

Optimal Estimator-Detector Receivers for Space-Time Block Coding

Olivier Roy, Sylvie Perreau and Alex Grant

Abstract—Most space-time coding schemes can be classified either as non-coherent (decoding is performed without forming an explicit channel estimate) or coherent (decoding is performed conditioned upon a channel estimate as if it were the actual channel realisation). In this paper we prove that optimal non-coherent decoding can be decomposed into a channel estimation step, followed by coherent decoding step. Surprisingly the required estimators do not in general minimise the mean squared error between the estimated and actual channel. We also investigate the role of training sequences in such systems.

Index Terms—Coherent / non-coherent detection, channel estimation, generalised likelihood ratio test, fading channels, multi-element antenna arrays, space-time block coding.

I. INTRODUCTION

Since the appearance of seminal works on space-time information theory and coding [1], [2], [3], [4], [5] two main philosophies have emerged for the design of codes and associated decoding algorithms.

The first strategy takes the view that fundamentally, the parameters of the space-time channel are unknown and information theoretic principles would direct us to design codes directly for the channel with unknown parameters. This is the non-coherent approach taken in [6], [7], [8], [9] and related works. Here the goal is to design codes and decoders which minimise the decoding error probability, without the possibly unnecessary restrictions due to the use of training sequences (which constrains code design) and channel estimation (which constrains decoder design). Within this framework, optimal detection, in term of minimising the decoder Word Error Rate (WER) is the non-coherent Maximum Likelihood (ML) rule given in [7]. Unless otherwise specified, this is what we shall mean by ML or optimal detection. Code design for this approach is hard since it is difficult to obtain an expression for the probability of error in closed-form and even more complex to find design rules to minimise it. In [7] the use of unitary codes was motivated by asymptotic capacity considerations. We shall refer to this first class of strategies as *non-coherent*.

The second strategy is to design the system such that the receiver can easily form some kind of estimate of the fading channel parameters, which is subsequently used within a coherent metric as if it were in fact the actual channel realisation [5], [10]. Such sub-optimal detection schemes combine well-known

channel estimation techniques with low-complexity decoding rules based on the invalid assumption of perfect channel knowledge. This assumption is usually justified through the use of training sequences or pilot symbols. The interest in this channel estimation based approach is largely motivated by the fact that the optimality provided by ML decoding often comes at the cost of prohibitive complexity. Although systems based on the principles of channel estimation appear to be attractive from a practical implementation point of view, the use of training sequences can only ever have a negative impact upon spectral efficiency. We shall refer to this second class of strategies as *coherent*. We emphasise that in this coherent case, the channel is not a-priori perfectly known, but is always estimated somehow from the received signals.

In this paper, we wish to compare these two approaches. We shall take the view that at the onset of decoding, the space-time channel gains are unknown. We do assume perfect knowledge of the second order statistics of the channel (i.e. the covariance of the channel gains). We shall regard any training sequences transmitted for the purpose of channel estimation to be part of the coded transmission (i.e. a deterministic prefix for each codeword). As pointed out in [7], code design for the coherent and non-coherent problems turn out to be extremely different since optimal codes for one case may give bad performance in the latter. Optimal transmitter designs, based on second order channel statistics, minimising an upper bound on the Symbol Error Rate (SER) are discussed in [11]. It is shown in [11] that trying to minimise the mean squared error of the channel estimate does not lead in general to the same optimal design.

In order to compare coherent and optimal non-coherent detection, we use an estimation-detection approach. Decoding will be performed as a two step process. First, a set of channel estimates is formed, one for each possible transmitted codeword. Secondly, a coherent metric is computed for each codeword, conditioned on its corresponding channel estimate. The decoder outputs the codeword with the best metric. Working directly on the decision metrics, we show that under the assumptions of unitary codewords and i.i.d. fading, Minimum Mean Square Error (MMSE) channel estimation preserves optimality. We relate this result to the Generalised Likelihood Ratio Test (GLRT) [14] and show that if the assumption of unitary codewords does not hold, optimality may still be retained but that in this case, the required channel estimators are no longer MMSE. Finally, we relate this work to the case where channel estimation is done using a training sequence.

This paper is organised as follows. In Section II we describe the system model and review optimal detection for this system. Section III discusses a possibly sub-optimal estimation-

Sylvie Perreau and Alex Grant are with the Institute of Telecommunications Research, University of South Australia, Mawson Lakes 5095, Australia, e-mail:{Sylvie.Perreau, Alex.Grant}@unisa.edu.au. Olivier Roy is with the Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland, e-mail:Olivier.Roy@epfl.ch. A. Grant is supported by the Australian Government under ARC Grants DP0209658 and DP0344856.

detection approach and introduces channel estimators that allow optimal detection. Section IV considers the use of training sequences. Finally, we offer some conclusions in Section V.

The following notation will be used throughout this paper. Lower case (e.g. x) denotes column vectors or scalars, depending on context whereas upper case (e.g. X) is used for matrices. The superscripts $(\cdot)^T$ and $(\cdot)^*$ denote transpose and conjugate transpose, respectively. $O_{N \times M}$ denotes an all zero matrix of size $N \times M$, I_N is the identity matrix of size $N \times N$ and $\text{diag}(d_1, d_2, \dots, d_N)$ is a $N \times N$ diagonal matrix with diagonal elements d_1, d_2, \dots, d_N . The squared Frobenius norm is denoted by $\|\cdot\|^2$. Finally, $\text{tr}(\cdot)$ and $\text{etr}(\cdot)$ stand for the trace and the exponential of the trace of a matrix, $|\cdot|$ for its determinant and $E\{\cdot\}$ for its expectation.

II. SYSTEM MODEL AND OPTIMAL DETECTION

Consider a t transmit, r receive space-time channel operating in Rayleigh flat fading environment with l consecutive channel uses ($l \geq t$),

$$Y = XH + N \quad (1)$$

where $Y \in \mathbb{C}^{l \times r}$ is the received matrix, X is the $l \times t$ transmitted codeword chosen equiprobably from a codebook, $X \in \{X_0, X_1\}$ (this restriction to two codewords is for simplicity only). The matrix $H \in \mathbb{C}^{l \times r}$ contains the channel gains and $N \in \mathbb{C}^{l \times r}$ is an additive noise matrix. The elements of H are i.i.d. circularly symmetric Gaussian with unit variance, those of N are i.i.d. circularly symmetric Gaussian with variance σ^2 .

Conditioned upon the received signal Y according to (1), the optimal detection problem is to decide between X_0 and X_1 with minimum probability of error. This is achieved by the Maximum Likelihood (ML) rule

$$p(Y|X_0) \stackrel{0}{\underset{1}{\gtrless}} p(Y|X_1). \quad (2)$$

Conditioned upon X_i , the output Y is a zero-mean complex Gaussian with covariance matrix $\Lambda_i = (X_i X_i^* + \sigma^2 I_l)$. The corresponding conditional densities are thus

$$p(Y|X_i) = \frac{1}{\pi^{lr} |\Lambda_i|^r} \text{etr}(-\Lambda_i^{-1} Y Y^*) \quad (3)$$

for $i = 0, 1$. Defining the decision statistic

$$\omega(Y) = \text{etr}((\Lambda_1^{-1} - \Lambda_0^{-1}) Y Y^*), \quad (4)$$

ML detection is done according to the rule

$$\omega(Y) \stackrel{0}{\underset{1}{\gtrless}} |\Lambda_0|^r |\Lambda_1|^{-r}. \quad (5)$$

III. ESTIMATION-DETECTION

In this section, we introduce the general estimation-detection approach considered in this paper and relate it to the GLRT. This latter allows optimal detection under the assumption of unitary codewords and i.i.d. fading. We show that if we use arbitrary codebooks, optimality can still be retained by changing the channel estimator used in the estimation-detection scheme.

A. Estimator-Detector Decomposition

Consider decoding according to the following (possible sub-optimal) two step process. First, compute two channel estimates

$$\hat{H}_0 = K_0 Y \quad (6)$$

$$\hat{H}_1 = K_1 Y \quad (7)$$

where K_0 and K_1 are $t \times l$ matrices. These channel estimates are computed without the use of any deterministic training such as Pilot Symbol Assisted Modulation (PSAM) [12] techniques. This will allow us to make a fair comparison of this estimation-detection approach to the optimal ML detection since no additional information is needed to compute the channel estimates.

Now, using these channel estimates, we can define a ‘‘coherent’’ detection approach to the originally non-coherent detection problem. Let

$$\mu_i(Y) = \exp \|Y - X_i \hat{H}_i\|^2 \quad (8)$$

$$= \text{etr} \left((Y - X_i \hat{H}_i)^* (Y - X_i \hat{H}_i) \right) \quad (9)$$

be the ‘‘coherent’’ metric for X_i , using \hat{H}_i as if it were the true channel realisation. Within this framework, the decision statistic is

$$\mu(Y) = \exp \left(\|Y - X_1 \hat{H}_1\|^2 - \|Y - X_0 \hat{H}_0\|^2 \right) \quad (10)$$

with decision rule

$$\mu(Y) \stackrel{0}{\underset{1}{\gtrless}} 1. \quad (11)$$

The rest of this paper will focus on finding channel estimators \hat{H}_i that allow the estimation-detection approach given by (10) and (11) to be equivalent to the optimal detection given by (4) and (5). More precisely, we will look for channel estimators that lead to the decision statistic (4), the decision threshold in (5) being easy to fix since the $|\Lambda_i|$ are easily calculated.

B. Generalised Likelihood Ratio Test

In our system model, a particular case of the approach taken in the previous subsection is the GLRT [14]. In fact, GLRT detection is done according to

$$\max_H p(Y|X_0, H) \stackrel{0}{\underset{1}{\gtrless}} \max_H p(Y|X_1, H) \quad (12)$$

which under our Gaussian assumptions turns out to be equivalent to the estimation-detection scheme described previously with \hat{H}_i being the ML channel estimate under hypothesis i

$$\hat{H}_i = \arg \max_H p(Y|X_i, H) \quad (13)$$

$$= (X_i^* X_i)^{-1} X_i^* Y. \quad (14)$$

Note that this channel estimate corresponds to a Zero-Forcing (ZF) estimate. It has been shown [14] that under both unitary codewords and i.i.d. fading assumptions, the GLRT performs optimal detection. In the following theorem we show that under the same assumptions, the use of an MMSE estimate

also retains optimality.

Theorem 1: Let the channel fading be i.i.d., X_0 and X_1 be unitary codewords and the MMSE channel estimate under hypothesis i be given by

$$\hat{H}_i = K_i Y$$

where

$$K_i = X_i^* (X_i X_i^* + \sigma^2 I_l)^{-1}$$

for $i = 0, 1$. Then the ML decision rule is given by (11).

Proof: We have that

$$\begin{aligned} Y - X_i \hat{H}_i &= Y - X_i K_i Y \\ &= (I - X_i K_i) Y \\ &= \sigma^2 (X_i X_i^* + \sigma^2 I_l)^{-1} Y \\ &= \sigma^2 \Lambda_i^{-1} Y. \end{aligned}$$

Thus, the decision statistic (10) becomes

$$\mu(Y) = \text{etr}(\sigma^4 (\Lambda_1^{-2} - \Lambda_0^{-2}) Y Y^*).$$

Using the matrix inversion lemma and the fact that $X_i^* X_i = I_t$, we have that

$$\begin{aligned} \Lambda_i^{-1} &= (X_i X_i^* + \sigma^2 I_l)^{-1} \\ &= \frac{1}{\sigma^2} I_l - \frac{1}{\sigma^4} X_i \left(\frac{1}{\sigma^2} X_i^* X_i + I_t \right)^{-1} X_i^* \\ &= \frac{1}{\sigma^2} I_l - \frac{1}{\sigma^2(\sigma^2 + 1)} X_i X_i^*. \end{aligned}$$

Similarly,

$$\begin{aligned} \Lambda_i^{-2} &= (\Lambda_i^{-1})^2 \\ &= \left(\frac{1}{\sigma^2} I_l - \frac{1}{\sigma^2(\sigma^2 + 1)} X_i X_i^* \right)^2 \\ &= \frac{1}{\sigma^4} I_l - \left(\frac{1}{\sigma^4} - \frac{1}{(\sigma^2 + 1)^2} \right) X_i X_i^*. \end{aligned}$$

Furthermore, using the determinant formula

$$|I + AB| = |I + BA|,$$

we obtain

$$\begin{aligned} |\Lambda_i^{-1}| &= \left| \frac{1}{\sigma^2} I_l - \frac{1}{\sigma^2(\sigma^2 + 1)} X_i X_i^* \right| \\ &= \left| \frac{1}{\sigma^2} I_t - \frac{1}{\sigma^2(\sigma^2 + 1)} X_i^* X_i \right| \\ &= \left(\frac{1}{\sigma^2 + 1} \right)^t \end{aligned}$$

which does not depend on i . Finally, since $(\Lambda_1^{-1} - \Lambda_0^{-1})$ and $(\Lambda_1^{-2} - \Lambda_0^{-2})$ differ only by a positive multiplicative constant and the decision threshold is equal to 1, the two decision rules (11) and (5) are equivalent. ■

Theorem 1 shows that the particular form of unitary codes allows us to some freedom in the design of the estimator that retains optimality. Furthermore, good channel estimation is not crucial in minimising the word error rate: we can do optimal detection at very low SNR with a ZF channel estimate which quality is poor compared to an MMSE estimate.

C. Arbitrary Codebooks and I.I.D. Fading

When no assumption is made on the codewords, an ML channel estimate, as used in the GLRT, does not preserve optimality. It has been shown [13] that even under asymptotic considerations (as the SNR goes to infinity), the GLRT performs worse than the ML receiver. In this subsection, we show that for arbitrary codebooks and i.i.d. fading, optimality can still be preserved using an estimation-detection approach. We exhibit the channel estimator needed, hereafter referred as the Minimum Codeword Error Probability (MCEP) estimator for i.i.d. fading.

Theorem 2: Let the channel fading be i.i.d. $X_i = U_i \Sigma_i V_i^*$ be the singular value decomposition (SVD) of X_i with U_i (resp. V_i) a $l \times l$ (resp. $t \times t$) unitary matrix and Σ_i a $l \times t$ matrix of the form

$$\Sigma_i = \begin{bmatrix} \sigma_{i,1} & & & & \\ & \sigma_{i,2} & & & \\ & & \ddots & & \\ & & & \sigma_{i,t} & \\ & & & & O_{(l-t) \times t} \end{bmatrix}$$

for $i = 0, 1$. The MCEP channel estimation for i.i.d. fading is performed using K_0 and K_1 given by

$$\begin{aligned} K_0 &= V_0 D_0 U_0^* \\ K_1 &= V_1 D_1 U_1^* \end{aligned}$$

where D_i is a $t \times l$ matrix of the form

$$D_i = \begin{bmatrix} d_{i,1} & & & & \\ & d_{i,2} & & & \\ & & \ddots & & \\ & & & d_{i,t} & \\ & & & & O_{t \times (l-t)} \end{bmatrix}$$

with

$$d_{i,j} = \begin{cases} \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \left(1 - \sqrt{\frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2}} \right) & \text{if } \sigma_{i,j} \neq 0 \\ 0 & \text{if } \sigma_{i,j} = 0 \end{cases}$$

for $j = 1, \dots, t$.

Proof: Let $X_i = U_i \Sigma_i V_i^*$ be the SVD of X_i , we have that

$$\begin{aligned} \Lambda_i &= X_i X_i^* + \sigma^2 I_l \\ &= U_i \Sigma_i V_i^* V_i \Sigma_i^* U_i^* + \sigma^2 I_l \\ &= U_i (|\Sigma_i|^2 U_i^* + \sigma^2 I_l) \\ &= U_i (|\Sigma_i|^2 + \sigma^2 I_l) U_i^* \end{aligned}$$

and thus

$$\Lambda_i^{-1/2} = U_i (|\Sigma_i|^2 + \sigma^2 I_l)^{-1/2} U_i^*.$$

Now we can find a matrix K_i of the form

$$K_i = V_i D_i U_i^*$$

such that

$$I_l - X_i K_i = k \Lambda_i^{-1/2}$$

for some constant k . We have that

$$\begin{aligned} I_l - X_i K_i &= k \Lambda_i^{-1/2} \\ I_l - U_i \Sigma_i V_i^* V_i D_i U_i^* &= k \Lambda_i^{-1/2} \\ I_l - U_i \Sigma_i D_i U_i^* &= k U_i \left(|\Sigma_i|^2 + \sigma^2 I_l \right)^{-1/2} U_i^*. \end{aligned}$$

Note that $\Sigma_i D_i$ is of the form

$$\begin{aligned} &\begin{bmatrix} \bar{\Sigma}_i & \\ & O_{(l-t) \times t} \end{bmatrix} \begin{bmatrix} \bar{D}_i & O_{t \times (l-t)} \\ & \end{bmatrix} \\ &= \begin{bmatrix} \bar{\Sigma}_i \bar{D}_i & O_{t \times (l-t)} \\ O_{(l-t) \times t} & O_{(l-t) \times (l-t)} \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} \bar{\Sigma}_i &= \text{diag}(\sigma_{i,1}, \sigma_{i,2}, \dots, \sigma_{i,t}) \\ \bar{D}_i &= \text{diag}(d_{i,1}, d_{i,2}, \dots, d_{i,t}). \end{aligned}$$

Coordinate-wise we obtain

$$\begin{cases} 1 - \sigma_{i,j} d_{i,j} = k \left(|\sigma_{i,j}|^2 + \sigma^2 \right)^{-1/2} & j = 1, \dots, t \\ 1 = k/\sigma & j = t+1, \dots, l \end{cases}$$

thus $k = \sigma$ and

$$d_{i,j} = \begin{cases} \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \left(1 - \sqrt{\frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2}} \right) & \text{if } \sigma_{i,j} \neq 0 \\ 0 & \text{if } \sigma_{i,j} = 0 \end{cases}$$

for $j = 1, \dots, t$. \blacksquare

Similarly we can show that if we look for $K_i = X_i^* \tilde{K}_i$ where \tilde{K}_i is a $l \times l$ matrix such that $\tilde{K}_i = U_i \tilde{D}_i U_i^*$ then the t diagonal coefficients of \tilde{D}_i which are not necessarily 0 are given by

$$\tilde{d}_{i,j} = \begin{cases} \frac{1}{|\sigma_{i,j}|^2} \left(1 - \sqrt{\frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2}} \right) & \text{if } \sigma_{i,j} \neq 0 \\ 0 & \text{if } \sigma_{i,j} = 0 \end{cases}$$

for $j = 1, \dots, t$.

Using the MCEP estimator for i.i.d. fading, the decision statistic becomes

$$\mu(Y) = \text{etr} \left(\sigma^2 \left(\Lambda_1^{-1} - \Lambda_0^{-1} \right) Y Y^* \right) \quad (15)$$

with decision rule

$$\mu(Y) \underset{1}{\overset{0}{>}} |\Lambda_0|^{r\sigma^2} |\Lambda_1|^{-r\sigma^2} \quad (16)$$

which is equivalent to the optimal decision rule (5). To summarise, optimal detection can be written as

$$\frac{|\Lambda_1|^{r\sigma^2} e^{-\|Y - X_0 K_0 Y\|^2}}{|\Lambda_0|^{r\sigma^2} e^{-\|Y - X_1 K_1 Y\|^2}} \underset{1}{\overset{0}{>}} 1. \quad (17)$$

From Theorem 2, we see that the channel estimator that retains optimality is not an MMSE one, i.e. minimising the channel estimation error does not result in minimising the decoder error rate. Note that the channel estimator that allows optimal detection might not be unique since, in the proof of Theorem 2,

we restrict ourselves to an estimator of the form $K_i = V_i D_i U_i^*$ where U_i and V_i are the unitary matrices of the SVD of X_i . We show on Fig. 1 the word error rates obtained by simulation using an MMSE, MCEP for i.i.d. fading (optimal) and ZF channel estimator. The parameters of this simulation are: $l = 5$, $t = 3$ and $r = 2$. X_0 and X_1 are chosen arbitrarily among the non-unitary codewords of size $l \times t$. Note that the SNR is in logarithmic scale for better readability only.

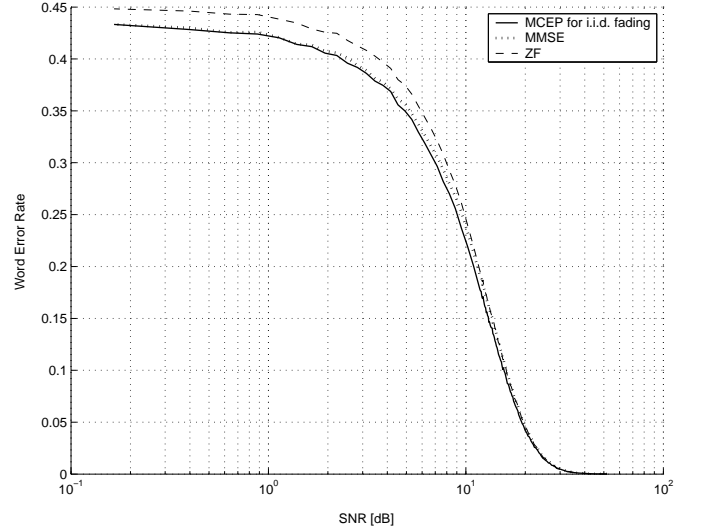


Fig. 1. Estimation-detection WER for MMSE, MCEP for i.i.d. fading (optimal) and ZF estimators.

D. Comparison of the MMSE, MCEP for I.I.D. Fading and ZF Estimators

We can compare the results we get for the MMSE, MCEP for i.i.d. fading and ZF estimators. Using the above SVD decomposition of X_i and writing the estimators as $K_i = V_i D_i U_i^*$, we get the following diagonal coefficients for D_i :

$$d_{(i,j),MMSE} = \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \alpha_{MMSE} \quad (18)$$

$$d_{(i,j),MCEP_{iid}} = \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \alpha_{MCEP_{iid}} \quad (19)$$

$$d_{(i,j),ZF} = \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \alpha_{ZF} \quad (20)$$

where

$$\begin{aligned} 0 &\leq \alpha_{MMSE} = \left(1 - \frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2} \right) \\ &\leq \alpha_{MCEP_{iid}} = \left(1 - \sqrt{\frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2}} \right) \\ &\leq \alpha_{ZF} = 1 \end{aligned}$$

and $j = 1, \dots, t$.

If we look at how much the power of the noise is taken into account compared to the signal itself in these three channel estimators, the above representation shows that the behaviour of

the MCEP channel estimator for i.i.d. fading is in between the MMSE and ZF case. More generally, these three estimators can be seen as elements of a class of estimators of the form

$$d_{(i,j),n} = \frac{\sigma_{i,j}^*}{|\sigma_{i,j}|^2} \left(1 - \left(\frac{\sigma^2}{|\sigma_{i,j}|^2 + \sigma^2} \right)^{1/n} \right) \quad (21)$$

where n is some positive number set to 1 for MMSE, 2 for MCEP and that tends to 0 for ZF. In this representation ZF appears as a limiting case. We show on Fig. 2 the channel estimation mean square error obtained by estimators corresponding to different values of n . Here $l = 5$, $t = 3$, $r = 1$ and some arbitrary training sequence of size $l \times t$ is used. Interestingly, the MCEP channel estimator for i.i.d. fading performs worse than ZF at high SNR while still retaining optimal detection. Note that the ZF is only partially represented to better compare the different curves.

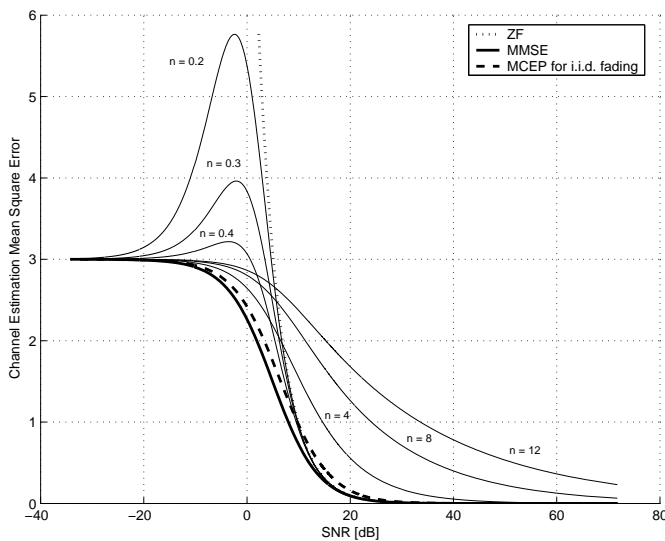


Fig. 2. Channel estimation mean square error for different values of n .

E. Extension to Transmit Correlated Fading

Under transmit correlated fading, the channel matrix can be written as

$$H = P_t^{1/2} \tilde{H} \quad (22)$$

where the element of \tilde{H} are i.i.d. circularly symmetric Gaussian with unit variance and P_t is the transmit correlation matrix. In this case, the GLRT does not preserve optimality. Nevertheless, it has been shown [13] that when the codewords are unitary, the GLRT achieves asymptotically the same decoding error probability as the ML receiver. The case of transmit correlated fading is equivalent to an i.i.d. fading case where the codewords X_i are simply replaced by $X_i P_t^{1/2}$. Thus, we can apply Theorem 2 using these new codewords and find the corresponding estimator that preserves optimal detection.

Future work will try to extend these results for both transmit and receive correlated fading.

IV. USE OF A TRAINING SEQUENCE

In the previous sections, we have considered the case where the channel can only be estimated conditioned on the observation Y . In this case, we have shown that the MMSE estimate of the channel does not provide in general optimal decoding error probability performance.

In this section, we now consider the use of a training sequence. In other words, we consider block codes which share a common prefix A , referred to as the training sequence. Again, assuming equiprobable codewords, optimal detection is done using ML decoding on the whole codeword, i.e. including A . In other words, no channel estimation using A as a training sequence followed by detection can outperform ML decoding in terms of WER. Another way to see it is that ML decoding makes the best use of the training sequence A in terms of minimising the error probability.

Let us assume that the codewords are of the form

$$X_i = \begin{bmatrix} A \\ B_i \end{bmatrix}$$

where the preamble or training A is a $l_1 \times t$ matrix common to all codewords and B_i is a $l_2 \times t$ matrix that depends on the codeword. Note that $l = l_1 + l_2$.

Let us partition the observation Y into the segments corresponding to A and B ,

$$Y = \begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$$

where

$$Y_A = AH + N_A \quad (23)$$

$$Y_B = BH + N_B. \quad (24)$$

Let us suppose that we use the training sequence A to compute a channel estimate \hat{H} which in turn is used in an ML decoding process as if it were the true channel realisation. A natural question that arises is: "Do we perform better if we also take into account the training sequence in the decoding process?". The following theorem shows that the word error probability remains unchanged.

Theorem 3: Suppose that we perform coherent ML decoding using the estimate \hat{H} as if it were the true channel realisation. Then, using the training sequence A in the decoding process does not change the word error rate.

Proof: Assuming \hat{H} to be the true channel realisation, the ML metric is given by $\|Y - X_i \hat{H}\|^2$ when the whole codeword is used and by $\|Y_{B_i} - B_i \hat{H}\|^2$ when the training sequence A is not taken into account. Detection is done by minimizing the

corresponding metric over all possible codewords. We have that

$$\begin{aligned}
& \|Y - X_i \hat{H}\|^2 \\
&= \text{tr} \left((Y - X_i \hat{H})(Y - X_i \hat{H})^* \right) \\
&= \text{tr} \left(\left(\begin{bmatrix} Y_A \\ Y_{B_i} \end{bmatrix} - \begin{bmatrix} A \\ B_i \end{bmatrix} \hat{H} \right) \left(\begin{bmatrix} Y_A \\ Y_{B_i} \end{bmatrix} - \begin{bmatrix} A \\ B_i \end{bmatrix} \hat{H} \right)^* \right) \\
&= \text{tr} \left(\begin{bmatrix} Y_A - A\hat{H} \\ Y_{B_i} - B_i\hat{H} \end{bmatrix} \begin{bmatrix} Y_A - A\hat{H} \\ Y_{B_i} - B_i\hat{H} \end{bmatrix}^* \right) \\
&= \text{tr} \left(\begin{bmatrix} Y_A - A\hat{H} \\ Y_{B_i} - B_i\hat{H} \end{bmatrix} \begin{bmatrix} Y_A^* - \hat{H}^* A^* & Y_{B_i}^* - \hat{H}^* B_i^* \end{bmatrix} \right) \\
&= \text{tr} \left((Y_A - A\hat{H})(Y_A^* - \hat{H}^* A^*) \right) \\
&\quad + \text{tr} \left((Y_{B_i} - B_i\hat{H})(Y_{B_i}^* - \hat{H}^* B_i^*) \right) \\
&\equiv \text{tr} \left((Y_{B_i} - B_i\hat{H})(Y_{B_i}^* - \hat{H}^* B_i^*) \right) \\
&= \|Y_{B_i} - B_i\hat{H}\|^2
\end{aligned}$$

where the fifth equality holds since $(Y_A - A\hat{H})(Y_A^* - \hat{H}^* A^*)$ (resp. $(Y_{B_i} - B_i\hat{H})(Y_{B_i}^* - \hat{H}^* B_i^*)$) is a $l_1 \times l_1$ (resp. $l_2 \times l_2$) matrix and the fact that the trace of a matrix is equal to the sum of its diagonal elements. Thus, these two metrics are equivalent and the ML detection is the same in both cases. The WER remains thus unchanged. ■

Another way to see Theorem 3 is that when the channel is assumed to be known, common information to all codewords turns out to be completely irrelevant. This is due to the fact that once the channel is known, minimising the probability of error is equivalent to maximising the distance between the codewords. And adding common information to every codeword does not increase this distance. Therefore, re-using the training sequence in the detection process does not improve the WER.

Figure 3 shows the WER results using four different methods. The parameters of this simulation are: $l_1 = 5$, $l_2 = 5$, $t = 3$ and $r = 2$. A , B_0 and B_1 are chosen arbitrarily. Note that the SNR is in logarithmic scale for better readability only. The first curb corresponds to optimal ML detection applied to the whole codeword. The three other curbs show the WER when first a channel estimation is done based on the training sequence using the MMSE, the MCEP and the ZF channel estimators and then ML detection is performed as if this estimate were the true channel realisation. Not surprisingly, the best one corresponds to the ML (or non-coherent approach). One can clearly see that the WERs corresponding to MMSE and MCEP estimates of the channel provide poorer results with the WER for the ZF method providing the worst results of the four methods. It is worth recalling that the MCEP method here does not lead to optimality because the channel estimate is obtained using the training sequence, and is not conditioned on the whole block code as it was in Section III-A.

V. CONCLUSION

The aim of this paper has been to compare coherent and non-coherent decoding schemes in the context of space-time block codes. We have in particular shown that optimal non-coherent

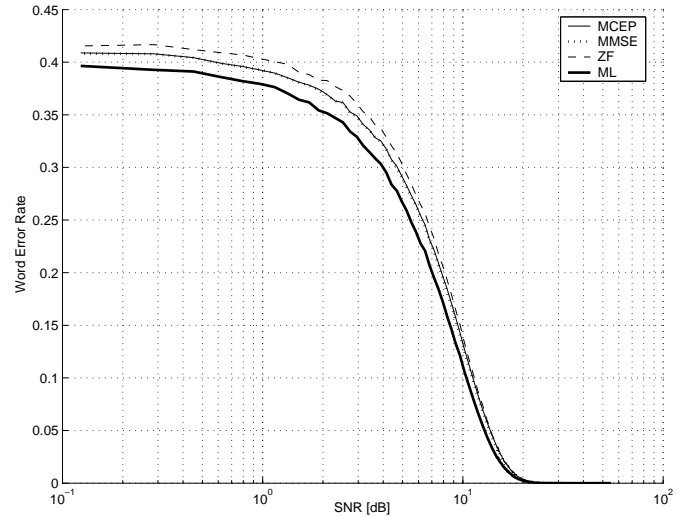


Fig. 3. WER for the optimal non-coherent ML detection and coherent ML detection using MMSE, MCEP and ZF channel estimate as the true channel realisation.

decoding can actually be decomposed in a “coherent” approach where a channel estimate is computed and used in the detection step. Under i.i.d. fading assumption, we have derived the expression for channel estimators that lead to optimal detection and proved that in the general case, these estimators do not correspond to an MMSE estimate of the channel. Therefore, we have shown that trying to find a channel estimate as close as possible to the actual channel realisation is not necessarily the best strategy to adopt in order to minimise the word error rate. Furthermore, we have related our approach to the GLRT and have proved that under i.i.d. fading and unitary codewords assumptions, the MMSE estimate leads to the optimal decision.

We have related the results found for the per-codeword processing approach to the case where a training sequence is used to estimate the channel. We have investigated the impact of this training sequence on the WER performance and showed that using it in both the channel estimation and the detection process does not improve the decoding error probability. We also compared the WER obtained with different channel estimators.

Finally, since we have derived expressions for channel estimators which ensure optimal decoding, it would be interesting as future work to investigate the performance of these channel estimates when applied in the context of Per-Survivor Processing (PSP) techniques.

REFERENCES

- [1] G.J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–159, Aug. 1996.
- [2] 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.
- [3] Siavash M. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE J. Selected Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
- [4] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999, Originally published as Tech. Memo., Bell Laboratories, Lucent Technologies, October 1995.

- [5] Vahid Tarokh, Nambi Seshadri and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, March 1998.
- [6] Thomas L. Marzetta and Bertrand M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
- [7] Bertrand M. Hochwald and Thomas L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 543–564, March 2000.
- [8] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.
- [9] Lizhong Zheng and D.N.C. Tse, "Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, 2002.
- [10] Vahid Tarokh, Ayman Naguib, Nambi Seshadri and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility and multiple paths," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 199–207, Feb. 1999.
- [11] A. Scaglione, P. Stoica, S. Barbarossa, G. Giannakis and H. Sampath, "Optimal designs for space time linear precoders and decoders," *IEEE Trans. Sig. Proc.*, vol. 50, no. 5, May 2002.
- [12] J. K. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh flat fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, 1991.
- [13] Matthias Brehler and Mahesh K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2383–2399, Sep. 2001.
- [14] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I*. New-York: Wiley, 1968.