

On the Impact of Antenna Correlation and CSI Errors on the Pilot-to-Data Power Ratio

Gábor Fodor^{*†}, Piergiuseppe Di Marco^{*}, Miklós Telek^{‡‡}

^{*}Ericsson Research, Stockholm, Sweden. E-mail: Gabor.Fodor|Piergiuseppe.Di.Marco@ericsson.com

[†]Royal Institute of Technology, Stockholm, Sweden. E-mail: gaborf@kth.se

[‡]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—In systems employing pilot-symbol aided channel estimation, the pilot-to-data power ratio is known to have a large impact on performance. Therefore, previous works proposed methods setting the pilot power such that either the weighted sum of the mean squared error (MSE) of the estimated data symbols is minimized or the overall spectral efficiency (SE) is maximized. However, previous works did not take into account the impact of correlated antennas and channel state information (CSI) errors on the optimal pilot power setting. In this paper we consider the uplink of a multi-user multiple-input multiple-output (MU MIMO) system employing a receiver that minimizes the MSE of the received data symbols in the presence of CSI errors and derive closed form expressions for the MSE and the achievable SE. These expressions take into account the impact of antenna correlation and CSI errors, and are a function of pilot power and the number of receive antennas. The analytical and numerical results can help set the pilot power, minimizing the MSE in multiple antenna systems.

Keywords: multi-antenna systems, channel state information, estimation techniques, receiver algorithms

I. INTRODUCTION

Communicating over an unknown wireless channel is subject to a penalty of channel uncertainty, sometimes in the form of training costs [1]¹. As it has been shown by [3] and [4], this penalty depends on the knowledge the receiver has of the channel and the rate of change of the channel, as well as on the number of transmit antennas. On the other hand, reducing this penalty by sending over only a fraction of the available degrees of freedom results in a loss of spectral efficiency. This fundamental insight has generated significant interest in designing channel state information (CSI) acquisition techniques and channel estimation algorithms since the late 1990s.

For example, the results of [5] and [6] established a lower bound for multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing systems with minimum mean squared error (MMSE) channel estimation. It was also shown that the optimal pilot-to-data power ratio (PDPR) setting that maximizes this lower bound or minimizes the average symbol

error rate can increase capacity by 10-20%, as compared with a system using a suboptimal pilot power setting.

Subsequently, the work reported in [7] provided a unified treatment of the optimum pilot overhead in multipath fading channels and gave closed form expressions for the fraction of the power budget that must be devoted to pilots, explicitly considering the dependence of the pilot overhead on the Doppler frequency and other factors. The impact of transmitter and receiver in-phase and quadrature imbalances and residual capacity offset on the pilot-to-data power allocation was analyzed, and a capacity bound maximizing power allocation was found in [8]. More recently, the pilot power ratio that maximizes the uplink sum-rate in zero-forcing based multi-user MIMO (MU MIMO) systems with a large number of antennas was studied in [9].

Along a related line of research, the results of [10],[11] and [12] indicate that the performance of MMSE receivers is sensitive to channel estimation errors. In particular, the often-used classical or *naïve* MMSE receiver does not, in fact, minimize the MSE of the estimated data symbols in the presence of CSI errors [12], [13]. It turns out that the difference between the naïve receiver and regularized (true) MMSE receiver, in terms of the achieved MSE, is significant in the large antenna regime [12].

These two lines of works suggest that for the purpose of determining the optimal pilot power setting it is important to take into account the operation of practical channel estimation and receiver algorithms. To the best of our knowledge, exact expressions for the achieved MSE and SE when using practical channel estimation (such as least squares, LS) and receiver algorithms (such as MMSE), and accounting for the pilot-to-data power ratio and antenna correlation, are not available. In this paper, we address this problem and derive closed form expressions for the uplink of a MU MIMO system, in which the base station (BS) uses LS or MMSE channel estimation and MMSE receiver. Throughout, we assume that the output of the MMSE detector, the residual signal plus interference from other spatial streams as well as the estimation error of the received data symbols can be approximated as Gaussian [10]. Because in practice the CSI estimation error is likely to be bounded, our design can be regarded as a worst-case design approach. Thereby, our contributions (detailed in Sections IV-VII and the Appendices) to the lines of works above can be

The work of G. Fodor has been supported by the Swedish Foundation for Strategic Research Strategic Mobility SM13-0008 Matthew Project.

The work of M. Telek was partially performed when he was visiting the Royal Institute of Technology, Stockholm. M. Telek was partially supported by the OTKA K 119750 grant.

¹Some parts of the model used in this paper have appeared in our conference paper [2]. However, the material presented in Sections IV-VIII and the related Appendices are novel contributions.

summarized as follows:

- We derive closed form exact expressions for both the MSE and the SE taking into account the CSI errors that are specific to the employed channel estimation technique;
- We explicitly take into account the impact of antenna correlation on these performance measures.

These formulas are then used to compare the performance of MU MIMO systems employing the naïve and MMSE receivers. An interesting insight is that when the system uses the MMSE receiver, the PDPR minimizing the MSE does not depend on the number of receive antennas at the BS but rather is dependent on the large-scale fading. This is in contrast to a system that employs the naïve receiver, for which the pilot power minimizing the MSE depends on the number of receive antennas. We believe that this insight can help set the pilot power almost optimally in practical systems in which the number of BS antennas can depend on the actual deployment scenario [14], [15]. In particular, our results show that when the optimal pilot power setting is employed at the terminal side, and the true MMSE receiver is used at the base station side, the system's performance is close to that of a hypothetical system that would have access to the perfect CSI.

The paper is structured as follows. The next section discusses related works. Section III describes the system model and summarizes preliminaries needed for development of the contributions of this paper. Sections IV and V analyze the MSE in the case of uncorrelated and correlated antennas at the receiver, respectively. Section VI derives closed form expressions for the MSE and SE when the receiver uses the MMSE receiver. Section VII presents numerical results on the MSE and SE, and Section VIII concludes the paper.

II. RELATED WORKS

In this section we review some of the relevant literature in the areas of information theoretical aspects of wideband communications, MIMO transceiver design in the presence of CSI errors and training based channel estimation techniques. We also point out our contributions to this line of research.

A. Information Theoretical Aspects of Wideband Communications and Capacity Analysis

An important insight from the works reported in [16] and [17] is that there is a continuum between the extremes of communicating in non-coherent (without CSI availability) and coherent (perfect CSI) fashions over wireless channels in terms of the achieved spectral efficiency. Specifically, communicating over a completely unknown channel is subject to a penalty of the channel uncertainty, sometimes in the form of training costs. (This penalty depends on the knowledge the receiver has of the channel and on the channel's rate of change.) On the other hand, reducing this penalty by sending over only a fraction of the available degrees of freedom results in a loss of spectral efficiency.

In practice, the channel coherence time might be long enough to both estimate the fading coefficients and use such

estimates to communicate coherently after the estimation period, as well as to achieve performance close to the fully coherent case (as emphasized in [1]).

Reference [17] studies the connection between the channel uncertainty penalty and the coherence length of the channel in MIMO systems. A key observation is that in the low signal-to-noise ratio (SNR) regime, estimating the channel at the receiver may not be possible and hence communication may be desirable without training. More exactly, if the channel coherence length is above a certain antenna- and SNR-dependent threshold, the noncoherent and coherent capacities become the same in the low-SNR regime.

The above results suggest that, depending on the SNR and the number of antennas, there may be a large gap between the coherent and noncoherent extremes in terms of achievable spectral efficiency, and channel learning is key in bridging this gap. Therefore, it is interesting to consider the ultrawide-band (UWB) regime and focus on the case when training signals are used for channel estimation at the receiver. The capacity of this scheme is studied in [18] to investigate the impact of multipath sparsity on achieving coherent capacity. The key results of this paper are a lower bound on the capacity of the training-based communication scheme and the coherence level that can be achieved, and the insights into the impact of channel sparsity on the achievable capacity in the UWB regime.

The work in [19] studies the impact of channel state feedback on the achievable rates in sparse wideband channels. A key insight is that a partial and/or limited feedback scheme, where only one bit per independent DoF is available at the transmitter, can nearly achieve the performance of a system in which perfect CSI is available at the transmitter. References [20] and [21] focus on acquiring channel state information at the transmitter in multi-user systems where the feedback from each user terminal must be limited. It is shown that the combination of long term channel statistics and instantaneous norm feedback provides sufficient information at the transmitter for efficient scheduling, beamforming and link adaptation in wide-area scenarios. More recently, the work in [22] considers a case in which a transmitter with two antennas broadcasts to two single-antenna users. It is assumed that the two receiving users have perfect channel information, whereas the transmitter has only statistical information of each user's link (covariance matrix of the vector channel). The paper focuses on the design of beamforming vectors that depend on such statistical information and maximize the ergodic sum-rate delivered to the two users.

B. MIMO Transceiver Design

Reference [23] deals with robust MIMO precoding design with deterministic imperfect channel state information at the transmitter (CSIT) such that the worst-case received SNR is maximized, or the worst-case error probability is minimized. Reference [4] is concerned with the design of linear MIMO transceivers that are robust to CSI perturbations at both sides of the link that is to errors in CSIT and channel state information at the receiver (CSIR). In this work, the design

of the transceiver is based on minimizing the average sum MSE of all data streams and users. This paper assumes a perturbation error (modelled as a Gaussian additive term), but this CSI error is not controlled by pilot power or the training scheme. Therefore, the pilot-data trade-off is not considered in this paper. The model used in [24] builds on the uplink-downlink duality in sum MSE under imperfect CSI. In this work, the imperfectness of the channel knowledge is taken into account in the joint MMSE design. The sum MSE minimization problem for the UL and DL is subject to power constraints. However, the aspect of pilot power is not considered and the MSE is not derived as a function of the pilot power under a constrained pilot-data budget.

C. Channel Estimation and the Pilot-Data Power Ratio

The seminal work reported in [25] evaluates the difference between the mutual information when the receiver has only an estimate of the channel and when it has perfect knowledge of the channel. Upper and lower bounds are established on this difference and are related to the variance of the channel measurement error. In [3] it is shown how training based channel estimation affects the capacity of the fading channel, recognizing that training imposes a substantial information-theoretic penalty, especially when the coherence interval T is only slightly larger than the number of transmit antennas or when the SNR is low. In these regimes, learning the entire channel is highly suboptimal. Conversely, if the SNR is high, and T is much larger than M , training-based schemes can come very close to achieving capacity. Therefore, the power that should be spent on training and data transmission depends on the relation between T and M . The work in [26] can be seen as a sequel of [3], taking into account intersymbol interference and the receiver technique (equalizer) used on the receiver side. However, none of these works consider the regularized MMSE receiver, and therefore the pilot power setting that minimizes the MSE of a regularized MMSE receiver is not discussed in these papers.

The MU MIMO setting is the focus of [27], in which the coherence interval of T symbols is expended for channel training, channel estimation, and precoder computation for DL transmission. Specifically, the optimum number of terminals in terms of the DL spectral efficiency is determined for a given coherence interval, number of base station antennas, and DL/UL signal-to-interference-plus-noise ratio. There is no receiver design involved and the pilot-to-data power trade-off is out of the scope of the optimization process.

Reference [7] investigates the optimization of the pilot overhead for single-user wireless fading channels, and the dependencies of this pilot overhead on various system parameters of interest (e.g. fading rate, SNR) are quantified. By finding an expansion of the spectral efficiency for the overhead optimization in terms of the fading rate around the perfect-CSI point, the square root dependence of both the overhead and the spectral efficiency penalty is cleanly identified.

D. Contributions

Our contributions to the above referenced works are establishing the pilot power when the receiver employs the regularized MMSE receiver for both uncorrelated and correlated receive antenna cases. Specifically:

- For the case of uncorrelated receive antennas at the BS, we give a closed form expression for the MSE of the estimated data symbols and for the pilot power that minimizes this MSE. (Lemma 2 and Proposition 3, respectively.)
- For the case of correlated receive antennas at the BS, we first identify the regularized MMSE receiver structure (Lemmas 4 and 5) and then give closed form expressions for the achieved MSE (Section VI).

These results allow us to study numerically the gains of using the regularized MMSE receiver and optimal pilot power levels over schemes that use the naïve receivers and/or suboptimal pilot power levels. A key insight is that the pilot power that minimizes the MSE does not depend on the number of antennas, but heavily depends on the path loss between the BS and the mobile terminal.

III. CHANNEL ESTIMATION AND RECEIVER MODEL

A. Channel Estimation Model

We consider the uplink of a MU MIMO system, in which the mobile stations (MS) transmit orthogonal pilot sequences of length τ_p : $\mathbf{s} = [s_1, \dots, s_{\tau_p}]^T \in \mathbb{C}^{\tau_p \times 1}$, in which each pilot symbol is scaled as $|s_i|^2 = 1$, for $i = 1, \dots, \tau_p$. The pilot sequences are constructed such that they remain orthogonal as long as the number of spatially multiplexed users is maximum τ_p [28]. In practice, such pilot sequences can be defined using the popular Zadoff-Chu sequences [29],[30]. Specifically, without loss of generality, we assume that the number MU MIMO users is $K \leq \tau_p$. In practice, $K \ll N_r$, where N_r is the number of antennas at the BS [31].

As emphasized in [31], MU MIMO differs from point-to-point MIMO in two respects: first, the terminals are typically separated by many wavelengths, and second, the terminals cannot collaborate among themselves, either to transmit or to receive data. That is, in MU MIMO systems, the terminals are autonomous so that we can assume that the transmit array is uncorrelated. However, it is important to capture the correlation structure at the receiver side so that we can evaluate the impact of CSIR errors on the optimal pilot power and the achieved MSE. In this paper we assume a comb type arrangement of the pilot symbols. Given F subcarriers in the coherence bandwidth, a fraction of τ_p subcarriers are allocated to the pilot and $\tau_d = F - \tau_p$ subcarriers are allocated to the data symbols. Each MS transmits at a constant power P_{tot} , however, this transmission power can be distributed unequally among the subcarriers. In particular, considering User- ℓ with a transmitted power $P_{p,\ell}$ for each pilot symbol and P_ℓ for each data symbol transmission, the sum constraint of:

$$\tau_p P_{p,\ell} + \tau_d P_\ell = P_{tot} \quad (1)$$

is enforced. In practice, this type of arrangement is suitable for time varying channels, so that channel estimation is facilitated at the same time instant that is used for data transmission. Thus, the $N_r \times \tau_p$ matrix of the received pilot signal from User- ℓ at the BS can be conveniently written as:

$$\mathbf{Y}_\ell^p = \alpha_\ell \sqrt{P_{p,\ell}} \mathbf{h}_\ell \mathbf{s}^T + \mathbf{N}, \quad (2)$$

where we assume that $\mathbf{h}_\ell \in \mathbb{C}^{N_r \times 1}$ is a circular symmetric complex normal distributed column vector with mean vector $\mathbf{0}$ and covariance matrix \mathbf{C}_ℓ (of size N_r), denoted as $\mathbf{h}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\ell)$, α_ℓ 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 noise power on the received pilot signal.

In this paper we assume that the BS uses the popular LS estimator that 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 (2) assuming:

$$\begin{aligned} \hat{\mathbf{h}}_\ell &= \mathbf{h}_\ell + \mathbf{w}_\ell = \frac{1}{\alpha_\ell \sqrt{P_{p,\ell}}} \mathbf{Y}_\ell^p \mathbf{s}^* (\mathbf{s}^T \mathbf{s}^*)^{-1} \\ &= \mathbf{h}_\ell + \frac{1}{\alpha_\ell \sqrt{P_{p,\ell} \tau_p}} \mathbf{N} \mathbf{s}^*, \end{aligned} \quad (3)$$

where $\mathbf{s}^* = [s_1^*, \dots, s_{\tau_p}^*]^T \in \mathbb{C}^{\tau_p \times 1}$ denotes the vector of pilot symbols and $(\mathbf{s}^T \mathbf{s}^*) = \tau_p$. By considering $\mathbf{h}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\ell)$, it follows that the estimated channel $\hat{\mathbf{h}}_\ell$ is a circular symmetric complex normal distributed vector $\hat{\mathbf{h}}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\ell)$, with

$$\mathbf{R}_\ell \triangleq \mathbb{E}\{\hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H\} = \mathbf{C}_\ell + \frac{\sigma_p^2}{\alpha_\ell^2 P_{p,\ell} \tau_p} \mathbf{I}_{N_r}. \quad (4)$$

By recognizing that \mathbf{h} and $\hat{\mathbf{h}}$ are jointly circular symmetric complex Gaussian (multivariate normal) distributed random variables, the distribution of the channel realization \mathbf{h}_ℓ conditioned on the estimate $\hat{\mathbf{h}}_\ell$ is normally distributed as follows [32], [33]:

$$(\mathbf{h}_\ell \mid \hat{\mathbf{h}}_\ell) \sim \mathcal{CN}(\mathbf{D}_\ell \hat{\mathbf{h}}_\ell, \mathbf{Q}_\ell), \quad (5)$$

where $\mathbf{D}_\ell \triangleq \mathbf{C}_\ell \mathbf{R}_\ell^{-1}$ and $\mathbf{Q}_\ell \triangleq \mathbf{C}_\ell - \mathbf{C}_\ell \mathbf{R}_\ell^{-1} \mathbf{C}_\ell$.

B. Received Data Signal Model

The MU MIMO received data signal at the BS can be written as:

$$\mathbf{y} = \underbrace{\alpha_\ell \mathbf{h}_\ell \sqrt{P_\ell} x_\ell}_{\text{User-}\ell} + \underbrace{\sum_{k \neq \ell}^K \alpha_k \mathbf{h}_k \sqrt{P_k} x_k}_{\text{Other users}} + \mathbf{n}_d, \quad (6)$$

where $\alpha_k \mathbf{h}_k$ is the $M \times 1$ 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 is the transmitted data symbol by User- k and \mathbf{n}_d denotes the Gaussian noise on the received data signal.

C. Employing an MMSE Receiver at the BS

In this paper the BS employs an MMSE receiver $\mathbf{G}_\ell \in \mathbb{C}^{1 \times N_r}$ to estimate the data symbol transmitted by User- ℓ . As it was shown in [12], in the case of a linear receiver \mathbf{G}_ℓ that requires the estimated channel of only User- ℓ as its input, the MSE of the estimated data symbols of User- ℓ can be conveniently expressed in the following quadratic form:

$$\begin{aligned} \text{MSE}(\mathbf{G}_\ell, \hat{\mathbf{h}}_\ell) &= \\ &\mathbf{G}_\ell \left(\alpha_\ell^2 P_\ell (\mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H + \mathbf{Q}_\ell) + \sum_{k \neq \ell}^K \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I} \right) \mathbf{G}_\ell^H \\ &- \alpha_\ell \sqrt{P_\ell} (\mathbf{G}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell + \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{G}_\ell^H) + 1. \end{aligned} \quad (7)$$

As we shall see later, our analysis allows for an arbitrary channel covariance matrix at the receiver side (\mathbf{C}_ℓ) in (7) that allows us to analyze the impact of CSI errors on the MSE performance with arbitrary correlation structure of the base station antennas. We recall that the MMSE receiver aims at minimizing the MSE between the estimate $\mathbf{G}_\ell \mathbf{y}$ and the transmitted symbol x_ℓ :

$$\mathbf{G}_\ell^* \triangleq \arg \min_{\mathbf{G}} \mathbb{E}\{\text{MSE}\} = \arg \min_{\mathbf{G}} \mathbb{E}\{|\mathbf{G}_\ell \mathbf{y} - x_\ell|^2\}. \quad (8)$$

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

$$\mathbf{G}_\ell^{\text{naïve}} = \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H (\alpha_\ell^2 P_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H + \sigma_d^2 \mathbf{I})^{-1}. \quad (9)$$

As it was shown in [12], this receiver does not minimize the MSE. Using the quadratic form in (7), it can be shown that the receiver that minimizes the MSE of the received data symbols, is constructed as:

$$\begin{aligned} \mathbf{G}_\ell^* &= \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \cdot \\ &\cdot \left(\alpha_\ell^2 P_\ell (\mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H + \mathbf{Q}_\ell) + \sum_{k \neq \ell}^K \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I} \right)^{-1}. \end{aligned} \quad (10)$$

D. Calculating the MSE When Employing the MMSE Receiver

In [12] it was shown that for the special case when the channel covariance matrices \mathbf{C}_ℓ and consequently the matrices \mathbf{R}_ℓ , \mathbf{D}_ℓ and \mathbf{Q}_ℓ are proportional to the identity matrix \mathbf{I}_{N_r} , with diagonal elements c_ℓ , r_ℓ , d_ℓ and q_ℓ respectively, the unconditional MSE $_\ell$ of the uplink estimated data symbols of User- ℓ when employing the \mathbf{G}_ℓ^* receiver can be calculated as follows.

Proposition 1. *The unconditional MSE_ℓ of the received data symbols of User-ℓ when the BS uses the optimal \mathbf{G}_ℓ^* receiver is as follows:*

$$\begin{aligned} \text{MSE}_\ell = & \frac{b_\ell \left(e^{\frac{b_\ell}{s_\ell r_\ell}} (b_\ell + N_r s_\ell r_\ell) E_{in} \left(N_r, \frac{b_\ell}{s_\ell r_\ell} \right) - s_\ell r_\ell \right)}{s_\ell^2 r_\ell^2} \\ & + \frac{N_r \left(e^{\frac{b_\ell}{s_\ell r_\ell}} (b_\ell + (1 + N_r) s_\ell r_\ell) E_{in} \left(1 + N_r, \frac{b_\ell}{s_\ell r_\ell} \right) - s_\ell r_\ell \right)}{s_\ell r_\ell} \\ & - 2 \cdot e^{\frac{b_\ell}{s_\ell r_\ell}} N_r E_{in} \left(1 + N_r, \frac{b_\ell}{s_\ell r_\ell} \right) + 1, \end{aligned} \quad (11)$$

where $E_{in}(n, z) \triangleq \int_1^\infty e^{-zt}/t^n dt$ is a standard exponential integral function, $s_\ell \triangleq d_\ell^2 p_\ell$, $b_\ell \triangleq q_\ell p_\ell + \sigma_d^2$ with $p_\ell \triangleq \alpha_\ell^2 P_\ell$.

The proof is in Appendix I.

Notice that specifically in the case of LS channel estimation and when \mathbf{C}_ℓ is of the form of $c_\ell \mathbf{I}_{N_r}$, from (4)-(5) we have:

$$r_\ell = c_\ell + \frac{\sigma_p^2}{\alpha_\ell^2 P_{p,\ell} \tau_p}; \quad d_\ell = \frac{c_\ell}{r_\ell}; \quad q_\ell = c_\ell - c_\ell d_\ell. \quad (12)$$

IV. ANALYSIS OF THE MSE IN THE CASE OF UNCORRELATED ANTENNAS

This section presents the optimal pilot power setting for the case when \mathbf{C}_ℓ is proportional to the identity matrix that is $\mathbf{C}_\ell = c_\ell \mathbf{I}_{N_r}$. We start with a further simplified version of Proposition 1.

Lemma 2. *When the BS uses the optimal \mathbf{G}_ℓ^* receiver and the channel can be assumed $\mathbf{C}_\ell = c_\ell \mathbf{I}_{N_r}$, the MSE of the estimated data symbols of each user can be calculated as follows:*

$$\text{MSE}(\mu_\ell) = \mu_\ell e^{\mu_\ell} E_{in}(N_r, \mu_\ell), \quad (13)$$

where $\mu_\ell = \mu(P_{p,\ell})$ is defined by

$$\mu_\ell \triangleq \frac{\sigma_d^2 \sigma_p^2 \tau_d + c_\ell \alpha_\ell^2 \left(\sigma_p^2 P_{tot} + \tau_p P_{p,\ell} (\sigma_d^2 \tau_d - \sigma_p^2) \right)}{c_\ell^2 \alpha_\ell^4 P_{p,\ell} \tau_p (P_{tot} - \tau_p P_{p,\ell})}. \quad (14)$$

The proof is in the Appendix II.

As it was underscored by [16] and [18], there is a gap in spectral efficiency between coherent and noncoherent communications and channel learning plays an important role in bridging this gap. Lemma 2 captures the training cost ($\tau_p P_{p,\ell}$) of communicating over an unknown channel specifically in the case of an uplink of MU MIMO system employing an MMSE receiver.

For the naïve ($\mathbf{G}_\ell^{\text{naive}}$) as well as for the optimal MMSE receiver (\mathbf{G}_ℓ^*), it is important to find the pilot power that minimizes the MSE. For the naïve receiver we obtain the optimal pilot power by numerical optimization, whereas for the optimal MMSE receiver the pilot power that minimizes the MSE has a closed form expression. The following proposition presents the optimal PDPR as a function of the total power and coherence budget and the large scale fading between the MS and the BS.

Proposition 3. *When employing the MMSE receiver \mathbf{G}_ℓ^* , in the case of $\mathbf{C}_\ell = c_\ell \mathbf{I}_{N_r}$, the pilot power that minimizes the MSE is independent of the number of receive antennas N_r and is given by:*

$$\begin{aligned} P_{p,\ell}^* = & \frac{\sigma_d \sigma_p \sqrt{(c_\ell P_{tot} \alpha_\ell^2 + \sigma_p^2)(c_\ell P_{tot} \alpha_\ell^2 + \sigma_d^2 \tau_d) \tau_d}}{c_\ell \alpha_\ell^2 \tau_p (\sigma_d^2 \tau_d - \sigma_p^2)} \\ & - \frac{\sigma_p^2 (c_\ell P_{tot} \alpha_\ell^2 + \sigma_d^2 \tau_d)}{c_\ell \alpha_\ell^2 \tau_p (\sigma_d^2 \tau_d - \sigma_p^2)}. \end{aligned} \quad (15)$$

The proof is in Appendix III.

Remark 1. *In the case of $\sigma_d = \sigma_p = \sigma$, expression (15) can be further simplified as*

$$\begin{aligned} P_{p,\ell}^* = & P_{tot} \left(\frac{\sqrt{\left(1 + \frac{\sigma^2}{c_\ell P_{tot} \alpha_\ell^2} \tau_d\right) \left(1 + \frac{\sigma^2}{c_\ell P_{tot} \alpha_\ell^2}\right) \tau_d}}{\tau_p (\tau_d - 1)} \right. \\ & \left. - \frac{\left(1 + \frac{\sigma^2}{c_\ell P_{tot} \alpha_\ell^2} \tau_d\right)}{\tau_p (\tau_d - 1)} \right). \end{aligned}$$

The optimal pilot power is a fraction of the power budget P_{tot} that depends on the number of pilot symbols τ_p and data symbols τ_d . It is also easy to verify that in the case of perfect channel knowledge (i.e., assuming $\sigma_p = 0$), expression (15) returns $P_{p,\ell}^ = 0$.*

V. ANALYSIS OF THE MSE IN THE CASE OF CORRELATED ANTENNAS

A. Determining \mathbf{G}^*

We now consider the general case when the channel covariance matrices (\mathbf{C}_k) are not diagonal, that is when we allow for an arbitrary correlation structure between the BS antennas. We assume that the BS employs the optimal MMSE equalizer according to (10) and write

$$\mathbf{G}_\ell^* = \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \left(\boldsymbol{\Psi}_\ell + \alpha_\ell^2 P_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \right)^{-1}, \quad (16)$$

where

$$\boldsymbol{\Psi}_\ell \triangleq \alpha_\ell^2 P_\ell \mathbf{Q}_\ell + \sum_{k \neq \ell}^K \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I}_{N_r}, \quad (17)$$

is a positive definite matrix which contains the covariance from all intra- and intercell interference sources that cause interference to the signal of User-ℓ and the self covariance term related with \mathbf{Q}_ℓ .

For an explicit inversion in (16) we introduce the SVD of $\boldsymbol{\Psi}_\ell$, that is $\boldsymbol{\Psi}_\ell = \boldsymbol{\Theta}_\ell^H \mathbf{S}_\ell \boldsymbol{\Theta}_\ell$. Since $\boldsymbol{\Psi}_\ell$ is positive definite, it is non singular and we can therefore define:

$$\boldsymbol{\nu}_\ell \triangleq \mathbf{S}_\ell^{-1/2} \boldsymbol{\Theta}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell, \quad (18)$$

which is a linear transformed version of $\hat{\mathbf{h}}_\ell$. It will be useful to notice that:

$$\hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \boldsymbol{\Psi}_\ell^{-1} = \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \boldsymbol{\Theta}_\ell^H \mathbf{S}_\ell^{-1} \boldsymbol{\Theta}_\ell = \boldsymbol{\nu}_\ell^H \mathbf{S}_\ell^{-1/2} \boldsymbol{\Theta}_\ell, \quad (19)$$

and

$$\hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \boldsymbol{\Psi}_\ell^{-1} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell = \|\boldsymbol{\nu}_\ell\|^2, \quad (20)$$

and note that from (18) we have

$$\mathbf{D}_\ell \hat{\mathbf{h}}_\ell = \boldsymbol{\Theta}_\ell^H \mathbf{S}_\ell^{1/2} \boldsymbol{\nu}_\ell. \quad (21)$$

With these notations, it is straightforward to prove the following useful lemma.

Lemma 4. *Given a channel estimate instance $\hat{\mathbf{h}}_\ell$, the MMSE weight matrix \mathbf{G}_ℓ , as a function of the number of receive antennas at the BS (N_r) can be expressed as follows:*

$$\mathbf{G}_\ell^* = \frac{\alpha_\ell \sqrt{P_\ell}}{\alpha_\ell^2 P_\ell \|\boldsymbol{\nu}_\ell\|^2 + 1} \boldsymbol{\nu}_\ell^H \mathbf{S}_\ell^{-1/2} \boldsymbol{\Theta}_\ell, \quad (22)$$

where $\|\boldsymbol{\nu}_\ell\|^2 = \boldsymbol{\nu}_\ell^H \boldsymbol{\nu}_\ell = \sum_{i=1}^{N_r} |\nu_{\ell_i}|^2$.

The proof is in Appendix IV.

To simplify the discussion we introduce

$$g_\ell \triangleq \frac{\alpha_\ell \sqrt{P_\ell}}{\alpha_\ell^2 P_\ell \|\boldsymbol{\nu}_\ell\|^2 + 1}. \quad (23)$$

B. Determining the MSE When Using \mathbf{G}^*

To determine the MSE, we first need to find the distribution of $\boldsymbol{\nu}_\ell$. The distribution of $\boldsymbol{\nu}_\ell$ is readable from (18) (notice that $\boldsymbol{\Psi}_\ell$ and thereby \mathbf{S}_ℓ are not random variables), and recall that $\hat{\mathbf{h}}_\ell$ is complex normal distributed with, $\hat{\mathbf{h}}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\ell)$. Therefore, for $\boldsymbol{\nu}_\ell$ we have

$$\boldsymbol{\nu}_\ell \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Omega}_\ell), \quad (24)$$

where

$$\begin{aligned} \boldsymbol{\Omega}_\ell &\triangleq \mathbb{E}(\boldsymbol{\nu}_\ell \boldsymbol{\nu}_\ell^H) = \mathbb{E}\left(\left(\mathbf{S}_\ell^{-1/2} \boldsymbol{\Theta}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell\right)\left(\mathbf{S}_\ell^{-1/2} \boldsymbol{\Theta}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell\right)^H\right) \\ &= \mathbf{S}_\ell^{-1/2} \boldsymbol{\Theta}_\ell \mathbf{D}_\ell \mathbb{E}\left(\hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H\right) \mathbf{D}_\ell^H \boldsymbol{\Theta}_\ell^H \mathbf{S}_\ell^{-1/2} \\ &= \mathbf{S}_\ell^{-1/2} \boldsymbol{\Theta}_\ell \mathbf{D}_\ell \mathbf{R}_\ell \mathbf{D}_\ell^H \boldsymbol{\Theta}_\ell^H \mathbf{S}_\ell^{-1/2}. \end{aligned}$$

We will need the SVD of $\boldsymbol{\Omega}_\ell$:

$$\boldsymbol{\Omega}_\ell = \boldsymbol{\Theta}_{\Omega_\ell}^H \mathbf{S}_{\Omega_\ell} \boldsymbol{\Theta}_{\Omega_\ell}, \quad (25)$$

where $\boldsymbol{\Theta}_{\Omega_\ell}$ is an orthogonal matrix ($\boldsymbol{\Theta}_{\Omega_\ell}^H \boldsymbol{\Theta}_{\Omega_\ell} = \mathbf{I}_{N_r}$). Furthermore we will need the linear transform of $\boldsymbol{\nu}_\ell$, which we denote with $\boldsymbol{\omega}_\ell$ whose covariance matrix is diagonal:

$$\boldsymbol{\omega}_\ell \triangleq \boldsymbol{\Theta}_{\Omega_\ell} \boldsymbol{\nu}_\ell. \quad (26)$$

Notice that (for ease of notation dropping the index ℓ):

$$\|\boldsymbol{\omega}\|^2 = \boldsymbol{\omega}^H \boldsymbol{\omega} = \boldsymbol{\nu}^H \boldsymbol{\Theta}_{\Omega}^H \boldsymbol{\Theta}_{\Omega} \boldsymbol{\nu} = \boldsymbol{\nu}^H \boldsymbol{\nu} = \|\boldsymbol{\nu}\|^2 \quad (27)$$

and

$$\begin{aligned} \mathbb{E}_\omega(\boldsymbol{\omega} \boldsymbol{\omega}^H) &= \mathbb{E}_\nu(\boldsymbol{\Theta}_{\Omega} \boldsymbol{\nu} \boldsymbol{\nu}^H \boldsymbol{\Theta}_{\Omega}^H) = \boldsymbol{\Theta}_{\Omega} \mathbb{E}_\nu(\boldsymbol{\nu} \boldsymbol{\nu}^H) \boldsymbol{\Theta}_{\Omega}^H = \\ &= \boldsymbol{\Theta}_{\Omega} \boldsymbol{\Theta}_{\Omega}^H \mathbf{S}_{\Omega} \boldsymbol{\Theta}_{\Omega} \boldsymbol{\Theta}_{\Omega}^H = \mathbf{S}_{\Omega}. \end{aligned} \quad (28)$$

It is now straightforward to prove the following lemma.

Lemma 5. *The MSE of User- ℓ as a function of $\boldsymbol{\nu}_\ell$, as defined in (18), is as follows.*

$$\text{MSE}(\boldsymbol{\nu}_\ell) = \frac{1}{\alpha_\ell^2 P_\ell \|\boldsymbol{\nu}_\ell\|^2 + 1} \quad (29)$$

The proof is in Appendix V.

Remark 2. *It is insightful to compare (29) with the MSE of a system with uncorrelated antennas and perfect channel estimation, that is when $\mathbf{D} = \mathbf{I}$, $\mathbf{Q} = \mathbf{0}$ and $\hat{\mathbf{h}} = \mathbf{h}$. In this case, from (43) we get:*

$$\text{MSE}(\mathbf{h}) = \frac{1}{\alpha_\ell^2 P_\ell \frac{\|\mathbf{h}_\ell\|^2}{\sigma_a^2} + 1}, \quad (30)$$

which indicates that $\boldsymbol{\nu}_\ell$ can be seen as an "equivalent channel" in the system of correlated antennas and partial CSI information, that is $\boldsymbol{\nu}_\ell$ captures the impact of both antenna correlation and CSI estimation errors on the MSE.

VI. CALCULATING THE UNCONDITIONAL MSE AND SE

Recall from Lemma 4 that $\|\boldsymbol{\nu}_\ell\|^2 = \boldsymbol{\nu}_\ell^H \boldsymbol{\nu}_\ell = \sum_{i=1}^{N_r} |\nu_{\ell_i}|^2$, where the ν_{ℓ_i} -s ($i = 1, \dots, N_r$) are, in general, not independent random variables. However, according to (28), the covariance matrix of $\boldsymbol{\omega}_\ell$ – that is \mathbf{S}_{Ω} – is diagonal, with not necessarily equal diagonal elements. Therefore, each $|\omega_{\ell_i}|^2$ (denoted by $|\omega_i|^2$ in the sequel) is exponentially distributed.

Assume that the variance of ω_i is ξ_i^2 , and consequently $|\omega_i|^2$ is exponentially distributed with parameter $\lambda_i = 1/\xi_i^2$. Therefore, $\sum_{i=1}^{N_r} |\omega_i|^2$ is the sum of N_r independent exponentially distributed random variables. The set of distributions composed by independent exponentially distributed phases are referred to as phase type distributions [34] and has a closed form description with matrix exponential functions. That is, the density function of $\sum_{i=1}^{N_r} |\omega_i|^2$ is

$$f(x) = e_1^T e^{\mathbf{A}x} e_{N_r} \lambda_{N_r}, \quad (31)$$

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

$$\mathbf{A} = \begin{pmatrix} -\lambda_1 & \lambda_1 & & & \\ & -\lambda_2 & \lambda_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \\ & & & & -\lambda_{N_r} \end{pmatrix}. \quad (32)$$

Based on $f(x)$ and (29) the MSE and the SE can be calculated as follows:

$$\begin{aligned} \text{MSE} &= \mathbb{E}_\nu(\text{MSE}(\boldsymbol{\nu})) = \mathbb{E}_\omega(\text{MSE}(\boldsymbol{\omega})) = \\ &= \int_x \frac{1}{\alpha_\ell^2 P_\ell x + 1} f(x) dx, \end{aligned} \quad (33)$$

$$\begin{aligned}\eta &= -\mathbb{E}_{\nu}(\log[\text{MSE}(\nu)]) = -\mathbb{E}_{\omega}(\log[\text{MSE}(\omega)]) \\ &= -\int_x \log\left[\frac{1}{\alpha_\ell^2 P_\ell x + 1}\right] f(x) dx.\end{aligned}\quad (34)$$

This general case simplifies to the following two special ones.

A. Case 1: Distinct Variances

We will now assume that N is the number of non-zero singular values in \mathbf{S}_Ω and all non-zero ξ_i (and λ_i) are distinct (different). In this case

$$f(x) = \sum_{i=1}^N \frac{\lambda_i e^{-\lambda_i x}}{\prod_{j=1, j \neq i}^N \left(1 - \frac{\lambda_i}{\lambda_j}\right)}, \quad (35)$$

and for the MSE we get:

$$\text{MSE} = \sum_{i=1}^N \frac{-\lambda_i^{\frac{2-N}{2}} e^{\frac{\lambda_i}{p}} E_{\text{in}}\left(1, \frac{-\lambda_i}{p}\right)}{p \prod_{j=1, j \neq i}^N \left(1 - \frac{\lambda_i}{\lambda_j}\right)}, \quad (36)$$

where recall from Proposition 1 that $p = \alpha^2 P_\ell$. For the SE we get:

$$\eta = \sum_{i=1}^N \frac{-\lambda_i^{\frac{2-N}{2}} e^{\frac{\lambda_i}{p}} E_{\text{in}}\left(1, \frac{-\lambda_i}{p}\right)}{\prod_{j=1, j \neq i}^N \left(1 - \frac{\lambda_i}{\lambda_j}\right)}. \quad (37)$$

B. Case 2: All Variances of ω are Equal

Suppose $\xi_i = \xi = \lambda^{-1/2}$, $\forall i \leq N$. In this case, the phase type distribution simplifies to the Erlang distribution:

$$f(x, N, \lambda) = \frac{\lambda^N x^{N-1} e^{-\lambda x}}{(N-1)!}, \quad (38)$$

and we get:

$$\text{MSE} = \frac{\lambda}{p} e^{\frac{\lambda}{p}} E_{\text{in}}\left(N, \frac{\lambda}{p}\right), \quad (39)$$

$$\eta = \frac{\mathcal{G}\left(\frac{\lambda}{p}\right)}{a^N (N-1)!},$$

where

$$\mathcal{G}(x) \triangleq \text{MeijerG}_{1,0}^{3,1} \left(\begin{matrix} -N_r; -(N_r-1) \\ -N_r, -N_r, 0; \end{matrix} \middle| x \right), \quad (40)$$

is the Meijer G function.

In case of identical variances in ω , (39) gives the same expression as (13) in accordance with the fact that \mathbf{S}_Ω is proportional to the identity matrix.

VII. NUMERICAL ANALYSIS OF THE MSE

A. Channel Model and Covariance Matrix

In this section we consider a single cell system, in which MSs use orthogonal pilots to facilitate the estimation of the uplink channel by the BS. Recall from Section III that the

Table I
SYSTEM PARAMETERS

Parameter	Value
Number of antennas	$N_r = 4, 16, 20, 64, 100, 500$
Path Loss	$\alpha_\ell = 40, 45, 50$ dB
Power budget	$\tau_p P_{p,\ell} + \tau_d P_\ell = P_{\text{tot}} = 250$ mW, as in Eq. (1).
Total number of symbols (per time slot)	$F = 12$
Antenna spacing	$D/\lambda = 0.15, \dots, 1.5$
Mean Angle of Arrival (AoA)	$\theta = 70^\circ$
Angular spread	$2 \cdot \theta_\Delta = 5, \dots, 45^\circ$

channel estimation process is independent for each MS and we can therefore focus on a single user. The covariance matrix \mathbf{C}_ℓ of the channel \mathbf{h}_ℓ as the function of the antenna spacing, mean angle of arrival and angular spread is modeled as by the well known spatial channel model, which is known to be accurate in non-line-of-sight environment with rich scattering and all antenna elements identically polarized, see [35]. For uniformly distributed angle of arrivals, the (m, n) ($m, n \in \{1, \dots, N_r\}$) element of the covariance matrix of User- ℓ \mathbf{C}_ℓ is given by

$$\mathbf{C}_{m,n} = \frac{1}{2\theta_\Delta} \int_{-\theta_\Delta}^{\theta_\Delta} e^{j \cdot 2\pi \cdot \frac{D}{\lambda} (n-m) \cos(\bar{\theta}+x)} dx, \quad (41)$$

where the system parameters are given in Table I. The covariance matrix \mathbf{C}_ℓ becomes practically diagonal as the antenna spacing and the angular spread grows beyond $D\lambda > 1$ and $\theta_\Delta > 30^\circ$. In contrast, with critically spaced antennas $D\lambda = 0.5$ and $\theta_\Delta < 10^\circ$, the antenna correlation in terms of the off-diagonal elements of \mathbf{C}_ℓ can be considered strong. Note that modeling the correlation matrices at the receiver side according to (41) corresponds to using the one-sided narrowband Kronecker model with receiver-side correlation, which is an appropriate model for the uplink of MU MIMO systems [31].

B. Numerical Results

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. Note that the performance characteristics of the proposed MMSE receiver 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 (10). The key input parameters to this system that are necessary to obtain numerical results using the MSE derivation in this paper are listed in Table I.

Figure 1 shows the cumulative distribution function (CDF) of the squared error of the estimated data symbols at the BS, i.e. the CDF of $\|\mathbf{G}\mathbf{y} - x\|^2$ using the naïve and the MMSE receiver when the number of antennas at the BS is $N_r = 500$. In all three cases in terms of path loss ($\alpha_\ell = 40$, $\alpha_\ell = 45$ and $\alpha_\ell = 50$ dB), the gain of the MMSE receiver is large in the entire region of the CDF. For example, at $\alpha_\ell = 40$, the

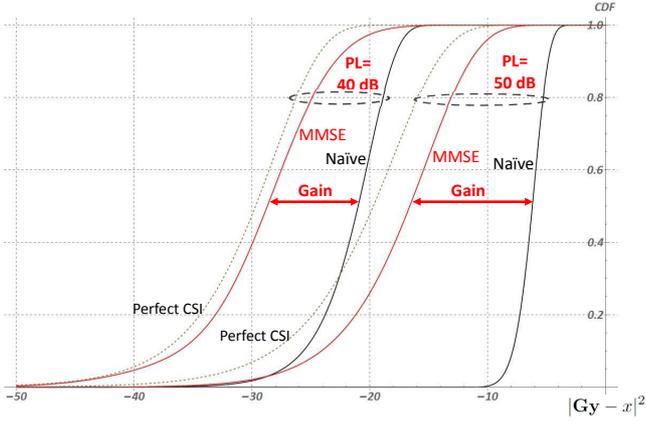


Figure 1. Cumulative distribution function of the squared error in a single user MIMO scenario when the path loss between the UE and the BS is set to 40 and 50 dB when using the naïve receiver, the MMSE receiver and the receiver which has access to the perfect CSI with $N_r = 500$ antennas.

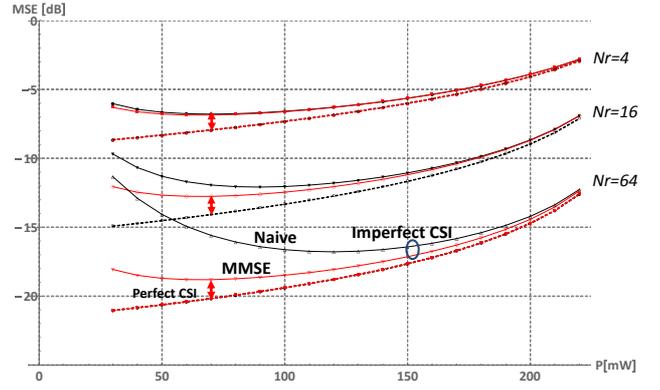


Figure 3. Comparing the MSE performance of the naïve and MMSE receivers with that of a receiver that uses perfect CSI. As the pilot power increases, the MSE achieved by the receiver that uses perfect CSI increases, because due to the sum power constraint the transmit power available for the data symbols decreases.

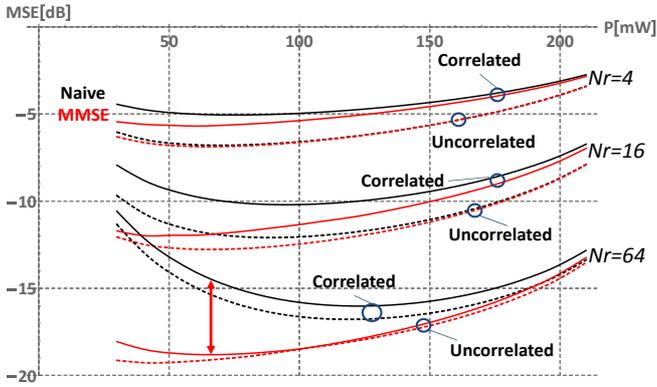


Figure 2. Comparing the performance of the naïve and the MMSE receiver in the case of correlated (solid lines) and uncorrelated (dashed lines) antennas (with $N_r = 4, 16, 64$).

median of the CDF is -21 dB with the naïve receiver and -29 dB with the MMSE receiver. This result indicates that using the MMSE receiver is advantageous not only in the average sense, but in virtually all channel states.

Figure 2 examines the impact of antenna correlation on the MSE with the naïve and the MMSE receivers. The impact of antenna correlation in terms of the achievable MSE decreases as the number of antennas increases from $N_r = 4$ to $N_r = 64$. An intuitive explanation of this insight is that the impact of correlation can be thought of as a factor that decreases the effective number of antennas, that is the number of antennas which contribute to the estimation of the transmitted data symbol. As the number of antennas grows large, antenna correlation decreases the effective number of antennas, but the loss due to this is not as significant as this loss when the number antennas is low. Instead, as the figure shows, at large number of antennas tuning the pilot power plays a more important role in minimizing the MSE than the effect of antenna correlation.

Figure 3 compares the performance of the naïve and the MMSE receivers with that of a receiver that has access to the perfect CSI, that is assuming that $\hat{\mathbf{h}}_\ell = \mathbf{h}_\ell$. This situation corresponds to $\mathbf{D}_\ell = \mathbf{I}$ and $\mathbf{Q}_\ell = \mathbf{0}$ and the structure of the naïve and the MMSE receivers coincide. Indeed, we recall that the naïve receiver does minimize the MSE in the case of perfect CSI. The key aspect to observe in Figure 3 is that the gap between the MMSE receiver and the receiver operating with perfect CSI does not depend on N_r . This is in sharp contrast with the gap between the naïve receiver and the receiver with perfect CSI, which largely increases as the number of antennas gets large.

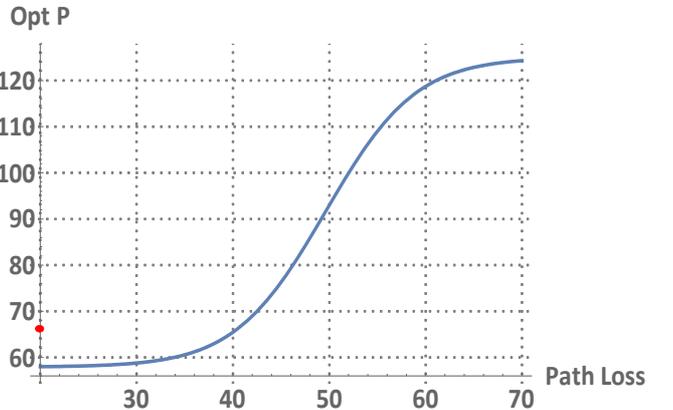


Figure 4. Optimum pilot power as a function of the path loss α_ℓ . The red dot indicates the optimum pilot power when using the MMSE receiver at $\alpha_\ell = 40$ dB, which is ≈ 68 mW.

Figure 4 shows the optimum pilot power as the function of the path loss when using the MMSE receiver. Recall that the optimum pilot power is independent of the number of antennas. This figure suggests that the pilot power should be tuned based on large scale fading. Although the optimum pilot

power is independent of N_r , the previous figures show that the importance of proper PDPR tuning increases when the number of antennas gets large. This is because, as visible in Figures 2, 3, and 5, the gap between the naïve and the MMSE receiver increases with the number of antennas.

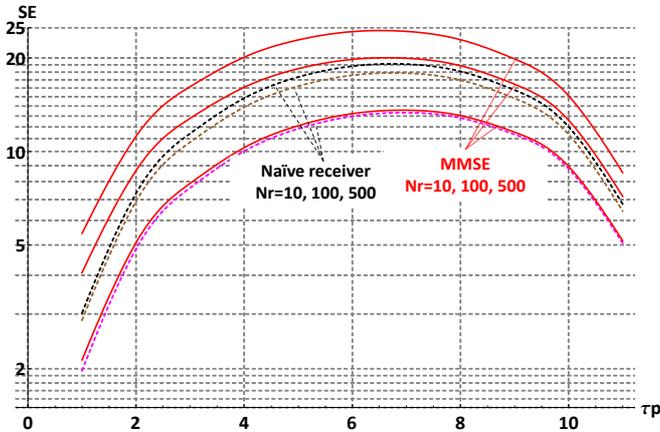


Figure 5. Spectral efficiency as a function of the employed pilot symbols τ_p . In this example, the number of users in MU MIMO system is set equally to τ_p , that is we assume that the number of users that can be spatially multiplexed equals the pilot sequence length.

Figure 5 shows the SE of a MU MIMO system, in which the number of spatially multiplexed users is equal to the length of the employed pilot sequence τ_p . Figure 5 illustrates the trade-off between increasing the number of MU MIMO users and the necessary number of pilot symbols used to create orthogonal pilot sequences. A greater number of users increases the SE of the system at the expense of spending more symbols on the pilot signals. Therefore, we can see that around $\tau_p = 6$ the SE reaches its highest value. The gain in terms of SE of using the MMSE receiver is around 25% when the number of antennas is large.

VIII. CONCLUSIONS

In this paper, we first derived an analytical expression (\mathbf{G}_ℓ^*) of a linear receiver structure that minimizes the MSE of the uplink estimated data symbols when the receive antennas possess a known correlation structure. We then derived closed form expressions for the MSE and the achievable SE when employing this MMSE receiver as a function of the pilot and data power, number of antennas, and path loss. We used Monte Carlo simulations to verify the analytical results and to gain insight into the system behavior when using \mathbf{G}_ℓ^* .

From the analysis we conclude that when employing the true MMSE receiver (\mathbf{G}^*) at the BS in a MU MIMO system, the pilot power that minimizes the MSE is independent of the number of receive antennas. This implies that the optimal training does not need to be adjusted for sites with different numbers of antennas or when upgrading existing antenna sites to a larger number of antennas. In the special, (but in practice, typical) case when the thermal noise power levels on the data

and pilot signals are equal, setting the pilot power by the terminal is easy, because the terminals continuously measure the path loss to the serving BS.

The simulation results provide the following insights:

- The performance difference between the naïve and the MMSE receiver increases with an increasing number of antennas. However, the performance gap between the MMSE receiver and the receiver that has access to a perfect CSI does not increase with the number of antennas.
- When the number of antennas is large, the impact of antenna correlation on the MSE is relatively small as compared with the impact of appropriately tuning the pilot power. When using the MMSE receiver (\mathbf{G}_ℓ^*), the optimal pilot power does not depend on the number of receive antennas, but is quite sensitive to large-scale fading.
- When the number of antennas is large, the gain of using the MMSE receiver over using the naïve receiver is large, not only in terms of MSE, but also in the entire CDF of the squared error of the estimated data symbols.

We also showed that the well known relation between the MSE and the SE that holds for the case when perfect CSI at the MMSE receiver is available is valid also for the case of imperfect CSI at the regularized MMSE receiver (\mathbf{G}^*). The deeper analysis of the impact of CSI errors in the case of non-separable channel models is an important topic for future research.

ACKNOWLEDGMENTS

We would like to thank the Associated Editor and the Reviewers for the constructive comments that helped us improve the contents and the presentation. The Authors thank Johan Söder and Yngve Selén for their valuable comments on an early version of the manuscript.

APPENDIX I

Proof of Proposition 1. If $\mathbf{C}_\ell = c_\ell \mathbf{I}$, implying $\mathbf{D}_\ell = d_\ell \mathbf{I}$, $\mathbf{Q}_\ell = q_\ell \mathbf{I}$ and the optimal \mathbf{G}_ℓ^* of (10) can be written as:

$$\mathbf{G}_\ell^* = \frac{\alpha_\ell \sqrt{P_\ell} d_\ell}{\alpha_\ell^2 P_\ell (d_\ell^2 \|\hat{\mathbf{h}}_\ell\|^2 + q_\ell) + \sum_{k \neq \ell}^K \alpha_k^2 P_k c_k + \sigma_d^2} \hat{\mathbf{h}}_\ell^H \triangleq g_\ell \cdot \hat{\mathbf{h}}_\ell^H. \quad (42)$$

Substituting \mathbf{G}_ℓ^* into (7) we get:

$$\text{MSE}(\hat{\mathbf{h}}_\ell) = -2\alpha_\ell \sqrt{P_\ell} g_\ell d_\ell \|\hat{\mathbf{h}}_\ell\|^2 + 1 + g_\ell^2 \cdot \left(\alpha_\ell^2 P_\ell d_\ell^2 \|\hat{\mathbf{h}}_\ell\|^4 + \left(\alpha_\ell^2 P_\ell q_\ell + \sum_{k \neq \ell}^K \alpha_k^2 P_k c_k + \sigma_d^2 \right) \|\hat{\mathbf{h}}_\ell\|^2 \right). \quad (43)$$

Recognizing that $\varphi_\ell \triangleq \|\hat{\mathbf{h}}_\ell\|^2$ is Gamma distributed, the density function of $\varphi_\ell \forall \ell$ is given by (dropping the index ℓ for convenience):

$$f_\varphi(x) = \frac{r^{-N_r} x^{N_r-1} e^{-x/r}}{(N_r - 1)!} \quad x > 0. \quad (44)$$

Proposition 1 follows from Lemma (43) by taking the average of MSE ($\hat{\mathbf{h}}_\ell$) using the the following integrals:

$$\int_{x=0}^{\infty} T_1 f_\varphi(x) dx = s_\ell \cdot \frac{N_r \left(-s_\ell r_\ell + e^{\frac{b_\ell}{s_\ell r_\ell}} (b_\ell + (1 + N_r) s_\ell r_\ell) E_{in} \left(1 + N_r, \frac{b_\ell}{s_\ell r_\ell} \right) \right)}{s_\ell^2 r_\ell}; \quad (45)$$

$$\int_{x=0}^{\infty} T_2 f_\varphi(x) dx = b_\ell \cdot \frac{-s_\ell r_\ell + e^{\frac{b_\ell}{s_\ell r_\ell}} (b_\ell + N_r s_\ell r_\ell) E_{in} \left(N_r, \frac{b_\ell}{s_\ell r_\ell} \right)}{s_\ell^2 r_\ell^2}; \quad (46)$$

$$\int_{x=0}^{\infty} T_3 f_\varphi(x) dx = 2 \cdot e^{\frac{b_\ell}{s_\ell r_\ell}} N_r E_{in} \left(1 + N_r, \frac{b_\ell}{s_\ell r_\ell} \right), \quad (47)$$

where $E_{in}(n, z) \triangleq \int_1^\infty e^{-zt}/t^n dt$ is a standard exponential integral function. \square

APPENDIX II

Proof of Lemma 2. We rewrite the MSE expression in (11), by making use of the following recursive relation, from [36] (also available at [37, 8.19.12]):

$$\mu_\ell E_{in}(N_r, \mu_\ell) + N_r E_{in}(N_r + 1, \mu_\ell) = e^{-\mu_\ell}. \quad (48)$$

Substituting $\mu_\ell = \frac{b}{rs}$ in this relation, using the terms of the MSE in (11) and rearranging, we obtain:

$$\text{MSE} = \frac{b_\ell}{r_\ell s_\ell} e^{\frac{b_\ell}{r_\ell s_\ell}} E_{in} \left(N_r, \frac{b_\ell}{r_\ell s_\ell} \right), \quad (49)$$

where, similarly to the notation used in Proposition 1, $b_\ell \triangleq q_\ell p_\ell + \sigma_d^2$ with $p_\ell \triangleq \alpha_\ell^2 P_\ell$ and $s_\ell \triangleq d_\ell^2 p_\ell$ and r_ℓ, d_ℓ and q_ℓ are defined in (12).

Finally, recognizing that:

$$\begin{aligned} \mu_\ell &= \frac{b_\ell}{r_\ell s_\ell} = \frac{q_\ell \alpha_\ell^2 P_\ell + \sigma_d^2}{d_\ell^2 \alpha_\ell^2 P_\ell r_\ell} = \\ &= \frac{\sigma_d^2 \sigma_p^2 \tau_d + c_\ell \alpha_\ell^2 (\sigma_p^2 P_{tot} + \tau_p P_{p,\ell} (\sigma_d^2 \tau_d - \sigma_p^2))}{c_\ell^2 \alpha_\ell^4 P_{p,\ell} \tau_p (P_{tot} - P_{p,\ell} \tau_p)} \end{aligned} \quad (50)$$

and substituting (1) into P_ℓ the lemma follows. \square

APPENDIX III

We first prove the following lemma that will be useful for the proof of Proposition 3.

Lemma 6. For $n > 0$ and $x \geq 0$, the following limit holds:

$$\lim_{x \rightarrow \infty} x^2 (1 - e^x (n+x) E_{in}(n, x)) = -n. \quad (51)$$

Proof of Lemma 6. Recalling the basic relationship between the incomplete Gamma function and the exponential integral function:

$$E_{in}(n, x) = x^{n-1} \Gamma(1-n, x), \quad (52)$$

and using the following expansion formula that is valid for large values of x (see [38]):

$$\begin{aligned} \Gamma(1-n, x) &\sim \\ &\sim x^{-n} e^{-x} \left(1 + \frac{-n}{x} + \frac{-n(-n-1)}{x^2} + \right. \\ &\quad \left. + \frac{-n(-n-1)(-n-2)}{x^3} + \dots \right), \end{aligned} \quad (53)$$

we have:

$$\begin{aligned} x^2 (1 - e^x (n+x) E_{in}(n, x)) &\sim x^2 (1 - e^x (n+x) x^{n-1} x^{-n} e^{-x} \\ &\cdot (1 + \frac{-n}{x} + \frac{-n(-n-1)}{x^2} + \frac{-n(-n-1)(-n-2)}{x^3} + \dots)). \end{aligned} \quad (54)$$

Rearranging terms, we finally get, for large x :

$$\begin{aligned} x^2 (1 - e^x (n+x) E_{in}(n, x)) &\sim \\ &-n + \frac{2n(1+n)}{x} - \frac{3n(1+n)(2+n)}{x^2} + \\ &+ \frac{4n(1+n)(2+n)(3+n)}{x^3} \mp \dots \end{aligned} \quad (55)$$

from which it follows:

$$\lim_{x \rightarrow \infty} x^2 (1 - e^x (n+x) E_{in}(n, x)) = -n. \quad (56)$$

\square

We can now prove Proposition 3.

Proof of Proposition 3. We begin by taking the first derivative of MSE as a function of $P_{p,\ell}$. To this end, we use (11) and take the derivative of the MSE with respect to μ_ℓ :

$$\begin{aligned} \text{MSE}'(\mu_\ell) &= -\mu_\ell e^{\mu_\ell} E_{in}(N_r - 1, \mu_\ell) + \\ &+ e^{\mu_\ell} E_{in}(N_r, \mu_\ell) + \mu_\ell e^{\mu_\ell} E_{in}(N_r). \end{aligned} \quad (57)$$

After some algebraic manipulation based on (48), we obtain:

$$\text{MSE}'(\mu_\ell) = e^{\mu_\ell} (N_r + \mu_\ell) E_{in}(N_r, \mu_\ell) - 1. \quad (58)$$

From [36] (also available at [37, 8.19.21]) we have

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

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

Next, we consider the first derivative of $\mu(P_{p,\ell})$ as defined in (14) with respect to $P_{p,\ell}$:

$$\begin{aligned} \mu'(P_{p,\ell}) &= \frac{\sigma_d^2 \sigma_p^2 \tau_d (2P_{p,\ell} \tau_p - P_{tot})}{c_\ell^2 P_{p,\ell}^2 \alpha_\ell^4 \tau_p (P_{tot} - \tau_p P_{p,\ell})^2} + \\ &+ \frac{c_\ell \alpha_\ell^2 \left(P_{p,\ell}^2 \tau_p^2 (\sigma_d^2 \tau_d - \sigma_p^2) + 2P_{p,\ell} P_{tot} \sigma_p^2 \tau_p - \sigma_p^2 P_{tot}^2 \right)}{c_\ell^2 P_{p,\ell}^2 \alpha_\ell^4 \tau_p (P_{tot} - \tau_p P_{p,\ell})^2}. \end{aligned} \quad (59)$$

The numerator of (59) is a second order polynomial of $P_{p,\ell}$ with the following coefficients $a_0 = -P_{tot}z$, $a_1 = 2\tau_p z$, $a_2 = c_\ell \alpha_\ell^2 \tau_p^2 (\sigma_d^2 \tau_d - \sigma_p^2)$, where $z = (c_\ell P_{tot} \alpha_\ell^2 + \sigma_d^2 \tau_d) \sigma_p^2$. a_0 is negative, a_1 is positive and the sign of a_2 depends on the sign of $\sigma_d^2 \tau_d - \sigma_p^2$. In reasonable cases a_2 is positive as well, because $\tau_d > 1$ and $\sigma_d \approx \sigma_p$. When a_2 is positive the numerator of (59) has one positive and one negative root, because $a_1 < \sqrt{a_1^2 - a_0 a_2}$ and the positive root is

$$P_{p,\ell}^* = \frac{-a_1 + \sqrt{a_1^2 - 4a_0 a_2}}{2a_2}. \quad (60)$$

Finally, the first derivative of the MSE with respect to $P_{p,\ell}$ is:

$$\frac{d}{dP_{p,\ell}} \text{MSE} = \text{MSE}'(\mu_\ell) \cdot \mu'(P_{p,\ell}). \quad (61)$$

Recall that $\text{MSE}'(\mu_\ell) \neq 0$, the roots of $\frac{d}{dP_{p,\ell}} \text{MSE}$ are identical with the roots of the numerator of (59) and the positive root of $\frac{d}{dP_{p,\ell}} \text{MSE}$ is $P_{p,\ell}^*$.

We still need to show that $P_{p,\ell}^*$ corresponds to a local minimum. To this end, we study the sign of $\lim_{P_{p,\ell} \rightarrow 0^+} \frac{d}{dP_{p,\ell}} \text{MSE}$. If the limit is negative then $P_{p,\ell}^*$ corresponds to a local minimum. Unfortunately, (61) is not directly applicable because $\lim_{P_{p,\ell} \rightarrow 0^+} \mu'(P_{p,\ell}) = 0$ and $\lim_{\mu_\ell \rightarrow \infty} \text{MSE}'(\mu_\ell) = \infty$. Instead, according to (49) we introduce $F(n, x) = x e^x E_{in}(n, x)$ and rewrite (50) as

$$\mu_\ell = \frac{b_1 + b_2 P_{p,\ell}}{P_{p,\ell} (b_3 - P_{p,\ell})},$$

where $b_1 = \frac{\sigma_d^2 \sigma_p^2 \tau_d + c_\ell \alpha_\ell^2 P_{tot} \sigma_p^2}{c_\ell^2 \alpha_\ell^4 \tau_p^2}$, $b_2 = \frac{c_\ell \alpha_\ell^2 \tau_p (\sigma_d^2 \tau_d - \sigma_p^2)}{c_\ell^2 \alpha_\ell^4 \tau_p^2}$, $b_3 = \frac{P_{tot}}{\tau_p}$ and note that b_1 and b_3 are positive. This way $\text{MSE} = F\left(N_r, \frac{b_1 + b_2 P_{p,\ell}}{P_{p,\ell} (b_3 - P_{p,\ell})}\right)$. Introducing $\tilde{P}_\ell = \frac{b_1 P_{p,\ell} (b_3 - P_{p,\ell})}{b_3 (b_1 + b_2 P_{p,\ell})}$ we also have $\text{MSE} = F\left(N_r, \frac{b_1}{\tilde{P}_\ell b_3}\right)$ and can rewrite the limit as

$$\begin{aligned} \lim_{P_{p,\ell} \rightarrow 0^+} \frac{d}{dP_{p,\ell}} \text{MSE} &= \lim_{P_{p,\ell} \rightarrow 0^+} \frac{d}{dP_{p,\ell}} F\left(N_r, \frac{b_1 + b_2 P_{p,\ell}}{P_{p,\ell} (b_3 - P_{p,\ell})}\right) = \\ &= \lim_{P_{p,\ell} \rightarrow 0^+} \frac{d}{dP_{p,\ell}} \tilde{P}_\ell \lim_{\tilde{P}_\ell \rightarrow 0^+} \frac{d}{d\tilde{P}_\ell} F\left(N_r, \frac{b_1}{\tilde{P}_\ell b_3}\right). \end{aligned}$$

The first term converges to 1, while the second term converges to $-\frac{N_r b_3}{b_1}$ based on (51). \square

APPENDIX IV

Proof of Lemma 4. According to the matrix inversion lemma for matrices \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} of size $n \times n$, $n \times m$, $m \times m$, $m \times n$,

respectively, we have

$$\begin{aligned} (\mathbf{A} + \mathbf{BCD})^{-1} &= \\ \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{DA}^{-1} \mathbf{B} + \mathbf{C}^{-1})^{-1} \mathbf{DA}^{-1}. \end{aligned}$$

Substituting $\mathbf{A} = \mathbf{\Psi}_\ell$, $\mathbf{B} = \alpha_\ell \sqrt{P_\ell} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell$, $\mathbf{C} = 1$, $\mathbf{D} = \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H$ we have:

$$\left(\mathbf{\Psi}_\ell + \alpha_\ell^2 P_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \right)^{-1} = \quad (62)$$

$$\begin{aligned} &= \mathbf{\Psi}_\ell^{-1} - \mathbf{\Psi}_\ell^{-1} \alpha_\ell \sqrt{P_\ell} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell \cdot \\ &\cdot \left(\alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{\Psi}_\ell^{-1} \alpha_\ell \sqrt{P_\ell} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell + 1 \right)^{-1} \cdot \\ &\cdot \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{\Psi}_\ell^{-1} = \\ &= \mathbf{\Psi}_\ell^{-1} - \frac{\alpha_\ell^2 P_\ell}{\alpha_\ell^2 P_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{\Psi}_\ell^{-1} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell + 1} \\ &\cdot \mathbf{\Psi}_\ell^{-1} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{\Psi}_\ell^{-1}. \end{aligned} \quad (63)$$

Substituting (63) into (16) gives:

$$\begin{aligned} \mathbf{G}_\ell^* &= \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \cdot \\ &\cdot \left(\mathbf{\Psi}_\ell^{-1} - \frac{\alpha_\ell^2 P_\ell}{\alpha_\ell^2 P_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{\Psi}_\ell^{-1} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell + 1} \mathbf{\Psi}_\ell^{-1} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{\Psi}_\ell^{-1} \right). \end{aligned} \quad (64)$$

Recall that $\mathbf{\Psi}_\ell = \mathbf{\Theta}_\ell^H \mathbf{S}_\ell \mathbf{\Theta}_\ell$ is the SVD of $\mathbf{\Psi}_\ell$. Substituting (19) - (20), and this SVD into (64) we get:

$$\begin{aligned} \mathbf{G}_\ell^* &= \alpha_\ell \sqrt{P_\ell} \left(\mathbf{\nu}_\ell^H \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell - \right. \\ &\left. \frac{\alpha_\ell^2 P_\ell}{\alpha_\ell^2 P_\ell \|\mathbf{\nu}_\ell\|^2 + 1} \|\mathbf{\nu}_\ell\|^2 \mathbf{\nu}_\ell^H \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell \right) = \\ &= \alpha_\ell \sqrt{P_\ell} \left(1 - \frac{\alpha_\ell^2 P_\ell \|\mathbf{\nu}_\ell\|^2}{\alpha_\ell^2 P_\ell \|\mathbf{\nu}_\ell\|^2 + 1} \right) \mathbf{\nu}_\ell^H \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell = \\ &= \frac{\alpha_\ell \sqrt{P_\ell}}{\alpha_\ell^2 P_\ell \|\mathbf{\nu}_\ell\|^2 + 1} \mathbf{\nu}_\ell^H \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell. \end{aligned} \quad (65)$$

\square

APPENDIX V

Proof of Lemma 5. Substituting (17) into (7) with optimal MMSE receiver, we get:

$$\begin{aligned} \text{MSE}(\hat{\mathbf{h}}_\ell) &= \mathbf{G}_\ell^* \left(\alpha_\ell^2 P_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H + \mathbf{\Psi}_\ell \right) \mathbf{G}_\ell^{*H} \\ &- \alpha_\ell \sqrt{P_\ell} (\mathbf{G}_\ell^* \mathbf{D}_\ell \hat{\mathbf{h}}_\ell + \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{G}_\ell^{*H}) + 1. \end{aligned} \quad (66)$$

From (22), (21) and the SVD of $\mathbf{\Psi}_\ell$ we have:

$$\mathbf{G}_\ell^* \mathbf{D}_\ell \hat{\mathbf{h}}_\ell = g_\ell \mathbf{\nu}_\ell^H \mathbf{\nu}_\ell = g_\ell \|\mathbf{\nu}_\ell\|^2, \quad (67)$$

$$\mathbf{G}_\ell^* \mathbf{\Psi}_\ell \mathbf{G}_\ell^{*H} = g_\ell^2 \mathbf{\nu}_\ell^H \mathbf{\nu}_\ell = g_\ell^2 \|\mathbf{\nu}_\ell\|^2, \quad (68)$$

substituting this into (66) we obtain

$$\begin{aligned} \text{MSE}(\mathbf{\nu}_\ell) &= \\ \alpha_\ell^2 P_\ell g_\ell^2 \|\mathbf{\nu}_\ell\|^4 + g_\ell^2 \|\mathbf{\nu}_\ell\|^2 - 2\alpha_\ell \sqrt{P_\ell} g_\ell \|\mathbf{\nu}_\ell\|^2 + 1, \end{aligned} \quad (69)$$

where g_ℓ is also a function of $\|\nu_\ell\|^2$ according to (23). Substituting (23) into (69) gives the lemma after some algebraic manipulations. \square

REFERENCES

- [1] D. Porrat, "Information Theory of Wideband Communications," *IEEE Communications Surveys and Tutorials*, vol. 9, no. 2, pp. 2–16, July 2007.
- [2] G. Fodor, P. D. Marco, and M. Telek, "On the Impact of Antenna Correlation on the Pilot-Data Balance in Multiple Antenna Systems," in *IEEE International Conference on Communications (ICC)*, London, UK, June 8-12 2015, pp. 2590–2596.
- [3] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.
- [4] X. Zhang, D. P. Palomar, and B. Ottersten, "Robust MAC MIMO Transceiver Design with Partial CSIT and CSIR," in *IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Pacific Grove, CA, USA, 4-7 November 2007, pp. 324–328.
- [5] T. Kim and J. G. Andrews, "Optimal Pilot-to-Data Power Ratio for MIMO-OFDM," *IEEE Globecom*, pp. 1481–1485, December 2005.
- [6] —, "Balancing Pilot and Data Power for Adaptive MIMO-OFDM Systems," *IEEE Globecom*, pp. 1–5, 2006.
- [7] N. Jindal and A. Lozano, "A Unified Treatment of Optimum Pilot Overhead in Multipath Fading Channels," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2939–2948, October 2010.
- [8] V. K. V. Gottomukkala and H. Minn, "Capacity Analysis and Pilot-Data Power Allocation for MIMO-OFDM With Transmitter and Receiver IQ Imbalances and Residual Carrier Frequency Offset," *IEEE Trans. Veh. Techn.*, pp. 553–565, 2012.
- [9] K. Min, M. Jung, T. Kim, Y. Kim, J. Lee, and S. Choi, "Pilot Power Ratio for Uplink Sum-Rate Maximization in Zero-Forcing Based MU-MIMO Systems with Large Number of Antennas," in *IEEE Vehicular Technology Conference (VTC-Fall)*, 2-5 September 2013, pp. 1–5.
- [10] E. Eraslan, B. Daneshmand, and C.-Y. Lou, "Performance Indicator for MIMO MMSE Receivers in the Presence of Channel Estimation Error," *IEEE Wireless Commun. Letters*, vol. 2, no. 2, pp. 211–214, 2013.
- [11] K. Guo, Y. Guo, G. Fodor, and G. Ascheid, "Uplink Power Control with MMSE Receiver in Multi-Cell MU-Massive-MIMO Systems," in *IEEE International Conference on Communications (ICC)*, 10-14 June 2014, pp. 5184–5190.
- [12] 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.
- [13] X. Li, E. Björnsson, E. G. Larsson, S. Zhou, and J. Wang, "Massive MIMO with Multi-cell MMSE Processing: Exploiting All Pilots for Interference Suppression," *arXiv:1505.03682v2 [cs.IT]*, May 2015.
- [14] X. Tang, M.-S. Alouini, and A. J. Goldsmith, "Effect of Channel Estimation Error on M-QAM BER Performance in Rayleigh Fading," *IEEE Transactions on Communications*, vol. 47, no. 12, pp. 1856–1864, December 1999.
- [15] Y. G. Li, J. H. Winters, and N. R. Sollenberger, "MIMO-OFDM for Wireless Communications: Signal Detection With Enhanced Channel Estimation," *IEEE Transactions on Communications*, vol. 50, no. 9, pp. 1471–1477, September 2002.
- [16] L. Zheng, D. Tse, and M. Médard, "Channel Coherence in the Low-SNR Regime," *IEEE Trans. Information Theory*, vol. 53, no. 3, pp. 976–997, March 2007.
- [17] S. Ray, M. Médard, and L. Zheng, "On Noncoherent MIMO Channels in the Wideband Regime: Capacity and Reliability," *IEEE Trans. Information Theory*, vol. 53, no. 6, pp. 1983–2009, June 2007.
- [18] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of Sparse Multipath Channels in the Ultra-Wideband Regime," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 357–371, October 2007.
- [19] G. Hariharan, V. Raghavan, and A. M. Sayeed, "Capacity of Sparse Wideband Channels with Partial Channel Feedback," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 475–493, June 2008.
- [20] D. Hammarwall, M. Bengtsson, and B. E. Ottersten, "Acquiring Partial CSI for Spatially Selective Transmission by Instantaneous Channel Norm Feedback," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 1188–1204, March 2008.
- [21] —, "Utilizing the Spatial Information Provided by Channel Norm Feedback in SDMA Systems," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 3278–3293, 2008.
- [22] V. Raghavan, S. V. Hanly, and V. Veeravalli, "Statistical Beamforming on the Grassmann Manifold for the Two-User Broadcast Channel," *IEEE Trans. Information Theory*, vol. 59, no. 10, pp. 6464–6489, October 2013.
- [23] J. Wang, M. Bengtsson, B. Ottersten, and D. Palomar, "Robust MIMO Precoding for Several Classes of Channel Uncertainty," *IEEE Trans. Signal Processing*, vol. 61, no. 12, pp. 3056–3070, April 2013.
- [24] M. Ding and S. D. Blostein, "Relation Between Joint Optimizations for Multiuser MIMO Uplink and Downlink with Imperfect CSI," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, USA, March 31-April 4 2008, pp. 3149 – 3152.
- [25] 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.
- [26] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the Capacity of Frequency-Selective Channels in Training-Based Transmission Schemes," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2572–2583, September 2004.
- [27] T. Marzetta, "How Much Training is Needed for Multiuser MIMO?" *IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)*, pp. 359–363, June 2006.
- [28] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A Coordinated Approach to Channel Estimation in Large-Scale Multiple-Antenna Systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, February 2013.
- [29] X. Hou, Z. Zhang, and H. Kayama, "DMRS Design and Channel Estimation for LTE-Advanced MIMO Uplink," *70th IEEE Vehicular Technology Conference (Fall)*, pp. 1–5, 2009.
- [30] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. WILEY, 2nd edition, 2011, ISBN-10: 0470660252.
- [31] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO - Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, 2013.
- [32] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*. Prentice Hall, 1993, no. ISBN: 0133457117.
- [33] 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.
- [34] M. F. Neuts, *Matrix Geometric Solutions in Stochastic Models*. Baltimore, MD, USA: The Johns Hopkins University Press, June 1981, no. ISBN-10: 0801825601.
- [35] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A Coordinated Approach to Channel Estimation in Large-scale Multiple-antenna Systems," <http://arxiv.org/abs/1203.5924>, no. arXiv:1203.5924v1 [cs.IT], March 2012.
- [36] 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.
- [37] P. D. Gallagher, "NIST Handbook of Mathematical Functions," <http://dlmf.nist.gov/8.19>, 2010, [Online; accessed 8-Oct-2015].
- [38] N. M. Temme, "Uniform Asymptotic Expansions of the Incomplete Gamma Functions and the Incomplete Beta Function," *Mathematics of Computation*, vol. 29, no. 132, pp. 1109–1114, 1975, American Mathematical Society.