TRANSACTIONS ON EMERGING TELECOMMUNICATIONS TECHNOLOGIES
Trans. Emerging Tel. Tech. 2016; **27**:1441–1453
Published online 19 July 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/ett.3069

RESEARCH ARTICLE

Optimal training for non-feedback adaptive modulation over Rayleigh fading channels

Khalid Zeineddine[1], Hussein Hammoud[2] and Ibrahim Abou-Faycal[2]*

[1] Department of Electrical Engineering & Computer Science, Northwestern University, USA
[2] Department of Electrical & Computer Engineering, American University of Beirut, Lebanon

ABSTRACT

Time-varying fading channels present a major challenge in the design of wireless communication systems. Adaptive schemes are often employed to adapt the transmission parameters to receiver-based estimates of the quality of the channel. We consider a pilot-based adaptive modulation scheme without the use of a feedback link. In this scheme, pilot tones (known by sender and receiver) are periodically sent through the channel for the purpose of channel estimation and coherent demodulation of data symbols at the receiver. We optimise the duration and power allocation of these pilot symbols to maximise the information-theoretic achievable rates using binary signaling. We analyze four transmission policies and numerically show how optimal training in terms of duration and power allocation varies with the channel conditions and from one transmission policy to another. We prove that for a causal estimation scheme with flexible power allocation, placing all the available power on one pilot is optimal. Copyright © 2016 John Wiley & Sons, Ltd.

***Correspondence**
Ibrahim Abou-Faycal, FEA, AUB, P.O. Box: 11-0236, Beirut 1107 2020, Lebanon.
E-mail: Ibrahim.abou-faycal@aub.edu.lb

Received 12 October 2015; Revised 29 April 2016; Accepted 12 June 2016

1. INTRODUCTION

In digital mobile communications, fast fading degrades the Bit Error Rate (BER) of the channel and inhibits coherent detection.[†] It is known that performance is limited by channel estimation errors [1–4]. Pilot Symbol Assisted Modulation (PSAM) is a technique that has been introduced in [5] to mitigate these effects. In this scheme, known training symbols (pilots) are periodically inserted into the data frame for the purpose of channel estimation and coherent demodulation of the data symbols [6].

Furthermore, channel-adaptive modulation dynamically adjusts certain transmission parameters such as the constellation size, transmitted power and code rate according to the channel quality. Adaptive signaling provides in general higher bit rates (relative to conventional nonadaptive methods) by increasing the transmission throughput under favorable channel conditions and reducing it as the channel condition is degraded.

Some of the previous adaptive schemes rely on a channel-feedback link to provide the transmitter with the Channel Side Information (CSI) [7, 8]. In [9], the authors consider employment of adaptive modulation with one pilot in addition to delayed feedback to the transmitter and prove that power adaptation via periodic feedback can increase the achievable rates. Similarly, in [10], authors consider pilot-based adaptive modulation where estimate is fed back to transmitter in order to adapt data and pilot power and study the optimal policy for power allocation for data and pilot symbols. Authors in [11] discuss adaptive modulation with feedback and develop an adaptive scheme that accounts for both channel estimation and prediction errors in order to meet a target BER. In [12], the authors attempt to optimise the spectral efficiency subject to a specific BER constraint in a pilot-based adaptive modulation setup with feedback. The earlier mentioned works study the performance of such systems and prove adaptive modulation using pilots can increase the achievable rates in general. However, systems that rely on a channel-feedback link present some disadvantages because of the modeling complexity on one hand and its infeasibility on the other hand when the channel is fading faster than it can be estimated (or predicted) and fed back to the sender. Optimising the pilot placement, power allocation and modulation schemes in a pilot-based setup is an active area of

[†]Coherent demodulation requires the extraction of a reliable phase reference from the received signal.

Copyright © 2016 John Wiley & Sons, Ltd.

research, whether in the case of a single receiver [9–11, 13–18] or multiple receivers [18–20].

A modified pilot-based adaptive modulation scheme over Rayleigh fading channels was presented in [21]. This scheme adapts the coded modulation strategy at the sender to the quality of the channel estimation (estimation error variance) at the receiver *without requiring any channel feedback*. In this work, we study the performance of this non-feedback adaptive modulation scheme over time-varying Rayleigh fading channels. Unlike the scheme in [21], we consecutively send a cluster of k pilots ($k \geqslant 1$) per data frame with k being an optimisation variable [22]. We determine the optimal duration and power allocation of the training period under different transmission policies for both causal and non-causal estimation. We study such systems at low Signal-to-Noise-Ratio (SNR) (we consider the received SNR) levels and the performance is measured in terms of achievable rates using binary signaling. We prove that the 'optimal' power allocation scheme which minimises the error variance of the estimates of the channel parameters –*which is set up offline without requiring feedback*– in case of causal estimation is the one in which all the available power is allocated on one pilot, if constraints allow it.

The organisation of this paper is as follows. In Section 2, we present the fading channel model, the adaptive transmission technique we use to transmit over the channel as well as the receiver details. The measure of performance is discussed in Section 3, the optimal power allocation for causal estimation is proved in Section 4 and the numerical results are presented in Section 5. In Section 6, we present possible extensions including other fading models and Section 7 concludes the paper.

2. PRELIMINARIES

2.1. The channel model

Consider the single-user discrete-time model for the Rayleigh fading channel,

$$Y_i = R_i X_i + N_i,$$

where i is the time index, $X_i \in \mathbb{C}$ is the channel input at time i, $Y_i \in \mathbb{C}$ is its output, and R_i and N_i are independent complex circular Gaussian[‡] random variables with zero mean and variance σ_R^2, and σ_N^2 respectively. The amplitude of the fading coefficient R_i is then Rayleigh distributed and its phase is uniform over $[-\pi, \pi)$. To account for power constraints, the input is subject to

$$\mathsf{E}\left[|X_i|^2\right] \leqslant P,$$

for some parameter P. Because from an information theoretic perspective scaling the output by $1/\sigma_R$ does not change the mutual information, we assume without loss of generality that $\sigma_R = 1$. The variance of the noise σ_N^2 is to be generally interpreted as $\left(\sigma_N^2/\sigma_R^2\right)$.

We assume in this study that the fading process follows a stationary first-order Gauss-Markov model introduced in [23], that is,

$$R_i = \alpha R_{i-1} + Z_i, \tag{1}$$

where the samples $\{Z_i\}$ are Independent and Identically-Distributed (IID) complex circular Gaussians with mean zero and variance equal to $\sigma_Z^2 = \left(1-\alpha^2\right)$ such that $\alpha \in [0, 1)$ to guarantee stationarity.

We assume in this work that the channel *statistics* are known to both transmitter and receiver. Whenever these statistics are not available to the transmitter, they may be obtained through feedback from the receiver, albeit at a much slower time scale.

The relation between this fading model (1) and the coherence time is established in [23]: For a coherence time of T_c and transmission over a bandwidth of W, α is determined by $\alpha^{T_c W} = \phi$ where ϕ is the level of decorrelation deemed necessary in the definition of coherence time.

While the Gauss-Markov model offers analytical flexibility, it was found in [24] to match well both, very slow fading as well fast fading channels. Additionally, even though we analyze the benefits of pilot clustering by assuming that the autocorrelation function of the fading process is derived from a stationary first-order Gauss-Markov model (1), we argue in Section 6 that the methodology may be readily adapted to other models – ones that may be more appropriate in some regimes, as long as the second-order statistics of the fading process are known. This is showcased by presenting the case of a Jakes' model [25] that takes into account higher orders of correlation.

2.2. The adaptive transmission scheme

At regular intervals, the transmitter successively sends k known pilot symbols whose purpose is to enable the estimation of the channel at the receiver. The channel estimation is solely based on the pilot symbols and no data-directed estimation is used. For each time sample i, the receiver computes the Minimum Mean-Square Estimate (MMSE) of the channel, the quality of which –measured through the estimate error variance– depends on its position with respect to the pilot symbols. After estimation, the channel, as seen by the receiver, is a Rician channel whose specular part is given by the estimate and whose Rayleigh component is given by the zero-mean Gaussian-distributed estimation error. A coherent device may then be used by the receiver and additional information may be relayed using the phase.

Being aware of the fact that the receiver is computing the MMSE, the transmitter adapts its modulation and coding scheme to the 'quality' of that channel estimate

[‡]A complex Gaussian random variable is circular if and only if it is zero-mean and its real and imaginary parts are independent with equal variances.

Trans. Emerging Tel. Tech. **27**:1441–1453 (2016) © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

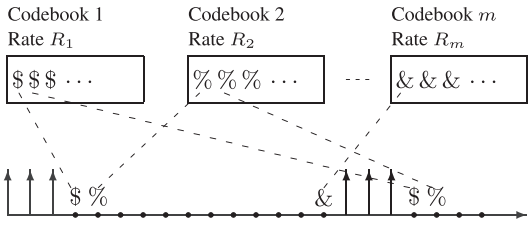

Figure 1. Multiple codebook interleaving.

as measured by the mean-square error. This mean-square error is independent of the value of the estimate –available only at the receiver– and can be computed at the transmitter which is possible because it is aware of the channel statistics (i.e. all model parameters are known). Indeed, the error variance may be even computed offline and the adaptive transmission scheme can then be determined offline as well. Even though there is no feedback to the transmitter, it is aware of the statistics of the estimation error beforehand. We emphasise that although the scheme is adaptive, the transmitter *does not use feedback* to determine its policy.

The transmitter employs multiple codebooks in an interleaved fashion as shown in Figure 1. Because the receiver will estimate the channel based on the sent pilots, the estimation error will depend on the distance from the pilots. A received codeword –when rearranged– would have been effectively transmitted in its entirety over a channel with identical statistics for each of its symbols.

The receiver adapts its throughput to the estimation error variance by coding the data symbols according to their distance from the training pilots. Symbols that are far away from the pilots encounter poorer channel estimates at the receiver and are therefore coded with lower rate codes, while closer symbols benefit from small estimation error variance and are coded with higher rate codes.

We only consider binary signaling. The motivation for this choice is multiple folds. First, in [26] the authors prove that for discrete-time memoryless Rayleigh fading channels subject to average power constraints, the capacity achieving distribution is discrete with a finite number of mass points. Moreover, a binary distribution was found to be optimal at low and moderate values of SNR [26–28]. Second, for a memoryless Rician fading channel, Luo [29] established a similar result that, combined with Gallager's in [27], implies that the binary input distribution is asymptotically optimal at low SNR [29]. Consequently, we choose the alphabet of every codebook to consist in general of two symbols:

$$\begin{cases} m_1 = a_1 + jb_1 & \text{with probability } p_1 \\ m_2 = a_2 + jb_2 & \text{with probability } p_2 = (1-p_1). \end{cases}$$

The rate of the codebooks is adjusted by modifying the probability distribution of the mass points. Numerical results in [13, 21] indicate that the optimal mass points always lie between the extremes of on-off keying (optimal for the IID Rayleigh fading case where no CSI is available at the receiver) and the antipodal signaling (optimal for a perfectly known channel). It is worth noting that some of the work in the literature consider these two extremes for designing the constellation mapping and try to optimise the transmission model in the case of imperfect CSI based on the SNR level [13]. Moreover, any rotational transformation of the two mass points will not affect the mutual information [26, 29]. Therefore, an optimal input distribution consists of two mass points $m_1, m_2 \in \mathbb{R}^*$ with $-\sqrt{P} \leqslant m_1 < 0$ and $m_2 \geqslant \sqrt{P}$.

2.3. Channel estimation at the receiver

Given a pilot spacing interval T, we send k pilots in the beginning of every data frame as shown in Figure 2. When transmitting a pilot at time index i, the input of the channel is $\sqrt{P_i}$ and its output is,

$$Y_i = \sqrt{P_i} R_i + N_i, \qquad i = 0, \cdots, k-1.$$

On the receiver side, we perform MMSE estimation based on the received signal during training. More precisely, we denote by $\mathcal{S}$ the set of indices corresponding to the received pilots $\{Y_s\}_{s\in\mathcal{S}}$ involved in estimating R_j for $j = k, \ldots, T-1$. Therefore, when $\mathcal{S} = \{0, \ldots, k-1\}$ we say we are performing causal MMSE estimation, and when $\mathcal{S} = \{0, \ldots, k-1, T, \ldots, T+k-1\}$ the MMSE estimate is said to be non-causal.

Next, we compute the MMSE estimate $\hat{R}_j\left(\{Y_s\}_{s\in\mathcal{S}}\right)$ of R_j for $j = k, \ldots, T-1$. Because the random variables $\left\{R_j, \{Y_s\}_{s\in\mathcal{S}}\right\}$ are jointly Gaussian, the MMSE estimator is linear and is identical to the Linear Least-Square Estimator (LLSE) the error variance v_j of which is,

$$v_j = 1 - \Lambda_{R_j, \{Y_s\}_{s\in\mathcal{S}}} \, \Lambda^{-1}_{\{Y_s\}_{s\in\mathcal{S}}} \, \Lambda^{T}_{R_j, \{Y_s\}_{s\in\mathcal{S}}}, \qquad (2)$$

where $\Lambda_{R_j, \{Y_s\}_{s\in\mathcal{S}}}$ is the cross-covariance matrix between R_j and $\{Y_s\}_{s\in\mathcal{S}}$ and $\Lambda_{\{Y_s\}_{s\in\mathcal{S}}}$ is the autocovariance of the vector of received pilots $\{Y_s\}_{s\in\mathcal{S}}$.

We note that the estimation error variance in equation (2) may be computed offline at design time –and therefore no feedback is needed to the encoder– and is only dependent on the autocorrelation function of $\{R_j\}$, the transmitted pilots and the noise spectral density.

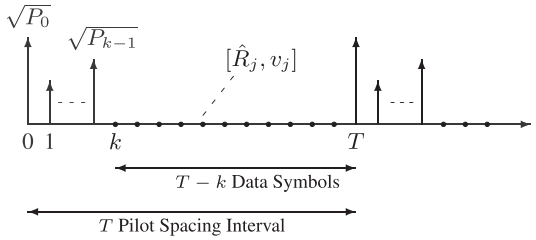

Figure 2. Pilot symbols and channel estimation.

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. ACHIEVABLE RATES

We consider the transmission scheme shown in Figure 2 with symbols sent with power P_j for $j = 0, \ldots, T-1$. Given a sample $\{y_s\}_{s\in\mathcal{S}}$, the received symbols can be written

$$\begin{aligned} Y_i &= R_i X_i + N_i \\ &= \left(\hat{R}_i + \Gamma_i\right) X_i + N_i, \quad \text{for } i = k, \ldots, T-1, \end{aligned}$$

where Γ_i is a zero-mean complex Gaussian error term that has a variance v_i. Therefore,

$$\begin{aligned} p\left(y_i | x_i, \{y_s\}_{s\in\mathcal{S}}\right) &= \mathcal{N}_{\mathbb{C}}\left(\hat{R}_i x_i,\, v_i\,|x_i|^2 + \sigma_N^2\right) \\ &= \frac{1}{\pi\left(v_i\,|x_i|^2 + \sigma_N^2\right)} e^{-\frac{|y_i - \hat{R}_i x_i|^2}{v_i|x_i|^2 + \sigma_N^2}}. \end{aligned}$$

When ignoring the fading correlation from one transmitted frame to another, the mutual information per symbol due to interleaving can be written as

$$\begin{aligned} &\frac{1}{T} I\left(\ \{X_i\}_{i=0}^{T-1}\ ;\ \{Y_i\}_{i=0}^{T-1}\ |\ \{Y_s\}_{s\in\mathcal{S}}\ \right) \\ &= \mathsf{E}_{\{Y_s\}_{s\in\mathcal{S}}}\left[\frac{1}{T}\sum_{i=k}^{T-1} I\left(X_i; Y_i | \{y_s\}_{s\in\mathcal{S}}\right)\right] \\ &= \frac{1}{T}\sum_{i=k}^{T-1} \mathsf{E}_{\hat{R}_i}\left[I\left(X_i; Y_i|\hat{R}_i\right)\right], \end{aligned} \tag{3}$$

where the expectation is now over the random variable $\hat{R}_i$. Note that $\hat{R}_i$ is a linear combination of the observations

$$\hat{R}_i = \sum_{m\in\mathcal{S}} \beta_m Y_m = R_i - \Gamma_i \sim \mathcal{N}_{\mathbb{C}}(0, 1 - v_i). \tag{4}$$

3.1. The computation method

The i^{th} term, $\mathsf{E}_{\hat{R}_i}\left[I\left(X_i; Y_i|\hat{R}_i\right)\right]$, in equation (3) depends on the choice of the corresponding binary probability distribution fully characterised by the three parameters $\{m_1, m_2, p_1\}_i$. This distribution (for $i = k, \ldots, T-1$) determines the rate of the corresponding codebook and should be chosen to maximise the mutual information quantity in (3). Therefore, we are interested in solving

$$\frac{1}{T}\sum_{i=k}^{T-1} \max_{\{m_1, m_2, p\}_i} \mathsf{E}_{\hat{R}_i}\left[I\left(X_i; Y_i|\hat{R}_i\right)\right], \tag{5}$$

subject to $\mathsf{E}\left[|X_i|^2\right] \leqslant P_i$ for all $i = k, \ldots, T-1$.

Furthermore, examining the probability distribution (4) of $\hat{R}_i$ indicates that the elementary quantity $I_{sub}(P_i,\, v_i) \triangleq \max_{\{m_1, m_2, p\}_i} \mathsf{E}_{\hat{R}_i}\left[I\left(X_i; Y_i|\hat{R}_i\right)\right]$ in (5) is only a function of the estimation error variance v_i and power P_i of the symbol. Thereafter the achievable rates become

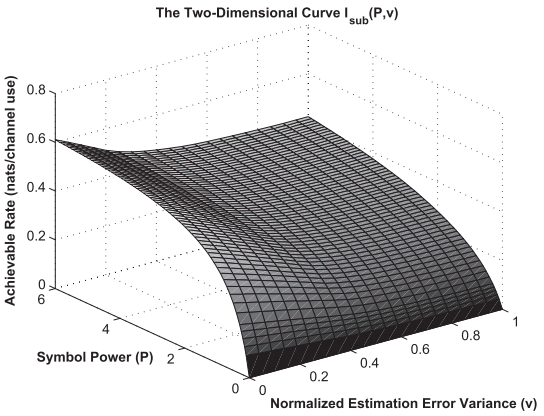

Figure 3. The two-dimensional curve $I_{sub}(P,\, v)$.

$$\frac{1}{T} I\left(\{X_i\}\,;\, \{Y_i\}\,|\,\{Y_s\}\right) = \frac{1}{T}\sum_{i=k}^{T-1} I_{sub}(P_i,\, v_i). \tag{6}$$

The two dimensional curve $I_{sub}(P,\, v)$ is numerically computed by solving the maximisation problem over a fine grid $\mathscr{V} = \{0 \leqslant P \leqslant P_{max}, 0 \leqslant v \leqslant 1\}$ as shown in Figure 3. Then given a transmission strategy consisting of an inter-pilot spacing T, k-pilot clustering, and a power allocation P_j for $j = 0, \ldots, T-1$, we calculate using equation (2) the estimation error variance v_j for $j = k, \ldots, T-1$. The corresponding elementary mutual information quantity $I_{sub}(P_j,\, v_j)$ can now be interpolated from the data set $\{\mathscr{V},\, I_{sub}(P,\, v)\}$ and used to compute the normalised sum in (6).

Finally, note that the error variance is a function of the power of the pilots $\{P_s\}_{s\in\mathcal{S}}$. Hence, equation (6) can also be written as

$$\frac{1}{T} I\left(\{X_i\}\,;\, \{Y_i\}\,|\,\{Y_s\}\right) = \frac{1}{T}\sum_{i=k}^{T-1} I_{sub}\left(P_i, \{P_s\}_{s\in\mathcal{S}}\right). \tag{7}$$

3.2. The transmission policy

We consider four types of transmission policies and we study how the optimal training strategy differs from one policy to another, analytically in Section 4 and numerically in Section 5.

3.2.1. Policy I.

The pilot symbols and the data symbols are transmitted with the same amount of power, that is,

$$P_s = P_i = P, \quad \forall s = 0, \ldots, k-1 \,\&\, i = k, \ldots, T-1.$$

Therefore, for a given channel model, k-pilot training, and an inter-pilot spacing T, the achievable rate in equation (7) is a function of P only.

Trans. Emerging Tel. Tech. **27**:1441–1453 (2016) © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2.2. Policy II.

In this policy, a flat power allocation is adopted for both the pilot symbols and the data symbols, but we allow the two levels to be different. More precisely,

$$P_s = P_{tr} \quad \forall s = 0, \ldots, k-1$$
$$P_i = P_d, \quad \forall\, i = k, \ldots, T-1.$$

The achievable rate is a function of P_{tr} & P_d which satisfy

$$\frac{1}{T}\sum_{j=0}^{T-1} P_j = \frac{1}{T}\left[kP_{tr} + (T-k)P_d\right] \leqslant P.$$

3.2.3. Policy III.

Following a flat power allocation for pilots ($P_s = P_{tr}, \forall s = 0, \ldots, k-1$), the data symbols are sent with power P_i for $i = k, \ldots, T-1$. These power levels satisfy

$$\frac{1}{T}\sum_{j=0}^{T-1} P_j = \frac{1}{T}\left[kP_{tr} + \sum_{j=k}^{T-1} P_j\right] \leqslant P.$$

3.2.4. Policy IV.

We send both the pilots and data symbols with variable power P_j for $j = 0, \ldots, T-1$. The constraint on the power levels is now given by $\frac{1}{T}\sum_{j=0}^{T-1} P_j \leqslant P$.

4. OPTIMAL POWER ALLOCATION FOR CAUSAL ESTIMATION

In this section, we find the optimal power allocation and training duration for policies II, III and IV under causal estimation. These optimal solutions are found by applying the result of Theorem 1 stated hereafter. The theorem implies that if we let kP_{tr} be the *total* power 'budget' for the training period, everything else being equal, among all the training power allocation schemes $\{P_s\}_{s=0}^{k-1}$ such that

$$\sum_{s=0}^{k-1} P_s = kP_{tr}, \tag{8}$$

the optimal one is the one where all the power is allocated to the last time slot $(k-1)$.

For causal MMSE estimation, $\mathcal{S} = \{0, \cdots, k-1\}$ and for $j = k, \cdots, (T-1)$ equation (2) can be written as

$$\begin{aligned} v_j &= 1 - \Lambda_{R_j,\{Y_s\}_{s\in\mathcal{S}}}\, \Lambda^{-1}_{\{Y_s\}_{s\in\mathcal{S}}}\, \Lambda^{T}_{R_j,\{Y_s\}_{s\in\mathcal{S}}} \\ &= 1 - \alpha^{2(j-k+1)} \begin{pmatrix} \alpha^{k-1} & \alpha^{k-2} & \cdots & 1 \end{pmatrix} \\ &\quad \times D^T \left[DAD^T + \sigma_N^2 I\right]^{-1} D \begin{pmatrix} \alpha^{k-1} \\ \vdots \\ 1 \end{pmatrix}, \end{aligned} \tag{9}$$

where A is the $k \times k$ symmetric, positive definite autocovariance matrix of the channel fading coefficients $\{R_s\}_{s\in\mathcal{S}}$, and D is the $k \times k$ 'input' matrix:

$$A = \begin{pmatrix} 1 & \alpha & \cdots & \alpha^{k-1} \\ \alpha & 1 & \cdots & \alpha^{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{k-1} & \alpha^{k-2} & \cdots & 1 \end{pmatrix}$$

$$D = \begin{pmatrix} \sqrt{P_0} & 0 & \cdots & 0 \\ 0 & \sqrt{P_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{P_{k-1}} \end{pmatrix}.$$

A power allocation that minimises the error variances of the estimates for all $\{j\}$'s –subject to the power constraint (8)– is naturally an optimal one. Examining (9), we note that a power allocation that minimises v_{j_o} for some j_o will also minimise v_j for all $\{j\}$'s, as it will be one that maximises

$$\begin{pmatrix} \alpha^{k-1} & \cdots & 1 \end{pmatrix} D^T \left[DAD^T + \sigma_N^2 I\right]^{-1} D \begin{pmatrix} \alpha^{k-1} \\ \vdots \\ 1 \end{pmatrix}. \tag{10}$$

The power allocation that maximises (10) is the subject of the following theorem, proven in Appendix A.

Theorem 1. *The expression (10) is maximised when all the available power is allocated to the last pilot, that is, $P_j = 0$, for all $0 \leqslant j \leqslant (k-2)$ and $P_{k-1} = kP_{tr}$.*

We note that Theorem 1 holds whenever one allows the power allocation during the training period to vary across the pilots. We also note that the result holds irrespective of how the power is allocated for the data.

4.0.5. Implications on the training duration.

- When considering policy IV, the powers of the individual training symbols are allowed to vary and the theorem states that all the power should be allocated to the last training symbol. Factoring in the loss of achievable rates due to training, it becomes clear that the optimal duration is that of *one* pilot transmission.
- Because the achievable rates using policy III are less or equal to those of policy IV, and because the optimal solution for policy IV is that of a 'flat' power allocation over the duration of the training –which is one, then the solution is also optimal for policy III.
- Finally, because the statement of the theorem is valid irrespective of how the power is allocated during data transmission and specifically even when a flat power allocation is used, the result implies that for policy II, using a training duration of one pilot is optimal as well.

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Naturally, these statements are true if the power level during training is optimised. In Section 5, we validate numerically these results.

5. NUMERICAL RESULTS

For a given channel model, a given SNR (power constraints), and estimation technique (causal or non-causal), we numerically determine the optimal training strategy consisting of:

(1) The duration of training or the number of pilots k.
(2) The inter-pilot spacing T.
(3) The power allocation for the pilots and data symbols in a transmitted frame, according to the transmission policy used.

In our work, the quality measure is the achievable rates which we compute for pilot clustering/training period of up to six pilots in each frame. We study the low *received* SNR regime (SNR values of -3dB, 0dB, 3dB and 6dB) for a first-order Gauss-Markov fading process with values of $\alpha = 0.9, 0.95, 0.97,$ and 0.99. On the receiver side, causal and non-causal estimation are investigated. We present hereafter graphs for some chosen test cases and compare the rates achieved using 1, 2, 3, 4, 5 and 6-pilot clustering strategies for different scenarios of SNR and fading correlation levels.

We note first that the numerical results confirm the observation previously made that the achievable rate in equation (3) depends on the choice of $\{m_1, m_2, p_1\}_i$, that is, the input distribution of the i-th symbol. As the symbol gets further away from the training pilots, the channel estimation quality (measured through the estimate error variance) is degraded and hence the amount of information sent over the channel decreases. This is translated by shifting $\{m_1, m_2, p_1\}_i$ from the antipodal distribution (optimal for a perfectly known channel) with $p_1 \approx p_2$ (high entropy) toward the other extreme of on-off keying (optimal for the IID Rayleigh fading case) with $p_1 \gg p_2$ (low entropy).

We also note that in the case of causal estimation, our numerical results are consistent with the results in Section 4.

5.1. Results for transmission policy I

For transmission policy I, pilot clustering proves to achieve higher rates under certain conditions compared with the 1-pilot scheme. In Figure 4, for an SNR = 0dB, $\alpha = 0.99$ and causal estimation, training with four pilots and inter-pilot spacing of T=29 symbols is optimal. A percent increase of 8.2% in information rate is achieved relative to the best rate achievable with a 1-pilot scheme. The results for other test cases are shown in Table I.

However, there are some scenarios when pilot-clustering is not useful. For the case when SNR = 6dB, $\alpha = 0.97$ and causal estimation, the 1-pilot scheme presents optimal rates.

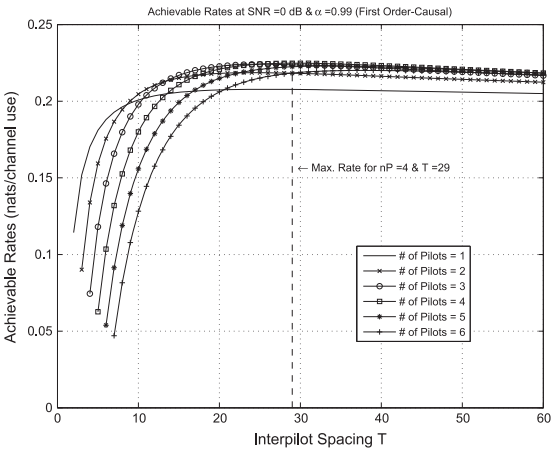

Figure 4. Achievable rates for policy I, for SNR = 0dB and $\alpha = 0.99$ with causal estimation.

Table I. Achievable rates for different transmission policies.

Policy		I	II	III	IV
α : 0.99	nP	4	1	1	1
SNR: 0dB	T	29	22	23	22
Causal Est.	Rate	0.2247	0.2418	0.2422	0.2422
		8.2%↑1	7.6%↑2		
α : 0.97	nP	1	1	1	1
SNR: 6dB	T	15	15	15	15
Causal Est.	Rate	0.3782	0.3829	0.3836	0.3836
			1.2%↑2		
α : 0.97	nP	3	1	1	1
SNR: -3dB	T	19	18	18	18
Non-Causal	Rate	0.1374	0.1470	0.1472	0.1472
		4.3%↑1	6.9%↑2		
Jakes' mod.					
f_d : 100 Hz	nP	2	1	1	1
f_s : 10 KHz	T	14	13	13	13
SNR: 3dB	Rate	0.3224	0.3508	0.3510	0.3510
Causal Est.		4%↑1	8.8%↑2		
Jakes' mod.					
f_d : 100 Hz	nP	4	1	1	1
f_s : 10 KHz	T	29	30	30	30
SNR: 0dB	Rate	0.2843	0.3064	0.3064	0.3064
Non-Causal		16%↑1	7.7%↑2		

[1]relative to the rate achieved by the 1-pilot scheme.
[2]relative to the achievable rate under Policy I.

In Figure 5 at an SNR = 0dB, and $\alpha = 0.9$ with causal estimation, we see that training is not beneficial in the first place because the information rate is less than that achieved over an IID Rayleigh fading channel.

As a conclusion, we can distinguish three cases. The first is when training is not applicable. The second is when the 1-pilot scheme gives the highest rates. And finally the third when pilot clustering is beneficial. From our numerical results, we note that as SNR increases and coherence time decreases, clustering becomes useless and the whole

Trans. Emerging Tel. Tech. **27**:1441–1453 (2016) © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

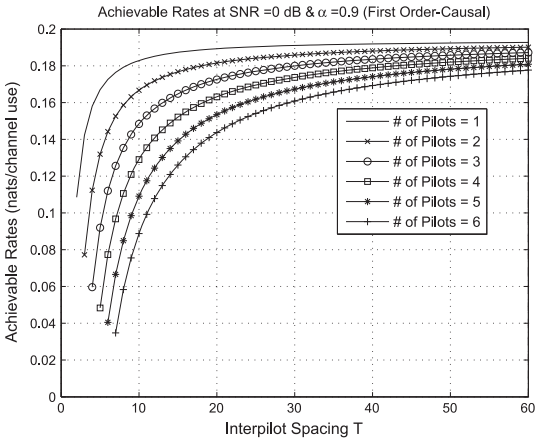

Figure 5. Achievable rates for policy I, for SNR = 0dB and $\alpha = 0.9$ with causal estimation.

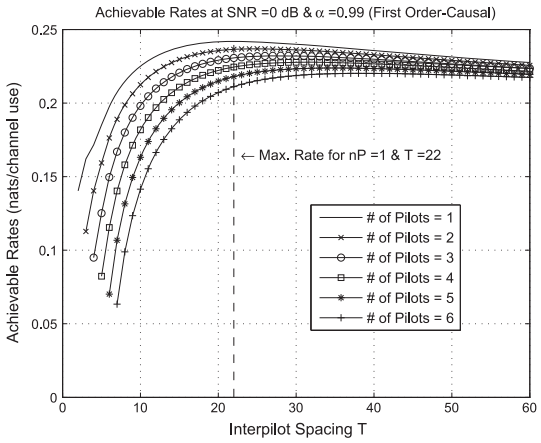

Figure 6. Achievable rates for policy II, for SNR = 0dB and $\alpha = 0.99$ with causal estimation.

scheme is pushed toward the 1-pilot training strategy and even to the extreme case of no training at all. This is directly related to the fact that training is inefficient (less CSI) when fading decorrelates quickly or when the SNR is high.

5.2. Results for transmission policy II

In this policy, the pilots are sent with fixed power P_{tr} (*per pilot*) and so are the symbols that are transmitted with power P_d (Section 3.2.2), such that $\frac{1}{T}\sum_{j=0}^{T-1} P_j \leqslant P$. Therefore, the optimal training strategy includes determining the optimal power allocation (P_{tr} and P_d) for the transmitted frame. Here, the notion of SNR is naturally associated with the average power P.

Figure 6 shows the achievable rates for SNR = 0dB, and $\alpha = 0.99$ with causal estimation. Unlike the results for policy I (Figure 4), training with four pilots is not optimal anymore. The 1-pilot scheme (with T=22) now offers 7.6% increase in the achievable rate compared with the 4-pilot scheme for policy I. The corresponding optimal power allocation across the transmission frame is shown in Figure 7.

The rest of the results are presented in Table I and they all confirm that, as expected pilot clustering is not optimal for policy II, and for any transmission strategy where the pilots' power is subject to optimisation for that matter. In this case, the transmitter decreases the estimation error variance (higher throughput) by boosting the power of the single pilot instead of increasing the number of pilots k and getting penalised by the normalising term $\frac{1}{T}$ in equation (7).

If a peak power constraint is imposed on the power of the pilots, the optimal training duration will not necessarily be one pilot. This can be seen from Figure 8 which shows the optimal power allocation across the transmission frame for SNR = 0dB, and $\alpha = 0.99$ with causal estimation whenever

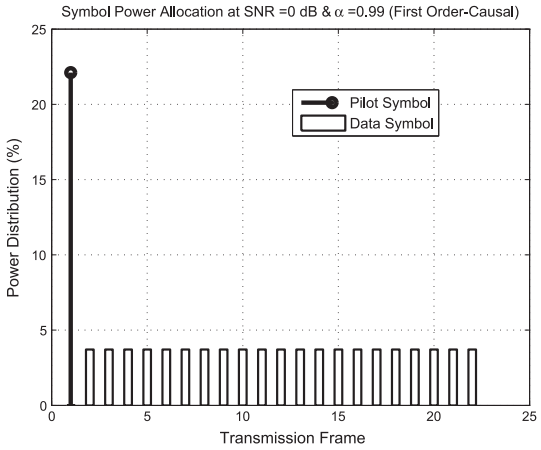

Figure 7. Optimal symbol power allocation (one frame) for policy II, for SNR = 0dB and $\alpha = 0.99$ with causal estimation.

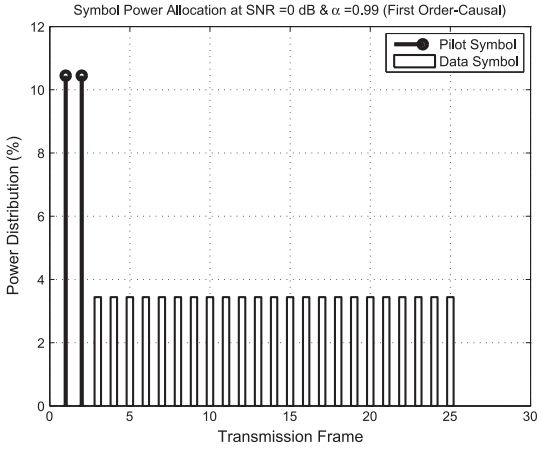

Figure 8. Optimal symbol power allocation (one frame) for policy II, for SNR = 0dB and $\alpha = 0.99$ with causal estimation and peak constraint $P_{tr} \leqslant 3P$.

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

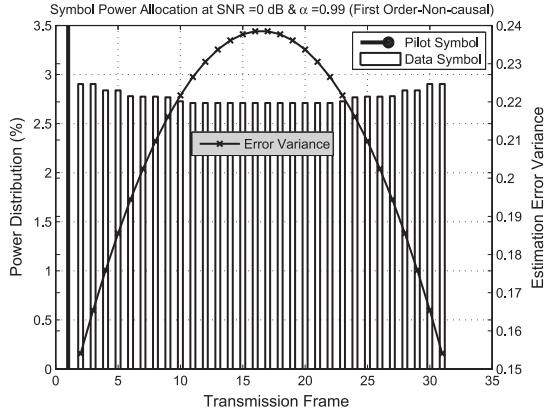

Figure 9. Optimal symbol power allocation (one frame) for policy III, for SNR = 0dB and α = 0.99 with non-causal estimation.

a peak constraint $P_{tr} \leqslant 3P$ is imposed. This constraint is effectively imposing a maximum Peak-to-Average Power Ratio (PAPR) value of 3.

As mentioned earlier, there are some scenarios where training is not useful and the rate is always less than that achieved over an IID Rayleigh fading channel. This is observed with causal estimation for an SNR = 0dB and α = 0.9 for example. In that case, all the power is allocated to the data symbols indicating that training is not beneficial.

5.3. Results for transmission policy III

For policy III, we send the data symbols with varying power as we hold on to a flat power allocation for the pilots. As already shown in the Section 4, clustering is not useful for this case as well. The transmitter boosts the power of the single pilot used in training to decrease the error variance and increase the achievable rate.

The numerical results are in accordance with those of Section 4 and they show how the power of the symbols is adapted to the estimation error variance. In Figure 9, the power allocated to each symbol and the variation of the error variance are presented for an SNR = 0dB, and α = 0.99 with non-causal estimation. This shows that symbols with lower variance are sent with higher power and vice versa. However, we should note that power variations among the data symbols is not profound.

The achievable rates for other cases are summarised in Table I. It is noticed that adapting the symbol power to the quality of estimation introduces a slight increase in achievable rates compared with policy II. As a result, one can say that uniform power allocation for the data symbols is sufficiently close to optimal and presents a more practical transmission strategy.

5.4. Results for transmission policy IV

Here, both the pilots and data symbols are sent with varying power (Section 3.2.4). However, from the results for transmission policy III, we already know that sending the data symbols with uniform power is very close to optimal.

Let us consider the case for an SNR = 0dB, and $\alpha = 0.99$. We choose a 4-pilot training scheme. For causal estimation, the power allocated to the pilots is shown in Figure 10. We notice that all of the power was found numerically to be allocated to the pilot closest to the symbols leaving the rest of the pilots that are further away with no power and therefore useless, which is consistent with the results of Theorem 1 and Section 4. Combining this result with the penalty factor $\frac{1}{T}$ in equation (3), we reach the conclusion that the 1-pilot scheme is always optimal (Table I).

A similar result is shown in Figure 11 for the non-causal estimation scenario. The powers of the first pilot (playing a prominent role in the non-causal part) and last pilot (with a prominent role in the causal part) are increased.

Whenever a peak power constraint is imposed on the power of the pilots, the optimal training duration will

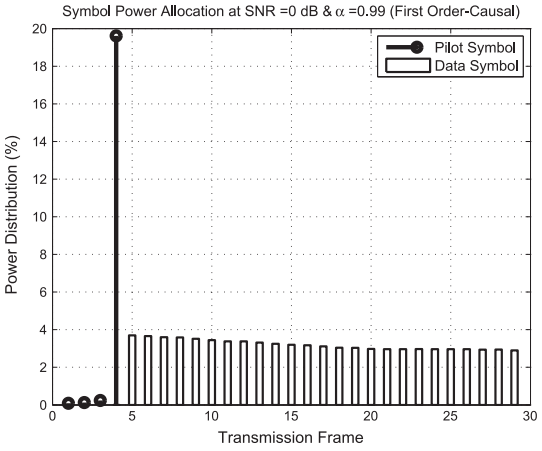

Figure 10. Optimal symbol power allocation (one frame) for policy IV, for SNR = 0dB, α = 0.99 and k = 4 with causal estimation.

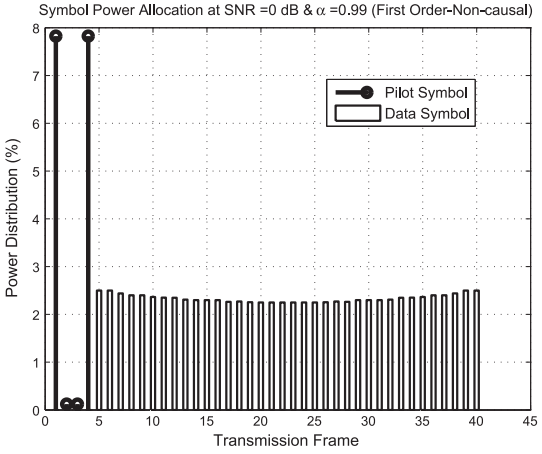

Figure 11. Optimal symbol power allocation (one frame) for policy IV, for SNR = 0dB, α = 0.99 and k = 4 with non-causal estimation.

Trans. Emerging Tel. Tech. **27**:1441–1453 (2016) © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

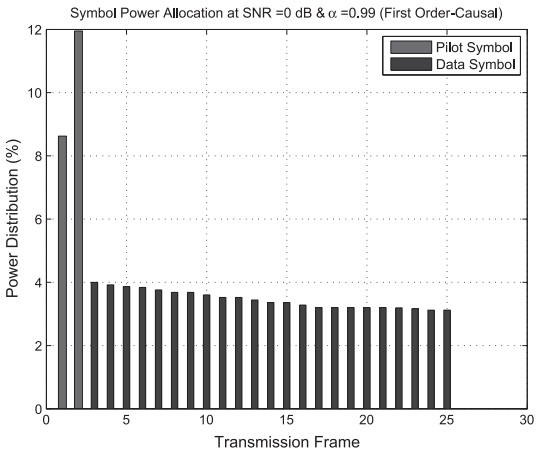

Figure 12. Optimal symbol power allocation (one frame) for policy IV, for SNR = 0dB and α = 0.99 with causal estimation and peak constraint $P_{tr} \leqslant 3P$.

potentially involve pilot clustering. The optimal duration and power allocation in Figure 12 are for an SNR = 0dB, and $\alpha = 0.99$ with causal estimation whenever a peak constraint $P_{tr} \leqslant 3P$ is imposed.

6. EXTENSIONS

6.1. Other fading process models

Whenever the fading process follows a different model, appropriate results may be readily derived as the numerical optimisation is only dependent on the autocovariance function of the process as seen from equation (2). In what follows, we present sample results using Jakes' model.

6.1.1. Jakes' model.

In Jakes' model [25], the normalised (unit variance) continuous-time autocorrelation function of the fading process is given by

$$\phi_{RR}(\tau) = J_0(2\pi f_d \tau),$$

where $J_0(.)$ is the zeroth-order Bessel function of the first kind and f_d is the maximum Doppler frequency. For the purposes of discrete-time simulation of this model [30], the autocorrelation sequence becomes

$$\phi_{RR}[l] = J_0(2\pi f_d T_s |l|),$$

where $1/T_s$ is the symbol rate.

In Table I, we list a sample of the results obtained for a bandwidth $f_s = 10$ kHz and a Doppler shift of $f_d = 100$ Hz. For example, optimal training consists of $k = 4$ and $T = 29$ when we have an SNR = 0dB and non-causal estimation. Throughput is improved by 16% in this case.

6.2. Higher order modulation

When considering large values of SNR, a larger constellation needs to be considered.

The presented analysis may be readily modified, and the optimisation procedure is adapted to constellations of higher order. For a ternary scheme for example, the constellation needs to be optimised over the five parameters:

- Locations: m_1, m_2 and m_3
- Probabilities: p_1 and p_2.

as per the optimisation problem in (5).

7. CONCLUSION

We studied the performance of the non-feedback pilot-based adaptive modulation scheme [21, 22, 31] over time-varying Rayleigh fading channels. We measured the performance in terms of achievable rates using binary signaling and we investigated the benefits of pilot clustering as well as power allocation.

We introduced a modular method to compute the rates in an efficient manner. Moreover, four types of transmission policies were analyzed. For each policy, we determined the optimal training strategy consisting of:

(1) The duration of training.
(2) The inter-pilot spacing.
(3) The power allocation for the pilots and data symbols in the frame.

Pilot clustering proved to be useful in the *low SNR–high coherence time* range where training is efficient (Policy I). However, when the pilot power is subject to optimisation (Policies II, III and IV), training for a smaller period but with boosted power becomes more beneficial than training with more pilots. We proved that the optimal training duration using causal estimation is indeed one whenever the power level during training is optimised and allowed to take arbitrary values. Numerical results suggest that this is also the case when using non-causal estimation at the receiver.

We also noted that the numerical computations indicate that a flat power allocation across the data slots in a frame is very close to optimal whenever the pilot power is subject to optimisation.

On the other hand, training is useless in the *high SNR–small coherence time* range and the rate is always less than that achieved over an IID Rayleigh fading channel. Several test cases are shown throughout this work to analyze how optimal training varies with channel conditions and from one transmission policy to another.

Extensions to this work can include adaptive schemes that integrate temporal and spatial components like the Multiple-Input Multiple-Output (MIMO) scenario.

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

APPENDIX

In this appendix we provide a proof for Theorem 1. For notational convenience, define

$$\phi \triangleq V^T D^T \left[DAD^T + \sigma_N^2 I \right]^{-1} DV,$$

where $V = \left(\alpha^{k-1}\ \alpha^{k-2}\ \cdots\ 1\right)^T$, A is the $k \times k$ symmetric, positive definite autocovariance matrix of the channel fading coefficients $\{R_s\}_{s\in\mathcal{S}}$, and D is the $k \times k$ 'input' matrix:

$$A = \begin{pmatrix} 1 & \alpha & \cdots & \alpha^{k-1} \\ \alpha & 1 & \cdots & \alpha^{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{k-1} & \alpha^{k-2} & \cdots & 1 \end{pmatrix},$$

$$D = \begin{pmatrix} \sqrt{P_0} & 0 & \cdots & 0 \\ 0 & \sqrt{P_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{P_{k-1}} \end{pmatrix},\ V = \begin{pmatrix} \alpha^{k-1} \\ \alpha^{k-2} \\ \vdots \\ 1 \end{pmatrix}.$$

Note that $\phi < 1$ for any $k \geqslant 1$ because $\phi = 1 - v_{k-1}$. We establish first the following lemma:

Lemma 1. *Let U be a $k \times k$ diagonal matrix with non-negative entries $\{x_i\}_{i=0}^{k-1}$ on the diagonal. Among all the permutations of the $\{x_i\}$'s, the one that maximises $V^T [A+U]^{-1} V$ is one where the diagonal entries are in non-increasing order.*

Proof. Assume that $\{x_i\}_{i=0}^{k-1}$ are in the following order: $0 \leqslant x_0 \leqslant x_1 \leqslant \cdots \leqslant x_{k-1}$. We prove in what follows that $U = diag(x_{k-1}, x_{k-2}, \cdots, x_0)$ maximises $V^T [A+U]^{-1} V$ using induction on k. To highlight the dependence on k we denote $\varphi_k = V_k^T [A_k + U_k]^{-1} V_k$, which is a positive quantity due to the positive definiteness of $[A_k + U_k]^{-1}$.

Base Cases For $k = 1$, $\varphi_1 = \frac{1}{1+x_0}$ and the statement holds. Examine now the case $k = 2$:

$$U = diag(x_0, x_1) \Longrightarrow \varphi_2 = \frac{(\alpha^2 x_1 + x_0 - \alpha^2 + 1)}{(x_0+1)(x_1+1) - \alpha^2}$$

$$U = diag(x_1, x_0) \Longrightarrow \varphi_2 = \frac{(\alpha^2 x_0 + x_1 - \alpha^2 + 1)}{(x_1+1)(x_0+1) - \alpha^2}.$$

Because $\alpha < 1$ and $x_1 \geqslant x_0$, the second value is larger.

Induction Step Suppose the property holds true up to $k-1$ ($k-1 \geqslant 2$) and we prove in what follows that it holds true for k:

$$\varphi_k = \left(\alpha^{k-1}\ \alpha^{k-2}\ \cdots\ 1\right) [A_k + U_k]^{-1} \begin{pmatrix} \alpha^{k-1} \\ \alpha^{k-2} \\ \vdots \\ 1 \end{pmatrix},$$

where A_k and U_k are square matrices of size k. We prove that φ_k is maximised when $\{x_i\}_{i=0}^{k-1}$ are placed in non-increasing order on the diagonal matrix U_k. The proof proceeds as follows: We first 'fix' x_{k-1} on the last diagonal entry of U_k and prove that $\{x_i\}_{i=0}^{k-2}$ should be in a non-increasing order to maximise φ_k. Next, we 'fix' $\{x_i\}_{i=0}^{k-3}$ on the first $(k-2)$ diagonal entries of U_k and we prove that, if $x_{k-2} \leqslant x_{k-1}$, having $U = diag\{x_0, x_1, \cdots, x_{k-3}, x_{k-1}, x_{k-2}\}$ (versus $U = diag\{x_0, x_1, \cdots, x_{k-3}, x_{k-2}, x_{k-1}\}$) gives us a larger value of φ_k, completing the proof.

- Using a block form, we write $[A_k + U_k]$ as:

$$A_k + U_k = \begin{bmatrix} E & F \\ F^T & G \end{bmatrix}$$

where

$$E = \begin{pmatrix} x_0+1 & \alpha & \cdots & \alpha^{k-2} \\ \alpha & x_1+1 & \cdots & \alpha^{k-3} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{k-2} & \alpha^{k-3} & \cdots & x_{k-2}+1 \end{pmatrix},$$

$$F = \begin{pmatrix} \alpha^{k-1} \\ \alpha^{k-2} \\ \vdots \\ \alpha \end{pmatrix} = \alpha V_{k-1} \quad \& \quad G = \left(x_{k-1} + 1\right).$$

This allows us to express $[A_k + U_k]^{-1}$ as [32]:

$$[A_k + U_k]^{-1} = \begin{bmatrix} E^{-1} + E^{-1}F[G - F^T E^{-1} F]^{-1} F^T E^{-1} & -E^{-1}F[G - F^T E^{-1}F]^{-1} \\ -[G - F^T E^{-1} F]^{-1} F^T E^{-1} & [G - F^T E^{-1}F]^{-1} \end{bmatrix}$$

and φ_k is equal to

$$F^T \left[E^{-1} + E^{-1}F[G - F^T E^{-1} F]^{-1} F^T E^{-1} \right] F$$
$$-[G - F^T E^{-1} F]^{-1} F^T E^{-1} F + [G - F^T E^{-1} F]^{-1}$$
$$-F^T E^{-1} F[G - F^T E^{-1} F]^{-1}$$

which reduces to:

$$\varphi_k = F^T E^{-1} F + \frac{\left(1 - F^T E^{-1} F\right)^2}{\left[x_{k-1} + 1 - F^T E^{-1} F\right]}$$
$$= F^T E^{-1} F + \frac{\left(x_{k-1} + 1 - F^T E^{-1} F - x_{k-1}\right)^2}{\left[x_{k-1} + 1 - F^T E^{-1} F\right]}$$
$$= (-x_{k-1} + 1) + \frac{x_{k-1}^2}{\left[x_{k-1} + 1 - F^T E^{-1} F\right]}.$$

Trans. Emerging Tel. Tech. **27**:1441–1453 (2016) © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The scalar $F^T E^{-1} F$ is equal to $\alpha^2 \varphi_{k-1}$. Indeed, F is of size $(k-1)\times 1$ and equal to αV_{k-1}, and E is a $(k-1)\times(k-1)$ sub-matrix of the form $[A_{k-1} + U_{k-1}]$. Because $\alpha < 1$, the scalar $F^T E^{-1} F$ is less than one and the denominator is a positive quantity. Therefore, with x_{k-1} fixed, φ_k is maximised when $F^T E^{-1} F$ is maximised. By the induction step, with a fixed x_{k-1} the remaining x_i's should be 'placed' in decreasing order on the diagonal of E –and U– to maximise φ_k.

- Now fix $\{x_i\}_{i=0}^{k-3}$. We prove that with $x_{k-2} \leqslant x_{k-1}$, $U = diag\{x_0, x_1, \cdots, x_{k-3}, x_{k-1}, x_{k-2}\}$ gives us a larger value for φ_k. To do this, we consider a different decomposition of the matrix $[A_k + U_k]$,

$$A_k + U_k = \begin{bmatrix} E & F \\ F^T & G \end{bmatrix},$$

where now

$$E = \begin{pmatrix} x_0+1 & \alpha & \cdots & \alpha^{k-3} \\ \alpha & x_1+1 & \cdots & \alpha^{k-4} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{k-3} & \alpha^{k-4} & \cdots & x_{k-3}+1 \end{pmatrix}$$

$$F = \begin{pmatrix} \alpha^{k-2} & \alpha^{k-1} \\ \alpha^{k-3} & \alpha^{k-2} \\ \vdots & \vdots \\ \alpha & \alpha^2 \end{pmatrix}$$

$$\& \quad G = \begin{pmatrix} x_{k-2}+1 & \alpha \\ \alpha & x_{k-1}+1 \end{pmatrix}.$$

Because $F = \begin{pmatrix} \alpha V_{k-2} & \alpha^2 V_{k-2} \end{pmatrix}$, φ_k is equal to

$$\alpha^4 V_{k-2}^T \left[E^{-1} + E^{-1} F \left[G - F^T E^{-1} F \right]^{-1} F^T E^{-1} \right] V_{k-2}$$
$$- 2\alpha^2 \begin{pmatrix} \alpha & 1 \end{pmatrix} \left[G - F^T E^{-1} F \right]^{-1} F^T E^{-1} V_{k-2}$$
$$+ \begin{pmatrix} \alpha & 1 \end{pmatrix} \left[G - F^T E^{-1} F \right]^{-1} \begin{pmatrix} \alpha \\ 1 \end{pmatrix}.$$

Noting that $\varphi_{k-2} = V_{k-2}^T E^{-1} V_{k-2}$, φ_k becomes

$$= \alpha^4 \varphi_{k-2} + \alpha^6 \varphi_{k-2}^2 \begin{pmatrix} 1 & \alpha \end{pmatrix} \left[G - F^T E^{-1} F \right]^{-1} \begin{pmatrix} 1 \\ \alpha \end{pmatrix}$$
$$- 2\alpha^3 \varphi_{k-2} \begin{pmatrix} \alpha & 1 \end{pmatrix} \left[G - F^T E^{-1} F \right]^{-1} \begin{pmatrix} 1 \\ \alpha \end{pmatrix}$$
$$+ \begin{pmatrix} \alpha & 1 \end{pmatrix} \left[G - F^T E^{-1} F \right]^{-1} \begin{pmatrix} \alpha \\ 1 \end{pmatrix}$$
$$= \alpha^4 \varphi_{k-2} + \xi(x_{k-2}, x_{k-1}),$$

where

$$\xi(x_{k-2}, x_{k-1}) = \xi$$
$$= \begin{pmatrix} \alpha(1-\alpha^2\varphi_{k-2}) & 1-\alpha^4\varphi_{k-2} \end{pmatrix} \left[G - F^T E^{-1} F \right]^{-1} \begin{pmatrix} \alpha(1-\alpha^2\varphi_{k-2}) \\ 1-\alpha^4\varphi_{k-2} \end{pmatrix}$$
$$= \begin{pmatrix} \alpha - \alpha^3\varphi_{k-2} & 1-\alpha^4\varphi_{k-2} \end{pmatrix} \begin{pmatrix} x_{k-2}+1-\alpha^2\varphi_{k-2} & \alpha-\alpha^3\varphi_{k-2} \\ \alpha-\alpha^3\varphi_{k-2} & x_{k-1}+1-\alpha^4\varphi_{k-2} \end{pmatrix}^{-1} \begin{pmatrix} \alpha-\alpha^3\varphi_{k-2} \\ 1-\alpha^4\varphi_{k-2} \end{pmatrix}. \quad \text{(A1)}$$

Checking the two possibilities, $\xi(x_{k-2}, x_{k-1}) - \xi(x_{k-1}, x_{k-2})$ has the same sign as

$$\Big[(\alpha-\alpha^3\varphi_{k-2})^2 x_{k-1} + (1-\alpha^4\varphi_{k-2})^2 x_{k-2} + (1-\alpha^2\varphi_{k-2})(1-\alpha^4\varphi_{k-2})(1-\alpha^2)\Big]$$
$$\Big[x_{k-2}x_{k-1} + (1-\alpha^4\varphi_{k-2})x_{k-1} + (1-\alpha^2\varphi_{k-2})x_{k-2} + (1-\alpha^2\varphi_{k-2})(1-\alpha^2)\Big]$$
$$- \Big[(\alpha-\alpha^3\varphi_{k-2})^2 x_{k-2} + (1-\alpha^4\varphi_{k-2})^2 x_{k-1} + (1-\alpha^2\varphi_{k-2})(1-\alpha^4\varphi_{k-2})(1-\alpha^2)\Big]$$
$$\Big[x_{k-2}x_{k-1} + (1-\alpha^4\varphi_{k-2})x_{k-2} + (1-\alpha^2\varphi_{k-2})x_{k-1} + (1-\alpha^2\varphi_{k-2})(1-\alpha^2)\Big],$$

which is zero if $\alpha = 1$ or $x_{k-2} = x_{k-1}$. Assuming $x_{k-2} < x_{k-1}$, it is of the same sign as

$$-\left(1-\alpha^6\varphi_{k-2}^2\right)x_{k-2}x_{k-1} - (1-\alpha^2\varphi_{k-2})^2(1-\alpha^2) - (1-\alpha^4\varphi_{k-2})(1-\alpha^2\varphi_{k-2})(x_{k-1}+x_{k-2}),$$

which is negative and hence

$$\xi(x_{k-2}, x_{k-1}) < \xi(x_{k-1}, x_{k-2}).$$

We conclude that, when fixing $\{x_i\}_{i=0}^{k-3}$, φ_k is maximised when the last two diagonal elements x_{k-2} and x_{k-1} are placed in non-increasing order.

The final step in the proof is to note that if the diagonal entries are not in non-increasing order, then either the first $(k-1)$ entries are not or the last two entries are not. This contradicts the previous two properties.

Before we state and prove the theorem, a couple of quantities that will come in handy hereafter are the partial derivatives of $\xi(x_{k-2}, x_{k-1})$ defined in (A1):

$$\frac{\partial}{\partial x_{k-2}}\xi \quad \propto \quad -\alpha^2(1-\alpha^2\varphi_{k-2})^2 x_{k-1}^2 \qquad \text{(A2)}$$

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\frac{\partial}{\partial x_{k-1}}\xi \quad \propto \quad -\Big[(1-\alpha^2\varphi_{k-2})(1-\alpha^2) +(1-\alpha^4\varphi_{k-2})x_{k-2}\Big]^2, \tag{A3}$$

where the expressions earlier are those of the respective numerators. We note that both quantities are non-positive and everything else being constant, the value of ξ decreases as x_{k-2} or x_{k-1} increases.

Theorem 2. *When maximising the scalar ϕ over all the choices of $\{P_j\}_0^{k-1}$ such that $\sum_{j=0}^{k-1} P_j = kP_{tr}$, the maximum is achieved when all the available power is allocated to the last pilot, i.e., $P_j = 0$, for all $0 \leqslant j \leqslant (k-2)$ and $P_{k-1} = kP_{tr}$.*

Proof. We start by imposing a lower bound on the powers $\{P_j\}$'s. More precisely, for some small enough $\epsilon > 0$, we assume that $P_j = \epsilon + P'_j$ and

$$D = \begin{pmatrix} \sqrt{\epsilon + P'_0} & 0 & \cdots & 0 \\ 0 & \sqrt{\epsilon + P'_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\epsilon + P'_{k-1}} \end{pmatrix},$$

and we optimise over the $\{P'_j\}$'s subject to the constraint

$$\sum_{j=0}^{k-1} P'_j \leqslant kP_{tr} - k\epsilon. \tag{A4}$$

The diagonal matrix D is non-singular, allowing us to express the objective function ϕ as:

$$\phi = V^T \left[A + \sigma_N^2 D^{-1} D^{-1}\right]^{-1} V.$$

Applying the result of Lemma 1 with $U = \sigma_N^2 D^{-1} D^{-1}$ –and diagonal entries $x_i = \sigma_N^2 \frac{1}{\epsilon + P'_i}$, yields that the optimal $\{P'_j\}$'s have to be non-decreasing. Additionally, the derivative (A3) indicates that the upperbound (A4) will be tight. Indeed, fixing $\{P'_0, \cdots, P'_{k-2}\}$ (or equivalently $\{x_0, \cdots, x_{k-2}\}$) and increasing P'_{k-1} (or equivalently decreasing x_{k-1}) will increase $\phi (= \varphi_k)$. This asserts that the power on the last pilot should be as large as possible so that the upper bound (A4) is met with equality.

The derivatives (A2) and (A3) allow us to make an even stronger statement: If $\{P'_0, \cdots, P'_{k-3}\}$ are fixed, among the choices of P'_{k-2} and P'_{k-1} such that

$$P'_{k-2} + P'_{k-1} \leqslant kP_{tr} - k\epsilon - \sum_{j=0}^{k-3} P'_j \triangleq M,$$

the one that maximises φ_k is $P'_{k-2} = 0$ and $P'_{k-1} = M$.

Indeed, because the bound will be met with equality and P'_{k-2} is less or equal to P'_{k-1} (by the result of Lemma 1), we let $P'_{k-2} = p$ and $P'_{k-1} = M - p$ and optimise over $p \in [0, M/2]$. Equivalently, $x_{k-2} = \sigma_N^2 \frac{1}{\epsilon + p}$, $x_{k-1} = \sigma_N^2 \frac{1}{\epsilon + M - p}$, and because φ_{k-2} is fixed, the derivative of φ_k with respect to p is

$$\begin{aligned} \frac{d}{dp}\varphi_k &= \frac{\partial \xi}{\partial x_{k-2}} \cdot \frac{dx_{k-2}}{dp} + \frac{\partial \xi}{\partial x_{k-1}} \cdot \frac{dx_{k-1}}{dp} \\ &\propto \frac{\alpha^2(1-\alpha^2\varphi_{k-2})^2}{(\epsilon+p)^2(\epsilon+M-p)^2} \\ &\quad - \frac{\left[(1-\alpha^2\varphi_{k-2})(1-\alpha^2)\frac{(\epsilon+p)}{\sigma_N^2} + (1-\alpha^4\varphi_{k-2})\right]^2}{(\epsilon+p)^2(\epsilon+M-p)^2}, \end{aligned}$$

which is of the same sign as

$$\begin{aligned} &\alpha(1-\alpha^2\varphi_{k-2}) - \Big[(1-\alpha^4\varphi_{k-2}) \\ &\qquad + (1-\alpha^2\varphi_{k-2})(1-\alpha^2)\frac{(\epsilon+p)}{\sigma_N^2}\Big] \\ &\leqslant \alpha(1-\alpha^2\varphi_{k-2}) - (1-\alpha^4\varphi_{k-2}) \\ &= (-1-\alpha^3\varphi_{k-2})(1-\alpha) \leqslant 0, \end{aligned}$$

for any p and therefore the maximum is attained when $p = 0$. Said differently, φ_k is maximum when $P'_{k-2} = 0$ and $P'_{k-1} = M$.

By Lemma 1, the optimal values of P'_i for all $i \in \{0, 1, \cdots, k-3\}$ are less or equal to P'_{k-2}. Because for an optimal power allocation P'_{k-2} is zero, $P'_i = 0$ for all $i \in \{0, 1, \cdots, k-2\}$ and $P'_{k-1} = kP_{tr} - k\epsilon$.

Finally, the same previous observations show that the smaller the ϵ the larger ϕ is. Consequently, taking the limit as ϵ goes to zero yields the optimal solution and the proof of the theorem is complete.

REFERENCES

1. Lapidoth A, Shamai S. Fading channels: how perfect need "perfect side information"? *IEEE Transactions on Information Theory* 2002; **48**(5): 1118–1134.
2. Vakili A, Sharif M, Hassibi B. The effect of channel estimation error on the throughput of broadcast channels. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, 2006; IV–IV.
3. Wang J, Li M, Zhang Y, Zhou Q. Effect of channel estimation error on the mutual information of MIMO fading channels. In *IEEE International conference on Wireless Communications, Networking and Mobile Computing WiCOM*, Dalian, 2008; 1–4.
4. Lozano A, Heath R, Andrews J. Fundamental limits of cooperation. *IEEE Transactions on Information Theory* 2013; **59**(9): 5213–5226.

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5. Cavers JK. An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *IEEE Transactions on Vehicular Technology* 1991; **40**(11): 686–693.
6. Toni L, Chan YS, Cosman P, Milstein L. Channel coding for progressive images in a 2-d time-frequency OFDM block with channel estimation errors. *IEEE Transactions on Image Processing* 2009; **18**(11): 2476 –2490.
7. Goldsmith AJ, Chua SG. Variable-rate variable-power MQAM for fading channels. *IEEE Transactions on Communications* 1997; **45**(10): 1218–1230.
8. Cai X, Giannakis GB. Adaptive PSAM accounting for channel estimation and prediction errors. *IEEE Transactions on Wireless Communications* 2005; **4**(1): 246–256.
9. Lamahewa TA, Sadeghi P, Kennedy R, Rapajic P. Model-based pilot and data power adaptation in psam with periodic delayed feedback. *IEEE Transactions on Wireless Communications* 2009; **8**(5): 2247–2252.
10. Agarwal M, Honig M, Ata B. Adaptive training for correlated fading channels with feedback. *IEEE Transactions on Information Theory* 2012; **58**(8): 5398–5417.
11. Cai X, Giannakis G. Adaptive PSAM accounting for channel estimation and prediction errors. *IEEE Transactions on Wireless Communications* 2005; **4** (1): 246–256.
12. Ekpenyong A, Huang YF. Feedback constraints for adaptive transmission. *IEEE Signal Processing Magazine* 2007; **24**(3): 69–78.
13. Misra S, Swami A, Tong L. Cutoff rate optimal binary inputs with imperfect CSI. *IEEE Transactions on Wireless Communications* 2006; **5**(10): 2903–2913.
14. Akin S, Gursoy M. Achievable rates and training optimization for fading relay channels with memory. In *Conference on Information Systems & Sciences CISS*, Princeton, 2008; 185–190.
15. Almustafa K, Primak S, Willink T, Baddour K. On achievable data rates and optimal power allocation in fading channels with imperfect CSI. In *International Symposium on Wireless Communication Systems ISWCS*, Trondheim, 2007; 282–286.
16. Akin S, Gursoy M. Training optimization for gauss-markov rayleigh fading channels. In *IEEE International Conference on Communications ICC*, Glasgow, 2007; 5999–6004.
17. Savazzi S, Spagnolini U. Optimizing training lengths and training intervals in time-varying fading channels. *IEEE Transactions on Signal Processing* 2009; **57**(3): 1098–1112.
18. Zhang H, Wei S, Ananthaswamy G, Goeckel D. Adaptive signaling based on statistical characterizations of outdated feedback in wireless communications. *Proceedings of the IEEE* 2007; **95**(12): 2337–2353.
19. Duong D, Holter B, Oien G. Optimal pilot spacing and power in rate-adaptive MIMO diversity systems with imperfect transmitter CSI. In *IEEE International Workshop on Signal Processing Advances in Wireless Communications*, New York, 2005; 47–51.
20. Maaref A, Aissa S. Optimized rate-adaptive PSAM for MIMO MRC systems with transmit and receive CSI imperfections. *IEEE Transactions on Communications* 2009; **57**(3): 821–830.
21. Abou-Faycal I, Médard M, Madhow U. Binary adaptive coded pilot symbol assisted modulation over Rayleigh fading channels without feedback. *IEEE Transactions on Communications* 2005; **53**(6): 1036–1046.
22. Zeineddine K, Abou-Faycal I. How much training is optimal in adaptive PSAM over Markov Rayleigh fading channels. In *IEEE International Symposium on Signal Processing and Information Technology ISSPIT*, Ajman, 2009; 366–371.
23. Médard M. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Transactions on Information Theory* 2000; **46**(3): 935–946.
24. Tan C, Beaulieu N. On first-rder Markov modeling for the Rayleigh fading channel. *IEEE Transactions on Communications* 2000; **48**(12): 2032–2040.
25. Jakes WC. *Microwave Mobile Communications*. Wiley: New York, 1974.
26. Abou-Faycal I, Trott MD, Shamai S. The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels. *IEEE Transactions on Information Theory* 2001; **47**(4): 1290–1301.
27. Gallager R. Power Limited Channels: Coding, Multiaccess, and Spread Spectrum. *MIT LIDS* 1987.
28. Verdu S. On Channel Capacity per unit cost. *IEEE Transactions on Information Theory* 1990; **36** (5): 1019–1030.
29. Luo C. Communication for Wideband Fading Channels: on Theory and Practice. *Ph.D. Thesis*, 2006.
30. Baddour KE, Beaulieu NC. Autoregressive Modeling for Fading Channel Simulation. *IEEE Transactions on Wireless Communications* 2005; **4**(4): 1650–1662.
31. Bdeir A, Abou-Faycal I, Médard M. Power Allocation Schemes for Pilot Symbol Assisted Modulation over Rayleigh Fading Channels with no Feedback. In *IEEE International Conference on Communications ICC*, Paris, 2004; 737–741.
32. Petersen KB, Pedersen MS. The matrix cookbook. Version 20121115, 2012.

21613915, 2016, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ett.3069 by American University Of Beirut, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License