Optimizing Training-based MIMO Systems: How Much Time is Needed for Actual Transmission?

Xiangyun Zhou, Parastoo Sadeghi and Tharaka A. Lamahewa
College of Engineering and Computer Science, The Australian National University, Canberra 0200 ACT, Australia
Email:{xiangyun.zhou, parastoo.sadeghi, tharaka.lamahewa}@anu.edu.au

***Abstract*—We study the design of training-based multiple-input multiple-output systems in two block-wise transmission schemes. The conventional transmission scheme has a fixed amount of energy to be used in each block, hence transmission takes place in every block. For this scheme, we study the optimality of using all available time in each block for transmission and provide bounds to significantly reduce the ranges of the possible values of the optimal training and data lengths. The second scheme, called the flashy transmission scheme, is constrained by an average amount of energy per block, and uses some but not necessarily all blocks for transmission. For this scheme, we find the optimal fraction of blocks to be used for transmission. When this optimal fraction is less than one, we show that the optimal training and data lengths are independent of the energy constraint.**

I. INTRODUCTION

Channel estimation plays a crucial role for high data rate transmission in wireless communications with coherent detection. Many practical systems periodically insert training symbols, which are known to the receiver, into data blocks to facilitate channel estimation [1]. In resource constrained communications, the total available transmission time and energy should be optimally distributed among training and data symbols to achieve maximum data rates. Numerous results on optimizing resource allocation in training-based transmissions have been obtained in the past few years. For constantly time-varying channels, the optimal training length and training interval were investigated in [2–5]. For block fading channels with fixed coherence block length, the training length for multiple-input multiple-output (MIMO) systems was optimized in [6, 7]. A common assumption in these works is that all available time is used for transmission.

In this paper, we revisit the common assumption used in aforementioned works by first relaxing it and then investigating its optimality in training-based transmissions. This is particularly important for many practical systems in which the transmit power for training symbols and data symbols is fixed to the same level. In this scenario, the optimal training and data lengths are a function of the transmit power and need to be found numerically [7]. We consider two block-wise transmission schemes with a short-term and a long-term energy constraint, respectively. The first scheme, named the conventional transmission scheme, is constrained by a *fixed* amount of energy to be used in each block, and hence the training and data transmissions take place in every block. The second scheme, named the flashy transmission scheme as proposed in [8], is constrained by an *average* amount of energy per block, and it uses some but not necessarily all blocks for transmission. We denote the fraction of all the blocks to be used for transmission as δ. Denoting the energy constraint by $\mathcal{E}$, the actual transmit power is given by

$$\mathcal{P}_t = \frac{\mathcal{E}}{\delta(L_p + L_d)},$$

where L_p and L_d are the training length and the data length in a transmission block of length L with $L_p + L_d \leq L$. In the conventional transmission scheme, $\mathcal{E}$ represents the *fixed* energy to be used in every block and $\delta = 1$. In this scheme, $\mathcal{P}_t$ can be boosted up by reducing L_p and/or L_d. In the flashy transmission scheme, $\mathcal{E}$ represents the *average* energy per block. In this scheme, the system designer has control over an additional parameter δ which gives an extra degree of freedom in transmission design. Therefore, there is clearly a trade-off between the transmit power and transmission time in optimizing system performance.

To study the optimal amount of time to be used for training and data transmissions, we consider training-based MIMO systems in block fading channels. Both the conventional and the flashy transmission schemes are considered. We use an average capacity lower bound as the figure of merit. The main results of this paper are summarized as follows.

- If a conventional transmission scheme allows the training and data lengths to be jointly optimized, we show that all available time should always be used for transmission. We also provide analytical bounds to significantly reduce the ranges of the possible values of the optimal training and data lengths.
- If a conventional transmission scheme is designed to have a fixed training length, we derive a threshold SNR above which it is optimal to use all available time for transmission.
- In the flashy transmission scheme, we analytically find the optimal fraction of coherence blocks for transmission, denoted by δ^*, as well as the critical SNR below which flashy transmission becomes necessary, *i.e.*, $\delta^* < 1$. We also show that the optimal training and data lengths are independent of the energy constraint when $\delta^* < 1$.

II. SYSTEM MODEL

We consider a MIMO flat-fading channel model with input-output relationship given by $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{y}$ is the $N_r \times 1$ received symbol vector, $\boldsymbol{x}$ is the $N_t \times 1$ transmitted

978-1-4244-2519-8/10/$26.00 ©2010 IEEE

symbol vector. $\boldsymbol{n}$ is the $N_r \times 1$ noise vector having independent and identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian (ZMCSCG) entries with variance normalized to unity. $\boldsymbol{H}$ is the $N_r \times N_t$ channel gain matrix whose entries are also i.i.d. ZMCSCG with unit variance.

We assume that the channel gains remain constant over one coherence block of L symbol periods and change to independent realizations in the next block. The conventional transmission scheme uses all the coherence blocks for transmission, while the flashy scheme only chooses a fraction of all the coherence blocks for transmission. We refer to this fraction as *flashiness* and denote it by δ. We also name the coherence blocks which are used for actual transmission as the *transmission blocks*. At the beginning of each transmission block, each transmit antenna sends L_p training symbols followed by L_d data symbols, with $L_p + L_d \leq L$. To obtain a meaningful estimate of $\boldsymbol{H}$, we assume that $L_p \geq N_t$ to ensure the receiver has at least as many measurements as unknowns [7].

In the conventional transmission scheme, $\mathcal{E}$ denotes the *fixed* amount of energy to be used in each block. While in the flashy transmission scheme, $\mathcal{E}$ denotes the *average* amount of energy per block. We also define the average transmit power per symbol as $\mathcal{P} = \mathcal{E}/(LT_s)$, where T_s is the symbol duration. Without loss in generality, we let $T_s = 1$. Due to the normalization in the variances of the noise and channel gains, we also refer to $\mathcal{P}$ as the average SNR. We assume that the actual transmit power is fixed to the same level for both training and data symbols, and hence is given by

$$\mathcal{P}_t = \frac{\mathcal{E}}{\delta(L_p + L_d)} = \frac{\mathcal{P}}{\delta}\frac{L}{L_p + L_d}. \tag{1}$$

A. Average Capacity Lower Bound

We adopt a widely-used lower bound for the average capacity with the minimum mean square error estimator (MMSE) for channel estimation derived in [7] and take the flashy transmission scheme into account as

$$\begin{aligned} C_{\text{LB}} &= \delta\frac{L_d}{L}E_{\boldsymbol{H}}\Big\{\log_2\Big|\boldsymbol{I}+\rho_{\text{eff}}\frac{\boldsymbol{H}\boldsymbol{H}^\dagger}{N_t}\Big|\Big\} \\ &= \delta n\frac{L_d}{L}E_\lambda\{\log_2(1+\rho_{\text{eff}}\lambda)\}, \end{aligned} \tag{2}$$

where $n = \min\{N_t, N_r\}$ and λ is an arbitrary eigenvalue of $\boldsymbol{H}\boldsymbol{H}^\dagger/N_t$. Note that $\delta = 1$ for the conventional transmission scheme. ρ_{eff} is named the effective SNR in [7] given by

$$\begin{aligned} \rho_{\text{eff}} &= \frac{(1-\sigma_e^2)\mathcal{P}_t}{1+\sigma_e^2\mathcal{P}_t} \\ &= \frac{L_p\mathcal{E}^2}{\delta(L_p+L_d)(N_t\mathcal{E}+\delta N_t L_p+\delta N_t L_d+L_p\mathcal{E})}, \end{aligned} \tag{3}$$

where $\sigma_e^2 = (1 + \mathcal{P}_t L_p/N_t)^{-1}$ is the variance of the channel estimation error. In following sections, we will use (2) as the figure of merit to investigate the optimal transmission design.

III. Conventional Transmission Scheme

We first study the conventional transmission scheme, *i.e.*, $\delta = 1$, with a fixed energy constraint $\mathcal{E}$ for each block. Two scenarios are considered: a) the training length L_p and the data length L_d are to be jointly optimized; b) the data length L_d is to be optimized for a given training length L_p.

A. Joint Training and Data Length Optimization

In general, the optimal training and data lengths, denoted by L_p^* and L_d^* respectively, need to be found numerically by evaluating the capacity lower bound for all possible values of L_p and L_d. Also, it is not clear whether $L_p + L_d = L$ is optimal for all SNR conditions. In the following, we provide analytical bounds on L_p^* and L_d^* as well as the optimality of $L_p + L_d = L$.

Theorem 1: When both L_p and L_d are allowed to be optimized, the optimal strategy is to use all available time for transmission at any SNR values, i.e., $L_p^* + L_d^* = L$. *The optimal value of L_d and L_p satisfy the following conditions.*

$$\begin{cases} L_d^* = L - N_t, \;\; L_p^* = N_t, & \textit{if } L \leq 2N_t \\ \zeta \leq L_d^* \leq L - N_t, \;\; N_t \leq L_p^* \leq L - \zeta, & \textit{else} \end{cases} \tag{4}$$

where

$$\zeta = \frac{2L\mathcal{P}+N_t\mathcal{P}+2N_t-\sqrt{N_t(N_t\mathcal{P}^2+4N_t\mathcal{P}+4N_t+4L\mathcal{P}^2+4L\mathcal{P})}}{2\mathcal{P}}$$

is the value of L_d at which we have

$$\sqrt{\frac{N_t L_d(L_d+L\mathcal{P})}{N_t+L\mathcal{P}}} + L_d = L, \tag{5}$$

Proof: see Appendix A.

When $\mathcal{P} \to 0$, we see from (5) that $\zeta \to L/2$. Therefore, the lower (upper) bound on L_d^* (L_p^*) in (4) approaches $L/2$. In fact, by using the first order approximation of the capacity lower bound at sufficiently low SNR, it can be shown that the optimal training and data lengths satisfy $L_p^* = L_d^* = L/2$ [7]. Therefore, we expect that the lower (upper) bound on L_d^* (L_p^*) given in *Theorem 1* is tight at sufficiently low SNR. When $\mathcal{P} \to \infty$, it can be shown that L_d should be chosen as large as possible, *i.e.*, $L_d^* = L - N_t$. (The proof follows Appendix A by letting $\mathcal{P} \to \infty$.) Therefore, we expect that the upper (lower) bound on L_d^* (L_p^*) given in *Theorem 1* is tight at sufficiently high SNR. We note that the optimality of $L_p = N_t$ at sufficiently high SNR was also commented in [7].

Fig. 1 shows the upper bound on L_p^* given in (4) as well as the exact values of L_p^* found numerically. We see that the upper bound significantly reduces the range of possible values of L_p^* at moderate to high SNR. For example in finding L_p^* for a 2×2 system with $L = 100$ operating at 5 dB, the analytical bounds tell that one only needs to search from 2 to 15, instead of searching every possible number from 2 to 99. In addition, we see that the upper bound gets tighter as SNR decreases.

B. Optimal Data Length for a Given Training Length

As we have seen, the training and data lengths, L_p and L_d, need be found numerically according to the operating SNR which usually varies with time. When the operating SNR changes rapidly with time, it is undesirable to frequently change L_p and L_d. In this scenario, we consider a transmission

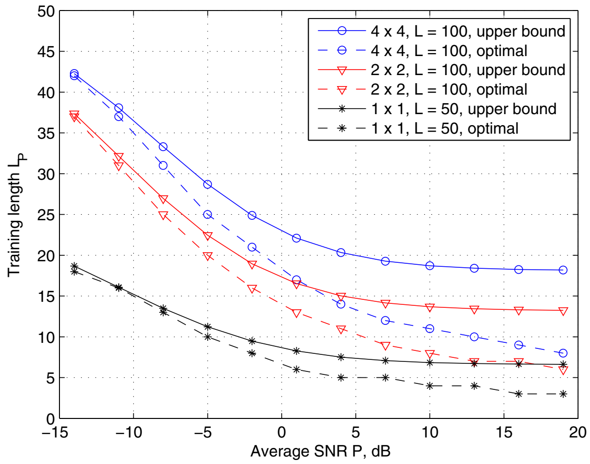

Fig. 1. The upper bound on L_p^* given in (4) versus average SNR for different block lengths and different numbers of transmit antennas.

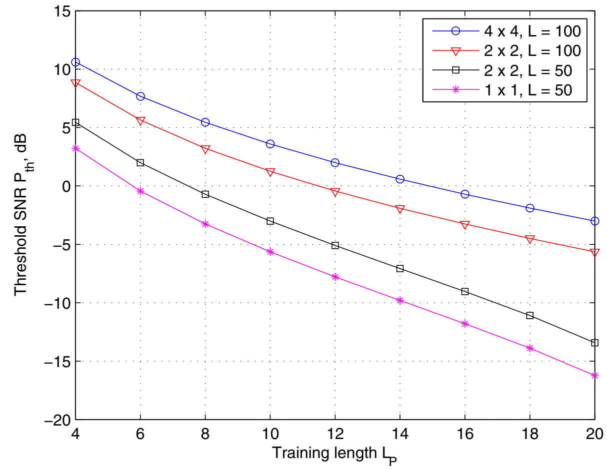

Fig. 2. The threshold SNR $\mathcal{P}_{\text{th}}$ given in (6) versus training length L_p for different block lengths and different numbers of transmit antennas.

strategy which optimizes L_p (and L_d) for a target SNR and fix it for a certain time period before redesigning becomes crucial.

From Fig. 1 we see that L_p^* decreases as SNR increases. If the operating SNR is higher than the target SNR, the system is using more training resource than the optimal amount, hence it is still desirable to use all available time for data transmission, *i.e.*, $L_d^* = L - L_p$. On the other hand, if the operating SNR is lower than the target SNR, the system is using insufficient amount of training which results in a degradation in channel estimation. In this scenario, the system may need to at least reduce the data length from $L - L_p$ to boost up the SNR.[1]

In the following, we find a threshold SNR $\mathcal{P}_{\text{th}}$, such that the optimal data length is given by $L_d^* = L - L_p$ for any given L_p as long as the operating SNR is above $\mathcal{P}_{\text{th}}$. That is, for a system with a fixed L_p which has been optimized according to some target SNR, it can keep using the original design of $L_d = L - L_p$ as long as the operating SNR is above $\mathcal{P}_{\text{th}}$.

Lemma 1: For any given L_p, a threshold SNR above which the optimal L_d equals $L - L_p$ is given by

$$\mathcal{P}_{\text{th}} = \frac{N_t(L - 2L_p)}{L_p(L_p + N_t)} \tag{6}$$

Proof: see Appendix B.

Fig. 2 shows the threshold SNR $\mathcal{P}_{\text{th}}$ given in (6) versus training length L_p. If a 2×2 system with $L = 100$ has a fixed training length of $L_p = 10$ which is optimal for a target SNR of 5 dB as shown in Fig. 1, the data length does not need to be reduced from $L - L_p = 90$ as long as the operating SNR is above $\mathcal{P}_{\text{th}} = 1.3$ dB as shown in Fig. 2. Note that the optimal data length may or may not equal $L - L_p$ when the operating SNR is below $\mathcal{P}_{\text{th}}$.

[1]For a given L_p, $L_d^* < L - L_p$ can happen at sufficiently low SNR. To see this, one can apply the first order approximation of the capacity lower bound and obtain the optimal data length by letting the derivative of the capacity approximation be zero.

IV. Flashy Transmission Scheme

Wideband wireless communication systems often operate in the low SNR regime. If the system is constrained by an average amount of energy per block $\mathcal{E}$, the flashy transmission scheme can be used to boost up the transmit power $\mathcal{P}_t$ by reducing the flashiness δ. We consider the scenario that the training and data lengths can be jointly optimized, and hence we have $L_p^* + L_d^* = L$. In this section, we aim to find the optimal flashiness, denoted by δ^*, which maximizes the capacity lower bound in the low SNR regime.

In the low SNR regime, *e.g.*, $\mathcal{P}_t \le -5$ dB, the second order approximation of the capacity lower bound is accurate, which is given by

$$\begin{aligned} C_{\text{LB}} &\approx \frac{nL_d}{L \ln 2} \delta E_\lambda \left\{ \rho_{\text{eff}} \lambda - \frac{\rho_{\text{eff}}^2 \lambda^2}{2} \right\} \\ &= \frac{nL_d}{L\ln 2}\Big(\frac{L_p M_1 \mathcal{P}^2}{\delta N_t + (N_t + L_p)\mathcal{P}} - \frac{M_2 \mathcal{P}^4}{2\delta(\delta N_t + (N_t + L_p)\mathcal{P})^2}\Big), \end{aligned} \tag{7}$$

where we have used $\mathcal{P} = \mathcal{E}/L$ and $M_k = E_\lambda\{\lambda^k\}$ is the kth raw moment of an unordered eigenvalue of $\boldsymbol{H}\boldsymbol{H}^\dagger/N_t$, which can be easily found in closed form by using the unordered eigenvalue distribution of $\boldsymbol{H}\boldsymbol{H}^\dagger$ given in [9].

By letting the first derivative of C_{LB} in (7) w.r.t. δ be zero, we obtain

$$\begin{aligned} &-M_1 N_t^2 \Big(\frac{\delta}{\mathcal{P}}\Big)^3 - M_1 N_t (N_t + L_p) \Big(\frac{\delta}{\mathcal{P}}\Big)^2 + \frac{3}{2} M_2 N_t L_p \frac{\delta}{\mathcal{P}} \\ &+ \frac{1}{2} M_2 L_p (N_t + L_p) = 0, \end{aligned} \tag{8}$$

which is a third order equation of $\delta/\mathcal{P}$. From the signs of the coefficients in (8), it can be shown that there are one positive root and two negative roots. Therefore, the positive root is a local maxima in $0 < \delta < \infty$. The optimal value of δ is then given by

$$\delta^* = \min\{1, \mathcal{P}F\}, \tag{9}$$

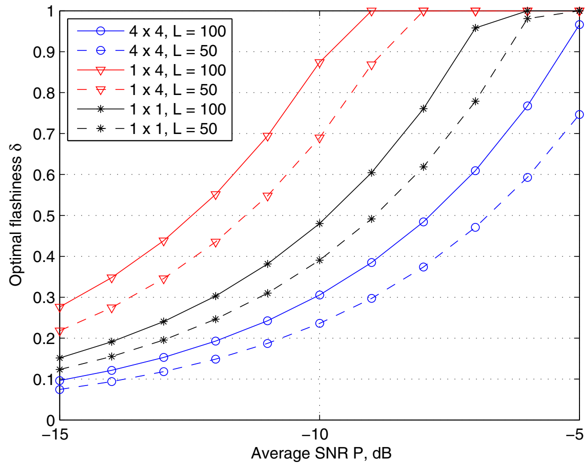

Fig. 3. The optimal value of δ versus average SNR for different block lengths and different numbers of antennas. The training length is optimized numerically.

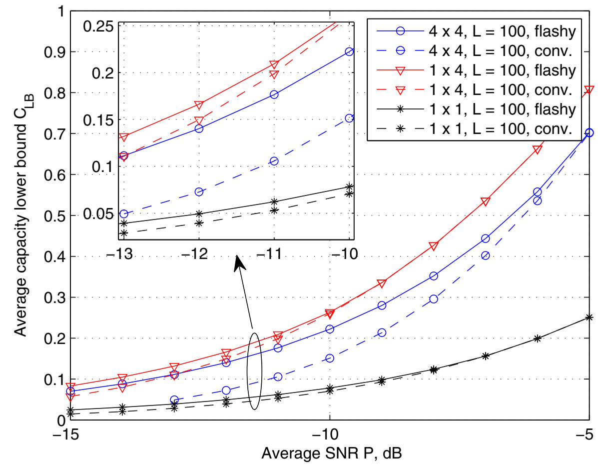

Fig. 4. Average capacity lower bound $C_{\rm LB}$ in (2) versus average SNR in both flashy and conventional transmission schemes with $L = 100$. The optimal values of δ shown in Fig. 3 are used in the flashy scheme. The training length is optimized numerically in both schemes.

where F is the positive root of (8)[2] which depends on N_t, N_r and L_p, but not on $\mathcal{P}$. From (9), we see that flashy transmission becomes necessary, *i.e.*, $\delta^* < 1$, when the average SNR falls below a critical value given by F^{-1}.

Since we have $L_p + L_d = L$, using (1) we get $\mathcal{P}_t = \mathcal{P}/\delta$. Hence the actual transmit power in flashy transmission scheme with δ^* in (9) is given by

$$\mathcal{P}_t = \mathcal{P}/\delta^* = \max\{\mathcal{P}, F^{-1}\}. \tag{10}$$

We see from (10) that the actual transmit power equals F^{-1}, which is a constant independent of the energy constraint, as long as the average power or SNR is below the critical value F^{-1}. This result directly leads to the following lemma.

Lemma 2: In the SNR regime where flashy transmission is necessary and the optimal δ given in (9) is used, the optimal training and data lengths are constants independent of the energy budget.

With the optimal δ given in (9), *Lemma 2* implies that L_p and L_d only need to be optimized once and can then be used at any values of $\mathcal{P}$ which satisfies $\mathcal{P} \leq F^{-1}$.

To give a flavor of the expression of δ^*, we provide the result for single-input single-output (SISO) systems as

$$\delta^* = \min\left\{1, \mathcal{P}\left(\frac{\epsilon^{1/3}}{3} + \left(\frac{L_p^2}{3} + \frac{11}{3}L_p + \frac{1}{3}\right)\epsilon^{-1/3} - \frac{1+L_p}{3}\right)\right\},$$

where $\epsilon = -(1+L_p)^3 + 3\sqrt{-3L_p - 39L_p^2 - 153L_p^3 - 39L_p^4 - 3L_p^5}$, and the critical SNR value below which the flashy scheme becomes necessary is hence given by

$$F^{-1} = \left(\frac{\epsilon^{1/3}}{3} + \left(\frac{L_p^2}{3} + \frac{11}{3}L_p + \frac{1}{3}\right)\epsilon^{-1/3} - \frac{1+L_p}{3}\right)^{-1}.$$

In the following, we present numerical results on the optimal value of δ as well as the capacity gain from flashy transmission with the optimal L_p found numerically. That is, we use δ given in (9) for any L_p and numerically find L_p^* which gives the highest capacity, then present the corresponding δ^* and $C_{\rm LB}$.

Fig. 3 shows the optimal value of δ. We see that δ^* decreases as $\mathcal{P}$ decreases which agrees with (9). In addition, δ^* tends to be smaller for systems with more transmit antennas or smaller block length.

Fig. 4 shows the average capacity lower bound in both flashy and conventional transmission schemes. We see that flashy transmission with optimal δ can significantly improve the capacity at low SNR. For example, the percentage capacity improvement from conventional transmission to the optimal flashy transmission at -12 dB is 100%, 11% and 26% for 4×4, 1×4 and 1×1 systems, respectively.

V. Conclusion

We have studied the design of training-based MIMO transmissions under two different energy constraints. We considered a conventional transmission scheme with a fixed amount of energy for each block and a flashy transmission scheme with an average amount of energy per block. In the conventional scheme, we showed that it is always optimal to use all available time for transmission only when the training length and the data length can be jointly optimized. We also provided bounds on the optimal training and data lengths. When the training length is fixed, we provided a threshold SNR above which it is still optimal to use all available time for transmission. For the flashy scheme, we analytically found the optimal fraction of coherence blocks to be used as well as the critical SNR below which flashy transmission becomes necessary. Our numerical result confirmed that the flashy transmission can achieve higher capacity than that in the conventional transmission at low SNR.

[2]The expression of F is rather long, hence we omit it for brevity. However, to give a flavor of its expression, we will show it for SISO systems.

Appendix A
Proof of Theorem 1

Finding the bounds on L_d^* and L_p^* involves two steps. First we find the optimal L_p for any given L_d. We then find the range of values in which L_d^* lies. For any given L_d, the capacity lower bound (2) is maximized when the effective SNR (3) is maximized. Letting the first derivative of ρ_{eff} w.r.t. L_p be zero, we obtain only one positive root for L_p given by

$$L_p^0 = \arg_{L_p}\left\{\frac{\mathrm{d}\rho_{\text{eff}}}{\mathrm{d}L_p}=0\right\} = \sqrt{\frac{N_t L_d(L_d+L\mathcal{P})}{N_t+L\mathcal{P}}}, \tag{11}$$

where we have used $\mathcal{E}=L\mathcal{P}$. Note that the training length must satisfy $N_t \le L_p \le L-L_d$. Therefore, for a given L_d the optimal training length is given by

$$L_p^* = \begin{cases} N_t, & \text{for } L_p^0 \le N_t \\ L_p^0, & \text{for } N_t \le L_p^0 \le L-L_d \\ L-L_d, & \text{for } L-L_d \le L_p^0, \end{cases} \tag{12}$$

which by referring to (11) is equivalent to

$$L_p^* = \begin{cases} N_t, & \text{for } 1 \le L_d \le N_t \\ L_p^0, & \text{for } N_t \le L_d \le \zeta \\ L-L_d, & \text{for } \zeta \le L_d \le L-N_t, \end{cases} \tag{13}$$

where

$$\zeta=\frac{2L\mathcal{P}+N_t\mathcal{P}+2N_t-\sqrt{N_t(N_t\mathcal{P}^2+4N_t\mathcal{P}+4N_t+4L\mathcal{P}^2+4L\mathcal{P})}}{2\mathcal{P}} \tag{14}$$

is the value of L_d at which $\sqrt{\frac{N_tL_d(L_d+L\mathcal{P})}{N_t+L\mathcal{P}}}+L_d=L$.

In order to find the range of values in which L_d^* lies, we consider two different cases, *i.e.*, $L \le 2N_t$ and $L > 2N_t$.

When $L \le 2N_t$, using (14) it can be shown that $L-N_t \le \zeta \le N_t$. Therefore, only the first range of L_d in (13) is feasible and $L_p^*=N_t$. In order to show the optimal data length takes the maximum possible value, *i.e.*, $L_d^*=L-L_p^*=L-N_t$, we need to show that the derivative of C_{LB} w.r.t. L_d is positive at $L_p^*=N_t$. Using (2) we have

$$\begin{aligned}\frac{\mathrm{d}C_{\text{LB}}}{\mathrm{d}L_d} &= \frac{n}{L\ln 2}E_\lambda\Big\{\ln(1+\rho_{\text{eff}}\lambda)+\frac{\lambda L_d}{1+\rho_{\text{eff}}\lambda}\frac{\mathrm{d}\rho_{\text{eff}}}{\mathrm{d}L_d}\Big\}\\ &= \frac{n}{L\ln 2}E_\lambda\Big\{\ln(1+\rho_{\text{eff}}\lambda)-\frac{\rho_{\text{eff}}\lambda}{1+\rho_{\text{eff}}\lambda}L_d\rho_{\text{eff}}\frac{\mathrm{d}\rho_{\text{eff}}^{-1}}{\mathrm{d}L_d}\Big\}.\end{aligned}$$

Since both ρ_{eff} and λ are non-negative, it can be shown that

$$\ln(1+\rho_{\text{eff}}\lambda)-\frac{\rho_{\text{eff}}\lambda}{1+\rho_{\text{eff}}\lambda}\ge 0.$$

Therefore, $L_d\rho_{\text{eff}}\frac{\mathrm{d}}{\mathrm{d}L_d}\rho_{\text{eff}}^{-1}<1$ implies $\frac{\mathrm{d}}{\mathrm{d}L_d}C_{\text{LB}}>0$. Using (3) it is then easy to show that $L_d\rho_{\text{eff}}\frac{\mathrm{d}}{\mathrm{d}L_d}\rho_{\text{eff}}^{-1}<1$ at $L_p^*=N_t$. Hence we obtain $L_d^*=L-N_t$.

When $L>2N_t$, we aim to show that the optimal data length resides in the last range in (13), *i.e.*, $\zeta \le L_d^* \le L-N_t$. This is proved by showing that L_d^* cannot reside in $1 \le L_d < N_t$ and $N_t \le L_d < \zeta$. When $1 \le L_d < N_t$, it is easy to show that $L_d\rho_{\text{eff}}\frac{\mathrm{d}}{\mathrm{d}L_d}\rho_{\text{eff}}^{-1}<1$ at $L_p^*=N_t$ given by (13). This implies $\frac{\mathrm{d}}{\mathrm{d}L_d}C_{\text{LB}}>0$ and hence L_d^* is not in $1 \le L_d < N_t$. When $N_t \le L_d < \zeta$, it is easy to show that $L_d\rho_{\text{eff}}\frac{\mathrm{d}}{\mathrm{d}L_d}\rho_{\text{eff}}^{-1}<1$ at $L_p^*=L_p^0$ given by (13). This implies $\frac{\mathrm{d}}{\mathrm{d}L_d}C_{\text{LB}}>0$ and hence L_d^* is not in $N_t \le L_d < \zeta$. Therefore, we have shown that $\zeta \le L_d^* \le L-N_t$, which also gives $L_p^*=L-L_d^*$ from (13).

Appendix B
Proof of Lemma 1

We need to show that for all SNR values above $\mathcal{P}_{\text{th}}$, we have $\frac{\mathrm{d}C_{\text{LB}}}{\mathrm{d}L_d}>0$ for $0 \le L_d \le L-L_p$. To find such a $\mathcal{P}_{\text{th}}$, we use the result in Appendix A which states that $L_d\rho_{\text{eff}}\frac{\mathrm{d}\rho_{\text{eff}}^{-1}}{\mathrm{d}L_d}<1$ implies $\frac{\mathrm{d}C_{\text{LB}}}{\mathrm{d}L_d}>0$. For any given L_p, we use ρ_{eff} in (3) with $\mathcal{E}=L\mathcal{P}$ to obtain

$$\begin{aligned}L_d\rho_{\text{eff}}\frac{\mathrm{d}\rho_{\text{eff}}^{-1}}{\mathrm{d}L_d} &= \frac{L_d(N_tL\mathcal{P}+2N_tL_d+2N_tL_p+L_pL\mathcal{P})}{(L_p+L_d)(N_tL\mathcal{P}+N_tL_p+N_tL_d+L_pL\mathcal{P})}\\ &= \frac{\beta+N_tL_d^2}{\beta+L_p^2L\mathcal{P}+N_tL_pL\mathcal{P}+N_tL_p^2},\end{aligned}$$

where $\beta=N_tL_dL\mathcal{P}+N_tL_d^2+2N_tL_pL_d+L_pL_dL\mathcal{P}$. We see that $L_d\rho_{\text{eff}}\frac{\mathrm{d}\rho_{\text{eff}}^{-1}}{\mathrm{d}L_d}<1$ reduces to $L_p^2L\mathcal{P}+N_tL_pL\mathcal{P}+N_tL_p^2>N_tL_d^2$, which needs to hold for $0 \le L_d \le L-L_p$. Therefore, we need to find the values of $\mathcal{P}$ which satisfy $L_p^2L\mathcal{P}+N_tL_pL\mathcal{P}+N_tL_p^2>N_t(L-L_p)^2$, given by

$$\mathcal{P}>\frac{N_tL-2N_tL_p}{L_p(L_p+N_t)}. \tag{15}$$

Therefore, a threshold SNR value, above which we have $\frac{\mathrm{d}C_{\text{LB}}}{\mathrm{d}L_d}>0$, is found as in (6).

Acknowledgements

This work was supported under the Australian Research Council's Discovery Projects funding scheme (project no. DP0773898).

References

[1] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[2] J. Baltersee, G. Fock, and H. Meyr, "An information theoretic foundation of synchronized detection," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2115–2123, Dec. 2001.

[3] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 712–720, Oct. 2002.

[4] V. Pohl, P. H. Nguyen, V. Jungnickel, and C. Helmolt, "Continuous flat-fading MIMO channels: Achievable rate and optimal length of the training and data phases," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1889–1900, July 2005.

[5] S. Savazzi and U. Spagnolini, "Optimizing training lengths and training intervals in time-varying fading channels," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 1098–1112, Mar. 2009.

[6] T. L. Marzetta, "BLAST training: estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, Sept. 1999, pp. 22–24.

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

[8] L. Zheng, D. N. C. Tse, and M. Médard, "Channel coherence in the low-SNR regime," *IEEE Trans. Inform. Theory*, vol. 53, no. 3, pp. 976–997, Mar. 2007.

[9] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.