A Comparison Analysis of Pre-filtering Methods in Space-Time Trellis-Based Reduced-State Turbo-Detection

Sami Chtourou†, Raphaël Visoz‡ and A.O. Berthet‡

†ISECS, route Menzel Chaker, km 0.5, Sfax, Tunisie
‡France Télécom, 38-40, rue du Général Leclerc, 92794 Issy les Moulineaux, France
§SUPELEC, Department of Telecommunications, Plateau de Moulon, 91192 Gif-sur-Yvette, France

Abstract

In this paper, we investigate two different prefiltering strategies, i.e., MIMO channel shortening and MIMO whitened matched filtering, to decode iteratively a STBICM transmitted over MIMO block fading ISI Channels using a trellis based post-detector. Indeed, resorting to such front-ends allows significant complexity gain when compared to the optimal MAP-based turbo-receiver. While those two approaches have been exhaustively visited in the past, their fair comparison in terms of trade-off between complexity and performance has - to our knowledge - never been tackled before. This paper aims at giving some valuable insights about the respective benefits of these two prefiltering methods with respect to the channel characteristics and overall receiver complexity.

Keywords: MIMO ISI channels, turbo-receivers, trellis-based detectors, channel shortening, whitened matched filtering
I. INTRODUCTION

I.1 Research context

The main drawback of STBICM lies in the complexity involved at the receiver side. As well-known, optimal joint decoding of STBICM is an intractable issue. An efficient way of approaching it is to divide the problem into two distinct functions, namely MIMO ISI detection and channel decoding, and to apply the turbo-principle [1]. Unfortunately, the complexity of APP MIMO ISI detection is directly related to the number of channel trellis states, which grows exponentially with the number of transmit antennas and with the channel memory. To resort to trellis state reduction and Per Survivor Processing (PSP) techniques enables to convert part of this memory exponential dependency into linear dependency (see [2] and the references therein). As long as the number of transitions per state does not explode, a class of qualizers able to approach the MAP criterion with much lower complexity is available. Contrary to linear equalizers based on MMSE criterion, whose stability is highly dependent on the matrix rank of the MIMO system, such trellis-based detectors (without pre-filtering) accept any kind of transmit and receive antenna configurations [2] [3]. However, for sharp channel dispersion, error propagation induced by massive trellis state reduction, albeit partially compensated by list-type PSP [4] [2] [3], entails the need of reshaping the channel impulse response by prefiltering.

I.2 In this paper

The purpose of this paper is to derive and analyze two different approaches proposed in the literature and based on channel prefiltering. Their complexity and performance will be compared in the context of STBICM transmitted over MIMO block fading ISI channel. The shortening approach proposes to design a receive prefilter which concentrates the energy of the channel in a small number of taps at the price of a colored noise. The equalizer sees a shortened channel and thus needs a smaller number of states. The advantage of this method lies in the fact that it allows MAP post-detection algorithm. As a result, the shortened channel does not need to be minimum phased. The second approach uses a sub-optimum trellis-based equalizer associated with a MIMO Whitened Matched Filter (WMF) concentrating the energy on the first channel tap. Unlike the shortening approach, the noise at the output of such prefilter remains white in principle.

The paper is organized as follows: the communication model is briefly introduced in Section II. We review then in Section III the first possible approach relying on MIMO channel shortening and post optimal MIMO MAP ISI detection of the shortened channel. Section IV develops another approach, using a MIMO Mean-squared Whitened Matched Filter (MS-WMF) front-end followed by reduced-state trellis-based MIMO ISI detection. In Section V, we compare both approaches’ complexity. Some simulation results and comments related to their performance are finally presented.

II. COMMUNICATION MODEL

II.1 Space-time bit interleaved coded modulation

Let $C$ be a linear code of dimension $k_c$, length $n_c$ and rate $r_c$ over $\mathbb{F}_2$ which transforms a message $m \in \mathbb{F}_2^{k_c}$ into a code word $c \in \mathbb{F}_2^{n_c}$. The produced code word enters a semi-random bitwise interleaver $\mathcal{I}$, whose output is segmented into a collection $D = \{D^b : b = 0, \ldots, N_B - 1\}$ of $N_B$ binary matrices $D^b$ of dimension $qN_T \times L$, where $q$ is the number of bits per constellation.
symbol per transmit antenna and $L$ the block length in channel use (c.u). The interleaver design meets two different rules: first, consecutive coded bits must be diagonally distributed among the $N_T \times N_T$ fading blocks for properly exploiting the available space-time diversity [5]. Second, consecutive coded bits must also be distributed among different channel uses in order to maximize the girth of the underlying STBICM factor graph [6] (cf. Section III) and, thus, to ensure a better convergence in iterative decoding (see, for example, [7] for the fading flat subcase). Columns of matrices $D^b$ are vectors $d^b_k \in \mathbb{F}_2^{N_T}$, $k = 0, \ldots, L - 1$ containing one subvector $d^b_{t,k}$ per channel input $t = 0, \ldots, N_T - 1$ with $q$ stacked binary components. Within each column $d^b_k$, all subvectors $\{d^b_{t,k}\}$ are mapped, through a labeling rule $\varphi_t : \mathbb{F}_2^{q} \to \mathcal{X} \subseteq \mathbb{C}$, into a complex symbol $x^b_{t,k}$ belonging to the complex signal set $\mathcal{X}$ of cardinality $|\mathcal{X}| = 2^q$ (average normalized energy). Note that we assume identical constellations on each transmit antenna, but authorize different labeling rules. After signal mapping, the matrix symbol digit $D^b$, is thus transformed into a complex matrix $X^b$ of dimension $N_T \times L$. Coding, interleaving, and per-antenna signal mappings can be regarded as a global coding-modulation process $\Psi : \mathbb{F}_2^{N_T \times L} \to \mathcal{S} \subseteq \mathcal{X}^{N_T \times L \times N_B}$ which maps the binary message $m$ into the codeword $f \in \mathcal{S}$. Since the labeling functions $\varphi_t$ are, in general, non-linear, so is $\Psi$. Falling into the general class of STCs, this architecture offers a spectral efficiency $\eta = \rho, q L$ bits per channel use (b/c.u) under ideal Nyquist band-limited filtering assumption.

II.2 MIMO block fading multipath channel

Let $H^b \in \mathbb{C}^{N_R \times N_T \times (M+1)}$ denote the MIMO fading multipath channel block $b$ and $\mathbf{H} = \{H^b : b = 0, \ldots, N_B - 1\}$ the collection of all channel blocks. Let $\mathbf{Y} = \{Y^b \in \mathbb{C}^{N_R \times L} : b = 0, \ldots, N_B - 1\}$ be the collection of all received matrices. The discrete-time base-band equivalent vector channel output $y^b_k \in \mathbb{C}^{N_R}$ at time $k = 0, \ldots, L - 1 + M$ can be written as

$$
y^b_k = \sum_{m=0}^{M} H^b_{m} x^b_{k-m} + w^b_k
$$

where $x^b_k \in \mathcal{X}^{N_T}$ is the vector constellation symbol transmitted at time $k$, $H^b_{m} \in \mathbb{C}^{N_R \times N_T}$ is the $m^{th}$ matrix tap of the channel Finite Impulse Response (FIR), $w^b_k \in \mathbb{C}^{N_R}$ is the vector of additive complex noise. The vectors of additive complex noise $w^b_k$ are assumed random independent identically distributed (i.i.d) zero-mean circularly-symmetric complex Gaussian and thus follow the pdf $\mathcal{CN} \left(0_{N_R}, \Theta_w = \sigma^2 I_{N_R} \right)$. Channel $H^b$, constant along the corresponding block duration $L$ (in c.u), has FIR of length $M + 1$, whose symbol-spaced taps $H^b_{0}, \ldots, H^b_{M}$ are $N_R \times N_T$ matrices with zero-mean circularly-symmetric complex Gaussian entries satisfying the normalization mean power constraint

$$
E \left[ \text{diag} \left( \sum_{m=0}^{M} H^b_{m} H^b_{m}^H \right) \right] = N_T I_{N_R}
$$

Due to the absence of CSI at transmitter, we further assume an equal-power system, i.e.,

$$
\Theta_{x^b_k} = E \left( x^b_k x^b_k^H \right) = I_{N_T}
$$
III. Turbo-detection with a MIMO channel shortening front-end

We present in this section an approach [8] which does not alter the BCJR MAP ISI detector while significantly reducing its complexity. This is made possible thanks to the Finite Impulse response (FIR) MIMO channel shortening front end concentrating the energy of the impulse response in a small number of adjacent taps. Thus, the cascade of the original MIMO channel and the prefilter can be viewed as a MIMO channel having $N_T$ transmit antennas, $N_{R_s}$ receive antennas and a smaller memory $M_s$. It is quite evident that taking $M_s < M$ reduces the number of the channel trellis states by a factor of $Q^{N_T(M-M_s)}$ since the complexity of the BCJR MAP equalizer (directly related to the number of the channel trellis states) grows exponentially not only with the number of transmit antennas and the constellation size but also with the channel memory. Moreover, we will see in Section III.2 how a part of the loss in multipath diversity due to channel shortening can be recovered via a higher level of virtual receive antenna diversity. Nevertheless, the main impairment of such approach lies in the fact that the noise at the output of the shortening front-end becomes temporally colored. We move now to detailing the prefilter computation.

III.1 Architecture of the associated turbo-receiver

As mentioned above, the shortening approach does not affect the BCJR MAP ISI detector. Thus, the architecture of the associated trellis based turbo-receiver can be easily depicted in Figure 1.

III.2 MIMO Channel Shortening

For each block $b$, in order to shorten the effective block channel realization $H^b$, the $N_{R_s}$-dimensional prefilter $F^s$ having $(l_p + 1)$ coefficients is applied to the received signal matrix $Y^b$. We choose to implement the shortening prefilter adopting a sliding window model (instead of a more optimal block model) for complexity reason. As the treatment is identical for all blocks, we omit index $b$ whenever it is possible for the sake of simplicity.

III.2.1 The MIMO channel Shortening front-end derivation: The resulting filtered signal $z_k \in \mathbb{C}^{N_{R_s}}$ at time $k$, $0 \leq k \leq L - 1$, is simply given by

$$z_k = F^s \overline{y}_k$$

where $\overline{y}_k$ is the stacked vector

$$\overline{y}_k = \left[ y_{k+l_p}^T \cdots y_k^T \right]^T \in \mathbb{C}^{N_{R_s}(l_p+1)}$$

and where

$$F^s = \left[ F_{-l_p}^s \cdots F_0^s \right] \in \mathbb{C}^{N_{R_s} \times N_{R_s}(l_p+1)}$$

denotes the prefilter coefficients gathered in a matrix form. Similarly, the shortened channel $H_s$ is defined by its matrix

$$H^s = \left[ H_0^s \cdots H_{M_s}^s \right] \in \mathbb{C}^{N_{R_s} \times N_T(M_s+1)}.$$
We introduce then the error vector $e_{k,\tau} \in \mathbb{C}^{N_{R_b}}$ defined as

$$e_{k,\tau} = z_k - H^* e_{k,\tau}$$ (5)

with $x_k \in \mathbb{C}^{N_T(l_p+M+1)}$ being the stacked vector

$$x_k = \begin{bmatrix} x_{k+l_p}^T \cdots x_{k-M}^T \end{bmatrix}^T.$$

The prefilter $F^*$ is optimized to minimize the Mean Square Error (MSE)

$$\delta_\tau = \text{tr} (\Theta_{e,\tau}).$$ (6)

where the $N_{R_b} \times N_{R_b}$ matrix $\Theta_{e,\tau}$

$$\Theta_{e,\tau} = E \left( e_{k,\tau} e_{k,\tau}^\dagger \right)$$ (7)

corresponds to the autocorrelation matrix of the error vector $e_{k,\tau}$.

The orthogonal projection theorem states that the error is orthogonal to the observed data, i.e.,

$$E \left( e_{k,\tau} y_k^\dagger \right) = 0_{N_{R_b} \times N_R(l_p+1)}.$$ (8)

This leads to the following relationship between the augmented shortened channel and the prefilter

$$F^* = H_e^* \Theta_{x,y} \Theta_{x}^{-1}.$$ (9)

In (9), the $N_T (l_p + M + 1) \times N_R (l_p + 1)$ matrix $\Theta_{x,y} = E \left( x_k y_k^\dagger \right)$ and the $N_R (l_p + 1) \times N_R (l_p + 1)$ matrix $\Theta_y = E \left( y_k y_k^\dagger \right)$ denote respectively the input-output cross correlation and the output autocorrelation matrices. Since the noise and data samples are uncorrelated, we obtain

$$\Theta_{x,y} = \Theta_{x} H^\dagger$$ (10)

and

$$\Theta_y = H \Theta_{x} H^\dagger + \Theta_w$$ (11)

where $\Theta_w = E \left( w_k w_k^\dagger \right)$ is the $N_R (l_p + 1) \times N_R (l_p + 1)$ autocorrelation matrix of the stacked noise vector

$$w_k = \begin{bmatrix} w_{k+l_p}^T \cdots w_k^T \end{bmatrix}^T \in \mathbb{C}^{N_T(l_p+1)}$$

and where $H$ corresponds to the $N_R (l_p + 1) \times N_T (l_p + M + 1)$ sylvester channel matrix

$$H = \begin{bmatrix} H_0 \cdots & H_M & 0_{N_R \times N_T} \cdots & 0_{N_R \times N_T} \\ 0_{N_R \times N_T} & H_0 \cdots & H_M \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots \\ 0_{N_R \times N_T} & \cdots & 0_{N_R \times N_T} & H_0 \cdots & H_M \end{bmatrix}.$$

Inserting then (10) and (11) in (9) yields

$$F^* = H_e^* \Theta_{x,y} H^\dagger \left( H \Theta_{x} H^\dagger + \Theta_w \right)^{-1}.$$ (12)
When further resorting to the matrix inversion lemma, the expression (12) of the prefilter turns out to
\[
F^s = H_r^s (H_r^\dagger \Theta_w^{-1} H_r + \Theta_x^{-1})^{-1} H_r^\dagger \Theta_w^{-1} .
\] (13)

Expression (13) shows that the optimum MIMO prefilter in the MMSE sense includes a noise whitening filter \( \Theta_w^{-1} \) followed by a matched filter \( H_r^\dagger \) in addition to the shortening component.

On the other hand, we see from the orthogonal projection theorem and from (13) that the expression of the error autocorrelation matrix defined by (7) becomes
\[
\Theta_{e_r} = -E \left( e_{k,r} x_k^\dagger \right) H_r^\dagger
= H_r^\dagger \Theta_x H_r^\dagger - F H \Theta_x H_r^\dagger
= H_r^\dagger \left( \Theta_x - \Theta_x H_r^\dagger (H_r \Theta_x H_r^\dagger + \Theta_w)^{-1} H_r \Theta_x \right) H_r^\dagger
\] (14)

Applying again the matrix inversion lemma on (14) yields
\[
\Theta_{\Delta r} = H_r^s (H_r^\dagger \Theta_w^{-1} H_r + \Theta_x^{-1})^{-1} H_r^s^\dagger
= H^s \Theta_r H^s^\dagger
\] (15)

III.2.2 Computing the shortened channel impulse response: The performance of the BCJR MAP equalizer closely depends on the SNRs at the outputs
\[
z_k = H_r^s x_k + e_{k,r}
\] (16), 0 \( \leq k \leq L - 1 \), of the prefilter. Thus, each SNR \( \gamma_j, j = 0 \cdots N_{Rs} - 1 \), is simply given by the \( j \)-th diagonal entry of the \( N_{Rs} \times N_{Rs} \) matrix \( H_r^s \Theta_x H_r^s^\dagger \Theta_{e_r}^{-1} \). The optimum shortened channel impulse response is derived in order to optimize those SNRs.

When assuming infinitely deep space-time interleaving, time independence between components of \( x_k \) holds\(^1\). Thus, taking into account the power allocation strategy chosen, i.e.,
\[
E \left( x_k x_k^\dagger \right) = I_{N_T},
\]
yields
\[
\Theta_x = I_{N_T(M + M + 1)}.
\]
As a consequence, each SNR \( \gamma_j \) becomes
\[
\gamma_j = \left[ H_r^s H_r^s^\dagger \Theta_{e_r}^{-1} \right]_{jj,j} .
\] (17)

The design criteria of the shortened channel is then strictly equivalent to
\[
\text{tr} \left\{ H_r^s H_r^s^\dagger \Theta_{e_r}^{-1} \right\} \rightarrow \text{max} .
\] (18)

We impose the constraint
\[
H^s H^s^\dagger = H^s H^s^\dagger = I_{N_{Rs}}
\] (19)
to exclude the trivial solution \( H^s = 0_{N_{Rs} \times N_T(M + 1)} \) since such constraint results in higher SNRs than the more conventional constraint \( H_m^s = I \) for any \( m, 0 \leq m \leq M_s \) [9]. Using (15) and (19) simplifies the design criteria to
\[
\text{tr} \left\{ \Theta_{e_r}^{-1} \right\} = \text{tr} \left\{ H^s \Theta_r H^s^\dagger - 1 \right\} \rightarrow \text{max} .
\] (20)

\(^1\)In practice, this independence property is proved even for finite (relatively small) interleaver depths
Let us now define the $e_j \in \mathbb{C}^{N_{R_s}}$ to be the $j^{th}$ unit vector having a one in its $j^{th}$ component and zeros elsewhere. Based on (15), the $j^{th}$ diagonal entry of $\Theta_{e,\tau}$ can be written as follows:

$$[\Theta_{e,\tau}]_{j,j} = e_j^\dagger H^* \Theta_{\tau} H^e_j = h^{s,j} \Theta_{\tau} h^{s,j\dagger}$$

with the $j^{th}$ row $h^{s,j}$ of $H^e$ being unit norm since $h^{s,j} h^{s,j\dagger} = e_j^\dagger H^* H^e_j = 1$. In order to satisfy the design criteria specified by (20), $h^{s,j}$ should minimize $[\Theta_{e,\tau}]_{j,j}$ for all $0 \leq j \leq N_{R_s} - 1$ subject to the constraint $h^{s,j} h^{s,j\dagger} = 1$. Therefore, the rows of $H^e$ have to be chosen as the eigenvectors of $\Theta_{\tau}$ corresponding to its $N_{R_s}$ smallest eigenvalues $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N_{R_s} - 1}$. Assuming this choice, it is quite evident that matrix $\Theta_{e,\tau}$ becomes diagonal and given by

$$\Theta_{e,\tau} = \text{diag}\{ \lambda_0 \lambda_1 \cdots \lambda_{N_{R_s} - 1} \}.$$  

(22)

The relative delay $\tau$ is then optimized by choosing the matrix $\Theta_{\tau}$ that allows maximizing

$$\sum_i 1/\lambda_i \rightarrow \text{max}.$$ 

It is important to notice here that we assume that the relative delay $\tau$ is the same for all transmit antennas. Lower MSE may be achieved by allowing a variable relative delay across channel inputs at the expense of increased computational complexity [9].

Since $\Theta_{\tau} \in \mathbb{C}^{N_T(M_s + 1) \times N_T(M_s + 1)}$, the MIMO shortened channel $H^e$ has a maximum of

$$N_{R_s} \leq N_T (M_s + 1)$$

outputs. Then, the shortened channel can have a larger number of outputs than the original one, assuming that

$$N_R < N_T (M_s + 1).$$

We can present this as a transformation of multipath diversity to receive antenna diversity [8]. We emphasize that taking $N_{R_s} > N_R$ does not increase the number of states in the MAP ISI detector. Indeed, only the branch metric calculation complexity grows up slightly as the number of receive antennas increases from $N_R$ to $N_{R_s}$.

On the other hand, the noise at the shortened channel output is spatially white as the matrix error autocorrelation matrix is now diagonal. However, the noise is still temporally colored, which conceptually moves this method from the detection theory. Better results are inevitably obtained when incorporating knowledge of the temporal noise autocorrelation matrix in the MAP ISI detector metric calculation. Nevertheless, we witness that neglecting temporal coloration has a limited effect on achieved performance [8].

III.2.3 MIMO channel shortening complexity: In this section, we enumerate the most complex tasks needed for deriving the shortened channel and the shortening prefilter, namely computing $(H^e \Theta^{-1} \Theta^e + \Theta^{-1})^{-1}$ and the single value decomposition of $\Theta_{\tau}$ for each possible delay $\tau$. Thanks to its block-Toeplitz structure, efficient algorithms can be used to invert $(H^e \Theta^{-1} \Theta^e + \Theta^{-1})$ with a complexity of $O(N_T^3 (l_p + M + 1)^2)$ [10]-[11]-[12]. The computation of the single value decomposition of $\Theta_{\tau}$ requires $O(N_T^3 (M_s + 1)^3)$ operations. Then, an estimate of the shortening approach complexity can be

$$(l_p + 1 + M - M_s) \left[ O(N_T^3 (l_p + M + 1)^2) + O(N_T^3 (M_s + 1)^3) \right]$$

operations.
III.3 Influence of the prefilter order

As mentioned above, we implement here the shortening approach adopting a sliding window model instead of the optimal block model. The, we try to see the influence of the shortening prefilter order on the performance of the turbo-receiver described in this section. Then, we consider a STBICM using a 4-PSK modulation. Transmission occurs over $2 \times 2$ block static channels. The shortened channel memory is set to $M_s = 1$ when the number of the virtual receive antennas is $N_{R_v} = 4$. The employed code is a rate-$1/2$ 64-state non-recursive convolutional code with $d_{free} = 10$ yielding a spectral efficiency of $\eta = 2$ bits p.c.u. The transmitted symbol block length is $L = 256$. In Figure 2, we address the transmission over a block-static $2 \times 2$ MIMO channel having 5 EQual Energy taps (EQ5 channel). The shortening turbo-receiver performs practically the same for $l_p = 10$ and $l_p = 15$. However, a degradation of near 2 dB appears at BLER $10^{-2}$ when decreasing the prefilter order to $l_p = 5$. In the same way, we reconsider in Figure 3 the same transmission scheme except that the channel is now a block-static $2 \times 2$ EQ-10 channel and that the prefilter order is taken respectively equal to $l_p = 15$, $l_p = 20$ and $l_p = 25$. The gap between the case $l_p = 15$ and both $l_p = 20$ and $l_p = 25$ is now near 1 dB at BLER $10^{-2}$. To conclude, it is important to keep in mind that when we increase the prefilter order, the shortening approach becomes more complex (see Section III.2.3). At the same time, keeping the prefilter order too low may result in performance degradation. Thus, we can say that this parameter choice plays here a key role in reaching a good trade-off between performance and complexity.

IV. Turbo-detection with a MIMO Whitened Matched Filter (WMF) front-end

In this section, we detail another possible way to alleviate the complexity of the optimal MAP ISI detector which consists in restricting all compound ISI trellises to sub-trellises. This is done by simply truncating the overall channel memory $M$ to an arbitrary reduced memory $M_r \leq M$ and by trying to recover the resulting sub-optimality via per-survivor processing (PSP) [13]-[14]. The complexity of the ISI detection task is approximately reduced by a factor of $Q^{N_{R_v}(M-M_r)}$

In order to fight back the well known resulting error propagation effect, the PSP technique normally requires the channels from all transmit antennas to all receive antennas to be minimum phase simultaneously [15]. Unfortunately, such a property is almost impossible to be met [16]. Instead, we propose to resort to the MIMO Whitened Matched Filter (WMF) which makes the MIMO channel minimum-phase, i.e., concentrates the energy in the first tap, while keeping a gaussian noise spatially and temporally white (contrary to the shortening approach). It is important to recall here that the MIMO channel is said minimum phase when the roots of the determinant of its $\mathcal{Z}$-transform are within the unit circle.

IV.1 Architecture of the associated turbo-receiver

The architecture of the turbo-receiver inspired from [17] and depicted in Figure 4 is based on the structure of the optimal MIMO receiver presented in [18] (see also [19] for the single transmit antenna case).

Since the Delayed Decision-Feedback Sequence Estimator (DDFSE) associated with pre-filtering is shown to ensure a reasonable trade-off between performance and complexity [13]-[20], a modified low complexity Soft Input Soft Output DDFSE-based detector (SISO DFSE)
[21] is used for iterative ISI detection. We witness that such association (MIMO WMF with SISO DFSE) represents a good alternative to the Generalized Soft Viterbi Algorithm (GSVA) [4] which compensates the error-propagation induced to the PSP by retaining more than one survivor path per state without resorting to a prefilter front-end [22]-[3]. This is perfectly illustrated in Figure 5 and Figure 6 where the performance of the proposed turbo-receiver with and without prefiltering is presented. We consider here a STBICM with a spectral efficiency $\eta = 2$ bits p.c.u using a 4-PSK modulation and a rate $1/2$ 64-state non-recursive convolutional code (with $d_{f, r e c} = 10$). The transmission occurs over $2 \times 2$ block static MIMO channels with respectively 3 and 5 Equal energy taps when the DFSE works with a reduced memory $m_r = 1$. The gain brought by prefiltering at BLER $10^{-2}$ is near $7$ dB for the EQ3 channel and reaches $14$ dB at the same BLER with the EQ5 channel. Such performance improvement shows the efficiency of the association between the whitening matched filter front-end and the DFSE.

IV.2 The Whitened Matched Filter

IV.2.1 Principle: For each channel block $b$, the $Z$-transform of the received signal can be written as

$$y^b(z) = H^b(z)x^b(z) + w^b(z) \quad (23)$$

where $x^b(z)$, $w^b(z)$ and $H^b(z) = \sum_{i=0}^{M} H^b_i z^{-i}$ denote the respective $Z$-transforms of the transmitted data, the noise samples and the block channel impulse response. When fed to the matched filter $H^b(\bar{z}^{-1})$, the received signal becomes

$$H^b(\bar{z}^{-1})y^b(z) = H^b(\bar{z}^{-1})x^b(z) + H^b(\bar{z}^{-1})w^b(z)$$

$$= \Theta_{ZF}^b(z)x^b(z) + w^b_c(z). \quad (24)$$

The noise $w^b_c(z)$ is now colored with spectrum $\Theta_{ZF}^b(z) = H^b(\bar{z}^{-1})H^b(z)$.

The spectral factorization theorem [23]-[24]-[25] states that, as soon as $N_R \geq N_T$, there exists a factorization of $\Theta_{ZF}^b(z)$

$$\Theta_{ZF}^b(z) = H^b(\bar{z}^{-1})H^b(z) = H^b_{min,ZF}(\bar{z}^{-1})H^b_{min,ZF}(z) \quad (25)$$

such as the minimum phase factorization $N_T \times N_T$ matrix filter $H^b_{min,ZF}(z)$ of $\Theta_{ZF}^b(z)$ is causal, stable and has a causal inverse. The $N_T \times N_T$ anticausal filter (if it exists)

$$G^b(z) = \left(H^b_{min,ZF}(\bar{z}^{-1})\right)^{-1} \quad (26)$$

is a whitening filter for a process with spectrum $\Theta_{ZF}^b(z)$. From (23) and (25), we see that applying the Zero-Forcing Whitened Matched Filter (ZF-WMF) $P^b_{ZF}(z)$ defined as

$$P^b_{ZF}(z) = G^b(z)H^b(\bar{z}^{-1}) \quad (27)$$

to the received signal $y^b(z)$ turns out

$$P^b_{ZF}(z)y^b(z) = H^b_{min,ZF}(z)x^b(z) + w^b_{ZF}(z). \quad (28)$$
It is quite evident from (24) and (25) that the resulting $N_T \times N_T$ noise vector

$$w^b_{w, ZF}(z) = P^b_{ZF}(z) w^b(z) = G^b(z) w^b_c(z)$$

becomes temporally white.

On the other hand, Gerstacker et al. provide in [26] a generalization of the well-known energy concentration property of minimum-phase equivalent systems in the SISO case to MIMO systems. Thus, the minimum-phase channel taps $H_{min, ZF,k}$, $0 \leq k \leq M$, verify

$$\sum_{k=0}^{M} \|H_{min, ZF,k}\|_F^2 \leq \sum_{m=0}^{M} \|H_m\|_F^2$$

(29)

where $\| \cdot \|_F$ denotes the Frobenius norm of a matrix. Thanks to this energy concentration property, prefiltering is clearly desirable for reduced state equalization. Indeed, only the front part of the channel impulse response is used in this case.

IV.2.2 Implementation of the WMF using the linear prediction theory: The Spectral Factorization theorem states that $P^b_{ZF}(z)$ may not exist if $\Theta^b_{ZF}(z)$ is singular on the unit circle. Thus, instead of implementing the ZF-WMF, we opt for the Mean-Squared Whitened Matched Filter (MS-WMF) [27] whose $Z$-transform is coined $P^b_{MS}(z)$. The expression of $P^b_{MS}(z)$ is identical with the $P^b_{ZF}(z)$ one given by relation (27) except that $G^b(z)$ becomes a whitening filter for a process with spectrum $\Theta^b_{MS}(z) = H^b(\cdot^{-1}) H^b(\cdot) + \sigma^2 I_{N_T}$.

(30)

Unlike the ZF-WMF, the MS-WMF always exists independently of the configuration of transmit and receive antennas; besides it approaches the ZF-WMF (if it exists) as the SNR goes to infinity. When applying the MS-WMF to the received signal, we obtain

$$P^b_{MS}(z) y^b(\cdot) = G^b(z) H^b(\cdot^{-1}) H(\cdot) \hat{x}(\cdot) + G^b(z) H^b(\cdot^{-1}) w^b(\cdot)$$

$$= G^b(z) (\Theta^b_{MS}(z) - \sigma^2 I_{N_T}) \hat{x}^b(\cdot) + G^b(z) H^b(\cdot^{-1}) w^b(\cdot)$$

$$= H^b_{min, MS}(\cdot) \hat{x}^b(\cdot) + w^b_{w, MS}(z)$$

(31)

where $H^b_{min, MS}(\cdot)$ stands for the spectral factor associated with $\Theta^b_{MS}(z)$. Then, it is straightforward to show that the $N_T \times N_T$ residual noise vector

$$w^b_{w, MS}(z) = -\sigma^2 G^b(z) \hat{x}^b(\cdot) + G^b(z) H^b(\cdot^{-1}) w^b(\cdot)$$

(32)

is white assuming that

$$\Theta^b_{\hat{x}}(z) \triangleq \mathbb{E}\left( \hat{x}^b(z) \hat{x}^b(\cdot^{-1}) \right) = I_{N_T}.$$ 

Such assumption is met under both deep space-time interleaving and the energy allocation strategy adopted (see Section III.2.2 for more details). Nevertheless, the residual noise $w^b_{w, MS}(z)$ related to the equivalent model described by (31) depends on transmitted data. Indeed, at the prefilter output, the MS approximation results in an anticausal component $-\sigma^2 G^b(z) \hat{x}^b(\cdot)$ which was included in the residual noise $w^b_{w, MS}(z)$. This is clearly a source of sub-optimality when compared to the Zero Forcing implementation of the Whitened Matched Filter.

We will omit index $b$ in the following since the treatment is identical for all blocks.
IV.2.2.1 Computing the FIR approximation of $G^b(z)$: Several algorithms have been presented in the literature in order to determine the spectral factor $H^b_{\text{min}, MS}(z)$ of $\Theta^b_{MS}(z)$ [28]. We rather choose here to compute an approximation of the Infinite Impulse Response (IIR) anti-causal filter $G^b(z)$ by a Finite Impulse Response filter using the linear prediction theory [29]-[30]-[17].

Therefore, let us consider $s(n)$ as a $N_T \times N_T$ wide sense stationary process with spectrum

$$\Theta_{\text{MS}}(z) = H^b_{\text{min}, MS}(z) H_{\text{min, MS}}(z^{-1})$$

(33)

and

$$\bar{s}(n) = \sum_{k=1}^{l_p} A_k s_{n-k}$$

(34)

be its estimation in the term of its $l_p$ most recent values. The $Z$-transform of the error vector $e(n) \triangleq s(n) - \bar{s}(n)$ is then given by

$$e(z) = A(z)s(z)$$

(35)

where

$$A(z) = I_{N_T} - \sum_{k=1}^{l_p} A_k z^{-k} = I_{N_T} - \bar{A}(z)$$

(36)

stands for the prediction error filter. The limit case $l_p \to \infty$ yields a white error process, i.e.,

$$A(z) \Theta_{\text{MS}}^b(z) A^\dagger(z^{-1}) = \Theta_{e_0}.$$  

(37)

with the $N_T \times N_T$ constant matrix $\Theta_{e_0}$ being the hermitian definite positive spatial autocorrelation matrix of the error process whose single value decomposition is

$$\Theta_{e_0} = U \Lambda U^\dagger = \left( U \Lambda^{1/2} \right) \left( \Lambda^{1/2} U \right)^\dagger = \Theta \Theta^\dagger.$$ 

(38)

Thus, combining (37) and (38) leads to

$$\Theta_{\text{MS}}^b(z) = \Theta^{-1} A(z)^{-1} (A^\dagger(z^{-1}) \Theta^{-1})^{-1} = H^b_{\text{min, MS}}(z) H_{\text{min, MS}}(z^{-1}).$$

However, the spatial factorization is unique up to a multiplication on the left by an arbitrary constant $N_T \times N_T$ unitary matrix. Thus, for a filter order $l_p$ not too low, the whitening filter $G(z)$ can be approximated by

$$G(z) \simeq \Theta^{-1} A(z^{-1}) = \Theta^{-1} \left( I_{N_T} - \sum_{k=1}^{l_p} A_k z^k \right).$$

(39)

Keeping in mind that $G(z) = \left( H^b_{\text{min, MS}}(z) \right)^{-1}$ is causal (according to the spectral factorization theorem), tending $z \to 0$ in (39) yields

$$\Theta = H^b_{\text{min, MS,0}}$$

(40)

and

$$\Theta_{e_0} = H^b_{\text{min, MS,0}} H_{\text{min, MS,0}}.$$ 

(41)
In (40), \( H_{\text{min,MS},0}^\dagger \) stands for the first tap of the equivalent channel \( H_{\text{min,MS}} \). Thus, the \( Z \)-transform \( P_{\text{MS}}(z) \) of the MS-WMF becomes
\[
P_{\text{MS}}(z) = \left( H_{\text{min,MS},0}^\dagger \right)^{-1} A(z) H_{\text{min,MS}}^\dagger (z^{-1})
\]
\[
= \left( H_{\text{min,MS},0}^\dagger \right)^{-1} \left( I_{N_T} - \sum_{k=1}^{l_p} A_k z^{-k} \right) H_{\text{min,MS}}^\dagger (z^{-1}) .
\] (42)

Practically, there is no need to compute explicitly \( H_{\text{min,MS},0} \). Indeed, let us first apply the filter \( A(z) H_{\text{min,MS}}^\dagger (z^{-1}) \). The remaining noise (based on the equivalent model given by equation 31) at the output of such prefilter is temporally white but spatially colored with its spatial correlation matrix being \( \Theta_{e,0} \) given by (41). This spatial correlation matrix \( \Theta_{e,0} \), i.e., \( H_{\text{min,MS},0} H_{\text{min,MS},0}^\dagger \), can be easily estimated. Then, applying the matrix \( L^{-1} \) obtained from the following Cholesky decomposition \( \Theta_{e,0} = L L^\dagger \) spatially whitens the noise while keeping the equivalent channel minimum phase. This is due to the fact that the matrix \( L^{-1} H_{\text{min,MS},0}^\dagger \) is unitary.

**IV.2.2 Computing the prediction error filter** \( \tilde{A}(z) \): The prediction filter \( \tilde{A}(z) \) is designed in a MMSE sense. Resorting to the orthogonality principle, i.e.,
\[
\mathbb{E}\left( e(n) s(n-k)^\dagger \right), \ 1 \leq k \leq l_p,
\]
we see that the optimum coefficients for this criterion are the solution of the Yule-Walker equations
\[
A \Theta_n = \theta_n .
\] (43)
with
\[
A = \begin{bmatrix} A_1 & \cdots & A_{l_p} \end{bmatrix} \in \mathbb{C}^{N_T \times N_T l_p}.
\]

In (43), \( \Theta_n \) and \( \theta_n \) denote respectively the \( l_p N_T \times l_p N_T \) block hermitian toeplitz autocorrelation matrix defined by its first row
\[
\Theta_n = \begin{bmatrix} \Theta_{n,0} & \Theta_{n,1} & \cdots & \Theta_{n,l_p-1} \end{bmatrix}
\]
and the \( N_T \times l_p N_T \) vector matrix
\[
\theta_n = \begin{bmatrix} \Theta_{n,1} & \Theta_{n,2} & \cdots & \Theta_{n,l_p} \end{bmatrix} .
\]
The \( N_T \times N_T \) autocorrelation matrices \( \Theta_{n,i} = \mathbb{E}(s_n s_{n-i}^\dagger) \), \( 0 \leq i \leq l_p \), are obtained by identification from
\[
\Theta_{MS}^\dagger (z) = \sum_i \Theta_{n,i} z^{-i} = H_{\text{min,MS}}^\dagger (z^{-1}) + \sigma^2 I_{N_T} .
\] (44)

**IV.2.3 Complexity of the proposed implementation**: The complexity of the MS-WMF approach is dominated by the resolution of the Yule-Walker equation. This can be done efficiently thanks to the generalized MIMO Levinson algorithm [31] which requires
\[
O \left( (l_p + 1)^2 N_T^3 \right)
\]
operations. Furthermore, the MIMO Levinson algorithm enables an adaptive selection of the prefilter order according to the given channel [26].
In order to validate the approximation made in the proposed implementation of the MS-WMF when computing the FIR whitening filter $G(z)$, we consider again the same transmissions scenarios as the ones taken previously in Figures 2 and 3. The reduced memory of the DFSE was taken $m_r = 1$. We further assume that the prefilter orders are the same, i.e., $l_p = 5, l_p = 10$ and $l_p = 15$ for the EQ3 channel and $l_p = 15, l_p = 20$ and $l_p = 25$ for the EQ5 channel. In both cases, we see from Figure 7 and Figure 8 that no performance loss is observed. This is clearly an important advantage compared to the shortening approach which is much more sensitive to the prefilter order value (refer to Section III.3).

V. COMPARISON OF BOTH APPROACHES

Our objective is to make a fair comparison of both approaches depicted in this paper in term of trade-off between performance and complexity.

V.1 Complexity comparison

We illustrate in Table I the computational complexity of both approaches based on the operations enumerated in Subsubsections III.2.3 and IV.2.3.

It is important to keep in mind here that this complexity comparison has a sense if and only if detectors and decoders used in both turbo-receivers have the same complexity and if prefilter orders are the same. Under this assumption, we see clearly from Table I that the shortening approach is more complex since the operations enumerated need to be repeated for each delay.

V.2 Performance comparison

V.2.1 Simulation setting: We move now to compare the performance of both proposed turbo-receivers. The prefilter order is equal to $l_p = 15$ and five iterations are performed. In a matter of fairness, the shortened channel is set to $M_s = 1$ when the DFSE (associated with the MS-WMF) works on a reduced memory of $m_r = 1$. A STBICM using 4-PSK is simulated. The transmitted block symbol length is $L = 256$. Transmission occurs over quasi-static $2 \times 2$ MIMO channel with perfectly uncorrelated taps. Unless otherwise stated, we employ a rate-1/2 64-state non-recursive convolutional code with $d_{free} = 10$. The spectral efficiency becomes then $\eta = 2$ bits p.c.u. We use the outage probability [32] and the coded Matched Filter Bound as benchmarks in order to evaluate the loss inherent to the suboptimality of both approaches.

A particular attention is still given to the region of BLER near to $10^{-2}$. Thus, a target BLER of $10^{-2}$ is systematically assumed when comparing the approaches. It is important to notice here that the $Eb/N_0$ appearing in all simulations refers to the SNR per receive antenna and per useful transmitted bit.
V.2.2 Effect of the channel selectivity: We start with the behavior of the proposed turbo-receivers with respect to the channel selectivity. The MS-WMF approach gives better results for very selective channels (EQ10 channel), cf. Figure 11. The shortening approach seems to be unable to deal with high selectivity and the use of turbo-equalization does not help much. For the EQ5 channel (Figure 10), the gap between both approaches at BLER $10^{-2}$ is reduced to 0.6 dB. The shortening approach becomes better from $Eb/N_0 = 7$dB. The same phenomenon occurs with the EQ3 channel where the shortening approach outperforms the MS-WMF approach from $Eb/N_0 = 2$ dB, cf. Figure 9. The gain (with shortening) at BLER $10^{-2}$ is 0.6dB. We also observe that the shortening approach is always better at the first iteration (and at the second iteration for the EQ3 and EQ5 channels). However, the gain between the second and the fifth iteration is more important with the MS-WMF approach and increases with channel selectivity. Moreover, it seems that, at high SNRs, the slope of the shortening approach is steeper. This demonstrates its higher capacity to recover the inherent system diversity.

V.2.3 Impact of the channel ISI profile: We reconsider here the same transmission scenarios as Figure 9 and Figure 10, except that the MIMO channel taps follow the EXPonential decreasing law described below: on each link $(r,t)$, the $M+1$ channel coefficients are i.i.d circularly symmetric complex Gaussian following pdfs $\mathcal{CN}(0, \sigma^2_m)$ with $\sigma^2_m \propto \exp(-2m)$, $\forall m$. Figure 12 treats the transmission over a 3 tap channel (EXP3 channel with $M = 2$) when Figure 13 deals with an EXP5 channel. Contrary to the equal energy tap channels, the shortening approach outperforms at all SNRs the whitening approach. In both cases, the gain brought by the turbo-receiver using a shortening front-end, when compared to the MS-WMF approach, is near to 1.5 dB at BLER $10^{-2}$. Thus, the shortening approach seems to be more appropriate for the ISI channel profile considered here. On the other hand, Figure 12 and 13 confirm that the shortening approach has a steeper slope than the MS-WMF approach. Moreover, this slope is parallel to the one of the MFB proving that all the available system diversity is captured here.

V.2.4 Effect of the spectral efficiency: We increase now the spectral efficiency of the simulated STBICM by using a rate-3/4 64-state non-recursive convolutional code with $d_{free} = 6$ instead. As we can see from Figure 16, the shortening approach is still unable to face high selectivity channel (an EQ10 channel was considered here). However, it gives better performance at BLER $10^{-2}$ with the EQ5 and the EQ3 channels. The gain brought by the shortening approach is close respectively to 2.5 dB and 1.5dB for the EQ3 (Figure 14) and the EQ5 (Figure 15) $2 \times 2$ MIMO channel. Furthermore, a phenomenon revealed previously is still present here independently of the channel memory. Indeed, the performance of the second iteration and the fifth one of the shortening approach is very close. The turbo process seems to be more beneficial for the MS-WMF approach.

VI. GENERAL REMARKS

Before concluding this paper, it is relevant to outline the following remarks:

- **Remark 1**: When considering the same prefilter order, the same iteration number and the same complexity for ISI detectors and channel decoders, the shortening turbo-receiver appears to be more complex than the MS-WMF one. However, simulations show clearly that the shortening approach ensures a faster convergence of the iterative process. We think that such phenomena is related to the fact that it uses an optimal MAP-based ISI detector. Moreover, this may help reducing the global shortening approach complexity by performing fewer iterations without significant performance loss.
Remark 2: We assume that neglecting the time correlation of the noise at the output of the shortening front-end has a limited effect on the performance achieved by the shortening approach. Intuitively, we think that the impact of noise coloration becomes more important as $M - M_r$ grows up. This may explain the difficulties faced by the shortening-based turbo-receiver when increasing the channel selectivity.

Remark 3: In the transmission scheme considered in this paper, we implicitly assume a guard interval to fight back block interference. The introduction of such guard interval whose length must be at least equal to the channel memory causes a slight spectral efficiency loss. From this point of view, the shortening approach appears to be more advantageous than the whitening one since it allows a smaller guard interval.

VII. CONCLUSIONS

In this paper, we have investigated two different prefiltering strategies, i.e., channel shortening and whitened matched filtering, to decode a STBICM over MIMO ISI Channel using a trellis based post-detector. As noted in the introduction, both prefiltering method performance highly depends on the MIMO channel rank. Simulation results tend to show that the MS-WMF approach outperforms the channel shortening approach at fixed complexity and high channel selectivity. However, the channel shortening approach appears as an interesting alternative for low frequency selective channel independently of the channel ISI profile considered. Moreover, the channel shortening obviously needs fewer iterations to converge and always outperforms the MS-WMF at first iteration. Finally, it is important to keep in mind that for large MIMO systems and/or high-order modulations, the number of transitions per state becomes the true limiting factor of both techniques detailed above. An interesting alternative can then be the resorting to soft interference cancellation algorithms.

REFERENCES

[7] Boutros (J.J), Gresset (N), and Brunel (L). Turbo Coding and Decoding for Multiple Antenna Channels. 3 rd International Symposium on Turbo Codes and Related Topics, Brest, France, September 2003.


Figure 1. Low complexity turbo-receiver using a shortening front-end - Récepteur turbo utilisant un filtre raccourcissant

Figure 2. Block-static System, $2 \times 2$ EQ5, $\eta=2$ bpcu, Shortening approach - Système statique par bloc, $2 \times 2$ EQ5, $\eta=2$ bpcu, Approche du filtre raccourcissant
Figure 3. Block-static System, $2 \times 2$ EQ10, $\eta=2$ bpcu, Shortening approach - Système statique par bloc, $2 \times 2$ EQ10, $\eta=2$ bpcu, Approche du filtre raccourcissant

Figure 4. Low complexity turbo-receiver using a MIMO WMF front-end - Récepteur turbo utilisant un filtre MIMO adapté blanchissant
Figure 5. Block-static System, $2 \times 2$ EQ3, $\eta=2$ bpcu, WMF approach vs DFSE - Système statique par bloc, $2 \times 2$ EQ3, $\eta=2$ bpcu, Approche du WMF vs DFSE

Figure 6. Block-static System $2 \times 2$ EQ5, $\eta=2$ bpcu, WMF approach vs DFSE - Système statique par bloc, $2 \times 2$ EQ5, $\eta=2$ bpcu, Approche du WMF vs DFSE
Figure 7. Block-static System, 2×2 EQ5, η=2 bpcu, WMF approach - Système statique par bloc, 2×2 EQ5, η=2 bpcu, Approche WMF

Figure 8. Block-static System, 2×2 EQ10, η=2 bpcu, WMF approach - Système statique par bloc, 2×2 EQ10, η=2 bpcu, Approche WMF
Figure 9. Block-static System, 2×2 EQ3, $\eta=2$ bpcu, WMF approach vs shortening approach - Système statique par bloc, 2×2 EQ3, $\eta=2$ bpcu, Approche du filtre WMF vs approche du filtre raccourcissant.

Figure 10. Block-static System, 2×2 EQ5, $\eta=2$ bpcu, WMF approach vs shortening approach - Système statique par bloc, 2×2 EQ5, $\eta=2$ bpcu, Approche du filtre WMF vs approche du filtre raccourcissant.
Figure 11. Block-static System, $2 \times 2$ EQ10, $\eta=2$ bpcu, WMF approach vs shortening approach - Système statique par bloc, $2 \times 2$ EQ10, $\eta=2$ bpcu, Approche du filtre WMF vs approche du filtre raccourcissant

Figure 12. Block-static System, $2 \times 2$ EXP3, $\eta=2$ bpcu, WMF approach vs shortening approach - Système statique par bloc, $2 \times 2$ EXP3, $\eta=2$ bpcu, Approche du filtre WMF vs approche du filtre raccourcissant
Figure 13. Block-static System, $2 \times 2$ EXP5, $\eta = 2$ bpcu, WMF approach vs shortening approach - Système statique par bloc, $2 \times 2$ EXP5, $\eta = 2$ bpcu, Approche du filtre WMF vs approche du filtre raccourcissant

Figure 14. Block-static System, $2 \times 2$ EQ3, $\eta = 3$ bpcu, WMF approach vs shortening approach - Système statique par bloc, $2 \times 2$ EQ3, $\eta = 3$ bpcu, Approche du filtre WMF vs approche du filtre raccourcissant
Figure 15. Block-static System, $2 \times 2$ EQ5, $\eta=3$ bpcu, WMF approach vs shortening approach - Système statique par bloc, $2 \times 2$ EQ5, $\eta=3$ bpcu, Approche du filtre WMF vs approche du filtre raccourcissant.

Figure 16. Block-static System, $2 \times 2$ EQ10, $\eta=3$ bpcu, WMF approach vs shortening approach - Système statique par bloc, $2 \times 2$ EQ10, $\eta=3$ bpcu, Approche du filtre WMF vs approche du filtre raccourcissant.