k}[n], \dots h_{Nk}[n]]^T$ denote the vector channel to the kth user.

We consider the channel to evolve over time as

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

where $\mathbf{z}_{h,k}[n]$ is the temporally and spatially white zero mean circularly symmetric complex Gaussian (ZMCSCG) innovation processes and $\rho_k[n]$ is the channel correlation coefficient defined as $\rho_k[n] = E[h_{ki}[\tau]h^*_{ki}[\tau - n]]$. A commonly used model for the evolution of the channel is a first order autoregressive process, with $\rho_k[n] = (\rho_k[1])^n \triangleq (J_0(2\pi f_{d,k}T_s))^n$ [11], [12], [35], where $J_0(.)$ is the Bessel function of the first kind and zeroth order [36], T_s is the sampling period at the BS, and $f_{d,k}$ is the Doppler frequency corresponding to the kth user. It is defined as $f_{d,k} = v_k f_c/c$, where f_c is the carrier frequency, and c is the speed of light.

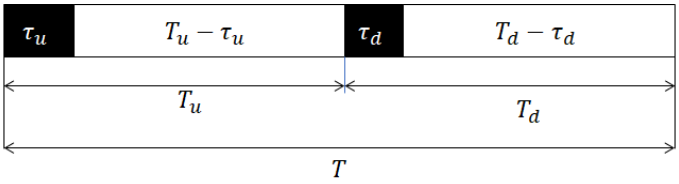

Fig. 1: The frame structure considered in this work.

Data is transmitted in frames consisting of T symbols. Each frame is comprised of an uplink subframe of duration T_u with $\tau_u \geq K$ training symbols followed by $T_u - \tau_u$ data symbols, and a downlink subframe of duration T_d with $\tau_d \geq 0$ training symbols[1] followed by $T_d - \tau_d$ data symbols (see Fig. 1.) For notational convenience, we assume the sample index n within each uplink/downlink subframe to take values between $n = -\tau_x + 1$ to $n = T_x - \tau_x$, with $x \in \{u, d\}$. In both these cases, the first τ_x symbols correspond to training, and data transmission starts from the sample index $n = 1$.

Now, during the uplink training phase, from $n = -\tau_u + 1$ to $n = 0$, the UEs transmit τ_u orthogonal pilot symbols in the uplink (thus, $\tau_u \geq K$), with the kth UE's pilot having a total energy $\mathcal{E}_{p,k}$. These pilots are used by the BS to obtain MMSE estimates $\hat{\mathbf{h}}_k[0]$ of $\mathbf{h}_k[0]$ related to the latter as

$$\mathbf{h}_k[0] = b_{u,k}[1,0]\hat{\mathbf{h}}_k[0] + \bar{b}_{u,k}[1,0]\tilde{\mathbf{h}}_k[0], \tag{2}$$

with $\tilde{\mathbf{h}}_k[n]$ being the channel estimation error, such that $E[\hat{\mathbf{h}}_k[0]\tilde{\mathbf{h}}^H_k[0]] = \mathbf{O}_N$. Here, $\mathbf{O}_N$ denotes the $N \times N$ all zero matrix. Also, $\bar{b}^2_{u,k}[n,p]$ denotes the mean squared estimation error in the channel vector $\mathbf{h}_k[n]$, at time n, and based on the signals received till time instant $p \leq n$. Specifically, if $\tau_u = K$ and the the pilot signal from the kth user is of the form $\delta[n-K+k]$ where $\delta[\cdot]$ is the Kronecker delta function, it is known that [17] $b_{u,k}[1,0] = \rho_k[K-k]\sqrt{\frac{\beta_k\mathcal{E}_{p,k}}{\beta_k\mathcal{E}_{p,k}+N_0}}$, where N_0 is the noise variance at the BS. Since $\rho_k[n]$ is a decreasing function of n, $b_{u,k}[1,0]$ can be lower bounded as

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

Note that $b_{u,k}[1,0]$ captures both the effect of channel aging as well as the channel estimation error.

In the uplink subframe, following pilot transmission, all the UEs simultaneously transmit data symbols during the next

[1]We will consider channel tracking at the BS using the data symbols transmitted during the uplink subframe. In this case, an up-to-date channel estimate is available at the BS at the end of the uplink subframe, using which it can transmit beamformed data to the users in the downlink. Therefore, downlink training may not be necessary, and we can set $\tau_d = 0$.

$T_u - K$ time instants. The data symbol transmitted by the kth user at the nth instant is denoted by $s_{u,k}[n]$, and is transmitted with an energy $\mathcal{E}_{u,s,k}$. The overall received symbol at the BS therefore becomes

$$\mathbf{y}_u[n] = \sum_{k=1}^{K} \sqrt{\beta_k \mathcal{E}_{u,s,k}} \mathbf{h}_k[n] s_{u,k}[n] + \sqrt{N_0}\mathbf{w}[n]. \quad (4)$$

The BS uses linear combining vectors $\mathbf{v}_k[n]$ $(1 \leq k \leq K)$ generated from the available channel estimates to estimate the symbols transmitted by the kth user as $\hat{s}_{u,k}[n]$.

Since the channel ages, the combining matrices based on the initial channel estimate available at the BS gradually become mismatched with the channel state. However, the recovered symbols can be used as pilot signals to update the available channel estimates, and thereby potentially prolong the usable uplink frame duration.

In the downlink subframe, the BS uses the available channel estimate to determine an appropriate precoding matrix $\mathbf{P} \in \mathbb{C}^{N \times K}$ for data transmission. This precoding matrix is then used to transmit beamformed pilots and data symbols simultaneously to all the UEs during the next T_d time instants. Denoting the data symbol transmitted to the kth user at the nth instant by $s_{d,k}[n]$, the precoded signal transmitted by the BS can be written as

$$\mathbf{x}[n] = \lambda \sum_{k=1}^{K} \sqrt{\epsilon_{d,s,k}} \mathbf{p}_k s_{d,k}[n], \quad (5)$$

with λ as the downlink power control coefficient such that $E[\|\mathbf{x}[n]\|_2^2] = \mathcal{E}_{d,s}$, $\mathbf{p}_k$ being the kth column of $\mathbf{P}$, and $\epsilon_{d,s,k} \geq 0$ being the fraction of the total downlink power allocated to the kth user such that $\sum_{k=1}^{K} \epsilon_{d,s,k} = 1$. Consequently, the symbol received at the kth UE is given by

$$y_{d,k}[n] = \lambda \sum_{l=1}^{K} \sqrt{\beta_k \epsilon_{d,s,l}} \mathbf{h}_k^T[n] \mathbf{p}_l s_{d,l}[n] + \sqrt{N_0}\nu_k[n], \quad (6)$$

with $\nu_k[n] \sim \mathcal{CN}(0,1)$ being the ZMCSCG additive noise. Now, letting $g_{kl}[n] = \frac{\lambda}{\sqrt{N}} \sqrt{\beta_k \epsilon_{d,s,l}} \mathbf{h}_k^T[n] \mathbf{p}_l$ denote the effective downlink channel for the lth user's data stream at the kth user, we can write

$$y_{d,k}[n] = \sqrt{N} g_{kk}[n] s_{d,k}[n] + \sqrt{N} \sum_{\substack{l=1 \\ l \neq k}}^{K} g_{kl}[n] s_{d,l}[n] + \sqrt{N_0}\nu_k[n]. \quad (7)$$

If no pilot signals are sent in the downlink, the UE recovers its data symbols by replacing the effective channel coefficient with its expected value, and treating interference as noise [2].

Since the channel varies over time, the precoding matrices based on the channel estimate available at the BS gradually become mismatched with the channel state. This results in the effective downlink channel drifting away from the positive real value assumed at the UE. However, the UEs can use the data symbol estimates as pilots to update their estimates of the effective channel, and thereby improve performance.

In the next sections, we discuss the Kalman filter based MMSE channel estimator and data detector for this scheme. We begin with the uplink case.

III. Uplink Channel Tracking and Symbol Detection

Let the MMSE estimate of the uplink channel from the kth user at the nth instant $(1 \leq n \leq T_u - K)$, and conditioned on the signals received till the $(n-1)$th instant be denoted by $\hat{\mathbf{h}}_k[n|\mathcal{Y}_{n-1}]$. By the orthogonality principle, $\hat{\mathbf{h}}_k[n|\mathcal{Y}_{n-1}]$ is related to the true channel at the nth instant, $\mathbf{h}_k[n]$, as

$$\mathbf{h}_k[n] = b_{u,k}[n, n-1]\hat{\mathbf{h}}_k[n|\mathcal{Y}_{n-1}] + \bar{b}_{u,k}[n, n-1]\tilde{\mathbf{h}}_k[n, n-1], \quad (8)$$

such that, $E[\hat{\mathbf{h}}_k[n|\mathcal{Y}_{n-1}]\tilde{\mathbf{h}}_k^H[n, n-1]] = \mathbf{O}_N$. Recall that $\bar{b}_{u,k}^2[n, n-1]$ is the mean squared error in the channel estimate at the nth instant based on the data received till the $(n-1)$th instant. Letting $\mathbf{H}[n] = [\mathbf{h}_1[n], \ldots, \mathbf{h}_K[n]]$, $\mathbf{b}_u[n, n-1] = [b_{u,1}[n, n-1], \ldots, b_{u,K}[n, n-1]]^T$, and $\mathbf{B}_u[n, n-1] = \text{diag}(\mathbf{b}_u[n, n-1])$, we can write the channel at the nth instant in terms of the available channel estimate as

$$\mathbf{H}[n] = \hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]\mathbf{B}_u[n, n-1] + \tilde{\mathbf{H}}[n, n-1]\bar{\mathbf{B}}_u[n, n-1]. \quad (9)$$

Also, letting $\boldsymbol{\rho}[n] = [\rho_1[n], \ldots, \rho_K[n]]^T$, the temporal evolution of $\mathbf{H}[n]$ can be expressed as

$$\mathbf{H}[n] = \mathbf{H}[n-1]\text{diag}(\boldsymbol{\rho}[1]) + \mathbf{Z}_h[n]\text{diag}(\bar{\boldsymbol{\rho}}[1]). \quad (10)$$

Let $\hat{s}_{u,k}[n]$ be the MMSE estimate of the symbol transmitted by the kth UE, such that $s_{u,k}[n] = a_{u,k}[n]\hat{s}_{u,k}[n] + \bar{a}_{u,k}[n]\tilde{s}_{u,k}[n]$, with $E[\hat{s}_{u,k}[n]\tilde{s}_{u,k}[n]^*] = 0$, and $E[|\hat{s}_{u,k}[n]|^2] = E[|\tilde{s}_{u,k}[n]|^2] = 1$. We discuss the derivation of the coefficient $\bar{a}_{u,k}[n]$ in Sec. IV (see Theorem 4). Also, defining $\hat{\mathbf{s}}_u[n] = [\hat{s}_{u,1}[n], \ldots, \hat{s}_{u,K}[n]]^T$ and $\tilde{\mathbf{s}}_u[n] = [\tilde{s}_{u,1}[n], \ldots, \tilde{s}_{u,K}[n]]^T$, we can equivalently write the signal vector received at the BS as

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

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]^T$, $\mathbf{a}_u[n] = [a_{u,1}[n], \ldots, a_{u,K}[n]]^T$, $\boldsymbol{\mathcal{E}}_{u,s} = [\mathcal{E}_{s,1}, \ldots, \mathcal{E}_{s,K}]^T$, and $\odot$ denotes the Hadamard product. Letting $\mathbf{h}[n] = \text{vec}(\mathbf{H}[n])$ and $\otimes$ denoting the Kronecker product, we rewrite (10) and (11) as

$$\mathbf{h}[n] = (\text{diag}(\boldsymbol{\rho}[1]) \otimes \mathbf{I}_N)\mathbf{h}[n-1] + (\text{diag}(\bar{\boldsymbol{\rho}}[1]) \otimes \mathbf{I}_N)\mathbf{z}_h[n], \quad (12)$$

$$\begin{aligned}\mathbf{y}_u[n] = {} & ((\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}} \odot \mathbf{a}_u[n] \odot \hat{\mathbf{s}}_u[n])^T \otimes \mathbf{I}_N)\mathbf{h}[n] \\ & + ((\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}} \odot \bar{\mathbf{a}}_u[n] \odot \tilde{\mathbf{s}}_u[n])^T \otimes \mathbf{I}_N)\mathbf{h}[n] \\ & + \sqrt{N_0}\mathbf{w}[n]. \quad (13)\end{aligned}$$

Viewing (12) and (13) as the process and observation equations for the channel state $\mathbf{h}[n]$ [37], we identify $\mathtt{P}[1] = (\text{diag}(\boldsymbol{\rho}[1]) \otimes \mathbf{I}_N)$ as the state evolution matrix and $\mathbf{C}[n] = (\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}} \odot \mathbf{a}_u[n] \odot \hat{\mathbf{s}}_u[n])^T \otimes \mathbf{I}_N$ as the observation matrix. Also, the term $(\bar{\boldsymbol{\rho}}[1] \otimes \mathbf{I}_N)\mathbf{z}_h[n]$ can be viewed as a temporally white process noise vector having a covariance matrix $\mathbf{Q}_1[n] = \text{diag}(\bar{\boldsymbol{\rho}}^2[1]) \otimes \mathbf{I}_N$. Similarly, with $\tilde{\mathbf{C}}[n] = ((\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}} \odot \bar{\mathbf{a}}_u[n] \odot \tilde{\mathbf{s}}_u[n])^T \otimes \mathbf{I}_N)$, we can write the observation noise covariance matrix, $\mathbf{Q}_2[n]$, as

$$\mathbf{Q}_2[n] = E\left[\tilde{\mathbf{C}}[n]\mathbf{h}[n]\mathbf{h}^H[n]\tilde{\mathbf{C}}^H[n]\right] + N_0\mathbf{I}_N$$

$$= \left(\sum_{k=1}^{K} \beta_k \mathcal{E}_{u,s,k} \bar{a}^2_{u,k}[n] + N_0\right) \mathbf{I}_N. \quad (14)$$

Using $\hat{\mathbf{h}}[n-1|\mathcal{Y}_{n-1}] \triangleq \text{vec}(\hat{\mathbf{H}}[n-1|\mathcal{Y}_{n-1}])$, $\tilde{\mathbf{h}}[n,n-1] \triangleq \text{vec}(\tilde{\mathbf{H}}[n,n-1])$ and $\hat{\mathbf{s}}_u[n]$, we can write the innovation component, denoted by $\tilde{\mathbf{y}}_u[n]$, as

$$\begin{aligned}\tilde{\mathbf{y}}_u[n] &= \mathbf{y}_u[n] - ((\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}} \odot \mathbf{a}_u[n] \odot \hat{\mathbf{s}}_u[n])^T \\ &\quad \times (\mathbf{B}_u[n,n-1]) \otimes \mathbf{I}_N)\hat{\mathbf{h}}[n|\mathcal{Y}_{n-1}] \\ &= \mathbf{C}[n](\mathbf{h}[n] - (\mathbf{B}_u[n,n-1] \otimes \mathbf{I}_N)\hat{\mathbf{h}}[n|\mathcal{Y}_{n-1}]) \\ &\quad + \tilde{\mathbf{C}}[n]\mathbf{h}[n] + \sqrt{N_0}\mathbf{w}[n] \\ &= \mathbf{C}[n](\bar{\mathbf{B}}_u[n,n-1] \otimes \mathbf{I}_N)\tilde{\mathbf{h}}[n,n-1] + \tilde{\mathbf{C}}[n]\mathbf{h}[n] \\ &\quad + \sqrt{N_0}\mathbf{w}[n]. \quad (15)\end{aligned}$$

The next step in deriving the Kalman filter based channel tracker is to derive the covariance matrix of the innovation component, which is given by the following Lemma.

Lemma 1. *The covariance matrix of the innovation process $\tilde{\mathbf{y}}_u[n]$ is given as*

$$\mathbf{R}_{\tilde{y}\tilde{y}}[n] = \left(\sum_{k=1}^{K} \beta_k \mathcal{E}_{u,s,k}(\bar{b}^2_{u,k}[n,n-1] + \bar{a}^2_{u,k}[n]) + N_0\right)\mathbf{I}_N. \quad (16)$$

Proof. See Appendix A. ∎

The expression for $\mathbf{R}_{\tilde{y}\tilde{y}}[n]$ contains three terms. The term $\bar{b}^2_{u,k}[n,n-1]$ depends on both the pilot power and on channel aging, and is explicitly derived in Theorem 3. The second term, $\bar{a}^2_{u,k}[n]$, depends on the MSE in the data symbols, and, as shown in Theorem 4, this in turn depends on the data SNR and $\bar{\mathbf{B}}^2_u[n,n-1]$. The third term is due to the AWGN.

To complete the derivation of the Kalman filtering based channel tracking procedure, we need to specify the MMSE channel predictor and the data symbol estimator, which is given by the following two Theorems.

Theorem 1. *The estimate $\hat{\mathbf{h}}[n+1|\mathcal{Y}_n]$ of $\mathbf{h}[n+1]$ can be obtained from $\tilde{\mathbf{y}}_u[n]$ and the available estimate $\hat{\mathbf{h}}[n|\mathcal{Y}_{n-1}]$ as*

$$\begin{aligned}\hat{\mathbf{h}}[n+1|\mathcal{Y}_n] &= (\mathbf{B}_u^{-1}[n+1,n]\mathbf{B}_u[n,n-1] \otimes \mathbf{I}_N) \\ &\quad \times \mathtt{P}[1]\hat{\mathbf{h}}[n|\mathcal{Y}_{n-1}] + \mathbf{G}[n]\tilde{\mathbf{y}}_u[n], \quad (17)\end{aligned}$$

where $\mathbf{G}[n]$ is the Kalman filtering gain, given as,

$$\begin{aligned}&\mathbf{G}[n] = (\mathbf{B}_u[n+1,n] \otimes \mathbf{I}_N)^{-1} \\ &\times \frac{(\boldsymbol{\rho}[1] \odot \bar{\mathbf{b}}^2_u[n,n-1] \odot \sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}} \odot \mathbf{a}_u[n] \odot \hat{\mathbf{s}}^*_u[n]) \otimes \mathbf{I}_N}{\sum_{k=1}^{K} \beta_k \mathcal{E}_{u,s,k}(\bar{b}^2_{u,k}[n,n-1] + \bar{a}^2_{u,k}[n]) + N_0}, \quad (18)\end{aligned}$$

with $\hat{\mathbf{h}}[1|\mathcal{Y}_{-1}]$ initialized via pilot training as discussed in [17], and $\mathbf{B}_u[1,0]$ is initialized using (3)*. Note that $\mathcal{Y}_{-1}$ corresponds to the received pilot signals.*

Proof. See Appendix B. ∎

Once the updated channel estimate is obtained at the BS at time $n-1$, the next received signal vector, $\mathbf{y}[n]$, is used to obtain an MMSE estimate of the transmitted symbols according to the following Theorem.

Theorem 2. *The MMSE estimate of $\mathbf{s}_u[n]$ can be obtained from $\mathbf{y}[n]$, and the available channel estimate $\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]$ as*

$$\begin{aligned}\hat{\mathbf{s}}_u[n] &= \mathbf{B}_u[n,n-1]diag(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}})\hat{\mathbf{H}}^H[n|\mathcal{Y}_{n-1}] \\ &\times \Bigg(\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]diag(\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s})\mathbf{B}^2_u[n,n-1]\hat{\mathbf{H}}^H[n|\mathcal{Y}_{n-1}] \\ &\quad + \left(\sum_{k=1}^{K} \bar{b}^2_{u,k}[n,n-1] + N_0\right)\mathbf{I}_N\Bigg)^{-1} \mathbf{y}_u[n]. \quad (19)\end{aligned}$$

Proof. See Appendix C. ∎

The MSE performance of the Kalman filter based channel tracker and the coefficients $b_{u,k}[n,n-1]$ in Theorem 1 can be obtained via Theorem 3 below.

Theorem 3. *The mean squared estimation error for the kth user's channel at the $(n+1)$th instant based on the observations till the nth instant is*

$$\begin{aligned}\bar{b}^2_{u,k}[n+1,n] &= \bar{\rho}^2_k[1] + \rho^2_k[1]\bar{b}^2_{u,k}[n,n-1] \times \Bigg(1 - \\ &\frac{\rho^2_k[1]a^2_{u,k}[n]\bar{b}^2_{u,k}[n,n-1]\beta_k\mathcal{E}_{u,s,k}}{b_{u,k}[n,n-1](\sum_{p=1}^{K} \beta_p \mathcal{E}_{u,s,p}(\bar{b}^2_{u,p}[n,n-1] + \bar{a}^2_{u,p}[n]) + N_0)}\Bigg). \quad (20)\end{aligned}$$

Proof. See Appendix D. ∎

Finally, the MSE in the symbol estimates is given by the following theorem.

Theorem 4. *The MSE of the received symbol estimate $\hat{\mathbf{s}}_{u,k}[n]$ is given by*

$$\bar{a}^2_{u,k}[n] = \frac{\sum_{l=1}^{K} \bar{b}^2_{u,l}[n,n-1] + N_0}{\beta_k \mathcal{E}_{u,s,k} b^2_{u,k}[n,n-1] + \sum_{l=1}^{K} \bar{b}^2_{u,l}[n,n-1] + N_0}. \quad (21)$$

Proof. See Appendix E. ∎

Also, for comparison, in the absence of Kalman filter based tracking (i.e., when the channel estimated from the pilot symbols is utilized for the rest of the frame, without tracking the channel variations), it is easy to show that the mean squared channel estimation error for the kth user can be reduced to

$$\bar{b}^2_{u,k}[n,0] = \rho^2_k[n]\bar{b}^2_{u,k}[1,0] + \bar{\rho}^2_k[n]. \quad (22)$$

We will use this expression later, to calculate the achievable rate of the system without channel tracking.

We summarize the channel tracking and data estimation procedure at the BS in Algorithm 1. We note that the calculation of $a_{u,k}[n]$ requires $\mathcal{O}(K)$ complex operations. Similarly, the use of Theorem 2 requires N^3 complex operations for inverting the $N \times N$ matrix, KN complex operations for generating the combining matrix and another KN complex operations for obtaining $\hat{s}_u[n]$ from $\mathbf{y}_u[n]$. Therefore, the overall computational complexity of the first step is $\mathcal{O}(KN+N^3)$. The second step has a computational complexity of $\mathcal{O}(K)$. Finally, in third step, the calculation of the Kalman gain

Algorithm 1: Data Aided Uplink Channel Tracking

Initialize $\hat{\mathbf{h}}_k[0]$ and $\mathbf{B}_u[1,0]$ using pilot based estimation.

for $n = 0, 1, \ldots, T_u - \tau_u$ **do**

1) Use (21) in Theorem 4 to compute $a_{u,k}[n]$, and use (19) in Theorem 2 to compute $\hat{\mathbf{s}}_u[n]$.
2) Compute $\mathbf{B}_u[n+1, n]$ using (20) in Theorem 3, and compute $\tilde{\mathbf{y}}_u[n]$ using (15).
3) Calculate $\mathbf{G}[n]$ from (18), and substitute it in (17) of Theorem 1 to obtain $\hat{\mathbf{h}}[n+1|\mathcal{Y}_n]$.

end

requires $\mathcal{O}(KN)$ complex operations and the channel update requires $\mathcal{O}((KN)^2)$ complex operations. Thus, the overall computational complexity of Algorithm 1 is $\mathcal{O}(N^3+(KN)^2)$ per symbol. We note that, when $K \ll N$, this is comparable to the $\mathcal{O}(N^3)$ complexity of the MMSE detector. Hence, Kalman filtering based channel tracking does not result in a significant increase in the computational load at the BS.

Based on the above results, in the next section, we theoretically analyze the performance of the channel tracking algorithm in terms of the uplink achievable rate.

IV. Uplink Achievable Rate Analysis

In the previous section, we derived the MSE performance of the Kalman filter based channel tracking technique. We now analyze the achievable rate of the system. Considering a generalized MMSE receiver based on the predicted channel, the BS uses $\mathbf{V}[n] = \mathbf{Q}^{-1}[n|\mathcal{Y}_{n-1}]\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]\text{diag}(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{\boldsymbol{u,s}}} \odot \mathbf{b}_u[n, n-1])$ as the receive combining matrix, where $\mathbf{Q}^{-1}[n|\mathcal{Y}_{n-1}] \triangleq (\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]\text{diag}(\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{\boldsymbol{u,s}} \odot \mathbf{b}_u^2[n, n-1])\hat{\mathbf{H}}^H[n|\mathcal{Y}_{n-1}] + \mu\mathbf{I}_N)^{-1}$ and μ is the regularization factor. Note that this generalized MMSE combiner reduces to the MMSE symbol estimator discussed previously for $\mu = \sum_{k=1}^{K} \bar{b}_{u,k}^2[n, n-1] + N_0$. Then, the combining vector for the kth stream is $\sqrt{\beta_k \mathcal{E}_{u,s,k}} b_{u,k}[n, n-1]\mathbf{Q}^{-1}[n|\mathcal{Y}_{n-1}]\hat{\mathbf{h}}_k[n|\mathcal{Y}_{n-1}]$. Therefore, $\hat{s}_{u,k}[n|\mathcal{Y}_{n-1}]$, which is the signal used at the BS to detect the kth user's data, can be expanded as

$$\begin{aligned}
\hat{s}_{u,k}[n|\mathcal{Y}_{n-1}] &= \beta_k \mathcal{E}_{u,s,k} b_{u,k}^2[n, n-1]\hat{\mathbf{h}}_k^H[n|\mathcal{Y}_{n-1}] \\
&\quad \times \mathbf{Q}^{-1}[n|\mathcal{Y}_{n-1}]\hat{\mathbf{h}}_k[n|\mathcal{Y}_{n-1}]s_{u,k}[n] \\
&\quad + \sum_{\substack{l=1 \\ l\neq k}}^{K} b_{u,l}[n, n-1]b_{u,k}[n, n-1]\sqrt{\beta_l \mathcal{E}_{u,s,l}\beta_k \mathcal{E}_{u,s,k}} \\
&\quad \times \hat{\mathbf{h}}_k^H[n|\mathcal{Y}_{n-1}]\mathbf{Q}^{-1}[n|\mathcal{Y}_{n-1}]\hat{\mathbf{h}}_l[n|\mathcal{Y}_{n-1}]s_{u,l}[n] \\
&\quad + \sum_{l=1}^{K} \bar{b}_{u,l}[n, n-1]b_{u,k}[n, n-1]\sqrt{\beta_l \mathcal{E}_{u,s,l}\beta_k \mathcal{E}_{u,s,k}} \\
&\quad \times \hat{\mathbf{h}}_k^H[n|\mathcal{Y}_{n-1}]\mathbf{Q}^{-1}[n|\mathcal{Y}_{n-1}]\tilde{\mathbf{h}}_l[n, n-1]s_{u,l}[n] \\
&\quad + \eta_k[n]. \qquad (23)
\end{aligned}$$

with $\eta_k[n] = \sqrt{\beta_k \mathcal{E}_{u,s,k}} b_{u,k}[n, n-1]\hat{\mathbf{h}}_k^H[n|\mathcal{Y}_{n-1}]\mathbf{Q}^{-1}[n|\mathcal{Y}_{n-1}]\mathbf{w}[n]$. Similarly, without the Kalman filter based channel tracking, the combining vector for the kth UE's signal is $\sqrt{\beta_k \mathcal{E}_{u,s,k}} b_{u,k}[1,0]\mathbf{Q}^{-1}[0]\hat{\mathbf{h}}_k[0]$, where $\mathbf{Q}^{-1}[0]$ equals $\mathbf{Q}^{-1}[n|\mathcal{Y}_{n-1}]$ defined above, with $n = 0$. Then, the received signal for the kth user becomes

$$\begin{aligned}
\hat{s}_{u,k}[n|\mathcal{Y}_{-1}] &= \beta_k \mathcal{E}_{u,s,k}\rho_k[n]b_{u,k}^2[1,0] \\
&\quad \times \hat{\mathbf{h}}_k^H[0]\mathbf{Q}^{-1}[0]\hat{\mathbf{h}}_k[0]s_{u,k}[n] \\
&\quad + \sum_{\substack{l=1 \\ l\neq k}}^{K} \rho_l[n]b_{u,l}[1,0]b_{u,k}[1,0]\sqrt{\beta_l \mathcal{E}_{u,s,l}\beta_k \mathcal{E}_{u,s,k}} \\
&\quad \times \hat{\mathbf{h}}_k^H[0]\mathbf{Q}^{-1}[0]\hat{\mathbf{h}}_l[0]s_{u,l}[n] \\
&\quad + \sum_{l=1}^{K} \rho_l[n]\bar{b}_{u,l}[1,0]b_{u,k}[1,0]\sqrt{\beta_k \mathcal{E}_{u,s,k}\beta_l \mathcal{E}_{u,s,l}} \\
&\quad \times \hat{\mathbf{h}}_k^H[0]\mathbf{Q}^{-1}[0]\tilde{\mathbf{h}}_l[0]s_{u,l}[n] \\
&\quad + \sum_{l=1}^{K} \bar{\rho}_l[n]b_{u,k}[1,0]\sqrt{\beta_k \mathcal{E}_{u,s,k}\beta_l \mathcal{E}_{u,s,l}} \\
&\quad \times \hat{\mathbf{h}}_k^H[0]\mathbf{Q}^{-1}[0]\mathbf{z}_{h,l}[n]s_{u,l}[n] + \eta_k[n], \qquad (24)
\end{aligned}$$

with $\eta_k[n] = \sqrt{\beta_k \mathcal{E}_{u,s,k}} b_{u,k}[1,0]\hat{\mathbf{h}}_k^H[0]\mathbf{Q}^{-1}[0]\mathbf{w}[n]$. In (24), the last term corresponds to the interference due to the aging of the available channel estimates without channel tracking and is absent when the Kalman filter is used, i.e., in (23).

In order to evaluate SINR of these signals, we need to compute the mean squared values of each of the terms in (23) and (24), and exploit the fact that the terms are all uncorrelated with each other. Since all the entries of the channel vectors are i.i.d. ZMCSCG random variables, we can use tools from random matrix theory [6], [17], to replace the expectation operation with its DE. Using results derived in [6], it can be shown that the DE of the achievable SINR with and without Kalman filter based channel tracking can be given as (25) and (26) on the next page, where the superscripts P and NP denote the Kalman filter based prediction and no prediction, respectively. Also, under channel tracking,

$$\varphi_k[n] \triangleq \left(\sum_{\substack{m=1 \\ m\neq k}}^{K} \frac{b_{u,m}^2[n, n-1]\beta_m \mathcal{E}_{u,s,m}}{1 + e_{k,m}[n]} + \mu \right)^{-1}, \qquad (27)$$

and $e_{k,m}[n]$ is iteratively computed as the limit $t \to \infty$ in

$$e_{k,m}^{(t)}[n] = \frac{b_{u,m}^2[n, n-1]\beta_m \mathcal{E}_{u,s,m}}{\sum_{i=1;i\neq k}^{K} \frac{b_{u,i}^2[n,n-1]\beta_i \mathcal{E}_{u,s,i}}{1+e_{k,i}^{(t-1)}[n]} + \mu}, \qquad (28)$$

with the initialization $e_{k,m}^{(0)}[n] = \frac{1}{\mu}$. Also,

$$\begin{aligned}
\dot{\epsilon}_{kl}[n] &= \dot{\varphi}_{kl}^2[n] + \frac{|b_{u,l}^2[n, n-1]\beta_l \mathcal{E}_{u,s,l}|^2 N^2 \dot{\varphi}_{kl}^4[n]}{|1 + b_{u,l}^2[n, n-1]\beta_l \mathcal{E}_{u,s,l} N \dot{\varphi}_{kl}[n]|^2} \\
&\quad - 2\Re\left\{ \frac{b_{u,l}^2[n, n-1]\beta_l \mathcal{E}_{u,s,l} N \dot{\varphi}_{kl}^3[n]}{1 + b_{u,l}^2[n, n-1]\beta_l \mathcal{E}_{u,s,l} N \dot{\varphi}_{kl}[n]} \right\}, \qquad (29)
\end{aligned}$$

where $\Re\{.\}$ denotes the real part of a complex number, and

$$\dot{\varphi}_{kl}[n] = \left(\sum_{m=1;m\neq l,k}^{K} \frac{b_{u,m}^2[n, n-1]\beta_m \mathcal{E}_{u,s,m}}{1 + \dot{e}_{k,l,m}[n]} + \mu \right)^{-1}, \qquad (30)$$

$$\gamma_{u,k}^{\mathrm{P}}[n] - \frac{N\beta_k\mathcal{E}_{u,s,k}b_{u,k}^2[n;n-1]}{\sum_{\substack{l=1\\l\neq k}}^{K}\beta_l\mathcal{E}_{u,s,l}b_{u,l}^2[n;n-1]\frac{\dot{\epsilon}_{kl}[n]}{\varphi_k^2[n]}+\sum_{l=1}^{K}\beta_l\mathcal{E}_{u,s,l}\bar{b}_{u,l}^2[n;n-1]+N_0} \xrightarrow{a.s.} 0. \tag{25}$$

$$\gamma_{u,k}^{\mathrm{NP}}[n] - \frac{N\rho_k^2[n]\beta_k\mathcal{E}_{u,s,k}b_{u,k}^2[1,0]}{\sum_{\substack{l=1\\l\neq k}}^{K}\rho_k^2[n]\beta_l\mathcal{E}_{u,s,l}b_{u,l}^2[1,0]\frac{\dot{\epsilon}_{kl}[n]}{\varphi_k^2[n]}+\sum_{l=1}^{K}\rho_l^2[n]\beta_l\mathcal{E}_{u,s,l}\bar{b}_{u,l}^2[1,0]+\sum_{l=1}^{K}\bar{\rho}_l^2[n]\beta_l\mathcal{E}_{u,s,l}+N_0} \xrightarrow{a.s.} 0. \tag{26}$$

with $\dot{e}_{k,l,m}[n]$ being iteratively computed as

$$\dot{e}_{k,l,m}^{(t)}[n] = \frac{b_{u,m}^2[n,n-1]\beta_m\mathcal{E}_{u,s,m}}{\sum_{i=1;i\neq m,k}^{K}\frac{b_{u,i}^2[n,n-1]\beta_i\mathcal{E}_{u,s,i}}{1+\dot{e}_{k,l,i}^{(t-1)}[n]}+\mu}, \tag{31}$$

with the initialization $\dot{e}_{k,l,m}^{(0)}[n] = \frac{1}{\mu}$. In case of no CSI tracking, the expressions in (27)–(31) remain the same, but with $\rho_k[n]b_{u,k}[1,0]$ replacing $b_{u,k}[n,n-1]$.

We note that a key difference between the SINRs with and without channel tracking ((25) and (26)) is in their dependence on channel aging. While (25) depends only on $\rho_k[1]$ (through $b_{u,k}[n,n-1]$ in (20)), (26) depends directly on $\rho_k[n]$. At high user velocities, $\rho_k[n]$ decreases rapidly with n, leading to significantly lower SINRs without channel tracking compared to the Kalman filtering based channel tracking algorithm.

Uplink Achievable Sum Rate: With the SINR expressions in hand, we can now write the uplink sum rate (x $\in$ {P,NP} with and without channel tracking, respectively) as

$$R^{\mathrm{x}} = \frac{1}{T_u}\sum_{k=1}^{K}\sum_{n=1}^{T_u-K}\log_2(1+\gamma_k^{\mathrm{x}}[n]). \tag{32}$$

From (29), as the SNR, and consequently, the term $N\beta_l\mathcal{E}_{u,s,l}$ increases, the term $\dot{\epsilon}_{kl}[n]$ converges to zero for all k and l, thus indicating that the inter-stream interference in MMSE based detection is negligibly small. However, from (24), we observe that a mismatch between the true and the estimated channel results in an additional interference term, which limits the performance of the system. Since the mismatch in the channel estimate and the true channel is smaller with Kalman filtering based channel tracking compared to the no tracking case, Kalman filtering not only helps to preserve the array gain of a massive MIMO system, but also leads to improved interference suppression in case of MMSE combining at the BS.

V. Blind Channel Tracking and Symbol Detection in the Downlink

Let the MMSE estimate of the effective downlink channel at the kth user, and at the nth ($1 \leq n \leq T_d - K$) instant, conditioned on the signals received till the $(n-1)$th instant be denoted by $\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]$. As a consequence of MMSE estimation, $\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]$ is related to $g_{kk}[n]$, the true effective downlink channel at the nth instant (see (7)), as

$$g_{kk}[n] = b_{d,k}[n,n-1]\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]+\bar{b}_{d,k}[n,n-1]\tilde{g}_{kk}[n,n-1], \tag{33}$$

with $\tilde{g}_{kk}[n,n-1]$ corresponding to the channel estimation error at the nth instant based on the signals received till the $(n-1)$th instant, such that, $E[\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]\tilde{g}_{kk}^*[n,n-1]] = 0$, and $E[|\tilde{g}_{kk}[n,n-1]|^2] = E[|\hat{g}_{kk}[n,n-1]|^2] = E[|g_{kk}[n,n-1]|^2]$, and $\bar{b}_{d,k}^2[n,n-1]$ is the normalized mean squared channel estimation error for the nth instant based on the data received till the $(n-1)$th instant.

Accounting for the effects of channel aging, $g_{kk}[n]$ can be alternatively expressed as

$$\begin{aligned}\sqrt{N}g_{kk}[n] &= \lambda\sqrt{\beta_k\epsilon_{d,s,k}}(\rho_k[1]\mathbf{h}_k[n-1]+\bar{\rho}[1]\mathbf{z}_{h,k}[1])^T\mathbf{p}_k\\ &= \sqrt{N}\rho_k[1]g_{kk}[n-1]+\sqrt{N}\bar{\rho}_k[1]\zeta_{kk}[n],\end{aligned} \tag{34}$$

with $\epsilon_{d,s,k} \geq 0$ being the fraction of the total downlink energy allocated to the kth user such that $\sum_{k=1}^{K}\epsilon_{d,s,k} = 1$, as before.

Letting $s_{d,k}[n] = a_{d,k}[n]\hat{s}_{d,k}[n] + \bar{a}_{d,k}[n]\tilde{s}_{d,k}[n]$, with $E[\hat{s}_{d,k}[n]\tilde{s}_{d,k}[n]^*] = 0$, and $E[|\tilde{s}_{d,k}[n]|^2] = 1$, the signal received at the kth UE can be equivalently written as

$$\begin{aligned}y_{d,k}[n] =& \sqrt{N}a_{d,k}[n]g_{kk}[n]\hat{s}_{d,k}[n]+\sqrt{N}\bar{a}_{d,k}[n]g_{kk}[n]\tilde{s}_{d,k}[n]\\ &+\sqrt{N}\sum_{\substack{l=1\\l\neq k}}^{K}g_{kl}[n]s_{d,l}[n]+\sqrt{N_0}\nu_k[n].\end{aligned} \tag{35}$$

Viewing (34) and (35) as the process and observation equations for the effective channel state $g_{kk}[n]$, we identify $\rho_k[1]$ as the state evolution coefficient and $\sqrt{N}a_{d,k}[n]\hat{s}_{d,k}[n]$ as the observation coefficient [37]. Also, $\bar{\rho}_k[1]\zeta_{kk}[n]$ is a temporally white process noise with variance

$$\sigma_1^2 = \lambda^2\beta_k\epsilon_{d,s,k}\bar{\rho}_k^2[1]\|\mathbf{p}_k\|_2^2. \tag{36}$$

Similarly, the variance of the observation noise is given by

$$\begin{aligned}\sigma_2^2 =& NE\left[|g_{kk}[n]|^2\right]E[|\tilde{s}_{d,k}[n]|^2]\,\bar{a}_{d,k}^2[n]\\ &+N\sum_{\substack{l=1\\l\neq k}}^{K}E\left[|g_{kl}[n]|^2|s_{d,l}[n]|^2\right]+N_0.\end{aligned} \tag{37}$$

Using the previously available estimate $\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]$ and $\hat{s}_{d,k}[n]$, we can define the prediction of $y_{d,k}[n]$ at the UE as $\hat{y}_k[n] = \sqrt{N}a_{d,k}[n]b_{d,k}[n,n-1]\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]\hat{s}_{d,k}[n]$, and write the innovation component $\tilde{y}_k[n]$ as

$$\begin{aligned}\tilde{y}_{d,k}[n] =& y_{d,k}[n]\\ &-\sqrt{N}a_{d,k}[n]b_{d,k}[n,n-1]\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]\hat{s}_{d,k}[n]\\ &=\sqrt{N}\left[\hat{s}_{d,k}[n]a_{d,k}[n]\bar{b}_{d,k}[n,n-1]\tilde{g}_{kk}[n,n-1]\right.\\ &\left.+\bar{a}_{d,k}[n]g_{kk}[n]\tilde{s}_{d,k}[n]+\sum_{\substack{l=1\\l\neq k}}^{K}g_{kl}[n]s_{d,l}[n]\right]+\sqrt{N_0}\nu_k[n].\end{aligned} \tag{38}$$

(39)

Since the transmitted symbols, the symbol errors, and the additive noise are all zero mean and uncorrelated, we can obtain the variance of the innovation process $\tilde{y}_k[n]$ as given by the following Lemma. We omit the proof as it is immediate.

Lemma 2. *The variance of the innovation process $\tilde{y}_{d,k}[n]$ is given by*

$$\begin{aligned}\sigma^2_{\tilde{y}\tilde{y},k}[n] = N\big(&a^2_{d,k}[n]\bar{b}^2_{d,k}[n,n-1]E[|\tilde{g}_{kk}[n,n-1]|^2] \\ &+ \bar{a}^2_{d,k}[n]E\left[|g_{kk}[n]|^2\right]\big) \\ &+ \lambda^2\sum_{\substack{l=1\\l\neq k}}^{K}\beta_k\epsilon_{d,s,l}E\left[|\mathbf{h}^T_k[n]\mathbf{p}_l|^2\right] + N_0. \end{aligned} \tag{40}$$

Similar to the uplink case, the innovation component can now be used to obtain an updated estimate of the effective downlink channel according to following theorem.

Theorem 5. *The estimate $\hat{g}_{kk}[n+1|\mathcal{Y}_n]$ of $g_{kk}[n+1]$ can be obtained from $y_{d,k}[n]$ and the previous estimate $\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]$ as*

$$\hat{g}_{kk}[n+1|\mathcal{Y}_n] = \frac{b_{d,k}[n,n-1]}{b_{d,k}[n+1,n]}\rho_k[1]\hat{g}_{kk}[n|\mathcal{Y}_{n-1}] + \Gamma_k[n]\tilde{y}_{d,k}[n], \tag{41}$$

where $\Gamma_k[n]$ is the Kalman filtering gain, and is given as

$$\Gamma_k[n] = \frac{\rho_k[1]\bar{b}^2_{d,k}[n,n-1]E[|g_{kk}[l]|^2]a_{d,k}[n]\hat{s}^*_{d,k}[n]}{b_{d,k}[n+1,n]\sigma^2_{\tilde{y}\tilde{y},k}[n]}. \tag{42}$$

Proof. See Appendix F. ∎

We use Theorem 5 to update the available channel estimates using the received data samples. Now, since we do not consider any explicit downlink training, we initialize the channel estimates at the UEs as $\hat{g}_{kk}[0] = E[g_{kk}[0]]$ and $\bar{b}^2_{d,k}[1,0] = E[|g_{kk}[0] - E[g_{kk}[0]]|^2]$. Once the effective downlink channel at the $(n+1)$th instant is estimated using the nth detected symbol, the UE then proceeds to obtain $\hat{s}_k[n+1]$ from $y_{d,k}[n+1]$, and repeats the cycle. In order to complete the process, we also need to obtain the MMSE estimate of the data symbol $s_{d,k}[n]$, which is given by the following theorem.

Theorem 6. *The MMSE estimate of $s_{d,k}[n]$ can be obtained from $y_{d,k}[n]$ and the available channel estimate $\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]$ as* (43) *on the next page.*

Proof. We can write the received signal $y_{d,k}[n]$ as,

$$\begin{aligned} y_{d,k}[n] = &\sqrt{N}b_{d,k}[n,n-1]\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]s_{d,k}[n] \\ &+ \sqrt{N}\bar{b}_{d,k}[n,n-1]\tilde{g}_{kk}[n,n-1]s_{d,k}[n] \\ &+ \sqrt{N}\sum_{\substack{l=1\\l\neq k}}^{K} g_{kl}[n]s_{d,l}[n] + \sqrt{N_0}\nu_k[n]. \end{aligned} \tag{44}$$

We can obtain (43) by following steps similar to those in the proof of Theorem 2. ∎

Theorem 7. *The mean squared estimation error of the kth user's channel at the $(n+1)$th instant based on the observations till nth instant is given by* (45)*, where $\sigma^2_{\tilde{y}\tilde{y},k}[n]$ is given by* (40)*.*

Algorithm 2: Data Aided Downlink Channel Tracking

Initialize $\hat{g}_{kk}[0]$ and $\mathbf{b}_{d,k}[1,0]$ their deterministic equivalent values.

for *n=0,1,...* $T_d - \tau_d$ **do**

1) Use (46) in Theorem 8 to compute $a_{d,k}[n]$, and use (43) in Theorem 6 to compute $\hat{s}_{d,k}[n]$.
2) Compute $b_{d,k}[n+1,n]$ using (45) in Theorem 7, and compute $\tilde{y}_{d,k}[n]$ using (38).
3) Calculate $\Gamma_k[n]$ from (42), and substitute it in (41) of Theorem 5 to obtain $\hat{g}_{kk}[n+1|\mathcal{Y}_n]$.

end

Proof. See Appendix G. ∎

We now present an expression for the the mean squared symbol estimation error at the UE, that is required to update the available estimate of the effective channel at the UEs.

Theorem 8. *The MSE of the received symbol estimate at the UE can be expressed as* (46)*.*

It can be observed that the symbol estimation error consists of three terms, with the first term corresponding to the estimation error for the effective downlink channel to the kth UE, the second term arising from the interference due to the data being transmitted to all the other UEs, and the third due to the additive noise.

In the absence of Kalman filter based tracking, it is easy to show that the mean squared channel estimation error of the kth user's channel can be reduced to

$$\bar{b}^2_{d,k}[n,0] = \rho^2_k[n]\bar{b}^2_{d,k}[1,0] + \bar{\rho}^2_k[n]. \tag{47}$$

An intuitive comparison of (47) with (45) is hard due to the large number of terms involved in (45). We therefore defer this discussion to the section on simulation results.

We summarize the overall channel tracking and data estimation procedure at the each of the users in Algorithm 2. We note that, since all the channel coefficients and corresponding symbols are scalars, this algorithm has a constant ($\mathcal{O}(1)$) computational complexity.

We use the above derived results to derive the achievable rate of the system in the next section.

VI. Downlink Achievable Rate Analysis

In the previous section, we derived the Kalman filtering based channel tracker and symbol detector, and its MSE performance. With these results in hand, we now proceed with analyzing achievable rate of the system in the downlink.

We consider that the BS employs matched filter (MF) precoding, such that the precoding matrix $\mathbf{P}$ takes the form $\mathbf{P} = \hat{\mathbf{H}}^*[0]$,[2] and consequently,

$$\mathbf{x}[n] = \lambda\sum_{k=1}^{K}\sqrt{\epsilon_{d,s,k}}\hat{\mathbf{h}}^*_k[0]s_k[n], \tag{48}$$

[2]We use the MF precoding here for simplicity, but the analysis can be extended to other schemes such as regularized zero-forcing precoding also.

$$\hat{s}_{d,k}[n] = \frac{\sqrt{N} b_{d,k}[n, n-1] \hat{g}^{*}_{kk}[n|\mathcal{Y}_{n-1}] y_{d,k}[n]}{N b^2_{d,k}[n, n-1] |\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]|^2 + N \bar{b}^2_{d,k}[n, n-1] E[|g_{kk}[n]|^2] + N(\sum_{\substack{l=1 \\ l \neq k}}^{K} E[|g_{kl}[n]|^2]) + N_0} \tag{43}$$

$$\bar{b}^2_{d,k}[n+1, n] = \bar{\rho}^2_k[1] \lambda^2 \beta_k \epsilon_{d,s,k} \|\mathbf{p}_k\|^2_2 + \rho^2_k[1] \bar{b}^2_{d,k}[n, n-1] \left(1 - \frac{\rho^2_k[1] \bar{b}^2_{d,k}[n, n-1] a^2_{d,k}[n]}{b_{d,k}[n, n-1] \sigma^2_{\tilde{y}\tilde{y},k}[n]} \right) \tag{45}$$

$$\bar{a}^2_{d,k}[n] = \frac{N \bar{b}^2_{d,k}[n, n-1] E[|g_{kk}[n]|^2] + N \sum_{\substack{l=1 \\ l \neq k}}^{K} E\left[|g_{kl}[n]|^2\right] + N_0}{N b^2_{d,k}[n, n-1] |\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]|^2 + N \bar{b}^2_{d,k}[n, n-1] E[|g_{kk}[n]|^2] + N \sum_{\substack{l=1 \\ l \neq k}}^{K} E\left[|g_{kl}[n]|^2\right] + N_0} \tag{46}$$

with $E[\|\mathbf{x}[n]\|^2_2] = N\lambda^2 \sum_{k=1}^{K} \epsilon_{d,s,k}$. Using $\sum_{k=1}^{K} \epsilon_{d,s,k} = 1$, we have $\lambda = \sqrt{\mathcal{E}_{d,s}/N}$. Defining $\mathcal{E}_{d,s,l} \triangleq \epsilon_{d,s,l} \mathcal{E}_{d,s}$, we can write

$$g_{kl}[n] = \frac{\sqrt{\beta_k \mathcal{E}_{d,s,l}}}{N} \mathbf{h}^T_k[n] \hat{\mathbf{h}}^{*}_l[0]. \tag{49}$$

When the number of BS antennas is large, we can use results from random matrix theory [6] to argue that $g_{kl}[n]$ converges to its DE, such that,

$$g_{kk}[n] - \rho_k[n] b_{u,k}[1, 0] \sqrt{\beta_k \mathcal{E}_{d,s,k}} \xrightarrow{\text{a.s.}} 0, \tag{50}$$

$$g_{kl}[n] \xrightarrow{\text{a.s.}} 0, \quad l \neq k, \tag{51}$$

$$|g_{kk}[n]|^2 - \left(\rho^2_k[n] b^2_{u,k}[1, 0] \beta_k \mathcal{E}_{d,s,k} + \frac{\beta_k \mathcal{E}_{d,s,k}}{N} \right) \xrightarrow{\text{a.s.}} 0, \tag{52}$$

and

$$|g_{kl}[n]|^2 - \frac{\beta_k \mathcal{E}_{d,s,l}}{N} \xrightarrow{\text{a.s.}} 0, \quad l \neq k. \tag{53}$$

Now, if the kth UE employs the Kalman filter based tracking of the effective downlink channel, then the signal received at the nth instant can be expressed as

$$\begin{aligned} y_{d,k}[n] = {} & \sqrt{N} b_{d,k}[n, n-1] \hat{g}_{kk}[n|\mathcal{Y}_{n-1}] s_k[n] \\ & + \sqrt{N} \tilde{b}_{d,k}[n, n-1] \tilde{g}_{kk}[n] s_k[n] \\ & + \sqrt{N} \sum_{\substack{l=1 \\ l \neq k}}^{K} g_{kl}[n] s_l[n] + \sqrt{N_0} w_k[n]. \end{aligned} \tag{54}$$

Also, without Kalman filter based channel tracking, the signal received at the kth UE can be expressed as

$$\begin{aligned} y_{d,k}[n] = {} & \sqrt{N} \hat{g}_{kk}[0] s_k[n] + \sqrt{N} (g_{kk}[n] - \hat{g}_{kk}[0]) s_k[n] \\ & + \sqrt{N} \sum_{\substack{l=1 \\ l \neq k}}^{K} g_{kl}[n] s_l[n] + \sqrt{N_0} w_k[n]. \end{aligned} \tag{55}$$

In both the above expressions, the first term corresponds to the desired signal received over a known channel, and all the other terms to noise and interference. It is important to note that both $\hat{g}_{kk}[n|\mathcal{Y}_{n-1}]$ and $\hat{g}_{kk}[0]$ are MMSE estimates of the true downlink channel based on independent sets of observation vectors. Hence, the noise and interference terms for both cases are uncorrelated with the desired signal. Using the worst case noise theorem [38], we can treat the interference from all sources as independent Gaussian noise. Also, from [17], we can write the effective SINR of the kth user's uplink signal at the nth instant with the Kalman filter channel tracking being employed at the UE as (56) at the top of the next page, with the superscript P indicating the use of Kalman filter based tracking. Similarly, the achievable SINR for the kth user without Kalman channel tracking is given as (57) on the next page, with the superscript NP indicating the absence of Kalman prediction. In (56), the primary effect of channel aging is captured by the first term in the denominator, $\bar{b}^2_{d,k}[n, n-1]$. When the channel tracking at the UEs is effective, the value of $\bar{b}^2_{d,k}[n, n-1]$ is small, which mitigates the impact of channel aging. However, in the absence of tracking, from (57), the noise and interference depend on $\bar{\rho}^2_k[n]$, making channel aging more harmful without channel tracking, especially at high user velocities and large values of n.

Now, for a downlink frame duration of T_d, the achievable sum rate with and without channel prediction is given by

$$R^{\text{x}} = \frac{1}{T_d} \sum_{k=1}^{K} \sum_{n=1}^{T_d - K} \log_2(1 + \gamma^{\text{x}}_k[n]), \tag{58}$$

with x $\in$ {P,NP}. The variable T_d can either be fixed, or optimized to satisfy a minimum quality of service constraint for the users. The latter is a one-dimensional optimization problem that can be easily solved using a simple line-search or bisection-based procedure.

VII. Simulation Results

In this section, we numerically quantify and compare the performance of the different channel tracking schemes, and contrast them against the case with no tracking. We consider a single cell system containing an $N = 256$ antenna BS serving $K = 16$ users transmitting at a carrier frequency $f_c = 3$ GHz with a signal bandwidth 100 kHz. We also assume that the BS samples the in-phase and quadrature components of the signal at the Nyquist rate of the complex baseband signal, i.e., at 100 kHz. The channel is assumed to age according to the AR-1 model, i.e., $\rho[n] = \rho^n$, with the correlation coefficient ρ lying in the range $0.999 \leq \rho \leq 0.99999$ [12]. More specifically, we consider $\rho = 0.99999$, $\rho = 0.9999$ and $\rho = 0.999$, corresponding to user velocities of 10, 70, and 270 km/h, respectively [12]. The frame duration is fixed at

$$\gamma_k^{\mathrm{P}}[n] - \frac{(1+Nb_{u,k}^2[1,0]\rho_k^2[n])\beta_k\mathcal{E}_{d,s,k}b_{d,k}^2[n,n-1]}{\bar{b}_{d,k}^2[n,n-1](1+Nb_{u,k}^2[1,0]\rho_k^2[n])\beta_k\mathcal{E}_{d,s,k}+\sum_{\substack{l=1\\l\neq k}}^{K}\beta_k\mathcal{E}_{d,s,l}+N_0} \xrightarrow{\text{a.s.}} 0. \tag{56}$$

$$\gamma_k^{\mathrm{NP}}[n] - \frac{N\beta_k\mathcal{E}_{d,s,k}b_{u,k}^2[0,0]}{(N\bar{\rho}_k^2[n]b_{u,k}^2[0,0]+\rho_k^2[n]\bar{b}_{d,k}^2[0,0]+\bar{\rho}_k^2[n])\beta_k\mathcal{E}_{d,s,k}+\sum_{\substack{l=1\\l\neq k}}^{K}\beta_k\mathcal{E}_{d,s,l}+N_0} \xrightarrow{\text{a.s.}} 0 \tag{57}$$

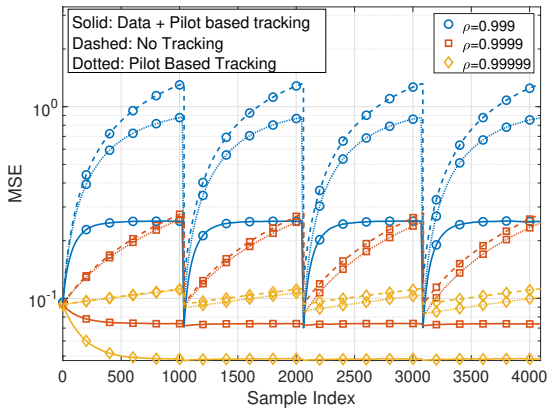

Fig. 2: MSE in uplink channel estimation vs. sample index, for different values of ρ, and for the first four frames.

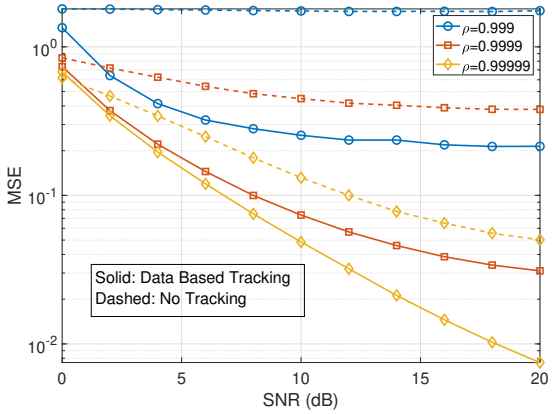

Fig. 3: MSE in the uplink channel estimate with and without Kalman filtering based tracking at the BS, as a function of the data SNR.

$T = 1024$ symbols [17]. However, for ease of presentation, while discussing the uplink case, we consider $T = T_u$ and $T_d = 0$, with the T_u uplink symbols consisting of $\tau_u = K$ training symbols followed by $T_u - K$ data symbols. In the downlink case, we consider $T_u = \tau_u = K$ and $\tau_d = 0$, such that $T = \tau_u + T_d$. Also, for simplicity, we consider that the path loss coefficients of all users are equal to unity, which corresponds to users employing path loss inversion based power control.[3] Also, unless specified otherwise, the data and pilot SNRs are assumed to be 10 dB.

A. Uplink Channel Tracking

In Fig. 2, we plot the MSE in the channel estimate obtained using the Kalman filter based channel tracking at the BS, as a function of the sample index, for the first four frames. We track the channel evolution via Theorem 1, which in turn involves updating the MSE using (20) in Theorem 3. We also compare the MSE against that obtained using the pilot only based tracking discussed in [34], [35] as well as without tracking, which corresponds to assuming the conventional block-fading channel to obtain pilot-only based channel estimates. We observe that, at all user mobilities, data-and-pilot based tracking achieves a lower MSE compared to the other two schemes as the sample index increases. In fact, at low and medium user velocities, the performance of the Kalman filter based tracking is constant across the frame, i.e., it is able to completely overcome the effect of channel aging. Further, the performance advantage offered by the data-aided Kalman filter based tracking increases with user mobility.

[3]The results for other forms of power control at the UEs are similar, but are not included here due to lack of space.

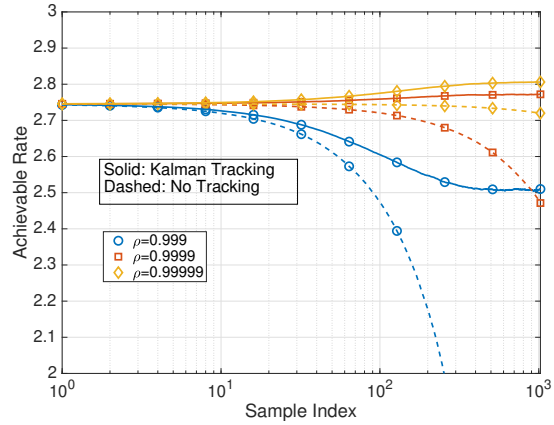

Fig. 4: Uplink rate with and without Kalman filtering based tracking vs. the sample index, for different values of ρ.

We plot the MSE at the end of the second frame as a function of data and pilot SNR for different values of ρ, in Fig. 3. For this simulation, we set the data SNR to be equal to the pilot SNR. We see that Kalman filtering offers significant advantage over no filtering in all the cases. Furthermore, at higher user mobility, the MSE begins to saturate as the SNR increases. This is because channel aging results in estimation errors, which cannot be overcome by data-aided channel tracking. On the other hand, at lower user mobility, Kalman tracking does result in lower MSE at higher SNRs, due to the lower symbol error rates and better channel tracking.

In Fig. 4, we plot the uplink achievable rate at the BS,

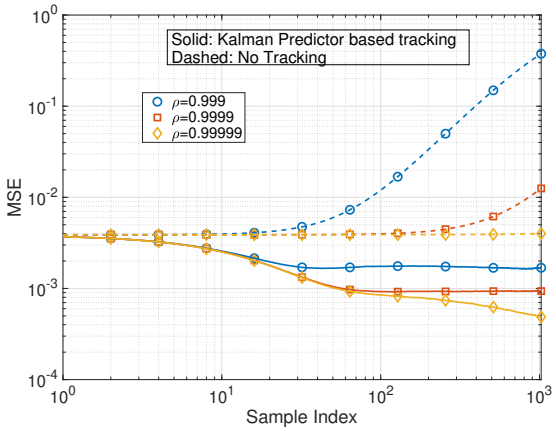

Fig. 5: MSE in the downlink channel estimate vs. sample index, with and without Kalman filtering based tracking, for different values of ρ.

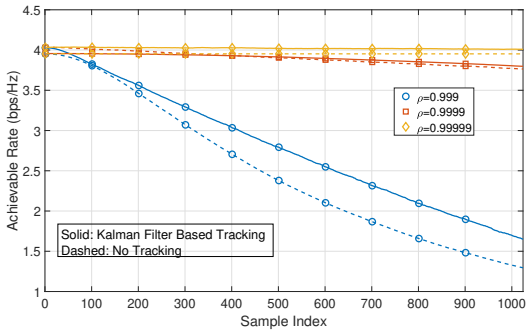

Fig. 6: Achievable downlink rate with and without Kalman filtering based downlink channel tracking vs. the sample index, for different values of ρ.

with and without Kalman filtering based channel tracking, as a function of the sample index, for different values of ρ. These correspond to the rate expression in (32), with the SINRs obtained using (25) and (26). When the channel is fast varying, Kalman filtering limits the deterioration in the uplink rates over time. On the other hand, for slowly varying channels, it actually leads to a marginal improvement in the achievable rates, as the channel estimates improve with the sample index. Further, in all cases, Kalman filtering offers much higher achievable rates compared to pilot-based channel estimation without channel tracking, particularly at higher sample indices.

B. Downlink Channel Tracking

In Fig. 5, we plot the MSE in the downlink channel estimate at a user, both with and without Kalman filter based blind channel tracking, obtained via Theorems 5 and 8, as a function of the sample index. Kalman filter based tracking improves the channel MSE in all the cases, and is particularly helpful in the case of high mobility users, where the channel ages rapidly. In Fig. 6, we plot the downlink rates achievable by a single user under matched filter precoding, with and without Kalman filtering based channel tracking, as a function of the sample index, and for different user mobilities. The rates are computed by substituting (56) and (57) in (58). Kalman filtering mitigates the performance loss due to aging for fast moving users, while offering a marginal improvement for slow moving users. However, unlike in the uplink case, the performance loss due to aging in fast moving users is larger, because the precoding matrix and the true downlink channel are still mismatched, and we are only able to track the variations in the *effective* downlink channel at the user.

VIII. Conclusions

In this work, we developed Kalman filter based data aided channel tracking algorithms in massive MIMO systems under aging channels for both uplink and downlink transmission. We also derived the mean squared error performance of channel tracking and received symbol estimation. We used these results to obtain expressions for the rates achievable with and without channel tracking in a massive MIMO system with time-varying channels. In the downlink, the gains due to channel tracking are lower, because the mismatch between the precoding matrix (which is based on the channel matrix available at the BS at the end of the training duration) and the true channel increases with time. We validated our derived results using Monte Carlo simulations, and found that Kalman filter based tracking indeed results in a substantial improvement in the achievable uplink rates, and a marginal improvement in the achievable downlink rates. Motivated by this difference in the tracking performance uplink and downlink channels, future work could consider the use of channel tracking in full duplex massive MIMO systems under channel aging.

Appendix

A. Proof of Lemma 1

From the definition of $\mathbf{R}_{\tilde{y}\tilde{y}}[n]$, we have

$$\begin{aligned}\mathbf{R}_{\tilde{y}\tilde{y}}[n] &= E[\tilde{\mathbf{y}}_u[n]\tilde{\mathbf{y}}_u^H[n]] \\ &= E[(\mathbf{C}[n](\bar{\mathbf{B}}_u[n,n-1]\otimes\mathbf{I}_N)\tilde{\mathbf{h}}[n,n-1]+\tilde{\mathbf{C}}[n]\mathbf{h}[n] \\ &+\sqrt{N_0}\mathbf{w}[n])\times(\mathbf{C}[n](\bar{\mathbf{B}}_u[n,n-1]\otimes\mathbf{I}_N)\tilde{\mathbf{h}}[n,n-1] \\ &+\tilde{\mathbf{C}}[n]\mathbf{h}[n]+\sqrt{N_0}\mathbf{w}[n])^H], \quad (59)\end{aligned}$$

where $(\bar{\mathbf{B}}_u[n,n-1]\otimes\mathbf{I}_N)\tilde{\mathbf{h}}[n,n-1] = \mathbf{h}[n] - (\mathbf{B}_u[n,n-1]\otimes\mathbf{I}_N)\hat{\mathbf{h}}[n|\mathcal{Y}_{n-1}]$. Since the data symbols, symbol errors and additive noise are all zero mean and uncorrelated,

$$\begin{aligned}\mathbf{R}_{\tilde{y}\tilde{y}}[n] &= E[\mathbf{C}[n]\bar{\mathbf{B}}_u[n,n-1]E[\tilde{\mathbf{h}}[n,n-1]\tilde{\mathbf{h}}^H[n,n-1]] \\ &\times\bar{\mathbf{B}}_u[n,n-1]\mathbf{C}^H[n]]+\mathbf{Q}_2[n] \\ &= E[\mathbf{C}[n]\bar{\mathbf{B}}_u^2[n,n-1]\mathbf{C}^H[n]]+\mathbf{Q}_2[n], \quad (60)\end{aligned}$$

where $\mathbf{Q}_2[n]$ was defined in (14). Also,

$$\begin{aligned}&E[\mathbf{C}[n]\bar{\mathbf{B}}_u^2[n,n-1]\mathbf{C}^H[n]] \\ &\quad=\left(\sum_{k=1}^{K}\beta_k\mathcal{E}_{u,s,k}\bar{b}_{u,k}^2[n,n-1]\right)\mathbf{I}_N, \quad (61)\end{aligned}$$

Substituting (61) and (14) into (60), we obtain (16).

B. Proof of Theorem 1

The standard Kalman filtering formulation [37] begins by writing the channel estimate as a linear combination of all the past innovation components, including the pilot signal, as

$$\hat{\mathbf{h}}[n+1|\mathcal{Y}_n] = \sum_{i=-K+1}^{n} \boldsymbol{\Psi}[n+1,i]\tilde{\mathbf{y}}_u[i], \tag{62}$$

where $\boldsymbol{\Psi}[n+1,i] \in \mathbb{C}^{KN\times N}$ is the gain associated with the ith innovation component. We know from the principle of orthogonality that,

$$E[(\mathbf{h}[n+1]-(\mathbf{B}_u[n+1,n]\otimes\mathbf{I}_N)\hat{\mathbf{h}}[n+1|\mathcal{Y}_n])\tilde{\mathbf{y}}_u^H[l]|\mathcal{Y}_n] = \mathbf{O}_{KN\times N}, \text{ for } 1\le l\le n. \tag{63}$$

Given $\mathcal{Y}_n$, $\mathbf{C}[n]$ is known, while $\tilde{\mathbf{C}}[n]$ is random and unknown. Therefore, the expectation above can be evaluated conditioned on $\mathbf{C}[n]$, and we get

$$\begin{aligned}E[\mathbf{h}[n+1]\tilde{\mathbf{y}}^H[l]|\mathbf{C}[n]] &= (\mathbf{B}_u[n+1,n]\otimes\mathbf{I}_N)\\ &\quad\times E[\hat{\mathbf{h}}[n+1|\mathcal{Y}_n]\tilde{\mathbf{y}}_u^H[l]|\mathbf{C}[n]]\\ &= (\mathbf{B}_u[n+1,n]\otimes\mathbf{I}_N)\\ &\quad\times E\left[\sum_{i=-K+1}^{n}\boldsymbol{\Psi}[n+1,i]\tilde{\mathbf{y}}_u[i]\tilde{\mathbf{y}}_u^H[l]\Big|\mathbf{C}[n]\right].\end{aligned} \tag{64}$$

Since the innovation process is temporally white, the right hand side of the above is simply $(\mathbf{B}_u[n+1,n]\otimes\mathbf{I}_N)\boldsymbol{\Psi}[n+1,l]\mathbf{R}_{\tilde{y}\tilde{y}}[l]$. Also, the left hand side can be simplified as

$$\begin{aligned}E[\mathbf{h}[n+1]\tilde{\mathbf{y}}^H[l]|\mathbf{C}[n]] &= \mathtt{P}[n+1-l]E[\mathbf{h}[l]\tilde{\mathbf{y}}_u^H[l]|\mathbf{C}[n]]\\ &\overset{(a)}{=} \mathtt{P}[n+1-l]E[\mathbf{h}[l](\mathbf{C}[l](\bar{\mathbf{B}}_u[l,l-1]\otimes\mathbf{I}_N)\tilde{\mathbf{h}}[l,l-1]\\ &\qquad+\tilde{\mathbf{C}}[l]\mathbf{h}[l]+\sqrt{N_0}\mathbf{w}[l])^H]\\ &\overset{(b)}{=} \mathtt{P}[n+1-l](\bar{\mathbf{B}}_u[l,l-1]\otimes\mathbf{I}_N)E[\tilde{\mathbf{h}}[l,l-1]\tilde{\mathbf{h}}^H[l,l-1]]\\ &\quad\times(\bar{\mathbf{B}}_u[l,l-1]\otimes\mathbf{I}_N)^H\mathbf{C}^H[l]+E[\mathbf{h}[l]\mathbf{h}^H[l]\tilde{\mathbf{C}}^H[l]]\\ &= \mathtt{P}[n+1-l]\bar{\mathbf{B}}_u^2[l,l-1]\mathbf{C}^H[l],\end{aligned} \tag{65}$$

where $\mathtt{P}[l] \triangleq (\mathrm{diag}(\boldsymbol{\rho}[l])\otimes\mathbf{I}_N)$, (a) follows by using the expression for $\tilde{\mathbf{y}}_u[n]$ from (15), and (b) by the orthogonality of $\hat{\mathbf{h}}[l,l-1]$ and $\tilde{\mathbf{h}}[l,l-1]$. Therefore,

$$\begin{aligned}\boldsymbol{\Psi}[n+1,l] &= (\mathbf{B}_u[n+1,n]\otimes\mathbf{I}_N)^{-1}\mathtt{P}[n+1-l]\\ &\quad\times\bar{\mathbf{B}}_u^2[l,l-1]\mathbf{C}^H[l]\mathbf{R}_{\tilde{y}\tilde{y}}^{-1}[l],\end{aligned} \tag{66}$$

and consequently

$$\begin{aligned}\hat{\mathbf{h}}[n+1|\mathcal{Y}_n] &= (\mathbf{B}_u[n+1,n]\otimes\mathbf{I}_N)^{-1}\\ &\quad\times\sum_{l=-K+1}^{n-1}\mathtt{P}[n+1-l]\bar{\mathbf{B}}_u^2[l,l-1]\mathbf{C}^H[l]\mathbf{R}_{\tilde{y}\tilde{y}}^{-1}[l]\tilde{\mathbf{y}}_u[l]\\ &\quad+(\mathbf{B}_u[n,n-1]\otimes\mathbf{I}_N)^{-1}\mathtt{P}[1]\bar{\mathbf{B}}_u^2[n+1,n]\mathbf{C}^H[n]\mathbf{R}_{\tilde{y}\tilde{y}}^{-1}[n]\tilde{\mathbf{y}}_u[n]\\ &= (\mathbf{B}_u[n+1,n]\otimes\mathbf{I}_N)^{-1}\\ &\quad\times\mathtt{P}[1]\sum_{l=-K+1}^{n-1}\mathtt{P}[n-l]\bar{\mathbf{B}}_u^2[l,l-1]\mathbf{C}^H[l]\mathbf{R}_{\tilde{y}\tilde{y}}^{-1}[l]\tilde{\mathbf{y}}_u[l]\\ &\quad+(\mathbf{B}_u[n+1,n]\otimes\mathbf{I}_N)^{-1}\mathtt{P}[1]\bar{\mathbf{B}}_u^2[n+1,n]\mathbf{C}^H[n]\mathbf{R}_{\tilde{y}\tilde{y}}^{-1}[n]\tilde{\mathbf{y}}_u[n]\\ &= (\mathbf{B}_u^{-1}[n+1,n]\mathbf{B}_u[n,n-1]\otimes\mathbf{I}_N)\mathtt{P}[1]\hat{\mathbf{h}}[n|\mathcal{Y}_{n-1}]\\ &\quad+(\mathbf{B}_u[n+1,n]\otimes\mathbf{I}_N)^{-1}\mathtt{P}[1]\bar{\mathbf{B}}_u^2[n+1,n]\mathbf{C}^H[n]\mathbf{R}_{\tilde{y}\tilde{y}}^{-1}[n]\tilde{\mathbf{y}}_u[n],\end{aligned} \tag{67}$$

where the last equation follows because the matrices $(\mathbf{B}_u[n,n-1]\otimes\mathbf{I}_N)$ and $\mathtt{P}[1]$ commute as they are diagonal. This completes the proof.

C. Proof of Theorem 2

The received signal $\mathbf{y}[n]$ can be written in terms of the available channel estimate $\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]$ as

$$\begin{aligned}\mathbf{y}_u[n] &= \hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]\mathbf{B}_u[n,n-1]\mathrm{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}}\odot\mathbf{s}_u[n])\\ &\quad+\tilde{\mathbf{H}}[n|\mathcal{Y}_{n-1}]\bar{\mathbf{B}}_u[n,n-1]\mathrm{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}}\odot\mathbf{s}_u[n])\\ &\quad+\sqrt{N_0}\mathbf{w}[n].\end{aligned} \tag{68}$$

For Gaussian distributed noise and channel estimation error, the MMSE estimator is same as the LMMSE estimator, and consequently,

$$\begin{aligned}\hat{\mathbf{s}}_u[n] &= E[\mathbf{s}_u[n]\mathbf{y}_u^H[n]|\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]]\\ &\quad\times\left(E[\mathbf{y}_u[n]\mathbf{y}_u^H[n]|\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]]\right)^{-1}\mathbf{y}[n].\end{aligned} \tag{69}$$

It is easy to show that,

$$\begin{aligned}&E[\mathbf{s}_u[n]\mathbf{y}_u^H[n]|\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]]\\ &\quad=\mathrm{diag}(\sqrt{\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s}})\mathbf{B}_u[n,n-1]\hat{\mathbf{H}}^H[n|\mathcal{Y}_{n-1}],\end{aligned} \tag{70}$$

and

$$\begin{aligned}&E[\mathbf{y}_u[n]\mathbf{y}_u^H[n]|\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]]\\ &\quad=\left(\sum_{k=1}^{K}\bar{b}_{u,k}^2[n,n-1]+N_0\right)\mathbf{I}_N\\ &\quad+\hat{\mathbf{H}}[n|\mathcal{Y}_{n-1}]\mathrm{diag}(\boldsymbol{\beta}\odot\boldsymbol{\mathcal{E}}_{u,s})\mathbf{B}_u^2[n,n-1]\hat{\mathbf{H}}^H[n|\mathcal{Y}_{n-1}],\end{aligned} \tag{71}$$

where we have used the fact that $\mathbf{B}_u[n,n-1]$ is a diagonal matrix. Substituting (70) and (71) into (69) completes the proof.

D. Proof of Theorem 3

In the Kalman filtering formulation, the estimation error covariance matrix is updated with time as [37]

$$\bar{\mathbf{B}}_u^2[n+1,n] = \mathtt{P}[1]\bar{\mathbf{B}}_u^2[n]\mathtt{P}^H[1]+\mathbf{Q}_1[n], \tag{72}$$

where $\mathbf{Q}_1[n]$ was defined in the paragraph above (14), and $\bar{\mathbf{B}}_u^2[n]$ is the error covariance matrix for the filtered channel state at the nth instant, $\hat{\mathbf{H}}[n|\mathcal{Y}_n] = \mathtt{P}^{-1}[1]\hat{\mathbf{H}}[n+1|\mathcal{Y}_n]$. We can write $\bar{\mathbf{B}}_u^2[n]$ as [37]

$$\bar{\mathbf{B}}_u^2[n] = E[(\mathbf{I}_{KN}-\mathtt{P}[1]\mathbf{G}[n]\mathbf{C}[n])]\bar{\mathbf{B}}_u^2[n,n-1], \tag{73}$$

and can be reduced to (74) on the next page, where the inverse in the $\mathbf{b}_u^{-1}[n,n-1]$ term is an element-wise inverse. The diagonal elements of $\bar{\mathbf{B}}_u^2[n]$ can be written as (75). Substituting this into (72), we obtain (76).

Considering the kth element of the vector $\bar{\mathbf{b}}_u^2[n+1,n]$ completes the proof.

$$\bar{\mathbf{B}}_u^2[n] = \Bigg(\Bigg(\mathbf{I}_K - \frac{E\left[(\boldsymbol{\rho}^2[1] \odot \bar{\mathbf{b}}_u^2[n,n-1] \odot \mathbf{b}_u^{-1}[n,n-1] \odot \sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}} \odot \mathbf{a}_u[n] \odot \hat{\mathbf{s}}_u^*[n])(\sqrt{\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s}} \odot \mathbf{a}_u[n] \odot \hat{\mathbf{s}}_u[n])^T\right]}{\sum_{k=1}^K \beta_k \mathcal{E}_{u,s,k}(\bar{b}_{u,k}^2[n,n-1] + \bar{a}_{u,k}^2[n]) + N_0} \Bigg) \otimes \mathbf{I}_N \Bigg) \bar{\mathbf{B}}_u^2[n, n-1] \tag{74}$$

$$\bar{\mathbf{B}}_u^2[n] = \left(\mathbf{I}_K - \frac{\mathrm{diag}(\boldsymbol{\rho}^2[1] \odot \bar{\mathbf{b}}_u^2[n,n-1] \odot \mathbf{b}_u^{-1}[n,n-1] \odot \boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s} \odot \mathbf{a}_u^2[n])}{\sum_{k=1}^K \beta_k \mathcal{E}_{u,s,k}(\bar{b}_{u,k}^2[n,n-1] + \bar{a}_{u,k}^2[n]) + N_0} \right) \bar{\mathbf{b}}_u^2[n, n-1] \tag{75}$$

$$\bar{\mathbf{b}}_u^2[n+1,n] = \bar{\boldsymbol{\rho}}^2[1] + \boldsymbol{\rho}^2[1] \odot \left(\mathbf{I}_K - \frac{\mathrm{diag}(\boldsymbol{\rho}^2[1] \odot \mathbf{b}_u^{-1}[n,n-1] \odot \bar{\mathbf{b}}_u^2[n,n-1] \odot \boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s} \odot \mathbf{a}_u^2[n])}{\sum_{k=1}^K \beta_k \mathcal{E}_{u,s,k}(\bar{b}_{u,k}^2[n,n-1] + \bar{a}_{u,k}^2[n]) + N_0} \right) \bar{\mathbf{b}}_u^2[n, n-1] \tag{76}$$

E. Proof of Theorem 4

We know that $E[\mathbf{s}_u[n]\mathbf{s}_u^H[n]] = \mathbf{I}_N$. Also, from the definition of $\hat{\mathbf{s}}_u[n]$ in (19), it is easy to show that

$$E[(\mathbf{a}_u[n] \odot \hat{\mathbf{s}}[n])(\hat{\mathbf{s}}^H[n] \odot \mathbf{a}_u^H[n])] = \mathrm{diag}(\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_s \odot \mathbf{b}_u^2[n,n-1]) \times \left(\mathrm{diag}(\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_s \odot \mathbf{b}_u^2[n,n-1]) + \left(\sum_{k=1}^K \bar{b}_{u,k}^2[n,n-1] + N_0 \right) \mathbf{I}_N \right)^{-1}$$

and consequently,

$$E[(\bar{\mathbf{a}}_u[n] \odot \tilde{\mathbf{s}}[n])(\tilde{\mathbf{s}}^H[n] \odot \bar{\mathbf{a}}_u^H[n])] = \left(\sum_{k=1}^K \bar{b}_{u,k}^2[n] + N_0 \right) \times \left(\mathrm{diag}(\boldsymbol{\beta} \odot \boldsymbol{\mathcal{E}}_{u,s} \odot \mathbf{b}_u^2[n,n-1]) + \left(\sum_{k=1}^K \bar{b}_{u,k}^2[n,n-1] + N_0 \right) \mathbf{I}_N \right)^{-1}. \tag{77}$$

We obtain (21) by considering the kth diagonal element of the above matrix.

F. Proof of Theorem 5

We can write the channel estimate in terms of all the past innovation components as

$$\hat{g}_{kk}[n+1|\mathcal{Y}_n] = \sum_{i=0}^n \Psi[n+1,i]\tilde{y}_{d,k}[i], \tag{78}$$

with $\Psi[n+1,i]$ being the gain associated with the innovation component at the ith instant. Invoking the principle of orthogonality, we have $E[(g_{kk}[n+1] - b_{d,k}[n+1,n]\hat{g}_{kk}[n+1|\mathcal{Y}_n])\tilde{y}_{d,k}^*[l]] = 0, 1 \le l \le n$. Given $\mathcal{Y}_n$, $\hat{s}_k[n]$ is known, while $\tilde{s}_k[n]$ is random and unknown. Therefore, we can replace the conditioning on $\mathcal{Y}_n$ with the conditioning on $\hat{s}_k[n]$ to write the above as

$$E[g_{kk}[n+1]\tilde{y}_{d,k}^*[l]|\hat{s}_k[n]] = b_{d,k}[n+1,n]E\left[\sum_{i=0}^n \Psi[n+1,i]\tilde{y}_{d,k}[i]\tilde{y}_{d,k}^*[l]\right]. \tag{79}$$

Since the innovation process is temporally white,

$$E[g_{kk}[n+1]\tilde{y}_{d,k}^*[l]|\hat{s}_{d,k}[n]] = \Psi[n+1,l]b_{d,k}[n+1,n]\sigma_{\tilde{y}\tilde{y},k}^2[l]. \tag{80}$$

Also,

$$E[g_{kk}[n+1]\tilde{y}_{d,k}^*[l]|\hat{s}_{d,k}[n]] = \rho_k[n+1-l]\bar{b}_{d,k}^2[l,l-1]E[|g_{kk}[l]|^2]a_{d,k}[l]\hat{s}_{d,k}^*[l]. \tag{81}$$

Therefore,

$$\Psi[n+1,l] = \frac{\rho_k[n+1-l]\bar{b}_{d,k}^2[l,l-1]E[|g_{kk}[l]|^2]a_{d,k}[l]\hat{s}_{d,k}^*[l]}{b_{d,k}[n+1,n]\sigma_{\tilde{y}\tilde{y},k}^2[l]}, \tag{82}$$

and consequently

$$\hat{g}_{kk}[n+1|\mathcal{Y}_n] = \rho_k[1]\frac{b_{d,k}[n,n-1]}{b_{d,k}[n+1,n]}\hat{g}_{kk}[n|\mathcal{Y}_{n-1}] + \frac{\rho_k[1]\bar{b}_{d,k}^2[n,n-1]E[|g_{kk}[l]|^2]a_{d,k}[n]\hat{s}_{d,k}^*[n]}{b_{d,k}[n+1,n]\sigma_{\tilde{y}\tilde{y},k}^2[n]}\tilde{y}_{d,k}[n], \tag{83}$$

which completes the proof.

G. Proof of Theorem 7

Similar to the uplink case, the mean squared estimation error can be updated with time as [37],

$$\bar{b}_{d,k}^2[n+1,n] = |\rho_k[1]|^2\bar{b}_{d,k}^2[n] + \sigma_1^2[n], \tag{84}$$

where $\bar{b}_{u,k}^2[n]$ is the variance of the filtered state error:

$$\bar{b}_{d,k}^2[n] = E[(1-\rho_k[1]\Gamma_k[n]a_{d,k}[n]\hat{s}_k[n])]\bar{b}_{d,k}^2[n,n-1]. \tag{85}$$

The above can be reduced to

$$\bar{b}_{d,k}^2[n] = \bar{b}_{d,k}^2[n,n-1] \times \left(1 - \frac{\rho_k^2[1]\bar{b}_{d,k}^2[n,n-1]a_{d,k}^2[n]}{b_{d,k}[n,n-1]\sigma_{\tilde{y}\tilde{y},k}^2}\right), \tag{86}$$

and the process noise variance $\sigma_1^2[n]$ can be obtained from (36). Substituting (86) and (36) into (84) yields (45) and completes the proof.

H. Tracking of Spatially Correlated Channels in the Uplink

We now consider the case where the channels from different users to the BS are spatially correlated, with the vector channel of the kth user distributed as $\mathbf{h}_k[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{T}_k)$ at time n, and evolving over time according to (1), with $\mathbf{z}_{h,k}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{T}_k)$. As before, denoting the MMSE channel estimate of $\mathbf{h}_k[n]$ by $\hat{\mathbf{h}}_k[n|\mathcal{Y}_n]$, we can write $\mathbf{h}_k[n]$ as

$$\mathbf{h}_k[n] = \hat{\mathbf{h}}_k[n|\mathcal{Y}_{n-1}] + \tilde{\mathbf{h}}_k[n, n-1], \tag{87}$$

with $E[\hat{\mathbf{h}}_k[n|\mathcal{Y}_{n-1}]\hat{\mathbf{h}}_k^H[n|\mathcal{Y}_{n-1}]] = \mathbf{B}_{u,k}[n, n-1]$, and $E[\tilde{\mathbf{h}}_k[n, n-1]\tilde{\mathbf{h}}_k^H[n, n-1]] = \mathbf{T}_k - \mathbf{B}_{u,k}[n, n-1]$. Note that the notation here is slightly different from (8), with the estimate and estimation error no longer having unit variance. Now, letting $\mathbf{h}[n] = [\mathbf{h}_1^T[n], \ldots, \mathbf{h}_K^T[n]]^T$, we can express $E[\mathbf{h}[n]\mathbf{h}^H[n]] = \mathbf{T}$, where $\mathbf{T}$ is a block diagonal matrix with $\mathbf{T}_k, k = 1, \ldots, K$ as its diagonal elements. We can similarly define the block diagonal matrix $\mathbf{B}_u$ as the covariance matrix of $\hat{\mathbf{h}}[n|\mathcal{Y}_{n-1}]$.

1) State Equations: Invoking the process equation from (12), the process noise covariance matrix is $\mathbf{Q}_1[n] = \text{diag}(\boldsymbol{\rho}^2[1]) \otimes \mathbf{T}$. Similarly, by inspecting (13), the observation noise covariance matrix can be reduced to

$$\mathbf{Q}_2[n] = \sum_{k=1}^{K} \beta_k \mathcal{E}_{u,s,k} \bar{a}_{u,k}[n] \mathbf{T}_k + N_0 \mathbf{I}_N. \tag{88}$$

Based on this, the covariance matrix of the innovation process, as calculated in Lemma 1, becomes

$$\mathbf{R}_{\tilde{y}\tilde{y}}[n] = \sum_{k=1}^{K} \beta_k \mathcal{E}_{u,s,k} (\bar{a}_{u,k}[n] \mathbf{T}_k + (\mathbf{T}_k - \mathbf{B}_k[n, n-1])) + N_0 \mathbf{I}_N. \tag{89}$$

2) Kalman Filtering: The estimate $\hat{\mathbf{h}}[n+1|\mathcal{Y}_n]$ of $\mathbf{h}[n+1]$ can be obtained from $\tilde{\mathbf{y}}_u[n]$ and the available estimate $\hat{\mathbf{h}}[n|\mathcal{Y}_{n-1}]$ as

$$\hat{\mathbf{h}}[n+1|\mathcal{Y}_n] = \mathtt{P}[n]\hat{\mathbf{h}}[n|\mathcal{Y}_{n-1}] + \mathbf{G}[n]\tilde{\mathbf{y}}_u[n] \tag{90}$$

with

$$\mathbf{G}[n] = \mathtt{P}[1](\mathbf{T} - \mathbf{B}_u[n, n-1])\mathbf{C}[n]\mathbf{R}_{\tilde{y}\tilde{y}}^{-1}[n]. \tag{91}$$

Similarly, $\mathbf{B}_u[n+1, n]$ can be updated via equations (72) and (73). These results, in conjunction with Algorithm 1, can be used to track the uplink spatially correlated masssive MIMO channel over time.

REFERENCES

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

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

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

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

[8] 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.

[9] 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.

[10] 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.

[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] 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.

[13] S. Ghosh and R. Chopra, "Training for massive MIMO systems with non-identically aging user channels," *Physical Communication*, vol. 35, p. 100710, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1874490718307122

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] T. Whitworth, M. Ghogho, and D. McLernon, "Optimized training and basis expansion model parameters for doubly-selective channel estimation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1490–1498, Mar. 2009.

[21] J. K. Tugnait, S. He, and H. Kim, "Doubly selective channel estimation using exponential basis models and subblock tracking," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1275–1289, Mar. 2010.

[22] H. Kim and J. K. Tugnait, "Turbo equalization for doubly-selective fading channels using nonlinear Kalman filtering and basis expansion models," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 2076–2087, Jun. 2010.

[23] H. Şenol and C. Tepedelenlioğlu, "Subspace-based estimation of rapidly varying mobile channels for OFDM systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 385–400, 2021.

[24] Y. Liu, S. Zhang, F. Gao, J. Ma, and X. Wang, "Uplink-aided high mobility downlink channel estimation over massive MIMO-OTFS system," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1994–2009, Sep. 2020.

[25] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.

[26] 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.

[27] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.
[28] A. Decurninge, M. Guillaud, and D. T. M. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *Proc. IEEE Globecom 2015 Workshops*, San Diego, CA, Dec. 2015, pp. 1–6.
[29] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.
[30] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *Future Network and Mobile Summit*, Jun. 2010, pp. 1–10.
[31] 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.
[32] 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.
[33] 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.
[34] 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.
[35] 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.
[36] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York: Dover, 1964.
[37] S. Haykin, *Adaptive Filter Theory*. Pearson Education, 2002.
[38] 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.

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) 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. His paper won the best paper award in the Communications Track at NCC 2014 and a paper co-authored with his student won the student best paper award at the IEEE ICASSP 2018. He has 70+ journal papers and 100+ conference papers to his credit. He was an associate editor for the IEEE SIGNAL PROCESSING LETTERS during 2012-16 and for the SADHANA JOURNAL during 2016-18. He is an elected member of the IEEE SPCOM Technical Committee for the years 2014-19. He is a past Chair of the IEEE Signal Processing Society, Bangalore Chapter. He is currently serving as an area editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and as an associate editor for the IEEE TRANSACTIONS ON INFORMATION THEORY and IEEE TRANSACTIONS ON COMMUNICATIONS.