Training sequence optimization for frequency selective channels with MAP equalization

Imed Hadj Kacem, Noura Sellami
Laboratoire LETI
ENIS, Route Sokra km 3.5
B.P 3038 Sfax, Tunisia

Aline Roumy
IRISA-INRIA
Campus de Beaulieu
35042 Rennes Cedex, France

Inbar Fijalkow
Laboratoire ETIS, UMR 8051
ENSEA-UCP-CNRS, 6 av. du Ponceau
95014 Cergy-Pontoise, France

Abstract—In this paper, we address the problem of optimization of the training sequence length for frequency selective channels when a Maximum *a posteriori* (MAP) equalizer is used. The optimal length of the training sequence is found by maximizing an effective signal-to-noise ratio (SNR) and an effective channel capacity of the training-based transmission scheme. We study these problems of optimization when the training and data powers are equal and when they are allowed to be different. When the powers can be different, we give the optimal power allocation.

I. Introduction

Equalization is used to combat intersymbol interference on frequency selective channels. The optimal equalizer [1] to be used is based on Maximum *a posteriori* (MAP) detection. It makes decision on a symbol-by-symbol basis and is optimum since it minimizes the bit error probability when the channel is known by the receiver. In practice, the channel impulse response is estimated by sending known training symbols. When the length of the training sequence increases, the variance of the channel estimation error decreases, but the information throughput decreases as well. Thus, a trade-off has to be found. Several methods have been proposed to design the optimal training sequence length. The solution presented in [2] and [3] is based on maximizing a lower bound of the capacity of the training-based scheme respectively for a single-input single-output (SISO) frequency selective channel and for a multiple-input multiple-output (MIMO) flat fading channel. Another approach tries to find the optimal sequence that minimizes the Mean Square Error (MSE) of the channel estimator for different systems [4]-[7]. All these works do not take into account the receiver used.

In this paper, we consider the particular case where a MAP equalizer is used for a transmission over a SISO frequency selective channel. We introduce simple expressions of the effective Signal-to-Noise Ratio (SNR) and the effective channel capacity for the training-based scheme. We find the training sequence lengths maximizing these quantities when the training and data powers are equal. When the powers can be different, we also give the optimal power allocation.

The paper is organized as follows. In Section II, we describe the transmission system model. Section III studies the optimization of the training sequence length when the training and data powers are equal. In Section IV, we find the optimal power allocation and the optimal training sequence length when the training and data powers are allowed to be different. Throughout this paper scalars and matrices are lower and upper case respectively and vectors are underlined lower case. The operator $(.)^T$ denotes the transposition, and I_m is the $m \times m$ identity matrix. $\lfloor x \rfloor$, $\lceil x \rceil$ and $|x|$ are respectively the greatest integer lower than x, the smallest integer greater than x and the absolute value of x.

II. Transmission system model

We consider a data transmission system over a frequency selective channel. The input information bit sequence is mapped to the symbol alphabet $\mathcal{A}$. For simplicity, we will consider only the BPSK modulation ($\mathcal{A} = \{-1, 1\}$). We assume that transmissions are organized into bursts of T symbols. The channel is supposed to be invariant during the transmission. The received baseband signal sampled at the symbol rate at time k is

$$x_k = \sum_{l=0}^{L-1} h_l s_{k-l} + n_k \qquad (1)$$

where L is the channel memory and s_k, for $1-L \leq k \leq T-1$, are the transmitted symbols. In this expression, n_k are modeled as independent random variables of real white Gaussian noise with normal probability density function $\mathcal{N}(0, \sigma^2)$ where $\mathcal{N}(\alpha, \sigma^2)$ denotes a Gaussian distribution with mean α and variance σ^2. The term h_l is the lth tap gain of the channel, which is assumed to be real valued.

The channel is estimated by using a training sequence of length $T_p \geq 2L - 1$. We assume that this sequence has ideal autocorrelation and crosscorrelation properties. Let $\widetilde{\underline{s}} = (s_{T_p-L}, ..., s_{1-L})^T$ be the vector of training symbols and $\underline{h} = (h_0, ..., h_{L-1})^T$ the vector of channel taps. The least square channel estimate $\widehat{\underline{h}} = \left(\widehat{h}_0, ..., \widehat{h}_{L-1} \right)^T$ is given by [8]:

$$\widehat{\underline{h}} = \left(H_L \left(\widetilde{\underline{s}} \right)^T H_L \left(\widetilde{\underline{s}} \right) \right)^{-1} H_L \left(\widetilde{\underline{s}} \right)^T \widetilde{\underline{x}} \qquad (2)$$

where $H_L \left(\widetilde{\underline{s}} \right)$ is the $(T_p - L + 1) \times L$ Hankel matrix having the first column $\left(s_{T_p-L}, ..., s_0 \right)^T$ and the last row $(s_0, ..., s_{1-L})$ and $\widetilde{\underline{x}} = (x_0, ..., x_{T_p-L})^T$ is the output of the channel corresponding to the training sequence.

Hence, we obtain

$$\widehat{\underline{h}} - \underline{h} \sim \mathcal{N}\left(0, \sigma^2 \left(H_L\left(\widetilde{\underline{s}}\right)^T H_L\left(\widetilde{\underline{s}}\right)\right)^{-1}\right)$$
$$= \mathcal{N}\left(0, \frac{\sigma^2}{(T_p - L + 1)\sigma_p^2} I_L\right) \tag{3}$$

where σ_p^2 is the transmit power during the training phase.

III. OPTIMIZATION OF THE TRAINING SEQUENCE LENGTH FOR EQUAL POWERS

We consider at the receiver a MAP equalizer using the BCJR algorithm [1]. We first assume that the transmit powers during the training and data transmission phases are equal to the unit. We will be interested in the optimization of the training sequence length by maximizing an effective SNR and an effective channel capacity that we will define.

A. Maximization of the effective SNR

When the channel is estimated, the Bit Error Rate (BER) at the output of the MAP equalizer can be approximated at high SNR as [9]

$$P \simeq Q\left(\frac{d_{min}}{2\sigma}\left(1 + \frac{L}{T_p - L + 1}\right)^{-\frac{1}{2}}\right) \tag{4}$$

where d_{min} is the channel minimum distance [10].
Hence, the equivalent signal to noise ratio at the output of the MAP equalizer using the channel estimate is given by

$$SNR_{eq,\hat{h}} = \frac{d_{min}^2}{4\sigma^2}\left(1 + \frac{L}{T_p - L + 1}\right)^{-1} \tag{5}$$

Increasing the training sequence length leads to an improvement of the channel estimate quality but also to a loss in data throughput. Thus, in order to take account this loss, we define an effective SNR at the output of the equalizer as

$$SNR_{eff,eq,\hat{h}} = \frac{T - T_p}{T} SNR_{eq,\hat{h}}$$
$$= \frac{T - T_p}{T} \frac{d_{min}^2}{4\sigma^2}\left(1 + \frac{L}{T_p - L + 1}\right)^{-1} \tag{6}$$

Our goal is to maximize $SNR_{eff,eq,\hat{h}}$ under the constraints $T_p \leq T$ and $T_p \geq r_0$, where $r_0 = 2L - 1$. Let $x \in \mathbb{R}_+$ and $f(x) = \frac{d_{min}^2}{4\sigma^2} \frac{T-x}{T}\left(1 + \frac{L}{x - L + 1}\right)^{-1}$. Notice that $SNR_{eff,eq,\hat{h}} = f(T_p)$. Let $f''(x)$ be the second derivative of $f(x)$. Since $f''(x) < 0$, for $x \in \mathbb{R}_+$, the function f is concave. Thus, it has a unique maximum reached for $x^* \in \mathbb{R}_+$, such as:

$$x^* = -1 + \sqrt{L + TL}. \tag{7}$$

We consider the two possible cases:
- If $x^* < r_0 \Leftrightarrow T < 4L - 1$ then the length of the training sequence T_p^* maximizing $SNR_{eff,eq,\hat{h}}$ is equal to r_0.
- If $r_0 \leq x^* \leq T \Leftrightarrow T \geq 4L - 1$ then $T_p^* = r_1^*$ where $r_1^* = \arg\max_{x \in \{\lfloor x^* \rfloor, \lceil x^* \rceil\}} f(x)$.
We can summarize the previous results as follows:
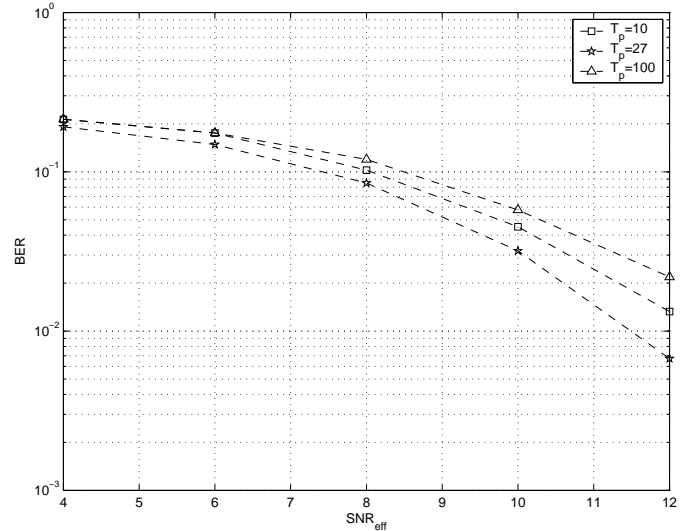
$$T_p^* = (r_1^* - r_0)^+ + r_0 \tag{8}$$

Fig. 1. MAP equalizer BER performance versus SNR_{eff} for different values of the length of the training sequence.

where $(x)^+ = \begin{cases} x \text{ if } x \geq 0; \\ 0 \text{ elsewhere}. \end{cases}$
When $T \geq 4L - 1$, the optimum value of $SNR_{eff,eq,\hat{h}}$ can be approximated by

$$SNR_{eff,eq,\hat{h}}^* \simeq \frac{d_{min}^2}{4\sigma^2}\left(1 + \frac{L+1}{T} - 2\frac{\sqrt{L + TL}}{T + 1}\right). \tag{9}$$

Simulation results

In our simulations, we consider $Channel 3$ with impulse response $(0.5; 0.71; 0.5)$. Figure 1 shows the BER performance of the MAP equalizer when the channel is estimated, for different values of the length of the training sequence ($T_p = 10$, 27 and 100) with respect to $SNR_{eff} = \frac{T - T_p}{T} SNR$, where SNR is the signal-to-noise ratio at the input of the MAP equalizer. We set the number of symbols per burst T to 256. According to the previous analytical results, the optimal length of the training sequence is $T_p^* = 27$. This is confirmed by the simulations since they show that the equalizer presents its best performance, at high SNR, when $T_p = T_p^*$.

B. Maximization of the effective channel capacity

In the case of channel estimation, the channel capacity of the training-based scheme using the MAP equalizer is given by

$$C = \frac{1}{2}\log\left(1 + SNR_{eq,\hat{h}}\right). \tag{10}$$

In order to take into account the loss in channel capacity due to the pilot symbols, we define an effective channel capacity as

$$C_{eff} = \frac{1}{2}\frac{T - T_p}{T}\log(1 + SNR_{eq,\hat{h}}). \tag{11}$$

We define $g(x) = (T - x)\log\left(1 + \frac{d_{min}^2}{4\sigma^2}\frac{x - L + 1}{x + 1}\right)$ for $x \in \mathbb{R}_+$, then $C_{eff} = g(T_p)$. Since $g''(x) < 0$, for $x \in \mathbb{R}_+$, g is concave. Hence, it has an unique maximum reached for $y^* \in \mathbb{R}_+$. As $g'(0)g'(T) < 0$, then according to the

theorem of intermediate values $y^* \in]0, T[$. Thus, two cases are considered:

- If $g'(r_0) \leq 0$ then $T_p^* = r_0$.
- If $g'(r_0) > 0$ then $y^* \in]r_0, T[$ and $T_p^* = \arg\max_{y \in \{\lfloor y^* \rfloor, \lceil y^* \rceil\}} g(y)$.

Hence,

$$T_p^* = (r_2^* - r_0)^+ + r_0 \qquad (12)$$

where $r_2^* = \arg\max_{y \in \{\lfloor y^* \rfloor, \lceil y^* \rceil\}} g(y)$.

IV. Joint optimization of the training sequence length and power allocation

We assume now that the training and data powers are allowed to be different. Thus, the T_p pilot symbols are transmitted with a power σ_p^2 and the $T_d = T - T_p$ data symbols are transmitted with a power σ_d^2.

A. Maximization of the effective signal-to-noise ratio

When the channel is estimated by the least square estimator and pilot and data symbols have different power levels, the expression of the BER at the output of the MAP equalizer is given by

$$P = Q\left(\sqrt{\frac{d_{min}^2}{4} \frac{\sigma_d^2}{\sigma^2}\left(1 + \frac{L\sigma_d^2}{(T_p - L + 1)\sigma_p^2}\right)^{-1}}\right). \qquad (13)$$

This result can be proved by using the same proof as in [9] while taking into account the pilot and data powers.
The equivalent signal-to-noise ratio at the output of the MAP equalizer becomes

$$SNR_{eq,\hat{h}} = \frac{d_{min}^2}{4} \frac{\sigma_d^2}{\sigma^2}\left(1 + \frac{L\sigma_d^2}{(T_p - L + 1)\sigma_p^2}\right)^{-1}. \qquad (14)$$

In this case, we define the effective signal-to-noise ratio $SNR_{eff,eq,\hat{h}}$ as:

$$SNR_{eff,eq,\hat{h}} = \frac{T - T_p}{T} \frac{d_{min}^2}{4} \frac{\sigma_d^2}{\sigma^2}\left(1 + \frac{L\sigma_d^2}{(T_p - L + 1)\sigma_p^2}\right)^{-1} \qquad (15)$$

Now, consider the following optimization problem:

$$\begin{cases} \max \ SNR_{eff,eq,\hat{h}}\left(T_p, \sigma_p^2, T_d, \sigma_d^2\right) \\ \text{s.t.} \\ \sigma_p^2 T_p + \sigma_d^2 T_d = \sigma_t^2 T \\ T_p + T_d = T \\ \sigma_p^2, \sigma_d^2 \geq 0 \\ r_0 \leq T_p \leq T \end{cases} \qquad (16)$$

where $\sigma_t^2 T$ is the total transmit energy per burst.
We denote the fraction of the total transmit energy used in the data transmission phase as

$$\sigma_d^2 T_d = \alpha \sigma_t^2 T, \ 0 < \alpha < 1 \qquad (17)$$

Thus, the effective SNR can be written as

$$SNR_{eff eq \hat{h}} = \frac{\alpha \sigma_t^2 d_{min}^2}{4\sigma^2}\left(1 + \frac{\alpha T_p L}{(1-\alpha)(T-T_p)(T_p - L + 1)}\right)^{-1} \qquad (18)$$

The problem (16) is equivalent to

$$\begin{cases} \max \ SNR_{eff,eq,\hat{h}}\left(T_p, \alpha\right) \\ \text{s.t.} \\ r_0 \leq T_p \leq T \\ 0 < \alpha < 1 \end{cases} \qquad (19)$$

Proposition 1 The optimal training sequence length and the optimal pilot symbol power maximizing the effective SNR under the constraints of (16) are given by

$$\begin{aligned} T_p^* &= (r_3^* - r_0)^+ + r_0 \\ \sigma_p^{*2} &= \frac{\left(1 - \alpha^*(T_p^*)\right)\sigma_t^2 T}{T_p^*} \end{aligned} \qquad (20)$$

where $r_3^* = \arg\max_{x \in \{\lfloor x^* \rfloor, \lceil x^* \rceil\}} f_1(x, \alpha^*(x))$, $f_1(x, \alpha) = \frac{\alpha \sigma_t^2 d_{min}^2}{4\sigma^2}\left(1 + \frac{\alpha x L}{(1-\alpha)(T-x)(x-L+1)}\right)^{-1}$, $x^* = \sqrt{TL} - T$, $\alpha^*(x) = \frac{A(x) - \sqrt{A(x)Lx}}{A(x) - Lx}$ and $A(x) = (T-x)(x-L+1)$.
The proof of Proposition 1 is given in the Appendix.

The power of data symbols maximizing $SNR_{eff,eq,\hat{h}}$ is then given by

$$\sigma_d^{*2} = \frac{\alpha^*(T_p^*)\sigma_t^2 T}{T - T_p^*} \qquad (21)$$

The optimum value of the effective signal-to-noise can be approximated by

$$SNR_{eff,eq,\hat{h}}^* \simeq \frac{d_{min}^2 \sigma_t^2}{4\sigma^2} \frac{A\left(T_p^*\right)\left(1 - \alpha^*\left(T_p^*\right)\right)^2}{LT_p^*} \qquad (22)$$

Simulation results
Figure 2 shows the BER at the output of the MAP equalizer with respect to SNR_{eff} for $Channel 3$, $T = 256$ and $\sigma_t^2 = 1$. We consider the scenarios given in Table I. According to (20), the theoretical values of the optimal length of the training sequence and the optimal pilot symbol power are respectively $T_p^* = 23$ and $\sigma_p^{*2} = 1.18$. Simulations in Figure 2 confirm that the MAP equalizer best performance are achieved when $T_p = T_p^*$ and $\sigma_p^2 = \sigma_p^{*2}$.

Scenario	T_p	σ_p^2
S1	T_p^*	σ_p^{*2}
S2	50	σ_p^{*2}
S3	10	σ_p^{*2}
S4	T_p^*	$0.75\sigma_p^{*2}$
S5	T_p^*	$1.5\sigma_p^{*2}$

TABLE I
Scenarios considered in Figure 2.

B. Maximization of the effective channel capacity

In the following, we define C_{eff}, the effective channel capacity as:

$$C_{eff} = \frac{1}{2}\frac{T - T_p}{T}\log\left(1 + SNR_{eq,\hat{h}}(T_p, \alpha)\right) \qquad (23)$$
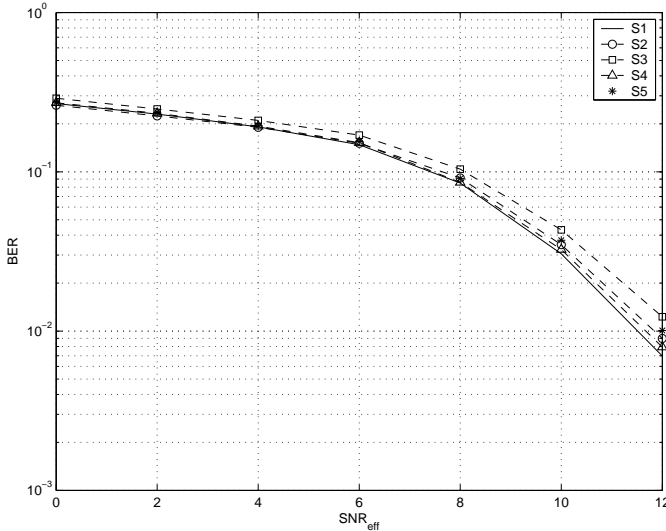
Fig. 2. BER versus SNR_{eff}: MAP equalizer performance for different values of T_p and σ_p^2 for $Channel3$, $T = 256$ and $\sigma_t^2 = 1$.

We want to solve the optimization problem given by:

$$\begin{cases} \max C_{eff}(T_p, \alpha) = \frac{1}{2} \frac{T-T_p}{T} \\ \log\left(1 + \frac{d_{min}^2}{4\sigma^2} \frac{\alpha\sigma_t^2 T}{T-T_p} \frac{(T-T_p)(T_p-L+1)(1-\alpha)}{(T-T_p)(T_p-L+1)(1-\alpha)+L\alpha T_p}\right) \\ \text{s.t.} \\ r_0 \leq T_p \leq T \\ 0 < \alpha < 1 \end{cases} \tag{24}$$

Let $x, \alpha \in \mathbb{R}_+$ and

$$g_1(x,\alpha) = \frac{T-x}{2T} \log\left(1 + \frac{d_{min}^2 \alpha \sigma_t^2 T}{4\sigma^2} \frac{(x-L+1)(1-\alpha)}{(T-x)(x-L+1)(1-\alpha)+L\alpha x}\right).$$

Hence, $C_{eff}(T_p, \alpha) = g_1(T_p, \alpha)$.

Proposition 2 The length of the training sequence and the training power maximizing the effective channel capacity are given by

$$\begin{aligned} T_p^* &= \arg\max_{y \in \{\lfloor y^* \rfloor, \lceil y^* \rceil\}} g_1(y, \alpha^*(y)) \\ \sigma_p^{*2} &= \frac{(1-\alpha^*(T_p^*))\sigma_t^2 T}{T_p^*} \end{aligned} \tag{25}$$

where $\alpha^*(x) = \frac{A(x) - \sqrt{A(x)Lx}}{A(x) - Lx}$, $A(x) = (T-x)(x-L+1)$, and $y^* = \arg\min_{r_0 \leq x \leq T} \left| \frac{\partial g_1}{\partial x}(x, \alpha^*(x)) \right|$.

The proof of Proposition 2 is given in the Appendix.

Simulation results

Figure 3 shows the effective channel capacity as a function of the pilot symbol power for $L = 3$, $T = 256$, $\sigma_t^2 = 0.2$ and $\sigma^2 = 1$ for different values of T_p. By using (25), we have $T_p^* = 13$ and $\sigma_p^{*2} = 0.46$. This is confirmed by the simulations since the effective channel capacity is maximized for these values. Figure 4 shows the effective channel capacity as a function of T_p for $L = 3$, $T = 256$ and $\sigma_t^2 = 1$ for different values of $\rho = 1/\sigma^2$. We notice that the length of the training sequence maximizing the effective channel capacity grows when the noise variance at the input of the MAP equalizer
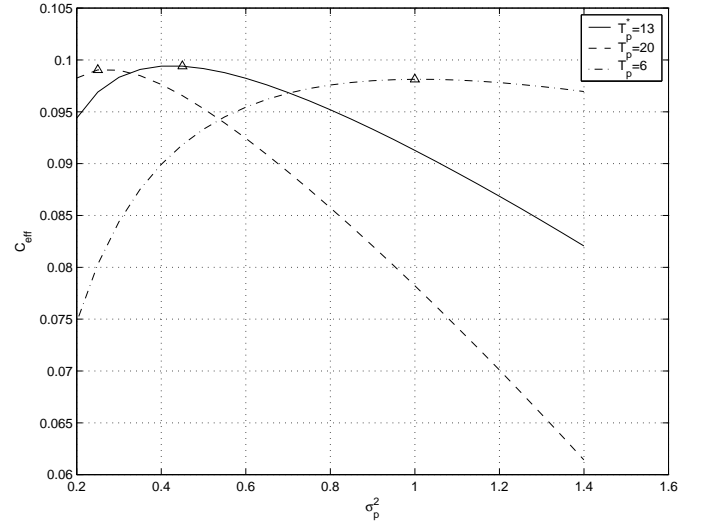

Fig. 3. Effective channel capacity versus σ_p^2 for different values of the length of the training sequence, $L = 3$, $T = 256$, $\sigma_t^2 = 0.2$ and $\sigma^2 = 1$.
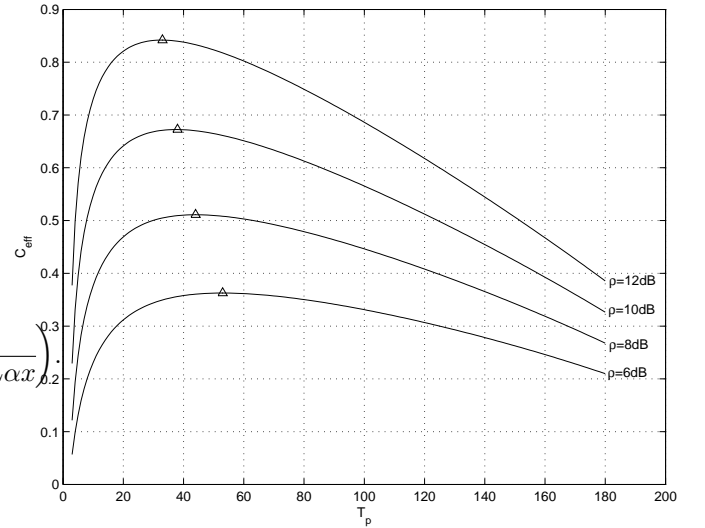

Fig. 4. Effective channel capacity versus T_p for $L = 3$, $T = 256$ and $\sigma_t^2 = 1$ for different values of ρ.

increases. Table II gives the values of T_p^* and σ_p^{*2} obtained by solving (24) for different values of σ_t^2. We suppose that the total length of the burst is $T = 256$, the channel memory length $L = 3$ and $\rho = 8dB$. When σ_t^2 increases, T_p^* decreases and σ_p^{*2} increases.

V. CONCLUSION

In this paper, we consider the problem of optimizing the training sequence length when a MAP equalizer is used. We study two cases: the case where the training and data powers are equal and the case where they are allowed to be different. We define an effective signal-to-noise ratio and an effective channel capacity for the training-based transmission scheme. We find the training sequence lengths maximizing these quantities in the case of equal powers. When the powers

σ_t^2	T_p^*	σ_p^{*2}
0.1	15	0.18
0.2	13	0.42
0.5	10	1.42
1	9	3.15
2	7	8.38
5	6	25.30

TABLE II
T_p^* AND σ_p^{*2} MAXIMIZING C_{eff} VERSUS σ_t^2 FOR $T = 256$, $L = 3$ AND $\rho = 8dB$.

are allowed to be different, we also give the optimal power allocation.

VI. APPENDIX

A. Proof of Proposition 1

Let $\{x, \alpha\} \in [L, +\infty[\times\mathbb{R}_+^*$ and

$$f_1(x, \alpha) = \frac{\alpha\sigma_t^2 d_{min}^2}{4\sigma^2}\left(1+\frac{\alpha x L}{(1-\alpha)(T-x)(x-L+1)}\right)^{-1}.$$

Thus, $SNR_{eff,eq,\hat{h}} = f_1(T_p, \alpha)$.
Now, consider the following relaxed optimization problem:

$$\begin{cases} \max f_1(x, \alpha) \\ \text{s.t.} \\ r_0 \leq x \leq T \\ 0 < \alpha < 1 \end{cases} \qquad (26)$$

Since, $\frac{\partial^2 f_1}{\partial \alpha^2} < 0$, f_1 is concave with respect to α, for a fixed value of x. Thus, when x is fixed, f_1 has a unique maximum achieved for $\alpha = \alpha^*(x)$ such as

$$\alpha^*(x) = \frac{A(x) - \sqrt{A(x)Lx}}{A(x) - Lx} \qquad (27)$$

where $A(x) = (T-x)(x-L+1)$. Notice that $0 < \alpha^*(x) < 1$. Hence, by using the combinatorial optimization [11], the solution of (26) is:

$$(T_p^*, \alpha^*) = \left(\arg\max_{r_0 \leq m \leq T} f_1(m, \alpha^*(m)), \alpha^*(T_p^*)\right) \qquad (28)$$

where m is an integer.
Let $x \in [L, +\infty[$ and

$$F_1(x) = f_1(x, \alpha^*(x)) = \frac{\sigma_t^2 d_{min}^2}{4\sigma^2}\frac{A(x)}{\sqrt{A(x)Lx} + Lx} \qquad (29)$$

As $F_1''(x) < 0$, for $x \in [L, +\infty[$, F_1 is concave. Thus, F_1 has a unique maximum reached for $x = x^* \in [L, +\infty[$. By calculating the derivative of F_1, we can show that it is maximized for $x^* = \sqrt{TL} - T$. Suppose that $T \geq \frac{r_0^2}{L-1}$, then $x^* \geq r_0$.
Let (T_1, α_1) be the solution of (19). Suppose that $T_1 \notin \{\lfloor x^* \rfloor, \lceil x^* \rceil\}$. Hence, there are the two possibilities: $T_1 < \lfloor x^* \rfloor = x_1$ or $T_1 > \lceil x^* \rceil = x_2$.
As $\frac{\partial^2 f_1}{\partial x^2}(x, \alpha) < 0$, f_1 is concave with respect to x for a fixed value of α.

- If $T_1 < x_1 < x_2$, we have $f_1(T_1, \alpha_1) \geq f_1(x_1, \alpha_1) \geq f_1(x_2, \alpha_1)$. Hence, $\frac{f_1(x_2,\alpha_1)-f_1(T_1,\alpha_1)}{x_2-T_1} < \frac{f_1(x_2,\alpha_1)-f_1(x_1,\alpha_1)}{x_2-x_1}$. On the other hand, $f_1(x, \alpha)$ is concave with respect to x for a fixed value of α. Thus, $\frac{f_1(x_2,\alpha_1)-f_1(T_1,\alpha_1)}{x_2-T_1} \geq \frac{f_1(x_2,\alpha_1)-f_1(x_1,\alpha_1)}{x_2-x_1}$ which is impossible.
- If $T_1 > x_2 > x^*$, $f_1(x^*, \alpha^*) \geq f_1(x_2, \alpha^*) \geq f_1(T_1, \alpha^*)$. Hence, $\frac{f_1(T_1,\alpha^*)-f_1(x^*,\alpha^*)}{T_1-x^*} < \frac{f_1(T_1,\alpha^*)-f_1(x_2,\alpha^*)}{T_1-x_2}$. On the other hand, $f_1(x, \alpha)$ is concave in x for a given value of α. Thus, $\frac{f_1(T_1,\alpha^*)-f_1(x^*,\alpha^*)}{T_1-x^*} \geq \frac{f_1(T_1,\alpha^*)-f_1(x_2,\alpha^*)}{T_1-x_2}$ which is impossible.

Thus, the training sequence length maximizing the effective signal-to-noise ratio is

$$r_3^* = \arg\max_{x\in\{\lfloor x^*\rfloor, \lceil x^*\rceil\}} f_1(x, \alpha^*(x)).$$

If $T < \frac{r_0^2}{L-1}$, then $x^* < r_0$. As F_1 is concave, $T_p^* = r_0$.
In conclusion,

$$T_p^* = (r_3^* - r_0)^+ + r_0. \qquad (30)$$

B. Proof of Proposition 2

The effective channel capacity can be written as $C_{eff}(T_p, \alpha) = g_1(T_p, \alpha)$ where $g_1(x, \alpha) = \frac{1}{2}\frac{T-x}{T}\log\left(1 + \frac{d_{min}^2}{4\sigma^2}\frac{\alpha\sigma_t^2 T}{T-x}\frac{(T-x)(x-L+1)(1-\alpha)}{(T-x)(x-L+1)(1-\alpha)+L\alpha x}\right)$, for $\{x, \alpha\} \in [r_0, +\infty[\times\mathbb{R}_+^*$.
Since, $\frac{\partial^2 g_1}{\partial \alpha^2} < 0$, g_1 is concave with respect to α for a fixed value of x. Thus, when x is fixed, g_1 has a unique maximum reached for $\alpha = \alpha^*(x)$ such as

$$\alpha^*(x) = \frac{A(x) - \sqrt{A(x)Lx}}{A(x) - Lx} \qquad (31)$$

where $A(x) = (T-x)(x-L+1)$. Note that $0 < \alpha^*(x) < 1$. Hence, the solution of (24) is [11]:

$$(T_p^*, \alpha^*) = \left(\arg\max_{r_0 \leq m \leq T} g_1(m, \alpha^*(m)), \alpha^*(T_p^*)\right) \qquad (32)$$

where m is an integer.
Let $x \in [r_0, +\infty[$ and

$$\begin{aligned} G_1(x) &= g_1(x, \alpha^*(x)) \\ &= (T-x)\log\left(1 + \frac{\sigma_t^2 T d_{min}^2}{4\sigma^2}\frac{x-L+1}{\sqrt{A(x)Lx}+Lx}\right). \end{aligned} \qquad (33)$$

As $G_1''(x) < 0$, for $x \in [r_0, +\infty[$, G_1 is concave. Besides, $G_1'(T) < 0$. Thus, we consider the two possible cases:

- If $G_1'(r_0) \leq 0 \Leftrightarrow |G_1'(r_0)| = \min_{r_0 \leq x \leq T}|G_1'(x)| \Leftrightarrow \left|\frac{\partial g_1}{\partial x}(r_0, \alpha^*(r_0))\right| = \min_{r_0 \leq x \leq T}\left|\frac{\partial g_1}{\partial x}(x, \alpha^*(x))\right|$, then $T_p^* = r_0$ and $\alpha^* = \alpha^*(r_0)$
- If $G_1'(r_0) > 0$, then according to the theorem of intermediate values, there exists a unique $y^* \in]r_0, T[$ such as $\frac{\partial g_1}{\partial x}(y^*, \alpha^*(y^*)) = 0$. Hence, $T_p^* = \arg\max_{y\in\{\lfloor y^*\rfloor, \lceil y^*\rceil\}} G_1(y)$ and $\alpha^* = \alpha^*(T_p^*)$.

REFERENCES

[1] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-32, pp. 1–100, March 1974.

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

[3] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequencey-selective channels in training-based transmission schemes," *IEEE Trans. on Signal Processing*, vol. 52, no. 9, pp. 2572–2583, September 2004.

[4] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. on Signal Processing*, vol. 51, no. 6, pp. 1615–1624, June 2003.

[5] T.-L. Tung, K. Yao, and R.E. Hudson, "Channel estimation and adaptive power allocation for performance and capacity improvement of multiple-antenna ofdm systems," *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 01)*, , no. 2, March 2001.

[6] T.F. Wong and B. Park, "Training sequence optimization in MIMO systems with colored interference," *IEEE Trans. on communications*, vol. 52, no. 11, pp. 1939–1947, November 2004.

[7] S. Buzzi, M. Lops, and S. Sardellitti, "Performance of iterative data detection and channel estimation for single-antenna and multiple-antennas wireless communications," *IEEE Trans. on Vehicular Technology*, vol. 53, no. 4, pp. 1085–1104, July 2004.

[8] S. Crozier, D. Falconer, and S. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Proceedings*, vol. 138, pp. 371–378, August 1991.

[9] N. Sellami, A. Roumy, and I. Fijalkow, "The impact of both *a priori* information and channel estimation errors on the MAP equalizer performance," *IEEE Trans. on Signal Processing*, vol. 54, no. 7, pp. 2716–2724, July 2006.

[10] G.D. Forney, "Maximum-likelihood sequence estimation for digital sequences in the presence of intersymbol interference," *IEEE Trans. on Inf. Theory*, vol. 18, pp. 363–378, May 1972.

[11] H.P. Christos and K. Steiglitz, *Combinatorial optimization: algorithms and complexity*, Prentice Hall, New Jersey, USA, 1982.