

# On Low-Complexity Space–Time Coding for Quasi-Static Channels

Giuseppe Caire, *Member, IEEE*, and Giulio Colavolpe, *Member, IEEE*

**Abstract**—We propose a new space–time coding scheme for the quasi-static multiple-antenna channel with perfect channel state information at the receiver and no channel state information at the transmitter. In our scheme, codewords produced by a trellis encoder are formatted into space–time codeword arrays such that decoding can be implemented efficiently by minimum mean-square error (MMSE) decision–feedback interference mitigation coupled with Viterbi decoding, through the use of per-survivor processing.

We discuss the code design for the new scheme, and show that finding codes with optimal diversity is much easier than for conventional trellis space–time codes (STCs). We provide an upper bound on the word-error rate (WER) of our scheme which is both accurate and easy to evaluate. Then, we find upper and lower bounds on the information outage probability with discrete independent and identically distributed (i.i.d.) inputs (as opposed to Gaussian inputs, as in most previous works) and we show that the MMSE front-end yields a large advantage over the whitened matched filter (i.e., zero-forcing) front-end. Finally, we provide a comprehensive performance/complexity comparison of our scheme with coded vertical Bell Labs layered space–time (V-BLAST) architecture and with the recently proposed threaded space–time codes. We also discuss the concatenation of our scheme with block space–time precoders, such as the linear dispersion codes.

**Index Terms**—Multiple-antenna systems, per-survivor processing, space–time coding.

## I. INTRODUCTION

IN recent years, transmission schemes based on multiple antennas have attracted much attention as a viable solution to increase spectral efficiency and performance of wireless channels.

Roughly speaking, works on multiple antennas can be classified according to the assumptions on the channel state information (CSI) available at the transmitter and at the receiver. The ergodic capacity [1] of a frequency-nonselctive channel with  $t$  transmit and  $r$  receive antennas and independent Rayleigh fading, no transmitter CSI and perfect receiver CSI has been calculated by Telatar in [2]. In this case, the capacity for high signal-to-noise ratio (SNR) is given by  $C = \min\{r, t\} \log_2 \text{SNR} + O(1)$  bit per channel use. The ergodic capacity with perfect CSI both at the transmitter and

at the receiver has the same high-SNR behavior [3]. The information outage probability [1] (related to the nonergodic, or “outage” capacity) with no transmitter CSI and perfect receiver CSI has been investigated numerically in [2] and in [4]. Transmit strategies minimizing the information outage probability with perfect CSI both at the transmitter and at the receiver have been considered in [5].

Other works assume perfect receiver CSI and partial transmitter CSI [6], frequency-selective fading, and a combination of multiple antennas and orthogonal frequency-division multiplexing (OFDM) [7], and large-system limits of physical scattering models, without any assumption of independent Rayleigh fading, based on large random matrix theory [8].

A different model for time-varying fading channels has been introduced by Hochwald and Marzetta in [9]. They considered a block-fading channel constant for  $T$  consecutive channel uses and independent from block to block, where both transmitter and receiver have no CSI. Tse and Zheng [10] have shown that the high-SNR capacity of such a channel is  $C = k(1 - k/T) \log_2 \text{SNR} + O(1)$ , where  $k = \min\{r, t, T/2\}$ . They also show that the capacity is maximized by using no more than  $k$  transmit antennas. In particular, having  $t > r$  antennas does not improve the high-SNR capacity behavior. In [10], [11] it is shown that the same high-SNR capacity behavior can be achieved by a “naive” scheme that allocates  $k$  input dimensions to channel estimation (e.g., by explicitly sending training symbols), and by a mismatched receiver that treats the training-based estimated channel as if it were the actual channel.

When the code block length  $N$  is much larger than  $\max\{r, t\}$ , and the channel coherence interval is  $T \geq N$ , one codeword spans a single channel realization. Then, we are in the presence of a compound channel, i.e., a collection of channels indexed by the different realizations of the  $r \times t$  channel matrix. In this case, the (nonergodic) capacity and information outage probability of the channel with or without CSI at the receiver coincide, since the channel matrix process is *strongly singular* [1].

In practice, the assumption of perfect CSI at the receiver holds approximately when the channel varies very slowly with respect to the duration of a codeword (quasi-static assumption). This is a realistic assumption in situations where the mobility of wireless terminals is limited or absent (e.g., indoor wireless local-area networks, wireless local loops). On the contrary, the assumption of perfect CSI at the transmitter holds only if a delay-free error-free feedback link from receiver to transmitter exists, or if time-division duplexing is used [12], where each end can estimate the channel from the incoming signal in the reverse direction. Motivated by the above considerations, we conclude

Manuscript received March 9, 2001; revised February 6, 2003. The material in this paper was presented in part at GLOBECOM 2001, San Antonio, TX, November 2001.

G. Caire is with Mobile Communications Group, Institut EURECOM, 06560 Valbonne, France (e-mail: giuseppe.caire@eurecom.fr).

G. Colavolpe is with the Dipartimento di Ingegneria dell’Informazione, Università di Parma, 43100 Parma, Italy (e-mail: giulio@unipr.it).

Communicated by R. Urbanke, Associate Editor for Coding Techniques.

Digital Object Identifier 10.1109/TIT.2003.811904

that assuming perfect CSI at the receiver and no CSI at the transmitter is reasonable and can be applied in several practical settings for which  $T \geq N \gg \max\{r, t\}$  and where feedback or time-division duplexing cannot be exploited. The results based on this assumption can be effectively approached in the high-SNR region by the naive scheme based on explicit training. Without further questioning the validity of this model, we adopt it as our starting point.

Coding schemes for the quasi-static multiple-antenna channel with perfect receiver CSI have been proposed in several works (see, for example, [13]–[19]). Recently, a general framework for space–time coding referenced to as *generalized space–time layering* was identified as follows [20]: a space–time encoder is formed by  $M$  encoders, producing  $M$  independent codewords, and by a space–time “formatter” that maps these codewords into a single  $t \times N$  codeword matrix. On the receiver side, since joint maximum-likelihood (ML) decoding of the  $M$  codewords might be too complex, some suboptimal low-complexity method combining decoding and signal processing is applied. The layered space–time known as diagonal Bell Labs space–time (D-BLAST) and vertical Bell Labs space–time (V-BLAST) [14], [21] belong to this class. Driven by this general framework, [20] proposed a scheme referenced to as *threaded space–time coding* (TSTC), suited to iterative decoding based on soft interference cancellation and linear minimum mean-square error (MMSE) interference mitigation at each iteration stage. Guidelines for the design of the TSTC component codes were given and the performance of the overall iterative decoder was studied by Monte Carlo simulation.

In this paper, we propose a new scheme nicknamed *wrapped space–time coding* (WSTC) which is also a particular instance of the generalized space–time layering framework of [20].<sup>1</sup> For this scheme, we propose a low-complexity suboptimal decoder combining decision-feedback equalization and *per-survivor processing* (PSP) [22]. More precisely, our decoding scheme is reminiscent of the “nulling and canceling” procedure proposed for V-BLAST [21], but interference is canceled by using the predecisions on the survivor history of the Viterbi decoder. It follows that our decoder does not suffer from decision feedback errors, and its complexity is equivalent to one iteration of the TSTC decoder. Interestingly, our scheme can be seen as the concatenation of trellis coding with delay diversity and the corresponding low-complexity decoder can be seen as a reduced-state sequence estimator [23]–[25] applied to the trellis of the augmented code, including the memory due to delay diversity as part of the encoder.

We discuss code design for WSTC, and show that finding codes with optimal diversity is actually much easier than for conventional trellis space–time codes (STCs). Namely, in many cases, the maximum possible diversity is achieved by well-known trellis codes. We provide a semianalytic upper bound on the word-error rate (WER) of our scheme that predicts accurately the simulation results and it is easy to evaluate

(at least for geometrically uniform codes). Then, in order to gain insight regarding the behavior of the WER, we find upper and lower bounds on the information outage probability with discrete independent and identically distributed (i.i.d.) inputs. Finally, we provide numerical examples illustrating some interesting facts about the proposed scheme. We compare WSTC with coded V-BLAST [21], that is also based on decision–feedback interference cancellation but it does not exploit PSP, and we show that the latter suffers from severe error propagation in the feedback decisions while the proposed scheme does not. We provide some comparisons with the TSTC scheme of [20], and we show that WSTC outperforms TSTC for the same decoder complexity (one decoder iteration), but it is generally outperformed by TSTC for more than one decoder iteration, i.e., at the cost of higher decoder complexity. Finally, for the case  $r < t$ , we discuss through some examples the opportunity of concatenating WSTCs with space–time block precoders such as the recently proposed *linear dispersion codes* [19].

The paper is organized as follows: in Section II, we introduce the WSTC scheme, in the general framework of [20]. In Section III, we describe the PSP-based low-complexity decoding scheme. In Section IV, we discuss the code design for WSTC and in Section V, we present the semianalytic bound on the WER. In Section VI, we present results on the outage probability with discrete inputs and in Section VII, we show some numerical results. Conclusions are summarized in Section VIII.

## II. CODING FOR THE QUASI-STATIC MULTIPLE-ANTENNA CHANNEL

The channel model with  $t$  transmitting and  $r$  receiving antennas considered in this paper is defined by [13], [26], [4], [2]

$$\mathbf{y}_n = \sqrt{\gamma} \mathbf{H} \mathbf{x}_n + \mathbf{z}_n, \quad n = 1, \dots, N \quad (1)$$

where  $\mathbf{x}_n \in \mathcal{X}^t$  is the vector of modulation symbols transmitted in parallel at time  $n$  by the  $t$  transmit antennas,  $\mathcal{X} \subseteq \mathbb{C}$  denotes a complex modulation signal set with unit average energy,  $\mathbf{z}_n \in \mathbb{C}^r$  is the noise vector  $\sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$ ,  $\mathbf{y}_n \in \mathbb{C}^r$  is the corresponding vector of received signal samples at the output of the receiving antennas,  $\mathbf{H} \in \mathbb{C}^{r \times t}$  is the channel matrix, and  $\gamma$  is the average SNR per transmit antenna. The channel matrix is normalized such that  $\frac{1}{rt} \text{trace}(E[\mathbf{H}\mathbf{H}^H]) = 1$ , so that the average received SNR per receiving antenna is given by  $t\gamma$ .

As anticipated in the Introduction, we consider the case where  $\mathbf{H}$  is random but constant over  $N \gg \max\{t, r\}$  channel uses, and we assume that the receiver knows  $\mathbf{H}$  perfectly, while the transmitter has no knowledge of  $\mathbf{H}$ .

An STC for the above channel is a set  $\mathcal{S} \subseteq \mathcal{X}^{t \times N}$  of  $t \times N$  complex matrices (codewords). Codeword matrices  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$  are transmitted by columns, in  $N$  consecutive channel uses. The STC spectral efficiency is given by  $\eta = \frac{1}{N} \log_2 |\mathcal{S}|$  bits per channel use. By definition, the information bit energy over noise power spectral density ratio is given by  $E_b/N_0 = t\gamma/\eta$ .

### A. Generalized Space–Time Layering

In [13], Tarokh *et al.* found criteria for STC design. They considered the pairwise-error probability (PEP) with ML de-

<sup>1</sup>At the time of the first submission of this work, [20] was still unpublished and we developed our scheme independently of the framework of [20]. However, following a comment by one of the anonymous reviewers, we found indeed useful to make explicit reference to [20].

coding and, in the case of Rayleigh/Rician fading coefficients, they showed that the PEP  $P(\mathbf{X} \rightarrow \mathbf{X}')$  averaged with respect to  $\mathbf{H}$  is upper-bounded as

$$P(\mathbf{X} \rightarrow \mathbf{X}') \leq \mathcal{K}\gamma^{-r\rho}$$

where  $\mathcal{K}$  is a factor that depends on the codeword difference matrix  $\mathbf{D} = \mathbf{X}' - \mathbf{X}$  and the statistics of the channel matrix, but it is independent of  $\gamma$ , and where  $\rho$  is the rank of  $\mathbf{D}$ . Driven by the above bound, they indicated the maximization of the minimum rank over all distinct  $\mathbf{X}, \mathbf{X}' \in \mathcal{S}$  as the most important criterion for STC design. For brevity, we shall refer to the minimum rank over all distinct codeword pairs as the *rank diversity* of the code.

For large  $N$ , STCs can be constructed from trellis codes [27] with trellis termination. Examples of such schemes are given in [13], [16]–[18]. Namely, consider a trellis code  $\mathcal{C}$  over  $\mathcal{X}$  of rate  $R = b/t$  bits per symbol, where each trellis step corresponds to  $b$  information (input) bits and  $t$  code (output) symbols, and the subcode of all  $\mathbf{c} \in \mathcal{C}$  leaving a given trellis state  $s_0$  and merging into a given trellis state  $s_N$  after  $N$  trellis steps. Then, a trellis STC can be obtained by formatting the codewords  $\mathbf{c}$  as  $t \times N$  matrices, i.e., by transmitting the  $t$  symbols produced by the trellis encoder at each trellis step in parallel on the  $t$  transmit antennas. The resulting spectral efficiency is given by  $\eta \approx b$  bits per channel use.<sup>2</sup>

The difficulty in constructing these codes is that the rank diversity is hard to evaluate and it is not easily related to the algebraic properties of the underlying trellis code. In [26], for the class of binary and quaternary trellis codes over  $\mathbb{Z}_2$  and  $\mathbb{Z}_4$  mapped onto binary phase-shift keying (BPSK) and quaternary phase-shift keying (QPSK), respectively, a condition on the underlying algebraic codes referred to as the *binary-rank* criterion is shown to imply the rank diversity of the resulting STC  $\mathcal{S}$  and it is used to construct STCs with full rank diversity (i.e., with  $\rho = t$ ). The binary-rank criterion is much easier to check than the rank diversity and yields some explicit general algebraic constructions.

ML decoding for trellis STCs can be implemented by the Viterbi algorithm applied to the trellis of the underlying trellis code. However, for a fixed code rate  $R$  and rank diversity  $\rho$ , the decoder complexity grows exponentially with the number of transmit antennas  $t$ . In fact, from [13, Lemma 3.3.2] we know that the trellis complexity is lower-bounded by  $2^{tR(\rho-1)}$ . Because of the difficulty of code design (with the exception of constructions given in [26]) and because of decoding complexity, trellis STCs are practically restricted to small  $t$ .

In order to handle a large number of antennas with moderate complexity, a generalized space–time layering approach is presented in [20]. In this framework, the space–time encoder is formed by  $M$  component encoders for codes  $\mathcal{C}_1, \dots, \mathcal{C}_M$  over  $\mathcal{X}$  and by a space–time formatter  $\mathcal{F}: \mathcal{C}_1 \times \dots \times \mathcal{C}_M \rightarrow \mathbb{C}^{t \times N}$ . The  $M$  encoders generate independent codewords  $\{\mathbf{c}^{(m)}: m = 1, \dots, M\}$ . Then, the codeword matrix  $\mathbf{X} = \mathcal{F}(\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(M)})$  is obtained by formatting the coded symbols as elements of  $\mathbf{X}$ , possibly including some zero symbols.

<sup>2</sup>For simplicity, we ignore the rate decrease due to trellis termination since it becomes negligible for  $N$  much larger than the trellis encoder memory.

Although ML joint decoding of the  $M$  codewords might be too complex, this approach lends itself to the following “divide-and-conquer” strategy for code design and decoder implementation: 1) the formatter is chosen in order to facilitate some low-complexity suboptimal decoding scheme based on individual decoding of the component codes and on mitigation of the mutual interference between the component codewords; 2) the component codes are designed under the assumption that mutual interference between their codewords is actually absent.

The layered space–time architectures known as D-BLAST [14] and V-BLAST [21]<sup>3</sup> and the TSTC scheme of [20] are particular instances of the above approach. Generalized space–time layering yields schemes suited to a large number of antennas. For example, the decoders of D- and V-BLAST and TSTC have complexity  $O(t^3)$ . For the sake of space limitation, and since these schemes are well known and remarkably well summarized in [20], we shall use them as a term of comparison throughout the rest of the paper, without getting into too many details.

### B. “Wrapped” Space–Time Codes (STCs)

The proposed scheme is also a particular instance of the framework of [20]. We motivate our scheme starting from D-BLAST [14]. Fig. 1 shows the D-BLAST formatting for parameters  $t = 4$ ,  $N = 36$ , with  $M = 3$  component codewords of length  $N' = 24$  and interleaving delay  $d = 6$ . The low-complexity decoder for D-BLAST consists of a linear front-end followed by decision–feedback interference cancellation [14]. The linear front-end is defined by the matrix  $\mathbf{F} \in \mathbb{C}^{r \times t}$  and produces the sequence of received vectors

$$\mathbf{v}_n = \mathbf{F}^H \mathbf{y}_n, \quad n = 1, \dots, N.$$

Then, each codeword  $\mathbf{c}^{(m)}$  is individually decoded by taking as observable the sequence of samples

$$r_\ell^{(m)} = v_{j,n} - \sqrt{\gamma} \sum_{k=j+1}^t b_{j,k} \hat{x}_{k,n}, \quad \ell = 1, \dots, N' \quad (2)$$

where the one-to-one index mapping  $(m, \ell) \leftrightarrow (j, n)$  is defined by the formatter  $\mathcal{F}$ ,  $v_{j,n}$  is the  $j$ th element of  $\mathbf{v}_n$ ,  $b_{j,k}$  is the  $(j, k)$ th element of the feedback filter matrix  $\mathbf{B} = \mathbf{F}^H \mathbf{H}$ , and  $\hat{x}_{k,n}$  is the decision about the  $(k, n)$ th symbol of  $\mathbf{X}$ . From Fig. 1, we observe that the elements  $x_{k,n}$  for  $k = j+1, \dots, t$  correspond to either zeros (for which no decision is needed) or to symbols belonging to codewords  $\mathbf{c}^{(m')}$  with  $m' < m$ . Therefore, by decoding the component codewords in the order  $m = 1, \dots, M$ , the decisions needed in (2) are provided by already decoded codewords.

In the original D-BLAST scheme of [14], the length of each component codeword is  $N' = dt$ . For a given  $t$ , a large delay  $d$  is needed in order to have long codewords. If interleaving delay is an issue, the D-BLAST scheme is forced to work with short component code block length  $N'$ . This might pose a serious problem for using trellis codes with a large number of states. In fact, the code memory might not be negligible with respect to  $N'$  thus, yielding a nonnegligible rate loss due to trellis termination.

<sup>3</sup>See also [28], where a comparison between “horizontal” and “diagonal” layered schemes is presented.

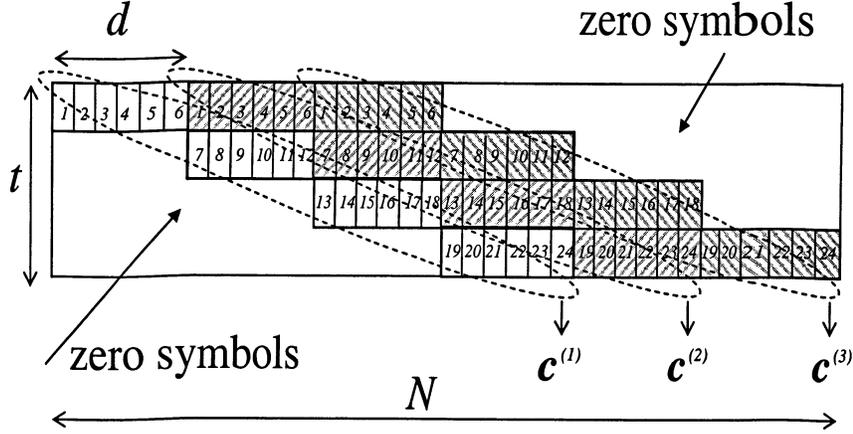


Fig. 1. D-BLAST formatting with  $t = 4$ ,  $d = 6$ ,  $N' = 24$ , and  $M = 3$ . The entries in the above array indicate the index of the symbols in the component codewords.

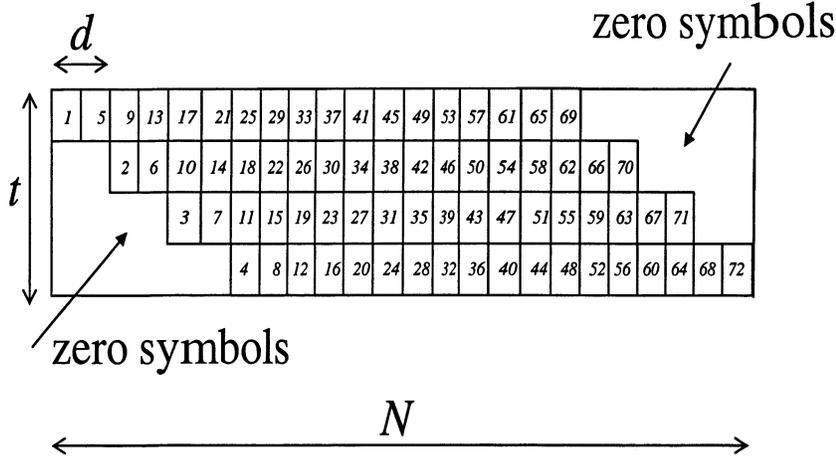


Fig. 2. WSTC formatting with  $t = 4$ ,  $d = 2$ , and  $N' = 72$ . The entries in the above array indicate the index of the symbols in the component codeword.

For this reason, we propose a scheme that keeps the simplicity of decision–feedback interference mitigation while allowing for arbitrarily long component codewords and small interleaving delay. In the proposed scheme, a single encoder for the component code  $\mathcal{C}$  over  $\mathcal{X}$  produces a codeword  $\mathbf{c}$  of length  $N'$ . This codeword is diagonally interleaved in order to form the  $t \times N$  codeword matrix  $\mathbf{X} = \mathcal{F}(\mathbf{c})$ , with  $N = N'/t + (t - 1)d$ . The formatter  $\mathcal{F}$  is defined by

$$x_{j,n} = \begin{cases} c_{\ell_j(n)}, & \text{if } 1 \leq \ell_j(n) \leq N' \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

for  $1 \leq j \leq t$  and  $1 \leq n \leq N$ , where

$$\ell_j(n) = [n - 1 - (j - 1)d]t + j. \quad (4)$$

The codeword matrix  $\mathbf{X}$  is filled by wrapping the codeword  $\mathbf{c}$  around the matrix diagonals, as illustrated by Fig. 2. For this reason, this scheme is referred to as *wrapped* space–time coding (WSTC). In this way, the interleaving delay  $d$  becomes a free parameter, independent of the component codeword block length  $N'$ .

As a limiting case, the interleaving delay may be also  $d = 0$ , i.e., a vertical interleaver may be used. For consistence with the case  $d > 0$ , where code symbols with a lower index take

the lower positions in each column of the codeword matrix  $\mathbf{X}$  (see Fig. 2), the space–time formatter for  $d = 0$  is defined by replacing (3) by

$$\ell_j(n) = (n - 1)t + t - j + 1. \quad (5)$$

*Remark 1:* When  $\mathcal{C}$  is a trellis code of rate  $R = b/t$ , the corresponding WSTC with  $d = 0$  coincides with a standard trellis STC. For  $d > 0$ , the corresponding WSTC can be seen as the concatenation of the trellis code  $\mathcal{C}$  with delay diversity [29], as shown in Fig. 3. Based on this observation, the ML decoder for the WSTC is easily obtained by applying the Viterbi algorithm on the trellis of the *augmented encoder*, i.e., the encoder obtained by including the memory due to delay diversity. However, for large product  $td$ , the complexity of ML decoding might be too large for practical applications.  $\diamond$

*Remark 2:* Because of the lower and upper triangles of zero symbols in the codeword matrix defined by (3), there is an inherent rate loss of  $(t - 1)d/N$ . This is negligible if  $N \gg td$ . Moreover, if the transmission of a long sequence of codewords is envisaged, the codeword matrices can be concatenated in order to fill the leading and tailing triangles of zeros, so that no rate loss is incurred.  $\diamond$

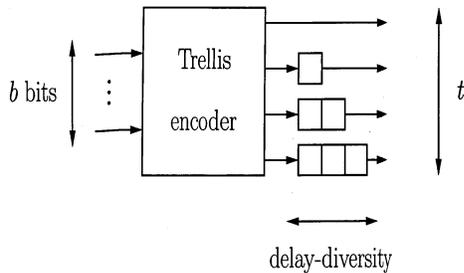


Fig. 3. WSTC seen as the concatenation of trellis coding and delay diversity.

### III. LOW-COMPLEXITY DECODING OF WSTC

In the case where  $\mathcal{C}$  is a trellis code, we propose a low-complexity suboptimal decoder obtained by combining decision-feedback interference cancellation with the Viterbi algorithm. The decoder works on the trellis of the component code  $\mathcal{C}$  (i.e., it neglects the state expansion due to delay diversity) and takes as observable the sequence of samples

$$r_\ell = v_{j,n} - \sqrt{\gamma} \sum_{k=j+1}^t b_{j,k} \hat{x}_{k,n}, \quad \ell = 1, \dots, N' \quad (6)$$

where  $v_{j,n}$  and  $b_{j,k}$  are defined as in (2) and  $1 \leq j \leq t$  and  $1 \leq n \leq N$  are the unique integers for which  $\ell_j(n) = \ell$ . From the index mapping (4) (or (5) for  $d = 0$ ), we see that the elements  $x_{k,n}$ , for  $k = j+1, \dots, t$ , correspond to either zeros (for which no decision is needed) or to symbols of  $\mathbf{c}$  with index  $\ell' \leq \ell - td + 1$  ( $\ell' \leq \ell - 1$  for  $d = 0$ ). These decisions are found in the Viterbi decoder survivor history, according to standard PSP [22].

As an example, consider the case  $d > 0$ . The delay for decision  $\hat{x}_{k,n}$  necessary to compute the observable for symbol  $c_{\ell_j(n)} = x_{j,n}$  is given by  $(k-j)(td-1)$ . If  $(k-j)(td-1)$  is larger than the Viterbi decoding delay (typically five or six times the code constraint length [30]), the corresponding decisions are reliably obtained from the Viterbi decoder output. If  $(k-j)(td-1)$  is not large enough, the corresponding *tentative* decisions  $\hat{x}_{k,n}$  are obtained separately for each trellis state of the Viterbi algorithm from the survivor path ending in that state. In this way, if the correct path is among the survivors, at least a subset of the paths are extended by the Viterbi algorithm with the correct feedback decisions.

*Choice of the Linear Front-End Filter:* For the sake of simplicity, the decoder treats the sequence  $r_\ell$  as if it were produced by the *virtual* scalar-input additive-noise channel

$$r_\ell = \sqrt{\beta_j} c_\ell + \nu_\ell, \quad \ell = 1, \dots, N' \quad (7)$$

where  $\nu_\ell$  is assumed i.i.d.  $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$  and, from (4) or (5), we have that

$$j = \begin{cases} t - |\ell - 1|_t, & \text{for } d = 0 \\ |\ell - 1|_t + 1, & \text{for } d > 0 \end{cases}$$

( $|\cdot|_t$  denotes a modulo  $t$  operation). The SNR  $\beta_j$  of the above channel is given by

$$\beta_j = \frac{\gamma |b_{j,j}|^2}{|\mathbf{f}_j|^2 + \gamma \sum_{k=1}^{j-1} |b_{j,k}|^2} \quad (8)$$

where  $\mathbf{f}_j$  is the  $j$ th column of the front-end filter  $\mathbf{F}$ . Because of diagonal interleaving, the codeword  $\mathbf{c}$  is cyclically interleaved over  $t$  virtual additive white Gaussian noise (AWGN) channels with SNRs  $\beta_1, \dots, \beta_t$ , so that exactly  $N'/t$  symbols are assigned to each channel  $j$ . Notice that (8) is the true SNR of channel  $j$  if decisions in (6) are correct, i.e., if the contribution of past symbols is canceled out exactly. Moreover, the noise samples  $\nu_\ell$  are not Gaussian and not independent, in general. However, provided that these assumptions hold, this scheme decomposes the channel (1) into  $t$  parallel channels with cyclic interleaving, as illustrated in Fig. 4.

Given the analogy between this scheme and decision-feedback equalization of intersymbol interference (ISI) channels, standard choices for the front-end filter matrix  $\mathbf{F}$  are also inspired by equalization [31]. These are as follows.

- 1) The *whitened matched filter* (WMF),  $\mathbf{F} = \mathbf{Q}$ , where  $\mathbf{H} = \mathbf{Q}\mathbf{B}$  is the “QR” decomposition [32] of the channel matrix  $\mathbf{H}$ ,  $\mathbf{Q} \in \mathbb{C}^{r \times t}$  has orthonormal columns and  $\mathbf{B} \in \mathbb{C}^{t \times t}$  is upper triangular.<sup>4</sup>
- 2) The *unbiased MMSE filter*, whose  $j$ th column is given by

$$\mathbf{f}_j = \alpha_j \left[ \mathbf{I} + \gamma \sum_{k=1}^{j-1} \mathbf{h}_k \mathbf{h}_k^H \right]^{-1} \mathbf{h}_j$$

where  $\alpha_j$  satisfies the unbiasedness constraint

$$|\mathbf{f}_j|^2 + \gamma \sum_{k=1}^{j-1} |\mathbf{f}_j^H \mathbf{h}_k|^2 = 1$$

and where  $\mathbf{h}_k$  denotes the  $k$ th column of  $\mathbf{H}$ .

The resulting PSP-based low-complexity decoders for WSTCs will be briefly denoted by WMF-PSP and by MMSE-PSP decoders, depending on the linear front end employed.

The WMF exists only if  $\mathbf{H}$  has rank  $t$ . Subject to mild assumptions on the statistics of  $\mathbf{H}$ , this condition holds with probability 1 if and only if  $t \geq r$ . With the WMF, the  $j$ th-channel SNR is given by  $\beta_j = \gamma |b_{j,j}|^2$  and the additive noise is exactly Gaussian i.i.d. (subject to the assumption of perfect feedback decisions).

The unbiased MMSE filter is always defined (for any  $t$  and  $r$  and all finite  $\gamma$ ) and the  $j$ th-channel SNR is given by

$$\beta_j = \gamma \mathbf{h}_j^H \left[ \mathbf{I} + \gamma \sum_{k=1}^{j-1} \mathbf{h}_k \mathbf{h}_k^H \right]^{-1} \mathbf{h}_j. \quad (9)$$

The additive noise in (7) is neither Gaussian nor i.i.d. (even assuming perfect feedback decisions).

### IV. CODE DESIGN FOR WSTC

Assuming that the parallel channel model with cyclic interleaving (7) holds, the PEP between two codewords  $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$  for given channel SNRs  $\beta_1, \dots, \beta_t$  is given by

$$P(\mathbf{c} \rightarrow \mathbf{c}' | \beta_1, \dots, \beta_t) = Q \left( \sqrt{2 \sum_{j=1}^t \beta_j w_j} \right) \quad (10)$$

<sup>4</sup>We notice here that the WMF corresponds to the front end of the “nulling and canceling” scheme proposed for the detection of V-BLAST [21], and it is analogous to the forced filter of the zero-forcing decision-feedback equalizer for ISI channels [31].

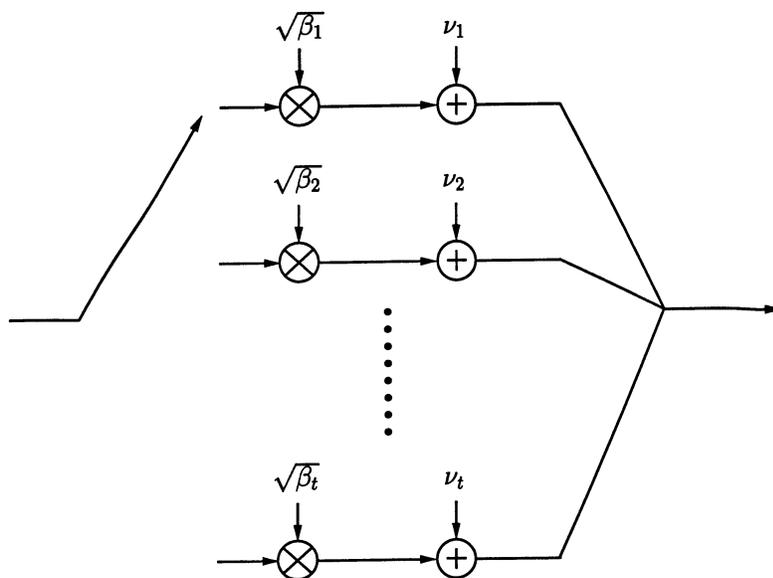


Fig. 4. Parallel channel model with cyclic interleaving originated by the PSP-based decoding of WSTCs subject to the assumption of perfect feedback decisions.

where

$$Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

is the Gaussian tail function and having defined the *squared euclidean weight* (SEW)  $w_j$  as

$$w_j = \frac{1}{4} \sum_{n=1}^N |x'_{j,n} - x_{j,n}|^2 \quad (11)$$

(the correspondence between symbols of  $\mathbf{c}$  and  $\mathbf{c}'$  and symbols of  $\mathbf{X} = \mathcal{F}(\mathbf{c})$  and  $\mathbf{X}' = \mathcal{F}(\mathbf{c}')$  is given by (3)).

A sensible criterion for the design of the component code  $\mathcal{C}$  is to maximize the code *block diversity*  $\delta$ , defined by

$$\delta = \min_{\mathbf{c}, \mathbf{c}' \in \mathcal{C}: \mathbf{c} \neq \mathbf{c}'} |\{j \in \{1, \dots, t\}: w_j \neq 0\}| \quad (12)$$

that is, to maximize the minimum number of nonzero rows in the matrix difference  $\mathbf{D} = \mathbf{X}' - \mathbf{X}$  for each pair of distinct codewords matrices  $\mathbf{X}, \mathbf{X}' \in \mathcal{F}(\mathcal{C})$ . The block-diversity criterion has been investigated in [33]–[35] for the design of trellis codes for cyclic interleaving and/or periodic puncturing.

The relationship between the rank diversity of a WSTC and the block diversity of its component code is given by the following.

*Proposition 1:* Consider a code  $\mathcal{C}$  over  $\mathcal{X}$  of rate  $R$  bits per symbol and block diversity  $\delta$ . Then, the rank diversity  $\rho$  of the corresponding WSTC  $\mathcal{S} = \mathcal{F}(\mathcal{C})$  satisfies

$$\rho \leq \delta \leq 1 + \left\lceil t \left( 1 - \frac{R}{\log_2 |\mathcal{X}|} \right) \right\rceil. \quad (13)$$

Moreover, there exist values of  $d$  for which  $\rho = \delta$ .

*Proof:* Consider a pair of distinct codewords  $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$ , the corresponding matrices  $\mathbf{X} = \mathcal{F}(\mathbf{c})$  and  $\mathbf{X}' = \mathcal{F}(\mathbf{c}')$ , and the difference matrix  $\mathbf{D} = \mathbf{X}' - \mathbf{X}$ . The rank of  $\mathbf{D}$  cannot be larger than the number of nonzero rows of  $\mathbf{D}$ , therefore  $\rho \leq$

$\delta$ . The WSTC  $\mathcal{S}$  can be seen as a block code of length  $t$  over the extended alphabet  $\mathcal{X}^N$ , where each row of  $\mathbf{X}$  is a symbol. The block-diversity  $\delta$  is the minimum Hamming distance of this block code. By applying the Singleton bound [36], we obtain

$$|\mathcal{S}| = 2^{NtR} \leq |\mathcal{X}|^{N(t-\delta+1)}$$

which implies the second inequality in (13).<sup>5</sup>

In order to prove the second statement of the proposition, consider pairs of distinct codewords  $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$  and write their difference  $\mathbf{c}' - \mathbf{c}$  by columns, as a  $t \times N'/t$  array  $\tilde{\mathbf{D}}$ . If  $\tilde{\mathbf{D}}$  has rank  $\geq \delta$  for all pairs  $\mathbf{c}, \mathbf{c}'$ , then the statement is satisfied for  $d = 0$ . Otherwise, for each  $d \geq 1$ , the difference matrix  $\mathbf{D}$  is obtained by appending to the right of each row of  $\tilde{\mathbf{D}}$  a tail of  $(t-1)d$  zeros and by shifting to the right each  $j$ th row by  $(j-1)d$  positions. Let  $\Omega \subseteq \{1, \dots, t\}$  denote the set of the indexes of nonzero rows of  $\tilde{\mathbf{D}}$  and for each row  $j \in \Omega$  let  $\ell_j$  and  $r_j$  denote the number of leading and trailing zeros. Now, let

$$d = \max\{1, \min\{d_1, d_2\}\} \quad (14)$$

where we define

$$d_1 = \max_{\substack{j, k \in \Omega \\ k < j}} \left\{ \left\lceil \frac{\ell_j - \ell_k}{k - j} \right\rceil \right\}$$

and

$$d_2 = \max_{\substack{j, k \in \Omega \\ k < j}} \left\{ \left\lceil \frac{r_k - r_j}{k - j} \right\rceil \right\}.$$

By construction, it is immediate to check that the resulting  $\mathbf{D}$  is either upper triangular or lower triangular and has rank equal to  $|\Omega|$ . By taking  $d$  to be the maximum of (14) over all distinct  $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$  we obtain that  $\mathcal{S} = \mathcal{F}(\mathcal{C})$  has rank diversity  $\delta$ .  $\square$

*Remark 3:* When the component code  $\mathcal{C}$  is a linear trellis code over  $\mathbb{Z}_2$  or  $\mathbb{Z}_4$  mapped onto BPSK or QPSK, respectively,

<sup>5</sup>This upper bound on the block diversity has been found in [37] and, independently, in [34].

TABLE I

BLOCK DIVERSITY FOR SOME WELL-KNOWN RATE-1/2 BINARY CONVOLUTIONAL CODES [38] MAPPED ONTO BPSK AND QPSK (WITH GRAY LABELING). THE CASES ACHIEVING THE BOUND (13) ARE DENOTED IN BOLD. CODE GENERATORS ARE EXPRESSED IN OCTAL NOTATION [38]

States	Generators	BPSK			QPSK		
		$t=2$	$t=4$	$t=8$	$t=2$	$t=4$	$t=8$
4	(5,7)	<b>2</b>	<b>3</b>	<b>4</b>	<b>2</b>	<b>3</b>	<b>3</b>
8	(15,17)	<b>2</b>	<b>3</b>	<b>4</b>	<b>2</b>	<b>3</b>	<b>4</b>
16	(23,35)	<b>2</b>	<b>3</b>	<b>5</b>	<b>2</b>	<b>3</b>	<b>4</b>
32	(53,75)	<b>2</b>	<b>3</b>	<b>5</b>	<b>2</b>	<b>3</b>	<b>4</b>
64	(133,171)	<b>2</b>	<b>3</b>	<b>5</b>	<b>2</b>	<b>3</b>	<b>5</b>

the observation in Remark 1 can be exploited in order to compute the rank diversity of the resulting WSTC via the binary-rank criterion. For example, consider the binary code  $\mathcal{C}$  with generators (5, 7, 7, 7) (octal notation [38]) mapped onto BPSK. The WSTC  $\mathcal{S} = \mathcal{F}(\mathcal{C})$  for  $t = 4$  antennas has rank diversity  $\rho = 4$  for all  $d > 0$ . In fact, the augmented generator matrix obtained by including delay diversity as part of the encoder (see Fig. 3) is given by

$$\mathbf{G}(D) = \begin{bmatrix} 1 + D^2, & D^d(1 + D + D^2), \\ D^{2d}(1 + D + D^2), & D^{3d}(1 + D + D^2) \end{bmatrix}. \quad (15)$$

By writing  $\mathbf{G}(D)$  as a  $4 \times (3d+3)$  matrix of binary coefficients (see [26]), it is immediate to check that the resulting matrix has rank 4 over the binary field for all  $d > 0$ .

*Remark 4:* From [13, Theorem 3.3.1], we know that for any STC over  $\mathcal{X}$  with  $t$  transmit antennas and spectral efficiency  $\eta = tR$  the rank diversity satisfies

$$\rho \leq 1 + \left\lfloor t \left( 1 - \frac{R}{\log_2 |\mathcal{X}|} \right) \right\rfloor. \quad (16)$$

Since this is the same upper bound on block diversity given in Proposition 1, we conclude that the wrapping construction incurs no loss of optimality in terms of rank diversity (for an appropriate choice of the delay  $d$ ). As a matter of fact, while it is difficult to construct codes with rank diversity equal to the upper bound (16), it is very easy to find trellis codes for which the upper bound (13) on  $\delta$  is met with equality, for several coding rates and values of  $t$ . Examples of these codes are tabulated in [33], [35]. Therefore, the wrapping construction is a powerful tool to construct STCs with optimal rank diversity.  $\diamond$

*Remark 5:* From [13, Lemma 3.3.1], we know that a trellis STC with rank diversity  $\rho$  must have constraint length  $L \geq \rho$ . This constraint applies to WSTCs as well, when we consider the constraint length of the augmented encoder including delay diversity.

On the other hand, the proposed low-complexity decoder works on the trellis of the component code  $\mathcal{C}$ .<sup>6</sup> Therefore, it is not *a priori* clear if WSTCs, even if optimal from the rank-diversity point of view, are going to pay a large penalty when low-complexity decoding is used instead of ML decoding. In the next sections, we shall see that the penalty incurred by

<sup>6</sup>Again, we stress the analogy of the problem at hand with the case of trellis coding over a *finite-memory* ISI channel, where the optimal ML decoder requires a number of states generally larger than that of the code alone.

TABLE II

BLOCK DIVERSITY FOR SOME WELL-KNOWN RATE-1/4 BINARY CONVOLUTIONAL CODES [38] MAPPED ONTO BPSK AND QPSK (WITH GRAY LABELING). THE CASES ACHIEVING THE BOUND (13) ARE DENOTED IN BOLD. CODE GENERATORS ARE EXPRESSED IN OCTAL NOTATION [38]

States	Generators	BPSK		QPSK	
		$t=4$	$t=8$	$t=4$	$t=8$
4	(5,7,7,7)	<b>4</b>	<b>5</b>	<b>3</b>	<b>5</b>
8	(13,15,15,17)	<b>4</b>	<b>6</b>	<b>4</b>	<b>6</b>
16	(25,27,33,37)	<b>4</b>	<b>7</b>	<b>4</b>	<b>6</b>
32	(53,67,71,75)	<b>4</b>	<b>7</b>	<b>4</b>	<b>7</b>
64	(135,135,147,163)	<b>4</b>	<b>6</b>	<b>3</b>	<b>5</b>

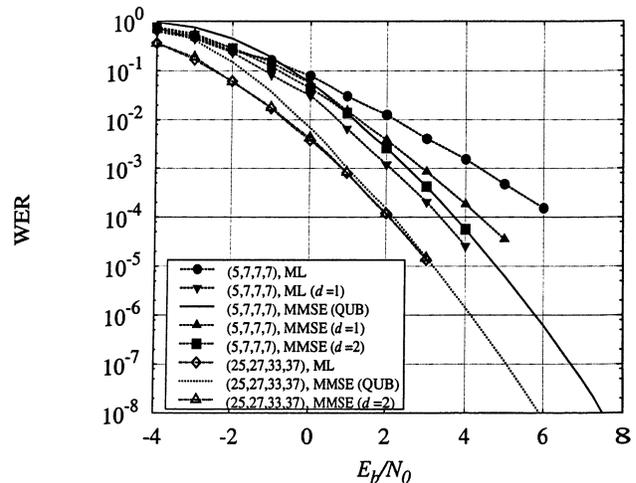


Fig. 5. WER of the WSTCs constructed from the convolutional codes with generators (5, 7, 7, 7) and (25, 27, 33, 37) mapped onto BPSK, independent Rayleigh fading,  $t = r = 4$ .

MMSE-PSP decoding with sufficiently large interleaving delay  $d$  may be either negligible or significant, depending on the code spectral efficiency. On the contrary, the penalty incurred by WMF-PSP decoding is significant for all  $d$  and all spectral efficiencies.  $\diamond$

*Remark 6:* In [39], a computationally efficient trellis-based algorithm for computing the block diversity of trellis codes with cyclic interleaving is given. In Section V, we give another computationally efficient method for calculating the block diversity. By using this method, we computed the block diversity for the binary codes of rate 1/2 and 1/4 tabulated in [38] and mapped onto BPSK and QPSK (with Gray labeling), for different values of  $t$ . Some of these results are reported in Tables I and II. We observe that several codes achieve maximum block diversity and, therefore, are good candidates for WSTC.  $\diamond$

*Example 1:* Fig. 5 shows the WER for WSTCs  $\mathcal{F}(\mathcal{C})$  obtained from the binary convolutional code  $\mathcal{C}$  with generators (5, 7, 7, 7) mapped onto BPSK and transmitted over a channel with  $t = r = 4$ , with interleaving delay  $d = 0, 1, 2$ . The channel matrix has i.i.d. elements  $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$  (independent Rayleigh fading) and each transmitted codeword corresponds to frames of 128 information bits.<sup>7</sup>

<sup>7</sup>All the numerical examples of this paper are obtained with independent Rayleigh fading and frames of 128 information bits. For the sake of brevity, we shall omit explicit mention of these parameters in the following.

The WER curves for  $d = 0$  and 1 with ML decoding and for  $d = 1$  and 2 with MMSE-PSP decoding are shown. It can be easily checked that for  $d = 0$  the resulting WSTC has  $\rho = 2$ , whereas for  $d \geq 1$  it achieves  $\rho = 4$  (see Remark 3). Hence,  $d = 1$  would be sufficient to achieve the best possible WER curve slope under ML decoding. However, the resulting ML decoder must work on the 32-state trellis of the augmented code including delay diversity, with generator matrix given in (15).

Interestingly, the MMSE-PSP decoder achieves a very similar performance by working on the four-state trellis of the underlying component code. However, this is obtained with delay  $d$  generally larger than the minimum required to have  $\rho = \delta$ . This fact is due to the effect of feedback decision errors in the reduced-complexity decoder. The MMSE-PSP decoder with  $d = 2$  performs very close to the solid curve denoted by ‘‘QUB,’’ which corresponds to the semianalytic *quasi-upper bound* (QUB) obtained in Section V by assuming perfect feedback decisions. It is also worthwhile noticing that, in all numerical experiments made, we found that small values of  $d$  suffice to achieve the best WER slope (e.g., in this case  $d = 2$  instead of  $d = 1$  is needed).

Fig. 5 shows also the WER of the WSTC based on the 16-state binary convolutional code with generators (25, 27, 33, 37) mapped onto BPSK. The WER curve for  $d = 0$  with ML decoding, for  $d = 2$  with MMSE-PSP decoding and the QUB assuming perfect feedback decisions are shown. From the binary-rank criterion [26] it is immediate to see that such WSTC scheme achieves rank diversity  $\rho = 4$  for all  $d \geq 0$ . In this case, expanding the code state space by delay diversity and using the low-complexity MMSE-PSP decoder is meaningless, since the component code achieves the optimal rank diversity already for  $d = 0$  and the corresponding ML decoder has the same trellis complexity as the MMSE-PSP decoder. However, it is interesting to notice that, also in this case, the PSP-based decoder achieves almost the same performance as the ML decoder.  $\diamond$

## V. WORD-ERROR RATE (WER) ANALYSIS

In this section, we derive a union upper bound on the average WER of a code  $\mathcal{C}$  over  $\mathcal{X}$  with block length  $N'$  cyclically interleaved over  $t$  parallel AWGN channels with random (but constant over each codeword) SNRs  $\beta_1, \dots, \beta_t$ . Given the parallel channel decomposition (7), this method provides also a *true* upper bound for the WSTC  $\mathcal{S} = \mathcal{F}(\mathcal{C})$  with WMF-PSP decoding, subject to the assumption of perfect feedback decisions. For MMSE-PSP decoding, the parallel AWGN channel model does not hold exactly even assuming perfect feedback decisions, since the noise is neither Gaussian nor independent. Nevertheless, numerical results show that the method generally provides an upper bound also for MMSE-PSP decoding. For this reason, in the following it will be referred to as QUB.

For simplicity, we assume that  $\mathcal{C}$  is geometrically uniform [40], so that we can take any codeword  $\mathbf{c} \in \mathcal{C}$  as the reference codeword. Consider the conditional PEP  $P(\mathbf{c} \rightarrow \mathbf{c}' | \beta_1, \dots, \beta_t)$  given in (10). By using the union bound we get

$$P_w(e | \beta_1, \dots, \beta_t) \leq \sum_{\mathbf{c}' \neq \mathbf{c}} P(\mathbf{c} \rightarrow \mathbf{c}' | \beta_1, \dots, \beta_t). \quad (17)$$

We are mainly interested in the case where  $\mathcal{C}$  is a trellis code with trellis termination. A pairwise error event  $\{\mathbf{c} \rightarrow \mathbf{c}'\}$  is said to be *simple* if the trellis paths corresponding to codewords  $\mathbf{c}$  and  $\mathbf{c}'$  split at a certain step  $1 \leq n_1 \leq N'$ , merge at step  $1 \leq n_2 \leq N'$ , coincide for all steps  $n \leq n_1$  and  $n \geq n_2$ , and remain separated for  $n_1 < n < n_2$ . If a pairwise error event is not simple, it is said to be composite. We have the following.

*Lemma 1:* For any arbitrary  $\beta_1, \dots, \beta_t$ , the right-hand side (RHS) of (17) remains an upper bound on the conditional WER if the sum is restricted to simple error events.

*Proof:* We have to prove that if  $\{\mathbf{c} \rightarrow \mathbf{c}'\}$  is not simple, then it can be eliminated from the union bound without changing the inequality relation. Consider the difference sequence  $\mathbf{d}' = \mathbf{c}' - \mathbf{c}$  (represented as a column vector), and let

$$\boldsymbol{\beta} = \text{diag} \left( \sqrt{\beta_1}, \dots, \sqrt{\beta_t}, \sqrt{\beta_1}, \dots, \sqrt{\beta_t}, \dots, \dots, \sqrt{\beta_1}, \dots, \sqrt{\beta_t} \right)$$

be an  $N' \times N'$  diagonal matrix containing the channel amplitudes  $\sqrt{\beta_1}, \dots, \sqrt{\beta_t}$  repeated periodically  $N'/t$  times (by definition,  $N'$  is an integer multiple of  $t$ ). Assume that  $\{\mathbf{c} \rightarrow \mathbf{c}'\}$  is composite. Then, there exist two codewords,  $\mathbf{c}_1$  and  $\mathbf{c}_2$ , such that  $\mathbf{d}' = \mathbf{d}_1 + \mathbf{d}_2$  and  $(\boldsymbol{\beta} \mathbf{d}_1)^H (\boldsymbol{\beta} \mathbf{d}_2) = 0$  for all  $\boldsymbol{\beta}$ , where  $\mathbf{d}_1 = \mathbf{c}_1 - \mathbf{c}$  and  $\mathbf{d}_2 = \mathbf{c}_2 - \mathbf{c}$ . This is because the path corresponding to a composite event can always be formed by the concatenation of two paths, which differ from the reference path on disjoint supports. Therefore, the nonzero elements in  $\mathbf{d}_1$  and  $\mathbf{d}_2$  are in different positions and the vectors  $\boldsymbol{\beta} \mathbf{d}_1$  and  $\boldsymbol{\beta} \mathbf{d}_2$  are orthogonal for any diagonal matrix  $\boldsymbol{\beta}$ . Let  $\mathbf{v}$  be the received signal corresponding to the transmission of  $\mathbf{c}$ , i.e.,  $\mathbf{v} = \boldsymbol{\beta} \mathbf{c} + \boldsymbol{\nu}$ , where  $\boldsymbol{\nu}$  is the AWGN vector. The pairwise error regions corresponding to the events  $\{\mathbf{c} \rightarrow \mathbf{c}'\}$ ,  $\{\mathbf{c} \rightarrow \mathbf{c}_1\}$ , and  $\{\mathbf{c} \rightarrow \mathbf{c}_2\}$  are given by the inequalities

$$\begin{aligned} \mathcal{E}' &= \{\mathbf{v}: -2\text{Re}\{\mathbf{v}^H \boldsymbol{\beta} \mathbf{d}'\} + |\boldsymbol{\beta} \mathbf{d}'|^2 \leq 0\} \\ \mathcal{E}_1 &= \{\mathbf{v}: -2\text{Re}\{\mathbf{v}^H \boldsymbol{\beta} \mathbf{d}_1\} + |\boldsymbol{\beta} \mathbf{d}_1|^2 \leq 0\} \\ \mathcal{E}_2 &= \{\mathbf{v}: -2\text{Re}\{\mathbf{v}^H \boldsymbol{\beta} \mathbf{d}_2\} + |\boldsymbol{\beta} \mathbf{d}_2|^2 \leq 0\}. \end{aligned} \quad (18)$$

The error event  $\{\mathbf{c} \rightarrow \mathbf{c}'\}$  can be removed from the union bound if  $\mathcal{E}' \subseteq \mathcal{E}_1 \cup \mathcal{E}_2$ . In order to show this, assume that  $\mathbf{v} \in \mathcal{E}' \cap (\mathcal{E}_1 \cup \mathcal{E}_2)^c$ . Then

$$\begin{aligned} -2\text{Re}\{\mathbf{v}^H \boldsymbol{\beta} \mathbf{d}_1\} + |\boldsymbol{\beta} \mathbf{d}_1|^2 &\geq 0 \\ -2\text{Re}\{\mathbf{v}^H \boldsymbol{\beta} \mathbf{d}_2\} + |\boldsymbol{\beta} \mathbf{d}_2|^2 &\geq 0. \end{aligned}$$

By summing the above two inequalities and by recalling the orthogonality property of  $\boldsymbol{\beta} \mathbf{d}_1$  and  $\boldsymbol{\beta} \mathbf{d}_2$ , we get

$$-2\text{Re}\{\mathbf{v}^H \boldsymbol{\beta} \mathbf{d}'\} + |\boldsymbol{\beta} \mathbf{d}'|^2 \geq 0$$

i.e.,  $\mathbf{v} \notin \mathcal{E}'$ , which contradicts the assumption.  $\square$

Next, we find a method to enumerate all simple error events. We can represent the code  $\mathcal{C}$  over a supertrellis, obtained by lumping together  $s$  trellis steps of the original trellis, so that each step of the supertrellis corresponds to  $s$  consecutive steps of the original trellis, and each step of the supertrellis has  $t$  output symbols. For example, if  $\mathcal{C}$  is a trellis code of rate  $b/p$  bits per

symbol, i.e., one transition of the original trellis corresponds to  $p$  output symbols, then  $s = t/p$  (where, for simplicity, we assume that  $p$  divides  $t$ ). The length of the terminated supertrellis is  $N'/t$  steps.

Let  $A_{w_1, \dots, w_t}^{(L)}$  denote the number of simple error events of length  $L$  in the supertrellis (the length  $L$  is measured in supertrellis steps) with SEWs  $w_1, \dots, w_t$  starting at step 1. Then, the total number of simple error events of length  $L$  and SEWs  $w_1, \dots, w_t$  is equal to  $(N'/t - L + 1)A_{w_1, \dots, w_t}^{(L)}$ . By restricting the sum in (17) to simple error events and by using (10) we have

$$\begin{aligned} P_w(e|\beta_1, \dots, \beta_t) &\leq \sum_{L=1}^{N'/t} (N'/t - L + 1) \sum_{w_1, \dots, w_t} A_{w_1, \dots, w_t}^{(L)} Q\left(\sqrt{2 \sum_{j=1}^t \beta_j w_j}\right) \\ &\leq \frac{N'}{t} \sum_{w_1, \dots, w_t} \left(\sum_{L=1}^{\infty} A_{w_1, \dots, w_t}^{(L)}\right) Q\left(\sqrt{2 \sum_{j=1}^t \beta_j w_j}\right) \\ &\stackrel{(a)}{=} \frac{N'}{t\pi} \int_0^{\pi/2} T\left(e^{-\beta_1/\sin^2 \phi}, \dots, e^{-\beta_t/\sin^2 \phi}\right) d\phi \end{aligned} \quad (19)$$

having defined the code multivariate weight enumerator function [41], [42]

$$T(W_1, \dots, W_t) = \sum_{w_1, \dots, w_t} \sum_{L=1}^{\infty} A_{w_1, \dots, w_t}^{(L)} \prod_{j=1}^t W_j^{w_j} \quad (20)$$

and used in (a) of (19) the *preferred* integral expression of the Gaussian tail function [43]

$$Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2 \sin^2 \phi}\right) d\phi.$$

The multivariate weight enumerator function counts all the simple error events starting at step 1 and ending after  $L$  steps of the supertrellis, for all possible lengths  $L = 1, 2, \dots$  (notice that in order to make the computation easier we extend the sum also to lengths  $L > N'/t$ ). A detailed example of the computation of  $T(W_1, \dots, W_t)$  is given in the Appendix.

In order to obtain the average WER, where expectation is done with respect to the joint statistics of the channel SNRs  $\beta_1, \dots, \beta_t$  (they need not be independent), we cannot average term-by-term the weight enumerator. In fact, the parallel channels with cyclic interleaving and random SNRs belong to the class of block-fading channels, for which the union bound averaged with respect to the fading statistics may be very loose or even not converge at all (see [34]). Then, since the conditional WER cannot be larger than 1, we follow the approach of [34] and obtain the final bound on the average WER as

$$P_w(e) \leq E \left[ \min \left\{ 1, \frac{N'}{t\pi} \int_0^{\pi/2} T\left(e^{-\beta_1/\sin^2 \phi}, \dots, e^{-\beta_t/\sin^2 \phi}\right) d\phi \right\} \right] \quad (21)$$

where the expectation is with respect to the joint statistics of  $\beta_1, \dots, \beta_t$ . For small  $t$  and if the joint pdf of  $\beta_1, \dots, \beta_t$  is known, the above expectation can be calculated by numerical integration methods (as in [34]). If  $t$  is large or if the joint statistics of  $\beta_1, \dots, \beta_t$  is not known explicitly, the bound (21) can

be evaluated by Monte Carlo averaging over a large number of realizations of the channel SNRs. In particular, if (21) is used as an upper bound on the WER of the WSTC scheme  $\mathcal{S} = \mathcal{F}(\mathcal{C})$ , the Monte Carlo average is obtained by generating a large number of channel matrices  $\mathbf{H}$  and by calculating the corresponding SNRs either via the QR decomposition (in the case of WMF-PSP decoding) or by using (9) (in the case of MMSE-PSP decoding). In this way, the method can be applied to any arbitrary channel statistics (for example,  $\mathbf{H}$  can be generated by ray-tracing software in order to model a particular scattering environment [3]).

*Remark 7:* The integral with respect to  $\phi$  can be efficiently computed by using Gauss–Chebychev quadrature rules. In fact, by letting

$$g(\sin^2 \phi) = T(e^{-\beta_1/\sin^2 \phi}, \dots, e^{-\beta_t/\sin^2 \phi})$$

we can write

$$\begin{aligned} \frac{1}{\pi} \int_0^{\pi/2} g(\sin^2 \phi) d\phi &= \frac{1}{2\pi} \int_{-1}^1 g(x^2) \frac{dx}{\sqrt{1-x^2}} \\ &\approx \frac{1}{2n} \sum_{i=0}^{n-1} g\left(\cos^2\left(\frac{2i+1}{2n}\pi\right)\right) \\ &\stackrel{n \text{ odd}}{=} \frac{1}{n} \sum_{i=0}^{(n-1)/2-1} g\left(\cos^2\left(\frac{2i+1}{2n}\pi\right)\right) \end{aligned} \quad (22)$$

where the last equality follows by noticing that  $g(0) = 0$ . Then, for an  $n$ -nodes quadrature rule only  $(n-1)/2$  evaluations of the integrand function are needed. Numerical experiments in Section VII show that very good accuracy is obtained already for  $n = 7$ .  $\diamond$

*Remark 8:* In [23], [24], [44], an analysis of the bit-error probability of reduced-state sequence estimation for coded ISI channels based on union-bounding techniques is presented. There, the main problem is to characterize the effect that wrong decisions in the survivors have on the current branch metric: clearly, wrong symbols on the survivor terminating in a given state affect the decision metrics of all paths stemming from that state.

We can provide an intuitive explanation of why this problem basically disappears in the analysis of the WER of PSP-based low-complexity decoding of WSTCs with sufficiently large interleaving delay  $d$ . Consider the situation depicted in Fig. 6 and the decision–feedback interference cancellation given in (6). There are two types of possible wrong decisions in any survivor: errors due to the unmerged section of the survivors and errors due to a past error event, occurred to the left of the survivors merge point (see Fig. 6). Decisions  $\hat{x}_{k,n}$  in (6) have delay  $(k-j)(td-1)$  for  $k = j+1, \dots, t$ . If the survivors always merge at delay smaller than  $td-1$ , only errors due to past error events affect the metric updating and the WER of the PSP-based decoder is equivalent to a genie-aided decoder with perfect feedback decisions. In fact, if  $\hat{x}_{k,n} \neq x_{k,n}$ , a codeword error already occurred for both the PSP-based and for the genie-aided decoders, while if  $\hat{x}_{k,n} = x_{k,n}$  the two decoders have the same branch metric for the  $(j, n)$ th symbol. In other words, the error

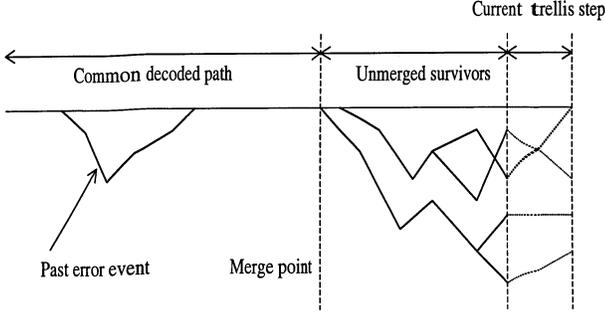


Fig. 6. Sketch representation of the trellis in the PSP-based decoder. Paths are extended from left to right and only survivors are shown.

propagation due to nonperfect feedback decisions has no influence on the WER, provided that  $td - 1$  is so large that the probability of path merge at delay  $< td - 1$  is very high. Notice that for large  $t$ , a large interleaving delay  $d$  is not needed to meet this condition.

On the contrary, in the case of reduced-state sequence estimation for ISI channels, errors are mostly due to unmerged survivors because of the time causality of the ISI channel. For this reason, the WER performance of WSTCs with PSP-based decoding can be predicted very well by the QUB (21) while neglecting feedback decision errors in the ISI setting yields overly optimistic results [23], [24], [44].  $\diamond$

*A Method for Calculating the Block Diversity:* A simple algebraic method to calculate the block diversity of a code  $\mathcal{C}$  with periodic interleaving over  $t$  parallel channels is given by the following.

*Lemma 2:* The block diversity of a code  $\mathcal{C}$  cyclically interleaved over  $t$  channels is equal to  $\delta$  if and only if  $T(W_1, \dots, W_t) = 0$  for all vectors  $(W_1, \dots, W_t) \in \{0, 1\}^t$  with Hamming weight less than  $\delta$  and there exists one vector  $(W_1, \dots, W_t) \in \{0, 1\}^t$  with Hamming weight  $\delta$  for which  $T(W_1, \dots, W_t) > 0$ .

*Proof:* Consider the parallel-channel model with cyclic interleaving (7) and Fig. 4. Let some of the SNRs  $\beta_j$  be equal to 0 and others to  $+\infty$ . If all codewords of  $\mathcal{C}$  are distinguishable after transmission over this parallel on-off channels, then  $P_w(e|\beta_1, \dots, \beta_t) = 0$ . Otherwise,  $P_w(e|\beta_1, \dots, \beta_t) > 0$ . Following the same steps yielding (19), we obtain a Chernoff union bound [30] on the WER as

$$P_w(e|\beta_1, \dots, \beta_t) \leq \frac{N'}{2t} T(e^{-\beta_1}, \dots, e^{-\beta_t}). \quad (23)$$

The Chernoff union bound is asymptotically tight for large SNR. Then, we conclude that the function  $T(e^{-\beta_1}, \dots, e^{-\beta_t})$  is zero if  $P_w(e|\beta_1, \dots, \beta_t) = 0$  and obviously it is positive if  $P_w(e|\beta_1, \dots, \beta_t) > 0$ . Finally, we notice that letting  $\beta_j$  be 0 or  $\infty$  is equivalent to letting  $W_j$  be equal to 1 or 0, respectively.  $\square$

## VI. MUTUAL INFORMATION AND OUTAGE PROBABILITY

Under the quasi-static regime, in the limit for large block length  $N$ , the best possible achievable WER of the multiple-antenna channel (1) is given by the *information outage probability* [1], defined by

$$P_{\text{out}}(\eta) = \Pr(I_{\mathcal{X}}(\mathbf{H}) \leq \eta) \quad (24)$$

where  $I_{\mathcal{X}}(\mathbf{H})$  denotes the *instantaneous* mutual information between the input  $\mathbf{x}_n$ , uniformly distributed over  $\mathcal{X}^t$ , and the output  $\mathbf{y}_n$  for a given channel matrix  $\mathbf{H}$ , given by

$$I_{\mathcal{X}}(\mathbf{H}) = t \log_2 |\mathcal{X}| - E_{\mathbf{x}, \mathbf{z}} \left[ \log_2 \left( \sum_{\mathbf{x}' \in \mathcal{X}^t} e^{2\sqrt{\gamma} \text{Re}\{\mathbf{z}^H \mathbf{H}(\mathbf{x}' - \mathbf{x})\} - \gamma(\mathbf{x}' - \mathbf{x})^H \mathbf{H}^H \mathbf{H}(\mathbf{x}' - \mathbf{x})} \right) \right] \quad (25)$$

( $E_{\mathbf{x}, \mathbf{z}}[\cdot]$  denotes expectation with respect to  $\mathbf{x}$  and  $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$ ).

The computation of  $I_{\mathcal{X}}(\mathbf{H})$  for a given finite signal set  $\mathcal{X}$  is a very demanding task for large  $t$ . In fact, (25) does not admit a simple closed form and requires the evaluation of a  $t$ -dimensional complex integral. Moreover, the integrand function requires the evaluation of the quadratic form  $(\mathbf{x}' - \mathbf{x})^H \mathbf{H}^H \mathbf{H}(\mathbf{x}' - \mathbf{x})$  in the  $|\mathcal{X}|^{2t}$  possible values of the difference vector  $\mathbf{x}' - \mathbf{x}$ . Hence, finding upper and lower bounds on  $I_{\mathcal{X}}(\mathbf{H})$  is essential for evaluating the limit performance of STC under the quasi-static regime.

Given the analogy between this problem and the ISI channel with i.i.d. discrete inputs, we shall make use of the bounds derived in [31]. A simple upper bound on  $I_{\mathcal{X}}(\mathbf{H})$  is obtained by assuming Gaussian i.i.d. inputs with the same average input constraint. We have [4], [2]

$$I_{\mathcal{X}}(\mathbf{H}) \leq I_G(\mathbf{H}) = \log_2 \det(\mathbf{I} + \gamma \mathbf{H} \mathbf{H}^H). \quad (26)$$

Other upper and lower bounds can be derived by considering the parallel channels (7) for appropriate choice of the SNRs  $\beta_j$ . In general, the instantaneous mutual information of the  $t$  AWGN parallel channels with inputs independent and uniform over  $\mathcal{X}$  is given by

$$I_{\text{P.ch.}}(\beta_1, \dots, \beta_t) = t \log_2 |\mathcal{X}| - \sum_{j=1}^t E_{x, \nu} \left[ \log_2 \left( \sum_{x' \in \mathcal{X}} e^{2\sqrt{\beta_j} \text{Re}\{\nu^*(x' - x)\} - \beta_j |x' - x|^2} \right) \right] \quad (27)$$

where  $\nu \sim \mathcal{N}(0, 1)$ . The *matched-filter bound* [31] is obtained by neglecting the mutual interference of the  $t$  transmit antennas. We have

$$I_{\mathcal{X}}(\mathbf{H}) \leq I_{\mathcal{X}}^{\text{mf}}(\mathbf{H}) = I_{\text{P.ch.}}(\gamma |\mathbf{h}_1|^2, \dots, \gamma |\mathbf{h}_t|^2). \quad (28)$$

A class of lower bounds can be obtained from the chain rule of mutual information as follows [31]. For given  $\mathbf{H}$  and any  $r \times t$  matrix  $\mathbf{F}$  for which  $\mathbf{v} = \mathbf{F}^H \mathbf{y}$  we can write

$$\begin{aligned} I_{\mathcal{X}}(\mathbf{H}) &= I(x_1, \dots, x_t; \mathbf{y}) \\ &= \sum_{j=1}^t I(x_j; \mathbf{y} | x_{j+1}, \dots, x_t) \\ &\stackrel{(a)}{\geq} \sum_{j=1}^t I(x_j; \mathbf{v} | x_{j+1}, \dots, x_t) \\ &\stackrel{(b)}{\geq} \sum_{j=1}^t I(x_j; v_j | x_{j+1}, \dots, x_t) \\ &= \sum_{j=1}^t I \left( x_j; v_j - \sqrt{\gamma} \sum_{k=j+1}^t b_{j,k} x_k \right) \end{aligned} \quad (29)$$

where the coefficients  $b_{j,k}$  are defined as in (2) and (6). The inequality (a) holds with equality if  $\mathbf{F}$  is information lossless.

In particular, for  $\mathbf{F}$  given by the WMF (when it exists) or by the MMSE filter defined in Section III equality holds. The inequality (b) holds with equality if  $\mathbf{F}$  is the MMSE filter and the input  $\mathbf{x}$  is a Gaussian white vector [45].<sup>8</sup> The  $j$ th term in the sum of the last line of (29) is the mutual information of the  $j$ th parallel channel in (7). If  $\mathbf{F}$  is the WMF, we get the lower bound

$$I_{\mathcal{X}}(\mathbf{H}) \geq I_{\mathcal{X}}^{\text{wmf}}(\mathbf{H}) = I_{\text{P.ch.}}(\gamma|b_{1,1}|^2, \dots, \gamma|b_{t,t}|^2). \quad (30)$$

If  $\mathbf{F}$  is the MMSE filter, the noise in (7) is not Gaussian since the  $x_j$ 's are discrete and (27) is not directly applicable. However, experimental results and some analytical arguments provided in [46] and in an asymptotic form in [47] show that the residual interference plus noise at the output of the MMSE filter is very close to Gaussian, even if the interfering symbols have a discrete distribution. In [31], experimental evidence shows that by making a Gaussian approximation (GA) of the MMSE filter output (referred to as MMSE-GA in the following) the resulting mutual information is a lower bound on the true mutual information (in [31] this is referred to as *conjectured lower bound*). Motivated by these arguments, we can write

$$I_{\mathcal{X}}(\mathbf{H}) \stackrel{\sim}{\geq} I_{\mathcal{X}}^{\text{mmse}}(\mathbf{H}) = I_{\text{P.ch.}}(\beta_1, \dots, \beta_t)$$

where  $\beta_j$  is given in (9) and  $\stackrel{\sim}{\geq}$  means that inequality is conjectured.

Upper and lower bounds on  $I_{\mathcal{X}}(\mathbf{H})$  yield lower and upper bounds on  $P_{\text{out}}(\eta)$ , respectively. Figs. 7 and 8 show the outage probability bounds versus  $E_b/N_0$  for  $t = r = 4$ , BPSK modulation, and independent Rayleigh fading, with  $\eta = 2$  and 3 bits/channel use, respectively. The curves are obtained by Monte Carlo simulation over several independent generations of the channel matrix  $\mathbf{H}$ . In the figure legend, "GI" denotes Gaussian inputs, "MFB," "WMF," and "MMSE-GA" denote the matched filter, the WMF, and the MMSE-GA bounds for BPSK inputs, respectively, and the points corresponding to the exact outage probability with BPSK are labeled by "BPSK."

For small spectral efficiency (e.g., Fig. 7), the MMSE-GA (conjectured) upper bound is very close to the true outage probability, while for larger  $\eta$  (e.g., Fig. 8) it becomes looser. Remarkably, in these examples the matched-filter bound provides a lower bound tighter than Gaussian inputs. The gap between Gaussian and BPSK inputs increases with  $\eta$  and, for all values of  $\eta$ , the WMF provides a very loose upper bound.

At this point, some qualitative considerations are in order.

- 1) In the derivation of (30) and of the MMSE-GA bound, perfect feedback decisions are a consequence of the mutual information chain rule (29). Hence, these bounds yield also the performance limit achievable by WMF-PSP and MMSE-PSP decoders with *perfect* decision feedback. In particular, since PSP-based decoding is very robust to feedback decision errors, we expect that the outage probability bounds predict well the performance limits of WSTCs under PSP-based decoding.

<sup>8</sup>This can be proved by the purely linear-algebraic identity

$$\det(\mathbf{I} + \gamma \mathbf{H} \mathbf{H}^H) = \prod_{j=1}^t (1 + \beta_j)$$

where  $\beta_j$  is given by (9).

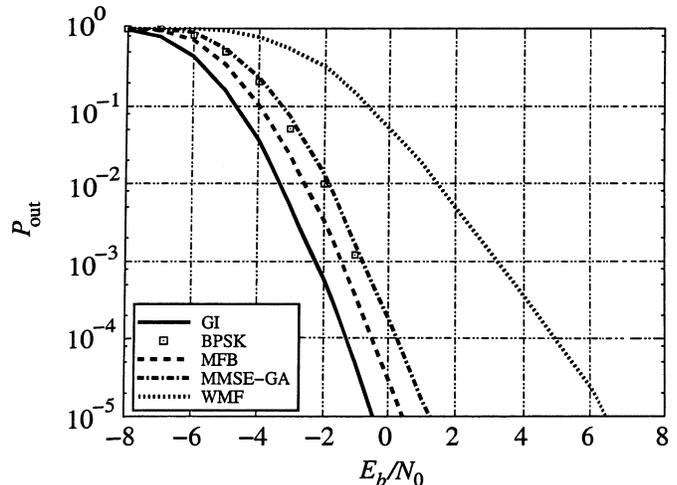


Fig. 7. Outage probability for BPSK, independent Rayleigh fading,  $t = r = 4$  and  $\eta = 2$  bits/channel use.

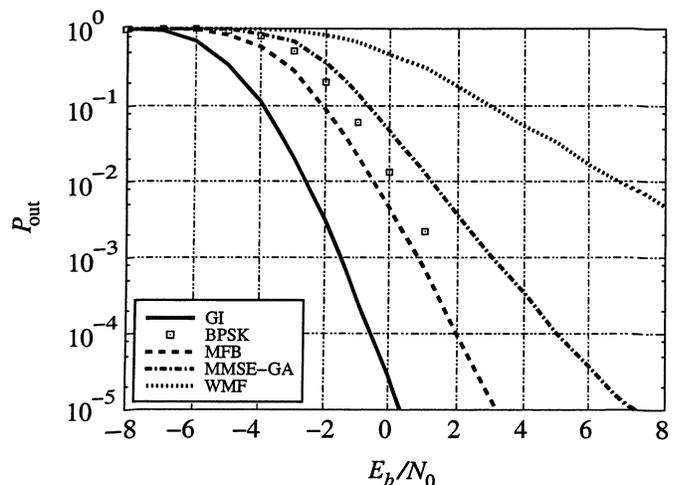


Fig. 8. Outage probability for BPSK, independent Rayleigh fading,  $t = r = 4$  and  $\eta = 3$  bits/channel use.

- 2) In terms of outage probability, the WMF front end with perfect feedback decisions suffers from a large gap with respect to the MMSE front end. On the contrary, it is well known that for any fixed  $\mathbf{H}$  of rank  $t$ , the two front ends (assuming perfect feedback decisions) are capacitywise equivalent as  $\gamma \rightarrow \infty$  (see [31] and references therein). This apparent paradox can be explained by noticing that the convergence  $I_{\mathcal{X}}^{\text{wmf}}(\mathbf{H}) \rightarrow I_{\mathcal{X}}^{\text{mmse}}(\mathbf{H})$  as  $\gamma \rightarrow \infty$  is nonuniform over the ensemble of channel matrices. Therefore,  $\Pr(I_{\mathcal{X}}^{\text{wmf}}(\mathbf{H}) \leq \eta)$  does not converge to  $\Pr(I_{\mathcal{X}}^{\text{mmse}}(\mathbf{H}) \leq \eta)$  even if  $I_{\mathcal{X}}^{\text{wmf}}(\mathbf{H}) \rightarrow I_{\mathcal{X}}^{\text{mmse}}(\mathbf{H})$  for any given  $\mathbf{H}$ .

For the above reason, we expect that also for practical codes the gap between WMF-PSP and MMSE-PSP decoding does not vanish as  $\gamma \rightarrow \infty$ . Curiously, several previous works on layered STCs [14], [21] reported results for the WMF front end with decision feedback, referred to as the "nulling and canceling" approach. Our results show that using the WMF instead of the MMSE front end in the quasi-static environment can be very dangerous.

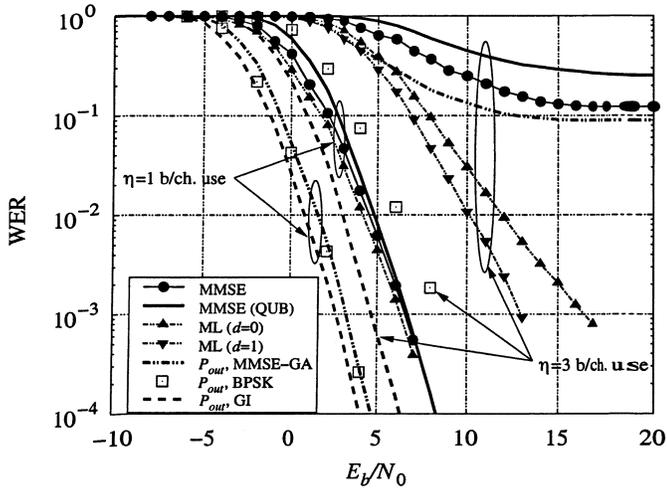


Fig. 9. WER of WSTCs and trellis STCs with MMSE-PSP and ML decoding, respectively, and related outage probabilities for  $t = 4$ ,  $r = 2$ , independent Rayleigh fading. The codes for spectral efficiency  $\eta = 1$  are based on the convolutional code with generators (25, 27, 33, 37) mapped onto BPSK. The codes for  $\eta = 3$  are based on the four-state code with rate 3/4 and generator matrix given in (31), mapped onto BPSK.

- 3) For values of  $\eta$  where the MMSE-GA is close to the true outage probability (e.g., Fig. 7), we expect that good WSTCs with MMSE-PSP decoding perform close to good STCs with ML decoding. On the contrary, for values of  $\eta$  at which the MMSE-GA is far from the true outage probability (e.g., Fig. 8) we expect that WSTCs with MMSE-PSP decoding suffer from a large performance degradation with respect to STCs with ML decoding.

*Example 2:* The last observation is well illustrated by the example of Fig. 9, comparing the WER of BPSK codes with spectral efficiency  $\eta = 1$  and 3 bits/channel use in the case  $t = 4$ ,  $r = 2$  with the corresponding  $P_{\text{out}}(\eta)$ . For  $\eta = 1$ , we consider the 16-states binary convolutional code of rate 1/4 of Table II. For  $\eta = 3$ , we consider the four-state Wyner–Ash code [48] of rate 3/4, whose generator matrix is given by

$$\mathbf{G}(D) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1+D & 1 & 0 & 0 \\ 0 & 1+D & 1 & 0 \end{bmatrix}. \quad (31)$$

We consider WSTCs with  $d = 2$  and MMSE-PSP decoding, and trellis STCs with ML decoding.

For  $\eta = 1$ , the MMSE-GA bound is very close to the true outage probability, while for  $\eta = 3$  it is very far away and shows an error floor. This behavior is reflected by the performance of the MMSE-PSP decoder: for  $\eta = 1$ , ML and MMSE-PSP decoding yield almost the same performance, while for  $\eta = 3$ , MMSE-PSP decoding compares very poorly with respect to ML decoding and shows an error floor similar to that of the MMSE-GA bound on the outage probability.

Such behavior was noticed in all cases examined. We conclude that the ability of MMSE-PSP decoding of performing close to ML, for given  $t$  and  $r$ , depends essentially on the code spectral efficiency and can be predicted very accurately by the behavior of the MMSE-GA bound on the outage probability.

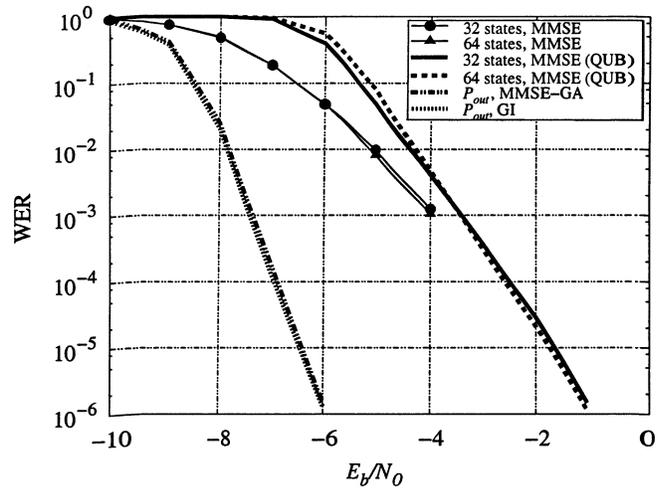


Fig. 10. WER of WSTCs based on the 32- and 64-states codes of rate 1/4 of Table II mapped onto QPSK, and related outage probabilities for  $t = r = 8$ , independent Rayleigh fading, and spectral efficiency  $\eta = 4$  bits/channel use.

A final remark is devoted to ML decoding. Since the 16-states code has rank diversity  $\rho = 4$  (see [26]), then  $d = 0$  is sufficient to achieve the optimal WER slope under ML decoding. In this case, the ML Viterbi decoder works on the 16-state trellis of the code. On the contrary, the four-state Wyner–Ash code has rank diversity  $\rho = 1$ ,<sup>9</sup> which is strictly less than the bound (13). However, the 64-state trellis STC resulting from the concatenation of the same code with delay diversity with  $d = 1$  achieves  $\rho = 2$ , which is optimal as it meets the bound (13) with equality.<sup>10</sup> In this case, the ML Viterbi decoder works on the 64-state trellis of the augmented encoder, including delay diversity. This effect can be observed by comparing the WER curves corresponding to  $\eta = 3$  and ML decoding for  $d = 0$  and  $d = 1$ , in Fig. 9.  $\diamond$

## VII. PERFORMANCE EXAMPLES

In this section, we present a selected set of numerical examples, aimed at pointing out some interesting facts.

*Relevance of the Rank-Diversity Criterion:* It has been recently observed that when both  $r$ ,  $t$  are large the rank diversity is not a very meaningful code design criterion [50]. This fact is illustrated in Fig. 10, showing the WER of the WSTCs based on the binary convolutional codes with 32 and 64 states and rate 1/4 of Table II with QPSK modulation, for  $t = r = 8$ . In this example, the interleaving delay is very large ( $d = 10$ ) in order to eliminate any effect of feedback decision errors and have  $\rho = \delta$  for the corresponding WSTCs. From Table II, we see that the 32-state code achieves optimal block diversity  $\delta = 7$ , while the 64-state achieves only  $\delta = 5$ . Nevertheless, in the range of WER shown, both simulation and the QUB are almost identical for the two codes.

On the contrary, it can be observed from the examples of Figs. 5 and 9 that for smaller values of  $t$  and  $r$ , the rank diversity is a very relevant criterion, at least for the family of codes considered here.

<sup>9</sup>This can be immediately seen from the fact that the trellis of the four-state Wyner–Ash code has parallel transitions.

<sup>10</sup>This statement can be proved by applying the generalized stacking construction result of [49]. We omit the proof for the sake of space limitation.

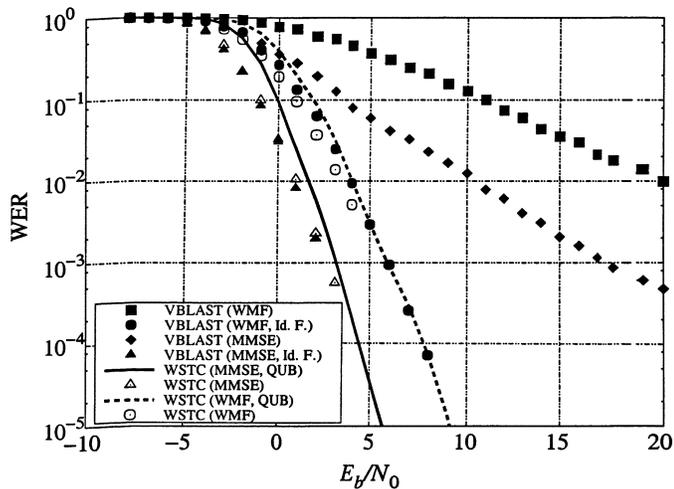


Fig. 11. WER of WSTC and V-BLAST schemes with code (23, 25) and QPSK for  $t = r = 4$ , independent Rayleigh fading, and spectral efficiency  $\eta = 4$  bits/channel use.

*Comparison With V-BLAST:* In [21], a space-time scheme nicknamed “V-BLAST” and its low-complexity decision-feedback detection scheme are presented. V-BLAST is equivalent to a WSTC with  $d = 0$ , but decision-feedback interference cancellation is obtained by feeding back symbol-by-symbol predecisions, without the aid of PSP. Since the order of symbol-by-symbol decisions is not dictated by the trellis time ordering of the underlying code, predecisions are made in an order that depends on  $\mathbf{H}$ , with the aim of limiting error propagation in the feedback. Namely, the columns of  $\mathbf{H}$  are permuted so that the “QR” decomposition of the permuted matrix yields WMF channel gains such that  $\min_{j=1, \dots, t} |b_{j,j}|^2$  is maximized. Remarkably, this ordering can be calculated by a simple “greedy” algorithm that maximizes at each step  $j = t, t-1, \dots, 1$  the SNR of the  $j$ th parallel channel [21]. In the case of MMSE front end, the same algorithm yields a sequence of SNRs  $\beta_j$  for which  $\min_{j=1, \dots, t} \beta_j$  is maximized asymptotically, for large SNR. As in classical decision feedback equalization, if the V-BLAST detector is concatenated with a decoder, hard predecisions are used only for the purpose of decision feedback, and soft values are fed to the decoder.

In Fig. 11, the WER of a WSTC obtained from the 16-states binary convolutional code with rate 1/2 of Table I mapped onto QPSK and  $t = r = 4$  antennas is compared with the performance of the scheme obtained by concatenating the same code with the V-BLAST detector. Both MMSE and WMF front ends are considered. Simulations of the PSP-based decoder for the WSTC (obtained for  $d = 2$ ) are in perfect agreement with the corresponding QUB, showing that the effect of feedback decision errors is negligible for the PSP-based decoder. The V-BLAST performance was obtained by simulation. For comparison, we also show the WER resulting from a genie-aided V-BLAST detector with ideal feedback decisions (curves denoted by “Id. F.”), that is very similar to the WER achieved by the WSTC without the genie. This shows that the large performance degradation of V-BLAST with respect to the WSTC is due to error propagation in the decision feedback: while the PSP-based decoder is very effective in preventing such propaga-

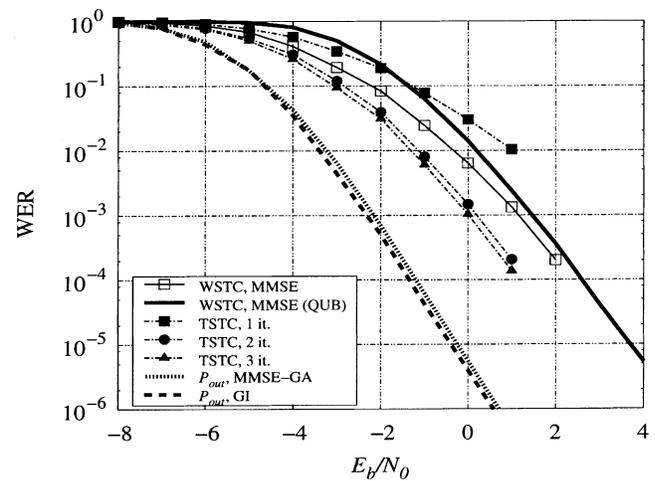


Fig. 12. WER of WSTC and TSTC schemes with code (25, 27, 33, 37) and QPSK for  $t = r = 4$ , independent Rayleigh fading, and spectral efficiency  $\eta = 2$  bits/channel use.

tion, the special detection ordering of V-BLAST is not. Moreover, the complexity of the “greedy” detection ordering algorithm is larger than the extra complexity of the Viterbi algorithm due to PSP, for large  $t$ . Then, WSTC with PSP-based decoding is not only more effective, but it is also simpler than V-BLAST.

*Comparison With TSTC:* Fig. 12 compares the performance of WSTC and TSTC [20] based on the 16-state binary convolutional code of rate 1/4 of Table II mapped onto QPSK, with  $t = r = 4$  antennas. Each iteration of the iterative decoder for TSTC proposed in [20] involves the calculation of a linear MMSE front-end filter (complexity  $O(t^3)$ ) and soft-in soft-out decoding of the underlying convolutional code, implemented by the forward-backward algorithm [51]. The computation complexity of the MMSE filter for TSTC and WSTC is almost identical, and the PSP decoder is slightly less complex than the forward-backward algorithm. Hence, one iteration of the TSTC decoder corresponds to (roughly) the same complexity of the whole WSTC decoder. Fig. 12 shows that the WSTC scheme yields a clear performance advantage over TSTC with one iteration (i.e., for the same decoder complexity), while the TSTC scheme achieves an overall better performance if more iterations are allowed (for this example, no improvement was observed for more than three TSTC decoder iterations). Interestingly, in all our numerical experiments we observed that TSTC with several decoder iterations achieves the same WER slope of WSTC. Therefore, the gain of TSTC over WSTC is constant in the high-SNR region (e.g., in Fig. 12 it is  $\approx 1.2$  dB).

*Handling the Case  $r < t$  via Space-Time Block Precoding:* When  $r < t$ , the WMF front end is not defined and the MMSE front end is not very effective in providing interference mitigation. Hence, it makes sense to consider some form of space-time block precoding to shape the overall channel (including precoding) so that it is better suited to the low-complexity decoding algorithm. We hasten to say that this general consideration can be applied to several situations where some specific signal-processing technique is used for low-complexity suboptimal decoding (e.g., in the case of TSTC), and that a comprehensive investigation of block precoders adapted

to some outer coding scheme for the quasi-static channel is far beyond the scope of this paper. However, in this example we provide some results based on the family of linear dispersion codes (LDCs) recently proposed by Hochwald and Hassibi in [19]. An LDC precoder takes blocks of  $Q$  modulation symbols and maps them onto the complex  $t \times T$  matrix signals

$$\mathbf{S} = \sum_{q=1}^Q (x_q \mathbf{C}_q + x_q^* \mathbf{D}_q)$$

where  $\mathbf{C}_q$  and  $\mathbf{D}_q$  are complex  $t \times T$  generator matrices. Then,  $\mathbf{S}$  is transmitted column-by-column over  $T$  channel uses. The spectral efficiency of the system with precoding is  $\eta = \frac{Q}{T}R$ , where  $R$  is the rate of the outer code (in bits per symbol). LDC precoding shapes the original  $t$ -input  $r$ -output complex channel into a virtual  $2Q$ -input  $2rT$ -output real channel.

Fig. 13 compares the WER performance of some schemes for the case  $t = 4$ ,  $r = 1$ , with spectral efficiency  $\eta = 1$  bit/channel use. The curves denoted by “MMSE, LD” refer to the concatenation of the WSTC obtained from the 16-states binary convolutional code of rate  $1/2$  in Table I mapped onto QPSK with a LDC designed in [19] for the  $t = 4$ ,  $r = 1$  channel, with  $Q = 4$  and  $T = 4$ .<sup>11</sup> We notice that this scheme is outperformed by the simpler WSTC obtained by using directly the 16-states code of rate  $1/4$  of Table II mapped onto BPSK, without precoding (curve denoted by “MMSE, No prec.”). Moreover, the latter scheme has almost the same performance (in the range of  $\text{WER} \geq 10^{-3}$ ) of a very simple scheme where the 16-states code of rate  $1/2$  of Table I mapped onto QPSK is transmitted by using only one transmit antenna in a round-robin fashion over all  $\binom{4}{1}$  combinations of one out of four antennas, and ML decoding is used at the receiver (curve denoted by “1 Tx, cyclic”).

The disappointing results of LDC precoding in this case can be explained by comparing the outage probability of the precoded channel with that of the original  $4 \times 1$  channel, showing a large gap even in the case of Gaussian inputs (i.e., assuming optimal outer coding and ML decoding).

Fig. 14 compares the WER performance of some schemes for the case  $t = 4$ ,  $r = 2$  with spectral efficiency  $\eta = 2$  bits/channel use. The curves denoted by “MMSE, LD” refers to the concatenation of the WSTC obtained from the 16-states binary convolutional code of rate  $1/2$  in Table II mapped onto QPSK with an LDC designed in [19] for the  $t = 4$ ,  $r = 2$  channel, with  $Q = 8$  and  $T = 4$ . We show also results for the simpler WSTC obtained by using directly the 16-states code of rate  $1/4$  of Table II mapped onto QPSK, without precoding (curve denoted by “MMSE, No prec.”) and for a scheme where the 16-states code of rate  $1/2$  of Table I mapped onto QPSK is transmitted by using only two transmit antennas in a round-robin fashion over all  $\binom{4}{2}$  combinations of two out of four antennas, and ML decoding is used at the receiver (curve denoted by “2 Tx, cyclic”).

In this case, the scheme with LDC precoding outperforms the other schemes. Again, this fact is explained by comparing the outage probability of the precoded channel with that of the original  $4 \times 2$  channel. These curves show that the gap between the LDC precoded channel with respect to the original channel in

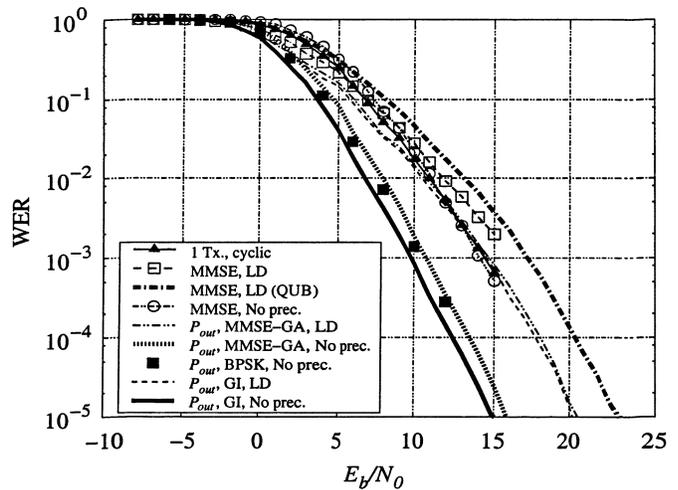


Fig. 13. WER of some WSTC schemes with and without LDC precoding in independent Rayleigh fading,  $t = 4$ ,  $r = 1$ , and spectral efficiency  $\eta = 1$  bit/channel use.

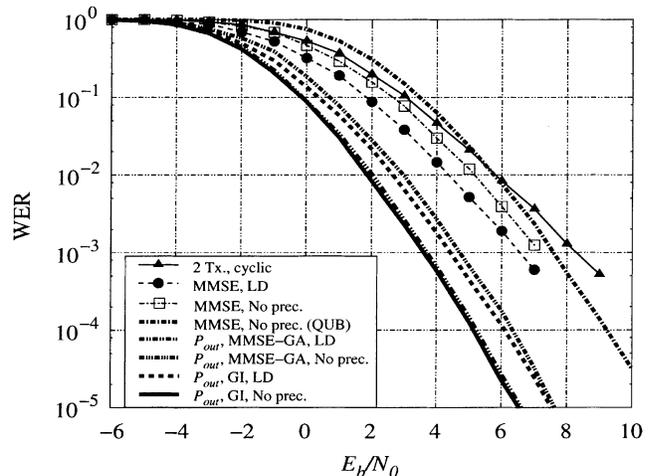


Fig. 14. WER of some WSTC schemes with and without LDC precoding in independent Rayleigh fading,  $t = 4$ ,  $r = 2$ , and spectral efficiency  $\eta = 2$  bits/channel use.

the case of Gaussian inputs (i.e., assuming optimal outer coding and ML decoding) is quite small, thus, showing that the LDC precoder is well suited for the quasi-static channel under the outage probability performance criterion.

The LDC precoders in [19], including those used in the above examples, were designed in order to maximize the average mutual information at a given SNR, and not to minimize the outage probability for a given spectral efficiency. Then, it is no surprise that they might yield disappointing performance in some cases. However, the example of Fig. 14 shows the potential interest of block precoding, especially if this is used in order to enable a low-complexity outer coding/decoding scheme, and calls for the design of good block precoders with respect to the outage probability criterion.

## VIII. CONCLUSION

A new scheme, nicknamed “wrapped” STC, was proposed. This scheme belongs to the class of generalized space-time layering schemes, and it is suited to a large number of antennas

<sup>11</sup>The generator matrices of the LDC precoders used in these examples were kindly provided to us by the authors of [19].

with low-complexity decoding, based on MMSE decision–feedback interference mitigation coupled with PSP-based Viterbi decoding. We showed that any trellis code with optimal block diversity can be turned into a WSTC with optimal rank diversity, with the advantage that block diversity is easy to check and optimal block diversity is easily achieved by several well-known trellis codes.

We provided a simple and accurate (quasi-)upper bound on the performance of WSTCs and an efficient method for calculating the block diversity of trellis codes, based on the multivariate weight enumerator function of the underlying component code.

Finally, we considered some upper and lower bounds on the information outage probability with discrete i.i.d. inputs, and we pointed out the relation between the tightness of the MMSE-GA (quasi-)upper bound with the ability of the MMSE-PSP decoding to perform close to ML decoding. This result is still in the form of numerical evidence, and it calls for a deeper information-theoretic investigation.

Performance examples of the proposed scheme constructed from well-known binary linear convolutional codes were provided. WSTCs compare favorably with respect to coded V-BLAST and TSTC of similar complexity. In particular, we showed that V-BLAST is prone to decision-feedback error propagation even if the detection ordering algorithm of [21] is used, while the PSP-based decoder is almost not affected by error propagation even for very small interleaving delay. The recently proposed TSTC scheme with iterative decoding yields generally better WER at the cost of larger decoding complexity. Indeed, one iteration of the TSTC decoder has complexity equivalent to the whole MMSE-PSP decoder for WSTC, and the TSTC scheme needs at least two iterations in order to outperforms the WSTC scheme.

Furthermore, we investigated the concatenation of WSTC with space–time block precoders, which might be useful in the case  $r < t$ . Through some examples based on LDCs we showed that block precoding might yield performance improvements, while retaining the low-complexity decoding scheme of the outer WSTC. However, new block precoders specifically designed for the quasi-static channel (i.e., according to the outage probability criterion) should be investigated.

The gap between the WER of WSTCs obtained from simple convolutional codes and the information outage probability indicates that more work is needed in order to design good component codes for the WSTC scheme. This is especially true for high-rate trellis codes, where the design of maximal block-diversity codes has been rarely addressed (see [33]).

#### APPENDIX

##### CALCULATION OF THE MULTIVARIATE WEIGHT ENUMERATOR FUNCTION

For the sake of completeness, we give here an example of calculation of  $T(W_1, \dots, W_t)$ . Further details and examples can be found in [41], [42]. Consider the four-state binary linear convolutional codes with rate 1/2 and generators (5, 7) and let  $t = 4$ . Fig. 15 shows the trellis of the original code and the supertrellis obtained by lumping together  $s = 2$  steps of the

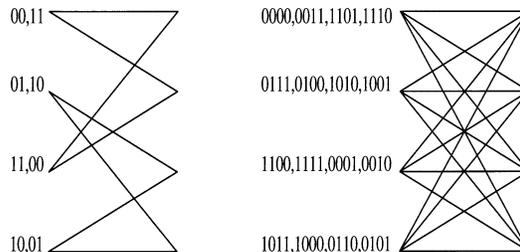


Fig. 15. Original trellis (left) and supertrellis (right) corresponding to  $s = 2$  consecutive steps for the binary linear convolutional code with generators (5, 7). The transitions labels next to each state, read from left to right, correspond to the transitions stemming from the state from top to bottom.

original trellis, so that  $t = 4$  output symbols correspond to each transition of the supertrellis (notice that the supertrellis has the same number of states of the original trellis but more transitions, and in general may have parallel transitions even if the original trellis has not).

Consider the (0, 0) transition, with code symbols

$$(c_1, c_2, c_3, c_4) = (1, 1, 1, 1)$$

(we assume BPSK modulation with the mapping  $\mu: \{0, 1\} \rightarrow \{\pm 1\}$ ). Any transition of the supertrellis with code symbols  $(c'_1, c'_2, c'_3, c'_4)$  is labeled by the monomial  $\prod_{j=1}^t W_j^{|c'_j - c_j|^2/4}$  (transitions between nonconnected states are labeled by the zero monomial). Then, we define the state variables  $V_1, \dots, V_3$  corresponding to states 1,  $\dots$ , 3, and we split state 0 into two states associated to the input variable  $X$  and to the output variable  $Y$ , respectively. The corresponding state equation is given by

$$\begin{bmatrix} Y \\ V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} 1 & W_2 W_3 W_4 & W_1 W_2 & W_1 W_3 W_4 \\ W_3 W_4 & W_2 & W_1 W_2 W_3 W_4 & W_1 \\ W_1 W_2 W_4 & W_1 W_3 & W_4 & W_2 W_3 \\ W_1 W_2 W_3 & W_1 W_4 & W_3 & W_2 W_4 \end{bmatrix} \cdot \begin{bmatrix} X \\ V_1 \\ V_2 \\ V_3 \end{bmatrix} \quad (\text{A1})$$

and it is obtained by letting each state/output variable equal the sum over all incoming transitions of the product of the transition weight by the state/input variable originating the transition. The multivariate weight enumerator function is finally given by

$$T(W_1, W_2, W_3, W_4) = Y/X - 1$$

where the term  $-1$  eliminates the contribution of the correct path, with Euclidean weight zero. This can be obtained by eliminating  $V_1, V_2, V_3$  from the system of linear equations (A1).

Generally speaking, by following the above recipe we can always put the state equations in the  $2 \times 2$  block form

$$\begin{bmatrix} Y \\ V_1 \\ \vdots \\ V_{S-1} \end{bmatrix} = \begin{bmatrix} D & C \\ B & A \end{bmatrix} \begin{bmatrix} X \\ V_1 \\ \vdots \\ V_{S-1} \end{bmatrix} \quad (\text{A2})$$

( $S$  denotes the number of states), where  $D$ ,  $\mathbf{C}$ ,  $\mathbf{B}$  and  $\mathbf{A}$  are  $1 \times 1$ ,  $1 \times (S-1)$ ,  $(S-1) \times 1$ , and  $(S-1) \times (S-1)$  polynomial matrices in the variables  $W_1, \dots, W_t$ , respectively. Then

$$T(W_1, \dots, W_t) = \mathbf{C}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + D - 1. \quad (\text{A3})$$

In order to compute  $T(W_1, \dots, W_t)$  at  $W_j = e^{-\beta_j/\sin^2 \phi}$ , for  $j = 1, \dots, t$ , for calculating the union bound (21), it is computationally more convenient to substitute the arguments inside the matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $D$  and then evaluate (A3) numerically.

#### ACKNOWLEDGMENT

The authors would like to thank B. Hochwald and B. Hassibi for providing the example of LDC precoder and the Associate Editor and the anonymous reviewers whose comments were very helpful in improving the paper. G. Colavolpe is greatly indebted with Prof. G. Prati, who made possible the cooperation from which this work arose.

#### REFERENCES

- [1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.
- [3] C. Chuah, D. Tse, and J. Kahn, "Capacity of multi-antenna array systems in indoor wireless environment," in *Proc. GLOBECOM '98*, Sydney, Australia, Nov. 1998.
- [4] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [5] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1273–1289, May 2001.
- [6] E. Visotsky and U. Madhow, "Space-time transmit strategies and channel feedback generation for wireless fading channels," in *Proc. 34th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2000.
- [7] H. Bolcskei, D. Gesbert, and A. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.
- [8] R. Muller, "A random matrix theory of communication via antenna arrays," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2495–2506, Sept. 2002.
- [9] T. Marzetta and B. 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.
- [10] L. Zheng and D. N. C. Tse, "Communicating on the Grassmann manifold: A geometric approach to the noncoherent multiple antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.
- [11] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.
- [12] R. Knopp and G. Caire, "Power control schemes for TDD systems with multiple transmit and receiving antennas," in *Proc. GLOBECOM '99*, Rio de Janeiro, Brasil, Dec. 1999.
- [13] V. Tarokh, N. Seshadri, and 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, Mar. 1998.
- [14] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Syst. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [15] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, p. 1456, July 1999.
- [16] J. Grimm, M. Fitz, and J. Krogmeyer, "Further results in space-time coding for Rayleigh fading," in *Proc. 36th Allerton Conf. Communication, Control and Computing*, Sept. 1998.
- [17] S. Baro, G. Bauch, and A. Hansmann, "Improved codes for space-time trellis-coded modulation," *IEEE Comm. Lett.*, vol. 4, no. 1, pp. 20–22, Jan. 2000.
- [18] X. Lin and R. Blum, "On the design of space-time codes employing multiple trellis coded modulation," in *Proc. 2000 Conf. Information Science and Systems*, Princeton, NJ, Mar. 2000.
- [19] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804–1824, July 2002.
- [20] H. El Gamal and R. Hammons, "A new approach to layered space-time coding and signal processing," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 2321–2334, Sept. 2001.
- [21] J. Foschini, G. Golden, R. Valenzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1841–1852, Nov. 1999.
- [22] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, pp. 354–364, Feb. 1995.
- [23] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, pp. 428–436, May 1989.
- [24] R. Ashkenazi and S. Shamai, "Suboptimal detection for intersymbol interference inflicted cable channels," *AEU*, vol. 51, no. 5, pp. 246–254, May 1997.
- [25] P. Chevillat and E. Eleftheriou, "Decoding of trellis-encoded signals in the presence of intersymbol interference and noise," *IEEE Trans. Commun.*, vol. 37, pp. 669–676, July 1989.
- [26] R. Hammons and H. El Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 524–542, Mar. 2000.
- [27] E. Biglieri, D. Divsalar, P. McLane, and M. Simon, *Introduction to Trellis-Coded Modulation With Applications*. New York: Macmillan, 1991.
- [28] D.-S. Shiu and J. Kahn, "Layered space-time codes for wireless communications using multiple transmit antennas," in *IEEE Int. Conf. Communications*, Vancouver, BC, Canada, June 1999.
- [29] N. Seshadri and J. Winters, "Two signaling schemes for improving the error performance of frequency-division-duplex transmission systems using transmitter antenna diversity," *Int. J. Wireless Inform. Networks*, vol. 1, no. 1, pp. 49–60, 1994.
- [30] A. Viterbi and J. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.
- [31] S. Shamai (Shitz) and R. Laroia, "The intersymbol interference channel: Lower bounds on capacity and channel precoding loss," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1388–1404, Sept. 1996.
- [32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.
- [33] R. Knopp and P. Humblet, "On coding for block fading channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 189–205, Jan. 2000.
- [34] E. Malkamaki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1643–1646, July 1999.
- [35] R. Wesel, X. Liu, and W. Shi, "Trellis codes for periodic erasures," *IEEE Trans. Commun.*, vol. 48, pp. 938–947, June 2000.
- [36] R. E. Blahut, *Principles and Practice of Information Theory*. New York: Addison-Wesley, 1987.
- [37] R. Knopp and P. Humblet, "Maximizing diversity in block fading channels," in *IEEE Int. Conf. Communications*, Montreal, QC, Canada, June 1997, pp. 647–651.
- [38] J. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [39] A. Lapidoth, "The performance of convolutional codes on the block erasure channel using various finite interleaving techniques," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1459–1473, Sept. 1994.
- [40] D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, Sept. 1991.
- [41] Y. Leung, S. Wilson, and J. Ketchum, "Multifrequency trellis coding with low delay for fading channels," *IEEE Trans. Commun.*, vol. 41, pp. 1450–1458, Oct. 1993.
- [42] R. Wesel and X. Liu, "Analytic techniques for periodic trellis codes," in *Proc. 36th Allerton Conf. Communications, Control and Computing*, Sept. 1998, pp. 39–48.

- [43] M. K. Simon and M. S. Alouini, "A unified approach to the performance analysis of digital communication over generalized fading channels," *Proc. IEEE*, vol. 86, pp. 1860–1877, Sept. 1998.
- [44] W. Sheen and G. Stuber, "Error probability for reduced-state sequence estimation," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 571–578, Apr. 1992.
- [45] M. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple access channel," in *Proc. Asilomar Conf.*, Pacific Grove, CA, Nov. 1997.
- [46] V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inform. Theory*, vol. 43, pp. 858–871, May 1997.
- [47] J. Zhang, E. Chong, and D. N. C. Tse, "Output MAI distribution of linear MMSE multiuser receivers in CDMA systems," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1128–1144, Mar. 2001.
- [48] R. McEliece, "The algebraic theory of convolutional codes," in *Handbook of Coding Theory*. Amsterdam, The Netherlands: North-Holland, 1998.
- [49] H. El Gamal and R. Hammons, "On the design of algebraic space–time codes for MIMO block fading channels," *IEEE Trans. Inform. Theory*, submitted for publication.
- [50] E. Biglieri, G. Taricco, and A. Tulino, "Performance of space–time codes for a large number of antennas," *IEEE Trans. Inform. Theory*, to be published.
- [51] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT–20, pp. 284–287, Mar. 1974.