Training interval length optimization for MIMO flat fading channels using decision-directed channel estimation

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 address the problem of optimization of the training sequence interval for MIMO (Multiple- Input Multiple-Output) flat fading channels when an iterative receiver composed of a likelihood generator and a Maximum *A Posteriori* (MAP) decoder is used. At each iteration of the receiver, the channel is estimated using the hard decisions on the transmitted symbols at the output of the decoder. The optimal length of the training interval is found by maximizing an effective signal-to-noise ratio (SNR) taking into account the data throughput loss due to the use of pilot symbols.**

I. Introduction

Wireless communications technologies are attracted by the Multiple-Input Multiple-Output (MIMO) systems since they provide significant increase in capacity [1]. In order to efficiently detect the transmitted symbols, the receiver needs a good estimate of the channel. The channel is classically estimated by using training sequences known at the receiver [2]. When the length of the training interval increases, the channel estimate becomes more reliable. However, this leads to a loss in terms of data throughput. Thus, instead of using the training sequences only, the information carried by the observations corresponding to the data symbols can also be used to improve iteratively the channel estimation. At each iteration, the channel estimator refines its estimation by using the hard or soft decisions on the data symbols at the output of the data detector or the channel decoder [3], [4], [5], [6].
A question that one can ask concerns the length of the training interval to choose in order to obtain a satisfactory initial channel estimate without decreasing significantly the data throughput. Several methods have been proposed to answer this question. The solution presented in [2] and [7] is based on maximizing a lower bound of the capacity of the training-based scheme respectively for a MIMO flat fading channel and a single-input single-output (SISO) frequency selective channel. Another approach tries to minimize the ratio of the channel Mean Square Error (MSE) to the data throughput [3] and therefore does not take into account the equalizer performance. In this paper, we consider a coded transmission over a MIMO flat fading channel. At the receiver, a turbo-detector consisting of a likelihood generator and a Maximum *A Posteriori* (MAP) decoder is considered. The channel is iteratively estimated by a decision-directed channel estimator using hard decisions on the coded bits at the output of the decoder. We derive the expression of the equivalent SNR at the output of the likelihood generator fed with the *a priori* information (from the decoder) and the channel estimate. We define, based on this expression, an effective SNR taking into account the loss in terms of data throughput due to the use of the training sequence. We propose to find the length of the training interval maximizing this expression. We show that when the decisions provided by the decoder are enough reliable, 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 we have done the study in [6] when the channel is Single Input Single Output (SISO) and frequency selective.
Throughout this paper scalars and matrices are lower and upper case respectively and vectors are underlined lower case. I_m is the $m \times m$ identity matrix. The operator $(.)^T$ denotes the transposition and $E(.)$ is the mathematical expectation.

II. Transmission system model

We consider a MIMO system composed of n_T transmit antennas and n_R receive antennas. As shown in Figure 1, the input data information bit sequence is encoded with a convolutional encoder. The output of the encoder is demultiplexed into n_T streams that are interleaved by different interleavers Π_i and mapped to the symbol alphabet $\mathcal{A} = \{-1, 1\}$. We assume that transmissions are organized into bursts. Each burst of T symbols is transmitted by one transmit antenna among n_T and the first T_p symbols are pilot symbols. The channel is supposed to be invariant during one burst and to change independently from burst to burst. The received baseband signal sampled at the symbol rate at time k at the receive antenna p is

$$y_k^{(p)} = \sum_{i=1}^{n_T} h_{pi} x_k^{(i)} + n_k^{(p)} \tag{1}$$

where $x_k^{(i)}$ is the k^{th} symbol transmitted by the i^{th} transmit antenna 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 real Gaussian variables. We assume that for a given receive antenna p, $E\left(\sum_{i=1}^{n_T}|h_{pi}|^2\right)=1$. 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.

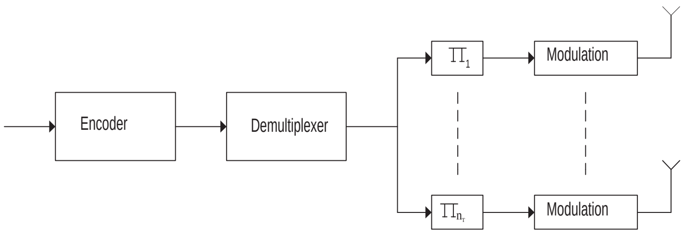

Fig. 1. Transmitter structure.

The initial channel estimate is provided to the receiver by a least square estimator using the training sequences with length T_p each, where $n_T \leq T_p \leq T$ [2] and T_p is the parameter to be optimized. We assume that the training sequences have ideal autocorrelation and crosscorrelation properties.

III. Iterative receiver with decision-directed channel estimation

As shown in Figure 2, we consider an iterative receiver composed of a likelihood generator, a MAP decoder and a decision-directed channel estimator. The likelihood generator computes the likelihoods on the coded bits $p(x_k^{(i)}=x|\underline{y}_k)$ for $1\leq i\leq n_T$ where $x\in\{-1,1\}$ and $\underline{y}_k=(y_k^{(1)},y_k^{(2)},...,y_k^{(n_R)})^T$ is the received signal at the n_R receive antennas at time k. These likelihoods are deinterleaved, multiplexed and used by the decoder to calculate the *A Posteriori* probabilities (APP) on the coded bits [8]. Then, based on these APPs, the decoder calculates extrinsic probabilities which will be interleaved, demultiplexed and then used by the likelihood generator at the next iteration as *a priori* probabilities. We start by presenting the likelihood generator. Then, we present the iterative channel estimator.

A. Likelihood generator

From (1), the received signal at the n_R receive antennas at time k is given by

$$\underline{y}_k = H\underline{x}_k + \underline{n}_k, \tag{2}$$

where H is the channel matrix, $\underline{x}_k=(x_k^{(1)},x_k^{(2)},...,x_k^{(n_T)})^T$ and $\underline{n}_k=(n_k^{(1)},n_k^{(2)},...,n_k^{(n_R)})^T$. The likelihood generator needs a good estimate $\widehat{H}$ of the channel in order to efficiently detect the transmitted symbols. We present in III-B the channel estimator.

The likelihood generator has to calculate the likelihood $p(\underline{y}_k|x_k^{(i)}=x)$ as

$$\begin{aligned} p(\underline{y}_k|x_k^{(i)}=x) &= \frac{p(\underline{y}_k, x_k^{(i)}=x)}{p(x_k^{(i)}=x)} \\ &= \sum_{\underline{x}_k\in X_0}\frac{p(\underline{y}_k|\underline{x}_k)p(\underline{x}_k)}{p(x_k^{(i)}=x)} \\ &= \sum_{\underline{x}_k\in X_0} p(\underline{y}_k|\underline{x}_k)\prod_{j=1,j\neq i}^{n_T} p\left(x_k^{(j)}\right) \end{aligned}$$

where X_0 is the set of all values that can be taken by $\underline{x}_k$ such that $x_k^{(i)}=x$ and $p(x_k^{(j)})$ are the *a priori* probabilities given by the MAP decoder to the likelihood generator.

Using the channel estimate, according to (2), the probability $p(\underline{y}_k|\underline{x}_k)$ is

$$p(\underline{y}_k|\underline{x}_k)=\frac{1}{(\pi\sigma^2)^{n_R}}\exp\left(\frac{-\|\underline{y}_k-\widehat{H}\underline{x}_k\|^2}{\sigma^2}\right). \tag{3}$$

B. Iterative channel estimator

At the first iteration of the iterative receiver, the channel estimate is provided by a least square estimator which uses the training symbols [2]. In order to refine the first channel estimate, the channel estimator uses the hard decisions on the transmitted coded symbols based on the APPs at the output of the decoder. Hence, the channel estimator is fed with $n_T T_p$ pilot symbols and $n_T\delta T$ estimates of the coded symbols coming from the decoder. Let $\underline{x}^{(i)}=\left(x_0^{(i)},...,x_{T_p+\delta T-1}^{(i)}\right)^T$ be the sequence transmitted by the i^{th} transmit antenna containing the T_p training symbols $x_k^{(i)}$ for $0\leq k\leq T_p-1$ and the δT data symbols $x_k^{(i)}$ for $T_p\leq k\leq T_p+\delta T-1$. The corresponding received vector at the p^{th} receive antenna $\underline{y}^{(p)}=\left(y_0^{(p)},y_1^{(p)},...,y_{T_p+\delta T-1}^{(p)}\right)^T$, is given by

$$\underline{y}^{(p)} = X\underline{h}^{(p)}+\underline{n}^{(p)} \tag{4}$$

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

$$X=\begin{pmatrix} x_0^{(1)} & ... & x_0^{(n_T)} \\ . & & . \\ . & & . \\ x_{T_{p-1}}^{(1)} & ... & x_{T_{p-1}}^{(n_T)} \\ x_{T_p}^{(1)} & & x_{T_p}^{(n_T)} \\ . & & . \\ . & & . \\ x_{T_p+\delta T-1}^{(1)} & ... & x_{T_p+\delta T-1}^{(n_T)} \end{pmatrix}$$

In order to estimate the channel, the observation vector $\underline{y}^{(p)}$ is approximated as follows:

$$\underline{y}^{(p)}\approx\widehat{X}\underline{h}^{(p)}+\underline{n}^{(p)} \tag{5}$$

where $\widehat{X}$ is the estimated version of the matrix X containing the hard decisions on the coded symbols at the output of the decoder. The iteration process can be repeated several times

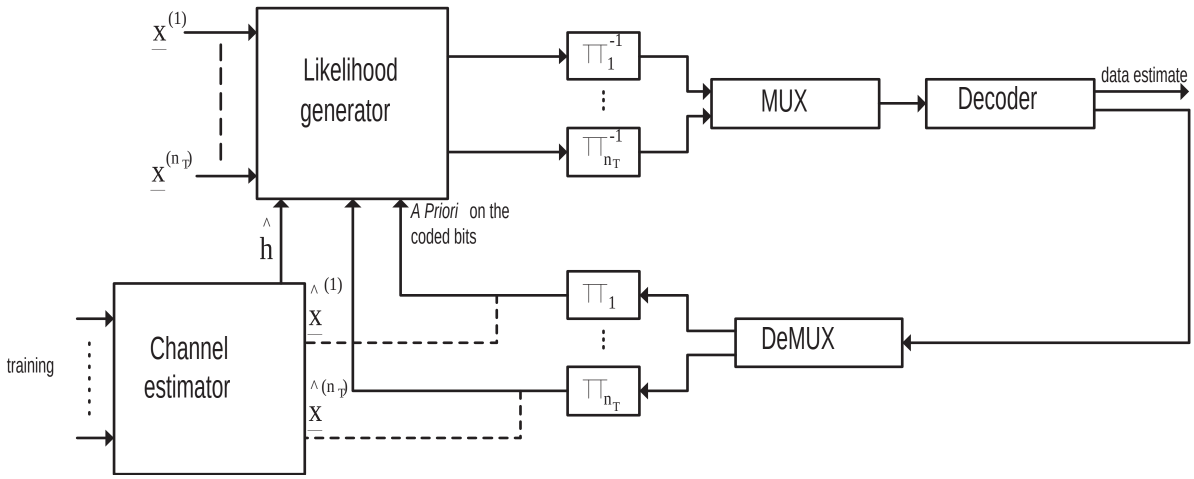

Fig. 2. Receiver structure

and here the matrix $\widehat{X}$ corresponds to the estimated symbols at the last iteration.
The least square channel estimate $\underline{\widehat{h}}^{(p)} = \left(\widehat{h}_{p1}, ..., \widehat{h}_{pn_T}\right)^T$ is given by

$$\underline{\widehat{h}}^{(p)} = \left(\widehat{X}^T \widehat{X}\right)^{-1} \widehat{X}^T \underline{y}^{(p)}. \tag{6}$$

IV. Equivalent SNR at the output of the likelihood generator

We assume that δT is fixed such that $\delta T >> n_T$. We also assume that the vector of errors on the coded symbols at the output of the decoder is independent of the noise vector. In average, the errors are assumed to be uniformly distributed over a burst. By generalizing the study of [4] to the MIMO flat fading channels, we can show that the channel estimation Mean Square Error (MSE) per receive antenna is given by

$$E\left(\|\delta \underline{h}\|^2\right) = \frac{\sigma^2 n_T}{T_p + \delta T} + 4\frac{\overline{\beta^2}\delta T^2 + (n_T - 1)\overline{\beta}\delta T}{\left(T_p + \delta T\right)^2} \tag{7}$$

where $\delta \underline{h} = \underline{h}^{(p)} - \underline{\widehat{h}}^{(p)}$, $\overline{\beta} = \frac{1}{n_T \delta T} E(n)$, $\overline{\beta^2} = \frac{1}{n_T \delta T^2} E(n^2)$ and n is the number of erroneous hard decisions on the coded symbols at the output of the decoder, used by the channel estimator. We give in the Appendix the proof of (7).
We assume that the *a priori* (extrinsic) Log Likelihood Ratios (LLRs) at the input of the likelihood generator, fed back from the decoder, are independent and identically distributed (iid) samples from a random variable with the conditional pdf $\mathcal{N}\left(\frac{\pm 2}{\sigma_a^2}, \frac{4}{\sigma_a^2}\right)$ [9], [10], [11]. By generalizing the study of [12], the equivalent signal-to-noise ratio at the output of the likelihood generator fed with the *a priori* LLRs from the decoder and the channel estimate can be approximated by

$$SNR_{eq} = \frac{1}{\sigma^2}(1 + \mu^2)\left(1 + \frac{E\left(\|\delta \underline{h}\|^2\right)}{\sigma^2(1 + \mu^2)}\right)^{-1} \tag{8}$$

where $\mu = \frac{\sigma}{\sigma_a}$ and $E\left(\|\delta \underline{h}\|^2\right)$ is the channel estimation MSE given in (7).

V. Optimization of the training interval length

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 [6] an effective SNR at the output of the likelihood generator as

$$\begin{aligned} SNR_{eff,eq} &= \tfrac{T - T_p}{T} SNR_{eq} \\ &= \tfrac{T - T_p}{T} \tfrac{(1+\mu^2)}{\sigma^2}\left(1 + \tfrac{E\left(\|\delta \underline{h}\|^2\right)}{\sigma^2(1+\mu^2)}\right)^{-1}. \end{aligned} \tag{9}$$

Our goal is to maximize the effective SNR when the channel is iteratively estimated by the decision-directed channel estimator. Hence, we define the following optimization problem

$$\begin{cases} \max\ SNR_{eff,eq} \\ \text{s.t.} \\ n_T \leq T_p \leq T - \delta T \end{cases} \tag{10}$$

Let $t \in \mathbb{R}_+^*$, $t \geq n_T$ and

$$f_1(t) = \frac{T - t}{T}\frac{(1 + \mu^2)}{\sigma^2}\left(1 + g(t)\right)^{-1} \tag{11}$$

where $g(t) = \frac{(1+\mu^2)}{\sigma^2}\left(\frac{\sigma^2 n_T}{t + \delta T} + 4\,\frac{\overline{\beta^2}\delta T^2 + (n_T - 1)\overline{\beta}\delta T}{(t + \delta T)^2}\right)$. Thus, $SNR_{eff,eq} = f_1(T_p)$.
When $g(t) << 1$, $f_1(t)$ can be approximated by

$$f_1(t) \approx \frac{T - t}{T}\frac{(1 + \mu^2)}{\sigma^2}\left(1 - g(t)\right) \tag{12}$$

which is a decreasing function.
When the δT decisions on the data symbols added to the training interval for the n_T transmit antennas are reliable, $g(T_p) << 1$. Thus, the optimal length of the training sequence is

$$T_p^* = n_T. \tag{13}$$

Remark 1: For the non-iterative receivers, the training interval length maximizing the effective SNR may be larger than n_T.

Remark 2: When the hard decisions used by the decision-directed channel estimator are not reliable, the approximation $g(T_p) << 1$ becomes inaccurate and the optimization problem can not be solved analytically. However, it is easy to show that when the decisions becomes less reliable, the optimal training interval length increases. Table I gives an idea on the reliability of the hard decisions used by the decision-directed channel estimator for $T = 512$, $\delta T = 100$, $n_T = 2$, $SNR = 10dB$ and $\sigma_a^2 = 0.5$. We suppose here that the channel estimator is fed with artificial *a posteriori* LLRs modeled as iid samples from a random variable with the conditional pdf $\mathcal{N}\left(\frac{\pm 2}{\sigma_x^2}, \frac{4}{\sigma_x^2}\right)$ [9], [10], [11].

σ_x^2	$\overline{\beta}$	$\overline{\beta^2}$	T_p^*
2	0.24	0.060	117
1	0.16	0.026	78
0.7	0.12	0.015	52
0.6	0.10	0.011	40
0.5	0.08	0.007	25
0.3	0.03	0.002	2
0.1	0.008	0.0001	2

TABLE I
OPTIMAL TRAINING INTERVAL LENGTH FOR DIFFERENT VALUES OF σ_x^2 FOR $T = 512$, $\delta T = 100$, $n_T = 2$ AND $\sigma_a^2 = 0.5$.

VI. SIMULATION RESULTS

We propose to validate the theoretical MSE expression given in (7). We simulate the decision directed channel estimator fed with artificial *a posteriori* LLRs modeled as iid samples from a random variable with the conditional pdf $\mathcal{N}\left(\frac{\pm 2}{\sigma_x^2}, \frac{4}{\sigma_x^2}\right)$ [9], [10], [11]. Figure 3 shows the channel estimation MSE curves with respect to SNR for different values of T, n_T and n_R. The channel tap gains h_{ji} are modeled as independent zero mean complex Gaussian variables with variance $1/n_T$. The training interval length is $T_p = n_T$, $\delta T = T - n_T$ and $\sigma_x^2 = 0.1$. The solid curves are obtained by simulations and the dotted curves are obtained using (7). We note that the theoretical curves approximate well the curves obtained by simulations.

Now, we consider the whole MIMO system with the channel coding at the transmitter and the iterative receiver composed of a likelihood generator and a MAP decoder. The information data are encoded using the rate $\frac{1}{2}$ convolutional code with generator polynomials $(7, 5)$ in octal. At the first iteration, the channel is estimated by using the training sequences. At the next iterations, it is estimated by using the decision-directed technique. Figure 4 shows the BER performance at the output of the MAP decoder at the convergence (after two iterations) with respect to $SNR_{eff} = \frac{T-T_p}{T}SNR$, where SNR is the signal to noise ratio at the input of the likelihood generator, for $n_T = n_R = 2$, $T = 512$, $\delta T = 450$ and for different values of T_p, the length of the training interval. The channel tap gains h_{ji} are modeled as independent zero mean complex Gaussian variables with variance 0.5. The δT estimates of the coded symbols at the input of the channel estimator are obtained by

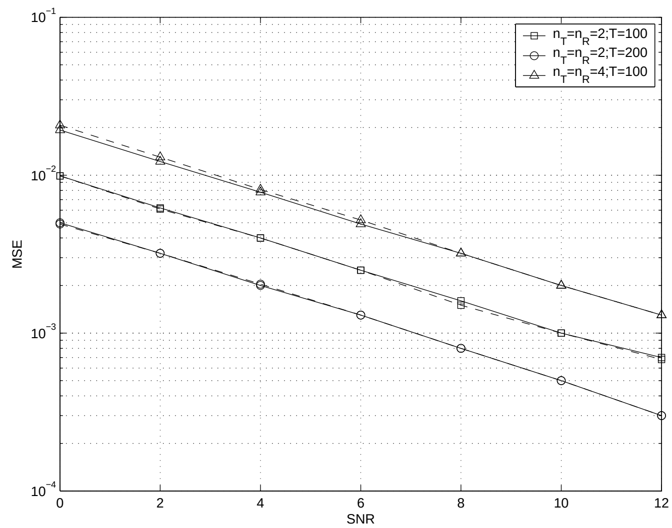

Fig. 3. Channel estimation MSE curves with respect to SNR for different values of T, n_T and n_R for $T_p = n_T$ and $\sigma_x^2 = 0.1$.

making hard decisions on the *a posteriori* LLRs at the output of the MAP decoder. From (13), $T_p^* = 2$. This is confirmed by simulations since they show that the MAP decoder presents its best performance when $T_p = 2$.

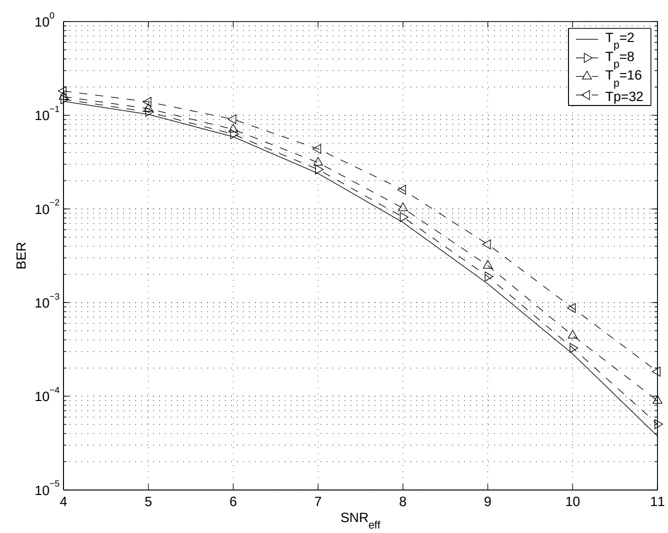

Fig. 4. BER at the output of the MAP decoder for different values of T_p.

VII. CONCLUSION

In this paper, we considered the problem of optimization of the training interval length for MIMO flat fading channels when an iterative receiver is used. The iterative receiver is composed of a likelihood generator, a MAP decoder and a decision-directed channel estimator. We calculated an effective SNR at the output of the likelihood generator taking into account the data throughput loss due to the use of pilot symbols. We proved that the optimal training interval length maximizing the effective signal-to-noise ratio is equal to the number of transmit antennas when the decisions provided by the decoder are reliable.

VIII. Appendix

Let $\underline{x}^{(i)} = \left(x_0^{(i)}, ..., x_{T_p+\delta T-1}^{(i)}\right)^T$ be the sequence transmitted by the i^{th} transmit antenna containing the T_p training symbols $x_k^{(i)}$ for $0 \leq k \leq T_p - 1$ and the δT data symbols $x_k^{(i)}$ for $T_p \leq k \leq T_p + \delta T - 1$. The corresponding received vector at the p^{th} receive antenna $\underline{y}^{(p)} = \left(y_0^{(p)}, y_1^{(p)}, ..., y_{T_p+\delta T-1}^{(p)}\right)^T$, is given by

$$\underline{y}^{(p)} = X\underline{h}^{(p)} + \underline{n}^{(p)} \tag{14}$$

where $\underline{h}^{(p)} = (h_{p1}, h_{p2}, ..., h_{pn_T})^T$, $\underline{n}^{(p)} = \left(n_0^{(p)}, n_1^{(p)}, ..., n_{T_p+\delta T-1}^{(p)}\right)^T$ and $X = \left[\underline{x}^{(1)}, ..., \underline{x}^{(n_T)}\right]$ is the matrix containing the pilots symbols and the $n_T\delta T$ data symbols. In order to estimate the channel, the observation vector $\underline{y}^{(p)}$ is approximated as follows:

$$\underline{y}^{(p)} \approx \widehat{X}\underline{h}^{(p)} + \underline{n}^{(p)} \tag{15}$$

where $\widehat{X}$ is the estimated version of the matrix X containing the hard decisions on the coded symbols at the output of the decoder. $\widehat{X}$ can be written as

$$\widehat{X} = X + \delta X \tag{16}$$

where $\delta X_{ij} \in \{-2, 0, 2\}$ for $0 \leq i \leq T_p + \delta T - 1$ and $1 \leq j \leq n_T$.

The least square channel estimate $\widehat{\underline{h}}^{(p)} = \left(\widehat{h}_{p1}, ..., \widehat{h}_{pn_T}\right)^T$ is given by

$$\widehat{\underline{h}}^{(p)} = \left(\widehat{X}^T\widehat{X}\right)^{-1}\widehat{X}^T\underline{y}^{(p)} \tag{17}$$

For the sake of conciseness, the exponent p is omitted from $\underline{y}^{(p)}$, $\underline{n}^{(p)}$, $\underline{h}^{(p)}$ and $\widehat{\underline{h}}^{(p)}$.

Under the assumptions that $X^TX = (T_p + \delta T)I_{n_T}$ and $\widehat{X}^T\widehat{X} = (T_p + \delta T)I_{n_T}$. The channel estimate can be rewritten as

$$\begin{aligned}\widehat{\underline{h}}^{(p)} &= \tfrac{1}{T_p+\delta T}I_{n_T}\left(X^T + \delta X^T\right)(X\underline{h} + \underline{n}) \\ &= \underline{h} + \tfrac{1}{T_p+\delta T}\left(X^T\underline{n} + \delta X^T X\underline{h} + \delta X^T\underline{n}\right).\end{aligned} \tag{18}$$

Hence,

$$\begin{aligned}\delta\underline{h} &= \widehat{\underline{h}} - \underline{h} \\ &= \tfrac{1}{T_p+\delta T}(M_1 + M_2 + M_3)\end{aligned} \tag{19}$$

where $M_1 = X^T\underline{n}$, $M_2 = \delta X^T X\underline{h}$ and $M_3 = \delta X^T\underline{n}$.

The channel estimation Mean Square Error (MSE) per receive antenna is given by

$$\begin{aligned}E\left(\|\delta\underline{h}\|^2\right) &= E\left[Tr(\delta\underline{h}\delta\underline{h}^T)\right] \\ &= Tr\left(E\left[\delta\underline{h}\delta\underline{h}^T\right]\right) \\ = \tfrac{1}{(T_p+\delta T)^2}&\left(Tr\left(E[M_1M_1^T]\right) + Tr\left(E[M_1M_3^T]\right) + Tr\left(E[M_2M_2^T]\right)\right) \\ + \tfrac{1}{(T_p+\delta T)^2}&\left(Tr\left(E[M_3M_1^T]\right) + Tr\left(E[M_3M_3^T]\right)\right)\end{aligned} \tag{20}$$

It is easy to show that

$$Tr\left(E[M_1M_1^T]\right) = \sigma^2 n_T(T_p + \delta T) \tag{21}$$

$$Tr\left(E[M_1M_3^T]\right) = -\sigma^2 n_T E(n) \tag{22}$$

$$Tr\left(E[M_3M_1^T]\right) = -\sigma^2 n_T E(n) \tag{23}$$

$$Tr\left(E[M_3M_3^T]\right) = 2\sigma^2 n_T E(n) \tag{24}$$

where n is the number of erroneous hard decisions on the coded symbols at the output of the decoder, used by the channel estimator.

To evaluate the term $Tr\left(E[M_2M_2^T]\right)$, we assume that the coefficients of $\underline{h}$ are decorrelated. Hence,

$$\begin{aligned}Tr\left(E[M_2M_2^T]\right) &= E\left[Tr\left(\delta X^T X D_h X^T\delta X\right)\right] \\ &= E\left[Tr\left(\textstyle\sum_{i=1}^{n_T} E_j\underline{u}_j\underline{u}_j^T\right)\right] \\ &= \textstyle\sum_{i=1}^{n_T} E_j E\left[\left\|\underline{u}_j\right\|^2\right].\end{aligned} \tag{25}$$

where $D_h = \underline{h}\,\underline{h}^T$, $E_j = E(|h_j|^2)$ and the vectors $\underline{u}_j$ for $1 \leq j \leq n_T$ are defined by $\delta X^T X \equiv [\underline{u}_1, ..., \underline{u}_{n_T}]$.

It can be checked that $E\left[\left\|\underline{u}_j\right\|^2\right]$ can be expressed as [4]

$$E\left[\left\|\underline{u}_j\right\|^2\right] = \frac{4}{n_T}E(n^2) + 4\overline{\beta}\sum_{i=1, i\neq j}^{n_T}\delta T - |i - j| \tag{26}$$

where $\overline{\beta} = \frac{1}{n_T\delta T}E(n)$.

Hence,

$$Tr\left(E[M_2M_2^T]\right) = 4E\left[\|\underline{h}\|^2\right]\left(\tfrac{1}{n_T}E(n^2) + (n_T - 1)\delta T\overline{\beta} - S\right) \tag{27}$$

where $S = \overline{\beta}\left(\frac{\sum_{j=1}^{n_T} jE_j}{\sum_{j=1}^{n_T} E_j} - (n_T - 1)\frac{\sum_{j=1}^{n_T} j^2E_j}{\sum_{j=1}^{n_T} E_j} + \frac{(n_T-1)n_T}{2}\right)$.

Now, we assume that $\delta T >> n_T$. Hence $S << \left(\frac{1}{n_T}E(n^2) + (n_T - 1)\delta T\overline{\beta}\right)$. As, $E\left[\|\underline{h}\|^2\right] = 1$, we have

$$Tr\left(E[M_2M_2^T]\right) = 4\left(\tfrac{1}{n_T}E(n^2) + (n_T - 1)\delta T\overline{\beta}\right) \tag{28}$$

Finally, the channel estimation MSE is given by

$$\begin{aligned}E\left(\|\delta\underline{h}\|^2\right) &= \tfrac{\sigma^2 n_T}{T_p+\delta T} + \tfrac{1}{(T_p+\delta T)^2}\left(\tfrac{1}{n_T}E(n^2) + (n_T - 1)\delta T\overline{\beta}\right) \\ &= \tfrac{\sigma^2 n_T}{T_p+\delta T} + \tfrac{\delta T^2\overline{\beta^2} + (n_T-1)\delta T\overline{\beta}}{(T_p+\delta T)^2}.\end{aligned} \tag{29}$$

where $\overline{\beta^2} = \frac{1}{n_T\delta T^2}E(n^2)$.

References

[1] G.J. Foschini and M.J. 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] S. Lasaulce and N. Sellami, "On the impact of using unreliable data on the bootstrap channel estimation performance," *4th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Rome, Italy*, pp. 348–352, June 2003.

[5] N. Sellami, S. Lasaulce, and I. Fijalkow, “Turbo channel estimation for coded DS-CDMA systems over frequency selective channels,” *4th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Rome, Italy*, pp. 80–84, June 2003.

[6] I. Hadj Kacem, N. Sellami, I. Fijalkow, and A. Roumy, “Training sequence length optimization for a turbo-detector using decision-directed channel estimation,” *Research Letters in Communications*, vol. 2008, pp. Article ID 860368, 5 pages, 2008. doi :10.1155/2008/860368, 2008.

[7] 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.

[8] 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.

[9] M. Tüchler, R. Koetter, and A. Singer, “Turbo equalization: principles and new results,” *IEEE Trans. on Communications*, vol. 50, no. 5, pp. 754–767, May 2002.

[10] S. Ten Brink, “Convergence of iterative decoding,” *IEEE Electronics Letters*, vol. 35, pp. 806–808, May 1999.

[11] N. Sellami, A. Roumy, and I. Fijalkow, “A proof of convergence of the MAP turbo-detector to the AWGN case,” *IEEE Trans. on Signal Processing*, vol. 56, no. 4, pp. 1548–1561, April 2008.

[12] 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.