Optimizing Pilot Spacing in MU-MIMO Systems Operating Over Aging Channels

Sebastian Fodor[♭], Gábor Fodor[⋆†], Doğa Gürgünoğlu[†], Miklós Telek[‡♯]
[♭]Stockholm University, Stockholm, Sweden. E-mail: sebbifodor@fastmail.com
[⋆]Ericsson Research, Stockholm, Sweden. E-mail: Gabor.Fodor@ericsson.com
[†]KTH Royal Institute of Technology, Stockholm, Sweden. E-mail: gaborf|dogag@kth.se
[‡]Budapest University of Technology and Economics, Budapest, Hungary. E-mail: telek@hit.bme.hu
[♯]ELKH-BME Information Systems Research Group, Budapest, Hungary. E-mail: telek@hit.bme.hu

***Abstract*—In the uplink of multiuser multiple input multiple output (MU-MIMO) systems operating over aging channels, pilot spacing is crucial for acquiring channel state information and achieving high signal-to-interference-plus-noise ratio (SINR). Somewhat surprisingly, very few works examine the impact of pilot spacing on the correlation structure of subsequent channel estimates and the resulting quality of channel state information considering channel aging. In this paper, we consider a fast-fading environment characterized by its exponentially decaying autocorrelation function, and model pilot spacing as a sampling problem to capture the inherent trade-off between the quality of channel state information and the number of symbols available for information carrying data symbols. We first establish a quasi-closed form for the achievable deterministic equivalent SINR when the channel estimation algorithm utilizes multiple pilot signals. Next, we establish upper bounds on the achievable SINR and spectral efficiency, as a function of pilot spacing, which helps to find the optimum pilot spacing within a limited search space. Our key insight is that to maximize the achievable SINR and the spectral efficiency of MU-MIMO systems, proper pilot spacing must be applied to control the impact of the aging channel and to tune the trade-off between pilot and data symbols.**

***Index terms*— autoregressive processes, channel estimation, estimation theory, multiple input multiple output, receiver design**

I. Introduction

In wireless communications, pilot symbol-assisted channel estimation and prediction are used to achieve reliable coherent reception, and thereby to provide a variety of high quality services in a spectrum efficient manner. In most practical systems, the transmitter and receiver nodes acquire and predict channel state information by employing predefined pilot sequences during the training phase, after which information symbols can be appropriately modulated and precoded at the transmitter and estimated at the receiver. Since the elapsed time between pilot transmissions and the transmit power level of pilot symbols have a large impact on the quality of channel estimation, a large number of papers investigated the optimal spacing and power control of pilot signals in both single and multiple antenna systems [1]–[12].

Specifically in the uplink of multiuser multiple input multiple output (MU-MIMO) systems, several papers proposed pilot-based channel estimation and receiver algorithms assuming that the complex vector channel undergoes block fading, meaning that the channel is constant between two subsequent channel estimation instances [13]–[16]. In the block fading model, the evolution of the channel is memoryless in the sense that each channel realization is drawn independently of previous channel instances from some characteristic distribution. While the block fading model is useful for obtaining analytical expressions for the achievable signal-to-interference-plus-noise ratio (SINR) and capacity [15], [17], it fails to capture the correlation between subsequent channel realizations and the aging of the channel between estimation instances [6], [7], [11], [12].

Due to the importance of capturing the evolution of the wireless channel in time, several papers developed time-varying channel models, as an alternative to block fading models, whose states are advantageously estimated and predicted by means of suitably spaced pilot signals. In particular, a large number of related works assume that the wireless channel can be represented as an autoregressive (AR) process whose states are estimated and predicted using Kalman filters, which exploit the correlation between subsequent channel realizations [3], [4], [6], [10], [12]. These papers assume that the coefficients of the related AR process are known, and the current and future states of the process (and thereby of the wireless channel) can be well estimated. Other important related works concentrate on estimating the coefficients of AR processes based on suitable pilot-based observations and measurements [18]–[20]. In our recent work [12], it was shown that when an AR process is a good model of the wireless channel and the AR coefficients are well estimated, not only the channel estimation can exploit the memoryful property of the channel, but also a new MU-MIMO receiver can be designed, which minimizes the mean squared error (MSE) of the received data symbols by exploiting the correlation between subsequent channel states. It is important to realize that the above references build on discrete time AR models, in which the state transition matrix is an input of the model and can be estimated by some suitable system identification technique, such as the one proposed in [20]. However, these papers do not ask the question of how often the channel state of an aging channel should be observed by suitably spaced pilot signals to realize a certain state transition matrix in the AR model of the channel.

G. Fodor and D. Gürgünoğlu were partially supported by the H2020 Marie Skłodowska-Curie Innovative Training Networks, GA Number: 956256. M. Telek is partially supported by the Hungarian Scientific Research Fund OTKA K-138208 project.

Specifically, a key characteristic of a continuous time Rayleigh fading environment is that the autocorrelation function of the associated stochastic process is a zeroth-order Bessel function, which must be properly modelled [21], [22]. This requirement is problematic when developing discrete-time AR models, since it is well-known that Rayleigh fading cannot be perfectly modelled with any finite order AR process (since the autocorrelation function of discrete time AR processes does not follow a Bessel function), although the statistics of an AR process can approximate those of Rayleigh fading [23], [24].

Recognizing the importance of modeling fast fading, including Rayleigh fading, channels with proper autocorrelation function as a basis for pilot spacing optimization, papers [25], [26] use a continuous time process as a representation of the wireless channel, and address the problem of pilot spacing as a sampling problem. According to this approach, pilot placement can be considered as a sampling problem of the fading variations, and the quality of the channel estimate is determined by the density and accuracy of channel sampling [26]. However, these papers consider single input single output (SISO) systems, and are not applicable to MU-MIMO systems employing a minimum mean squared error (MMSE) receiver, which was proposed in, for example, [12]. On the other hand, paper [6] analyzes the impact of channel aging on the performance of multiple input multiple output (MIMO) systems, without investigating the interplay between pilot spacing and the resulting state transition matrix of the AR model of the fast fading channel. Since that paper proposes three important channel estimation and prediction schemes and establishes closed forms for the deterministic equivalent SINR [6], it serves as the benchmark for the pilot spacing algorithm proposed in the present paper.

In this paper, we are interested in determining the average SINR in the uplink of MU-MIMO systems operating in fast fading as a function of pilot spacing, pilot/data power allocation, number of antennas and spatially multiplexed users. Specifically, we ask the following two important questions, which are not answered by previous works:

- What is the average SINR in a closed or quasi-closed form in the uplink of MU-MIMO systems in fast fading in the presence of antenna correlation? How does the average SINR depend on pilot spacing and pilot/data power control?
- What is the optimum pilot spacing and pilot/data power allocation as a function of the number of antennas and the Doppler frequency associated with the continuous time – modeled as a piece-wise constant – fast fading channel?

In the light of the above discussion and questions, the main contributions of the present paper, which are provided in Sections IV and V, are as follows:

- Theorem 1 and Proposition 2 establish an upper bound on the achievable SINR as a function of pilot spacing, which is instrumental for determining the optimum pilot spacing.
- Proposition 3, building on Proposition 2, provides an upper bound on the average achievable spectral efficiency, which is instrumental in limiting the search space for the optimal frame size as a function of the Doppler frequency.

In addition, we believe that the engineering insights drawn from the numerical studies are useful when designing pilot spacing, for example in the form of determining the number of reference signals in an uplink frame structure, for MU-MIMO systems.

Specifically, to answer the above questions, we proceed as follows. In the next section, we present our system model, which admits correlated wireless channels between any of the single-antenna mobile terminal and the receive antennas of the base station (BS). Next, in Section III, we apply well-known results developed for block fading channels in [15], [27]–[30] to the system model developed in Section II. Section IV studies the impact of pilot spacing on the achievable SINR and the spectral efficiency (SE) of all users in the system, and establishes an upper bound on this SINR. We show that this upper bound is monotonically decreasing as the function of pilot spacing. This property is very useful, because it enables to limit the search space of the possible pilot spacings when looking for the optimum pilot spacing in Section V. That section also considers the special case when the channel coefficients associated with the different receive antennas are uncorrelated and identically distributed. It turns out that in this special case a simplified SINR expression can be derived. Section VI presents numerical results and discusses engineering insights. Finally, Section VII draws conclusions.

In this paper we use the notation $[\mathbf{v}]_n = \mathbf{w}_n$ and $[\mathbf{A}]_{n,m} = \mathbf{B}_{n,m}$ to denote the elements of the block vectors and block matrices $\mathbf{v} = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_N \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} \mathbf{B}_{1,1} & \dots & \mathbf{B}_{1,M} \\ \vdots & \ddots & \vdots \\ \mathbf{B}_{N,1} & \dots & \mathbf{B}_{N,M} \end{bmatrix}$, respectively.

II. System Model

A. Uplink Pilot Signal Model

By extending the single antenna channel model of [25], each transmitting mobile station (MS) uses a single time slot to send F pilot symbols, followed by Δ time slots, each of which containing F data symbols according to Figure 1. Each symbol is transmitted within a coherent time slot of duration T. Thus, the total frame duration is $(1+\Delta)T$, such that each frame consists of 1 pilot and Δ data time slots, which we will index with $i = 1 \dots \Delta$. User-k transmits each of the F pilot symbols with transmit power $P_{p,k}$, and each data symbol in slot-i with transmit power $P_k(i),\ k = 1 \dots K$. To simplify notation, in the sequel we tag User-1, and will drop index $k = 1$ when referring to the tagged user.

Assuming that the coherence bandwidth accommodates at least F pilot symbols, this system allows to create F orthogonal pilot sequences. To facilitate spatial multiplexing and channel state information at the receiver (CSIR) acquisition at the BS, the MSs use orthogonal complex sequences, such as shifted Zadoff-Chu sequences of length $\tau_p = F$, which we denote as:

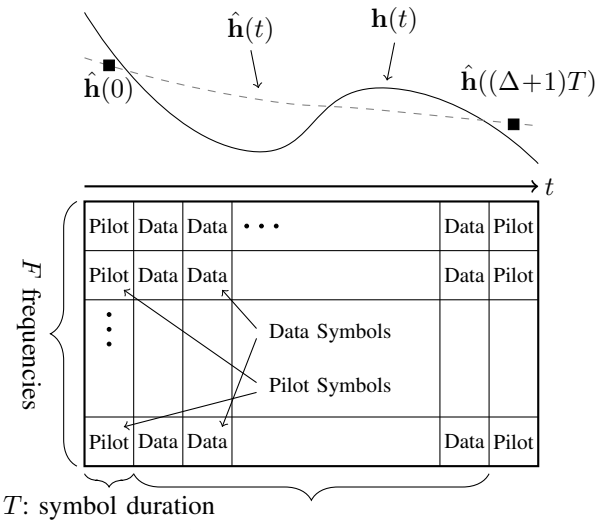

Figure 1. Pilot (P) and data (D) symbols in the time-frequency domains of the system in the $(0,(\Delta+1)T)$ interval. The solid line above the time-frequency resource grid represents the piece-wise constant complex channel $\mathbf{h}(t)$, while the dashed line represents the MMSE channel estimate $\hat{\mathbf{h}}(t)$. Notice that in each time slot of length T all symbols are either pilot or data symbols. This arrangement is used by the block type pilot allocation in 4G and 5G systems.

$$\mathbf{s} \triangleq \left[s_1, ..., s_{\tau_p}\right]^T \in \mathbb{C}^{\tau_p \times 1}, \tag{1}$$

whose elements satisfy $|s_i|^2 = 1$. Under this assumption, the system can spatially multiplex $K \leq F$ MSs. Focusing on the received pilot signal from the tagged user at the BS, the received pilot signal takes the form of [12]:

$$\mathbf{Y}(t) = \alpha\sqrt{P_p}\mathbf{h}(t)\mathbf{s}^T + \mathbf{N}(t) \quad \in \mathbb{C}^{N_r \times \tau_p}, \tag{2}$$

where $\mathbf{h}(t) \in \mathbb{C}^{N_r \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, that is, $\mathbf{h}(t)$ is a complex normal distributed column vector with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C} \in \mathbb{C}^{N_r \times N_r}$. Furthermore, α denotes large scale fading, P_p denotes the pilot power of the tagged user, and $\mathbf{N}(t) \in \mathbb{C}^{N_r \times \tau_p}$ is the additive white Gaussian noise (AWGN) with element-wise variance σ_p^2.

It will be convenient to introduce $\tilde{\mathbf{Y}}(t)$ by stacking the columns of $\mathbf{Y}(t)$ as:

$$\tilde{\mathbf{Y}}(t) = \mathbf{vec}\big(\mathbf{Y}(t)\big) = \alpha\sqrt{P_p}\mathbf{S}\mathbf{h}(t) + \tilde{\mathbf{N}}(t) \ \in \mathbb{C}^{\tau_p N_r \times 1}, \tag{3}$$

where **vec** is the column stacking vector operator, $\tilde{\mathbf{Y}}(t)$, $\tilde{\mathbf{N}}(t) \in \mathbb{C}^{\tau_p N_r \times 1}$ and $\mathbf{S} \triangleq \mathbf{s} \otimes \mathbf{I}_{N_r} \in \mathbb{C}^{\tau_p N_r \times N_r}$ is such that $\mathbf{S}^H\mathbf{S} = \tau_p \mathbf{I}_{N_r}$, where $\mathbf{I}_{N_r}$ is the identity matrix of size N_r.

B. Channel Model

In (2), the channel $\mathbf{h}(t)$ evolves according to a multivariate complex stochastic process with stationary covariance matrix $\mathbf{C}$. That is, for symbol duration T, the channel $(\mathbf{h}(t))$ evolves according to the following AR process:

$$\mathbf{h}(t+T) = \mathbf{A}\mathbf{h}(t) + \boldsymbol{\vartheta}(t), \tag{4}$$

where the transition matrix of the AR process is denoted by $\mathbf{A}$, and $\boldsymbol{\vartheta}(t) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Theta})$ denotes the random process noise vector with zero mean and $\boldsymbol{\Theta}$ covariance matrix. This AR model has been commonly used to approximate Rayleigh fading channels in e.g. [31]. Equation (4) implies that the autocorrelation function of the channel process is:

$$\mathbb{E}\left(\mathbf{h}(t)\mathbf{h}^H(t+iT)\right) = \mathbf{C}\left(\mathbf{A}^H\right)^i, \quad \forall i. \tag{5}$$

Consequently, the autocorrelation function of the fast fading channel $(\mathbf{h}(t))$ is modelled as:

$$\mathbf{R}(i) \triangleq \mathbb{E}\left(\mathbf{h}(t)\mathbf{h}^H(t+iT)\right) = \begin{cases} \mathbf{C}e^{\mathbf{Q}^H iT} & \text{if } i \geq 0, \\ \mathbf{C}e^{-\mathbf{Q}iT} & \text{if } i < 0, \end{cases} \tag{6}$$

where matrix $\mathbf{Q}$ describes the correlation decay, such that: $e^{\mathbf{Q}T} = \mathbf{A}$. From (6), we have $\mathbf{R}(i) = \mathbf{R}^H(-i)$. Similarly, for user k,

$$\mathbf{R}_k(i) \triangleq \mathbb{E}\left(\mathbf{h}_k(t)\mathbf{h}_k^H(t+iT)\right) = \begin{cases} \mathbf{C}_k e^{\mathbf{Q}_k^H iT} & \text{if } i \geq 0, \\ \mathbf{C}_k e^{-\mathbf{Q}_k iT} & \text{if } i < 0, \end{cases} \tag{7}$$

In each pilot slot, the BS utilizes MMSE channel estimation to obtain the channel estimate of each user, as it will be detailed in Section III. Without loss of generality, to simplify the notation, hereafter we assume that the time unit is T and $iT = i$.

C. Data Signal Model

When spatially multiplexing K MU-MIMO users, the received data signal at the BS at time t is [12]:

$$\mathbf{y}(t) = \underbrace{\alpha\mathbf{h}(t)\sqrt{P}x(t)}_{\text{tagged user}} + \underbrace{\sum_{k=2}^{K}\alpha_k\mathbf{h}_k(t)\sqrt{P_k}x_k(t)}_{\text{co-scheduled MU-MIMO users}} + \mathbf{n}_d(t), \tag{8}$$

where $\mathbf{y}(t) \in \mathbb{C}^{N_r \times 1}$; and $x_k(t)$ denotes the transmitted data symbol of User-k at time t with transmit power P_k. Furthermore, $\mathbf{n}_d(t) \sim \mathcal{CN}\left(\mathbf{0}, \sigma_d^2\mathbf{I}_{N_r}\right)$ is the AWGN at the receiver. Since each slot contains either only pilot or data symbols, the pilot and data power levels can be set independently.

III. Overview of Previous Results and Prerequisites

In this section, we are interested in

- calculating the MMSE estimation of the channel in each data slot $1 \leq i \leq \Delta$, based on received pilot signals, as a function of the frame size corresponding to pilot spacing (see Δ in Figure 1), and
- determining the slot-by-slot SINR of a tagged user $(\bar{\gamma}(i))$ of a MU-MIMO system operating over fast fading channels modelled as AR processes.

To this end, we apply well-known results developed for block fading channels in [15], [27]–[30] to the system model developed in the previous section.

Estimating the channel at the receiver can be based on multiple received pilot signals both before and after the actual data slot i. While using pilot signals that are received before data slot i requires to store the samples of the received

pilot, using pilot signals that arrive after data slot i requires to store the received data signals and necessarily induces some delay in estimating the transmitted data symbol. In the sequel we use the general case of "p before, q after" to illustrate the operation of the MMSE channel estimation scheme, that is when the receiver uses the p pilot signals $\tilde{\mathbf{Y}}(0)$, $\tilde{\mathbf{Y}}\big(-(\Delta+1)\big)\dots\tilde{\mathbf{Y}}\big(-(p-1)(\Delta+1)\big)$ before the data slot and the q pilot signals $\tilde{\mathbf{Y}}(\Delta+1)$, $\tilde{\mathbf{Y}}\big(2(\Delta+1)\big)\dots\tilde{\mathbf{Y}}\big(q(\Delta+1)\big)$ after the data slot for CSIR acquisition. We are also interested in determining the distribution of the resulting channel estimation error, whose covariance matrix, denoted by $\mathbf{Z}(\Delta,i)$, will play an important role in subsequently determining the deterministic equivalent of the SINR in each slot i.

A. MMSE Channel Estimation and Channel Estimation Error

In each data slot i, the BS utilizes the pilot signals obtained in the pilot slots. The BS waits for q pilot slots to occur before calculating this estimation, and uses the pilot signals from these later pilot slots and p previous pilot slots to estimate the channel during the data slot. Since there are Δ data slots between two pilot slots, the utilized pilot signals are $\tilde{\mathbf{Y}}(n(\Delta+1))$ where n runs from $-(p-1)$ to q.

Lemma 1. *The* MMSE *channel estimator approximates the autoregressive fast fading channel in time slot i based on the received pilots at $n(\Delta+1)$, $n=-(p-1),\dots,q$ as*

$$\hat{\mathbf{h}}_{\text{MMSE}}(\Delta,i)=\mathbf{H}^{\star}(\Delta,i)\hat{\mathbf{Y}}(\Delta), \tag{9}$$

where

$$\begin{aligned}
\mathbf{H}^{\star}(\Delta,i)&=\frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbf{E}(\Delta,i)\Big(\mathbf{M}(\Delta)+\mathbf{\Sigma}_{p+q}\Big)^{-1}\\
&\quad\cdot\Big(\mathbf{s}^H\otimes\mathbf{I}_{(p+q)N_r}\Big),\\
\Big[\hat{\mathbf{Y}}(\Delta)\Big]_n&\triangleq\tilde{\mathbf{Y}}\big(n(\Delta+1)\big),\\
\mathbf{\Sigma}_{p+q}&\triangleq\frac{\sigma_p^2}{\alpha^2P_p\tau_p}\mathbf{I}_{(p+q)N_r},\\
\Big[\mathbf{E}(\Delta,i)^T\Big]_m&\triangleq\mathbf{R}\big(m(\Delta+1)-i\big), \qquad (10)\\
\Big[\mathbf{M}(\Delta)\Big]_{n,m}&\triangleq\mathbf{R}\big((m-n)(\Delta+1)\big), \qquad (11)
\end{aligned}$$

where n and m run from $-(p-1)$ to q.

Proof. The lemma can be established using standard techniques for deriving the MMSE estimator [12], [27], and rewriting $\hat{\mathbf{Y}}(\Delta)$ as $\hat{\mathbf{Y}}(\Delta)=(\mathbf{I}_{p+q}\otimes\mathbf{S})\bar{\mathbf{h}}(\Delta)+\tilde{\bar{\mathbf{N}}}$, where $\Big[\bar{\mathbf{h}}(\Delta)\Big]_n\triangleq\mathbf{h}\big(n(\Delta+1)\big)$ and $\Big[\tilde{\bar{\mathbf{N}}}(\Delta)\Big]_n\triangleq\tilde{\mathbf{N}}\big(n(\Delta+1)\big)$. □

From Lemma 1, it follows that the MMSE estimate of the channel is expressed as:

$$\begin{aligned}
\hat{\mathbf{h}}_{\text{MMSE}}(\Delta,i)&=\mathbf{H}^{\star}(\Delta,i)\hat{\mathbf{Y}}(\Delta)\\
&=\mathbf{H}^{\star}(\Delta,i)\left(\alpha\sqrt{P_p}(\mathbf{I}_{p+q}\otimes\mathbf{S})\bar{\mathbf{h}}(\Delta)+\tilde{\bar{\mathbf{N}}}(\Delta)\right)\\
&=\frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbf{E}(\Delta,i)\left(\mathbf{M}(\Delta)+\mathbf{\Sigma}_{p+q}\right)^{-1}\\
&\quad\cdot\left(\alpha\sqrt{P_p}\tau_p\bar{\mathbf{h}}(\Delta)+\left(\mathbf{I}_{p+q}\otimes\mathbf{S}^H\right)\tilde{\bar{\mathbf{N}}}(\Delta)\right). \qquad (12)
\end{aligned}$$

Next, we are interested in deriving the distribution of the estimated channel and the channel estimation error, since these will be important for understanding the impact of pilot spacing on the achievable SINR and spectral efficiency of the MU-MIMO system. To this end, the following two corollaries of Lemma 1 – which follow directly from the Lemma and (12) – will be important in the sequel.

Corollary 1. *The estimated channel $\hat{\mathbf{h}}_{\text{MMSE}}(\Delta,i)$ is a circular symmetric complex normal distributed vector $\hat{\mathbf{h}}_{\text{MMSE}}(\Delta,i)\sim\mathcal{CN}\Big(\mathbf{0},\hat{\mathbf{\Phi}}_{\text{MMSE}}(\Delta,i)\Big)$, with*

$$\begin{aligned}
\hat{\mathbf{\Phi}}_{\text{MMSE}}(\Delta,i)&\triangleq\mathbb{E}_{\mathbf{h},\mathbf{n}}\{\hat{\mathbf{h}}_{\text{MMSE}}(\Delta,i)\hat{\mathbf{h}}^H_{\text{MMSE}}(\Delta,i)\}\\
&=\mathbf{E}(\Delta,i)\big(\mathbf{M}(\Delta)+\mathbf{\Sigma}_{p+q}\big)^{-1}\mathbf{E}^H(\Delta,i). \qquad (13)
\end{aligned}$$

An immediate consequence of Corollary 1 is the following corollary regarding the covariance of the channel estimation error, as a function of pilot spacing.

Corollary 2. *The channel estimation error in slot i, $\hat{\mathbf{h}}_{\text{MMSE}}(\Delta,i)-\mathbf{h}(\Delta,i)$, is complex normal distributed with zero mean vector and covariance matrix given by:*

$$\mathbf{Z}(\Delta,i)\triangleq\mathbf{C}-\mathbf{E}(\Delta,i)\big(\mathbf{M}(\Delta)+\mathbf{\Sigma}_{p+q}\big)^{-1}\mathbf{E}^H(\Delta,i). \tag{14}$$

In the following section we will calculate the SINR of the received data symbols. For simplicity of notation, we use $\hat{\mathbf{h}}_{\text{MMSE}}(\Delta,i)=\hat{\mathbf{h}}(\Delta,i)$, and introduce

$$\mathbf{b}(\Delta,i)\triangleq\alpha\sqrt{P(i)}\hat{\mathbf{h}}(\Delta,i), \tag{15}$$

with covariance matrix

$$\begin{aligned}
\mathbf{\Phi}(\Delta,i)&\triangleq\mathbb{E}\left(\mathbf{b}(\Delta,i)\mathbf{b}^H(\Delta,i)\right)\\
&=\alpha^2P(i)(\mathbf{C}-\mathbf{Z}(\Delta,i)). \qquad (16)
\end{aligned}$$

To summarize, this subsection derived the MMSE channel estimator (Lemma 1) that uses the received pilot signals both before and after a given data slot i and depends on the frame size Δ (pilot spacing). As important corollaries of the channel estimation scheme, we established the distribution of both the estimated channel (Corollary 1) and the associated channel estimation error in each data slot i (Corollary 2), as functions of both the employed pilot spacing and pilot power. These results serve as a starting point for deriving the achievable SINR and spectral efficiency in the next subsection.

B. SINR Calculation

We start with recalling an important lemma from [30], which calculates the per-slot SINR in an AR fast fading environment when the BS uses the MMSE estimation of the fading channel, and employs the optimal linear receiver:

$$\mathbf{G}^{\star}(\Delta,i)=\mathbf{b}^H(\Delta,i)\mathbf{J}^{-1}(\Delta,i), \tag{17}$$

where $\mathbf{J}(\Delta,i)\in\mathbb{C}^{N_r\times N_r}$ is defined as

$$\mathbf{J}(\Delta,i)\triangleq\sum_{k=1}^{K}\mathbf{b}_k(\Delta,i)\mathbf{b}_k^H(\Delta,i)+\boldsymbol{\beta}(\Delta,i),$$

where

$$\boldsymbol{\beta}(\Delta,i) \triangleq \sum_{k=1}^{K} \alpha_k^2 P_k \mathbf{Z}_k(\Delta,i) + \sigma_d^2 \mathbf{I}_{N_r}. \tag{18}$$

When using the above receiver, which minimizes the MSE of the received data symbols in the presence of channel estimation errors, the following result from [30] will be useful in the sequel:

Lemma 2 (See [30], Lemma 3)**.** *Assume that the receiver employs* MMSE *symbol estimation, that is it employs the optimal linear receiver* $\mathbf{G}^{\star}(\Delta,i)$ *given in* (17)*. Then the per-slot* SINR *of the estimated data symbols of the tagged user,* $\gamma(\Delta,i)$ *is given as:*

$$\gamma(\Delta,i) = \mathbf{b}^H(\Delta,i)\bar{\mathbf{J}}^{-1}(\Delta,i)\mathbf{b}(\Delta,i), \tag{19}$$

where

$$\bar{\mathbf{J}}(\Delta,i) \triangleq \mathbf{J}(\Delta,i) - \mathbf{b}(\Delta,i)\mathbf{b}^H(\Delta,i). \tag{20}$$

For the AR fading case considered in this paper, based on the definitions of $\mathbf{b}(\Delta,i)$, $\mathbf{J}(\Delta,i)$ and $\bar{\mathbf{J}}(\Delta,i)$, the per-slot SINR of the tagged user is then expressed as:

$$\begin{aligned}\gamma(i) &= \mathbf{b}^H(\Delta,i)\bar{\mathbf{J}}^{-1}(\Delta,i)\mathbf{b}(\Delta,i)\\ &= \mathrm{tr}\left(\mathbf{b}(\Delta,i)\mathbf{b}^H(\Delta,i)\bar{\mathbf{J}}^{-1}(\Delta,i)\right).\end{aligned} \tag{21}$$

C. Slot-by-Slot Deterministic Equivalent of the SINR as a Function of Pilot Spacing Δ

We can now state the following important proposition that gives the deterministic equivalent of the SINR in data slot i, $\bar{\gamma}(\Delta,i)$, when the number of antennas N_r approaches infinity. This deterministic equivalent SINR gives a good approximation of averaging the per-slot SINR of the tagged user, and can be proved by invoking [28, Theorem 1] or [15, Theorem 1].

Proposition 1. *Assuming,* $1 \leq \lim_{N_r,K\to\infty} N_r/K \leq \infty$ *and* $\mathbf{C}_k/N_r, \mathbf{Z}_k(\Delta,i)/N_r$ *have uniformly bounded spectral norms, the deterministic equivalent SINR of the tagged user in data slot* i *can be calculated as:*

$$\bar{\gamma}(\Delta,i) = \mathrm{tr}\Big(\boldsymbol{\Phi}(\Delta,i)\mathbf{T}(\Delta,i)\Big), \tag{22}$$

where $\mathbf{T}(\Delta,i)$ *is defined as:*

$$\mathbf{T}(\Delta,i) \triangleq \left(\sum_{m=2}^{K} \frac{\boldsymbol{\Phi}_m(\Delta,i)}{1+\delta_m(\Delta,i)} + \boldsymbol{\beta}(\Delta,i)\right)^{-1}, \tag{23}$$

and $\delta_m(\Delta,i)$ *are the solutions of the following system of* K *equations*

$$\delta_m(\Delta,i) = \mathrm{tr}\left(\boldsymbol{\Phi}_m(\Delta,i)\left(\sum_{l=2}^{K} \frac{\boldsymbol{\Phi}_l(\Delta,i)}{1+\delta_l(\Delta,i)} + \boldsymbol{\beta}(\Delta,i)\right)^{-1}\right) \tag{24}$$

for $\forall m = 1,\ldots,K$.

The above system of K equations gives the deterministic equivalent of the SINR of the tagged user, and a different set of K equations must be used for each user.

To summarize, this section established the slot-by-slot SINR of a tagged user ($\bar{\gamma}(i)$) of a MU-MIMO system operating over a fast fading channels modelled as AR processes, by applying our previous result obtained for discrete-time AR channels reported in [12]. Next, we invoked [28, Theorem 1], to establish the deterministic equivalent SINR for each slot, as a function of the frame size (pilot spacing) Δ, see Proposition 1. These results serve as a basis for formulating the pilot spacing optimization problem over the frame size and pilot power as optimization variables.

IV. Impact of Pilot Spacing on the SINR and Spectral Efficiency

In this section, we study the impact of pilot spacing on the achievable SINR and the SE of all users in the system. The approximate SE, based on the deterministic equivalent SINR, associated with the i-th data symbol of user k is

$$\mathrm{SE}_k(\Delta,i) \triangleq \log\Big(1+\bar{\gamma}_k(\Delta,i)\Big), \tag{25}$$

where $\bar{\gamma}_k(\Delta,i)$ denotes the average SINR of user k when sending the i-th data symbol, and when Δ data symbols are sent between every pair of pilot symbols. Consequently, the average SE of user k over the $(\Delta+1)$ slot long frame is

$$\frac{\sum_{i=1}^{\Delta} \mathrm{SE}_k(\Delta,i)}{\Delta+1}, \tag{26}$$

which can be optimized over Δ. More importantly, the aggregate average SE of the MU-MIMO system for the K users can be expressed as:

$$\mathrm{SE}(\Delta) = \frac{\sum_{k=1}^{K}\sum_{i=1}^{\Delta} \mathrm{SE}_k(\Delta,i)}{\Delta+1}. \tag{27}$$

A. An Upper Bound of the Deterministic Equivalent SINR and the SE

Let us assume that $\mathbf{Q}_k = q_k\mathbf{I}_{N_r}$, that is the channel vector $\mathbf{h}_k(t)$ consists of independent AR processes in the spatial domain, implying that:

$$\mathbf{R}_k(i) \triangleq \mathbb{E}\left(\mathbf{h}_k(t)\mathbf{h}_k^H(t+i)\right) = \mathbf{C}_k e^{q_k^* i}, \tag{28}$$

where q_k is a scalar, q_k^* denotes complex conjugation, and let $\bar{q}_k \triangleq Re(q_k) < 0$.

Note that the exponential approximation of the autocorrelation function of the fast fading process expressed in (28) is related to the Doppler frequency of Rayleigh fading through:

$$\underbrace{\mathbf{C}J_0(2\pi f_D i)}_{\text{True autocorrelation of Rayleigh fading}} \approx \mathbf{R}(i), \tag{29}$$

where $J_0(.)$ is the zeroth order Bessel function [32]. Based on the exponential approximation of this Rayleigh fading process in (28), the Doppler frequency of the approximate model is obtained from $2\pi f_D i = Re(q_k^* i)$, i.e. $f_D = 2\pi/\bar{q}_k$.

To optimize (27), we first find an upper bound of $\mathrm{SE}_k(\Delta,i)$ via an upper bound of $\bar{\gamma}_k(\Delta,i)$. To simplify the notation, the following discussion refers to the tagged user, and later we

utilize that the same relations hold for all users. We introduce the following upper bound of $\bar{\gamma}(\Delta,i)$:

$$\bar{\gamma}^{(u)}(\Delta,i) \triangleq \mathrm{tr}\left(\mathbf{\Phi}^{(u)}(\Delta,i)\left(\sum_{l=1}^{K}\alpha_l^2 P_l \mathbf{Z}_l^{(u)}(\Delta,i)+\sigma_d^2\mathbf{I}_{\mathbf{N_r}}\right)^{-1}\right), \tag{30}$$

where $\mathbf{Z}^{(u)}(\Delta,i)$ and $\mathbf{\Phi}^{(u)}(\Delta,i)$ are given by

$$\mathbf{Z}^{(u)}(\Delta,i) \triangleq \mathbf{C} - \rho(\Delta,i)\mathbf{C}\left(\eta\mathbf{C}+\mathbf{\Sigma}\right)^{-1}\mathbf{C}, \tag{31}$$

$$\mathbf{\Phi}^{(u)}(\Delta,i) \triangleq \alpha^2 P \rho(\Delta,i)\mathbf{C}\left(\eta\mathbf{C}+\mathbf{\Sigma}\right)^{-1}\mathbf{C}, \tag{32}$$

with η being a properly set constant (see Theorem 1), $\mathbf{\Sigma} \triangleq \frac{\sigma_p^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r}$, and

$$\rho(\Delta,i) \triangleq \sum_{\ell=-(p-1)}^{q} e^{2\bar{q}|i-\ell(\Delta+1)|}. \tag{33}$$

As mentioned, for $\bar{\gamma}^{(u)}$ defined in (30) to be a suitable upper bound of $\bar{\gamma}$, we need to set the constant η in (31) and (32) properly. To this end, the following theorem is helpful.

Theorem 1. *If $\bar{q} < 0$, and*

$$\begin{aligned}
&p+q=1 \text{ and } 0<\eta<1, \text{ or}\\
&p+q=2 \text{ and } 0<\eta<1-a, \text{ or}\\
&p+q=3 \text{ and } 0<\eta<1-\tfrac{a(\sqrt{8+a^2}-a)}{2}, \text{ or}\\
&p+q=4 \text{ and } 0<\eta<1-\tfrac{a(a^2+1+(a-1)\sqrt{5-2a+a^2})}{2},
\end{aligned} \tag{34}$$

with $a \triangleq e^{2\bar{q}(\Delta+1)}$ then $\bar{\gamma}(\Delta,i) \le \bar{\gamma}^{(u)}(\Delta,i)$.

Proof. We prove the theorem based on the following inequalities

$$\begin{aligned}
\bar{\gamma}(\Delta,i) &\overset{(a)}{\le} \mathrm{tr}\left(\mathbf{\Phi}(\Delta,i)\boldsymbol{\beta}(\Delta,i)^{-1}\right)\\
&\overset{(b)}{\le} \mathrm{tr}\left(\mathbf{\Phi}^{(u)}(\Delta,i)\boldsymbol{\beta}(\Delta,i)^{-1}\right) \overset{(c)}{\le} \bar{\gamma}^{(u)}(\Delta,i),
\end{aligned} \tag{35}$$

which are proved in consecutive lemmas, that is Lemmas 3-6 below. □

Remark 1. *While Theorem 1 is not possible to generalize symbolically for cases when $p+q > 4$, by numerical experiments we found that when $0 < \eta < L(a)$, and $L(a) = c_3a^3 + c_2a^2 - c_1a + c_0$ with $c_3 = -0.329$, $c_2 = 1.154$, $c_1 = -1.810$ and $c_0 = 0.985$, then $\bar{\gamma}(\Delta,i) \le \bar{\gamma}^{(u)}(\Delta,i)$ holds at least for the cases when $p+q \le 50$, that is for virtually all practically relevant cases.*

Lemma 3. *Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be positive definite matrices and $\mathbf{D}$ be any matrix, such that $\mathbf{A} \preceq \mathbf{B}$ (i.e. $\mathbf{B}-\mathbf{A}$ is a positive semidefinite matrix), then*

$$\mathbf{A}^{-1} \succeq \mathbf{B}^{-1}, \tag{36}$$

$$\mathrm{tr}\left(\mathbf{D}^H\mathbf{A}\mathbf{D}\right) \le \mathrm{tr}\left(\mathbf{D}^H\mathbf{B}\mathbf{D}\right) \tag{37}$$

$$\mathrm{tr}\left(\mathbf{A}\mathbf{C}\right) \le \mathrm{tr}\left(\mathbf{B}\mathbf{C}\right) \tag{38}$$

$$\mathrm{tr}\left(\mathbf{C}\mathbf{A}^{-1}\right) \ge \mathrm{tr}\left(\mathbf{C}\mathbf{B}^{-1}\right). \tag{39}$$

Proof. $\mathbf{A}^{-1} \succeq \mathbf{B}^{-1}$ is given in [33, p. 495, Corollary 7.7.4(a)]. (37) follows from the fact that $\mathbf{D}^H(\mathbf{B}-\mathbf{A})\mathbf{D}$ is a positive semidefinite matrix since $\mathbf{B}-\mathbf{A}$ is a positive semidefinite matrix and for any $\mathbf{x}$

$$\mathbf{x}^H\mathbf{D}^H(\mathbf{B}-\mathbf{A})\mathbf{D}\mathbf{x} = \mathbf{y}^H(\mathbf{B}-\mathbf{A})\mathbf{y} \ge 0 \tag{40}$$

where $\mathbf{y} \triangleq \mathbf{D}\mathbf{x}$. Let $\mathbf{C} = \mathbf{D}^H\mathbf{D}$ be the Cholesky decomposition of $\mathbf{C}$ then (38) and (39) follows from (37), by utilizing the cyclic property of the trace operator. □

Lemma 4. *For $\bar{q} < 0$ and η satisfying* (34)*, the following relation holds*

$$\begin{aligned}
&\mathbf{E}(\Delta,i)\left(\mathbf{M}(\Delta)+\mathbf{\Sigma}_{p+q}\right)^{-1}\mathbf{E}^H(\Delta,i)\\
&\qquad \preceq \rho(\Delta,i)\mathbf{C}\left(\eta\mathbf{C}+\mathbf{\Sigma}\right)^{-1}\mathbf{C}
\end{aligned} \tag{41}$$

Proof. The proof for the case $p=2$ and $q=1$ is in Appendix A, the proof of the general case is analogous. □

Having prepared with Lemma 3 and Lemma 4, we can prove the (a), (b) and (c) inequalities in (35) by Lemma 5 ((a) part) and Lemma 6 ((b) and (c) parts) as follows.

Lemma 5. *The deterministic equivalent SINR of the tagged user satisfies*

$$\bar{\gamma}(\Delta,i) \le \mathrm{tr}\left(\mathbf{\Phi}(\Delta,i)\boldsymbol{\beta}(\Delta,i)^{-1}\right).$$

Proof. The proof is in Appendix B. □

Lemma 6. *When the conditions of Theorem 1 hold, we have*

$$\mathrm{tr}\left(\mathbf{\Phi}(\Delta,i)\boldsymbol{\beta}(\Delta,i)^{-1}\right) \le \mathrm{tr}\left(\mathbf{\Phi}^{(u)}(\Delta,i)\boldsymbol{\beta}(\Delta,i)^{-1}\right) \tag{42}$$

$$\mathrm{tr}\left(\mathbf{\Phi}^{(u)}(\Delta,i)\boldsymbol{\beta}(\Delta,i)^{-1}\right) \le \bar{\gamma}^{(u)}(\Delta,i). \tag{43}$$

Proof. When the conditions of Theorem 1 hold, Lemma 4 implies that $\mathbf{\Phi}(\Delta,i) \preceq \mathbf{\Phi}^{(u)}(\Delta,i)$ and $\mathbf{Z}(\Delta,i) \succeq \mathbf{Z}^{(u)}(\Delta,i)$. Using the first relation and the Lemma 3 gives (42), while using the second relation and Lemma 3 gives (43). □

B. Useful Properties of the Upper Bounds on the Deterministic Equivalent SINR and Overall System Spectral Efficiency

Theorem 1 is useful, because it establishes an upper bound, denoted by $\bar{\gamma}^{(u)}(\Delta,i)$, of the deterministic equivalent of the SINR, $\bar{\gamma}(\Delta,i)$.

To use the $\bar{\gamma}^{(u)}(\Delta,i)$ upper bound for limiting the search space for an optimal $\bar{\gamma}(\Delta,i)$ in Section V, we need the following properties of the upper bound.

Proposition 2. *The $\bar{\gamma}^{(u)}(\Delta,i)$ upper bound has the following properties: $\partial\bar{\gamma}^{(u)}(\Delta,i)/\partial\rho(\Delta,i) \ge 0$ and $\rho(\Delta,i) \to 0 \Rightarrow \bar{\gamma}^{(u)}(\Delta,i) \to 0$.*

Proof. The proof is in Appendix C. □

Similarly, the SINR of user k satisfies the inequality $\bar{\gamma}_k(\Delta,i) \leq \bar{\gamma}_k^{(u)}(\Delta,i)$ where $\bar{\gamma}_k^{(u)}(\Delta,i)$ is defined in a similar way as $\bar{\gamma}_1^{(u)}(\Delta,i)$. The $\bar{\gamma}_k^{(u)}(\Delta,i)$ upper bound is such that $\partial\bar{\gamma}_k^{(u)}(\Delta,i)/\partial\rho_k(\Delta,i) \geq 0$ and $\rho_k(\Delta,i) \to 0 \Rightarrow \bar{\gamma}_k^{(u)}(\Delta,i) \to 0$.

Since our most important performance measure is the overall SE, we are interested in establishing a corresponding upper bound on the overall SE of the system. To this end, we introduce the related upper bound on the SE of user k:

$$\mathrm{SE}_k^{(u)}(\Delta) \triangleq \frac{\sum_{i=1}^{\Delta} \log\left(1+\bar{\gamma}_k^{(u)}(\Delta,i)\right)}{\Delta}. \tag{44}$$

and bound the aggregate average SE of the MU-MIMO system (c.f. (27)). Notice that the denominator in $\mathrm{SE}_k^{(u)}$ is Δ while the denominator in SE_k is $\Delta+1$. This will be necessary for the monotonicity property in Proposition 3.

Proposition 3.

$$\mathrm{SE}^{(u)}(\Delta) \triangleq \sum_{k=1}^{K} \mathrm{SE}_k^{(u)}(\Delta) \geq \mathrm{SE}(\Delta), \tag{45}$$

and $\mathrm{SE}^{(u)}(\Delta)$ *decreases with* Δ *and approaches* 0 *when* Δ *approaches infinity.*

Proof. The proof is in Appendix D. □

C. Summary

This section first established an upper bound on the deterministic equivalent SINR in Theorem 1. Next, Proposition 2 and Proposition 3 have stated some useful properties of this upper bound and a corresponding upper bound on the overall system spectral efficiency. Specifically, Proposition 3 suggests that the upper bound on the spectral efficiency of the system is monotonically decreasing in Δ and tends to zero as Δ approaches infinity. As we will see in the next section, this property can be exploited to limit the search space for finding the optimal Δ.

V. A Heuristic Algorithm to Find the Optimal Frame Size (Pilot Spacing)

A. A Heuristic Algorithm for Finding the Optimal Δ

In this section we build on the property of the system-wide spectral efficiency, as stated by Proposition 3, to develop a heuristic algorithm to find the optimal Δ. While we cannot prove a convexity or non-convexity property of $\mathrm{SE}(\Delta)$, we can utilize the fact that $\mathrm{SE}(\Delta) \leq \mathrm{SE}^{(u)}(\Delta)$ as follows. As Algorithm 1 scans through the possible values of Δ, it checks if the current best Δ (that is Δ_{opt}) is one less than the currently examined Δ (Line 17). As it will be exemplified in Figure 6 in the numerical section, the key is to notice that the SE upper bound determines the search space of the possible Δ values, where the associated SE can possibly exceed the currently found highest SE. Specifically, the search space can be limited to (Line 18):

$$\Delta_{\max} = {\mathrm{SE}^{(u)}}^{-1}(\mathrm{SE}_\Delta), \tag{46}$$

Algorithm 1: Optimum frame size algorithm using an SE upper bound

Input: $\mathbf{Q}$, $\mathbf{C}$, $\mathbf{\Sigma}$, α^2, P_{tot}

1 $\mathrm{SE}_1 = \mathrm{SE}(1)$ using (27), $\Delta_{\max} = {\mathrm{SE}^{(u)}}^{-1}(\mathrm{SE}_1)$
2 $\Delta = 1$, $\Delta_{\text{opt}} = \Delta_{\max}$, $\mathrm{SE}_{\text{opt}} = \mathrm{SE}(\Delta_{\text{opt}})$ using (27)
3 **while** $\Delta < \Delta_{\max}$ **do**
4 **for** $k = 1 \ldots K$ **do**
5 **for** $i = 1 \ldots \Delta$ **do**
6 Calculate $\mathbf{R}_k(i), \mathbf{R}_k(\Delta+1),$
7 $\mathbf{R}_k(\Delta+1\pm i), \mathbf{R}_k(2\Delta+2)$ using (7)
8 Calculate $\mathbf{E}_k(\Delta,i)$ using (10)
9 Calculate $\mathbf{Z}_k(\Delta,i)$ using (14)
10 Calculate $\mathbf{\Phi}_k(\Delta,i)$ using (16)
11 Calculate $\boldsymbol{\beta}_k(\Delta,i)$ using (18)
12 Calculate $\bar{\gamma}_k(\Delta,i)$ using (22)
13 Calculate $\mathrm{SE}_k(\Delta,i)$ using (25)
14 $\mathrm{SE}_\Delta = \mathrm{SE}(\Delta)$ using (27)
15 **if** $\mathrm{SE}_\Delta > \mathrm{SE}_{\text{opt}}$ **then**
16 $\Delta_{\text{opt}} = \Delta$, $\mathrm{SE}_{\text{opt}} = \mathrm{SE}_\Delta$
17 **if** $\Delta_{\text{opt}} = \Delta - 1$ **then**
18 $\Delta_{\max} = {\mathrm{SE}^{(u)}}^{-1}(\mathrm{SE}_\Delta)$
19 $\Delta = \Delta + 1$

Output: Δ_{opt}

where ${\mathrm{SE}^{(u)}}^{-1}$ denotes the inverse function of $\mathrm{SE}^{(u)}(.)$ and $\mathrm{SE}_\Delta \triangleq \mathrm{SE}(\Delta)$ as calculated in (27). The computational complexity of Algorithm 1 is $\mathcal{O}(\Delta_{\max} K((p+q)N_r)^3)$, where the complexity of the matrix inversion in computing $Z(\Delta,i)$ is $\mathcal{O}(((p+q)N_r)^3)$.

B. The Case of Independent and Identical Channel Coefficients

In the special case where the elements of the vector $\mathbf{h}(i)$ are independent stochastically identical stochastic processes, the covariance matrices become real multiples of the identity matrix $\mathbf{C} \triangleq c\mathbf{I}_{N_r}$, $\mathbf{\Sigma} = s\mathbf{I}_{N_r}$, $\mathbf{R}(i) = r(i)\mathbf{I}_{N_r}$, $\mathbf{Z}(i) = z(i)\mathbf{I}_{N_r}$, $\mathbf{\Phi}(i) = \phi(i)\mathbf{I}_{N_r}$, $\boldsymbol{\beta}(i) = \beta(i)\mathbf{I}_{N_r}$, further more $\mathbf{E}(i) = \mathbf{e}(i) \otimes \mathbf{I}_{N_r}$, with:

$$s \triangleq \frac{\sigma_p^2}{\alpha^2 P_p \tau_p}, \tag{47}$$

$$r(i) \triangleq \begin{cases} ce^{q^* i} & \text{if } i \geq 0, \\ ce^{-qi} & \text{if } i < 0, \end{cases} \tag{48}$$

$$\mathbf{e}(\Delta,i) \triangleq \begin{bmatrix} r(i+(p-1)(\Delta+1)) & \ldots & r(i-q(\Delta+1)) \end{bmatrix} \tag{49}$$

$$\mathbf{m}(\Delta) \triangleq \tag{50}$$

$$\begin{bmatrix} c & r(\Delta+1) & \ldots & r((\Delta+1)(p+q-1)) \\ r(\Delta+1)^* & c & \ddots & \vdots \\ \vdots & \ddots & c & r(\Delta+1) \\ r((\Delta+1)(p+q-1))^* & \ldots & r(\Delta+1)^* & c \end{bmatrix}$$

$$z(i) \triangleq c - \mathbf{e}(\Delta, i)(\mathbf{m}(\Delta) + s\mathbf{I})^{-1}\mathbf{e}^H(\Delta, i), \quad (51)$$

$$\phi(i) \triangleq \alpha^2 P(i)(c - z(i)), \quad (52)$$

$$\beta(i) \triangleq \left(\sum_{k=1}^{K} \alpha_k^2 P_k z_k(i) + \sigma_d^2\right). \quad (53)$$

In this special case, calculating the deterministic equivalent of the SINR by Proposition 1 simplifies to solving a set of scalar equations as stated in the following corollary.

Corollary 3. *In this special case, the deterministic equivalent of the* SINR *in slot i, $\bar{\gamma}(i)$, can be obtained as the solution of the scalar equation*

$$\beta(i) = \frac{N_r \phi(i)}{\bar{\gamma}(i)} - \sum_{k=2}^{K} \frac{\phi_k(i)}{1 + \frac{\bar{\gamma}(i)\phi_k(i)}{\phi(i)}}. \quad (54)$$

Proof. The proof follows from [28] and [29]. Since the matrices $\mathbf{\Phi}_k(i)$ and $\mathbf{Z}_k(i)$ are constant multiple of identity matrices, (24) can then be rewritten as

$$\delta_k(i) = N_r \phi_k(i) \left(\sum_{l=2}^{K} \frac{\phi_l(i)}{1 + \delta_l(i)} + \beta(i)\right)^{-1} \quad (55)$$

for $k = 1, \ldots, K$. Using $\bar{\gamma}(i) = \delta_1(i)$ and comparing (55) for different values of k we get

$$\delta_k(i) = \frac{\phi_k(i)}{\phi_1(i)}\delta_1(i) = \frac{\phi_k(i)}{\phi_1(i)}\bar{\gamma}(i). \quad (56)$$

Substituting the rightmost expression of (56) into (55) with $k = 1$ and rearranging gives the corollary. □

Notice that calculations inside the inner for loop of Algorithm 1, that is the calculations in Lines 6-13 can be substituted by equations (48), (49), (51), (52) and (53). Since these substitutions eliminate the step of inverting a $(p+q)N_r$ size matrix, the time complexity of Algorithm 1 decreases to $\mathcal{O}(\Delta_{\max} K(p+q)^3)$.

VI. Numerical Results

Table I
System Parameters

Parameter	Value
Number of receive antennas at the BS antennas	$N_r = 10, 100$
Path loss of the tagged MS	$\alpha = 90$ dB
Frame size	$\Delta = 2 \ldots 50$
Pilot and data power levels	$P_p = 50...125$ mW; $P = 125$ mW
MIMO receivers	MMSE receiver given by (17)
Channel estimation	MMSE channel estimation given by Lemma 1
Maximum Doppler frequency	$f_D = 50, 500, 1500$ Hz
Slot duration (T)	32μs
Number of users	$K = 2$
Noise variance	$\sigma_p^2 = \sigma_d^2 = -121.45$ dBm

In this section, we consider a single cell of a MU-MIMO ($K = 2$) system with $N_r = 10$ and $N_r = 100$ receive antennas, in which the wireless channel between the served MS and the BS is modelled as (4) and (28).

The MU-MIMO case with greater number of users ($K > 2$) gives similar results albeit with somewhat lower SINR values from the point of view of the tagged user. The BS estimates the state of the wireless channel based on the properly (i.e. $\Delta \times T$) spaced pilot signals using MMSE channel estimation and interpolation according to Lemma 1, and uses MMSE symbol estimation employing the optimal linear receiver $\mathbf{G}^\star(iT)$ in each slot as given in (17). Specifically, in each time slot $i = 1 \ldots \Delta$, the BS uses $p \geq 1$ pilot signals transmitted by the MS prior to the data symbols and q pilot symbols sent after the data symbols, where $q = 0$ or $q = 1$. That is, we study the performance of the "p before, q after" schemes, where q is either zero or one. In practice, when $q = 1$, the BS can store the received data signals until it receives the pilot signal in slot $i = \Delta + 1$ before using an MMSE interpolation of the channel states between $i = 0$ and $i = \Delta + 1$. Furthermore, we will assume that the BS estimates perfectly the autocorrelation function of the channel, including the associated maximum Doppler frequency and, consequently, the characterizing zeroth order Bessel function. The most important system parameters are listed in Table I. Here we assume that the slot duration (T) corresponds to a symbol duration in 5G orthogonal frequency division multiplexing (OFDM) systems using 122 MHz clock frequency, which can be used up to 20 GHz carrier frequencies [34]. Note that the numerical results presented below – except for the benchmarking results in Figures 8 and 9 – are obtained by the analytical calculation and numerical evaluation of the deterministic equivalent of the SINR and the corresponding average spectral efficiency. The benchmarking results shown in Figures 8 and 9 are obtained by Monte Carlo simulations.

Figure 2 shows the achieved spectral efficiency averaged over the data slots $i = 1 \ldots \Delta$, that is averaged over the data slots of a frame of size $\Delta + 1$ when using the 2 before 1 after scheme ($p = 2$, $q = 1$). Short frames imply that the pilot overhead is relatively large, which results in poor spectral efficiency. On the other hand, too large frames (that is when Δ is too large) make the channel estimation quality in the "middle" time slots poor, since for these time slots both available channel estimates $\hat{\mathbf{h}}(0)$ and $\hat{\mathbf{h}}(\Delta + 1)$ convey little useful information, especially at high Doppler frequencies when the channel ages rapidly. Indeed, as seen in Figure 2, the frame size has a large impact on the achievable spectral efficiency, suggesting that the optimum frame size depends critically on the Doppler frequency. As we can see, the spectral efficiency as a function of the frame size is in general neither monotone nor concave, and is hence hard to optimize.

The average spectral efficiency as a function of the pilot/data power ratio and the frame size is shown in Figure 3. This figure clearly shows that setting the proper frame size and tuning the pilot/data power ratio are both important to maximize the average spectral efficiency of the system. The optimal frame size and power configuration are different for different Doppler frequencies, which in turn emphasizes the importance

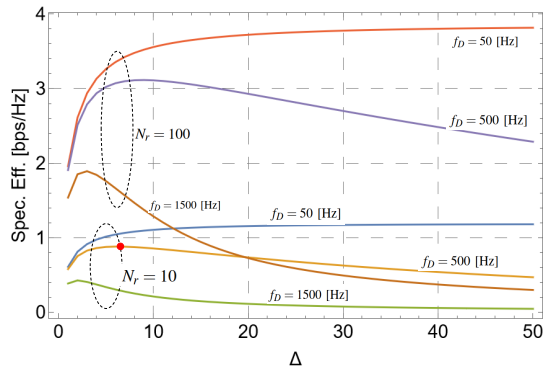

Figure 2. Spectral efficiency as a function of frame size (Δ) with maximum Doppler frequency $f_D = 50, 500, 1500$ Hz with $N_r = 10$ (lower three curves) and $N_r = 100$ (upper three curves). At higher maximum Doppler frequency, the optimum frame size is smaller than at low Doppler frequency. (The red dot indicates the optimal spectral efficiency for the 2 before 1 after scheme with $N_r = 10$ receive antennas when $f_D = 500$ Hz.)

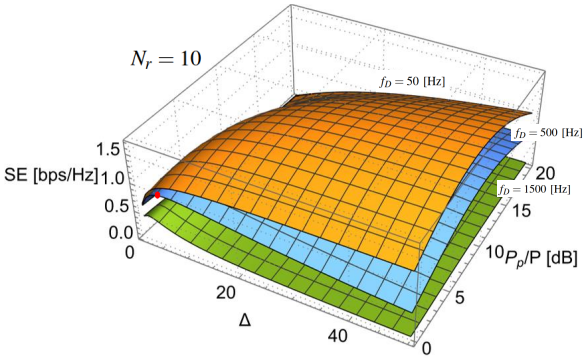

Figure 3. Spectral efficiency with the 2 before 1 after method as a function of the pilot/data power ratio and the frame size with Doppler frequency $f_D = 50$Hz, $f_D = 500$ Hz and $f_D = 1500$ Hz when $N_r = 10$. In all three cases, the spectral efficiency depends heavily on the employed pilot power and pilot spacing (frame size).

of accurate Doppler frequency estimates.

Figure 4 shows the optimal frame size as a function of the maximum Doppler frequency, and 5 shows the achieved spectral efficiency when using the optimal frame size. At $f_D = 500$ Hz, for example, when the optimal frame size is 6, the achieved spectral efficiency when using $N_r = 10$ antennas is a bit below 1 bps/Hz. We can see that setting the optimal frame size is indeed important, because it helps to make the achievable spectral efficiency quite robust with respect to even a significant increase in the Doppler frequency.

Figure 6 illustrates the upper bounds on spectral efficiency as a function of the frame size for different Doppler frequencies. Recall from Figure 2 that the spectral efficiency of the system is a non-concave function of the frame size. Therefore, limiting the possible frame sizes that can optimize spectral efficiency is useful, which can be achieved by the upper bounds shown in the figure. Since the upper bound is monotonically decreasing, finding a point of the spectral efficiency curve (see the curve marked with $f_D = 500$ Hz and its upper bounding curve) with a negative derivative helps to find the range of possible

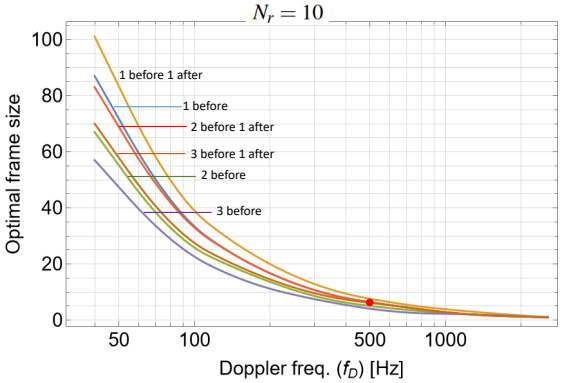

Figure 4. Optimal frame size as a function of the maximum Doppler frequency, and the channel estimation scheme employed.

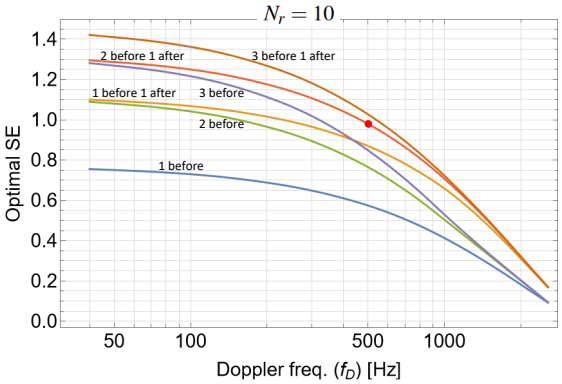

Figure 5. Optimal spectral efficiency as a function of the maximum Doppler frequency, that is the spectral efficiency when using the optimal frame size.

frame sizes that maximize spectral efficiency. For $f_D = 500$ Hz, as illustrated in the figure, larger frame sizes than $\Delta = 34$ would lead to a lower upper bound than the spectral efficiency achieved at $\Delta = 6$. Therefore, when searching for the optimal Δ, once we found that the spectral efficiency at $\Delta = 7$ is less than at $\Delta = 6$ (negative derivative), the search space is limited to $[6, 34]$.

Figure 7 shows the maximum achievable spectral efficiency by setting the optimum frame size as a function of the utilized pilots when using the "p before" and "p before 1 after" schemes for channel estimation. (For an easy reference, in this figure the "2 before 1 after" scheme for the case when f_D is 500 Hz, is represented by the red dot.) At a high Doppler frequency, the optimal spectral efficiency is practically insensitive to increasing p beyond 5, while at a lower Doppler frequency, the optimal spectral efficiency by setting the frame size to its optimal value benefits from utilizing a greater number of past pilots. Since using a Kalman filter gives the same performance as when $p \to \infty$, we see that setting p to a large finite number achieves practically identical performance as that of the Kalman filter.

Figure 8 compares the "1 before", "2 before" and the "2 before, 1 after" schemes with the three schemes ("current", "aged" and "predicted") proposed in [6]. The "current" channel estimation method of [6] uses a pilot in each data slot, and

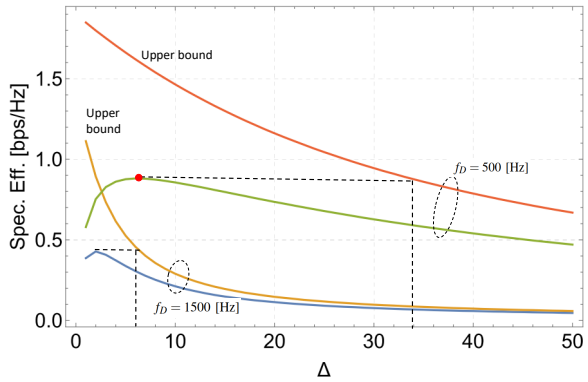

Figure 6. Upper bounding the achievable spectral efficiency as a function of the frame size (Δ) at $f_D = 500$ Hz and $f_D = 1500$ Hz. Note that the upper bound is monotonically decreasing, which helps to limit the search space for the optimum frame size.

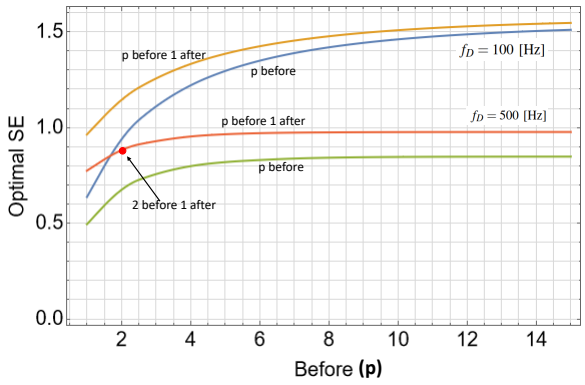

Figure 7. Maximum achievable spectral efficiency vs p when using the "p before" and "p before 1 after" schemes for channel estimation at $f_D = 100$ Hz and $f_D = 500$ Hz Doppler frequency. At a high Doppler frequency, the optimal spectral efficiency is practically insensitive to increasing p beyond 5. (The red dot indicates the 2 before 1 after scheme.)

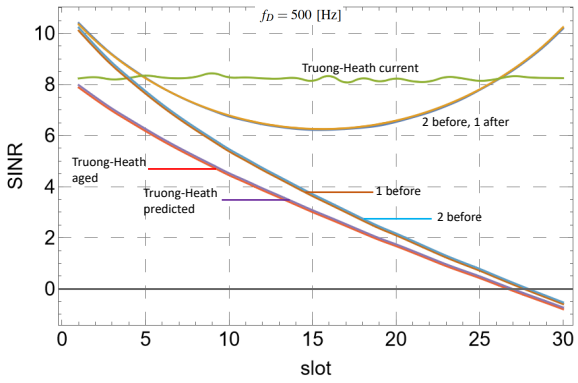

Figure 8. Benchmarking some of the "p before q after" schemes with the three channel estimation and prediction schemes proposed in [6] in terms of the achieved per-slot SINR. Since the Truong-Heath "current" scheme uses a pilot in each slot, it can achieve higher SINR in between pilots when the frame size is set suboptimally (here $\Delta = 30$ instead of the optimal $\Delta = 6$.)

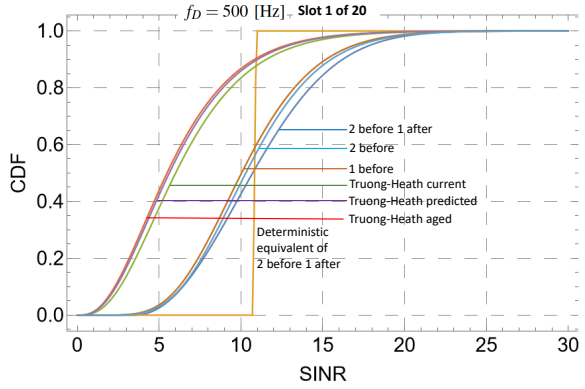

Figure 9. Benchmarking in terms of the CDF of the achieved SINR in slot 1 when the frame size is set to 20. Notice that the deterministic equivalent of the SINR for the 2 before 1 after scheme is an approximation of the average SINR over the channel realizations.

therefore, it achieves the same SINR in each slot. The "aged" and "predicted" methods use only the pilot at the beginning of each frame. All three methods proposed in [6] use maximum ratio combining for data estimation. In contrast, the "2 before 1 after" scheme uses three pilots and achieves higher SINR in the beginning and at the end of each frame. The "aged" and "predicted" methods reach their respective highest SINR at the beginning of the frame, after which both the channel estimation quality and consequently the achieved SINR degrades due to channel aging.

Figure 9 shows the CDF of the achieved SINR in slot 1 when the frame size is set to 20. The 2 before 1 after scheme produces a somewhat higher SINR in the entire support of the SINR, while the Truong-Heath schemes achieve lower SINR values due to employing maximum ratio combining reception as opposed to the MMSE reception used by the p before q after schemes used in this paper.

VII. Conclusions

This paper investigated the fundamental trade-off between using resources in the time domain for pilot signals and data signals in the uplink of MU-MIMO systems operating over fast fading wireless channels that age between subsequent pilot signals. While previous works indicated that when the autocorrelation coefficient between subsequent channel realization instances in discrete time is high, both the channel estimation and the MU-MIMO receiver can take advantage of the memoryful property of the channel in the time domain. However, previous works do not answer the question how often the channel should be observed and estimated such that the subsequent channel samples are sufficiently correlated while taking into account that pilot signals do not carry information bearing symbols and degrade the overall spectral efficiency. To find the optimal pilot spacing, we first established the deterministic equivalent of the achievable SINR and the associated overall spectral efficiency of the MU-MIMO system. We then used some useful properties of an upper bound of this spectral efficiency, which allowed us to limit the search space for the optimal pilot spacing (Δ). The numerical results indicate that the optimal pilot spacing is sensitive to the Doppler frequency of the channel and that proper pilot spacing has a significant impact on the achievable spectral efficiency.

APPENDIX A
PROOF OF LEMMA 4

Proof. Notice that

$$[\mathbf{M}(\Delta)]_{n,m} = \begin{cases} e^{q(\Delta+1)(m-n)}\mathbf{C} & \text{if } n \leq m \\ e^{q(\Delta+1)(n-m)^*}\mathbf{C} & \text{if } n > m \end{cases} \tag{57}$$

that is

$$\mathbf{M}(\Delta) = \tag{58}$$
$$\underbrace{\begin{bmatrix} 1 & e^{q(\Delta+1)} & \dots & e^{q(\Delta+1)(p+q-1)} \\ \left(e^{q(\Delta+1)}\right)^* & 1 & \ddots & \vdots \\ \vdots & \ddots & 1 & \left(e^{q(\Delta+1)}\right)^* \\ e^{q(\Delta+1)(p+q-1)^*} & \dots & \left(e^{q(\Delta+1)}\right)^* & 1 \end{bmatrix}}_{\triangleq \mathbf{M}_{p+q}(\Delta)} \otimes \mathbf{C}.$$

The smallest eigenvalue of $\mathbf{M}_{p+q}(\Delta)$ is analytical for $p+q \leq 4$ and it is provided in (34). For $p+q > 4$, it is not analytical and Remark 1 provides an order 3 polynomial lower bound of the smallest eigenvalue, which is numerically tested up to $p+q = 50$. Let

$$\mathbf{M}^{(u)}(\Delta) \triangleq \underbrace{\begin{bmatrix} \eta & 0 & \cdots & 0 \\ 0 & \eta & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \eta \end{bmatrix}}_{\triangleq \mathbf{M}^{(u)}_{p+q}(\Delta)} \otimes \mathbf{C}. \tag{59}$$

When η satisfies to (34), we have

$$\mathbf{M}^{(u)}_{p+q}(\Delta) \preceq \mathbf{M}_{p+q}(\Delta). \tag{60}$$

Utilizing that the spectrum of a Kronecker product $\sigma(\mathbf{A} \otimes \mathbf{B})$ is [35]

$$\sigma(\mathbf{A} \otimes \mathbf{B}) = \{\, \mu_A \mu_B \mid \mu_A \in \sigma(\mathbf{A}), \mu_B \in \sigma(\mathbf{B}) \,\}, \tag{61}$$

we further have

$$\mathbf{M}^{(u)}(\Delta) \preceq \mathbf{M}(\Delta), \tag{62}$$

which implies

$$\left(\mathbf{M}^{(u)}(\Delta) + \mathbf{\Sigma}_{p+q}\right)^{-1} \succeq \left(\mathbf{M}(\Delta) + \mathbf{\Sigma}_{p+q}\right)^{-1}, \tag{63}$$

according to (36). The statement of the lemma comes from (63) using (37), $\mathbf{M}^{(u)}(\Delta) = \eta \mathbf{I}_{p+q} \otimes \mathbf{C}$, and noting that

$$\begin{aligned} &\mathbf{E}(\Delta,i)\left(\mathbf{M}^{(u)}(\Delta) + \mathbf{\Sigma}_{p+q}\right)^{-1}\mathbf{E}^H(\Delta,i) \\ &= \mathbf{E}(\Delta,i) \begin{bmatrix} \eta\mathbf{C}+\mathbf{\Sigma} & & \\ & \ddots & \\ & & \eta\mathbf{C}+\mathbf{\Sigma} \end{bmatrix}^{-1} \mathbf{E}^H(\Delta,i) \\ &= \sum_{\ell=-(p-1)}^{q} \mathbf{R}(i-\ell(\Delta+1))(\eta\mathbf{C}+\mathbf{\Sigma})^{-1}\mathbf{R}^H(i-\ell(\Delta+1)) \\ &= \rho(\Delta,i)\mathbf{C}\left(\eta\mathbf{C}+\mathbf{\Sigma}\right)^{-1}\mathbf{C}, \end{aligned} \tag{64}$$

where $\mathbf{R}(i)$ and $\rho(\Delta,i)$ are defined in (28) and (33). □

APPENDIX B
PROOF OF LEMMA 5

Proof.

$$\begin{aligned} &\bar{\gamma}(\Delta,i) \\ &= \mathbb{E}\left(\text{tr}\left(\mathbf{\Phi}(\Delta,i)\left(\sum_{k=2}^{K}\mathbf{b}_l(\Delta,i)\mathbf{b}_l^H(\Delta,i) + \boldsymbol{\beta}(\Delta,i)\right)^{-1}\right)\right) \\ &= \int_{\mathbf{v}_2\in\mathbb{R}^{N_r}}\cdots\int_{\mathbf{v}_K\in\mathbb{R}^{N_r}}\prod_{k=2}^{K} Pr(\mathbf{b}_l(\Delta,i)=\mathbf{v}_l) \\ &\qquad \cdot \text{tr}\left(\mathbf{\Phi}(\Delta,i)\left(\sum_{k=2}^{K}\mathbf{v}_l\mathbf{v}_l^H + \boldsymbol{\beta}(\Delta,i)\right)^{-1}\right) d\mathbf{v}_K\dots d\mathbf{v}_2 \\ &\leq \int_{\mathbf{v}_2\in\mathbb{R}^{N_r}}\int_{\mathbf{v}_K\in\mathbb{R}^{N_r}}\prod_{k=2}^{K} Pr(\mathbf{b}_l(\Delta,i)=\mathbf{v}_l) \\ &\qquad \cdot \text{tr}\left(\mathbf{\Phi}(\Delta,i)\boldsymbol{\beta}(\Delta,i)^{-1}\right) d\mathbf{v}_K\dots d\mathbf{v}_2 \\ &= \text{tr}\left(\mathbf{\Phi}(\Delta,i)\boldsymbol{\beta}(\Delta,i)^{-1}\right), \end{aligned}$$

where we used that $\sum_{l=2}^{K}\mathbf{v}_l\mathbf{v}_l^H$ is a positive definite matrix, $\sum_{l=2}^{K}\mathbf{v}_l\mathbf{v}_l^H + \boldsymbol{\beta}(\Delta,i) \succeq \boldsymbol{\beta}(\Delta,i)$ and Lemma 3. □

APPENDIX C
PROOF OF PROPOSITION 2

Proof. To prove monotonicity in ρ first notice that

$$\rho(\Delta_1,i_1) > \rho(\Delta_2,i_2) \Rightarrow \mathbf{Z}^{(u)}(\Delta_1,i_1) \preceq \mathbf{Z}^{(u)}(\Delta_2,i_2),$$
$$\rho(\Delta_1,i_1) > \rho(\Delta_2,i_2) \Rightarrow \mathbf{\Phi}^{(u)}(\Delta_1,i_1) \succeq \mathbf{\Phi}^{(u)}(\Delta_2,i_2).$$

and so

$$\begin{aligned} &\rho(\Delta_1,i_1) > \rho(\Delta_2,i_2) \\ &\Downarrow \\ &\mathbf{\Phi}^{(u)}(\Delta_1,i_1)\left(\sum_{l=1}^{K}\alpha_l^2 P_l \mathbf{Z}_l^{(u)}(\Delta_1,i_1) + \sigma_d^2\mathbf{I}_{N_r}\right)^{-1} \\ &\quad \succeq \mathbf{\Phi}^{(u)}(\Delta_2,i_2)\left(\sum_{l=1}^{K}\alpha_l^2 P_l \mathbf{Z}_l^{(u)}(\Delta_2,i_2) + \sigma_d^2\mathbf{I}_{N_r}\right)^{-1} \\ &\Downarrow \\ &\text{tr}\left(\mathbf{\Phi}^{(u)}(\Delta_1,i_1)\left(\sum_{l=1}^{K}\alpha_l^2 P_l \mathbf{Z}_l^{(u)}(\Delta_1,i_1) + \sigma_d^2\mathbf{I}_{N_r}\right)^{-1}\right) \\ &\geq \text{tr}\left(\mathbf{\Phi}^{(u)}(\Delta_2,i_2)\left(\sum_{l=1}^{K}\alpha_l^2 P_l \mathbf{Z}_l^{(u)}(\Delta_2,i_2) + \sigma_d^2\mathbf{I}_{N_r}\right)^{-1}\right), \end{aligned}$$

that is: $\bar{\gamma}^{(u)}(\Delta_1,i_1) \geq \bar{\gamma}^{(u)}(\Delta_2,i_2)$.

Finally, to prove convergence to 0, notice that

$$\rho(\Delta,i) \to 0 \Rightarrow \mathbf{Z}^{(u)}(\Delta_1,i_1) \to \mathbf{C},$$

$$\rho(\Delta,i) \to 0 \Rightarrow \mathbf{\Phi}^{(u)}(\Delta_1,i_1) \to \mathbf{0}.$$

And so, when $\rho(\Delta,i)\to 0$, we have

$$\bar{\gamma}^{(u)}(\Delta,i) = \operatorname{tr}\left(\mathbf{\Phi}^{(u)}(\Delta,i)\left(\sum_{l=1}^{K}\alpha_l^2 P_l \mathbf{Z}_l^{(u)}(\Delta,i)+\sigma_d^2\mathbf{I}_{N_r}\right)^{-1}\right)$$

$$\overset{\rho(\Delta,i)\to 0}{\to} \operatorname{tr}\left(\mathbf{0}\left(\sum_{l=1}^{K}\alpha_l^2 P_l \mathbf{C}+\sigma_d^2\mathbf{I}_{N_r}\right)^{-1}\right)=0.$$

□

Appendix D
Proof of Proposition 3

Proof. From Theorem 1 and (44) the inequality follows. For monotonicity, notice that $\rho_k(\Delta+1,i)<\rho_k(\Delta,i)$ and $\rho_k(\Delta+1,i+1)<\rho_k(\Delta,i)$. Since by Proposition 2 the upper bound of the SINR is increasing with ρ_k we have

$$\begin{aligned}\bar{\gamma}_k^{(u)}(\Delta+1,i)&\le\bar{\gamma}_k^{(u)}(\Delta,i)\\ \bar{\gamma}_k^{(u)}(\Delta+1,i+1)&\le\bar{\gamma}_k^{(u)}(\Delta,i),\end{aligned}\tag{65}$$

from which it follows that

$$\log\left(1+\bar{\gamma}_k^{(u)}(\Delta+1,i)\right)\le\log\left(1+\bar{\gamma}_k^{(u)}(\Delta,i)\right)\tag{66}$$

$$\log\left(1+\bar{\gamma}_k^{(u)}(\Delta+1,i+1)\right)\le\log\left(1+\bar{\gamma}^{(u)}(\Delta,i)\right).\tag{67}$$

Let $\ell=\arg\min_i\bar{\gamma}_k^{(u)}(\Delta+1,i)$, we then have

$$\begin{aligned}&\frac{1}{\Delta+1}\times\sum_{i=1}^{\Delta+1}\log\left(1+\bar{\gamma}_k^{(u)}(\Delta+1,i)\right)\\ &\le\frac{1}{\Delta}\times\left(\sum_{i=1}^{\ell-1}\log\left(1+\bar{\gamma}_k^{(u)}(\Delta+1,i)\right)\right.\\ &\qquad\left.+\sum_{i=\ell+1}^{\Delta+1}\log\left(1+\bar{\gamma}_k^{(u)}(\Delta+1,i)\right)\right),\end{aligned}$$

since on the right hand side we are removing the smallest term before calculating the mean. Invoking (66) and (67) on the first and second sum, respectively, it follows that

$$\begin{aligned}&\frac{1}{\Delta+1}\times\sum_{i=1}^{\Delta+1}\log\left(1+\bar{\gamma}_k^{(u)}(\Delta+1,i)\right)\\ &\le\frac{1}{\Delta+1}\times\left(\sum_{i=1}^{\ell-1}\log\left(1+\bar{\gamma}_k^{(u)}(\Delta,i)\right)\right.\\ &\qquad\left.+\sum_{i=\ell}^{\Delta}\log\left(1+\bar{\gamma}_k^{(u)}(\Delta,i)\right)\right)\\ &=\frac{1}{\Delta}\times\sum_{i=1}^{\Delta}\log\left(1+\bar{\gamma}_k^{(u)}(\Delta,i)\right).\end{aligned}\tag{68}$$

From which it follows that

$$\begin{aligned}\mathrm{SE}_k^{(u)}(\Delta+1)&=\frac{\sum_{i=1}^{\Delta+1}\log\left(1+\bar{\gamma}_k^{(u)}(\Delta+1,i)\right)}{\Delta+1}\\ &\le\frac{\sum_{i=1}^{\Delta}\log\left(1+\bar{\gamma}_k^{(u)}(\Delta,i)\right)}{\Delta}=\mathrm{SE}_k^{(u)}(\Delta),\end{aligned}\tag{69}$$

that is $\mathrm{SE}_k^{(u)}(\Delta)$ is decreasing in Δ.

To prove convergence to zero, recall from Proposition 2 that $\partial\bar{\gamma}_k^{(u)}(\Delta,i)/\partial\rho_k(\Delta,i)\ge 0$ and

$$\begin{aligned}&\rho_k(\Delta,i)\to 0\Rightarrow\bar{\gamma}_k^{(u)}(\Delta,i)\to 0\\ &\Rightarrow\log\left(1+\bar{\gamma}_k^{(u)}(\Delta,i)\right)\to 0,\end{aligned}\tag{70}$$

where

$$\rho_k(\Delta,i)=e^{2\bar{q}_k(\Delta+1+i)}+e^{2\bar{q}_k i}+e^{2\bar{q}_k(\Delta+1-i)}.$$

We show that for any $\varepsilon>0$, there is some M such that

$$\mathrm{SE}^{(u)}(M)<\varepsilon.\tag{71}$$

Due to $\bar{q}_k<0$, we have $\rho_k(\Delta,i)<\rho_k(1,1)$, which implies

$$\log\left(1+\bar{\gamma}_k^{(u)}(\Delta,i)\right)<\log\left(1+\bar{\gamma}_k^{(u)}(1,1)\right),\tag{72}$$

for all Δ and i. Let $A\triangleq\log\left(1+\bar{\gamma}_k^{(u)}(1,1)\right)$ and N such that $N\varepsilon-2A>0$, and set

$$\epsilon\triangleq\frac{N\varepsilon-2A}{N-2}.\tag{73}$$

Since $\bar{q}_k<0$, we have

$$\begin{aligned}&\rho_k(\Delta,i)\\ &<(p+q)\max(r(i+(p-1)(\Delta+1)),\ldots,r(i-q(\Delta+1)))\\ &=(p+q)e^{2\bar{q}_k\min(i,\Delta+1-i)},\end{aligned}$$

and it follows that for $\frac{\Delta}{N}\le i\le\frac{(N-1)\Delta}{N}$

$$\rho_k(\Delta,i)<(p+q)e^{2\bar{q}_k\frac{\Delta}{N}}.\tag{74}$$

Notice that by equation (70) we can choose some large M, such that

$$\frac{M}{N}\le i\le\frac{(N-1)M}{N}\Rightarrow\log(1+\bar{\gamma}_k^{(u)}(M,i))<\epsilon.\tag{75}$$

We can now show that when $M=\Delta$, then $\mathrm{SE}_k^{(u)}(\Delta)<\varepsilon$. To this end, we split up the sum in the numerator of (69), that is $\sum_{i=1}^{\Delta}\log(1+\bar{\gamma}^{(u)}(\Delta,i))$, into three terms, and bound the first and third terms using the general upper bound A, and the middle term by ϵ:

$$\begin{aligned}\mathrm{SE}_k^{(u)}(\Delta)&=\frac{\sum_{i=1}^{\Delta}\log(1+\bar{\gamma}^{(u)}(\Delta,i))}{\Delta}\\ &=\frac{\sum_{i=1}^{\Delta/N}\log(1+\bar{\gamma}^{(u)}(\Delta,i))}{\Delta}\\ &\quad+\frac{\sum_{i=\Delta/N+1}^{(N-1)\Delta/N}\log(1+\bar{\gamma}^{(u)}(\Delta,i))}{\Delta}\end{aligned}$$

$$
\begin{aligned}
&+ \frac{\sum_{i=(N-1)\Delta/N+1}^{\Delta} \log(1+\bar{\gamma}^{(u)}(\Delta,i))}{\Delta} \\
&< \frac{(\Delta/N)A}{\Delta} + \frac{((N-2)\Delta/N)\epsilon}{\Delta} + \frac{(\Delta/N)A}{\Delta} \\
&= \frac{2A+(N-2)\epsilon}{N} = \varepsilon, \qquad (76)
\end{aligned}
$$

where the last equation is due to the definition of ϵ in (73), which completes the proof. □

REFERENCES

[1] M. Yan and D. Rao, "Performance of an array receiver with a Kalman channel predictor for fast Rayleigh flat fading environments," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 6, pp. 1164–1172, 2001.

[2] Y. Zhang, S. B. Gelfand, and M. P. Fitz, "Soft-output demodulation on frequency-selective Rayleigh fading channels using AR channel models," *IEEE Transactions on Communications*, vol. 55, no. 10, pp. 1929–1939, Oct. 2007.

[3] H. Abeida, "Data-aided SNR estimation in time-variant Rayleigh fading channels," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5496–5507, Nov. 2010.

[4] H. Hijazi and L. Ros, "Joint data QR-detection and Kalman estimation for OFDM time-varying Rayleigh channel complex gains," *IEEE Transactions on Communications*, vol. 58, no. 1, pp. 170–177, Jan. 2010.

[5] S. Ghandour-Haidar, L. Ros, and J.-M. Brossier, "On the use of first-order autorgegressive modeling for Rayleigh flat fading channel estimation with Kalman filter," *Elsevier Signal Processing*, no. 92, pp. 601–606, 2012.

[6] 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, 2013.

[7] 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 Transactions on Communications*, vol. 63, no. 12, pp. 4879–4893, 2015.

[8] L.-K. Chiu and S.-H. Wu, "An effective approach to evaluate the training and modeling efficacy in MIMO time-varying fading channels," *IEEE Transactions on Communications*, vol. 63, no. 1, pp. 140–155, 2015.

[9] S. Kashyap, C. Mollén, E. Björnson, and E. G. Larsson, "Performance analysis of (TDD) massive MIMO with Kalman channel prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, LA, USA: IEEEE, Mar. 2017.

[10] 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*, pp. 1–1, 2020, early access.

[11] J. Yuan, H. Q. Ngo, and M. Matthaiou, "Machine learning-based channel prediction in massive MIMO with channel aging," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 2960–2973, 2020.

[12] G. Fodor, S. Fodor, and M. Telek, "Performance analysis of a linear MMSE receiver in time-variant rayleigh fading channels," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 4098–4112, 2021.

[13] R. Couillet, J. Hoydis, and M. Debbah, "Random beamforming over quasi-static and fading channels: A deterministic equivalent approach," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6392–6425, 2012.

[14] C. Wen, G. Pan, K. Wong, M. Guo, and J. Chen, "A deterministic equivalent for the analysis of non-Gaussian correlated MIMO multiple access channels," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 329–352, 2013.

[15] J. Hoydis, S. T. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need ?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[16] R. K. Mallik, M. R. Bhatnagar, and S. P. Dash, "Fractional pilot duration optimization for SIMO in Rayleigh fading with MPSK and imperfect CSI," *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1732–1744, 2018.

[17] L. Hanlen and A. Grant, "Capacity analysis of correlated MIMO channels," *IEEE Transactions on Information Theory*, vol. 58, no. 11, pp. 6773–6787, 2012.

[18] A. Mahmoudi and M. Karimi, "Inverse filtering based method for estimation of noisy autoregressive signals," *Signal Processing*, vol. 91, no. 7, pp. 1659–1664, 2011.

[19] Y. Xia and W. X. Zheng, "Novel parameter estimation of autoregressive signals in the presence of noise," *Automatica*, vol. 62, pp. 98–105, 2015.

[20] M. Esfandiari, S. A. Vorobyov, and M. Karimi, "New estimation methods for autoregressive process in the presence of white observation noise," *Signal Porcessing (Elsevier)*, vol. 2020, no. 171, pp. 10 780–10 790, 2020.

[21] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, no. 6, pp. 920–928, 2003.

[22] C.-X. Wang, M. Pätzold, and Q. Yao, "Stochastic modeling and simulation offrequency-correlated wideband fading channels," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 3, pp. 1050–1063, May 2007.

[23] M. McGuire and M. Sima, "Low-order Kalman filters for channel estimation," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, Victoria, BC, Canada, Aug. 2005, pp. 352–355.

[24] W. X. Zheng, "Fast identification of autoregressive signals from noisy observations," *IEEE Transactions on Circuits and Systems*, vol. 52, no. 1, pp. 43–48, Jan. 2005.

[25] S. Savazzi and U. Spagnolini, "On the pilot spacing constraints for continuous time-varying fading channels," *IEEE Transactions on Communications*, vol. 57, no. 11, pp. 3209–3213, 2009.

[26] ——, "Optimizing training lengths and training intervals in time-varying fading channels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1098–1112, 2009.

[27] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*, ser. Prentice Hall Signal Processing. Prentice Hall, 1993, no. ISBN: 013-345711-7.

[28] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[29] A. Abrardo, G. Fodor, M. Moretti, and M. Telek, "MMSE receiver design and SINR calculation in MU-MIMO systems with imperfect CSI," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 269–272, Feb. 2019.

[30] G. Fodor, S. Fodor, and M. Telek, "On the achievable SINR in MU-MIMO systems operating in time-varying Rayleigh fading," *IEEE Transactions on Communications (Early Access)*, 2022, DoI: 10.1109/TCOMM.2021.3126760.

[31] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wirel. Commun.*, vol. 4, no. 4, pp. 1650–1662, 2005.

[32] C.-X. Wang and M. Patzold, "Efficient simulation of multiple cross-correlated rayleigh fading channels," in *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003.*, vol. 2, 2003, pp. 1526–1530 vol.2.

[33] R. A. Horn and C. R. Johnson, *Matrix Analysis (2nd ed.)*. Cambridge University Press, 2013.

[34] A. A. Zaidi, R. Baldemair, H. Tullberg, H. Bjorkegren, L. Sundstrom, J. Medbo, C. Kilinc, and I. Da Silva, "Waveform and numerology to support 5g services and requirements," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 90–98, 2016.

[35] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.