On Space-Time Coding for Quasi-Static Multiple-Antenna Channels

Giuseppe Caire   and   **Giulio Colavolpe**

Mobile Communications Group   Università di Parma

Institut EURECOM   Dipartimento di Ingegneria dell'Informazione

2229 Route des Cretes   Parco Area delle Scienze 181/A

06560 Valbonne, FRANCE   43100 Parma - ITALY

Tel: +33 (0)4 93002904   Tel: +39 0521 905744

Fax: +33 (0)4 93002627   Fax: +39 0521 905758

E-mail: giuseppe.caire@eurecom.fr   Email: giulio@unipr.it

March 7, 2001

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. The new scheme includes both trellis space-time codes and layered space-time codes as special cases. When the number of transmit antennas is not larger than the number of receive antennas our scheme can be efficiently decoded by minimum mean-square error (MMSE) decision-feedback interference cancellation coupled with Viterbi decoding through the use of per-survivor processing (PSP). 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. We provide an upper bound to the word-error rate of our scheme which is both accurate and easy to evaluate. Then, we find upper and lower bounds to the information outage probability with *discrete 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 front-end when coupled with the decision-feedback PSP-based decoder. Our scheme yields a large performance gain with respect to coded V-BLAST of similar complexity, and it can be easily coupled with the recently proposed *linear-dispersion precoding* to handle the case of a number of receive antennas smaller than the number of transmit antennas.

Keywords: Space-time coding, multiple-antenna systems, per-survivor processing.

1   Introduction

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

Roughly speaking, works on multiple antennas can be classified depending on the assumptions on the channel state information (CSI) available at the transmitter and at the receiver. The *ergodic* capacity [1] of a frequency non-selective 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-SNR is given by $C = \min\{r, t\} \log_2 \mathrm{SNR} + O(1)$ bit/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 non-ergodic, 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 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 more realistic 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 channel is $C = k(1 - k/T) \log_2 \mathrm{SNR} + O(1)$, where $k = \min\{r, t, T/2\}$. They show also that capacity is maximized by using no more than k transmit antennas. In particular, having $t > r$ antennas does not improve capacity. 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 for channel estimation (e.g., by explicitly sending training symbols), and by a mismatched receiver that treats the training-based estimated channel as if it was 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 (non-ergodic) 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 quite realistic assumption in several 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 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, which is actually the way the large majority of existing wireless systems work today. Without further questioning about 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, 14, 15, 16, 17, 18, 19]). In this paper we propose a new scheme nicknamed "wrapped" space-time coding that includes both the trellis space-time codes of [13] and the layered space-time codes of [14] as special cases. For the case $r \geq t$, our scheme can be efficiently decoded by minimum mean-square error (MMSE) decision-feedback interference cancellation coupled with Viterbi decoding, through the use of per-survivor processing (PSP) [20]. We discuss the code design for the new scheme, and show that finding codes with optimal diversity is actually much easier than for conventional trellis space-time codes. We show many examples where the maximum possible diversity is achieved by well-known trellis codes. We provide an upper bound to the word-error rate of our scheme which is both very close to simulations and easy to evaluate (at least for geometrically uniform codes). Then, in order to understand better the behavior of the word-error rate, we find upper and lower bounds to the information outage probability with discrete i.i.d. inputs. We show that the MMSE front-end yields a large advantage over the whitened matched filter (WMF) front-end if coupled with the decision-feedback PSP-based decoder. Finally, we provide numerical examples showing the accuracy of our analytical bound and the effectiveness of the proposed space-time coding scheme. We compare our scheme with V-BLAST [21], which is also based on decision-feedback interference cancellation but it does not exploit PSP, and we show that the latter suffers severely from error propagation in the feedback decisions while the proposed scheme does not. Also, we consider the case $r < t$ and we show that our scheme can work as outer coding where inner coding (or "precoding") is obtained by the recently proposed linear dispersion codes [19].

The paper is organized as follows: in Section 2 we review the basic trellis and layered space-time coding schemes and we introduce the new "wrapped" scheme. In Section 3 we describe the details of the PSP-based decoding scheme. In Section 4 we discuss the code design for our scheme and in Section 5 we present the semi-analytic bound on the word-error rate. In Section 6 we present results on the outage probability with discrete inputs and in Section 7 we show some numerical results, the comparison with V-BLAST and the use of linear-dispersion precoding. Conclusions are summarized in Section 8 and some side results are collected in Appendices A and B for the sake of completeness.

2 Coding for the Quasi-Static Multiple-Antenna Channel

The multiple-input multiple-output (MIMO) system with t transmitting (Tx) and r receiving (Rx) antennas considered in this paper is defined by [13, 22, 4, 2]

$$\mathbf{y}_n = \sqrt{\gamma}\mathbf{H}\mathbf{x}_n + \mathbf{z}_n, \quad n = 1, \ldots, N \tag{1}$$

where $\mathbf{x}_n \in \mathcal{X}^t$ is the vector of modulation symbols transmitted in parallel at time n by the Tx 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 i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$, $\mathbf{y}_n \in \mathbb{C}^r$ is the corresponding vector of received signal samples at the output of the Rx antennas, $\mathbf{H} \in \mathbb{C}^{r \times t}$ is the channel matrix, and γ is the SNR per Tx antenna. We assume that the channel matrix is normalized such that $\frac{1}{rt}\mathrm{trace}E[\mathbf{H}\mathbf{H}^H] = 1$, so that the average received SNR per Rx 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 $\mathbf{H}$.

A space-time code (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 $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ are transmitted by columns, in N consecutive channel uses. The STC spectral efficiency is given by $\eta = \frac{1}{N}\log_2 |\mathcal{S}|$ bit/channel use. By definition, the information bit-energy over noise power spectral density ratio is given by $E_b/N_0 = t\gamma/\eta$.

2.1 Trellis Space-Time Codes

In [13], Tarokh *et al.* found criteria to design STC. They considered the pairwise-error probability (PEP) with maximum-likelihood (ML) decoding and, in the case of Rayleigh/Rician fading coefficients, they showed that the PEP $P(\mathbf{X} \to \mathbf{X}')$ averaged with respect to $\mathbf{H}$ is upper bounded as

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

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

For large N, STC can be constructed from multidimensional trellis codes (M-TCM) [23] with trellis termination. Examples of such schemes are given in [13, 16, 17, 18]. Namely, consider a M-TCM code $\mathcal{C}$ over $\mathcal{X}$ of rate $R = b/t$ bit/symbol, where each trellis step corresponds to b information (input) bits and to 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 simply formatting the codewords

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 Tx antennas. The resulting spectral efficiency is given by $\eta \approx b$ bit/channel use.[1]

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 M-TCM code. In [22], for the class of binary and quaternary trellis codes over $\mathbb{Z}_2$ and $\mathbb{Z}_4$ mapped onto BPSK and 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.

The maximum-likelihood (ML) decoder for trellis STC can be implemented by the Viterbi Algorithm (VA) applied on the trellis of the underlying M-TCM code. However, for fixed code rate R and rank-diversity ρ the decoder complexity grows exponentially with the number of Tx antennas t. In fact, from Lemma 3.3.2 in [13] we know that the trellis complexity is lower bounded by $2^{tR(\rho-1)}$.

Because of the difficulty in code design (with the exception of constructions given in [22]) and because of decoding complexity, trellis STC are practically restricted to small t.

2.2   Layered Space-Time Codes

In [14], Foschini proposed a STC scheme suited to handle a very large number of Tx and Rx antennas. In this scheme, M codes $\mathcal{C}^{(1)}, \dots, \mathcal{C}^{(M)}$ over $\mathcal{X}$ produce independently codewords $\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(M)}$ of length $N' = td$, where $d \geq 1$ and M are given integers. These codewords are diagonally interleaved (or "layered"),[2] as shown in Fig. 1, in order to form the $t \times N$ codeword matrix (with $N = d(M + t - 1)$)

$$\mathbf{X} = \mathcal{L}_d(\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(M)})$$

where $\mathcal{L}_d$ indicates the "layered" diagonal interleaver with interleaving delay d (see Fig. 1).

A reduced-complexity suboptimal decoder for this scheme is obtained by a linear front-end followed by decision-feedback interference cancellation [14]. The linear front-end, defined by a matrix $\mathbf{F} \in \mathbb{C}^{r \times t}$, 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 decoded by taking as observable the sequence of samples

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

---

[1]For simplicity, we ignore the rate decrease due to trellis termination since it becomes negligible for N much larger than the TCM encoder memory.

[2]In [21] another layered scheme based on standard rectangular (or "vertical") interleaving is considered, and in [24] a comparison between "horizontal" and diagonal layered schemes is presented.

where the one-to-one index mapping $(m, \ell) \leftrightarrow (j, n)$ is induced by the interleaver $\mathcal{L}_d$, $v_{j,n}$ is the j-th element of $\mathbf{v}_n$, $b_{j,k}$ is the (j, k)-th element of the matrix $\mathbf{B} = \mathbf{F}^H \mathbf{H}$ and $\widehat{x}_{k,n}$ is a decision on the (k, n)-th symbol of $\mathbf{X}$. From Fig. 1, we see that the elements $x_{k,n}$ for $k = j + 1, \ldots, t$ correspond to either zeros (for which no decision is needed) or to symbols of codewords $\mathbf{c}^{(m')}$ with index $m' < m$. Therefore, by decoding the codewords in the order $m = 1, \ldots, M$, the decisions needed in (2) are provided by earlier decoded codewords.

2.3 "Wrapped" Space-Time Codes

In the original layered construction of [14], the length of each component codeword is $N' = dt$. Because of practical hardware complexity limitations t cannot be a very large number. This implies that in order to have long codewords, the interleaving delay d must be large. If interleaving delay is an issue, the layered 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 non-negligible rate loss due to trellis termination.

For this reason, we propose a scheme which keeps the the simplicity of decision-feedback decoding while allows for arbitrarily long component codewords and small interleaving delay. Interestingly, trellis STC and layered STC are found to be special cases. In the proposed scheme, a single 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 $\mathbf{X} = \mathcal{D}_d(\mathbf{c})$, with $N = N''/t + (t - 1)d$. The diagonal interleaver $\mathcal{D}_d$ is defined by

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

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

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

The codeword matrix $\mathbf{X}$ is filled by wrapping the codeword $\mathbf{c}$ around its diagonals, as illustrated by Fig. 2. Hence, this STC scheme shall be referred to as *wrapped* space-time code (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 lower index take the lower positions in each column of the codeword matrix $\mathbf{X}$ (see Fig. 2), for $d = 0$ we assume that the codeword matrix $\mathbf{X}$ is filled with the elements of the codeword $\mathbf{c}$ as in Fig. 3. In this case, index $\ell_j(n)$ in (3) becomes

$$\ell_j(n) = (n - 1)t + t - j + 1 . \tag{5}$$

Remark 1 If $\mathcal{C}$ is a M-TCM code of rate $R = b/t$, then $\mathcal{S} = \mathcal{D}_0(\mathcal{C})$ is a trellis STC. If $\mathcal{C} = \mathcal{C}^{(1)} \times \cdots \times \mathcal{C}^{(M)}$ (i.e., $\mathcal{C}$ is given by the Cartesian product of codes $\mathcal{C}^{(1)}, \ldots, \mathcal{C}^{(M)}$), then $\mathcal{S} = \mathcal{D}_d(\mathcal{C}) = \mathcal{L}_d(\mathcal{C}^{(1)}, \ldots, \mathcal{C}^{(M)})$ is a layered STC. Hence, WSTC is a generalization of both trellis and layered STC schemes. $\diamondsuit$

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. ◇

3 Decoding of Wrapped Space-Time Codes

In the case where $\mathcal{C}$ is a trellis code, we propose a reduced-complexity suboptimal decoder obtained by combining decision-feedback interference cancellation with the VA, in analogy with the delayed decision-feedback detection (DDFD) approach for ISI channels with coding (see [25, 26, 27, 28] and references therein). The decoder takes as observable the sequence of samples

$$r_\ell = v_{j,n} - \sqrt{\gamma} \sum_{k=j+1}^{t} b_{j,k} \widehat{x}_{k,n} \ , \quad \ell = 1, \ldots, N'' \tag{6}$$

where $v_{j,n}$ and $b_{j,k}$ are defined as in (2) and where $1 \le j \le t$ and $1 \le n \le N$ 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 $k = j+1, \ldots, t$ correspond to either zeros (for which no decision is needed) or to symbols of $\mathbf{c}$ with index $\ell' \le \ell - td + 1$ ($\ell' \le \ell - 1$ for $d = 0$). These decisions are found in the survivor history according to *per-survivor processing* (PSP) [20]. As an example, consider the case $d > 0$. The delay for decision $\widehat{x}_{k,n}$ 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 VA decoding delay (typically 5 or 6 times the code constraint length [29]), the corresponding decisions are reliably obtained from the VA output. If $(k-j)(td-1)$ is not large enough, the corresponding decisions can be obtained from the survivors ending in each state of the VA. In this way, if the correct path is among the survivors, at least a subset of the paths are extended by the VA with the correct feedback decisions.

Choice of the linear front-end filter. For the sake of simplicity, the decoder treats the sequence r_ℓ as if it was produced by the *virtual* scalar-input additive-noise channel

$$r_\ell = \sqrt{\beta_j} c_\ell + \nu_\ell, \quad \ell = 1, \ldots, N'' \tag{7}$$

where ν_ℓ 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 β_j of the above channel is given by

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

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 $\beta_1, \ldots, \beta_t$, 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 exactly. Moreover, the noise samples ν_ℓ are not Gaussian and not independent, in general. However, provided that these assumptions hold, this scheme decomposes the MIMO 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 ISI channels, standard choices for the front-end filter matrix $\mathbf{F}$ are also inspired by equalization [30]. If $\mathbf{H}$ has rank t, an information-lossless front-end is given by the WMF $\mathbf{F} = \mathbf{Q}$, where $\mathbf{H} = \mathbf{QB}$ is the "QR" decomposition [31] of the channel matrix $\mathbf{H}$, where $\mathbf{Q} \in \mathbb{C}^{r \times t}$ has orthonormal columns and $\mathbf{B} \in \mathbb{C}^{t \times t}$ is upper triangular. In this case, 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). Another information-lossless front-end is the unbiased MMSE filter whose j-th column $\mathbf{f}_j$ is the solution of the SNR maximization problem

$$\begin{cases} \text{maximize} & \beta_j \\ \text{subject to} & |\mathbf{f}_j|^2 + \gamma \sum_{k=1}^{j-1} |b_{j,k}|^2 = 1 \end{cases}$$

and it is given explicitly by

$$\mathbf{f}_j = \frac{\left[\mathbf{I} + \gamma \sum_{k=1}^{j-1} \mathbf{h}_k \mathbf{h}_k^H\right]^{-1} \mathbf{h}_j}{\sqrt{\mathbf{h}_j^H \left[\mathbf{I} + \gamma \sum_{k=1}^{j-1} \mathbf{h}_k \mathbf{h}_k^H\right]^{-1} \mathbf{h}_j}} \tag{9}$$

where $\mathbf{h}_j$ denotes the j-th column of $\mathbf{H}$. In this case, the j-th channel SNR is given by

$$\beta_j = \gamma \mathbf{h}_j^H \left[\mathbf{I} + \gamma \sum_{k=1}^{j-1} \mathbf{h}_k \mathbf{h}_k^H\right]^{-1} \mathbf{h}_j \tag{10}$$

and the additive noise is neither Gaussian nor i.i.d. (even assuming perfect feedback decisions).

If $\mathbf{H}$ has rank less than t, the WMF is not defined and the MMSE filter is information-lossy. Subject to mild conditions on the statistics of $\mathbf{H}$, the probability that $\mathbf{H}$ has rank less than t when $r \geq t$ is zero. Therefore, the above schemes can be practically applied whenever $r \geq t$. In the rest of this paper we restrict our treatment to this case, and we briefly address the case $r < t$ in Section 7.

4 Code Design for the 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 $\beta_1, \ldots, \beta_t$ is given by

$$P(\mathbf{c} \to \mathbf{c}'|\beta_1, \ldots, \beta_t) = Q\left(\sqrt{2\sum_{j=1}^{t}\beta_j w_j}\right) \tag{11}$$

where $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$ is the Gaussian tail function and where we define the *squared Euclidean weight* (SEW) w_j as

$$w_j = \frac{1}{4}\sum_{n=1}^{N}|x'_{j,n} - x_{j,n}|^2 \tag{12}$$

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

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

$$\delta = \min_{\mathbf{c}, \mathbf{c}' \in \mathcal{C}: \mathbf{c}' \neq \mathbf{c}} |\{j \in \{1, \ldots, t\} \ : \ w_j \neq 0\}| \tag{13}$$

that is, to maximize the minimum number of non-zero rows in the matrix difference $\mathbf{D} = \mathbf{X}' - \mathbf{X}$ for each pair of distinct codewords matrices $\mathbf{X}, \mathbf{X}' \in \mathcal{D}_d(\mathcal{C})$. The block-diversity criterion has been investigated in [32, 33, 34] 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 bit/symbol and block-diversity δ. Then, the rank-diversity ρ of the corresponding WSTC $\mathcal{S} = \mathcal{D}_d(\mathcal{C})$ satisfies:*

$$\rho \leq \delta \leq 1 + \left\lfloor t\left(1 - \frac{R}{\log_2|\mathcal{X}|}\right)\right\rfloor . \tag{14}$$

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

Proof. Consider a pair of distinct codewords $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$, the corresponding matrices $\mathbf{X} = \mathcal{D}_d(\mathbf{c})$ and $\mathbf{X}' = \mathcal{D}_d(\mathbf{c}')$ and the difference matrix $\mathbf{D} = \mathbf{X}' - \mathbf{X}$. The rank of $\mathbf{D}$ cannot be larger than the number of non-zero 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 δ is then the minimum Hamming distance of this block code. By applying the Singleton bound [35], we obtain

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

which implies the second inequality in (14).[3]

In order to prove the second statement of the proposition, consider a distinct pair of 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 $\widetilde{\mathbf{D}}$. If any such $\widetilde{\mathbf{D}}$ has rank $\geq \delta$, 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 $\widetilde{\mathbf{D}}$ a tail of $(t-1)d$ zeros and by shifting to the right each j-th row by $(j-1)d$ positions. Let $\Omega \subseteq \{1, \ldots, t\}$ denote the set of the indexes of non-zero rows of $\widetilde{\mathbf{D}}$ and for each row $j \in \Omega$ let ℓ_j and r_j denote the number of leading and tailing zeros. Now, let

$$d = \max\left\{1, \min\{d_1, d_2\}\right\} \tag{15}$$

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 (15) over all distinct $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$ we obtain that $\mathcal{S} = \mathcal{D}_d(\mathcal{C})$ has rank-diversity δ. □

Remark 3 From Theorem 3.3.1 of [13], we know that for any STC over $\mathcal{X}$ with t Tx antennas and spectral efficiency $\eta = tR$ the rank-diversity satisfies

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

Since this is the same upper bound on block-diversity given in Proposition 1, we get 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 (14) on δ is met with equality, for several coding rates and values of t. Examples of these codes are tabulated in [32, 34]. Therefore, the wrapping construction is a powerful tool to construct STC with maximum rank-diversity. ◊

Remark 4 From Lemma 3.3.1 of [13] we know that a trellis STC with rank-diversity ρ must have constraint length $L \geq \rho$.[4] This constraint does not apply to WSTC. For example, the binary 4-state convolutional code (CC) of rate 1/4 with generators $(5, 7, 7, 7)$ (octal notation [30]) has constraint length 3, but the corresponding WSTC for $t = 4$ antennas with interleaver delay $d \geq 2$ achieves $\rho = 4$. This fact is explained by noticing

---

[3]This upper bound on the block-diversity has been found in [36] and independently in [33].

[4]Following [37], we define the constraint length of a trellis code as $L = \nu_{\max} + 1$, where $\nu_{\max}$ is the maximal length of the shift-registers in the canonical feedforward encoder.

Table 1: Block-diversity for rate 1/2 binary codes tabulated in [30] mapped onto BPSK/QPSK. Bold numbers denote diversity achieving the bound (14). Code generators are expressed in octal notation [30].

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

that the diagonal interleaver *expands* the state space of the overall interleaved code. On the other hand, the PSP-based VA decoder proposed for WTSCs works on the trellis of the underlying code $\mathcal{C}$, i.e., it ignores the state space expansion.[5] 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 PSP-based VA decoding is used instead of optimal ML decoding. In next sections we shall see that the penalty incurred by the MMSE front-end and sufficiently large interleaving delay d is practically negligible, while the penalty incurred by the WMF front-end and any d can be very large. ◇

Remark 5 In [38], a computational efficient trellis-based algorithm for computing the block-diversity of trellis codes with cyclic interleaving is given. In Section 5, 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 1/4 and 1/2 tabulated in [30] and mapped onto BPSK and QPSK, for different values of t. Some of these results are reported in Tables 1 and 2. We observe that several codes achieve maximum block-diversity and therefore are good candidates for WSTC. ◇

Example 1 Fig. 5 shows the word-error rate (WER) vs. E_b/N_0 for WTSCs obtained from the binary CC with generators $(5, 7, 7, 7)$ mapped onto BPSK and transmitted over a $t = r = 4$ channel, with $d = 0, 1, 2$. The channel matrix has i.i.d. elements $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$ (independent Rayleigh fading). Each transmitted codeword corresponds to 128 information bits. Decoding is performed by the PSP-based VA working on the underlying 4-state trellis, with MMSE front-end. As d increases, the slope of the WER curve becomes steeper and steeper. In fact, it can be checked that for $d = 0$ the resulting WSTC has $\rho = 2$, for $d = 1$ has $\rho = 3$ and for $d = 2$ has $\rho = 4$. The solid curve denoted as "QUB" is

---

[5]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 2: Block-diversity for rate 1/4 binary codes tabulated in [30] mapped onto BPSK/QPSK. Bold numbers denote diversity achieving the bound (14). Code generators are expressed in octal notation [30].

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

an analytical *Quasi-Upper Bound* developed in Section 5 assuming the parallel channel model induced by perfect decision-feedback.

For the sake of comparison, we show the WER performance of the trellis STC or, equivalently, WSTC with $d = 0$, obtained by formatting the 16-state binary CC with generators $(25, 27, 33, 37)$ mapped onto BPSK. This code is shown to achieve rank-diversity $\rho = 4$ in [22]. As expected, the slope of the error curve for this code is the same of the 4-state WSTC with $d = 2$. We also show the QUB for the 16-state trellis STC assuming the parallel channel model induced by perfect decision-feedback with MMSE front-end. The fact that the simulated ML curve gets asimptotically very close to the QUB makes us conjecture that the PSP-based VA with MMSE front-end pays almost no penalty with respect to optimal ML decoding. An information-theoretic explanation of this fact will be provided in Section 6. $\diamond$

5 Word-Error Rate 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 $\beta_1, \ldots, \beta_t$. Given the parallel channel decomposition (7), this method provides also a *true* upper bound for the WSTC $\mathcal{S} = \mathcal{D}_d(\mathcal{C})$ with WMF front-end, subject to the assumption of perfect feedback decisions. For the MMSE front-end, 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 provides always an upper bound also for MMSE. For this reason, in the following it will be referred to as *Quasi-Upper Bound* (QUB).

For simplicity, we assume that $\mathcal{C}$ is geometrically uniform [39], so that we can take any codeword $\mathbf{c} \in \mathcal{C}$ as the reference codeword. Consider the conditional PEP $P(\mathbf{c} \rightarrow$

$\mathbf{c}'|\beta_1,\ldots,\beta_t)$ given in (11). By using the union bound we get

$$P_w(e|\beta_1,\ldots,\beta_t) \leq \sum_{\mathbf{c}' \neq \mathbf{c}} P(\mathbf{c} \to \mathbf{c}'|\beta_1,\ldots,\beta_t) \qquad (17)$$

We are mainly interested in the case where $\mathcal{C}$ is a trellis code with trellis termination. A pairwise error event $\{\mathbf{c} \to \mathbf{c}'\}$ 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, is said to be composite. We have the following:

Lemma 1 *For any arbitrary $\beta_1,\ldots,\beta_t$, the 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 $\{\mathbf{c} \to \mathbf{c}'\}$ 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

$$\boldsymbol{\beta} = \operatorname{diag}(\sqrt{\beta_1},\ldots,\sqrt{\beta_t},\sqrt{\beta_1},\ldots,\sqrt{\beta_t},\ldots,\ldots,\sqrt{\beta_1},\ldots,\sqrt{\beta_t})$$

be a $N'' \times N''$ diagonal matrix containing the channel amplitudes $\sqrt{\beta_1},\ldots,\sqrt{\beta_t}$ repeated periodically N''/t times (by definition, N'' is an integer multiple of t). Assume that $\{\mathbf{c} \to \mathbf{c}'\}$ 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 be always be formed by the concatenation of two paths, which differ from the reference path on disjoint supports. Therefore, the non-zero 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 $\{\mathbf{c} \to \mathbf{c}'\}, \{\mathbf{c} \to \mathbf{c}_1\}$ and $\{\mathbf{c} \to \mathbf{c}_2\}$ are given by the inequalities

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

The error event $\{\mathbf{c} \to \mathbf{c}'\}$ 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,

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

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

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

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

Next, we find a method to enumerate all simple error events. We can represent the code $\mathcal{C}$ over a super-trellis, obtained by lumping together s trellis steps of the original trellis, so that each step of the super-trellis corresponds to s consecutive steps of the original trellis, and each step of the super trellis has t output symbols. For example, if $\mathcal{C}$ is a trellis code of rate b/p bit/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 super-trellis is N''/t steps.

Let $A^{(L)}_{w_1,\ldots,w_t}$ denote the number of simple error events of length L in the super-trellis (the length L is measured in super-trellis steps) with SEWs $w_1,\ldots,w_t$ starting at step 1. Then, the total number of simple error events of length L and SEWs $w_1,\ldots,w_t$ is equal to $(N''/t - L + 1)A^{(L)}_{w_1,\ldots,w_t}$. By restricting the sum in (17) to simple error events and by using (11) we have

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

where we define the code multivariate weight enumerator function [40, 41]

$$
T(W_1,\ldots,W_t) = \sum_{w_1,\ldots,w_t}\sum_{L=1}^{\infty}A^{(L)}_{w_1,\ldots,w_t}\prod_{j=1}^{t}W_j^{w_j}
\tag{20}
$$

and where in (a) of (19) we used the *preferred* integral expression of the Gaussian tail function [42]

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

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 length $L = 1, 2, \ldots$ (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 $T(W_1,\ldots,W_t)$ is given in Appendix A.

In order to obtain the average WER, where expectation is done with respect to the joint statistics of the channel SNRs $\beta_1,\ldots,\beta_t$ (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 belongs 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 [33]). Then, since the conditional WER cannot be larger than 1,

we follow the approach of [33] and obtain the final bound on the average WER as

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

where the expectation is with respect to the joint statistics of $\beta_1, \ldots, \beta_t$. For small t and if the joint pdf of $\beta_1, \ldots, \beta_t$ is known, the above expectation can be calculated by numerical integration methods (as in [33]). If t is large or if the joint statistics of $\beta_1, \ldots, \beta_t$ 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 wrapped scheme $\mathcal{S} = \mathcal{D}_d(\mathcal{C})$, 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 front-end) or by using (10) (in the case of MMSE front-end). In this way, the method can be applied to any arbitrary channel statistics (for example, $\mathbf{H}$ can be generated by a ray-tracing software in order to model a particular scattering environment [3]).

Remark 6 The integral with respect to ϕ can be efficiently computed by using Gauss-Chebychev quadrature rules. In fact, by letting $g(\sin^2\phi) = T\left(e^{-\beta_1/\sin^2\phi}, \ldots, e^{-\beta_t/\sin^2\phi}\right)$ we can write

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

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

Remark 7 In [25, 26, 43], an analysis of the bit-error probability of DDFD 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 VA 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 $\widehat{x}_{k,n}$ in (6) have delay $(k-j)(td-1)$ for $k = j+1, \ldots, t$. If the survivors merge always at delay smaller than $td-1$, only errors due to past error events affect the metric updating and the WEP of the PSP-based VA is equivalent to a genie-aided decoder with perfect feedback decisions. In fact, if $\widehat{x}_{k,n} \neq x_{k,n}$ a codeword error already occurred for both the PSP-based and for the genie-aided decoders, while if $\widehat{x}_{k,n} = x_{k,n}$ the two decoders have the same branch metric for the (j,n)-th symbol. In other words, the error propagation due to non-perfect feedback decisions has no influence on the WER, provided that $td-1$ is large enough in order to have a very high probability of path merge at delay $< td-1$. Notice that for large t, a large interleaving delay d is not needed to meet this condition.

On the contrary, in the case of DDFD for ISI 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 VA decoding can be predicted very well by the QUB (21) while neglecting feedback decision errors in DDFD yields overly optimistic results [25, 26, 43]. $\diamondsuit$

A method for evaluating 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 δ if and only if $T(W_1, \ldots, W_t) = 0$ for all vectors $(W_1, \ldots, W_t) \in \{0,1\}^t$ with Hamming weight less than δ and there exists one vector $(W_1, \ldots, W_t) \in \{0,1\}^t$ with Hamming weight δ for which $T(W_1, \ldots, W_t) > 0$.*

Proof. Consider the parallel channel model with cyclic interleaving (7) and Fig. 4. Let some of the SNRs β_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 $P_w(e|\beta_1, \ldots, \beta_t) = 0$. Otherwise, $P_w(e|\beta_1, \ldots, \beta_t) > 0$. Following the same steps yielding (19), we obtain a Chernoff union bound [29] on the WEP as

$$P_w(e|\beta_1, \ldots, \beta_t) \leq \frac{N''}{2t} T\left(e^{-\beta_1}, \ldots, e^{-\beta_t}\right) \tag{23}$$

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

6   Mutual Information and Outage Probability

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

defined by

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

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

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

($E_{\mathbf{x},\mathbf{z}}\{\cdot\}$ 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 and discrete 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 to $I_{\mathcal{X}}(\mathbf{H})$ 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 [44]. A simple upper bound on $I_{\mathcal{X}}(\mathbf{H})$ is obtained by assuming Gaussian i.i.d. inputs (GI) with the same average input constraint. We have

$$I_{\mathcal{X}}(\mathbf{H}) \le I_{\mathcal{G}}(\mathbf{H}) = \log_2 \det \left(\mathbf{I} + \gamma \mathbf{H}\mathbf{H}^H\right) \tag{26}$$

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

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

where $\nu \sim \mathcal{N}(0, 1)$. The *matched-filter bound* (MFB) [44] is obtained by neglecting the mutual interference of the t transmit antennas. We have

$$I_{\mathcal{X}}(\mathbf{H}) \le I_{\mathcal{X}}^{\text{mfb}}(\mathbf{H}) = I_{\text{P.ch.}}\left(\gamma|\mathbf{h}_1|^2, \ldots, \gamma|\mathbf{h}_t|^2\right) \tag{28}$$

A class of lower bounds can be obtained from the chain-rule of mutual information as

follows [44]. For given $\mathbf{H}$ and any $r \times t$ matrix $\mathbf{F}$ for which $\mathbf{v} = \mathbf{F}^H \mathbf{y}$ we can write

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

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 or by the MMSE filter defined in Section 2.3 equality holds (recall that we assume $r \geq t$). The inequality (b) holds with equality if $\mathbf{F}$ is the MMSE filter and the inputs $x_1, \ldots, x_t$ are Gaussian i.i.d. [45, 46].[6] 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

$$
I_{\mathcal{X}}(\mathbf{H}) \geq I_{\mathcal{X}}^{\mathrm{wmf}}(\mathbf{H}) = I_{\mathrm{P.ch.}}\left(\gamma|b_{1,1}|^2, \ldots, \gamma|b_{t,t}|^2 \right)
\tag{30}
$$

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 [47] and in an asymptotic form in [48] 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 [44], 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 to the true mutual information (in [44] this is referred to as *conjectured lower bound*). Motivated by these arguments, we can write $I_{\mathcal{X}}(\mathbf{H}) \overset{\sim}{\geq} I_{\mathcal{X}}^{\mathrm{mmse}}(\mathbf{H}) = I_{\mathrm{P.ch.}}(\beta_1, \ldots, \beta_t)$ where β_j is given in (10) and $\overset{\sim}{\geq}$ means that inequality is conjectured.

Upper and lower bounds on $I_{\mathcal{X}}(\mathbf{H})$ yield lower and upper bounds on $P_{\mathrm{out}}(\eta)$, respectively. Figs. 7, 8 and 9 show a comparison of the various outage probability bounds versus E_b/N_0 for $t = r = 4$, BPSK modulation and independent Rayleigh fading. The curves are

---

[6]This can be proved in a purely linear-algebraic way by showing the determinant decomposition

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

where β_j is given by (10).

obtained by Monte Carlo simulation over several independent generations of the channel matrix **H**. Some points for the true outage probability are also shown for comparison.

For small spectral efficiency (Fig. 7 and 8) the MMSE-GA (quasi-)upper bound is very close to the true outage probability, while as η increases (Fig. 9) it becomes looser. Remarkably, in the cases considered here the MFB provides a lower bound tighter than the GI. The gap between GI and BPSK increases with η and, for all values of η, the WMF suffers from a large degradation with respect to MMSE-GA.

At this point, some qualitative considerations are in order:

1. In the derivation of (30) and of the MMSE-GA (quasi-)upper bound the perfect feedback decisions are not *assumed*, but they are a consequence of the chain-rule of mutual information (29). Hence, these bounds yield also the performance limit achievable by the WMF or MMSE front-end with *perfect* decision-feedback. In particular, since the PSP-based VA decoding is very robust to feedback decision errors, we expect that these outage probability bounds predict well the performance limits of WSTCs with PSP-based VA decoding.

2. In the range of η where the MMSE-GA is close to the true outage probability (e.g., in Fig. 7 and 8), we expect that *good* WSTCs designed with respect to the block-diversity criterion and decoded by PSP-based VA decoding and MMSE front-end perform close to good STC designed with respect to the rank-diversity criterion and decoded by ML decoding. On the contrary, we expect that PSP-based VA decoding suffers from a performance penalty with respect to ML decoding over the range of η where the MMSE-GA is far from the true outage probability (e.g., in Fig. 9). This explain why in Example 1 the QUB assuming MMSE front-end and perfect feedback decisions is very close to the simulated ML performance for trellis STCs of rate 1/4.

3. 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 **H** of rank t, the two front-ends (assuming perfect feedback decisions) are capacitywise equivalent as $\gamma \to \infty$ (see [44] and references therein). This apparent paradox can be explained by noticing that the convergence $I_{\mathcal{X}}^{\text{wmf}}(\mathbf{H}) \to I_{\mathcal{X}}^{\text{mmse}}(\mathbf{H})$ as $\gamma \to \infty$ is non-uniform over the ensemble of random **H**. Therefore, $P(I_{\mathcal{X}}^{\text{wmf}}(\mathbf{H}) \leq \eta)$ does not converge to $P(I_{\mathcal{X}}^{\text{mmse}}(\mathbf{H}) \leq \eta)$ even if $I_{\mathcal{X}}^{\text{wmf}}(\mathbf{H}) \to I_{\mathcal{X}}^{\text{mmse}}(\mathbf{H})$ for any given **H**.

   For the above reason, we expect that also for practical codes the gap between WMF and MMSE does not vanish as $\gamma \to \infty$. Curiously, several previous works on layered STCs [14, 21] reported results for the WMF front-end, referred to as the "nulling and canceling" approach. Our results show that it is actually very dangerous to use WMF rather than MMSE with decision-feedback interference cancellation in a quasi-static environment.

Example 2 In this example we show that the MMSE and the WMF front-ends have in fact two opposite behaviors with respect to the interleaving delay d: while the performance with MMSE front-end improves as the interleaving delay d increases (making feedback decisions more and more reliable) the performance with WMF front-end improves by making d small (non-perfect interference cancellation), provided that the resulting WSTC preserves its rank-diversity as d decreases. Fig. 10 and 11 show WER vs. E_b/N_0 for the WSTC obtained from the CC with generators $(23, 35)$ mapped onto BPSK, in the same conditions of Example 1, for $d = 0, 1, 2, 3$ with PSP-based VA decoding with MMSE and WMF front-ends, respectively. In Appendix B we show that this code has rank-diversity equal to 3 for any $d \geq 0$, i.e., always equal to the maximum possible rank-diversity for $t = 4$ Tx antennas and binary coding of rate $R = 1/2$. For $d = 0$ the WSTC reduces to a trellis STC, where the underlying trellis code is given by the equivalent rate 2/4 CC obtained by lumping two trellis steps of the original rate 1/2 CC. Therefore, exact ML decoding is possible. With MMSE front-end (Fig. 10), the PSP-based decoder approaches exact ML as d increases. On the contrary, with WMF front-end (Fig. 11), the PSP-based decoder suffers from an increasing performance gap with respect to exact ML.

This effect can be explained by noticing that the WMF front-end is capacitywise suboptimal in the presence of perfect feedback decisions. On the contrary, the MMSE front-end is optimal with perfect feedback decisions.[7]

Finally, this effect is clearly visible only for WSTCs for which $\rho = \delta$ for all $d \geq 0$. If ρ increases with d, the performance with WMF front-end is the result of two contrasting effects, and its behavior is difficult to predict. $\Diamond$

7   Performance Examples

In this Section, the WER performance of various WSTCs are assessed for quasi-static fading channels with independent Rayleigh fading. The union bound on the WER described in Section 5 is employed (curves denoted by "QUB") and the results are compared with computer simulation of the full PSP-based VA decoder (curves denoted by "SIM"). In our simulations, each codeword corresponds to 128 information symbols. As a consequence, the codeword length is variable, depending on the code rate.

Results for known convolutional codes.   Figs. 12, 13 and 14 show the QUB on the WER for some WSTC based on binary CCs of rate 1/2 and 1/4 mapped onto BPSK and QPSK, with $t = r = 4$ and $t = r = 8$. Some points obtained by simulating the full PSP-based VA decoder (with interleaving delay $d = 2$) is shown for the sake of comparison. Also, we included the outage probability with GI and the MMSE-GA conjectured upper bound with discrete-inputs. The difference in the slope of the WER curves for the various codes reflects the different block-diversities in Tables 1 and 2. Remarkably, there is still a

---

[7]Strictly speaking, this is true only for Gaussian inputs, but as argued before for discrete inputs and coding rate in an appropriate range Gaussian and discrete inputs behave similarly.

consistent gap (ranging from 1.5 to 4 dB at WER $= 10^{-3}$) between these simple off-the-shelf codes and the outage curve. This calls for the design of good component codes for the WSTC scheme.

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

Fig. 15 compares the WER vs. E_b/N_0 of a WSTC with either MMSE and WMF front-end obtained from the binary CC with generators $(23, 35)$ mapped onto QPSK, with $t = r = 4$ antennas with the scheme obtained by concatenating the same trellis code with the V-BLAST detector. Simulations of the PSP-based VA decoder for the WSTC (obtained for $d = 2$) are in perfect agreement with the corresponding QUB. The V-BLAST performance was obtained by simulation. For comparison, we show also the WER resulting from a genie-aided V-BLAST detector with ideal feedback decisions (curves denoted by "Id. F."), which is very similar to the WER achieved by the WSTC without 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, and that the PSP-based decoder is very effective in preventing such propagation while 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 VA due to PSP, for large t. Then, WSTC with PSP-based VA decoding is not only more effective, but might be also simpler than V-BLAST.

Fig. 16 shows analogous results for WSTC and V-BLAST schemes based on the binary CC with generators (23,35) mapped onto BPSK, with $t = r = 8$ antennas. We see that also for V-BLAST the advantage of the MMSE versus the WMF front-end can be very large.

Handling the case $r < t$ via LD precoding.   When $r < t$ the low-complexity PSP-based decoding scheme cannot be applied. Recently, Hochwald and Hassibi proposed a scheme called "linear dispersion" (LD) coding [19]. This scheme takes blocks of Q

modulation symbols and map 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 $t \times T$ matrices defining the LD code. Then, $\mathbf{S}$ is transmitted column-by-column over T channel uses. The resulting spectral efficiency is $\eta = \frac{Q}{T} R$, where R is the rate of the outer code. We can think of LD coding as a precoder which shapes the original $t \times r$ complex MIMO channel into a virtual $2Q \times 2rT$ real MIMO channel. As long as Q and T are chosen such that $Q \leq rT$, WSTC with PSP-based decoding can be applied as an outer coding/decoding scheme to the LD-precoded channel.

As an example, Fig. 17 shows the WER vs. E_b/N_0 of a coding scheme obtained by concatenating the WSTC obtained from the binary CC with generators (23,35) mapped onto QPSK with a LD precoder designed in [19] for the $t = 4$, $r = 1$ channel, with $Q = 4$ and $T = 4$.[8] The resulting spectral efficiency is $\eta = 1$ bit/channel use. For the sake of comparison, we show also the outage curve with GI on the original 4×1 channel (curve denoted by "No prec."), the outage curve with GI of the LD-precoded channel, and the MFB and MMSE-GA outage curve bounds with QPSK inputs and LD-precoded channel. We notice that the gap between the GI outage of the original 4×1 channel and the WER of the concatenated WSTC with LD-precoding is almost entirely due to the LD precoder, since the simulated WER with PSP-based VA decoding is less than 1.5 dB away from the outage of the LD-precoded channel, while the gap between LD-precoded and original channels is about 5 dB for WER $\approx 10^{-4}$.

The LD precoders in [19], included the one used in this example, where 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. We believe that a better WER performance for the LD-precoded channel can be achieved by explicitly designing the precoder in order to minimize the outage probability.

8   Conclusions

A new scheme, nicknamed "wrapped" STC was proposed. This scheme generalizes both trellis and layered STCs and it is suited to a large number of antennas and low complexity decoding, based on MMSE decision-feedback coupled with PSP-based Viterbi Decoding. We showed that any trellis codes with maximal block-diversity can be turned into a WSTC with maximal rank-diversity, with the advantage that block-diversity is easy to check and maximal block-diversity is easily achieved by several well-known trellis codes. We also showed via numerical experiments and information-theoretical arguments that the MMSE decision-feedback receiver coupled with the PSP-based VA, providing very reliable decisions, performs very close to optimal ML decoding. On the contrary, a similar scheme with WMF instead of MMSE front-end is very suboptimal in the quasi-stationary regime.

---

[8]This precoder was kindly provided to us by the authors of [19].

We provided a simple and rather accurate (quasi-)upper bound on the performance of WSTCs and an efficient method for calculating the block-diversity of trellis codes, based on a multivariate weight enumerator. Finally, we provided some upper and lower bounds to the information outage probability with discrete i.i.d. inputs.

Performance examples of the proposed scheme constructed from well-known binary linear convolutional codes were provided. WSTCs compare very favorably with respect to coded V-BLAST of similar complexity. In particular, we showed that V-BLAST is prone to error propagation in the feedback 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.

In the case $r < t$, where the decision-feedback scheme cannot be used directly, our scheme is naturally suited to work as outer code where the inner code (pre-coder) is a linear dispersion code as studied in [19].

The gap between the WER of WSTCs obtained from simple off-the-shelf codes and the information outage probability indicates that some work needs still to be done in designing good component codes for the WSTC scheme. This is especially true for high-rate trellis codes, where the design for maximal block-diversity has been rarely addressed (see [32]).

Appendices

A Calculating the Multivariate Weight Enumerator Function

For the sake of completeness, we give here an example of calculation of $T(W_1, \ldots, W_t)$. Further details and examples can be found in [40, 41]. Consider the rate $1/2$ binary linear CC with 4 states and generators $(5, 7)$ and let $t = 4$. Fig. 18 shows the trellis of the original code and the super-trellis obtained by lumping together $s = 2$ steps of the original trellis, so that $t = 4$ output symbols correspond to each transition of the super-trellis (notice that the super-trellis 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 $\mu : \{0, 1\} \to \{\pm 1\}$). Any transition of the super-trellis 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 non-connected states are labeled by the zero monomial). Then, we define the state variables $V_1, \ldots, V_3$ corresponding to states $1, \ldots, 3$, 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

$$
\begin{bmatrix} Y \\ V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} 1 & W_2 W_3 W_4 & W_1 W_2 & W_1 W_3 W_4 \\ W_3 W_4 & W_2 & W_1 W_2 W_3 W_4 & W_1 \\ W_1 W_2 W_4 & W_1 W_3 & W_4 & W_2 W_3 \\ W_1 W_2 W_3 & W_1 W_4 & W_3 & W_2 W_4 \end{bmatrix} \begin{bmatrix} X \\ V_1 \\ V_2 \\ V_3 \end{bmatrix} \tag{A1}
$$

and it is obtained by letting each state/output variable equal to 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).

More in general, by following the above recipe we can always put the state equations in the 2×2 block form

$$\begin{bmatrix} Y \\ V_1 \\ \vdots \\ V_{S-1} \end{bmatrix} = \begin{bmatrix} D & \mathbf{C} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \begin{bmatrix} X \\ V_1 \\ \vdots \\ V_{S-1} \end{bmatrix} \tag{A2}$$

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

$$T(W_1, \ldots, W_t) = \mathbf{C}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + D - 1 \tag{A3}$$

In order to compute $T(W_1, \ldots, W_t)$ at $W_j = e^{-\beta_j / \sin^2 \phi}$, for $j = 1, \ldots, t$, 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.

B   Rank-diversity of the WSTCs build from code (25,35) for $t = 4$

In this appendix we show that the WSTCs build from the binary CC with generators $(23, 35)$ mapped onto BPSK for transmission over $t = 4$ antennas have all rank-diversity $\rho = \delta = 3$, which is the maximum possible (see (14)). Interestingly, this proof shows through an example that checking the rank-diversity of a trellis STC is a quite involved task even in a simple binary case.

Proof. We start by proving the statement for $d = 0$. By lumping together two trellis steps of the original trellis we obtain an equivalent rate $2/4$ 16-states code whose generator matrix is given by

$$\mathbf{G}(D) = \begin{bmatrix} 1 + D^2 & 1 + D + D^2 & D & 1 \\ D^2 & D & 1 + D^2 & 1 + D + D^2 \end{bmatrix}$$

where D is the delay operator. Let $b_1(D) = \sum_{i=0}^{\infty} b_{1i} D^i$ and $b_2(D) = \sum_{i=0}^{\infty} b_{2i} D^i$ denote the two information-bit sequences entering the equivalent encoder. The corresponding

binary codeword, formatted as a $4 \times N$ binary matrix, is given by

$$\mathbf{c}(D) = \begin{bmatrix} (1 + D^2)b_1(D) + D^2b_2(D) \\ (1 + D + D^2)b_1(D) + Db_2(D) \\ Db_1(D) + (1 + D^2)b_2(D) \\ b_1(D) + (1 + D + D^2)b_2(D) \end{bmatrix}. \tag{B1}$$

The resulting WSTC $\mathcal{S}$ is obtained by applying componentwise the BPSK modulation mapping $\mu : \{0, 1\} \rightarrow \{+1, -1\}$ to all such $\mathbf{c}(D)$. In order to show that $\mathcal{S}$ has rank-diversity $\rho = 3$, we have to show that for all $\mathbf{c}(D) \neq \mathbf{0}$ the matrix $\mu(\mathbf{c}(D)) - \mathbf{1}(D)$ (seen as a $4 \times N$ matrix of coefficients) has rank ≥ 3, i.e., that there exist a set of 3 linearly independent rows (here, $\mathbf{1}(D)$ denotes the matrix of all ones). This is equivalent to check that for all $\mathbf{c}(D) \neq \mathbf{0}$ the matrix $\mu(\mathbf{c}(D)) - \mathbf{1}(D)$ contains a $3 \times N$ submatrix of rank 3.

From [22, Th. 5] we know that a matrix $\mathbf{A} \in \mathbb{F}_2^{L \times N}$ has rank L over $\mathbb{F}_2$ if and only if all matrices $\widetilde{\mathbf{A}} \in \mathbb{R}^{L \times N}$ obtained from $\mathbf{A}$ by multiplying its elements by ± 1 in any arbitrary way have rank L over $\mathbb{R}$. As a corollary, $\mu(\mathbf{c}(D)) - \mathbf{1}(D)$ contains a $3 \times N$ submatrix of rank 3 if $\mathbf{c}(D)$ contains a $3 \times N$ binary submatrix of rank 3 over $\mathbb{F}_2$. From [22, Th. 14] we know that a binary matrix in the form $\mathbf{A}(D) = [a_1(D), \ldots, a_L(D)]^T q(D)$ for $q(D) \neq 0$ has rank L if the matrix $[a_1(D), \ldots, a_L(D)]^T$ seen as a $L \times (\nu_{\max} + 1)$ matrix of coefficients has rank L (here, $\nu_{\max}$ denotes the maximum degree of $a_1(D), \ldots, a_L(D)$).

Armed with these results we are ready to prove the statement. We consider separately the following cases:

- $b_1(D) \neq 0$, $b_2(D) = 0$. In this case, the resulting subcode is formed by the matrices $\mathbf{G}_1(D)b_1(D)$, where

$$\mathbf{G}_1(D) = \begin{bmatrix} 1 + D^2 \\ 1 + D + D^2 \\ D \\ 1 \end{bmatrix},$$

  seen as a matrix of coefficients, has rank 3.

- $b_1(D) = 0$, $b_2(D) \neq 0$. Again, the resulting subcode is formed by the matrices $\mathbf{G}_2(D)b_2(D)$, where

$$\mathbf{G}_2(D) = \begin{bmatrix} D^2 \\ D \\ 1 + D^2 \\ 1 + D + D^2 \end{bmatrix},$$

  seen as a matrix of coefficients, has rank 3.

- $b_1(D) \neq 0$, $b_2(D) \neq 0$. In this case, since the polynomial ring over $\mathbb{F}_2$ has no zero divisors, the equation

$$A(D)b_1(D) + B(D)b_2(D) = 0 \tag{B2}$$

  for relatively prime $A(D)$ and $B(D)$ implies that $b_1(D) = B(D)q(D)$ and $b_2(D) = A(D)q(D)$, for an arbitrary $q(D) \neq 0$.

By letting equal to zero every single row, sum of two rows, sum of three rows or sum of four rows of (B1) we find equations of the form (B2). If $A(D)$ (resp. $B(D)$) is equal to zero, the equation is satisfied for $b_2(D)$ (resp. $b_1(D)$) equal to zero, obtaining one of the previous cases. If both $A(D)$ and $B(D)$ are non-zero, we obtain a subcode formed by the matrices $\mathbf{G}_{\mathrm{AB}}(D)q(D)$, where

$$
\mathbf{G}_{\mathrm{AB}}(D) = \begin{bmatrix} (1+D^2)B(D) + D^2A(D) \\ (1+D+D^2)B(D) + DA(D) \\ DB(D) + (1+D^2)A(D) \\ B(D) + (1+D+D^2)A(D) \end{bmatrix} \tag{B3}
$$

By representing the $\mathbf{G}_{\mathrm{AB}}(D)$'s as matrices of binary coefficients it can be verified that all of them have rank 3. Since the union of all such subcodes forms the set of codeword matrices $\mathbf{c}(D)$ with rank less than 4 over $\mathbb{F}_2$, we conclude that the rank-diversity of $\mathcal{S}$ is 3.

The extensive calculation of all $A(D)$ and $B(D)$ is lengthy and repetitive. We give just an example and we skip the rest for the sake of space limitation. Consider for example a linear dependence constraint on the first and third row. We have

$$
(1+D^2)b_1(D) + D^2b_2(D) + Db_1(D) + (1+D^2)b_2(D) = 0
$$

which yields

$$
(1+D+D^2)b_1(D) + b_2(D) = 0
$$

Then, by letting $A(D) = 1+D+D^2$ and $B(D) = 1$ in (B3) we obtain the 4×1 matrix

$$
\mathbf{G}_{\mathrm{AB}}(D) = \begin{bmatrix} 1+D^3+D^4 \\ 1+D^3 \\ 1+D^3+D^4 \\ D^2+D^4 \end{bmatrix}
$$

which corresponds to the matrix of binary coefficients

$$
\begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}
$$

with rank 3.

The proof for arbitrary $d > 0$ follows along the same lines by noticing that the WSTC $\mathcal{S}$ resulting from diagonal interleaving with delay d is obtained by applying componentwise the BPSK modulation mapping μ to all binary codeword arrays

$$
\mathbf{c}(D) = \begin{bmatrix} (1+D^2)b_1(D) + D^2b_2(D) \\ ((1+D+D^2)b_1(D) + Db_2(D))D^d \\ (Db_1(D) + (1+D^2)b_2(D))D^{2d} \\ (b_1(D) + (1+D+D^2)b_2(D))D^{3d} \end{bmatrix} . \tag{B4}
$$

Hence, the same reasoning as for $d = 0$ can be applied to (B4).

Acknowledgment

References

[1] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, October 1998.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, November 1999.

[3] C. Chuah, D. Tse, and J. Kahn, "Capacity of multi-antenna array systems in indoor wireless environment," in *Globecom '98*, Sydney, Australia, November 1998.

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

[5] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *to appear on IEEE Trans. on Inform. Theory*, 2001.

[6] E. Visotsky and U. Madhow, "Space-time transmit strategies and channel feedback generation for wireless fading channels," in *34-th Asilomar Conf. Signals, Systems and Computers (Asilomar 2000)*, Pacific Grove, CA, October 2000.

[7] H. Bolcskei, D. Gesber, and A. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," submitted to IEEE Trans. on Inform. Theory, October 1999.

[8] R. Muller, "A random matrix theory of communication via antenna arrays," submitted to IEEE Trans. on Inform. Theory, August 2000.

[9] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. on Inform. Theory*, vol. 45, no. 1, pp. 139–157, January 1999.

[10] L. Zheng and D. Tse, "Packing spheres in the Grassmann manifold: a geometric approach to the non-coherent multi-antenna channel," submitted to IEEE Trans. on Inform. Theory, April 2000.

[11] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," submitted to IEEE Trans. on Inform. Theory, August 2000.

[12] R. Knopp and G. Caire, "Power control schemes for TDD systems with multiple transmit and receiving antennas," in *Globecom '99*, Rio de Janeiro, December 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. on Inform. Theory*, vol. 44, no. 2, pp. 744–765, March 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. on Inform. Theory*, vol. 45, no. 5, pp. 1456, July 1999.

[16] J. Grimm, M. Fitz, and J. Krogmeyer, "Further results in space-time coding for Rayleigh fading," in *36-th Allerton conf. on Commun., Control and Comp.*, September 1998.

[17] S. Baro, G. Bauch, and A. Hansmann, "Improved codes for space-time trellis-coded modulation," *IEEE Comm. Letters*, vol. 4, no. 1, pp. 20–22, January 2000.

[18] X. Lin and R. Blum, "On the design of space-time codes employing multiple trellis coded modulation," in *2000 Conf. on Inform. Sc. and Syst., CISS 2000*, Princeton Uuniversity, NJ, March 2000.

[19] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," submitted to IEEE Trans. on Inform. Theory, August 2000.

[20] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. on Commun.*, vol. 43, no. 2, pp. 354–364, February 1995.

[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, no. 11, pp. 1841–1852, November 1999.

[22] R. Hammons and H. El Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Trans. on Inform. Theory*, vol. 46, no. 2, pp. 524–542, March 2000.

[23] E. Biglieri, D. Divsalar, P. McLane, and M. Simon, *Introduction to trellis-coded modulation with applications*, Macmillan, New York, 1991.

[24] D.-S. Shiu and J. Kahn, "Layered space-time codes for wireless communications using multiple transmit antennas," in *IEEE Intern. Conf. on Commun., ICC '99*, Vacouver, Canada, June 1999.

[25] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. on Commun.*, vol. 37, no. 5, pp. 428–436, May 1989.

[26] R. Ashkenazi and S. Shamai, "Suboptimal detection for intersymbol interference inflicted cable channels," *AEU*, vol. 51, no. 5, pp. 246–254, May 1997.

[27] V. Eyuboglu and S. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. on Commun.*, vol. 36, no. 1, pp. 13–20, January 1988.

[28] P. Chevillat and E. Eleftheriou, "Decoding of trellis-encoded signals in the presence of intersymbol interference and noise," *IEEE Trans. on Commun.*, vol. 37, no. 7, pp. 669–676, July 1989.

[29] A. Viterbi and J. Omura, *Principles of digital communication and coding*, McGraw-Hill, New York, 1979.

[30] J. Proakis, *Digital communications, 3rd Ed.*, McGraw-Hill, New York, 1995.

[31] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, New York, 1985.

[32] R. Knopp and P. Humblet, "On coding for block fading channels," *IEEE Trans. on Inform. Theory*, vol. 46, no. 1, pp. 189–205, January 2000.

[33] E. Malkamaki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1643–1646, July 1999.

[34] R. Wesel, X. Liu, and W. Shi, "Trellis codes for periodic erasures," *IEEE Trans. on Commun.*, vol. 48, no. 6, pp. 938–947, June 2000.

[35] R. E. Blahut, *Principles and practice of information theory*, Addison-Wesley, New York, 1987.

[36] R. Knopp and P. Humblet, "Maximizing diversity in block fading channels," in *IEEE Intern. Conf. on Commun. ICC '97*, Montreal, June 1997, pp. 647–651.

[37] D. Forney and M. Trott, "The dynamics of group codes: state spaces, trellis diagrams, and canonical encoders," *IEEE Trans. on Inform. Theory*, vol. 39, no. 5, pp. 1491–1513, September 1993.

[38] A. Lapidoth, "The performance of convolutional codes on the block erasure channel using various finite interleaving techniques," *IEEE Trans. on Inform. Theory*, vol. 40, no. 5, pp. 1459–1473, September 1994.

[39] D. Forney Jr., "Geometrically uniform codes," *IEEE Trans. on Inform. Theory*, vol. 37, no. 5, pp. 1241–1260, September 1991.

[40] Y. Leung, S. Wilson, and J. Ketchum, "Multifrequency trellis coding with low delay for fading channels," *IEEE Trans. on Commun.*, vol. 41, no. 10, pp. 1450–1458, October 1993.

[41] R. Wesel and X. Liu, "Analytic techniques for periodic trellis codes," in *36-th Allerton conf. on Commun., Control and Comp.*, September 1998, pp. 39–48.

[42] M. K. Simon and M. S. Alouini, "A unified approach to the performance analysis of digital communication over generalized fading channels," *Proc. of the IEEE*, vol. 86, no. 9, pp. 1860–1877, Sept. 1998.

[43] W. Sheen and G. Stuber, "Error probability for reduced-state sequence estimation," *IEEE J. Select. Areas Commun.*, vol. 10, no. 3, pp. 571–578, April 1992.

[44] S. Shamai and R. Laroia, "The intersymbol interference channel: lower bounds on capacity and channel precoding loss," *IEEE Trans. on Inform. Theory*, vol. 42, no. 5, pp. 1388–1404, September 1996.

[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 Conference*, Pacific Groove, CA, November 1997.

[46] T. Guess and M. Varanasi, "An information-theoretic derivation of the MMSE decision-feedback equalizer," in *36-th Allerton conf. on Commun., Control and Comp.*, September 1998.

[47] V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. on Inform. Theory*, vol. 43, pp. 858–871, May 1997.

[48] J. Zhang, E. Chong, and D. Tse, "Output MAI distribution of linear MMSE multiuser receivers in CDMA systems," *to appear on IEEE Trans. on Inform. Theory*, 2001.
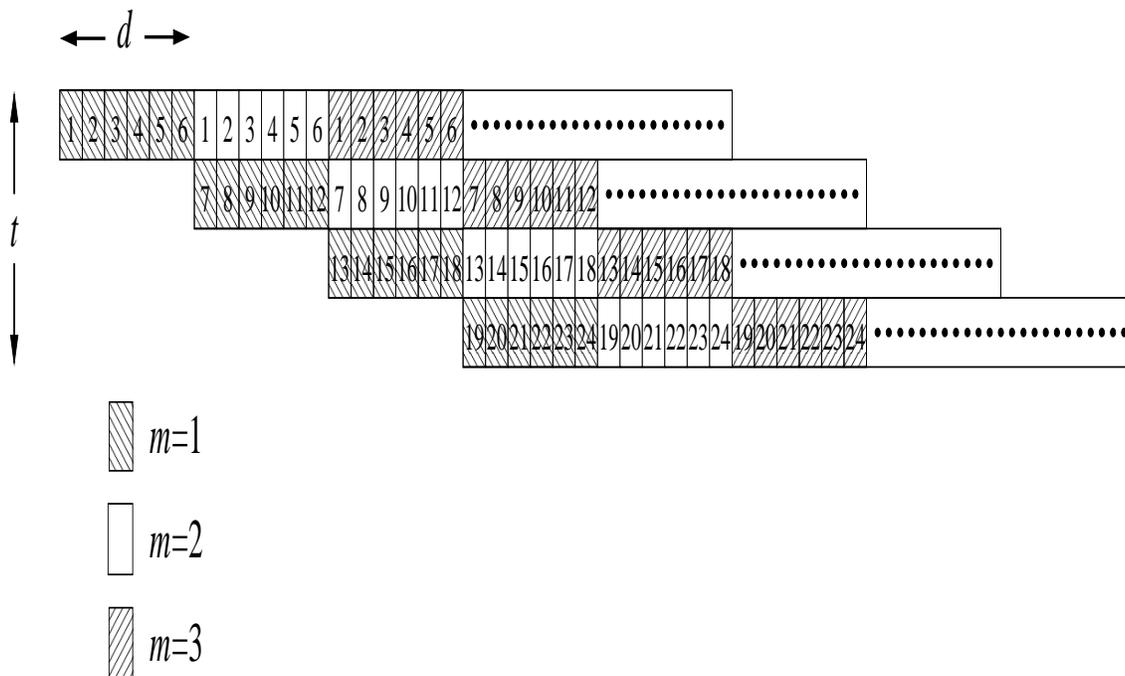
Figure 1: Layered ST code with $t = 4$, $d = 6$ and $N' = 24$. The index m indicates the component codewords $\mathbf{c}^{(m)}$ and the integer entries in the array indicate the index of the ℓ-th element $c_\ell^{(m)}$ of $\mathbf{c}^{(m)}$.
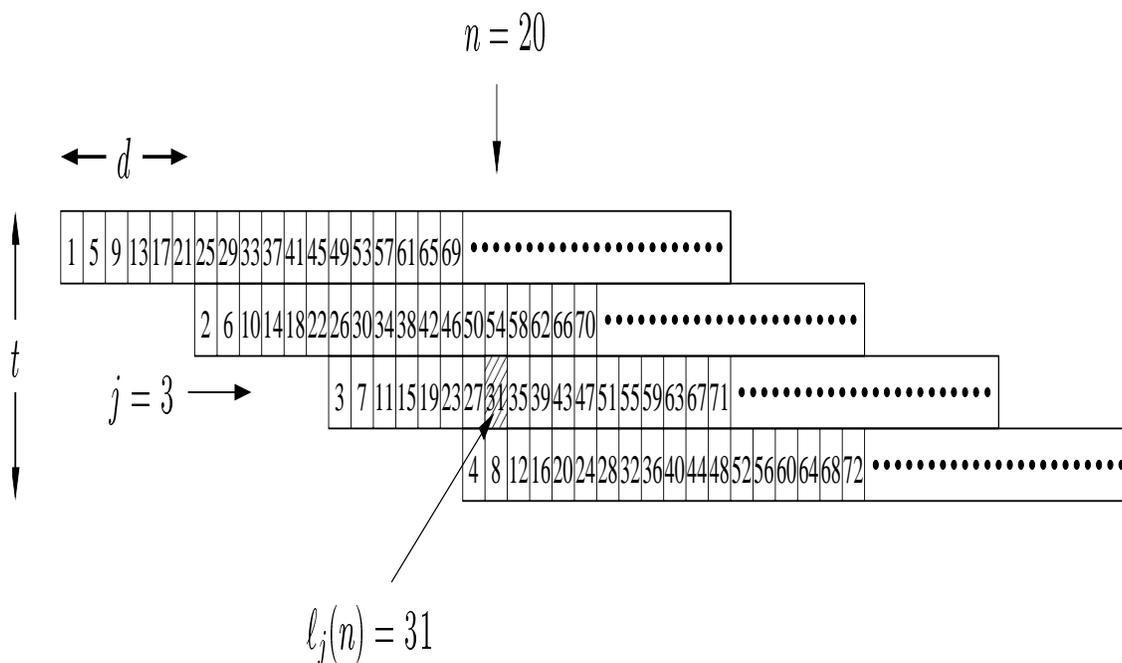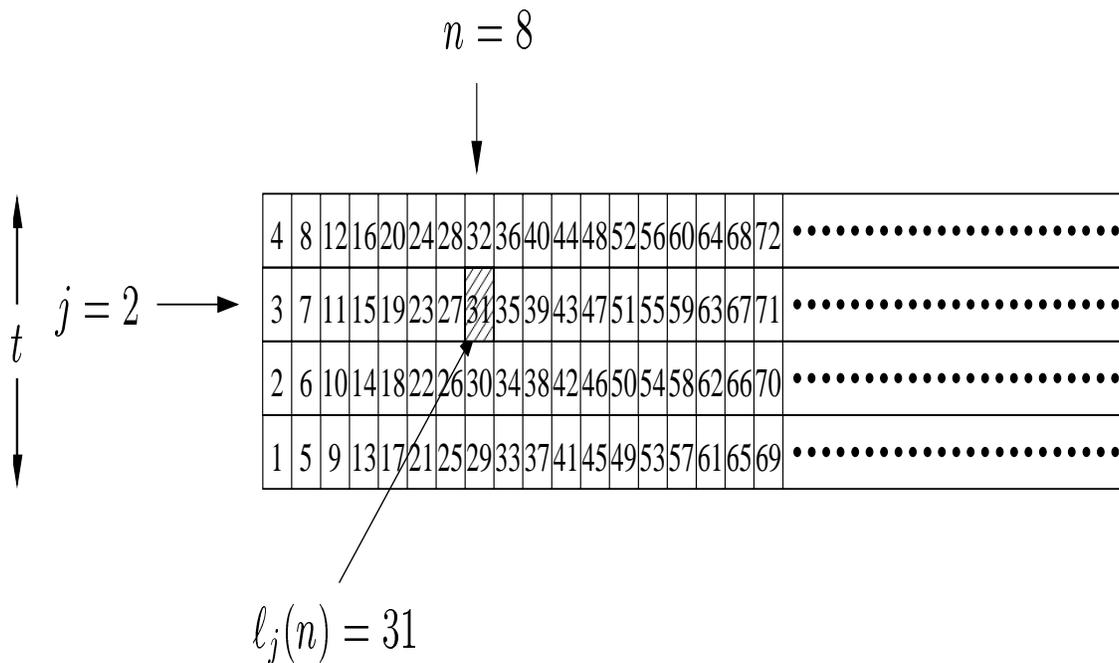
Figure 2: Wrapped ST code with $t = 4$ and $d = 6$. The integer entries in the array indicates the index of the ℓ-th element c_ℓ of the component codeword $\mathbf{c}$. An example of the indexing rule $\ell_j(n)$ is given for $j = 3$ and $n = 20$.

$$n = 8$$

4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68	72	• •
3	7	11	15	19	23	27	31	35	39	43	47	51	55	59	63	67	71	• •
2	6	10	14	18	22	26	30	34	38	42	46	50	54	58	62	66	70	• •
1	5	9	13	17	21	25	29	33	37	41	45	49	53	57	61	65	69	• •

$j = 2$

t

$$\ell_j(n) = 31$$

Figure 3: Wrapped ST code with $t = 4$ and $d = 0$. The integer entries in the array indicates the index of the ℓ-th element c_ℓ of the component codeword $\mathbf{c}$. An example of the indexing rule $\ell_j(n)$ is given for $j = 2$ and $n = 8$.
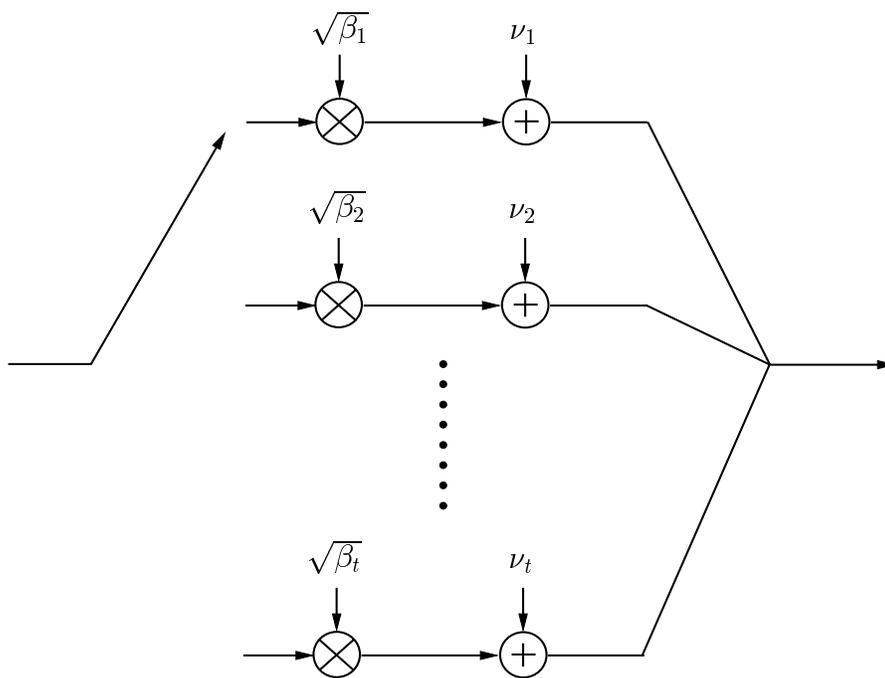
Figure 4: Parallel channel model with cyclic interleaving originated by the PSP-based VA decoding of WSTCs subject to the assumption of perfect feedback decisions.
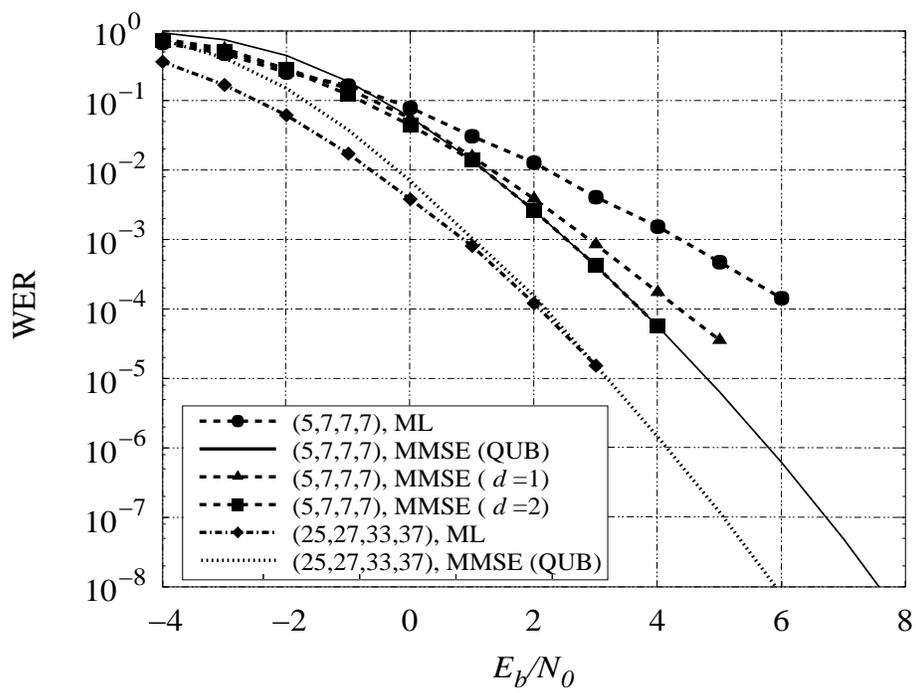
Figure 5: WER of the WSTCs constructed from the CCs with generators (5,7,7,7) and (25,27,33,37) mapped onto BPSK, independent Rayleigh fading, $t = r = 4$.

Current trellis step

Common decoded path

Unmerged survivors

Past error event

Merge point
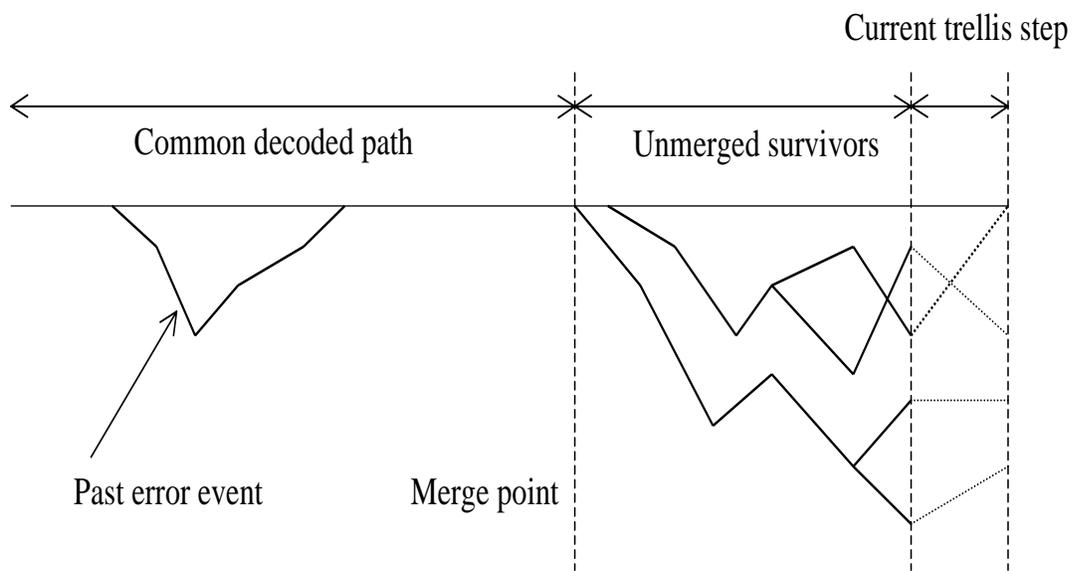
Figure 6: Sketch representation of the trellis in the PSP-based VA decoder. Paths are extended from left to right and only survivors are shown.
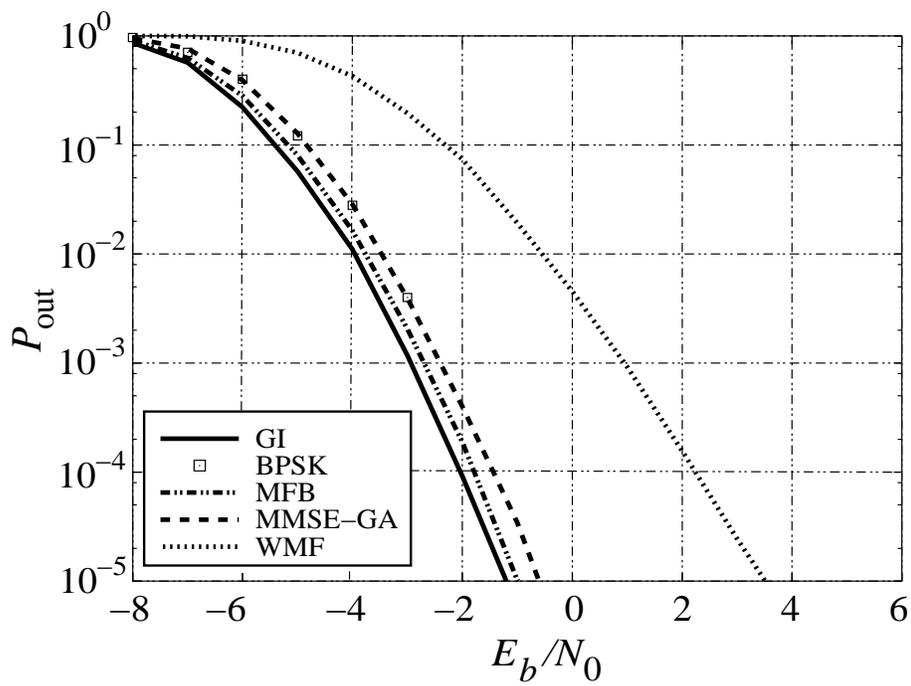
Figure 7: Outage probability vs. E_b/N_0 for BPSK, independent Rayleigh fading, $t = r = 4$ and $\eta = 1$ bit/channel use.
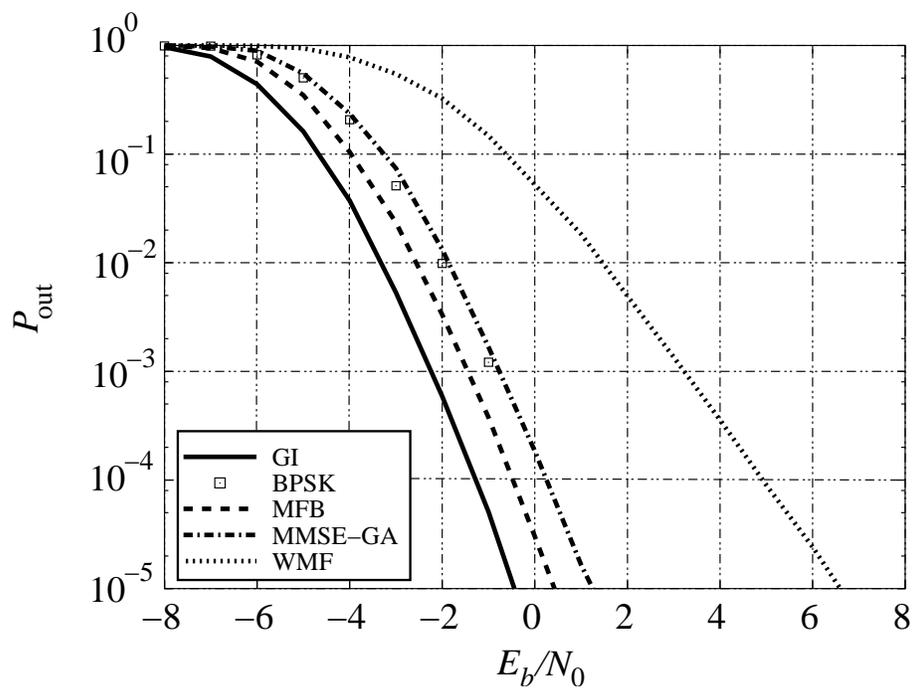
Figure 8: Outage probability vs. E_b/N_0 for BPSK, independent Rayleigh fading, $t = r = 4$ and $\eta = 2$ bit/channel use.
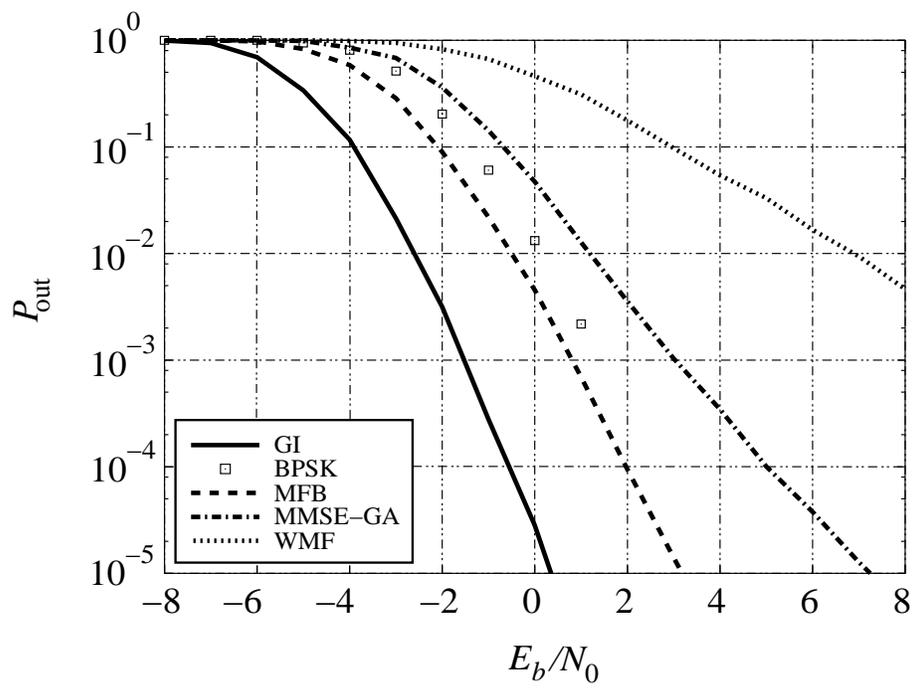
Figure 9: Outage probability vs. E_b/N_0 for BPSK, independent Rayleigh fading, $t = r = 4$ and $\eta = 3$ bit/channel use.
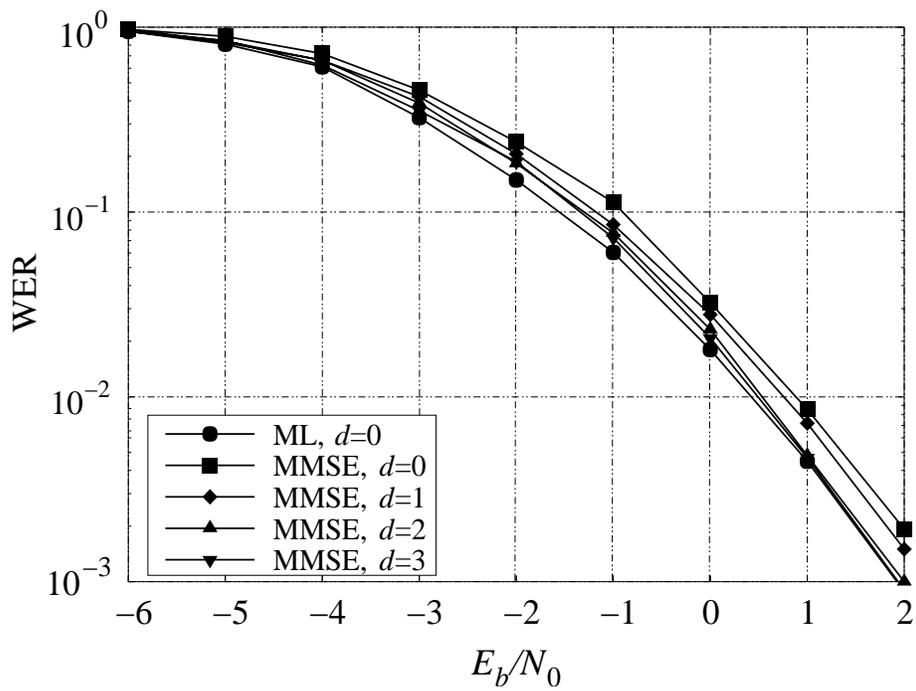
Figure 10: WER of WSTCs obtained from the code (23,25) and BPSK, with MMSE front-end and various values of d. Independent Rayleigh fading, $t = r = 4$. Spectral efficiency $\eta = 2$ bit/channel use.
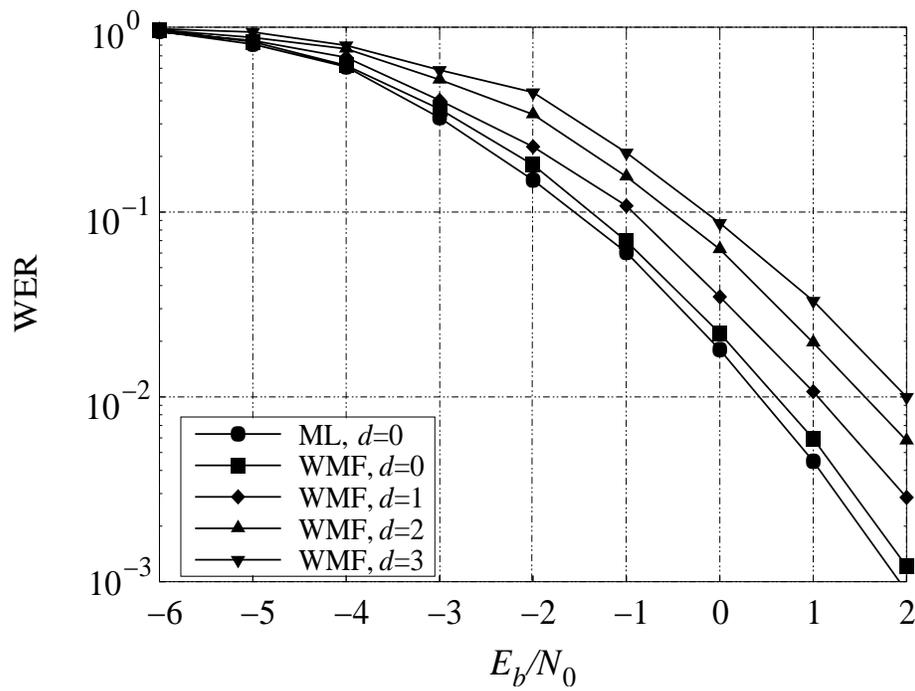
Figure 11: WER of WSTCs obtained from the code (23,25) and BPSK, with WMF front-end and various values of d. Independent Rayleigh fading, $t = r = 4$. Spectral efficiency $\eta = 2$ bit/channel use.
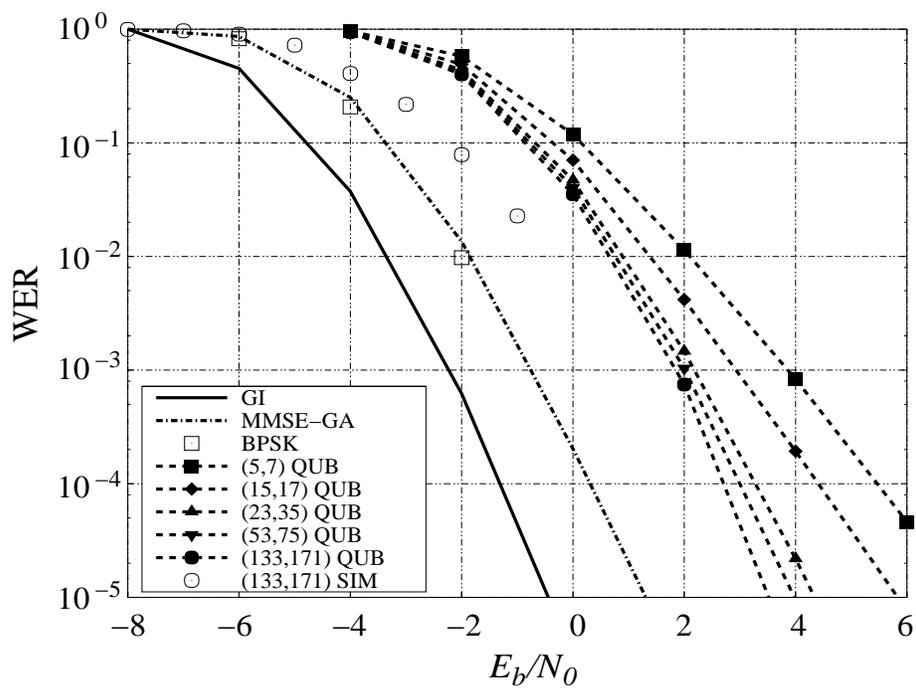
Figure 12: WER vs. E_b/N_0 for WSTCs based on binary CCs of rate 1/2 mapped onto BPSK, independent Rayleigh fading, with $t = r = 4$.
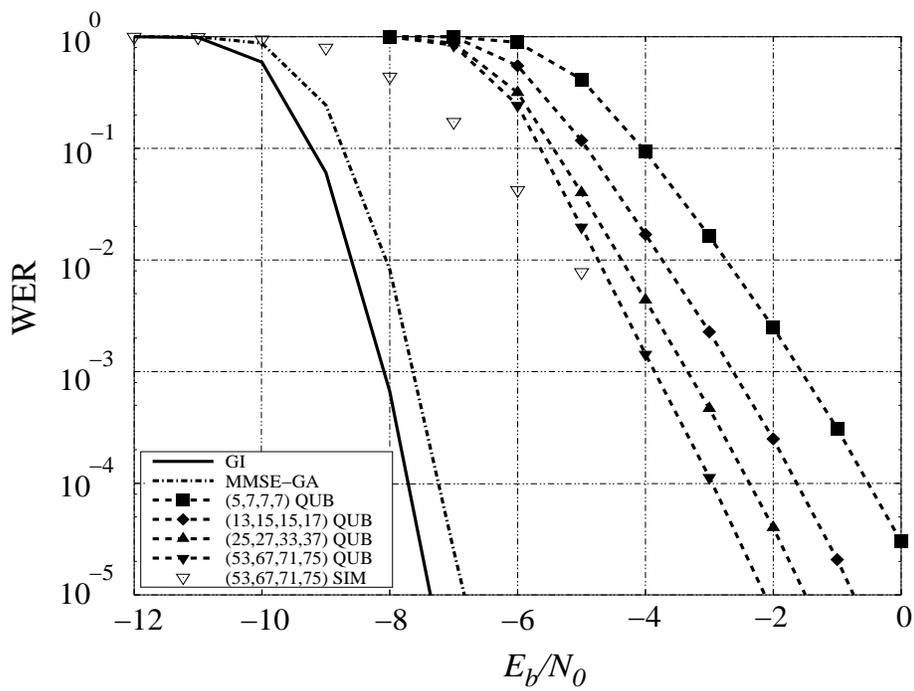
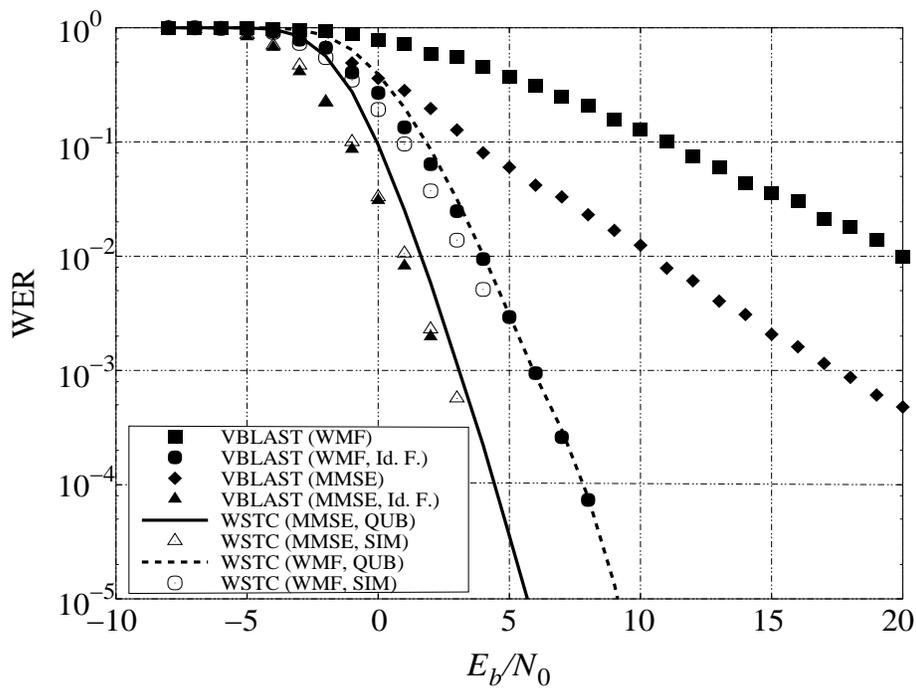Figure 13: WER vs. E_b/N_0 for WSTCs based on binary CCs of rate 1/4 mapped onto BPSK, independent Rayleigh fading, with $t = r = 8$.
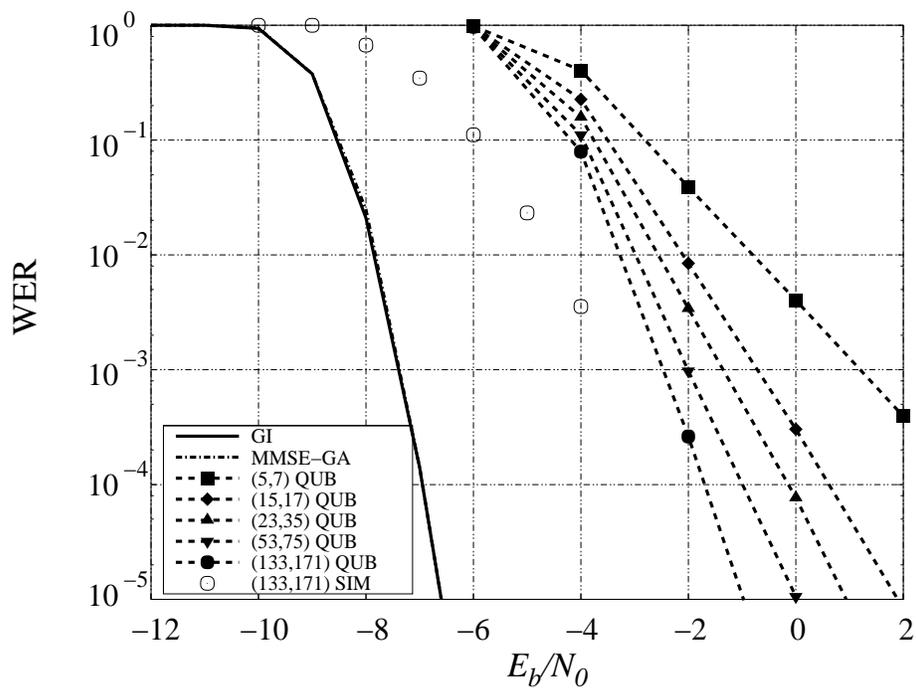
Figure 14: WER vs. E_b/N_0 for WSTCs based on binary CCs of rate 1/2 mapped onto QPSK, independent Rayleigh fading, with $t = r = 8$.

Figure 15: WER of WSTC and V-BLAST schemes with code (23,25) and QPSK. Independent Rayleigh fading, $t = r = 4$. Spectral efficiency $\eta = 4$ bit/channel use.
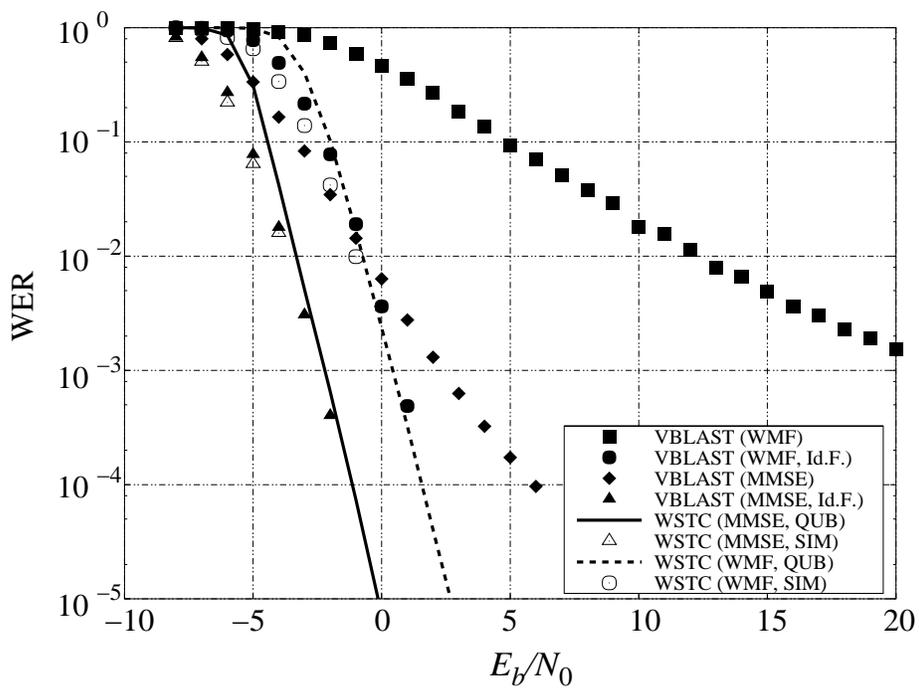
Figure 16: WER of WSTC and V-BLAST schemes with code (23,25) and BPSK. Independent Rayleigh fading, $t = r = 8$. Spectral efficiency $\eta = 4$ bit/channel use.
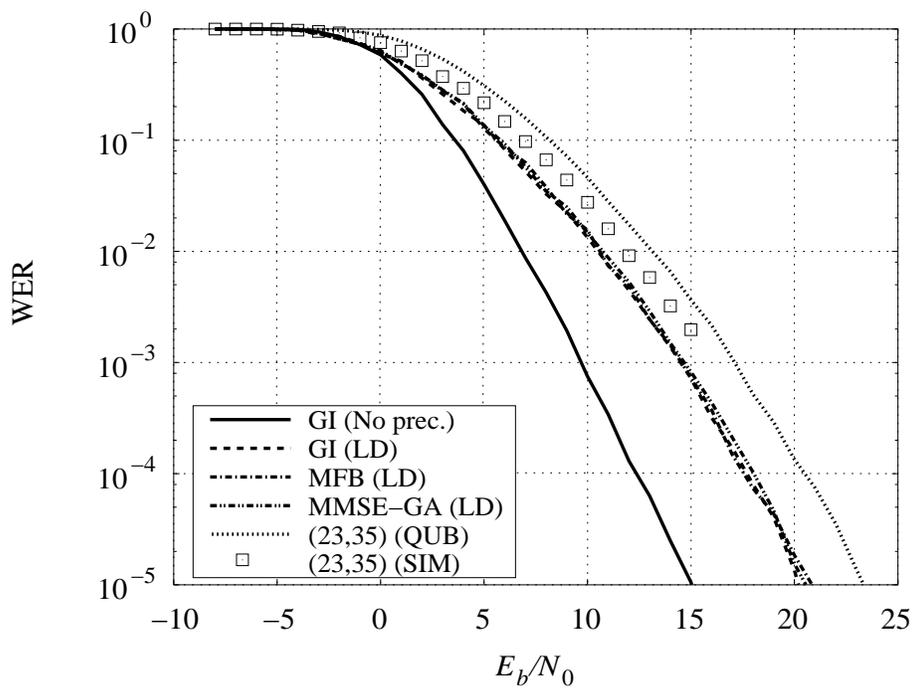
Figure 17: WER of WSTC with code (23,25) and QPSK concatenated with LD-precoding. Independent Rayleigh fading, $t = 4, r = 1$. Spectral efficiency $\eta = 1$ bit/channel use.
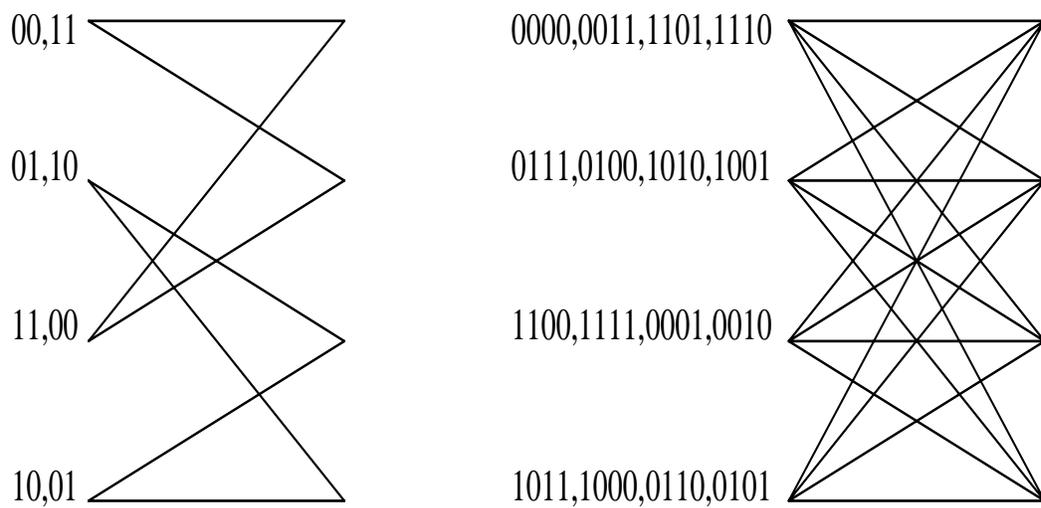
Figure 18: Original trellis (left) and super-trellis (right) corresponding to $s = 2$ consecutive steps for the binary linear CC with generators $(5, 7)$. The transitions labels next to each state, read form left to right, correspond to the transitions stemming from the state from top to bottom.