Joint training interval length and power allocation optimization for MIMO flat fading channels

Imed Hadj-Kacem[1,2], Noura Sellami[1], Inbar Fijalkow[2], and Aline Roumy[3]
[1] Laboratoire LETI, ENIS, Route Sokra km 3.5, B.P 3038 Sfax, Tunisia
[2] ETIS, UMR 8051, CNRS, ENSEA, Univ Cergy-Pontoise, F-95000 Cergy, France
[3] INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

***Abstract*—In this paper, we jointly optimize the training interval length and the power allocation for MIMO (Multiple-Input Multiple-Output) flat fading channels when a Maximum *A Posteriori* (MAP) detector is used at the receiver and the training and data powers are allowed to vary. We calculate the equivalent Signal-to-Noise Ratio (SNR) at the output of the MAP detector. Based on this expression, we define an effective SNR taking into account the data throughput loss due to the use of pilot symbols. We find that the optimal length of the training interval maximizing this quantity is equal to the number of transmit antennas. When the values of the pilot and data powers are not allowed to be different, we give the optimal training interval length maximizing the effective SNR and we show that it can be larger than the number of transmit antennas.**

I. Introduction

Multiple-Input Multiple-Output (MIMO) systems provide significant increase in capacity especially when the channel is known [1]. In practice, the channel estimation procedure is performed by transmitting training symbols that are known at the receiver. When the length of the training interval increases, the channel estimation becomes more reliable. However, this leads to a loss in terms of data throughput. Thus, a trade-off has to be found. Several methods have been proposed to design the optimal training interval length. The solution presented in [2] is based on maximizing a lower bound of the capacity of the training-based scheme for a MIMO flat fading channel. Another approach proposed to find the optimal training interval length that minimizes the Mean Square Error (MSE) of the channel estimator [3] and therefore does not take into account the equalizer performance. In [4], we considered a transmission over a single-input single-output (SISO) frequency selective channel where a Maximum *A Posteriori* (MAP) equalizer is used. We proposed to maximize an effective Signal-to-Noise Ratio (SNR) computed at the output of the MAP equalizer and that takes into account the loss in terms of data throughput due to the use of the pilot symbols. In this paper, we propose to generalize this study to MIMO flat fading channels. At the receiver, we consider a Maximum *A Posteriori* (MAP) detector. The channel is estimated by the least square estimator [5]. We derive the expression of the equivalent SNR at the output of the MAP detector when the training and data powers are allowed to be different. Based on this expression, we define an effective SNR taking into account the loss in terms of data throughput due to the use of the training symbols. We propose to jointly optimize the length of the training interval and the power allocation by maximizing this effective SNR. We show that the optimal training interval length is equal to its minimum value n_T, where n_T is the number of the transmit antennas. Notice that a similar result was found in [2] by maximizing a lower bound on the training-based channel capacity.

The paper is organized as follows. In Section II, we describe the transmission system model. In Section III, we give the expression of the equivalent SNR at the output of the MAP detector. Section IV studies the joint optimization of the training interval length and the power allocation. Section V gives simulation results.

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. The quantities $\lfloor t \rfloor$, $\lceil t \rceil$ and $|t|$ are respectively the greatest integer lower than t, the smallest integer greater than t and the absolute value of t.

II. Transmission system model

We consider a MIMO system composed of n_T transmit antennas and n_R receive antennas. The input data information bit sequence is mapped to the symbol alphabet $\mathcal{A}$. For simplicity, we consider the BPSK modulation ($\mathcal{A} = \{-1, 1\}$). We assume that transmissions are organized into bursts of T symbols and that the first T_p ones are pilot symbols. The channel is supposed to be invariant during one burst and to change independently from burst to burst. We also assume that the training and data powers are allowed to be different. The received baseband signal sampled at the symbol rate at time k at the receive antenna p is

$$y_k^{(p)} = \begin{cases} \sqrt{\sigma_p^2} \sum_{i=1}^{n_T} h_{pi} x_k^{(i)} + n_k^{(p)} & \text{for } 0 \leq k \leq T_p - 1 \\ \sqrt{\sigma_d^2} \sum_{i=1}^{n_T} h_{pi} x_k^{(i)} + n_k^{(p)} & \text{for } T_p \leq k \leq T - 1 \end{cases} \quad (1)$$

where $x_k^{(i)}$ is the k^{th} symbol transmitted by the i^{th} transmit antenna, σ_p^2 and σ_d^2 are respectively the powers of the pilot symbols and the data symbols and h_{ji} is the channel tap gain between the j^{th} transmit antenna and the i^{th} receive antenna. The channel tap gains h_{ji} are modeled as independent zero mean complex Gaussian variables. We assume that for a given receive antenna p, $E\left(\sum_{i=1}^{n_T} |h_{pi}|^2\right) = 1$ where $E(.)$

is the mathematical expectation. In (1), $n_k^{(p)}$ are modeled as independent samples from a random variable with normal probability density function (pdf) $\mathcal{N}(0,\sigma^2)$ where $\mathcal{N}(\alpha,\sigma^2)$ denotes a Gaussian distribution with mean α and variance σ^2. The received vector at the p^{th} receive antenna during the training phase, $\underline{y}^{(p)}=\left(y_0^{(p)},y_1^{(p)},...,y_{T_p-1}^{(p)}\right)^T$, is given by

$$\underline{y}^{(p)} = X\underline{h}^{(p)} + \underline{n}^{(p)} \tag{2}$$

where $\underline{h}^{(p)} = (h_{p1},h_{p2},...,h_{pn_T})^T$, $\underline{n}^{(p)} = \left(n_0^{(p)},n_1^{(p)},...,n_{T_p-1}^{(p)}\right)^T$ and

$$X=\sqrt{\sigma_p^2}\begin{pmatrix} x_0^{(1)} & ... & x_0^{(n_T)} \\ x_1^{(1)} & & x_1^{(n_T)} \\ . & & . \\ . & & . \\ . & & . \\ x_{T_p-1}^{(1)} & & x_{T_p-1}^{(n_T)} \end{pmatrix}.$$

Under the assumption that enough pilot symbols have been transmitted ($T_p \geq n_T$), the least square channel estimate $\underline{\widehat{h}}^{(p)}=\left(\widehat{h}_{p1},...,\widehat{h}_{pn_T}\right)^T$ is given by [5]

$$\underline{\widehat{h}}^{(p)} = \left(X^TX\right)^{-1}X^T\underline{y}^{(p)}. \tag{3}$$

We assume that the training sequences have ideal autocorrelation and crosscorrelation properties which means that $X^TX=\sigma_p^2T_pI_{n_T}$. Hence, we obtain

$$\begin{aligned}\delta\underline{h}^{(p)} = \underline{\widehat{h}}^{(p)} - \underline{h}^{(p)} &\sim \mathcal{N}\left(0,\sigma^2(X^TX)^{-1}\right)\\ &= \mathcal{N}\left(0,\tfrac{\sigma^2}{T_p\sigma_p^2}I_{n_T}\right)\end{aligned} \tag{4}$$

III. Equivalent SNR at the output of the MAP detector

We consider a MAP detector at the receiver. From (1), the received signal $\underline{y}_k=(y_k^{(1)},y_k^{(2)},...,y_k^{(n_R)})^T$ at the n_R receive antennas at time k, for $T_p \leq k \leq T-1$, is given by

$$\underline{y}_k = \sqrt{\sigma_d^2}H\underline{x}_k + \underline{n}_k, \tag{5}$$

where H is the channel matrix, $\underline{x}_k=(x_k^{(1)},x_k^{(2)},...,x_k^{(n_T)})^T$ and $\underline{n}_k=(n_k^{(1)},n_k^{(2)},...,n_k^{(n_R)})^T$. In order to detect the transmitted symbols, the MAP detector calculates the probabilities $p(x_k^{(i)}=x|\underline{y}_k)$ for $1\leq i\leq n_T$ where $x\in\{-1,1\}$. According to the Bayes formula

$$p(x_k^{(i)}=x|\underline{y}_k) = \frac{p(\underline{y}_k|x_k^{(i)}=x)p(x_k^{(i)}=x)}{p(\underline{y}_k)} \tag{6}$$

Since the transmitted symbols are equiprobable, the MAP detector has only to calculate the probability $p(\underline{y}_k|x_k^{(i)}=x)$ which is given by

$$p(\underline{y}_k|x_k^{(i)}=x) = \sum_{\underline{x}_k\in X_1} p(\underline{y}_k|\underline{x}_k) \tag{7}$$

where X_1 is the set of all values that can be taken by $\underline{x}_k$ such that $x_k^{(i)}=x$. According to (5), the probability $p(\underline{y}_k|\underline{x}_k)$ is

$$p(\underline{y}_k|\underline{x}_k) = \frac{1}{(\pi\sigma^2)^{n_R}}\exp\left(\frac{-\|\underline{y}_k-\sqrt{\sigma_d^2}\widehat{H}\underline{x}_k\|^2}{\sigma^2}\right) \tag{8}$$

where $\widehat{H}$ is the estimated version of H.

Proposition 1: When the channel is estimated by the least square estimator and pilot and data symbols have different power levels, respectively σ_p^2 and σ_d^2, the equivalent signal-to-noise ratio at the output of the MAP detector is given by

$$SNR_{eq} = \frac{\sigma_d^2}{\sigma^2}\left(1+\frac{n_T\sigma_d^2}{T_p\sigma_p^2}\right)^{-1} \tag{9}$$

where σ^2, n_T and T_p are respectively the noise variance, the number of transmit antennas and the training interval length. The proof of Proposition 1 is given in the Appendix.

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

Increasing the training interval length leads to an improvement of the channel estimate quality but also to a loss in terms of data throughput. Thus, in order to take this loss into account, we define as in [4] an effective SNR at the output of the MAP detector as

$$\begin{aligned}SNR_{eff,eq} &= \tfrac{T-T_p}{T}SNR_{eq}\\ &= \tfrac{T-T_p}{T}\tfrac{\sigma_d^2}{\sigma^2}\left(1+\tfrac{n_T\sigma_d^2}{T_p\sigma_p^2}\right)^{-1}\end{aligned} \tag{10}$$

Our goal is to maximize $SNR_{eff,eq}$ under a sum energy constraint and a total blocklength constraint. Hence, we define the following optimization problem

$$\begin{cases} \max\ SNR_{eff,eq}\left(T_p,\sigma_p^2,T_d,\sigma_d^2\right)\\ \text{s.t.}\\ \sigma_p^2T_p+\sigma_d^2T_d=\sigma_t^2T\\ T_p+T_d=T\\ \sigma_p^2,\sigma_d^2\geq 0\\ n_T\leq T_p\leq T-1\end{cases} \tag{11}$$

where T_d is the length of data symbols block and σ_t^2T is the total transmit energy per block.

We denote the fraction of the total transmit energy used in the data transmission phase as

$$\sigma_d^2T_d = \alpha\sigma_t^2T, \quad 0<\alpha<1 \tag{12}$$

The effective SNR can then be written as

$$SNR_{eff,eq} = \frac{\sigma_t^2(T-T_p)\alpha(1-\alpha)}{(1-\alpha)(T-T_p)+n_T\alpha} \tag{13}$$

Interestingly, the effective SNR depends only on the energy ratio α defined in (12) and on the training interval length, for fixed σ_t^2, T and n_T. Hence, the problem (11) is equivalent to

$$\begin{cases} \max SNR_{eff,eq}(T_p,\alpha)\\ \text{s.t.}\\ n_T\leq T_p\leq T-1,\ 0<\alpha<1\end{cases} \tag{14}$$

Even more interestingly, the constraints are now independent in the sense that each constraint function depends on α or T_p [6, p133]. This will allow the simplification of the resolution of the optimization problem as stated in the following Proposition.

Proposition 2: When $T \neq 2n_T$, the optimal training interval length and the optimal pilot symbol power maximizing the effective SNR under the constraints of (11) are given by

$$\begin{aligned} T_p^* &= n_T \\ \sigma_p^{*^2} &= \frac{\sigma_t^2 T\left(-n_T+\sqrt{n_T(T-n_T)}\right)}{n_T(T-2n_T)} \end{aligned} \tag{15}$$

When $T = 2n_T$, the solution of (11) is

$$\begin{aligned} T_p^* &= n_T \\ \sigma_p^{*^2} &= \frac{\sigma_t^2 T}{2n_T} \end{aligned} \tag{16}$$

The proof of Proposition 2 is omitted for the sake of space. The power of data symbols maximizing the effective SNR is then given by

$$\sigma_d^{*^2} = \frac{\sigma_t^2 T - \sigma_p^{*^2} n_T}{T - n_T} \tag{17}$$

Note that $\sigma_p^{*^2} = \sigma_d^{*^2} = \sigma_t^2$ when $T = 2n_T$.

Remark 1: For a given value of T_p ($T_p \neq T - n_T$), $SNR_{eff,eq}$ is maximized for $\sigma_p^2 = \sigma_p^{*^2}(T_p)$ given by

$$\sigma_p^{*^2}(T_p) = \frac{\sigma_t^2 T\left(-n_T + \sqrt{n_T(T-T_p)}\right)}{T_p(T - T_p - n_T)} \tag{18}$$

when $T_p = T - n_T$, $\sigma_p^{*^2}(T_p) = \frac{\sigma_t^2 T}{2T_p}$.

Remark 2: When the values of the pilot and data powers are not allowed to be different and then are not considered in the optimization problem, the training interval length maximizing the effective SNR may be larger than n_T and is given by

$$T_p^* = (r^* - n_T)^+ + n_T \tag{19}$$

where $(u)^+ = \frac{|u|+u}{2}$, $r^* = \arg\max_{t\in\{\lfloor t^*\rfloor, \lceil t^*\rceil\}} f(t)$, $f(t) = \frac{T-t}{T\sigma^2}\left(1 + \frac{n_T}{t}\right)^{-1}$ and $t^* = -n_T + \sqrt{n_T^2 + n_T T}$.
Notice that for $T > 3n_T + 4$, $T_p^* > n_T$.

V. Simulation results

In this section, we propose to validate our analytical results by simulations. Figure 1 shows the Bit Error Rate (BER) at the output of the MAP detector 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 detector for $T = 256$, $\sigma_t^2 = 4dB$, $n_T = n_R = 2$. The channel tap gains h_{ji} are modeled as independent zero mean complex Gaussian variables with variance 0.5. According to (15), $T_p^* = n_T = 2$ and $\sigma_p^{*^2} = 14.18dB$. We consider four scenarios given in Table I where $\sigma_p^{*^2}(T_p)$ is the value of the pilot power maximizing the effective SNR for a given value of T_p (see Remark 1). Simulations in Figure 1 confirm that the MAP detector best performance are achieved when T_p is equal to its minimum value $n_T = 2$ and $\sigma_p^2 = \sigma_p^{*^2} = 14.18dB$.

Scenario	T_p	σ_p^2
S1	$T_p^* = 2$	$\sigma_p^{*^2} = 14.18dB$
S2	8	$\sigma_p^{*^2}(8)$
S3	16	$\sigma_p^{*^2}(16)$
S4	32	$\sigma_p^{*^2}(32)$

TABLE I
Scenarios considered in Figure 1.

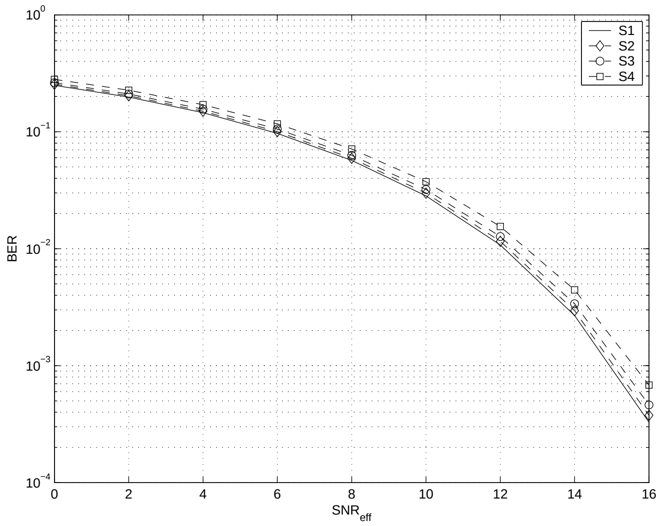

Fig. 1. BER versus SNR_{eff} at the output of the MAP detector for $T = 256$, $\sigma_t^2 = 4dB$ and $n_T = n_R = 2$ (joint optimization of the training interval length and the power allocation).

Figure 2 shows the optimal training and data powers maximizing the effective SNR with respect to T, the burst length, for $n_T = 5$ and $\sigma_t^2 = 6dB$. Notice that $\sigma_p^{*^2}$ is given by (15) when $T \neq 10$ and by (16) when $T = 10$. The expression of $\sigma_d^{*^2}$ is given by (17). The solid lines are obtained using our study. The dotted ones are obtained using the results of [2]. We notice that the criteria of the maximization of the effective SNR gives the same results as the one based on maximizing a lower bound of the capacity [2]. We verify that when T increases, $\sigma_p^{*^2}$ increases as well and when $T = 2n_T$, $\sigma_p^{*^2} = \sigma_d^{*^2} = \sigma_t^2$. This result can be proved by using (17).

Now, we consider the case where the pilot and data powers are not allowed to be different, $\sigma_p^2 = \sigma_d^2 = \sigma_t^2 = 1$ (see Remark 2). We plot in Figure 3 the BER at the output of the MAP detector with respect to SNR_{eff} for $T = 256$ and for different values of T_p. From (19), $T_p^* = 21$. This confirms that when the pilot and data powers are not considered in the optimization problem, the training sequence interval maximizing the effective SNR may be larger than n_T.

VI. Conclusion

In this paper, we considered the joint optimization problem of the training interval length and power allocation for MIMO flat fading channels when a MAP detector is used at the receiver. We defined an effective SNR at the output of the

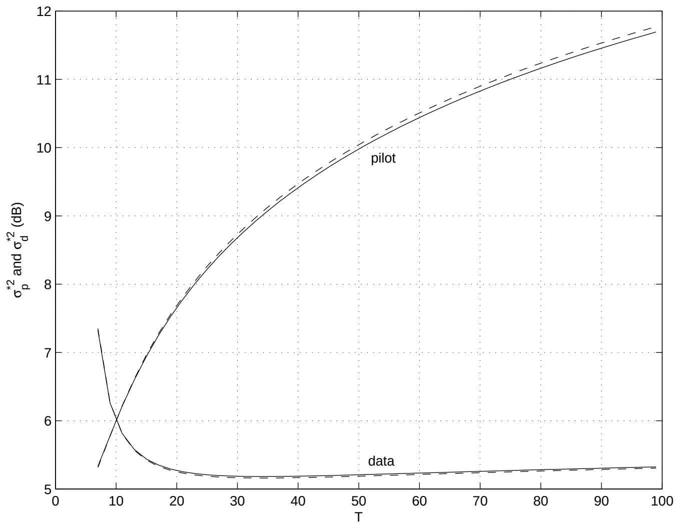

Fig. 2. σ_p^{*2} and σ_d^{*2} with respect to T for $n_T = 5$ and $\sigma_t^2 = 6dB$.

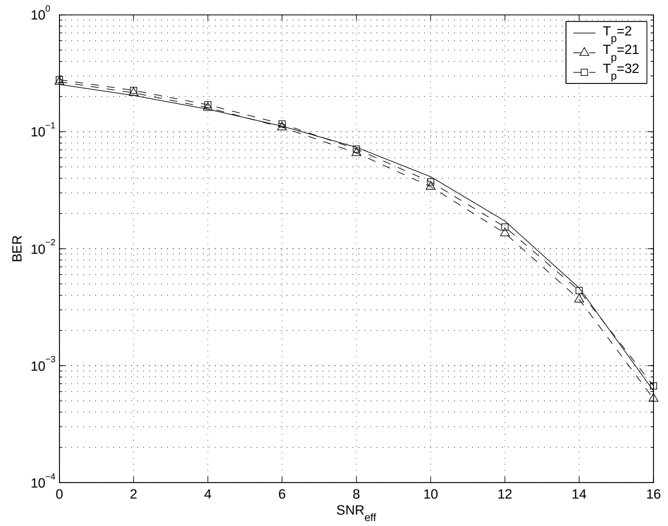

Fig. 3. BER versus SNR_{eff} at the output of the MAP detector for $T = 256$, $\sigma_p^2 = \sigma_d^2 = \sigma_t^2 = 1$ and $n_T = n_R = 2$ (optimization of the training interval length for equal powers).

MAP detector. We proved that the optimal training interval length maximizing the effective signal-to-noise ratio is equal to the number of transmit antennas n_T and we gave the optimal power allocation.

VII. Appendix: Proof of Proposition 1

The received signal at the p^{th} receive antenna corresponding to the data transmission phase, $\underline{y}^{(p)} = \left(y_{T_p}^{(p)}, ..., y_{T-1}^{(p)}\right)^T$, is given by

$$\underline{y}^{(p)} = \sqrt{\sigma_d^2} H(\underline{h}^{(p)})\underline{x} + \underline{n}^{(p)} \tag{20}$$

where T is the burst length, T_p is the length of the training interval, $\underline{x} = \left(x_{T_p}^{(1)}, ..., x_{T_p}^{(n_T)}, ..., x_{T-1}^{(1)}, ..., x_{T-1}^{(n_T)}\right)^T$ is the vector of the transmitted symbols during the data phase, n_T is the number of the transmitted antennas, σ_d^2 is the power of the data symbols, $\underline{h}^{(p)} = (h_{p1}, h_{p2}, ..., h_{pn_T})^T$ is the flat fading channel between the n_T transmitted antennas and the receive antenna p and $H(\underline{h}^{(p)}) =$

$$\begin{pmatrix} h_{p1} & ... & h_{pn_T} & 0 & & ... & & 0 \\ 0 & h_{p1} & ... & h_{pn_T} & 0 & & ... & 0 \\ . & . & . & & . & . & & . \\ . & & . & . & & . & . & . \\ . & & & . & . & & . & . \quad . \\ 0 & & ... & & 0 & h_{p1} & ... & h_{pn_T} \quad 0 \\ 0 & & & ... & & 0 & h_{p1} \quad ... & h_{pn_T} \end{pmatrix}$$

The data estimate according to the MAP criterion is given by

$$\underline{\widehat{x}} = \arg\min_{\underline{u}} \left(\left\| \underline{y}^{(p)} - \sqrt{\sigma_d^2} H(\underline{\widehat{h}}^{(p)})\underline{u} \right\|^2 : \underline{u} \in \mathcal{A}^{T+L-1} \right) \tag{21}$$

where $\underline{\widehat{h}}^{(p)} = \left(\widehat{h}_{p1}, \widehat{h}_{p2}, ..., \widehat{h}_{pn_T}\right)^T$ is the estimate of $\underline{h}^{(p)}$. For the sake of conciseness, the exponent p is omitted from $\underline{y}^{(p)}$, $\underline{n}^{(p)}$, $\underline{h}^{(p)}$ and $\underline{\widehat{h}}^{(p)}$.

Let $\underline{x}^{(q)} = \left(x_{T_p}^{(q)}, ..., x_{T-1}^{(q)}\right)^T$ be the subvector of $\underline{x}$ transmitted by the q^{th} antenna and $\underline{\widehat{x}}^{(q)} = \left(\widehat{x}_{T_p}^{(q)}, ..., \widehat{x}_{T-1}^{(q)}\right)^T$ its estimate. In the following, we consider an error event characterized by its length m_q. Hence, we suppose that there exists an interval of size m_q such that all the symbols of $\underline{\widehat{x}}^{(q)}$ are different from the corresponding symbols of $\underline{x}^{(q)}$ while the preceding and the following symbols are the same for $\underline{\widehat{x}}^{(q)}$ and $\underline{x}^{(q)}$. Let $\underline{x}_{m_q}^{(q)}$ and $\underline{\widehat{x}}_{m_q}^{(q)}$ be respectively the vector of transmitted symbols and estimated ones corresponding to the error interval.

Now, let $\underline{\widehat{x}}_m = \left(\underline{\hat{x}}_{m_1}^{(1)}, ..., \underline{\hat{x}}_{m_{n_T}}^{(n_T)}\right)^T$ and $\underline{x}_m = \left(\underline{x}_{m_1}^{(1)}, ..., \underline{x}_{m_{n_T}}^{(n_T)}\right)^T$. A subevent E_m of the error event of length m is that $\underline{\widehat{x}}_m$ is better than $\underline{x}_m$ in the sense of the MAP criterion. Hence

$$E_m : \left\| \underline{y}_m - \sqrt{\sigma_d^2} H_m(\underline{\widehat{h}})\underline{\widehat{x}}_m \right\|^2 \leq \left\| \underline{y}_m - \sqrt{\sigma_d^2} H_m(\underline{\widehat{h}})\underline{x}_m \right\|^2 \tag{22}$$

where $m = \sum_{q=1}^{n_T} m_q$, $\underline{y}_m$ and $H_m(\underline{\widehat{h}})$ are the subvector of $\underline{y}$ and the estimated channel matrix corresponding to the error intervals.

Let $\underline{e}_m = \underline{\widehat{x}}_m - \underline{x}_m$ be the vector of errors. The event (22) is equivalent to

$$\sigma_d^2 \left\| H_m(\underline{\widehat{h}})\underline{e}_m \right\|^2 \leq 2 \left(\underline{e}_m^T H_m(\underline{\widehat{h}})^T \left(\underline{y}_m - \sqrt{\sigma_d^2} H_m(\underline{\widehat{h}})\underline{x}_m \right) \right) \tag{23}$$

Let $H_m(\Delta\underline{h}) = H_m(\underline{\widehat{h}}) - H_m(\underline{h})$. The expression (23) is then equivalent to

$$\sigma_d^2 \| H_m(\underline{\widehat{h}})\underline{e}_m \|^2 \leq 2 \left(\underline{e}_m^T H_m(\underline{\widehat{h}})^T \underline{n}_m - \sqrt{\sigma_d^2} \underline{e}_m^T H_m(\underline{\widehat{h}})^T H_m(\Delta\underline{h})\underline{x}_m \right) \tag{24}$$

where $\underline{n}_m$ is the subvector of $\underline{n}$ corresponding to the error event.

Using the assumptions given in [7], we obtain that

$\|H_m(\widehat{\underline{h}})\underline{e}_m\| \to \|\underline{\varepsilon}_m\|(1+\xi_{T_0})$ where ξ_{T_0} tends in probability to 0. Hence, $\|H_m(\widehat{\underline{h}})\underline{e}_m\| \to \|\underline{\varepsilon}_m\|$.
Thus, we obtain

$$\sigma_d^2\|\underline{\varepsilon}_m\|^2 \leq 2\left(\underline{\varepsilon}_m^T\underline{n}_m - \sqrt{\sigma_d^2}\underline{\varepsilon}_m^T H_m(\Delta\underline{h})\underline{x}_m\right) \quad (25)$$

We suppose that $\Delta\underline{h} \sim \mathcal{N}(0, \mathcal{C})$, $\mathcal{C}$ being the covariance matrix of $\Delta\underline{h}$. Defining $\mathcal{C}_m(\underline{x}) = H_m(\underline{x}_m)\mathcal{C}H_m(\underline{x}_m)^T$, $H_m(\underline{x}_m)$ being the Hankel matrix such as $H_m(\underline{x}_m)\Delta\underline{h} = H_m(\Delta\underline{h})\underline{x}_m$, we obtain

$$\|\underline{\varepsilon}_m\|^2 \leq \chi_x \quad (26)$$

where $\chi_x \sim \mathcal{N}(0, \Delta_x)$ with

$$\Delta_x = 4\frac{\|\underline{\varepsilon}_m\|^2\sigma^2}{\sigma_d^2}\left(1 + \frac{\sigma_d^2\underline{\varepsilon}_m^T\mathcal{C}_m(\underline{x})\underline{\varepsilon}_m}{\sigma^2\|\underline{\varepsilon}_m\|^2}\right). \quad (27)$$

Hence, the probability of the error event $P(E_m)$ is given by

$$P(E_m) = Q\left(\frac{\|\underline{\varepsilon}_m\|\sigma_d}{2\sigma}\left(1 + \frac{\sigma_d^2\underline{\varepsilon}_m^T\mathcal{C}_m(\underline{x})\underline{\varepsilon}_m}{\sigma^2\|\underline{\varepsilon}_m\|^2}\right)^{-\frac{1}{2}}\right). \quad (28)$$

We suppose that a perfect training sequence of length T_p is used and then $\widehat{h}_{pj} = h_{pj} + \sigma_e k_j$, $1 \leq j \leq n_T$, where k_j are modeled as independent real Gaussian random with zero mean and variance 1. From (4), $\sigma_e = \frac{\sigma}{\sqrt{T_p}\sigma_p}$. Thus, $C_m(\underline{x}) \longrightarrow n_T\sigma_e^2 I_m$, and $\underline{\varepsilon}_m^T\mathcal{C}_m(\underline{x})\underline{\varepsilon}_m \longrightarrow n_T\sigma_e^2\|\underline{\varepsilon}_m\|^2$. This leads to

$$P(E_m) = Q\left(\frac{\|\underline{\varepsilon}_m\|\sigma_d}{2\sigma}\left(1 + \frac{\sigma_d^2 n_T}{\sigma_p^2 Tp}\right)^{-\frac{1}{2}}\right). \quad (29)$$

The overall error probability $P_e(\Sigma)$ can then be approximated by

$$P_e(\Sigma) \simeq Q\left(\frac{\|\underline{\varepsilon}_m\|\sigma_d}{2\sigma}\left(1 + \frac{\sigma_d^2 n_T}{\sigma_p^2 Tp}\right)^{-\frac{1}{2}}\right). \quad (30)$$

By comparing the error probability obtained at the output of the MAP detector when the channel is perfectly known [8] and the one given in (30) when the channel is estimated, we conclude that the equivalent SNR at the input of the MAP detector is [7]

$$SNR_{eq} = \frac{\sigma_d^2}{\sigma^2}\left(1 + \frac{\sigma_d^2 n_T}{\sigma_p^2 Tp}\right)^{-1}. \quad (31)$$

References

[1] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[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] 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.

[4] I. Hadj Kacem, N. Sellami, A. Roumy, and I. Fijalkow, "Training sequence optimization for frequency selective channels with MAP equalization," *IEEE ISCCSP*, pp. 532–537, Malta, March 2008.

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

[6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[7] A. Gorokhov, "On the performance of the viterbi equalizer in the presence of channel estimation errors," *IEEE Signal Process. Letters*, vol. 5, no. 12, pp. 321–324, December 1998.

[8] G. 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.