Performance Analysis of a Linear MMSE Receiver in Time-Variant Rayleigh Fading Channels

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

***Abstract*—The performance of the uplink of single and multiuser multiple input multiple output (MIMO) systems depends crucially on the receiver architecture and the quality of channel state information at the receiver. Therefore, several previous works have developed minimum mean squared error (MMSE) receivers and proposed balancing the resources spent on acquiring channel state information and transmitting the payload of data packets. Somewhat surprisingly, the most popular MIMO linear MMSE receivers do not exploit the correlation structure that is present in autoregressive Rayleigh fading environments. Therefore, in this paper we first develop a new linear receiver that not only takes channel state information errors into account in minimizing the MSE of the received data symbols, but it also utilizes that the subsequent noisy channel coefficients are correlated. For this new linear MMSE receiver, we derive the achieved MSE as a function of the number of receive antennas and the pilot-to-data power ratio. Interestingly, we find that the pilot power that minimizes the MSE of the data symbols does not depend on the number of antennas and that the new linear MMSE receiver outperforms previously proposed MIMO receivers when the autocorrelation coefficient of the channel is high.**

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

I. Introduction

In multi-user multiple input multiple output (MU-MIMO) systems, the base station (BS) usually estimates the state of the channels by means of reference or pilot signals transmitted by mobile stations (MSs) in the uplink. For example, MSs in long term evolution (LTE) systems use cyclically shifted Zadoff-Chu sequences to form demodulation reference signals that are orthogonal in the code domain. These reference signals enable the BS to obtain channel state information at the receiver (CSIR), which is necessary for uplink data reception [1]. In general, in systems that employ pilot aided CSIR acquisition, determining the number of pilot symbols and tuning the pilot power ratio (PPR) are important for optimizing the system performance, see [2]–[13], and more recently [14].

The seminal work by [2] evaluated the difference between the mutual information when the receiver has only an estimate of the channel and when CSIR is perfect. The results in [3] showed that training based channel estimation significantly affects the capacity and that training imposes a substantial information-theoretic penalty.

Subsequently, references [4], [5] established a lower bound specifically for multiple input multiple output (MIMO) orthogonal frequency division multiplexing (OFDM) systems employing minimum mean squared error (MMSE) channel estimation. It was also shown that the optimal PPR that minimizes the mean symbol error rate increases the capacity by 10-20% as compared with a system using suboptimal PPR setting. Reference [7] optimized the pilot overhead for single-user fading channels in the case of block-fading and continuous fading wireless channels.

In our previous work, we analysed the performance of a receiver that minimizes the mean squared error (MSE) of the uplink received data symbols and treats the estimated channel as if it was the actual channel, and consequently minimizes the MSE only when perfect channel state information is available at the receiver [15]. The MU-MIMO setting is analyzed in [6], in which the coherence interval of T symbols is expended for transmitting pilot symbols (channel training), CSIR acquisition, and precoder computation for downlink (DL) transmission. More recently, references [10], [11] and [12] considered the uplink power control and PPR setting problem in MU-MIMO systems assuming practical (zero-forcing (ZF) and MMSE-based) multiantenna receiver structures, while the fundamental limits of training-based uplink MU-MIMO systems are established in [13]. The results of [13] determine upper and lower bounds of the mutual information in training-based uplink MU-MIMO systems and propose the optimal training design for the maximization of the lower bound of the mutual information. For multi-cell massive MIMO systems, paper [16] developed and analysed a pilot power allocation scheme through user grouping, and derived closed-form expressions for a relative channel estimation error metric and the achievable uplink rate. Notice that none of the above papers developed a MU-MIMO receiver that minimizes the expectation of the data symbol error in the presence of CSIR errors.

In contrast, papers [17]–[19] and more recently [20] proposed linear MMSE receivers that take into account the statistical properties of CSIR errors when minimizing the MSE of the estimated uplink data symbols. Consequently, the specific structure of the MU-MIMO MMSE receiver depends on both the channel estimation scheme (e.g. MMSE used in [17], [21] or least squares (LS) used in [19], [20]) and whether the regularization takes into account the channel estimates of interfering users [18], [20] or is based on the covariance matrices of the interfering users [19]. In either of these cases,

the linear MMSE receiver shows significant gains over naive receivers, which minimize the MSE of the received data symbols only when perfect CSIR is assumed.

A promising technique for large antenna systems employs data-aided channel estimation, according to which partially decoded data are used to aid pilot-based channel state acquisition [22], [23]. Paper [22] proposes an analytical approach that can predict the performance of data-aided channel estimation schemes with good accuracy, especially for matched filter detectors. An iterative data-aided channel estimation receiver is designed in [23], whose performance is analysed using orthogonal or superimposed pilots. However, these papers do not address autoregressive wireless channels. Also, iterative linear MMSE detection has been proposed in [24], and more recently in [25]. In the latter, it was proved that a matched iteative linear MMSE detector achieves the capacity of MIMO non-orthogonal multiple access systems with any number of users. The analysis of these papers focuses on block-fading channels rather than assuming autoregressive fading processes.

Along a closely related line of research, several important contributions showed that exploiting the correlation between the subsequent complex channel coefficients in time-variant Rayleigh fading environments is important. This is because when the channel can be modeled as an autoregressive (AR) process with Gaussian process noise, its state can be better estimated when the correlation between subsequent channel realizations is taken into account [26]–[28]. In fact, this correlation between subsequent channel realizations can also be used to predict the evolution of channel parameters in time, which helps overcome channel aging affects [29]–[33]. However, the receiver design is out of the scope of these papers. Papers [29] and [31], for example, use maximum ratio combining, while papers [32] and [33] focus on channel prediction and are not concerned with receiver design. In contrast, equalizer and receiver algorithms for single input single output (SISO) and MIMO systems are studied in [34], [35] and [36]–[38], respectively. Unfortunately, the algorithms developed for SISO systems do not aim to minimize the MSE of the data symbols and do not easily generalize to MIMO systems. Likewise, the design of an MMSE receiver is not the topic of papers [36], [37] and [38].

Although it is intuitively appealing that an MMSE receiver design should take into account both the CSIR errors and the memoryfull property of the AR Rayleigh channel, to the best of our knowledge, such a receiver has not been designed and analyzed previously. More specifically, knowing that the linear MMSE receiver far outperforms the naive receiver (see [18], [19], [39]) in terms of the achieved MSE, in this paper, we ask the question if is it possible to further improve the performance by regularizing the linear MMSE receiver by means of the correlation structure of the time-variant channel. We will assume that the system can implement a state of the art scheme, such as those developed in [40] or more recently in [41] for estimating the AR parameter of the wireless channel, as it evolves in time. This is an important question, because it gives an indication to system designers on the potential of exploiting the evolution of the channel in time and also of tuning the PPR when using the optimal linear MMSE receiver. Indeed, we intuitively expect that when the channel is memoryfull, capturing its correlation structure in time can lead to spending a smaller proportion of the time and power resources on CSIR acquisition, which leaves more resources to improve the signal-to-interference-plus-noise ratio (SINR) of the transmitted data symbols. However, to the best of our knowledge, this intuition has not been rigorously examined previously.

To summarize the contributions of the present paper, we highlight the following main results: Our main contribution is thus the identification of a linear MMSE receiver that minimizes the MSE of the received data symbols if the channel can be modelled by an AR(1) process, and the quantitative (closed) form analysis of the mean squared error of the data symbols when such a receiver is employed. Specifically, the following are important and highly non-trivial new results, compared with the current state of the art literature:

- Propositions 1 and 2: Give the optimum MIMO receiver when the channel realizations are correlated.
- Theorem 1: Gives the MSE in a closed form, which allows to calculate the MSE without lengthy simulations.
- Proposition 3 and Remark 2: Give the optimum pilot power and help determine the potential gain that comes from tuning the pilot power, rather than using a fixed predetermined pilot power.

We believe that the proposed receiver and the insights obtained in the performance analysis section are useful in the design of MU-MIMO cellular systems and for tuning the pilot and data transmit power levels in single and multiuser MIMO systems.

The next section describes our system model, which comprises the channel model and the received data signal model. Section III provides the analytical description of the channel estimation procedure. Section IV develops the optimal linear MMSE receiver for AR time-variant Rayleigh fading channels and shows that the previously derived linear MMSE receivers are special cases of this general receiver structure. Section V calculates the MSE assuming independent antennas when using the proposed receiver, as a function of the number of antennas and the PPR. Again, we show that previously derived MSE formulas are special cases of this more general formula. Section VI generalizes the results of the previous section for the case when the channel coefficients at the different antennas are correlated. Section VII presents numerical results and highlights engineering insights. Finally, Section VIII draws conclusions and discusses open questions for future work.

II. System Model

A. General Considerations and Pilot Signal Model

We focus on the uplink of a cellular MU-MIMOs system, in which the MSs transmit orthogonal pilot sequences

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

to facilitate CSIR acquisition at the BS. Each pilot symbol is scaled appropriately according to $|s_i|^2 = 1$, for $i = 1, .., \tau_p$. To

insure spatial separability, the pilot sequences are constructed such that they remain orthogonal as long as the number of spatially multiplexed users is maximum τ_p [1]. Specifically, without loss of generality, we assume that the number of MU-MIMO users is $K \leq \tau_p$.

In this paper we assume a comb type arrangement of the pilot symbols [26], [42], which is an appropriate pilot symbol arrangement if the coherence bandwidth of the channel accommodates multiple coherent frequency channels or subcarriers, as noted in [29]. In the comb type pilot arrangement, pilot and data symbols are sent in the same time slot using different subcarriers within the coherence bandwidth. In LTE systems, for example, a 180 kHz chunk (resource block) can often be considered coherent, and it carries 12 subcarriers when 15 kHz subcarrier spacing is used [1].[1]

Given $\tau_p + \tau_d$ subcarriers in the coherence bandwidth, a fraction of τ_p subcarriers are allocated to construct the pilot sequences, while τ_d subcarriers are reserved for data symbols. Each MS transmits at a constant total power P_{tot}, where this power budget can be distributed unequally among the subcarriers. Specifically, for each user, the sum power constraint of

$$\tau_p P_p + \tau_d P = P_{\text{tot}}$$

is enforced, where P_p denotes the pilot power and P denotes the transmit power used for data transmission. (Since the notation P frequently appears in the sequel, we use notation P instead of P_d.) That is, when employing τ_p pilot symbols and a total of $\tau_p P_p$ pilot power for channel estimation, the transmit power for each data symbol is limited to:

$$P = \frac{P_{\text{tot}} - \tau_p P_p}{\tau_d}. \tag{2}$$

As (2) suggests, the trade-off between pilots and data signals includes both the transmit power and time (or number of pilot and data symbols), as it was examined by several papers, see for example [9], [42]. For ease of presentation, and without loss of generality, we refer to User-1 as the tagged user, and avoid indexing the users unless such indexing is necessary to avoid confusion. Thus, the $N_r \times \tau_p$ matrix of the received pilot signal from the tagged user at the BS can be conveniently written as:

$$\mathbf{Y}^p = \alpha\sqrt{P_p}\mathbf{h}\mathbf{s}^T + \mathbf{N} \quad \in \mathbb{C}^{N_r \times \tau_p}, \tag{3}$$

where we assume that $\mathbf{h} \in \mathbb{C}^{N_r \times 1}$ is a circular symmetric complex normal distributed column vector with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C}$ (of size N_r), denoted as $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, α accounts for the large scale fading, $\mathbf{N} \in \mathbb{C}^{N_r \times \tau_p}$ is the spatially and temporally additive white Gaussian noise (AWGN) with element-wise variance σ_p^2, where the index p refers to the pilot signal.

B. Channel Model

As an extension of the memoryless channel model, in this paper we assume that in the consecutive CSIR acquisition periods (denoted as $\ldots, t-1, t, t+1, \ldots$) the complex channel vector evolves according to the following discrete time AR(1) equation [26], [34]:

$$\mathbf{h} = \mathbf{h}(t) = \mathbf{A}\mathbf{h}(t-1) + \boldsymbol{\vartheta}(t) \quad \in \mathbb{C}^{N_r \times 1}, \tag{4}$$

where $\boldsymbol{\vartheta}(t) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Theta})$ is a complex normal distributed process noise vector with zero mean and covariance matrix $\boldsymbol{\Theta}$, which is identically and independently distributed in consecutive CSIR acquisition periods, and $\mathbf{A}$ is the transition matrix of the AR(1) process, which characterizes the dependence of the consecutive channel vectors [37]. We recall that N_r is the number of antennas at the BS. Altough Rayleigh fading cannot be perfectly modelled with any finite order AR process, the coefficients of the AR models can be determined such that the statistics of the AR model closely match those of the Rayleigh fading [43], [44]. In this paper, we assume that the AR(1) model in (4) is a good model of the wireless fading channel, and that the model parameters can be determined by existing system identification techniques (see for example [43], [45] and more recently [46]).

In this model we assume that $\mathbf{h}(t)$ is stationary, implying that its mean vector and covariance matrix are constant:

$$\mathbf{h}(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}) \quad \forall t,$$

which, according to (4), leads to

$$\mathbf{C} = \mathbf{A}\mathbf{C}\mathbf{A}^H + \boldsymbol{\Theta} \quad \in \mathbb{C}^{N_r \times N_r}. \tag{5}$$

C. Received Data Signal Model

Assuming K spatially multiplexed users, the MU-MIMO received data signal at the BS at time t can be written as:

$$\mathbf{y}(t) = \underbrace{\alpha\mathbf{h}(t)\sqrt{P}x(t)}_{\text{tagged user}} + \underbrace{\sum_{k=2}^{K} \alpha_k \mathbf{h}_k(t)\sqrt{P_k}x_k(t)}_{\text{other users}} + \mathbf{n}_d(t), \tag{6}$$

where $\mathbf{y}(t) \in \mathbb{C}^{N_r \times 1}$; and α_k $\mathbf{h}_k(t)$ denotes the $N_r \times 1$ vector channel including large and small scale fading between User-k and the BS, P_k is the data transmit power of User-k, $x_k(t)$ is the transmitted data symbol by User-k and $\mathbf{n}_d(t)$ denotes the AWGN on the received data signal with element-wise variance σ_d^2.

The most important system parameters are summarized in Table I.

III. Channel Estimation

A. Least Squares Channel Estimation

For benchmarking purposes, we first assume that the BS uses the popular LS channel estimator, which relies on correlating the received signal with the known pilot sequence. Note that our methodology to determine the MSE of the received data is not confined to the LS estimator, but is directly applicable to an MMSE or other linear channel estimation techniques as well. For each MS, the BS utilizes pilot sequence orthogonality and estimates the channel based on (3) assuming:

$$\hat{\mathbf{h}} = \frac{1}{\alpha\sqrt{P_p}}\mathbf{Y}^p\mathbf{s}^*(\mathbf{s}^T\mathbf{s}^*)^{-1},$$

[1] In this paper, we do not study the situation in which the channel changes both in time and frequency on a symbol-by-symbol level.

Table I
NOTATION AND TERMINOLOGY

Notation	Meaning
K	Number of single-antenna users
N_r	Number base station antennas
τ_p, τ_d	Number of pilot and data symbols, respectively, within a coherent block of symbols
$\mathbf{s} \in \mathbb{C}^{\tau_p \times 1}$	Sequence of pilot symbols
x	Data symbol
P_p, P, P_{tot}	Pilot power per symbol, data power per symbol, and total transmit power
$\mathbf{Y}^p \in \mathbb{C}^{N_r \times \tau_p}, y(t) \in \mathbb{C}^{N_r}$	Received pilot and data signal, respectively
$\mathbf{N}, \mathbf{n}_d(t)$	Additive white Gaussian noise at the received pilot and data signal, respectively
α	Large scale fading between the mobile station and the base station
$\mathbf{h}(t), \hat{\mathbf{h}}(t) \in \mathbb{C}^{N_r}$	Fast fading channel and estimated channel
$\sigma_p^2\mathbf{I}_{N_r}, \sigma_d^2\mathbf{I}_{N_r}, \mathbf{C} \in \mathbb{C}^{N_r \times N_r}$	Covariance of $\mathbf{N}$, $\mathbf{n}_d$, $\mathbf{h}(t)$, respectively
$\mathbf{A} \in \mathbb{C}^{N_r \times N_r}$	State transition matrix of the fast fading channel as an AR process parameter
$\boldsymbol{\vartheta}(t) \in \mathbb{C}^{N_r}, \boldsymbol{\Theta} \in \mathbb{C}^{N_r \times N_r}$	Process noise of the channel AR process and its covariance matrix
$\mathbf{w}(t) \in \mathbb{C}^{N_r}, \boldsymbol{\Sigma} \in \mathbb{C}^{N_r \times N_r}$	Channel estimation error (measurement noise) and its covariance matrix
$\mathbf{G}, \mathbf{G}^{\text{naive}}, \mathbf{G}^{\text{conv}}, \mathbf{G}^{\star}$	Generic, naive, conventional and optimal (i.e. minimum MSE) receiver

that is:

$$\hat{\mathbf{h}} = \mathbf{h} + \mathbf{w} = \mathbf{h} + \frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbf{N}\,\mathbf{s}^*, \tag{7}$$

where $\mathbf{s}^* = \left[s_1^*, ..., s_{\tau_p}^*\right]^T \in \mathbb{C}^{\tau_p \times 1}$ denotes the vector of conjugate pilot symbols and $(\mathbf{s}^T\mathbf{s}^*) = \tau_p$.

By considering $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, it follows that the estimated channel $\hat{\mathbf{h}}$ is a circular symmetric complex normal distributed vector $\hat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, with

$$\mathbf{R} \triangleq \mathbb{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\} = \mathbf{C} + \frac{\sigma_p^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r}, \tag{8}$$

where $\mathbf{I}_{N_r}$ is the identity matrix of size N_r. The channel estimation error is defined as

$$\mathbf{w} \triangleq \hat{\mathbf{h}} - \mathbf{h},$$

so that $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\Sigma} = \frac{\sigma_p^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} \triangleq s\mathbf{I}_{N_r} \tag{9}$$

and the MSE of the least squares channel estimation is derived as

$$\varepsilon_{\text{LS}} \triangleq \mathbb{E}_{\mathbf{h},\mathbf{N}}\{||\mathbf{w}||_F^2\} = \text{Tr}\left\{\boldsymbol{\Sigma}\right\} = \frac{N_r\sigma_p^2}{\alpha^2 P_p \tau_p}, \tag{10}$$

where Tr denotes the trace operation and $||\cdot||_F^2$ is the Frobenius norm. As it was shown in [19], [39], the distribution of the channel realization $\mathbf{h}$ conditioned on the estimate $\hat{\mathbf{h}}$ is normally distributed as follows:

$$(\mathbf{h} \mid \hat{\mathbf{h}}) \sim \mathbf{D}\hat{\mathbf{h}} + \underbrace{\mathcal{CN}\Big(\mathbf{0}, \mathbf{Q}\Big)}_{\text{channel estimation noise}}, \tag{11}$$

where $\mathbf{D} \triangleq \mathbf{C}\mathbf{R}^{-1}$ and $\mathbf{Q} \triangleq \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C}$. Equation (11) suggests that when using LS channel estimation, increasing the pilot power improves the quality of the acquired CSIR by reducing the channel estimation noise $\mathbf{Q}$. As we will see in the sequel, reducing the channel estimation noise is also possible by exploiting the dependency in the evolution of the channel.

B. Kalman Filtering Assisted Channel Estimation

As described in (4), in this paper we model the wireless channel as a stochastic time-variant linear system. The state of the channel is estimated in the form of the observation vector as follows:

$$\hat{\mathbf{h}}(t) = \mathbf{h}(t) + \mathbf{w}(t), \tag{12}$$

where the estimation error $\mathbf{w}(t)$ can be thought of as measurement noise or observation error. In this paper we assume that the channel estimator at the BS uses Kalman filtering, as an alternative to LS channel estimation, to forecast the channel as:

$$\mathbf{h}^f(t) = \mathbf{A}\mathbf{h}^a(t-1),$$

where $\mathbf{h}^a(t-1) \triangleq \mathbb{E}_{\mathbf{h}(t-1)}\{\mathbf{h}(t-1)|\hat{\mathbf{h}}(t-1)\} = \mathbf{D}\hat{\mathbf{h}}(t-1)$ is the best estimate of $\mathbf{h}(t-1)$ based only on the current observation $\hat{\mathbf{h}}(t-1)$ at time instance $(t-1)$. When using Kalman filtering, the best estimate of the channel state at t combines the channel forecast and the channel observation at t, that is:

$$\mathbf{h}^a(t) = \mathbf{h}^f(t) + \mathbf{K}\Big(\underbrace{\hat{\mathbf{h}}(t) - \mathbf{h}^f(t)}_{\text{innovation}}\Big), \tag{13}$$

where $\mathbf{K}$ is the Kalman gain matrix, and $\mathbf{h}^a(t)$ is the optimal estimate of $\mathbf{h}(t)$ at time t. The Kalman gain matrix $\mathbf{K}$ is determined according to:

$$\begin{aligned}\mathbf{P}^f &= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T + \boldsymbol{\Theta},\\ \mathbf{K} &= \mathbf{P}^f\left(\mathbf{P}^f + \boldsymbol{\Sigma}\right),\end{aligned} \tag{14}$$

where $\mathbf{P}^f$ is the forecast error covariance matrix.

C. Some Useful Properties of $\mathbf{h}(t)$, $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{h}}(t-1)$ When Using Kalman Filtering

According to (4) and (12), the random vector composed of $\mathbf{h}(t)$, $\hat{\mathbf{h}}(t)$, $\hat{\mathbf{h}}(t-1)$ is multivariate zero-mean complex normal distributed. Consequently, its distribution is determined by the joint covariance matrix. Since the joint distribution will be needed in the subsequent MSE calculation, the following two lemmas will be useful to determine the structure of the MMSE receiver.

Lemma 1. *The covariance matrix of the* $\left[\hat{\mathbf{h}}(t), \hat{\mathbf{h}}(t-1), \mathbf{h}(t)\right]^T$ *complex normal distributed random vector is*

$$\mathbf{\Psi} = \begin{bmatrix} \mathbf{C}+\mathbf{\Sigma} & \mathbf{AC} & \mathbf{C} \\ \mathbf{CA^H} & \mathbf{C}+\mathbf{\Sigma} & \mathbf{CA^H} \\ \mathbf{C} & \mathbf{AC} & \mathbf{C} \end{bmatrix}. \tag{15}$$

The proof is in Appendix I. Lemma 1 is useful, because according to Theorem 10.2 of [47], if we know the joint probability density function of multivariate complex Gaussian vectors, which in our case is completely determined by the above covariance matrix, the conditional probability density functions can be easily determined. This is important, because it will help us to eliminate the dependence of the MSE $(\mathbf{G}, \mathbf{h})$ expression on $\mathbf{h}$ in the sequel. We can now state the following lemma, which will be important in determining the MMSE receiver and the achieved MSE.

Lemma 2. *The channel realization* $\mathbf{h}(t)$ *conditioned on the current and previous estimates* $\hat{\mathbf{h}}(t)$ *and* $\hat{\mathbf{h}}(t-1)$ *is normally distributed as follows:*

$$\left(\mathbf{h}(t)\middle|\hat{\mathbf{h}}(t), \hat{\mathbf{h}}(t-1)\right) \sim \mathbf{E}\boldsymbol{\zeta}(t) + \underbrace{\mathcal{CN}\left(\mathbf{0}, \mathbf{Z}\right)}_{\text{channel estimation noise}}, \tag{16}$$

where

$$\boldsymbol{\zeta}(t) \triangleq \begin{bmatrix} \hat{\mathbf{h}}(t) \\ \hat{\mathbf{h}}(t-1) \end{bmatrix} \in \mathbb{C}^{2N_r \times 1},$$

$$\mathbf{E} \triangleq \begin{bmatrix} \mathbf{C} & \mathbf{AC} \end{bmatrix} \begin{bmatrix} \mathbf{C}+\mathbf{\Sigma} & \mathbf{AC} \\ \mathbf{CA}^H & \mathbf{C}+\mathbf{\Sigma} \end{bmatrix}^{-1} \in \mathbb{C}^{N_r \times 2N_r}, \tag{17}$$

$$\mathbf{Z} \triangleq \mathbf{C} - \mathbf{E}\begin{bmatrix} \mathbf{C} \\ \mathbf{CA}^H \end{bmatrix} \in \mathbb{C}^{N_r \times N_r}. \tag{18}$$

The proof is in Appendix II. Comparing (11) and (16), notice that Lemma 2 suggests that when channel estimation utilizes both $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{h}}(t-1)$, that is when we use Kalman filtering, the channel estimation noise is characterized by the covariance matrix $\mathbf{Z}$ rather than by $\mathbf{Q}$.

Remark 1. *In the special case when* $\mathbf{A} = \mathbf{0}$*, it holds that* $\mathbf{E} = \mathbf{D}$ *and* $\mathbf{Z} = \mathbf{Q}$ *and:*

$$\mathbb{E}\left(\mathbf{h}(t)\middle|\hat{\mathbf{h}}(t), \hat{\mathbf{h}}(t-1)\right) = \mathbf{C}\mathbf{R}^{-1}\hat{\mathbf{h}}(t),$$

and

$$\mathbf{C}_{\mathbf{h}(t)\middle|\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)} = \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C}. \tag{19}$$

Note that Lemma 2 suggests that when the channel follows an AR(1) process, and the channel estimator uses Kalman filtering, the receiver structure should be adjusted as compared with the case when the channel estimator uses LS channel estimation based only on the currently received pilot signal. That is, to fully take advantage of the memory of the channel, not only the channel estimator should be modified (using Kalman filtering), but also the receiver should be adjusted. We elaborate more on this idea in the next section.

IV. Deriving the Optimal MMSE Receiver for Autoregressive Rayleigh Channels

A. Employing an MMSE Receiver at the BS

In this paper the BS employs an MMSE receiver $\mathbf{G} \in \mathbb{C}^{1\times N_r}$ to estimate the transmitted data symbols. We recall that the MMSE receiver aims to minimize the MSE between the estimate $\mathbf{Gy}$ and the transmitted symbol x:

$$\mathbf{G}^\star \triangleq \arg\min_{\mathbf{G}} \mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}},\mathbf{n},x}\{|\mathbf{Gy} - x|^2\} \quad \in \mathbb{C}^{1\times N_r}. \tag{20}$$

When the BS employs a naïve receiver, the estimated channel is taken as if it was the actual channel:

$$\mathbf{G}^{\text{naive}} = \alpha\sqrt{P}\hat{\mathbf{h}}^H(\alpha^2 P\hat{\mathbf{h}}\hat{\mathbf{h}}^H + \sigma_d^2\mathbf{I}_{N_r})^{-1}. \tag{21}$$

As we shall see, this receiver does not minimize the MSE. For block fading channels, the conventional receiver that uses a single channel estimate ($\hat{\mathbf{h}}$) and minimizes the MSE of the received data symbols can be written as [19]:

$$\mathbf{G}^{\text{conv}} = \alpha\sqrt{P}\hat{\mathbf{h}}^H\mathbf{D}^H\left(\alpha^2 P\left(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}\right) + \sigma_d^2\mathbf{I}_{N_r}\right)^{-1}. \tag{22}$$

This receiver will serve as an important benchmark to the MMSE receiver that we develop in the sequel.

B. Determining the Actual MMSE Receiver Vector

This section is concerned with determining the MMSE receiver vector $\mathbf{G}$ that the BS should use to demodulate the received data signal such that the data symbol estimation error for the tagged user is minimized. This minimization should take explicitly account that the BS has access only to the estimated channels $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{h}}(t-1)$. This receiver can be opposed to the naïve receiver that minimizes the MSE only when perfect channel estimation is assumed. To this end, we consider the MSE of the estimated data symbols of the tagged user, obtained from the signal model of (6) using a receiver vector $\mathbf{G}$:

$$\begin{aligned} &\text{MSE}\left(\mathbf{G}, \mathbf{h}(t), \mathbf{h}_2(t)\ldots, \mathbf{h}_K(t)\right) = \mathbb{E}_{x,\mathbf{n}_d}\{|\mathbf{Gy} - x|^2\} = \\ &= \mathbb{E}_{x,\mathbf{n}_d}\Big|(\mathbf{G}\alpha\mathbf{h}(t)\sqrt{P} - 1)x + \sum_{k=2}^{K}\mathbf{G}\alpha_k\mathbf{h}_k(t)\sqrt{P_k}x_k + \\ &+ \mathbf{G}\mathbf{n}_d\Big|^2 = \mathbb{E}_{x,\mathbf{n}_d}\Big|(\mathbf{G}\alpha\mathbf{h}(t)\sqrt{P} - 1)x\Big|^2 + \\ &+ \sum_{k=2}^{K} P_k\mathbb{E}_{x,\mathbf{n}_d}|\mathbf{G}\alpha_k\mathbf{h}_k(t)x_k|^2 + \mathbb{E}_{x,\mathbf{n}_d}|\mathbf{G}\mathbf{n}_d|^2, \end{aligned} \tag{23}$$

where we utilized that $\mathbb{E}\{x_k\} = 0$ and $\mathbb{E}\{\mathbf{n}_d\} = \mathbf{0}$. Additionally, utilizing $\mathbb{E}\{x_k x_k^*\} = 1$ and $\mathbb{E}\{\mathbf{n}_d\mathbf{n}_d^H\} = \sigma_d^2\mathbf{I}_{N_r}$, we have:

$$\begin{aligned} &\text{MSE}\left(\mathbf{G}, \mathbf{h}(t), \mathbf{h}_2(t), \ldots, \mathbf{h}_K(t)\right) = \\ &= \Big|\mathbf{G}\alpha\mathbf{h}(t)\sqrt{P} - 1\Big|^2 + \sum_{k=2}^{K} P_k|\mathbf{G}\alpha_k\mathbf{h}_k(t)|^2 + \sigma_d^2\mathbf{G}\mathbf{G}^H, \end{aligned} \tag{24}$$

from which the MSE of the tagged user can be expressed as follows:

$$\text{MSE}\left(\mathbf{G},\mathbf{h}(t)\right)=\mathbb{E}_{\mathbf{h}_2(t),\ldots,\mathbf{h}_K(t)}\Big\{\text{MSE}\left(\mathbf{G},\mathbf{h}_1(t),\ldots,\mathbf{h}_K(t)\right)\Big\}$$

$$=\alpha^2P\mathbf{G}\mathbf{h}(t)\mathbf{h}^H(t)\mathbf{G}^H-\alpha\sqrt{P}\Big(\mathbf{G}\mathbf{h}(t)+\mathbf{h}^H(t)\mathbf{G}^H\Big)$$
$$+\underbrace{\sum_{k=2}^{K}\alpha_k^2P_k\mathbf{G}\mathbf{C}_k\mathbf{G}^H}_{\text{multi-user interference}}+\sigma_d^2\mathbf{G}\mathbf{G}^H+1. \tag{25}$$

Equation (25) expresses the MSE as a function of the generic receiver $\mathbf{G}$, the actual channel realization of he tagged user, and the multiuser interference. In practice, we can only design a receiver $\mathbf{G}$ which is a function of the estimated channels, and therefore we need an expression for the MSE as a function of $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{h}}(t-1)$, rather than $\mathbf{h}(t)$. This is stated in the following proposition, which invokes Lemma 1:

Proposition 1. *The MU-MIMO receiver vector that minimizes the MSE is as follows.*

$$\mathbf{G}^\star(t)=\arg\min_{\mathbf{G}}\text{MSE}\left(\mathbf{G},\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)\right)=\mathbf{b}(t)^H\mathbf{F}(t)^{-1}, \tag{26}$$

where $\mathbf{b}(t)\in\mathbb{C}^{N_r\times1}$ *and* $\mathbf{F}(t)\in\mathbb{C}^{N_r\times N_r}$ *are defined as*

$$\mathbf{b}(t)\triangleq\alpha\sqrt{P}\mathbf{E}\boldsymbol{\zeta}(t), \tag{27}$$

$$\mathbf{F}(t)\triangleq\Bigg(\mathbf{b}(t)\mathbf{b}(t)^H+\underbrace{\alpha^2P\Big(\mathbf{Z}+\sum_{k\neq\kappa}\mathbf{C}_k+\frac{\sigma_d^2}{\alpha^2P}\mathbf{I}_{N_r}\Big)}_{\triangleq\mathbf{M}}\Bigg), \tag{28}$$

and the $\boldsymbol{\zeta}(t)$, $\mathbf{E}$ *and* $\mathbf{Z}$ *matrices were introduced in Lemma 2.*

The proof is in Appendix III. The following remark relates the proposed MMSE receiver to the traditional receiver, which was proposed in, for example, [19].

Remark 2. *In the special case when* $\mathbf{A}=\mathbf{0}$*, it holds that* $\mathbf{E}=\mathbf{D}$*,* $\mathbf{Z}=\mathbf{Q}$ *and* $\boldsymbol{\zeta}(t)=\hat{\mathbf{h}}$*. Therefore, in this special case it also holds that* $\mathbf{b}(t)=\alpha\sqrt{P}\mathbf{D}\hat{\mathbf{h}}$ *and:*

$$\mathbf{G}^\star(t)=\alpha\sqrt{P}\hat{\mathbf{h}}^H\mathbf{D}^H\Big(\alpha^2P\Big(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H+\mathbf{Q}\Big)+$$
$$+\sum_{k=2}^{K}\alpha_k^2P_k\mathbf{C}_k+\sigma_d^2\mathbf{I}_{N_r}\Big)^{-1}, \tag{29}$$

which is identical with the $\mathbf{G}^{\text{conv}}$ *conventional receiver expressed in* (22).

C. Summary

In this section, we derived the optimal linear receiver $\mathbf{G}^\star(t)$, which minimizes the MSE of the data symbols if the wireless channel can be modeled as an AR(1) process, whose transition matrix $\mathbf{A}$ can be estimated by state of the art system identification methods [27], [48]. Under this assumption, the MMSE receiver derived in Proposition 1 takes advantage of (regularized by) the statistical properties of the estimates $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{h}}(t-1)$, rather than operating on $\hat{\mathbf{h}}(t)$ only. As we will see, using Kalman filtering for channel estimation and the proposed MMSE regularization leads to improved overall performance in terms of achieved MSE and SINR.

V. Calculating the MSE When Using the MMSE Receiver with Uncorrelated and Identical Channel Coefficients

In this section, our goal is to derive a closed form expression for the MSE, when the receiver uses the MMSE receiver derived in the preceding section and the antennas are properly spaced and the channel coefficients at the different antennas can be assumed uncorrelated and identically distributed. That is, we consider the case of $\mathbf{C}=c\mathbf{I}_{N_r}$, $\mathbf{A}=a\mathbf{I}_{N_r}$, $\boldsymbol{\Sigma}=s\mathbf{I}_{N_r}=\frac{\sigma_p^2}{\alpha^2P_p\tau_p}\mathbf{I}_{N_r}$ according to (9), and $\boldsymbol{\Theta}=\theta\mathbf{I}_{N_r}=(c-aca^*)\mathbf{I}_{N_r}$ according to (5). These assumptions allow a simplified computation of the MSE based on scalar coefficients. In the next section, we handle the general case with dependent antennas using matrix expressions.

To compute the MSE, we first calculate the conditional expectation of the squared symbol error when using the optimal $\mathbf{G}(t)^\star$ receiver as a function of the estimated channels $\hat{\mathbf{h}}(t-1)$ and $\hat{\mathbf{h}}(t)$. Next, utilizing the joint distribution of the estimated channels, we determine the unconditional MSE of the received data symbols.

We start by noticing the following corollary of Proposition 1.

Corollary 1. *When the receiver uses the* $\mathbf{G}^\star(t)$ *MMSE receiver, the achieved MSE is as follows.*

$$\min_{\mathbf{G}}\text{MSE}\left(\mathbf{G},\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)\right)=$$
$$=\text{MSE}\left(\mathbf{G}^\star(t),\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)\right)=$$
$$=1-\mathbf{b}(t)^H\mathbf{F}(t)^{-1}\mathbf{b}(t). \tag{30}$$

Proof. Substituting $\mathbf{G}^\star(t)=\mathbf{b}(t)^H\mathbf{F}(t)^{-1}$ into (67) in Appendix III yields the corollary. □

A. Calculating $\mathbf{G}^\star(t)$ in the Case of Uncorrelated and Identical Channel Coefficients

Recalling the definitions of $\mathbf{E}$ in (17) and using the uncorrelated and identically distributed channel coefficients assumption we get:

$$\mathbf{E}=\begin{bmatrix}e_1\mathbf{I}_{N_r} & e_2\mathbf{I}_{N_r}\end{bmatrix}\in\mathbb{C}^{N_r\times2N_r}, \tag{31}$$

where:

$$e_1=\frac{c(c+s-aca^*)}{c(c+s-aca^*)+s(c+s)},\qquad e_2=\frac{acs}{(c+s)^2-ac^2a^*}. \tag{32}$$

Furthermore, due to the definition of $\mathbf{Z}$ in (18), we have that $\mathbf{Z}=z\mathbf{I}$, where

$$z=\frac{cs(c+s-aca^*)}{(c+s)^2-ac^2a^*}.$$

Likewise, using the definition of $\mathbf{b}(t)$ in (27), we get:

$$\mathbf{b}(t) = \alpha\sqrt{P}\left(e_1\hat{\mathbf{h}}(t) + e_2\hat{\mathbf{h}}(t-1)\right) = \alpha\sqrt{P}\mathbf{v}(t) \quad \in \mathbb{C}^{N_r\times 1}, \tag{33}$$

where we introduced:

$$\mathbf{v}(t) \triangleq \mathbf{E}\boldsymbol{\zeta}(t) = e_1\hat{\mathbf{h}}(t) + e_2\hat{\mathbf{h}}(t-1). \tag{34}$$

Finally, using the definition of $\mathbf{F}(t)$ in (28), it simplifies to:

$$\begin{aligned}\mathbf{F}(t) &= \alpha^2 P\Bigg((e_1\hat{\mathbf{h}}(t) + e_2\hat{\mathbf{h}}(t-1))(e_1\hat{\mathbf{h}}(t) + e_2\hat{\mathbf{h}}(t-1))^H \\ &\quad + \Bigg(\underbrace{z + \sum_{k\neq\kappa} c_k + \frac{\sigma_d^2}{\alpha^2 P}}_{\triangleq f}\Bigg)\mathbf{I}_{N_r}\Bigg) \\ &= \alpha^2 P\big(\mathbf{v}(t)\mathbf{v}(t)^H + f\mathbf{I}_{N_r}\big).\end{aligned} \tag{35}$$

Using the above assumptions and notation, we can derive the MMSE receiver for the special case of uncorrelated and identically distributed channel coefficients at the different antennas.

Proposition 2. *When the channel covariance matrices are diagonal with equal elements, the MMSE receiver takes the following form:*

$$\mathbf{G}^\star(t) = \mathbf{b}(t)^H\mathbf{F}(t)^{-1} = \frac{(\alpha\sqrt{P})^{-1}}{f + ||\mathbf{v}(t)||^2}\mathbf{v}(t)^H, \tag{36}$$

where $\mathbf{v}(t)$ *was introduced in* (34) *and* f *was introduced in* (35) *above.*

The proof is in Appendix IV. From this proposition, recalling (30), the following corollary directly follows.

Corollary 2. *When using the MMSE receiver defined by* (36) *in the case of uncorrelated channel coefficients, the MSE can be calculated as follows.*

$$\begin{aligned}&\mathrm{MSE}\left(\mathbf{G}^\star(t), \hat{\mathbf{h}}(t), \hat{\mathbf{h}}(t-1)\right) = 1 - \mathbf{b}(t)^H\mathbf{F}(t)^{-1}\mathbf{b}(t) = \\ &= \frac{f}{f + ||\mathbf{v}(t)||^2}.\end{aligned} \tag{37}$$

The proof of this corollary is in Appendix V.

B. Calculating the Unconditional MSE of the Received Data Symbols

Due to (37), to calculate the unconditional MSE, we first need to determine the distribution of $||\mathbf{v}||^2$. This is derived in the following lemma.

Lemma 3. *The squared norm of the complex random vector* $\mathbf{v}(t)$ *follows the Gamma distribution with parameters* N_r *and* λ*, that is, its probability density function is given as:*

$$f_{||\mathbf{v}||^2}(x) = \frac{\lambda^{N_r}x^{N_r-1}e^{-\lambda x}}{(N_r-1)!},$$

where:

$$\lambda \triangleq \frac{1}{|e_1a^2 + e_2a|^2c + |e_1a + e_2|^2\theta + |e_1|^2\theta + |e_1|^2s + |e_2|^2s}, \tag{38}$$

where s *was defined in* (9) *and* $\theta = c - aca^* = c(1-|a|^2)$.

The proof of this lemma is in Appendix VI. Using the MSE expression (30) and invoking Lemma 3, we can prove the following theorem.

Theorem 1. *When using the MMSE receiver, the unconditional MSE is*

$$MSE = f\lambda e^{f\lambda}E_{in}(N_r, f\lambda), \tag{39}$$

where $E_{in}(n,w) \triangleq \int_{t=1}^{\infty} e^{-wt}/t^n dt$ *is the standard exponential integral function.*

The proof is in Appendix VII.

C. Optimum Pilot Power

According to Theorem 1, the MSE for a given number of antennas N_r is fully determined by the product $f\lambda$. The definition of $\mathbf{F}(t)$ in (35) suggests that the multiuser interference is captured in the term $\sum_{k\neq\kappa} c_k$. Since in this section we are interested in determining the optimal pilot power in the single user case, we start with rewriting the product $f\lambda$ for the single user case, that is when the multiuser interference is zero. Determining the optimum PPR in the multiuser case is an important topic, which is left for future work. In order to preserve symbolic tractability, and to gain engineering insights, in this subsection we assume that the AR parameter a, that is the diagonal elements of $\mathbf{A}$ are real, in which case $aa^* = a^2$. Under this assumption, and using the definition of f in (35) and λ in (38), let us write $f\lambda$ in the following form:

$$f\lambda = \frac{a_2P_p^2 + a_1P_p + a_0}{b_3P_p^3 + b_2P_p^2 + b_1P_p + b_0}, \tag{40}$$

where after some algebraic manipulation, it is straightforward to see that:

$$\begin{aligned}a_0 &= c\alpha^2\sigma^4P_{\text{tot}} + \tau_d\sigma^6 \\ a_1 &= c^2\alpha^4\sigma^2\tau_pP_{\text{tot}} + c\alpha^2\sigma^4(2\tau_d - 1)\tau_p \\ a_2 &= c^2\alpha^4\sigma^2(\tau_d - 1)\tau_p(1-a^2)\end{aligned} \tag{41}$$

and

$$\begin{aligned}b_0 &= 0 \\ b_1 &= c^2\alpha^4\sigma^2\tau_pP_{\text{tot}} \\ b_2 &= c^3\alpha^6\tau_p^2P_{\text{tot}} - c^2\alpha^4\sigma^2\tau_p^2(a^2+1) \\ b_3 &= c^3\alpha^6\tau_p^3(a^2-1).\end{aligned} \tag{42}$$

Now we can state the following proposition, which, together with the subsequent remark, provides a useful engineering insight.

Proposition 3. *In a single user MIMO system, when $a \in \mathbb{R}$, if the following quartic equation has positive real roots, then the optimal pilot power is in the set of such positive real roots:*

$$c_0 + c_1 P_p + c_2 P_p^2 + c_3 P_p^3 + c_4 P_p^4 = 0, \tag{43}$$

where

$$\begin{aligned}
c_0 &= (c\alpha^2 P_{\text{tot}} + \sigma_d^2 \tau_d)\sigma_p^6 P_{\text{tot}}(1+a^2) \\
c_1 &= 2\left(c\alpha^2 P_{\text{tot}}(1-a^2) - (a^2+1)\sigma_p^2\right)\left(c\alpha^2 P_{\text{tot}} + \sigma_d^2\tau_d\right)\sigma_p^4\tau_p \\
c_2 &= c\alpha^2\sigma_p^2\tau_p^2\Big(c^2\alpha^4 P_{\text{tot}}^2(a^2-1)^2 + \\
&\quad + c\alpha^2 P_{\text{tot}}(a^2-1)\left(4\sigma_p^2 + (a^2-1)\sigma_d^2\tau_d\right) + \\
&\quad - \sigma_p^2\sigma_d^2\tau_d\left(5-a^2\right) + (1+a^2)\sigma_p^4\Big). \\
c_3 &= 2c^2\alpha^4\sigma_p^2\tau_p^3(1-a^2)\cdot \\
&\quad \cdot\left(\sigma_p^2 - 2\sigma_d^2\tau_d - c\alpha^2 P_{\text{tot}}(1-a^2)\right). \\
c_4 &= (1-a^2)^2 c^3\alpha^6\left(\sigma_p^2 - \sigma_d^2\tau_d\right)\tau_p^4.
\end{aligned} \tag{44}$$

The proof is in Appendix VIII. Finding the roots of (43) is straightforward both symbolically and numerically. Notice that a direct consequence of Proposition 3 is that the optimal pilot power does not depend on the number of antennas (N_r), since the coefficients of (43) and (44) do not contain N_r. On the other hand, the coefficients clearly depend on the large scale fading α and AR parameter a. As expected, the achieved MSE clearly depends on N_r, through the $f\lambda$ term as expressed in (40).

Remark 3. *The number of real roots and their respective signs of the real-coefficient quartic equation is determined by its discriminant [49], [50], and it is difficult to show symbolically that* (43) *has exactly one positive root. However, notice that c_0 is positive, implying that the quartic polynomial in* (43) *is positive when $P_p = 0$. Furthermore, c_4 is negative, since $\sigma_p \approx \sigma_d$ and $\tau_d > 1$, implying that the quartic polynomial tends to negative infinity when P_p tends to the negative or positive infinity. This in turn implies that* (43) *has at least one negative and at least one positive root.*

Furthermore, by numerical experiments we found that the discriminant of the first derivative of the polynomial in (43) with respect to P_p, which is a cubic polynomial, is negative in all cases that are meaningful from an engineering point of view. This implies that the quartic polynomial in (43) has only one extreme value, which is a maximum due to the fact the polynomial tends to negative infinity in both the negative and the positive infinities. This suggests that the quartic in (43) has one negative, one positive and two complex roots (which are conjugate complex pairs). This finding, using numerical experiments, is also confirmed by observing that the second derivative of the quartic in equation (43) is negative in practically relevant cases. This implies that it is always concave, which, together with Remark 3, implies that it has two distinct real roots, one of which is positive, giving the optimum pilot power.

VI. Calculating the MSE in the Case of Correlated Channel Coefficients

A. Calculating the MSE as a Function of the Estimated Channels

To calculate the MSE when the channel coefficients are correlated, that is when the covariance matrix of $\mathbf{h}$ contains off-diagonal elements, the following corollary of Proposition 1 will be useful.

Corollary 3. *When the receiver uses the $\mathbf{G}^\star(t)$ MMSE receiver, the achieved MSE is as follows.*

$$\begin{aligned}
&\min_{\mathbf{G}} \text{MSE}\left(\mathbf{G}, \hat{\mathbf{h}}(t), \hat{\mathbf{h}}(t-1)\right) = \\
&\quad = \text{MSE}\left(\mathbf{G}^\star(t), \hat{\mathbf{h}}(t), \hat{\mathbf{h}}(t-1)\right) = \frac{1}{1+\omega(t)},
\end{aligned} \tag{45}$$

where $\omega(t) = \mathbf{b}(t)^H \mathbf{M}^{-1}\mathbf{b}(t)$.

Proof. Substituting $\mathbf{G}^\star(t) = \mathbf{b}(t)^H \mathbf{F}(t)^{-1}$ into (67) in Appendix III, and using (28) we get:

$$\begin{aligned}
&\text{MSE}\left(\mathbf{G}^\star(t), \hat{\mathbf{h}}(t), \hat{\mathbf{h}}(t-1)\right) = \\
&\quad = 1 - \mathbf{b}(t)^H \mathbf{F}(t)^{-1}\mathbf{b}(t) = \\
&\quad = 1 - \mathbf{b}(t)^H \left(\mathbf{b}(t)\mathbf{b}(t)^H + \mathbf{M}\right)^{-1}\mathbf{b}(t).
\end{aligned} \tag{46}$$

Using the Sherman-Morrison formula, we further have

$$\begin{aligned}
&\text{MSE}\left(\mathbf{G}^\star(t), \hat{\mathbf{h}}(t), \hat{\mathbf{h}}(t-1)\right) = \\
&= 1 - \mathbf{b}(t)^H \left(\mathbf{b}(t)\mathbf{b}(t)^H + \mathbf{M}\right)^{-1}\mathbf{b}(t) \\
&= 1 - \mathbf{b}(t)^H \left(\mathbf{M}^{-1} - \frac{\mathbf{M}^{-1}\mathbf{b}(t)\mathbf{b}(t)^H\mathbf{M}^{-1}}{1+\mathbf{b}(t)^H\mathbf{M}^{-1}\mathbf{b}(t)}\right)\mathbf{b}(t) \\
&= 1 - \left(\omega(t) - \frac{\omega(t)^2}{1+\omega(t)}\right) = \\
&= 1 - \left(\frac{\omega(t)}{1+\omega(t)}\right) = \frac{1}{1+\omega(t)}.
\end{aligned} \tag{47}$$

□

Remark 4. *We can rewrite* (47) *in the following form:*

$$\begin{aligned}
&\text{MSE}\left(\mathbf{G}^\star(t), \hat{\mathbf{h}}(t), \hat{\mathbf{h}}(t-1)\right) = \\
&\quad = \frac{1}{1+\alpha^2 P(\mathbf{E}\boldsymbol{\zeta}(t))^H(\alpha^2 P\mathbf{Z} + \sigma_d^2\mathbf{I}_{N_r})^{-1}(\mathbf{E}\boldsymbol{\zeta}(t))} = \\
&\quad = \frac{1}{1+\alpha^2 P||\boldsymbol{\nu}(t)||^2},
\end{aligned} \tag{48}$$

where

$$\boldsymbol{\nu}(t) \triangleq \left(\alpha^2 P\mathbf{Z} + \sigma_d^2\mathbf{I}_{N_r}\right)^{-1/2}\mathbf{E}\boldsymbol{\zeta}(t). \tag{49}$$

Notice that in the case of perfect channel estimation (i.e. $\mathbf{E}\boldsymbol{\zeta}(t) = \mathbf{h}(t)$ and $\mathbf{Z} = \mathbf{0}$), from (48) *we get:*

$$\text{MSE}(\mathbf{h}(t)) = \frac{1}{1+\alpha^2 P\frac{||\mathbf{h}(t)||^2}{\sigma_d^2}}, \tag{50}$$

which indicates that $\boldsymbol{\nu}(t)$ in (49) *can be considered an equivalent channel and is in line with our expectation that*

$\alpha^2 P||\boldsymbol{\nu}(t)||^2$ *is an equivalent signal-to-noise ratio (SNR) when using the proposed* $\mathbf{G}^\star(t)$ *receiver.*

B. Calculating the Unconditional MSE

Due to (45), to calculate the unconditional MSE, we first need to determine the distribution of $\omega(t)$. This is derived in this section. As it was introduced in (16), we have that:

$$\begin{aligned}\mathbf{E} &= \left[\begin{array}{cc}\mathbf{E_1} & \mathbf{E_2}\end{array}\right] \\ &= \left[\begin{array}{cc}\mathbf{C} & \mathbf{AC}\end{array}\right]\left[\begin{array}{cc}\mathbf{C}+\mathbf{\Sigma} & \mathbf{AC} \\ \mathbf{CA}^H & \mathbf{C}+\mathbf{\Sigma}\end{array}\right]^{-1}. \end{aligned} \quad (51)$$

Then, considering (27),

$$\mathbf{b}(t) = \alpha\sqrt{P}\mathbf{E}\boldsymbol{\zeta}(t) = \alpha\sqrt{P}\left(\mathbf{E_1}\hat{\mathbf{h}}(t) + \mathbf{E_2}\hat{\mathbf{h}}(t-1)\right), \quad (52)$$

and

$$\begin{aligned}&\mathbf{E_1}\hat{\mathbf{h}}(t) + \mathbf{E_2}\hat{\mathbf{h}}(t-1) \\ &= \mathbf{E_1}(\mathbf{h}(t) + \mathbf{s}(t)) + \mathbf{E_2}(\mathbf{h}(t-1) + \mathbf{s}(t-1)) \\ &= \mathbf{E_1}(\mathbf{A}\mathbf{h}(t-1) + \boldsymbol{\vartheta}(t) + \mathbf{s}(t)) + \\ &\quad + \mathbf{E_2}(\mathbf{A}\mathbf{h}(t-2) + \boldsymbol{\vartheta}(t-1) + \mathbf{s}(t-1)) \\ &= \mathbf{E_1}(\mathbf{A}^2\mathbf{h}(t-2) + \mathbf{A}\boldsymbol{\vartheta}(t-1) + \boldsymbol{\vartheta}(t) + \mathbf{s}(t)) + \\ &\quad + \mathbf{E_2}(\mathbf{A}\mathbf{h}(t-2) + \boldsymbol{\vartheta}(t-1) + \mathbf{s}(t-1)) \\ &= (\mathbf{E_1}\mathbf{A}^2 + \mathbf{E_2}\mathbf{A})\mathbf{h}(t-2) + (\mathbf{E_1}\mathbf{A} + \mathbf{E_2})\boldsymbol{\vartheta}(t-1) + \\ &\quad + \mathbf{E_1}\boldsymbol{\vartheta}(t) + \mathbf{E_1}\mathbf{s}(t) + \mathbf{E_2}\mathbf{s}(t-1). \end{aligned} \quad (53)$$

From this expression, the covariance of $\mathbf{b}(t)$, denoted as $\mathbf{B}$, can be obtained as

$$\begin{aligned}&\mathbf{B}(\alpha^2 P)^{-1} = \mathbb{E}(\mathbf{b}(t)\mathbf{b}(t)^H)(\alpha^2 P)^{-1} \\ &= (\mathbf{E_1}\mathbf{A}^2 + \mathbf{E_2}\mathbf{A})\mathbf{C}(\mathbf{E_1}\mathbf{A}^2 + \mathbf{E_2}\mathbf{A})^H + \\ &\quad (\mathbf{E_1}\mathbf{A} + \mathbf{E_2})\mathbf{\Theta}(\mathbf{E_1}\mathbf{A} + \mathbf{E_2})^H + \\ &\quad + \mathbf{E_1}\mathbf{\Theta}\mathbf{E_1}^H + \mathbf{E_1}\mathbf{\Sigma}\mathbf{E_1}^H + \mathbf{E_2}\mathbf{\Sigma}\mathbf{E_2}^H. \end{aligned} \quad (54)$$

Since $\mathbf{B}$ is Hermitian positive definite, we can use its Cholesky decomposition as $\mathbf{B} = \mathbf{U}\mathbf{U}^H$, and write:

$$\omega(t) = \mathbf{b}(t)^H\mathbf{M}^{-1}\mathbf{b}(t) = \mathbf{v}(t)^H\mathbf{U}^H\mathbf{M}^{-1}\mathbf{U}\mathbf{v}(t), \quad (55)$$

where the elements of $\mathbf{v}(t)$ are independent complex normal random variables, i.e., $\mathbf{v}(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$. However, $\mathbf{M}$ in general is not diagonal. We would like to write $\omega(t)$ as the sum of (the squared norms of) independent random variables. Therefore, let

$$\mathbf{U}^H\mathbf{M}^{-1}\mathbf{U} = \mathbf{V}^H\mathbf{\Gamma}\mathbf{V} \quad (56)$$

be the singular value decomposition of $\mathbf{U}^H\mathbf{M}^{-1}\mathbf{U}$, i.e. $\mathbf{V}$ is unitary and $\mathbf{\Gamma}$ is diagonal. Then

$$\omega(t) = \mathbf{v}(t)^H\mathbf{U}^H\mathbf{M}^{-1}\mathbf{U}\mathbf{v}(t) = \mathbf{v}(t)^H\mathbf{V}^H\mathbf{\Gamma}\mathbf{V}\mathbf{v}(t), \quad (57)$$

where the elements of $\mathbf{V}\mathbf{v}(t)$ are independent complex normal random variables:

$$\mathbf{V}\mathbf{v} \sim \mathcal{CN}\Big(\mathbf{0}, \underbrace{\mathbf{V}\mathbf{V}^H}_{\mathbf{I_{N_r}}}\Big), \quad (58)$$

Therefore,

$$\omega(t) = \sum_{i=1}^{N_r}\gamma_i \underbrace{\left|(\mathbf{V}\mathbf{v}(t))_i\right|^2}_{\triangleq|\omega(t)_i|^2}, \quad (59)$$

where γ_i is the i^{th} singular value (diagonal element) in $\mathbf{\Gamma}$ and each $|\omega(t)_i|^2$ is exponentially distributed with parameter 1.

C. Calculating the Unconditional MSE

Since $|\omega(t)_i|^2$ is exponentially distributed with parameter 1, $\gamma_i|\omega(t)_i|^2$ is exponentially distributed with parameter $\lambda_i \triangleq 1/\gamma_i$, and $\sum_{i=1}^{N_r}\gamma_i|\omega(t)_i|^2$ follows the phase type distribution with density function

$$f(x) = \mathbf{e_1}^T e^{\mathbf{A}x}\mathbf{e}_{N_r}\lambda_{N_r},$$

where $\mathbf{e_i}$ is the i-th unit vector (whose only nonzero element is 1 at position i), and the matrix $\mathbf{A}$ is given as:

$$\mathbf{A} = \begin{pmatrix} -\lambda_1 & \lambda_1 & & \\ & -\lambda_2 & \lambda_2 & \\ & & \ddots & \ddots \\ & & & -\lambda_{N_r} \end{pmatrix}.$$

In the special case when all singular values are identical ($\gamma_i = \gamma$), $\omega(t)$ is Gamma distributed with parameters N_r and $\lambda = 1/\gamma$, as in (38). In the other special case when all singular values are distinct, $\omega(t)$ is hypo-exponentially distributed with parameters $\lambda_i = 1/\gamma_i, (i = 1, \ldots, N_r)$ and

$$f(x) = \sum_{i=1}^{N_r} c_i\lambda_i e^{-\lambda_i x}, \quad (60)$$

where $c_i = \prod_{j=1, j\neq i}^{j=N_r}\frac{\lambda_j}{\lambda_j - \lambda_i}$.

Based on $f(x)$ and (59), the MSE can be calculated as follows:

$$\text{MSE} = \int_x \frac{1}{\alpha^2 P x + 1} f(x)dx, \quad (61)$$

which has the form of (39) if all singular values are identical, and the form of

$$\text{MSE} = \sum_{i=1}^{N_r} c_i \frac{\lambda_i}{\alpha^2 P} e^{\frac{\lambda_i}{\alpha^2 P}} E_{in}\left(1, \frac{\lambda_i}{\alpha^2 P}\right), \quad (62)$$

if all singular values are distinct.

VII. Numerical Results

Table II
System Parameters

Parameter	Value
Autoregressive parameter	$a = 0, 0.1, \ldots 0.95$
Number of antennas	$N_r = 20, 100$
Path Loss of tagged user	$\alpha = 90$ dB
Number of pilot and data symbols	$\tau_p = 1;\ \tau_d = 11$
Power budget	$\tau_p P_p + \tau_d P = P_{tot}$ =250 mW.
MIMO receivers	Naive, conventional [19] and proposed MMSE, see (26)

In this section we consider a single cell single user MIMO system, in which the mobile terminal is equipped with a single

transmit antenna, whereas the BS employs N_r receive antennas and we assume the case of independent and identically distributed channel coefficients from Section V. Note that the performance characteristics of the proposed MMSE receiver operating with Kalman filtering assisted channel estimation as compared with the naïve receiver are similar in the multi-user MIMO case from the perspective of the tagged user, since the proposed receiver treats the multi-user interference as noise according to (28). The key input parameters to this system that are necessary to obtain numerical results using the MSE derivation in this paper (ultimately relying on Theorem 1 and Proposition 3) are listed in Table II. For benchmarking purposes, we also examine the performance of the conventional MMSE receiver that does not take advantage of autoregressive property of the channel, that is it relies on the current estimate $\hat{\mathbf{h}}(t)$ [19], as opposed to the proposed MMSE receiver that uses Kalman filtering and thereby it utilizes both $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{h}}(t-1)$. In this section, we refer to these three receiver structures as the 'naive', 'conventional' and the proposed 'MMSE' receivers. According to Remarks 1 and 2, 'MMSE' is identical with 'conventional' when $a = 0$.

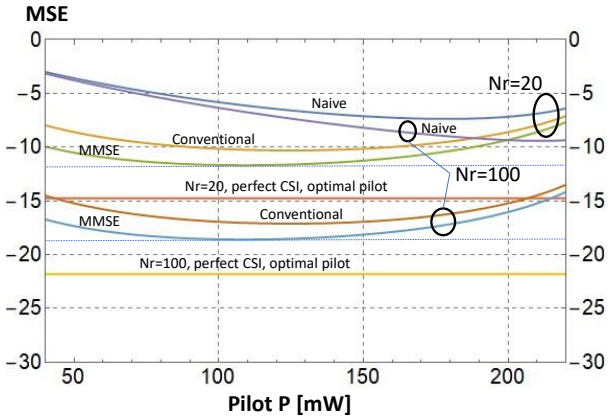

Figure 1. MSE as a function of the pilot power (P_p) when using the naive and conventional receivers and the proposed MMSE receiver. As a benchmark, the figure also indicates the MSE that is achieved when using perfect CSI with $N_r = 20$ and $N_r = 100$ antennas and assuming $a = 0.95$.

Figure 1 shows the MSE as a function of the pilot power under the constraint of (2). The upper 3 curves, marked with $N_r = 20$ show the performance with $N_r = 20$ receive antennas, while the lower 3 curves show the results obtained with $N_r = 100$ antennas. Notice that the gain by using the proposed MMSE receiver is much higher when the number of antennas is large. This result may be counterintuitive at first sight, since the performance gain by employing sophisticated precoders in massive MIMO transmissions in the downlink decreases with an increasing number of antennas. However, as it is clearly visible in Figure 1, in the presence of CSIR errors, a receiver that compensates for the CSIR errors is increasingly superior to poor receiver designs. This figure also shows the 'ultimate' MSE performance with a given number of antennas, that is when perfect CSIR is available and the pilot power is set to its optimal value (discussed further in conjunction with Figures 3-4).

Figure 2 focuses on the performance of the proposed MMSE

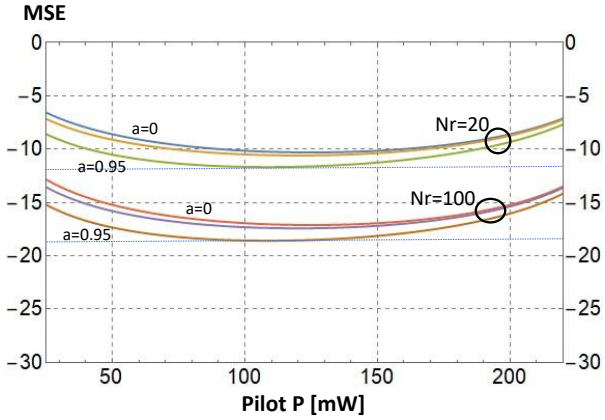

Figure 2. MSE as a function of the pilot power (P_p) when using the proposed MMSE receiver. with $N_r = 20$ and $N_r = 100$ antennas. As the autoregressive parameter increases from 0 (upper curve) to 0.5 (middle curve) and 0.95 (lower curve), the MSE decreases in both cases.

receiver when the autoregressive parameter of the channel is $a = 0$, $a = 0.5$ and $a = 0.95$. When $a = 0$, the MMSE receiver performance is identical with that of the conventional MMSE receiver. Similarly to Figure 1, we are interested in the MSE performance as a function of the pilot power under the constraint of (2). Here we can observe that higher a values, which characterize the memoryfull property of the evolution of the channel, enable to reduce the MSE. This MSE gain is present with both $N_r = 20$ and $N_r = 100$ number of antennas.

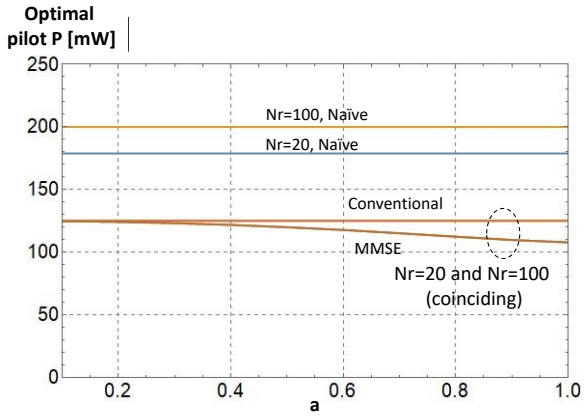

Figure 3. Optimum pilot power as a function of the autoregressive parameter when using the naive, conventional and MMSE receivers. with $N_r = 20$ and $N_r = 100$ antennas. Notice that in the case of the conventional and MMSE receivers, the optimum pilot power curves for $N_r = 20$ and $N_r = 100$ overlap. When $a = 0$, the optimum pilot setting of the MMSE receiver becomes identical with those of the conventional receiver.

Figure 3 examines the optimal setting of the pilot power as a function of a. As it is visible already in Figure 1, when using the naive receiver, somewhat higher pilot power is needed to minimize the MSE with $N_r = 100$ antennas then with $N_r = 20$ antennas. In this example, we need to increase the pilot power from around 160 mW to around 180 mW in order to minimize the MSE when using the naive receiver. In contrast, a much lower pilot power (around 125 mW at $a = 0$ and around 110 mW at $a = 0.95$) is sufficient with the proposed MMSE receiver. Interestingly, as it was discussed after Proposition

3, this optimum pilot power does not depend on number of antennas, therefore the curves for both the conventional and the MMSE receiver with $N_r = 20$ and $N_r = 100$ coincide.

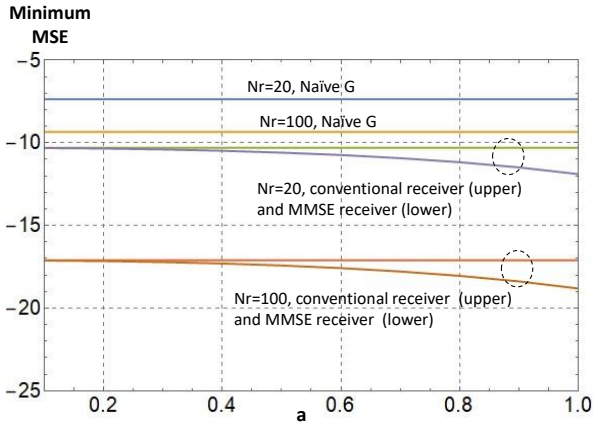

Figure 4. Minimum MSE, that is when using the optimum pilot power, as a function of the autoregressive parameter and when using $N_r = 20$ and $N_r = 100$ antennas. Notice that at high a, and when using the proposed MMSE receiver, and the optimum pilot power, the achieved MSE with $N_r = 20$ antennas is lower than when using the naive receiver with $N_r = 100$ antennas.

Figure 4 examines the achievable minimum MSE, that is the MSE that can be achieved by appropriately setting the pilot power and thereby the PPR. First, notice that the gain in terms of the achievable minimum MSE when using the MMSE receiver is larger with $N_r = 100$ than with $N_r = 20$. Also, notice that when using a properly designed receiver, the minimum achievable MSE with a moderate number of antennas can be less than the achievable MSE with a large number of antennas if the autoregressive parameter of the channel is high. However, even with low a, the system with $N_r = 20$ antennas using the proposed MMSE receiver performs similarly to the system with a large number of antennas using the naive MIMO receiver. It is very important to realize that this insight applies to the lowest achievable MSE, that is when the PPR is set optimally according to Figure 3. When $a = 0$, the performance and optimum pilot setting of the MMSE receiver becomes identical with those of the conventional receiver.

Figure 5 focuses on the impact of improved channel estimation on the channel estimation noise, which is an important metric of the available CSIR. We can see that when we exploit the autoregressive evolution of the channel by employing the proposed MMSE receiver, the channel estimation noise, as defined in (16), decreases. By way of example, notice that if the wireless channel evolves with a strong autocorrelation, we can reduce the channel estimation noise by employing proper receiver design, which is equivalent with significantly increasing the pilot power (by around 20-30 mW).

Figure 6 shows the cumulative distribution function (CDF) of the spectral efficiency that can be achieved when using the naive, conventional and MMSE receivers at the base station when it employs $N_r = 100$ receive antennas. This figure shows that the MSE gain harvested by the MMSE receiver when the channel can be modeled as an autoregressive process carries over to a spectral efficiency gain over the naive and

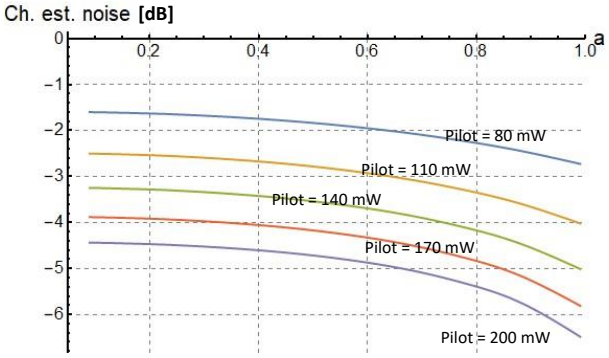

Figure 5. Channel estimation noise ($\mathbf{Z}$ as defined in (16), which can be compared with $\mathbf{Q}$), as a function of the autoregressive parameter. This figure indicates that the channel estimation noise (diagonal elements of $\mathbf{Z}$) decreases at higher a values, which indicates the availability of a higher quality CSIR as t

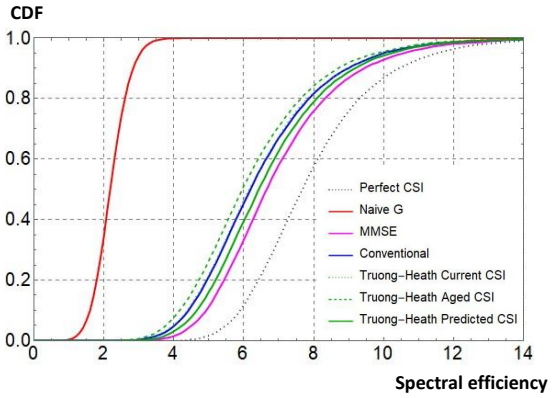

Figure 6. Cumulative distribution function (CDF) of the achievable spectral efficiency with $N_r = 100$ antennas when the AR parameter of the channel is $a = 0.9$ and the pilot power is set to 125 mW. The proposed MMSE receiver achieves higher spectral efficiency over the entire CDF region than previously proposed schemes, including the schemes proposed by Truong and Heath in [29].

conventional receivers.

In some settings the parameters of the AR process might not be known by receiver and some estimation of a must be used [40], [41]. It is therefore a question of interest how robust the performance of the MMSE receiver is to estimation errors. Figure 7 shows the average spectral efficiency of the MMSE receiver as a function of both the actual AR parameter and the estimated value of the parameter. It is compared to the average spectral efficiency of the conventional receiver, that is when the estimated value of a is equal to 0. Notice that for any value of a the marginal distribution of $\mathbf{h}(t)$ is $\mathcal{CN}(\mathbf{0}, \mathbf{C})$. Hence the achieved spectral efficiency does not depend on a when the conventional receiver is used, which motivates the comparison in Figure 7. As expected, when $a = 0$ the performances of the conventional and the MMSE receivers coincide. Furthermore, when $a > 0.6$, any estimate of a gives a better performance than the conventional receiver, and if $a = 0.4$ any estimate lower than 0.7 also gives a better performance. Hence, the MMSE receiver is robust to estimation errors of the AR parameter.

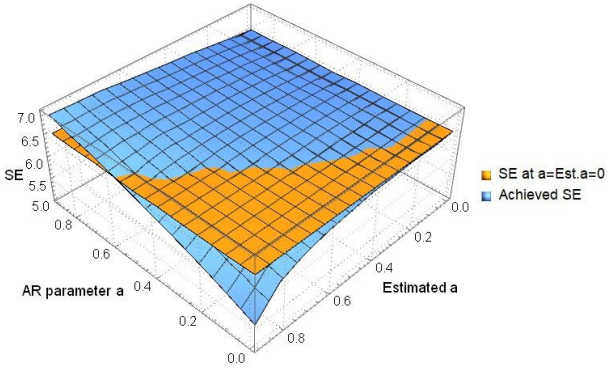

Figure 7. Achieved spectral efficiency (SE) when the system employs $N_r = 100$ antennas as a function of the actual AR parameter a and the estimated a parameter at the receiver. The spectral efficiency is highest when a is close to 1 and the estimated a is an accurate estimate. The figure also indicates the achieved spectral efficiency when $a = 0$ and it is known perfectly at the receiver (that is when the system employs the conventional receiver).

VIII. Conclusions and Outlook

In this paper, we made the point that exploiting the memoryfull property of autoregressive fading channels enables to reduce the MSE of the received data symbols. To achieve this MSE reduction, not only the channel estimation process must be adjusted, but also the receiver structure must take into account the statistical properties of the estimated channel. Specifically, when the autoregressive channel is estimated by a Kalman filter, a linear receiver that minimizes the MSE over all linear receivers was proposed. This new MMSE receiver exploits the memoryfull property of the fading channel and provides better performance than both the naive receiver, which assumes that the erroneous channel estimate can be used as if it was the actual channel, and a previously proposed receiver that does not utilize multiple and correlated channel estimates. We derived a closed form expression for calculating the MSE, which helps to properly set the pilot-to-data power ratio, which is also important for reaching high performance. These results extend previous results in the literature concerning receiver designs and tuning the pilot and data power levels. As noted in Subsection V.C, determining the optimum PPR in the multiuser case is left for future work. Other important future works include investigating the performance of the proposed MMSE receiver in other propagation conditions, determining the autoregressive model parameters in the presence of user mobility and analyzing the sensitivity of the performance with regards to model parameter estimation errors. An important research direction is to pose the question whether and how exploiting the memoryfull property can be integrated into schemes that employ data-aided channel estimation and iterative linear MMSE detection, such as those proposed in [22]–[25].

Appendix I

Proof of Lemma 1. The joint covariance matrix of vectors $\mathbf{h}(t)$, $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{h}}(t-1)$ is by definition:

$$\mathbf{\Psi} \triangleq \begin{bmatrix} \mathbf{C}_{\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t)} & \mathbf{C}_{\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)} & \mathbf{C}_{\hat{\mathbf{h}}(t),\mathbf{h}(t)} \\ \mathbf{C}_{\hat{\mathbf{h}}(t-1),\hat{\mathbf{h}}(t)} & \mathbf{C}_{\hat{\mathbf{h}}(t-1),\hat{\mathbf{h}}(t-1)} & \mathbf{C}_{\hat{\mathbf{h}}(t-1),\mathbf{h}(t)} \\ \mathbf{C}_{\mathbf{h}(t),\hat{\mathbf{h}}(t)} & \mathbf{C}_{\mathbf{h}(t),\hat{\mathbf{h}}(t-1)} & \mathbf{C}_{\mathbf{h}(t),\mathbf{h}(t)} \end{bmatrix}. \tag{63}$$

Using the definitions of the respective covariance matrices, and utilizing that $\mathbb{E}\left(\mathbf{w}(t)\hat{\mathbf{h}}(t-1)^H\right) = \mathbb{E}\left(\mathbf{w}(t)\left(\mathbf{h}(t-1)+\mathbf{w}(t-1)\right)^H\right) = \mathbf{0}$, the lemma follows. □

Appendix II

Proof of Lemma 2. Substituting $\mathbf{\Psi}$, as defined in (15) into equations (10.24) and (10.25) of [47], we get

$$\mathbb{E}\left(\mathbf{h}(t)|\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)\right) = \underbrace{\mathbb{E}\left(\mathbf{h}(t)\right)}_{\mathbf{0}} + \underbrace{+\begin{bmatrix} \mathbf{C} & \mathbf{AC} \end{bmatrix}\begin{bmatrix} \mathbf{C}+\mathbf{\Sigma} & \mathbf{AC} \\ \mathbf{CA}^H & \mathbf{C}+\mathbf{\Sigma} \end{bmatrix}^{-1}}_{\triangleq \mathbf{E}\ \in \mathbb{C}^{N_r \times 2N_r}} \cdot \boldsymbol{\zeta}(t),$$

and

$$\mathbf{C}_{\mathbf{h}(t)|\hat{\mathbf{h}}(t)\hat{\mathbf{h}}(t-1)} = \mathbf{C} - \mathbf{E}\begin{bmatrix} \mathbf{C} \\ \mathbf{CA}^H \end{bmatrix} \triangleq \mathbf{Z}\ \in \mathbb{C}^{N_r \times N_r},$$

where $\mathbf{E}$ is defined in (64). □

Appendix III

Proof of Proposition 1. We will make use of the following

$$\mathbb{E}\left(\mathbf{h}(t)|\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)\right) = \mathbf{E}\boldsymbol{\zeta}(t), \tag{64}$$

and

$$\mathbb{E}\left(\mathbf{h}(t)\mathbf{h}(t)^H|\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)\right) = \mathbf{E}\boldsymbol{\zeta}(t)\boldsymbol{\zeta}(t)^H\mathbf{E}^H + \mathbf{Z}. \tag{65}$$

Using these expectations and recalling (25), we can calculate the MSE:

$$\begin{aligned} &\mathrm{MSE}\left(\mathbf{G},\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)\right) = \mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)}\{\mathrm{MSE}\left(\mathbf{G},\mathbf{h}\right)\} = \\ &= \alpha^2 P\mathbf{G}\left(\mathbf{E}\boldsymbol{\zeta}(t)\boldsymbol{\zeta}(t)^H\mathbf{E}^H + \mathbf{Z}\right)\mathbf{G}^H + \\ &\quad + \sigma_d^2\mathbf{G}\mathbf{G}^H + \sum \alpha^2 P\mathbf{G}\mathbf{C}\mathbf{G}^H + 1 - \\ &\quad - \alpha\sqrt{P}\left(\mathbf{G}\mathbf{E}\boldsymbol{\zeta}(t) + \boldsymbol{\zeta}(t)^H\mathbf{E}^H\mathbf{G}^H\right), \end{aligned} \tag{66}$$

which can be written in the following quadratic form:

$$\begin{aligned} &\mathrm{MSE}\left(\mathbf{G},\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)\right) = \\ &= \mathbf{G}\mathbf{F}(t)\mathbf{G}^H - \mathbf{b}(t)^H\mathbf{G}^H - \mathbf{G}\mathbf{b}(t) + 1. \end{aligned} \tag{67}$$

The $\mathbf{G}$ that minimizes this quadratic form is $\mathbf{G}^\star(t) = \mathbf{b}(t)^H\mathbf{F}(t)^{-1}$. □

APPENDIX IV

Proof of Proposition 2. Starting from (35) and applying the Sherman-Morrison formula, we get:

$$\mathbf{F}(t)^{-1} = \frac{1}{\alpha^2 P}\left(f^{-1}\mathbf{I}_{N_r} - \frac{f^{-1}\mathbf{v}(t)\mathbf{v}(t)^H}{f+\mathbf{v}(t)^H\mathbf{v}(t)}\right) = \frac{1}{\alpha^2 P f}\left(\mathbf{I}_{N_r} - \frac{\mathbf{v}(t)\mathbf{v}(t)^H}{f+||\mathbf{v}(t)||^2}\right). \quad (68)$$

Substituting this $\mathbf{F}(t)^{-1}$ into (26), we get:

$$\mathbf{G}^\star(t) = \frac{1}{\alpha\sqrt{P}f}\mathbf{v}(t)^H\left(\mathbf{I}_{N_r} - \frac{\mathbf{v}(t)\mathbf{v}(t)^H}{f+||\mathbf{v}(t)||^2}\right) = \frac{1}{\alpha\sqrt{P}f}\left(\mathbf{v}(t)^H - \frac{||\mathbf{v}(t)||^2\mathbf{v}(t)^H}{f+||\mathbf{v}(t)||^2}\right)$$

$$= \frac{1}{\alpha\sqrt{P}f}\left(1-\frac{||\mathbf{v}(t)||^2}{f+||\mathbf{v}(t)||^2}\right)\mathbf{v}(t)^H = \frac{1}{\alpha\sqrt{P}f}\left(\frac{f}{f+||\mathbf{v}(t)||^2}\right)\mathbf{v}(t)^H = \frac{(\alpha\sqrt{P})^{-1}}{f+||\mathbf{v}(t)||^2}\mathbf{v}(t)^H, \quad (69)$$

where

$$||\mathbf{v}(t)||^2 = |e_1|^2||\hat{\mathbf{h}}(t)||^2 + |e_2|^2||\hat{\mathbf{h}}(t-1)||^2 + + e_1e_2^*\hat{\mathbf{h}}^H(t-1)\hat{\mathbf{h}}(t) + e_1^*e_2\hat{\mathbf{h}}(t)^H\hat{\mathbf{h}}(t-1). \quad (70)$$

□

APPENDIX V

Proof of Corollary 2. Starting from the MSE expression of (30) and substituting $\mathbf{b}(t)$ and $\mathbf{F}(t)^{-1}$ from (33) and (68) respectively, we get:

$$\text{MSE}\left(\mathbf{G}^\star(t),\hat{\mathbf{h}}(t),\hat{\mathbf{h}}(t-1)\right) = 1-\mathbf{b}(t)^H.\mathbf{F}(t)^{-1}.\mathbf{b}(t)$$

$$= 1 - f^{-1}\mathbf{v}(t)^H\left(\mathbf{I}-\frac{\mathbf{v}(t)\mathbf{v}(t)^H}{f+||\mathbf{v}(t)||^2}\right)\mathbf{v}(t)$$

$$= 1 - f^{-1}\left(||\mathbf{v}(t)||^2 - \frac{||\mathbf{v}(t)||^2||\mathbf{v}(t)||^2}{f+||\mathbf{v}(t)||^2}\right)$$

$$= 1 - f^{-1}||\mathbf{v}(t)||^2\left(1-\frac{||\mathbf{v}(t)||^2}{f+||\mathbf{v}(t)||^2}\right)$$

$$= 1 - f^{-1}||\mathbf{v}(t)||^2\left(\frac{f}{f+||\mathbf{v}(t)||^2}\right)$$

$$= 1 - \frac{||\mathbf{v}(t)||^2}{f+||\mathbf{v}(t)||^2} = \frac{f}{f+||\mathbf{v}(t)||^2}$$

□

APPENDIX VI

Proof of Lemma 3. To determine the distribution of $||\mathbf{v}(t)||$, let us first focus on the distribution of $\mathbf{v}(t)$. To this end, we write:

$$\begin{aligned}\mathbf{v}(t) &= e_1\hat{\mathbf{h}}(t) + e_2\hat{\mathbf{h}}(t-1)\\ &= e_1(\mathbf{h}(t)+\mathbf{s}(t)) + e_2(\mathbf{h}(t-1)+\mathbf{s}(t-1))\\ &= e_1(\mathbf{A}\mathbf{h}(t-1)+\boldsymbol{\vartheta}(t)+\mathbf{s}(t)) + \\ &\quad + e_2(\mathbf{A}\mathbf{h}(t-2)+\boldsymbol{\vartheta}(t-1)+\mathbf{s}(t-1))\\ &= e_1(\mathbf{A}^2\mathbf{h}(t-2)+\mathbf{A}\boldsymbol{\vartheta}(t-1)+\boldsymbol{\vartheta}(t)+\mathbf{s}(t)) + \\ &\quad + e_2(\mathbf{A}\mathbf{h}(t-2)+\boldsymbol{\vartheta}(t-1)+\mathbf{s}(t-1))\\ &= (e_1\mathbf{A}^2+e_2\mathbf{A})\mathbf{h}(t-2) + (e_1\mathbf{A}+e_2)\boldsymbol{\vartheta}(t-1) + \\ &\quad + e_1\boldsymbol{\vartheta}(t) + e_1\mathbf{s}(t) + e_2\mathbf{s}(t-1), \end{aligned} \quad (71)$$

where $\mathbf{h}(t-2)\sim\mathcal{CN}(\mathbf{0},\mathbf{C})$, $\boldsymbol{\vartheta}(i)\sim\mathcal{CN}(\mathbf{0},\boldsymbol{\Theta})$, and $\mathbf{s}(i)\sim\mathcal{CN}(\mathbf{0},\boldsymbol{\Sigma})$ and $\mathbf{A}=a\mathbf{I}_{N_r}$. That is:

$$\mathbf{v}(t)\sim\mathcal{CN}(\mathbf{0},c^*\mathbf{I}_{N_r})$$

with

$$c^* = |e_1a^2+e_2a|^2c + |e_1a+e_2|^2\theta + |e_1|^2\theta + |e_1|^2s + + |e_2|^2s, \quad (72)$$

where, due to (4) and (5), $\theta = c(1-|a|^2)$.

Therefore, each $|v_i|^2$ is exponentially distributed, with parameter $\lambda \triangleq \frac{1}{c^*}$, and Therefore, $||\mathbf{v}(t)||^2 = \sum_i |\mathbf{v}(t)_i|^2$, which is the sum of N_r such independent and identically distributed random variables follows the Gamma distribution as the lemma states. □

APPENDIX VII

Proof of Theorem 1. Recall that the MSE, when using $\mathbf{G}^\star(t)$, is given by (30). Due to Lemma 3, we know the distribution of $||\mathbf{v}(t)||^2$, and we can therefore write:

$$\text{MSE} = \int_{x=0}^{\infty} g(x)f_{||\mathbf{v}(t)||^2}(x)dx = \int_{x=0}^{\infty}\frac{f}{f+x}\,\frac{\lambda^{N_r}x^{N_r-1}e^{-\lambda x}}{(N_r-1)!}dx = f\lambda e^{f\lambda}E_{in}(N_r,f\lambda), \quad (73)$$

where $E_{in}(n,z)\triangleq\int_{t=1}^{\infty}e^{-zt}/t^n dt$ is the standard exponential integral function. □

APPENDIX VIII

Proof of Proposition 3. We start with rewriting the MSE expression in (39) by introducing $\mu\triangleq f\lambda$ and making use of the following recursive relation from [51] (also available at [52, 8.19.12]):

$$\mu E_{\text{in}}(N_r,\mu) + N_rE_{\text{in}}(N_r+1,\mu) = e^{-\mu}. \quad (74)$$

We would like to take the first derivative of the MSE as a function of P_p. To this end, we use (39), and take the derivative of the MSE with respect to μ

$$\text{MSE}'(\mu) = -\mu e^{\mu}E_{\text{in}}(N_r-1,\mu) + + e^{\mu}E_{\text{in}}(N_r,\mu_\kappa) + \mu_\kappa e^{\mu}E_{\text{in}}(N_r). \quad (75)$$

After some algebraic manipulation based on (74), we obtain:

$$\text{MSE}'(\mu) = e^{\mu}\big(N_r+\mu\big)E_{\text{in}}(N_r,\mu) - 1. \quad (76)$$

From [51] (also available at [52, 8.19.21]) we have

$$1 < (x+n)e^x E_{\text{in}}(n,x) < \frac{x+n}{x+n-1}. \tag{77}$$

Substituting $x = \mu$ and $n = N_r$ shows that $\text{MSE}'(\mu) \neq 0$ if $0 < \mu$.

Next, we consider the first derivative of μ as defined in (40) with respect to P_p:

$$\begin{aligned}\mu'(P_p) &= \\ &= \frac{c_0 + c_1 P_p + c_2 P_p^2 + c_3 P_p^3 + c_4 P_p^4}{d_0 + d_1 P_p + d_2 P_p^2 + d_3 P_p^3 + d_4 P_p^4 + d_5 P_p^5 + d_6 P_p^6},\end{aligned} \tag{78}$$

where the coefficients of the numerator are defined in (44), and the coefficients of the denominator are given as follows:

$$\begin{aligned}
d_0 &= 0 \\
d_1 &= 0 \\
d_2 &= c^2\alpha^4\sigma^4\tau_p P_{\text{tot}}^2(a^2+1)^2 \\
d_3 &= 2c^2\alpha^4\sigma^2\tau_p^2 P_{\text{tot}}(a^2+1)\left(c\alpha^2 P_{\text{tot}}(1-a^2) - (a^2+1)\sigma^2\right) \\
d_4 &= c^2\alpha^4\tau_p^3\left(c^2\alpha^4 P_{\text{tot}}(a^2-1)^2\right. \\
&\quad \left. +4c\alpha^2\sigma^2 P_{\text{tot}}(a^4-1) - \sigma^2(a^2+1)\right) \\
d_5 &= 2c^3\alpha^6\tau_p^4(a^2-1)\left(c\alpha^2 P_{\text{tot}}(1-a^2) - (a^2+1)\sigma^2\right) \\
d_6 &= c^4\alpha^8\tau_p^5(a^2-1)^2.
\end{aligned} \tag{79}$$

Finally, the first derivative of the MSE with respect to P_p is:

$$\text{MSE}'(P_p) = \underbrace{\text{MSE}'(\mu)}_{\neq 0} \cdot \mu'(P_p). \tag{80}$$

Recall that $\text{MSE}'(\mu) \neq 0$. Consequently, the roots of $\text{MSE}'(P_p)$ are identical with the roots of the numerator of (78). □

REFERENCES

[1] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. WILEY, 2nd edition, 2011, ISBN-10: 0470660252.

[2] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

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

[4] T. Kim and J. G. Andrews, "Optimal pilot-to-data power ratio for MIMO-OFDM," in *IEEE Globecom*, St. Louis, MO, USA, Dec. 2005, pp. 1481–1485.

[5] ——, "Balancing pilot and data power for adaptive MIMO-OFDM systems," in *IEEE Globecom*, San Francisco, CA, USA, Dec 2006.

[6] T. Marzetta, "How much training is needed for multiuser MIMO ?" *IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)* , pp. 359–363, June 2006.

[7] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhear in multipath fading channels," *IEEE Trans. on Communications*, vol. 58, no. 10, pp. 2939–2948, October 2010.

[8] N. Sun and J. Wu, "Maximizing spectral efficiency for high mobility systems with imperfect channel state information," *IEEE Trans. Wireless Communication*, vol. 13, no. 3, pp. 1462–1470, March 2014.

[9] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 605–608, Dec. 2014.

[10] K. Guo, Y. Guo, G. Fodor, and G. Ascheid, "Uplink power control with mmse receiver in multi-cell MU-Massive-MIMO systems," in *Proc. of IEEE International Conference on Communications (ICC)*, Jun. 2014, pp. 5184–5190.

[11] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2363–2378, May 2017.

[12] T. C. Mai, H. Q. Ngo, M. Egan, and T. Q. Duong, "Pilot power control for cell-free massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 11 264–11 268, 2018.

[13] X. Yuan, C. Fan, and Y. J. Zhang, "Fundamental Limits of Training-Based Uplink Multiuser MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7544–7558, 2018.

[14] I. Atzeni, B. Gouda, and A. Tölli, "Distributed Precoding Design via Over-the-Air Signaling for Cell-Free Massive MIMO," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2020, early Access.

[15] G. Fodor and M. Telek, "On the Pilot-Data Power Trade Off in Single Input Multiple Output Systems," *European Wireless '14*, vol. Barcelona, Spain, May 2014.

[16] P. Liu, S. Jin, T. Jiang, Q. Zhang, and M. Matthaiou, "Pilot Power Allocation Through User Grouping in Multi-Cell Massive MIMO Systems," *IEEE Transactions on Communications*, vol. 65, no. 4, pp. 1561–1574, 2017.

[17] 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.

[18] X. Li, E. Björnson, E. G. Larsson, S. Zhou, and J. Wang, "A multi-cell MMSE detector for massive MIMO systems and new large system analysis," in *IEEE Global Communications Conference (Globecom)*. San Diego, CA, USA: IEEE, Dec. 2015.

[19] G. Fodor, P. D. Marco, and M. Telek, "On minimizing the MSE in the presence of channel information errors," *IEEE Communications Letters*, vol. 19, no. 9, pp. 1604 – 1607, September 2015.

[20] 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.

[21] X. Li, E. Björnson, E. G. Larsson, S. Zhou, and J. Wang, "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression," *EURASIP Journal on Wireless Communications and Networking*, vol. 117, May 2017, DoI: 10.1186/s13638-017-0879-2.

[22] J. Ma and L. Ping, "Data-Aided Channel Estimation in Large Antenna Systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3111–3124, 2014.

[23] J. Ma, C. Liang, C. Xu, and L. Ping, "On Orthogonal and Superimposed Pilot Schemes in Massive MIMO NOMA Systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2696–2707, 2017.

[24] X. Yuan, L. Ping, C. Xu, and A. Kavcic, "Achievable Rates of MIMO Systems With Linear Precoding and Iterative LMMSE Detection," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7073–7089, 2014.

[25] L. Liu, Y. Chi, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-Achieving MIMO-NOMA: Iterative LMMSE Detection," *IEEE Transactions on Signal Processing*, vol. 67, no. 7, pp. 1758–1773, 2019.

[26] 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.

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

[28] S. Kashyap, C. Mollen, 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.

[29] 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.

[30] 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.

[31] A. Papazafeiropoulos and T. Ratnarajah, "Modeling and Performance of Uplink Cache-Enabled Massive MIMO Heterogeneous Networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8136–8149, 2018.

[32] 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.
[33] 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.
[34] 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.
[35] Y. Zhang, S. B. Gelfand, and M. P. Fitz, "Soft-output demodulation on frequency-selective rayleigh fading channels uing AR channel models," *IEEE Transactions on Communications*, vol. 55, no. 10, pp. 1929–1939, Oct. 2007.
[36] 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.
[37] F. Lehmann, "Blind estimation and detection of space-time trellis coded transmissions over the rayleigh fading MIMO channel," *IEEE Transactions on Communications*, vol. 56, no. 3, pp. 334–338, Mar. 2008.
[38] 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, Jan. 2015.
[39] G. Fodor, P. D. Marco, and M. Telek, "On the Impact of Antenna Correlation and CSI Errors on the Pilot-to-Data Power Ratio," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2622 – 2633, April 2016.
[40] A. Mahmoudi and M. Karimi, "Estimation of the parameters of multichannel autoregressive signals from noisy observations," *Signal Porcessing (Elsevier)*, vol. 2008, no. 88, pp. 2777–2783, 2008.
[41] 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.
[42] G. Fodor, P. D. Marco, and M. Telek, "Performance analysis of block and comb type channel estimation for massive MIMO systems," in *First International Conference on 5G*, Akaslompolo, Finland, 2014.
[43] 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.
[44] 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.
[45] Z. B. Krusevac and P. B. Rapajic, "Adaptive AR channel model identification of time-varying communication systems," in *IEEE 10th International Symposium on Spread Spectrum Techniques and Applications*, Bologna, Italy, Aug. 2008.
[46] S. Mekki, M. Amara, A. Feki, and S. Valentin, "Channel gain prediction for wireless links with Kalman filters and expectation-maximization," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Doha, Qatar, Apr. 2016.
[47] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*. Prentice Hall, 1993, no. ISBN: 0133457117.
[48] A. Barbieri, A. Piemontese, and G. Colavolpe, "On the ARMA approximation for fading channels described by the Clarke model with applications to Kalman-based receivers," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 535 – 540, Feb. 2009.
[49] E. L. Rees, "Graphical discussion of the roots of a quartic equation," *The American Mathematical Monthly*, vol. 29, no. 2, pp. 51–55, 1922.
[50] G. Brookfield, "Factoring quartic polynomials: A lost art," *Mathematics Magazine*, vol. 8, no. 1, pp. 67–70, Feb. 2007.
[51] F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, Eds., *NIST Handbook of Mathematical Functions*, 1st ed. New York, NY, USA: Cambridge University Press, 2010.
[52] "NIST handbook of mathematical functions," http://dlmf.nist.gov/8.19, 2010, [Online; accessed 8-Oct-2015].