2016

Optimal pilot matrix design for training-based channel estimation in MIMO communications

Y El Mouden
University of Wollongong in Dubai

Nidhal Abdulaziz
University of Wollongong in Dubai, nidhal@uow.edu.au

Publication Details

El Mouden, Y. & Abdulaziz, N. 2016, 'Optimal pilot matrix design for training-based channel estimation in MIMO communications', Journal of Telecommunications, vol. 33, no. 1, pp. 1-7.

Optimal Pilot Matrix Design for Training-Based Channel Estimation in MIMO Communications

Y. EL Mouden, N. Abdulaziz

Abstract— This paper considers pilot-based (or training-based) channel matrix estimation in downlink multiuser multiple input multiple output (MIMO) wireless communication systems under Rician fading. The Bayesian approach is used to analyse and derive closed-form mathematical expressions for the minimum mean square estimator (MMSE) and the mean square error (MSE) in estimating the channel state information (CSI) where the long-term channel statistics are known a priori. It is shown how the pilot matrix can be designed to maximize the estimation performance. MATLAB simulation examples are used to evaluate the performance of the MMSE channel estimator for different training sequences and system statistics and to show how both the optimal length of the training sequence (number of columns of the pilot matrix) and the channel estimation error decreases with the spatial correlation.

Index Terms— Channel Estimation, MIMO Systems, Optimal Training Sequence, Pilot Matrix, Rician Fading.

1 Introduction

THE performance of MIMO wireless communication systems is mainly governed by the wireless channel environment [1].

The wireless channel is dynamic and unpredictable, which makes an exact analysis of the wireless communication system often difficult [1]. In fact, the understanding of wireless channels is the foundation for the development of high performance and bandwidth-efficient wireless transmission technology. A major limiting factor in large-scale MIMO is the availability of accurate instantaneous channel state information (CSI) [2]. This is since high spatial resolution can only be exploited if the propagation environment is precisely known. CSI is typically acquired by transmitting predefined pilot signals and estimating the channel coefficients from the received signals [2], [3]. The instantaneous channel matrix is acquired from the received pilot signal by applying an appropriate estimation scheme . The Bayesian minimum mean square error (MMSE) estimator is optimal if the channel statistics are known [2], while the minimum-variance unbiased (MVU) estimator is applied otherwise [4]. These channel estimators basically solve a linear system of equations, or equivalently multiply the received pilot signal with an inverse of the covariance matrices.

In this paper, training-based estimation of instantaneous CSI in multiple- input multiple-output (MIMO) systems is considered. Thus, the estimation is conditioned on the received signal from a known training sequence, which potentially can be adapted to the long-term statistics. By nature, the channel is stochastic, which motivates Bayesian estimation—that is, modelling of the current channel state as a realization from a known multi-variate probability density function (PDF). There is also a large amount of literature on estimation of deterministic MIMO channels which are analytically tractable but in general provide less accurate channel estimates, as shown in [5], [6]. Herein, the concentration is on minimum mean square error (MMSE) estimation of the channel matrix given the first and second order system statistics. It is shown in this work how the pilot matrix can be designed to maximize the estimation performance by minimizing the mean square error (MSE) in estimating the channel matrix. MATLAB simulation examples are used to evaluate the performance of the MMSE channel estimator for different training sequences and system statistics and to show how both the optimal length of the training sequence (number of columns of the pilot matrix) and the channel estimation error decreases with the spatial correlation.

2 General system Functionality

The MIMO techniques and multicell coordination schemes require accurate channel state information (CSI). At the same time, the channels are continuously changing due to small-scale fading [7], [8]. It is therefore necessary to have a mechanism that acquires channel knowledge at regular intervals to keep it up-to-date. Each interval witnesses a constant behavior of the channel (i.e., coherence time). The common way is to use pilot signaling; that is, sending a known signal and trying to estimate channel properties by comparing the transmitted signal with the received signal. Training signaling provides the receiver with channel knowledge[9]. This information can be fed back to the base

- *Y. EL Mouden is with the Engineering Research Centre, University of Wollongong in Dubai, Dubai, PO 20183.*
- *N. Abdulaziz is an Associate Professor with the Faculty of Engineering and Information Sciences, University of Wollongong in Dubai, Dubai, PO 20183.*

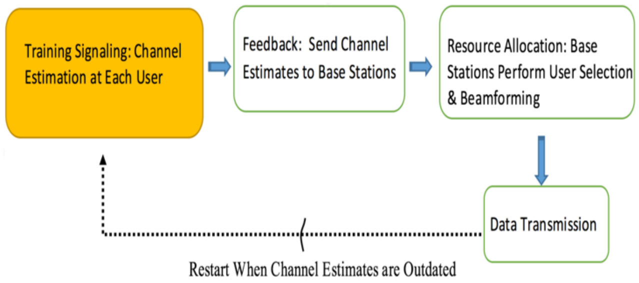

Fig. 1. Illustrative block diagram of the system operation in a wireless communication system. This paper focuses on the orange (shaded) block, i.e., Training-based channel estimation.

station, but it should be done in a concise way to save resources. When both the base station and the users have learned the channel, this information is used for resource allocation, multiuser MIMO transmission, multicell coordination, and processing of the received data signals. After a while (e.g., a few milliseconds), the small-scale fading has changed the channel and made the acquired channel information outdated. It is time for new training signaling and the system operation starts all over again. The cyclic operation is illustrated in Fig. 1 and described in details in the following:

1. *Training*: The base station sends a training sequence that enables each user to estimate its channel. This thesis focuses on developing optimal method for acquiring instantaneous channel state information. This major component of the system is the major topic of this paper.
2. *Feedback*: Each active user feeds back a quantized version of its channel estimate to the base station, using an uplink sub channel.
3. *Resource Allocation*: The base station performs resource allocation (i.e., selects users and their signal correlation matrices). The allocation is either fixed throughout the coherence time or consists of a collection of scheduling/precoding strategies.
4. *Training*: The base station can send a second training sequence, to enable selected users to. adapt to the signal correlation matrices (both their own and the interfering ones). This is often necessary to achieve coherent reception.
5. *Data:* Transmission of data until the end of the coherence time.

The following section will cover the mathematical modelling of the system and its analysis.

3 Problem Formulation and Analysis

The aim at this point is to introduce the mathematical system model used for multiuser MIMO (MU-MIMO) downlink communications and formulate the main system assumptions and problems covered in the paper.

3.1 Mathematical system modelling

A downlink MU-MIMO communication system with n_T antennas base station communicating with r users is considered such as $r \geq n_T$, as illustrated by Fig. 2. Let the kth user be denoted by u_k, where $k \in \{1,2,\dots,r\}$, and has n_R antennas. Also, let $\mathbf{H}_k \in \mathbb{C}^{n_R \times n_T}$ denote the narrowband channel matrix represented in the complex baseband. Each complex entry $[\mathbf{H}_k]_{ij}$ in the channel matrix describes the channel from the jth transmit antenna (base station) to the ith receive antenna (user device); its norm represents the gain/strength introduced by the channel, while its argument represents the phase-shift introduced by the channel.

The multipath propagation is modeled as quasi-static block fading meaning that the channel matrix $\mathbf{H}_k$is constant for a set of consecutive discrete time instants $\mathcal{X}$ and then replaced with a new independent realization. The coherence time, which is the duration over which the channel matrix is constant due to flat fading, is therefore $|\mathcal{X}|$ the number of elements in the set $\mathcal{X}$. The transceiver hardware is assumed to be ideal, without other impairments or distortions than can be included in the channel matrix and background noise, as discussed in [10], [11], and [7].
Based on above assumptions, the discrete time model that relates $\mathbf{y}_k(t) \in \mathbb{C}^{n_R \times 1}$ the symbol-sampled complex baseband received signal at user u_k at time instant $t \in \mathcal{X}$ to the transmitted symbol-sampled complex baseband signal $\mathbf{x}(t) \in \mathbb{C}^{n_T \times 1}$ is given by the equation:

$$\mathbf{y}_k(t) = \mathbf{H}_k \mathbf{x}(t) + \mathbf{n}_k(t) \tag{1}$$

where $\mathbf{n}_k(t) \in \mathbb{C}^{n_R \times 1}$ is the complex vector modeled with a circular symmetric complex distribution representing both additive noise and interference, this can be written as $\mathbf{n}_k(t) \in \mathcal{CN}\left(\bar{\mathbf{n}}_k(t), \mathbf{\Sigma}_k(t)\right)$.
$\mathrm{n}_k(t) \in \mathrm{CN}\left(\bar{\mathrm{n}}_k(t), \Sigma_k(t)\right)\bar{\mathbf{n}}_k(tt) \in \mathbb{C}^{n_R \times 1}$ is the mean value of the disturbance and $\mathbf{\Sigma}_k(t) \in \mathbb{C}^{n_R \times n_R}$ is the covariance matrix.

Let $\mathbf{s}_k(t) \in \mathbb{C}^{n_T \times 1}$ denote the stochastic data signal corresponding to user u_k, this signal has constellation points as its elements after digital modulation where each constellation point represents a symbol. These symbols result from parsing the original data stream (sequence of bits) intended to user u_k into subsequences where each subsequence is named a symbol.
Since $\mathbf{x}(t)$ contains all data signals designated to each of the users, then $\mathbf{x}(t)$ can be expressed as:

$$\mathbf{x}(t) = \sum_{k=1}^{r} \mathbf{s}_k(t) \tag{2}$$

$\mathbf{s}_k(t)$ are zero-mean with $n_T \times n_T$ complex-valued signal correlation matrices:

$$\mathbf{S}_k(t) = \mathbb{E}(\mathbf{s}_k(t)\, \mathbf{s}_k(t)^{H}) \tag{3}$$

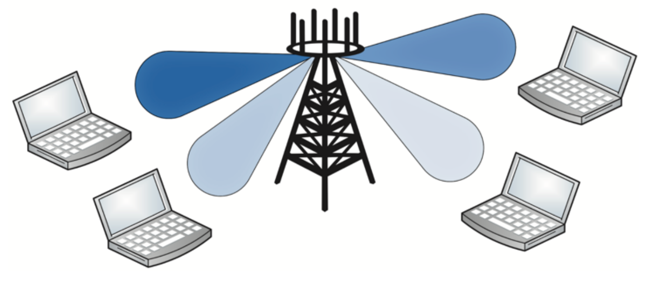

Fig. 2. MU-MIMO downlink communication scenario considered in this paper.

These matrices are important design parameters that are used in the research paper to optimize the system performance. The selection of $\mathbf{S}_1(t), \mathbf{S}_2(t), \ldots, \mathbf{S}_r(t)$ is called resource allocation and implicitly includes selecting which users to transmit to at a given time instant, the design of beamforming directions, and power allocation. Basically, $\mathrm{tr}\{\mathbf{S}_k(t)\}$ describes the power allocated for transmission to user u_k, while the eigenvectors and eigenvalues of $\mathbf{S}_k(t)$ describe the spatial distribution of this power [7]. Space division multiple access (SDMA) is the general case when multiple users are served simultaneously, while the special case when only one user is given non- zero power at each time instant is known as time division multiple access (TDMA). To enable efficient multiuser SDMA transmission, the resource allocation should preferably be based on the current channel state information (CSI)—that is, the collection of current channel matrices $\mathbf{H}_k$.

3.2 Rician fading channel model

The small-scale fading is modeled probabilistically by having a channel matrix $\mathbf{H}_k$ with circular-symmetric complex Gaussian entry elements. This model is proper for scenarios with rich multipath propagation and has been validated by extensive measurements[12], [13] , [14].

The channel matrix $\mathbf{H}_k$ is transformed into a vector vec($\mathbf{H}_k$) by stacking its columns to achieve a proper multivariate channel distribution [6], [7].
Let then:

$$\mathrm{vec}(\,\mathbf{H}_k\,) \in \mathcal{CN}(\mathrm{vec}(\bar{\mathbf{H}}_k), \mathbf{R}_k) \tag{4}$$

where the mean $\bar{\mathbf{H}}_k \in \mathbb{C}^{n_R \times n_T}$ describes the line-of-sight component and the positive semidefinite matrix $\mathbf{R}_k \in \mathbb{C}^{n_T r \times n_T r}$ describes the spatial properties of the multipath propagation. This matrix depends on the large-scale fading and varies at a much slower pace than the small-scale fading. Throughout this thesis, it is assumed that both the base station and user u_k can keep track of $\bar{\mathbf{H}}_k$ and $\mathbf{R}_k$ perfectly, either using a negligible feed-back overhead or reverse-link estimation [15].

It should be noted that the statistical model in (1) is known as Rician fading, because the norm of each element in $\mathbf{H}_k$ is Rician distributed. Also, the stochastic additive noise and interference modeled by $\mathbf{n}_k(t) \in \mathcal{CN}\left(\bar{\mathbf{n}}_k(t), \mathbf{\Sigma}_k(t)\right)$ is called Rician disturbance because the magnitude of each element in $\mathbf{n}_k(t)$ is Rician distributed.

An important property to know is the fact that the disturbance is statistically independent from the channel [9], [16] which will help simplifying derivations for the MMSE channel estimator using statistical signal processing theory [17].

3.1 Downlink transmission vs Uplink transmission

The problem of downlink multiuser MIMO transmission (also known as the broadcast channel) is commonly regarded as more challenging than uplink transmission (also known as the multiple access channel) [32], and there are many strong arguments supporting this statement. First of all, efficient multiantenna transmission requires accurate channel information at both sides in the downlink (to achieve the full multiplexing gain), while channel information is only critically needed at the base station during uplink transmission [1], [33]. Secondly, user devices have to contain power-efficient hardware and are therefore limited to low-complexity signal processing algorithms, while the base station can apply advanced algorithms for signal reception in the uplink. Thirdly, many services primarily create downlink traffic (i.e., video streaming), making the downlink throughput the limiting factor for the user experience. However, there are important connections between the downlink and uplink, which have enabled researchers to gain intuition on the design of downlink transmission by solving mathematically more convenient uplink problems [34], [35]. Many results in this paper could therefore be useful also for the design of uplink transmissions. Moreover, the input-output model in (1) can describe many other types of systems than narrowband MIMO communication. In OFDM systems, (1) can model each of the subchannels [7], [18]. As a result, the analysis in this paper can be applied to many problems other than narrowband communications.

4 Training-Based channel estimation

In this section, both the transmitter and the receiver are assumed to know the long-term statistics of the channel and of the disturbance (i.e., their mean value and covariance matrices). In order to estimate properties of the current channel realization $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$, the transmitter can send a sequence of known training vectors. Sequences of arbitrary length $B \geq 1$ are considered and represented by the training matrix $\mathbf{P} \in \mathbb{C}^{n_T \times B}$. This matrix fulfills a total training power constraint $\mathrm{tr}(\mathbf{P}^H\mathbf{P}) \leq \mathcal{P}$ and its maximal rank is $m \triangleq \min\,(n_T, B)$, which represents the maximal number of spatial channel directions that the training can excite. The columns of $\mathbf{P}$ are used as transmit signal in (1) for B channel uses (i.e., $t = 1, 2, \ldots, B$).
Considering the mathematical model developed in equation (1), let:

$$\mathbf{P} = [\mathbf{x}(1), \ldots, \mathbf{x}(B)]$$

and

$$\mathbf{N} = [\mathbf{n}(1), \ldots, \mathbf{n}(B)] \in \mathbb{C}^{n_R \times B}$$

Therefore, the combined received matrix $\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(B)] \in \mathbb{C}^{n_R \times B}$ of the training transmission will be:

$$\mathbf{Y} = \mathbf{HP} + \mathbf{N} \tag{5}$$

where the disturbance $\mathbf{N}$ is assumed to be uncorrelated with the channel $\mathbf{H}$. The disturbance is modelled as:

$$\mathrm{vec}(\mathbf{N}) \in \mathcal{CN}(\mathrm{vec}(\overline{\mathbf{N}}), \mathbf{S})$$

where $\mathbf{S} \in \mathbb{C}^{Bn_R \times Bn_R}$ is the positive definite covariance matrix and $\overline{\mathbf{N}} \in \mathbb{C}^{n_R \times B}$ is the mean disturbance.

4.1 Channel matrix estimation using MMSE

The estimation of the channel matrix at the receiver is required because instantaneous CSI can be used for receive processing to improve interference suppression and simplify the detection of the original data signal. CSI can be used also for feedback to employ beamforming and rate adaptation [16].

In this section, MMSE estimation of the channel matrix from the observation during training signaling is considered. Basically, the MMSE estimator of a vector $\mathbf{h}$ from an observation $\mathbf{y}$ is expressed as:

$$\hat{\mathbf{h}}_{\mathrm{MMSE}} = \mathbb{E}\{\mathbf{h}|\mathbf{y}\} = \int \mathbf{h} g(\mathbf{h}|\mathbf{y})\, d\mathbf{h} \quad (6)$$

where $g(\mathbf{h}|\mathbf{y})$ is the conditional (posterior) PDF of $\mathbf{h}$ given $\mathbf{y}$ [16], [17].

The MMSE estimator minimizes the $MSE \triangleq \mathbb{E}\{\|\mathbf{h} - \hat{\mathbf{h}}_{\mathrm{MMSE}}\|^2\}$ and the optimal MSE can be calculated as the trace of the covariance matrix $\mathbf{C}_{\mathrm{MMSE}}$ of $f(\mathbf{h}|\mathbf{y})$ averaged over $\mathbf{y}$. The MMSE estimator is the Bayesian counterpart to the minimum variance unbiased (MVU) estimator developed for deterministic channels [17].

By vectorizing the received signal in (8) and applying:

$$\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{N}) \quad (7)$$

the received training signal will be:

$$\mathrm{vec}(\mathbf{Y}) = \widetilde{\mathbf{P}}\mathrm{vec}(\mathbf{H}) + \mathrm{vec}(\mathbf{N}) \quad (8)$$

where $\widetilde{\mathbf{P}} = \mathbf{P}^T \otimes \mathbf{I}$. It follows directly from applying the formulas in chapter 15 of [17] that the MMSE estimator, $\widehat{\mathbf{H}}_{\mathrm{MMSE}}$, of the Rician fading channel matrix can be expressed as:

$$\begin{aligned}\mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MMSE}}) &= \mathrm{vec}(\overline{\mathbf{H}}) + (\mathbf{R}^{-1} + \widetilde{\mathbf{P}}^H\mathbf{S}^{-1}\widetilde{\mathbf{P}})^{-1}\widetilde{\mathbf{P}}^H\mathbf{S}^{-1}\mathbf{v} \\ &= \mathrm{vec}(\overline{\mathbf{H}}) + \mathbf{R}\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{-1}\mathbf{v} \quad (9)\end{aligned}$$

where $\mathbf{v} = \mathrm{vec}(\mathbf{Y}) - \widetilde{\mathbf{P}}\,\mathrm{vec}(\overline{\mathbf{H}}) - \mathrm{vec}(\overline{\mathbf{N}})$.

The error covariance is defined as:

$$\mathbf{C}_{\mathrm{MMSE}} = \mathbb{E}\{(\mathrm{vec}(\mathbf{H}) - \mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MMSE}})(\mathrm{vec}(\mathbf{H}) - \mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MMSE}})^H\}$$

and becomes, after applying formulas of chapter 15 of [17],

$$\begin{aligned}\mathbf{C}_{\mathrm{MMSE}} &= (\mathbf{R}^{-1} + \widetilde{\mathbf{P}}^H\mathbf{S}^{-1}\widetilde{\mathbf{P}})^{-1} \\ &= \mathbf{R} - \mathbf{R}\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{-1}\widetilde{\mathbf{P}}\mathbf{R}\end{aligned}$$

Since $MSE \triangleq \mathbb{E}\{\|\mathrm{vec}(\mathbf{H}) - \mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MMSE}})\|^2\} = \mathrm{tr}\{\mathbf{C}_{\mathrm{MMSE}}\}$

Therefore, $MSE = \mathrm{tr}\{\mathbf{R} - \mathbf{R}\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{-1}\widetilde{\mathbf{P}}\mathbf{R}\}$ (10)

It should be noted that the computation of MMSE estimate in (9) only requires a multiplication of $\mathrm{vec}(\mathbf{Y})$ with a matrix and adding a vector, both of which depend only on the system statistics. Hence, the computational complexity of the estimator is limited [8].

4.2 Pilot Matrix optimization for channel estimation

In this section, it will be shown how to design the pilot matrix $\mathbf{P}$ which minimizes the mean square error (MSE) in estimating the channel matrix using MMSE. The optimization problem is formulated as:

$$\begin{aligned}&\min_{\mathbf{P}} \quad \mathrm{tr}\{\mathbf{R} - \mathbf{R}\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{-1}\widetilde{\mathbf{P}}\mathbf{R}\} \\ &\text{subject to} \quad \mathrm{tr}(\mathbf{P}^H\mathbf{P}) \le \mathcal{P}\end{aligned} \quad (11)$$

where MSE is the objective function and the inequality represents the constraint on the power required for pilot signaling. It is clear that the MSE depends on the pilot matrix $\mathbf{P}$ and on the covariance matrices of the channel and disturbance statistics, while it is unaffected by the mean values. Hence, the pilot matrix can be designed to optimize the estimation performance by adaptation to the second order statistics. For general channel and disturbance statistics, the optimum pilot matrix will not have any special form that can be exploited when solving (11). However, if the covariance matrices $\mathbf{R}$ and $\mathbf{S}$ are structured, the optimal $\mathbf{P}$ may be written or expressed in the same way $\mathbf{R}$ and $\mathbf{S}$ are structured. For mathematical tractability, it is common to use the so-called Kronecker-structre in which $\mathbf{R}$ and $\mathbf{S}$ are expressed as:

$$\mathbf{R} = {\mathbf{R}_T}^T \otimes \mathbf{R}_R, \; \mathbf{S} = {\mathbf{S}_T}^T \otimes \mathbf{S}_R \quad (12)$$

where $\mathbf{R}_T \in \mathbb{C}^{n_T \times n_T}$ and $\mathbf{R}_R \in \mathbb{C}^{n_R \times n_R}$ represent the spatial covariance matrices at the transmitter and receiver side, respectively. $\mathbf{S}_T \in \mathbb{C}^{B \times B}$ and $\mathbf{S}_R \in \mathbb{C}^{n_R \times n_R}$ represent the temporal covariance matrix and the received spatial covariance matrix.

Emil and Bjorn [7]have proved mathematically, through the use of convex optimization and majorization theory, that when $\mathbf{R}_T$ and $\mathbf{S}_T$ are expressed in terms of their eigenvalue decompositions, the solution to a typical optimization problem, as in (11), is the pilot matrix that has a singular value decomposition as:

$$\mathbf{P} = \mathbf{U}_T \mathrm{diag}(\sqrt{p_1}, \dots, \sqrt{p_{n_T}}){\mathbf{V}_T}^H \quad (13)$$

where $\mathbf{U}_T$ and $\mathbf{V}_T$ are matrices that contain the eigenvalues of $\mathbf{R}_T$ and $\mathbf{S}_T$ in opposite orders, while $p_1, \dots, p_{n_T}$ are the ordered training powers that can be found using equation(13) of theorem 1 in [7].

On the other hand, we call $\mathbf{P}$ a heuristic pilot matrix when the covariance matrices $\mathbf{R}$ and $\mathbf{S}$ are not forcibly Kronecker-structured as in (12) and are defined by their general definition using the mathematical expectation. It will be shown through MATLAB simulation that this heuristic pilot matrix produces good estimation performance, even when the covariance matrices are not Kronecker-structured.

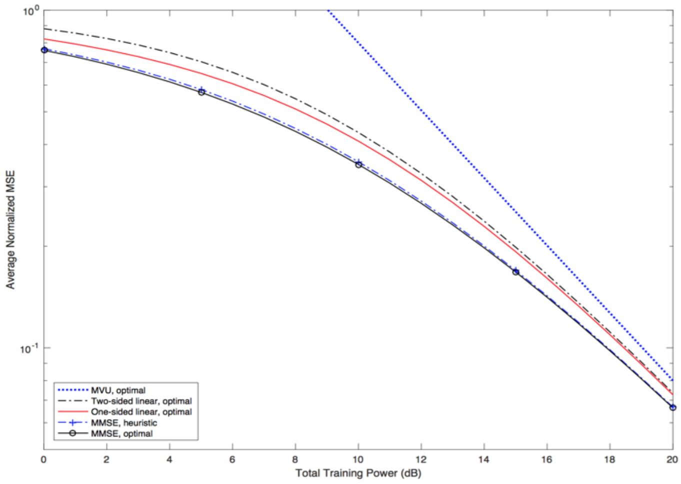

Fig. 3. The average normalized MSEs of estimation of H as a function of the total training power ($n_T = 8$, $n_R = 4$). The performance of four different estimators with MSE-minimizing training matrices is compared. The performance with the heuristic training is also given.

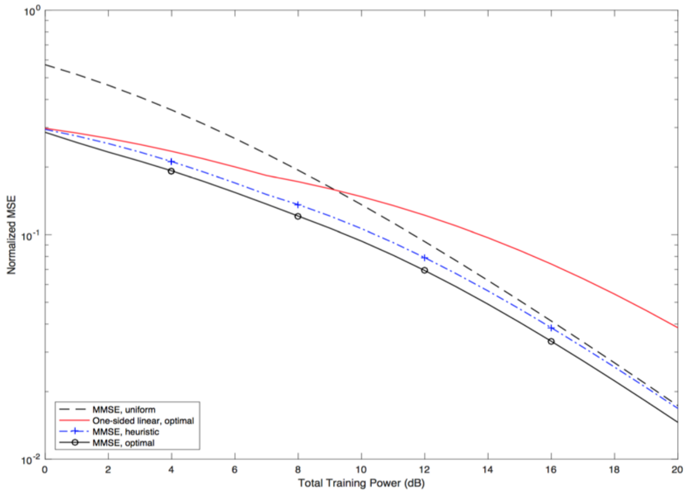

Fig. 4. The normalized MSEs of estimation of H as a function of the total training power. The performance of four different estimators with MSE-minimizing training matrices is compared. The MMSE estimator with three different training matrices is compared with the one sided linear estimator.

5 SIMULATION RESULTS

In this part, the mathematical results of previous sections and previous work from literature will be illustrated and evaluated using MATLAB simulation. The MMSE estimator of the channel matrix will be compared with other recently proposed estimators and the potential gain of training sequence optimization is exemplified. Moreover, it will be shown how the optimal length of training sequence depends on the spatial correlation and available training power.

The MSE performance of the channel matrix estimator was thoroughly evaluated in [19] for interference-limited Kronecker-structured systems. Thus, the opposite setting of a noise-limited non-Kronecker-structured system will be considered.

To illustrate the performance of the training sequence design for channel matrix estimation under general channel conditions, the Weichselberger model [20]is considered. This model has recently attracted much attention for its accurate representation of measurement data [16]. According to this model, the channel matrix can be expressed as $\mathbf{H} = \mathbf{U}_A \widetilde{\mathbf{H}} \mathbf{U}_B^H$ where $\mathbf{U}_A$ and $\mathbf{U}_B$ are unitary matrices and $\widetilde{\mathbf{H}} \in \mathbb{C}^{n_R \times n_T}$ has independent elements with variances given by the corresponding elements of the coupling matrix $\mathbf{\Omega}$. The unitary matrices will not affect the performance when MSE minimizing precoding design is employed, and can therefore be selected as identity matrices [21]. Without loss of generality, the coupling matrices are always scaled as $\mathrm{tr}(\mathbf{\Omega}) = n_R n_T$ to make sure that the SINR can be described by the training power constraint: $SINR_{\text{training}} = \frac{\mathcal{P}\mathrm{tr}(\mathbf{R})}{\mathrm{tr}(\mathbf{S})} = \mathcal{P}$.

5.1 Comparison of different channel estimators

To enable comparison with other estimators, the channel is zero-mean, although it is known based on (10) that the performance is unaffected by non-zero mean components. The normalized MSE is defined as $\mathbb{E}\left\{\left\|\mathrm{vec}(\mathbf{H}) - \mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MMSE}})\right\|^2\right\} / \mathrm{tr}(\mathbf{R})$.

In Fig. 3, the normalized MSEs averaged over 2000 scenarios are given with different coupling matrices with $n_T = 8$, $n_R = 4$, and independent chi-distributed elements. The performance of four different estimators with MSE minimizing training matrices are compared. The MVU/ML channel estimator $\widehat{\mathbf{H}} = \mathbf{Y}\mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1}$ [22], the one sided linear estimator in [7], [22], the two-sided linear Bayesian linear estimator proposed in [23], and the MMSE estimator in (10).

The MVU/ML estimator is unaware of the channel statistics (i.e., non-Bayesian), and it is clear from Fig. 3 that this leads to poor estimation performance. The two-sided linear estimator also performs poorly under the given premises, but can provide good performance in special cases [22]. The performance gap between the one-sided linear estimator and the MMSE estimator (which is also linear) is noticeable, while the difference between employing the optimal training matrix and the one proposed in the heuristic is small. It should be pointed out that the use of independent chi-distributed elements in the coupling matrix induces a spatially correlated environment with a few dominating paths. In less correlated scenarios, the difference between the estimators decreases, but the order of quality is usually the same.

5.2 Comparison of pilot matrices

In Fig. 4, the performance of the MMSE estimator is shown for a uniform training matrix ($\mathbf{P} = \sqrt{\mathcal{P}/n_T}\,\mathbf{I}$), MSE minimizing training matrix (achieved numerically), and the simple explicit heuristic training matrix. The one-sided linear estimator is given as a reference. In this simulation, the coupling matrix that was proposed in [29, Eq. 28] is used to describe an environment with two small scatterers, two big scatterers, and one large cluster. It is clear that the gain of employing an MSE minimizing pilot matrix is noticeable, and the heuristic approach mentioned earlier captures most of this gain although uniform training is asymptotically optimal at high training power.

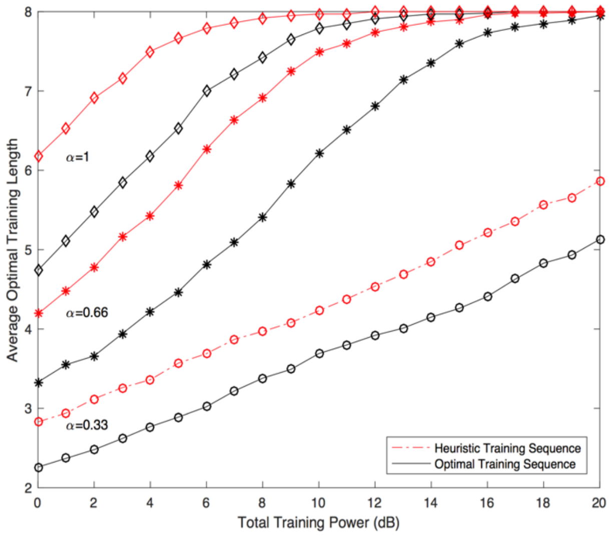

Fig. 5. Smallest training sequence length, necessary to minimize the MSE, in estimating the channel matrix in a system with $n_T = 8$ and uncorrelated receive antennas. The results for different correlation, α, between transmit antennas.

5.3 Pilot matrix rank

In this section, it is illustrated how the optimal length of the training sequence varies with the spatial correlation and training power. Based on Theorem 3 in [7], it can be shown that the optimal length is the smallest B that can achieve the minimal MSE which is equal to the rank of $\mathbf{P}$ (in noise-limited systems).

Jointly-correlated statistics is considered with the system dimensions $n_T = 8$, $n_R = 4$, and with coupling matrices with independent chi-distributed elements. To induce random transmit-side correlation, the jth column of the coupling matrix is scaled by $\alpha^{j-1}$for different values on α: 0.33, 0.66, and 1. The average optimal training sequence length (i.e., average rank of $\mathbf{P}$) is shown in Fig. 5 for both an MSE- minimizing training matrix and the heuristic pilot matrix. In the case of identically distributed elements of the coupling matrix ($\alpha = 1$), there is sufficient spatial correlation to have rank($\mathbf{P}$) $< n_T$ at low training power. As the spatial correlation increases (i.e., α decreases), the optimal training length decreases and the convergence towards full rank becomes slower. The simulation shows that the heuristic method on average produces a slightly longer training sequence than necessary. The simulation depicts also how the optimal B can be smaller than n_T, thus the results of [24] (i.e., $B = n_T$ in uncorrelated systems) are not applicable for generalised cases.

It can be concluded that rigorous system analysis is always required to determine the optimal length under general statistics. The loss in in performance by employing an even shorter training sequence may be minor compared with the gain of having more data symbols.

6 Conclusion and Future Work

Training signalling can be used to estimate accurate channel information at the receiver. Closed-form expression for MMSE estimation of the channel matrix and its MSE have been mathematically derived in this paper under Rician channel and disturbance statistics. It was also shown how the pilot matrix can be designed to optimise the estimation performance.

This optimisation problem is convex under certain statistical conditions and can sometimes even have closed-form solutions. A heuristic training method was discussed based on these insights and it shows close-to-optimal performance and large potential improvements over uniform training. Finally, it was shown that both the pilot matrix rank (or the optimal length of the training sequence) and the estimation error decrease with the spatial correlation.

Future work will focus more on the complexity issues that arise from the use of conventional channel estimators in massive MIMO where the number of antennas at each side of the link can exceed 100. This future research work will attempt to design a new estimator that satisfy the low complexity and high performance requirements that could be used for the future 5G wireless systems.

Acknowledgment

Y. EL Mouden wishes to thank with much appreciation Dr Emil Bjornson for his significant work in the area of multi-user MIMO communications which inspired the coordination of this research work.

References

[1] "MIMO and 802.11n | Air-Stream." [Online]. Available: http://www.air-stream.org.au/technical-references/mimo-and-80211n. [Accessed: 11-Dec-2015].

[2] I. Poole, "MIMO Technology | Multiple Input Multiple Output | Radio-Electronics.Com." [Online]. Available: http://www.radio-electronics.com/info/antennas/mimo/multiple-input-multiple-output-technology-tutorial.php. [Accessed: 11-Dec-2015].

[3] A. Chochalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge University Press, 2014, 2014.

[4] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*. John Wiley & Sons, 2010.

[5] S. K. Mohammed, A. Zaki, A. Chockalingam, and B. S. Rajan, "High-Rate Space-Time Coded Large-MIMO Systems: Low-Complexity Detection and Channel Estimation," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 6, pp. 958–974, 2009.

[6] N. Shariati, E. Bjornson, M. Bengtsson, and M. Debbah, "Low-Complexity Polynomial Channel Estimation in Large-Scale MIMO With Arbitrary Statistics," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 815–830, 2014.

[7] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, 2010.

[8] N. Shariati, E. Bjornson, M. Bengtsson, and M. Debbah, "Low-complexity channel estimation in large-scale MIMO using polynomial expansion," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, 2013, pp. 1157–1162.

[9] J. Flam, E. Bjornson, and S. Chatterjee, "Pilot design for MIMO channel estimation: An alternative to the Kronecker structure assumption," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 5061–5064.

[10] B. Goransson, S. Grant, E. Larsson, and Z. Feng, "Effect of transmitter and receiver impairments on the performance of MIMO in HSDPA," in *2008 IEEE 9th Workshop on Signal Processing Advances in Wireless Communications*, 2008, pp. 496–500.

[11] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *2010 International ITG Workshop on Smart Antennas (WSA)*, 2010, pp. 189–196.

[12] J. W. Wallace and M. A. Jensen, "Measured characteristics of the

MIMO wireless channel," in *IEEE 54th Vehicular Technology Conference. VTC Fall 2001. Proceedings (Cat. No.01CH37211)*, 2001, vol. 4, pp. 2038–2042.

[13] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[14] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-input - multiple-output measurements and modeling in Manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 321–331, Apr. 2003.

[15] T. Aste, P. Forster, L. Fety, and S. Mayrargue, "Downlink beamforming avoiding DOA estimation for cellular mobile communications," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, 1998, vol. 6, pp. 3313–3316.

[16] E. Bjornson and B. Ottersten, "Training-based Bayesian MIMO channel and channel norm estimation," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 2701–2704.

[17] S. M. Kay, "Fundamentals of Statistical Signal Processing: Estimation Theory," *Englewood Cliffs NJ Prentice Hall*, vol. I, no. 4, p. 513, Mar. 1993.

[18] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for Next-Generation Wireless Systems," *Proc. IEEE*, vol. 95, no. 7, pp. 1430–1469, Jul. 2007.

[19] Y. Liu, T. F. Wong, and W. W. Hager, "Training Signal Design for Estimation of Correlated MIMO Channels With Colored Interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486–1497, Apr. 2007.

[20] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wirel. Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.

[21] E. Bjornson, E. Jorswieck, and B. Ottersten, "Impact of Spatial Correlation and Precoding Design in OSTBC MIMO Systems," *IEEE Trans. Wirel. Commun.*, vol. 9, no. 11, pp. 3578–3589, Nov. 2010.

[22] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, 2006.

[23] D. Katselis, E. Kofidis, and S. Theodoridis, "On Training Optimization for Estimation of Correlated MIMO Channels in the Presence of Multiuser Interference," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4892–4904, Oct. 2008.

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

Y. EL Mouden is a student in the final year of his Bachelor's Degree in Electrical Engineering and a research assistant at the University of Wollongong in Dubai. His research and academic interests include digital and statistical signal processing, image processing, wireless communication systems, Electronics, and control systems. He won the best poster award for his bachelor's thesis in innovation fair. His professional memberships include the following but not limited to; Institute of Electrical and Electronics Engineers (IEEE), and Engineers Australia (EA). computing environments.

N. Abdulaziz is currently working as Associate Professor at the Faculty of Engineering and Information Sciences, University of Wollongong in Dubai. Before that she was the Director for Electronics and Electrical Engineering, Heriot Watt University (HWU) Dubai Campus. Dr Abdulaziz's research interests include image and video coding, line coding, wavelet transform, digital watermarking and engineering education and she has published over 37 journal and conference papers in these subject areas. She has received her PhD from Monash University, Australia, in Electrical and Computer Systems Engineering. She won the Lampard prize for the best PhD thesis submitted to the Department of Electrical and Computer Engineering at her University.