Space–Time Coding Techniques With Bit-Interleaved Coded Modulations for MIMO Block-Fading Channels

Nicolas Gresset, Member, IEEE, Loïc Brunel, Member, IEEE, and Joseph J. Boutros, Member, IEEE

Abstract—The space–time bit-interleaved coded modulation (ST-BICM) is an efficient technique to obtain high diversity and coding gain on a block-fading multiple-input multiple-output (MIMO) channel. Its maximum-likelihood (ML) performance is computed under ideal interleaving conditions, which enables a global optimization taking into account channel coding. Thanks to a diversity upper bound derived from the Singleton bound, an appropriate choice of the time dimension of the space–time coding is possible, which maximizes diversity while minimizing complexity. Based on the analysis, an optimized interleaver and a set of linear precoders, called dispersive nucleo algebraic (DNA) precoders are proposed. The proposed precoders have good performance with respect to the state of the art and exist for any number of transmit antennas and any time dimension. With turbo codes, they exhibit a frame error rate which does not increase with frame length.

Index Terms—Bit-interleaved coded modulation, interleaving, multiple-antenna channels, Singleton bound, space–time coding.

I. INTRODUCTION

The wide panel of today’s wireless transmission contexts makes implausible the existence of a miraculous universal solution which always exhibits good performance with low complexity. Different scenarios (indoor, outdoor with low velocity, outdoor with high velocity) correspond to different amounts of time and frequency diversity. The success of the multicarrier modulation as a solution for future wireless systems is partly due to the low receiver complexity even over large frequency bands. In this paper, we focus on an indoor environment and design a system approaching the optimal performance taught by information theory. In a wireless indoor environment, both time and frequency diversities may be poor due to small terminal velocity and possibly very short channel impulse response. These particularly tricky low-diversity channels are modeled as block-fading channels. Over low-diversity multiple-input multiple-output (MIMO) channels, space–time coding techniques often enable transmission with improved data rate and diversity, within a limit given by the rank of the MIMO system [1], [18], [19], [21]. These open-loop schemes only require the knowledge of the channel long-term statistics. Besides, closed-loop techniques such as beamforming take benefit from a short-term channel knowledge to improve the performance/complexity tradeoff at the cost of additional signalling overhead. As a first step in providing increased data rates in future generations of indoor wireless local-access networks (WLANs), we study how to appropriately choose the channel coding, the channel interleaving, and the space–time coding.

For frame sizes of practical interest, coded modulations have to be considered since space–time codes employed with uncoded modulations exhibit a frame error rate (FER), which is dramatically degraded [24, Appendix A]. Thus, we focus on the bit-interleaved coded modulation (BICM) structure, which is the concatenation of a channel encoder, an interleaver and a modulator. The analysis of the BICM maximum-likelihood (ML) performance is tractable and eases the coded modulation design. Furthermore, thanks to the interleaver, iterative processing at the receiver achieves quasi-ML performance with reduced complexity. On a MIMO channel, the BICM may be concatenated with a simple full-rate space-division multiplexing scheme (SDM) [21]. In this paper, we improve performance of this space–time BICM (ST-BICM) by replacing the SDM by a more efficient full-rate linear space–time code: a linear precoding or equivalently a space–time spreading. Linear precoding is performed by multiplying the complex multiple-antenna signal by a square complex space–time matrix. The space–time matrix enhances the diversity by mixing the symbols of different time periods and antennas together.

The choice of the ST-BICM structure may also be explained as follows: We aim at optimizing a full-rate space–time code based on linear precoding, taking into account the structure of the whole transmitter, which inevitably includes an error-correcting code, an interleaver, and a symbol mapper. Usually, space–time codes are designed independently from the other elements of the transmitter. However, frames of bits are linked through the error-correction code and optimizing the space–time code taking into account the whole transmitter is equivalent to optimizing a BICM concatenated with a space–time code, i.e., an ST-BICM.

On an ergodic channel, the achieved diversity order is equal to the code minimum distance multiplied by the number of receive antennas. In most cases, the minimum distance is high enough and increasing diversity through linear precoding does not bring
much improvement. For a block-fading channel, the diversity is upper-bounded by the number of channel realizations in a code-word multiplied by the number of transmit antennas and the number of receive antennas. Using the Singleton bound, we will exhibit an additional upper bound on the diversity order, which may be very limiting without precoding. Hence, in this paper, we will study ST-BICM with linear precoding, focusing on the block-fading channel and optimize the linear precoding using the ST-BICM ML performance in order to achieve full diversity and maximum coding gain. First, we derive the coding gain of an ideal ST-BICM. It is related to the notion of Shannon code and sphere hardening [39]. Indeed, the ideal Shannon code for additive white Gaussian noise (AWGN) channels is located near a sphere, called the Shannon’s sphere. Thanks to the interleaver, the squared Euclidean norm of BICM codewords has low variance, which implies that codewords lie close to the Shannon’s sphere. The BICM may be seen as a quantization of this sphere, which should be as uniform as possible to maximize the size of Voronoi regions. On MIMO fading channels, the Shannon’s sphere becomes a Shannon’s ellipsoid [20] and BICM codewords are randomly located close to the ellipsoid. We show that the ideal BICM configuration maximizes the Voronoi region volume whatever the channel realization. We present a practical system that approaches the ideal BICM configuration including the so-called dispersive nucleo algebraic (DNA) precoder and compare its performance to the ideal ST-BICM performance. The DNA precoder exists for any numbers of space and time dimensions. We finally design a practical interleaver, which approximates the ideal interleaving conditions.

The paper is organized as follows: In Section II, the ST-BICM transmitter and the associated iterative receiver are presented. In Section III, we derive the analytical ML performance under ideal interleaving assumption for an ergodic channel without precoding and a block-fading channel with and without precoding. Using the Singleton bound, we show in Section IV that ideal interleaving conditions cannot be achieved on a block-fading channel with any kind of parameters and that linear precoding may be mandatory in some configurations. Section V describes the linear precoding optimization for a block-fading channel and Section VI the interleaver design for convolutional codes and its application to turbo codes. Finally, simulation results are presented in Section VII, which confirm the behavior which was expected from the analytical study. Furthermore, they show the good performance of the DNA precoders and the advantage of using turbo codes to get a nonincreasing FER when the frame size increases.

II. SYSTEM MODEL AND NOTATIONS

A. Transmitter Scheme

The transmitter scheme is built from the following fundamental block concatenation: A binary error-correcting code \( C \) followed by a deterministic interleaver \( \Pi \), a symbol mapper (e.g., for a quadrature amplitude modulation (QAM)), a full-rate space–time period \( S \) (i.e., a linear precoder) and a set of \( n_t \) transmit antennas. Fig. 1 illustrates the BICM transmitter structure.

Without loss of generality, we assume that the error-correcting code \( C \) is a convolutional code with rate \( R_C \). The encoder associates with the input information word \( b \) the codeword \( c \in C \). Sequence \( b \) (resp., \( c \)) has length \( K_C L_C \) (resp., \( N_C L_C \)) bits, where \( L_C \) is the codeword length in trellis branches. The interleaver \( \Pi \), which scrambles the \( L_C N_C \) coded bits, is a crucial function in the BICM structure, as it allows the receiver to perform iterative joint detection and decoding. Indeed, it ensures independence between extrinsic and \textit{a priori} probabilities, in both the detector and the decoder. Furthermore, when ML decoding is tractable, interleaving prevents erroneous bits of a same error event from interfering with each other in the same precoded symbol. The interleaver \( \Pi \) may be pseudorandom (PR) or semi-deterministic with some deterministic constraints as described in Section VI.

In the symbol mapper, \( m \) consecutive interleaved coded bits are mapped together onto a modulation symbol, according to a bijection between bit vectors and modulation symbols called mapping or labeling. The number of modulation symbols is equal to \( M = 2^m \). For each channel use, i.e., in each time period, the mapper reads \( m n_t \) coded bits and generates \( n_t \) modulation symbols. In order to make the reading easier, the obtained \( n_t \)-dimensional constellation \( \Omega \) will denote both the set of symbols and the set of binary labelings. Throughout this paper, we will consider QAM modulations as they achieve a good compromise between spectral efficiency (in bits per second per hertz (bits/s/Hz) or bits per dimension (bits/dim)) and performance. Moreover, with QAM modulation, the system is easily modeled using a lattice constellation structure [10], which gives access to the lattice theory toolbox, both for transmitter and receiver optimizations. We assume that the QAM modulation has unit energy. The linear precoder \( S \) spreads the QAM symbols over \( s \) time periods. It converts the \( n_t \times n_r \) vector channel into an \( N_t \times N_r \) vector channel, where \( N_t = n_t s \) and \( N_r = n_r s \). The \( N_t \times N_r \) matrix \( S \) multiplies a vector of \( N_t \) QAM symbols \( z_i = (z_{i1}, z_{i2}, \ldots, z_{in}) \) at the mapper output, generating \( N_t \) symbols to be transmitted during \( s \) time periods. Vector \( z_i \) is the
\( \cdot \)th vector to be precoded. The precoder \( S \) spreads the transmitted symbols over a higher number of channel states in order to exploit diversity. The \( S \) matrix is normalized as follows:

\[
\sum_{i=1}^{N_c} \sum_{t=1}^{N_t} |S_{it,j}|^2 = N_t.
\]

(1)

In this paper, we assume a block-fading channel with \( n_c \) distinct channel realizations during a codeword. We denote \( n_s \) the number of distinct channel realizations during a precoded symbol. To simplify notations, we assume that \( n_s \) divides \( n_c \). We will call \( channel \ state \) the \( 1 \times n_c \) single-input multiple-output (SIMO) channel associated with one of the \( n_t \) transmit antennas and one of the \( n_c \) channel realizations. The channel experienced by precoded symbol \( \hat{i} \) is represented by an \( N_t \times N_r \) block-diagonal matrix \( H_i \) with \( s \) blocks of size \( n_t \times n_r \). During one precoded symbol, we assume that each of the \( n_s \) channel realizations is repeated \( s/n_s \) times. The \( H_i \) matrix is organized as follows:

\[
H_i = \text{diag} \left( H_i^{[1]}, \ldots, H_i^{[s]} \right),
\]

(2)

where \( H_i^{[t]} \) denotes the \( n_t \times n_r \) complex matrix representing the \( t \)th channel realization experienced by the \( \hat{i} \)th precoded symbol. The \( H_i^{[t]} \) matrix is repeated \( s/n_s \) times. Elements of \( H_i^{[t]} \) are independent complex Gaussian variables with zero mean and unit variance. Let \( \mathcal{H} \) denote the set of channel realizations observed during the transmission of a codeword. Thanks to the extended channel matrix, we write the channel input–output relation as

\[
y_i = x_i + \eta_i = z_i S H_i + \eta_i,
\]

(3)

where \( y_i \in \mathbb{C}^{N_r} \) and each receive antenna is perturbed by an additive white complex Gaussian noise \( \eta_{i,j} \), \( j = 1, \ldots, N_r \), with zero mean and variance \( 2N_0 \). We define the signal-to-noise ratio \( E_b/N_0 \), where \( E_b \) is the total energy of an information bit at the receiver. Thanks to linear precoding, the \( n_t \times n_r \) MIMO \( n_c \)-block-fading channel is converted into an \( N_t \times N_r \) MIMO \( N_c \)-block-fading channel where \( N_c = n_c / n_s \). If \( n_s = 1 \), the precoder experiences a quasi-static \( n_t \times n_r \) MIMO channel. In the following, index \( \hat{i} \) will be omitted if a single precoded symbol is considered and precoding time period will refer to a transmission over \( SH \), i.e., over \( s \) time periods.

The concatenation of the binary error-correcting code \( C \), the interleaver \( \Pi \), the mapper \( \Omega \), the linear precoder \( S \), and the channel describes a global Euclidean code \( C_E \) which converts \( L_C K_c \) information bits into a complex \( L_C N_c/m \)-dimensional point.

### B. Iterative Receiver Scheme

An ideal BICM receiver would directly perform an ML decoding on the set \( C_E \) of transmitted codewords. However, it requires an exhaustive search among the \( 2^{K_c L_c} \) codewords, which is intractable. All existing receivers use the concatenated structure of the BICM to split the reception into several steps. In this paper, we assume perfect synchronization and channel estimation. Thus, the receiver, as depicted in Fig. 2, is divided in two main elements: a soft-input soft-output (SISO) a posteriori probability (APP) QAM detector, which acts as a soft-output equalizer for both the space–time spreader and the MIMO channel, converting the received point \( \mathbf{y} \) into information on the coded bits in the estimated coded sequence \( \hat{c} \), and a SISO decoder for \( C \), improving the information on coded bits and estimating the information bit sequence \( \hat{b} \). The depicted iterative joint detection and decoding process is based on the exchange of soft values between these two elements. The SISO detector computes extrinsic probabilities \( \xi(c) \) on coded bits thanks to the conditional likelihoods \( p(y_i|z_i) \) and the a priori probabilities \( \pi(c) \) fed back from the SISO decoder

\[
\xi(c) = \frac{\sum_{z_i} \exp \left( -\frac{1}{2N_0} |y_i - z_i S H_i|^2 \right) \prod_{\ell \neq i} \pi(c_{\ell})}{\sum_{z_i} \exp \left( -\frac{1}{2N_0} |y_i - z_i S H_i|^2 \right) \prod_{\ell \neq i} \pi(c_{\ell})}
\]

(4)

where \( \Omega \) is the Cartesian product \( (M\text{-QAM})^{N_c} \), i.e., the set of all vectors \( \mathbf{z} \) generated by the QAM mapper, \( |\mathbf{z}| = 2^m N_r \). The subset \( \Omega(c_{\ell}) = 1 \), for \( \ell = 0, 1, \ldots, n_c - 1 \), is restricted to the vectors \( \mathbf{z} \) in which the \( \ell \)th coded bit is equal to 1. The detector independently computes the soft outputs for each precoding time period. At the first iteration, there is no a priori information available at the detector input. Through the iterations, the a priori probability on constellation points computed from the probabilities fed back by the SISO decoder becomes more
and more accurate. Ideal convergence is achieved when a priori probabilities provided by the decoder are perfect, i.e., equal to 0 or 1. The decoder uses a forward–backward algorithm [2], which computes the exact extrinsic probability using the trellis structure of the code.

III. THEORETICAL PERFORMANCE FOR IDEALLY INTERLEAVED BICM

A lot of study has been done to estimate the frame or bit-error rate of the BICM with ML decoding, in particular using Gaussian approximations or numerical integrations [5], but a closed-form expression of the pairwise error probability had not been derived yet. This section first describes an accurate computation of bit- and frame-error rates of BICM ML performance over ergodic MIMO channel with ideal interleaving and without precoding. A more detailed description of the derivation may be found in [23] and [24]. Under ideal interleaving condition, we are able to derive a closed-form expression of the probability density of the log-likelihood ratio (LLR) at the output of the detector and then a closed-form expression of the pairwise error probability at the output of the decoder. It is then straightforward to use well-known techniques to estimate the bit- or frame-error rate of a coded modulation from pairwise error probability. This subject has been extensively discussed for coded modulations over AWGN channels. Examples are the union bound on the transfer function of a convolutional code and the more accurate tangential sphere bound [36] for spherical constellations.

Subsequently, we extend the study to the block-fading MIMO channel when linear precoding is used at the transmitter. Note that the method is also valid for correlated MIMO channels. We extract from the bit-error rate (BER) expression some design criteria on BICM precoder, interleaver, and error-correcting code.

A. Ideal Interleaving Condition

The evaluation of the BER or FER of a coded modulation is usually based on the derivation of an upper bound on the actual performance obtained by a balanced summation of pairwise error probabilities. Each pairwise error probability involves the Euclidean distance between two codewords with a Hamming distance \(w\).

With an \(n_t \times n_r\) MIMO block-fading channel with \(n_r\) blocks, the minimum diversity recovered at the detector output, and thus at the decoder output, is always equal to the reception diversity \(n_r\). Let us consider an error event with \(w\) erroneous bits and assume that the maximum diversity order is \(T_{\text{max}}\). If \(w \geq T_{\text{max}}/n_r\), we achieve full diversity if each of the \(T_{\text{max}}/n_r\) independent fading random variables is experienced by at least one bit among \(w\). In a precoding time period \(k\) in which at least one erroneous bit is transmitted, the transmitted and competing points are called \(x_k = z_kSH_k\) and \(x_k' = z_k'SH_k\). When performing ML decoding or APP detection, we are interested in the equivalent binary shift keying (BSK) modulation defined by the two points \(x_k\) and \(x_k'\). The vector \((z_k - z_k')SH_k\) has \(sn_r\) circular symmetric Gaussian components. Thus, whatever the number of erroneous bits on a precoding time period, the obtained diversity is limited to \(sn_r\). Having several erroneous bits per precoding time period is useless. On the contrary, if the erroneous bits are located on different precoding time periods and experience different fading random variables, a higher diversity is achieved. This is what we call the noninterference property. Furthermore, we will see in Section V that an equi-distribution of erroneous bits on channel states is required to achieve a maximum coding gain. We call it the equi-distribution property. The ideal interleaver is defined as follows.

**Definition 1:** (Ideal Interleaving) For any pair of codewords with \(w\) different bits at positions \(i_1, \ldots, i_w\), an ideal interleaver allocates the bits to transmitted symbols as follows:

- noninterference property: \(\forall i_k, i_{k'}, \text{bits at positions } i_k \text{ and } i_{k'} \text{ are transmitted on different precoding time periods;}
- equi-distribution property: the bits at positions \(i_1, \ldots, i_w\) are as equiprobably distributed over all channel states as allowed by \(w\).

In practice, such an interleaver does not always exist. We will see in the following that the Singleton bound gives an existence condition of the ideal interleaver. In Section VI, we present optimized interleavers that approach the ideal condition.

B. Exact Pairwise Error Probability For Ergodic Channels Without Precoding

In [23], we have established a closed-form expression for the conditional pairwise error probability on ergodic MIMO channels under ML decoding of the BICM and ideal channel interleaving. The mathematical derivation in this subsection follows [23]. Transmitted symbols are not precoded: \(s = 1\), \(S = I_{n_t}\), the \(n_t \times n_t\) identity matrix. Thus, (2) reduces to \(H_k = H_k^H\).

Consider the pairwise error probability that a codeword \(c \in C\) is transmitted and a codeword \(c' \in C\) is decoded. The \(w\) different bits between the two codewords are transmitted in \(w\) different time periods, complementing one bit in the mapping of one of the \(n_t\) QAM symbols. The transmitted noiseless vectors corresponding to the two codewords \((c, c')\) only differ in \(w\) positions. Let us define \(Z = (z_1, \ldots, z_{n_t})\) and \(Z' = (z'_1, \ldots, z'_{n_t})\) the \(wn_t\)-dimensional vectors corresponding to these positions and \(X = (x_1, \ldots, x_{n_t})\) and \(X' = (x'_1, \ldots, x'_{n_t})\) the \(wn_t\)-dimensional vectors corresponding to \(Z\) and \(Z'\) and filtered by the channel matrix \(H_k\).

We define \(d_k = ||z_k - z_k'||\). The Euclidean distance \(||X - X'||\) depends on both the set of distances \((d_1, \ldots, d_w)\) and the set of channel realizations \(\Omega\). Let \(D\) denote the set of all Euclidean distances obtained under flipping one bit in the constellation \(\Omega\). Define the set \(\Delta = \{\delta_1, \ldots, \delta_{n_d}\} \subset D\) with distinct elements from the sequence \((d_1, d_2, \ldots, d_w) \in \Delta \subset D\), i.e., the Euclidean distance \(d_k\) takes its values from the set \(\Delta\). Obviously, \(n_d = |\Delta| \leq |D|\). Let the integer \(\lambda_k\) denote the frequency of \(\delta_k\) in the sequence \((d_1, d_2, \ldots, d_w)\). If \(\sum_{k=1}^{n_d} \lambda_k = w\), and \(\Lambda = (\lambda_1, \ldots, \lambda_{n_d})\). The pairwise error probability conditioned on the channel realization set \(\Omega\) and the Hamming weight \(w\) is expressed as

\[
P_w;\Omega(c \rightarrow c') = P_w;\Omega(X \rightarrow X') = P \left( \sum_{k=1}^{w} \text{LLR}_k < 0 \right)
\]
where $LLR_k$ is the $k$th LLR, corresponding to the $k$th error position, and is equal to

$$LLR_k = \frac{|y_k - x'_k|^2 - |y_k - x_k|^2}{2N_0} \sim \mathcal{N} \left( \frac{R_k}{2N_0}, \frac{R_k}{N_0} \right)$$  \hspace{1cm} (6)$$

$R_k = \frac{||z_k - z'_k||^2}{2N_0}$ has a chi-square distribution of order $2n_r$. Averaging over $\mathcal{H}$, we calculate the characteristic function of $E_{\mathcal{H}} \{ \sum_{k=1}^w LLR_k \}$

$$\psi(j\nu) = \left( \prod_{k=1}^w \left( \frac{-j\nu \alpha_k^*}{2N_0} \right)^{n_r} \right) \left( \prod_{n=\max(0, \mu_n) \neq 0}^{n_d} \frac{\mu_n + \beta_n}{-n \lambda n} \right)$$  \hspace{1cm} (7)$$

where

$$\begin{cases} 
  n > 0, & \beta_n = \frac{1}{2} \left( 1 + \sqrt{1 + \frac{8N_0}{\gamma_k}} \right) \\
  n < 0, & \beta_n = \frac{1}{2} \left( 1 - \sqrt{1 + \frac{8N_0}{\gamma_k}} \right)
\end{cases}$$  \hspace{1cm} (8)$$

Applying a partial fraction expansion, we obtain the expression of the pairwise error probability

$$P_w(X \rightarrow X') = P_w(\Delta, \Lambda) = \prod_{k=1}^w \left( \frac{-2N_0}{d^2_k} \right)^n_r$$

$$\times \sum_{n=1}^{n_d} \sum_{n=1}^{n_r} \frac{\alpha_{n_r}}{\alpha_{n}} \left( \frac{1}{2} + \frac{1}{2} \sqrt{1 + \frac{8N_0}{\gamma_k}} \right)^{n_r}$$  \hspace{1cm} (9)$$

where the coefficients $\alpha_{n_r}$ are given by an identification of the coefficients of two series expansions in $\epsilon$ as in [23].

We compute the asymptotic expression when the noise level is low. Indeed, the coding gain and diversity are measured for high signal-to-noise ratios, where the performance has a linear asymptote on logarithmic scales

$$P_w(\Delta, \Lambda) \sim \left( \frac{2n_r w - 1}{n_r w} \right) \left( \prod_{k=1}^w \frac{2N_0}{d^2_k} \right)^{n_r}$$

$$= \left( \frac{2n_r w - 1}{n_r w} \right) \left( \frac{2N_0}{G_{\text{ergo}}(\Delta, \Lambda)} \right)^{n_r}$$  \hspace{1cm} (10)$$

with $\binom{w}{n_r} = n_r! ((w - n_r)!)$. The diversity associated with the considered pairs of Hamming weight $w$ is the exponent of $2N_0$, equal to $u n_r$. We define the coding gain or coding advantage as the coefficient dividing $2N_0$, i.e.,

$$G_{\text{ergo}}(\Delta, \Lambda) = \left( \prod_{k=1}^w d^2_k \right)^{1/w} = \left( \prod_{k=1}^w d^2_k \right)^{1/w}$$  \hspace{1cm} (11)$$

All sequences $(d_1, \ldots, d_w)$ corresponding to the same pair $(\Delta, \Lambda)$ yield the same pairwise error probability. By averaging over all possible pairs $(c, c')$ or equivalently, over all sets of distances $D_w$, we obtain $P_w = E_{D_w} \{ P_w(\Delta, \Lambda) \}$, the conditional probability that an error event of Hamming weight $w$ occurs. From this pairwise error probability, it is easy to estimate the FER or BER of the BICM with ideal interleaving thanks to a classical union bound on the weight enumeration function of the error-correcting code. Moreover, we may derive a design criterion of the BICM from the coding gain $G_{\text{ergo}}(\Delta, \Lambda)$ expression. In the following, we derive the coding gain for block-fading channels and linear precoding in order to obtain the ML design criterion of the ST-BICM.

### C. Exact Pairwise Error Probability for MIMO Block-Fading Channels Without Precoding

We assume that Definition 1 is satisfied. For a block-fading channel with $n_c$ independent realizations in a frame, the decision variable between $X$ and $X'$ is still given by (5). However, the involved channel matrices are not independent as for an ergodic channel. The conditions of independence are the following.

- If two LLR random variables depend on two different channel realizations, they are independent.
- If two LLR random variables depend on the same channel realization but on different transmit antennas, the random variables are independent.

The maximum number of independent LLR variables is $n_c n_t$, the transmit diversity order. We choose the error-correcting code so that $w \geq n_c n_t$. We now group the $w$ random variables LLR into $\min(n_c n_t, w) = n_c n_t$ independent blocks. Let $LLR_{k,l,i}$ be the $i$th LLR corresponding to the BSK transmission on the $l$th antenna of the $k$th block, $k = 1, \ldots, n_c$, $l = 1, \ldots, n_t$ and $i = 1, \ldots, \kappa_k l$, where $\kappa_k l$ is the number of bits transmitted on the $l$th antenna of the $k$th block. We have $\sum_{l=1}^{n_t} \sum_{i=1}^{\kappa_k l} = w$. Finally, $LLR$ is the sum of the $n_c n_t$ independent random variables $LLR_{k,l,i} = \sum_{i=1}^{\kappa_k l} LLR_{k,l,i}$. Let $d_{k,l,i}$ denote the distance associated with $LLR_{k,l,i}$, and define $\gamma_{k,l,i} = \sum_{i=1}^{\kappa_k l} d_{k,l,i}^2$, the distance associated with $LLR_{k,l,i}$. We have

$$LLR_{k,l,i} \sim \mathcal{N} \left( \frac{R_{k,l,i}}{2N_0}, \frac{R_{k,l,i}}{N_0} \right)$$  \hspace{1cm} (12)$$

where $R_{k,l,i} = \gamma_{k,l,i}^2 ||H_{k,l}(i)||^2$ and $H_{k,l}(i)$ is the $i$th row of $H_k$. For all $i$, $LLR_{k,l,i}$ are transmitted over the equivalent $1 \times n_r$ SIMO channel defined by $H_k(i)$, which is chi-square distributed with degree $2 n_r$. The $LLR_{k,l,i}$ variables are transmitted on independent channel states, as for the ergodic channel case, we directly apply (9) and obtain the conditional pairwise error probability closed-form expression

$$P_w(X \rightarrow X') = P_w(\Delta, \Lambda) = \prod_{k=1}^w \prod_{l=1}^{n_t} \left( \frac{-2N_0}{\gamma_{k,l}} \right)^n_r$$

$$\times \sum_{n=1}^{n_d} \sum_{n=1}^{n_r} \frac{\alpha_{n_r}}{\alpha_{n}} \left( \frac{1}{2} + \frac{1}{2} \sqrt{1 + \frac{8N_0}{\gamma_{k,l}}} \right)^{n_r}$$  \hspace{1cm} (13)$$

where $\delta_\Omega \in \Delta$ and $(\Delta, \Lambda)$ is the pair of sets representing the sequence $(\gamma_{1,1}, \ldots, \gamma_{n_c n_t,1})$. The $\alpha_{n,i}$ coefficients are computed as for (9).

The asymptotic expression of $P_w(\Delta, \Lambda)$ is

$$P_w(\Delta, \Lambda) \sim \left( \frac{2n_r n_t n_c - 1}{n_r n_t n_c} \right) \prod_{k=1}^w \prod_{l=1}^{n_t} \left( \frac{2N_0}{\gamma_{k,l}} \right)^n_r$$  \hspace{1cm} (14)$$
The diversity associated with the considered pairs of Hamming weight \( w \) is then equal to the exponent \( n_k n_r n_r \). The coding gain is given by the geometrical mean of the \( \gamma_{k,t}^2 \), and is equal to

\[
G_{df}(\Delta, \Lambda) = \left( \prod_{k=1}^{N_r} \prod_{l=1}^{N_t} \gamma_{k,t}^2 \right)^{1/n_k n_r n_r}.
\]  

We will see in the following how to use this coding gain as a design criterion for the ST-BICM optimization. We now consider an equivalent computation of the coding gain for a linearly precoded ST-BICM.

D. Exact Pairwise Error Probability for MIMO Block-Fading Channels With Precoding

When a linear precoder \( S \) of size \( N_t \times N_t \) is used, the detector computes soft outputs on the \( N_t \) transmitted symbols using the equivalent channel matrix \( S H_k \) of size \( N_t \times N_r \). The structure of \( H_k \) is described in (2). \( S H_k \) can be seen as a correlated MIMO channel [43]. Under the ideal interleaving condition, we consider at most a single erroneous bit per block of \( s \) time periods in position \( 1 \leq \ell \leq m N_t \) inside the binary mapping of the transmitted symbol \( z \), leading to symbol \( z^\ell \). For simplicity reasons, we assume that the error weight \( w \) satisfies \( w \leq N_t N_r \). Moreover, we assume that the mapping is mono-dimensional: the BSks are transmitted on a selected input of the matrix \( S H_k \). Let \( L \), \( R_{k,t}, \ell \), be the \( \ell \)th variable among \( R_{k,t} \), corresponding to the transmission of a BSk on the \( 1 \times N_t \) channel \( S H_k \), where \( S \) corresponds to the row of \( S \). We have \( \sum_{k=1}^{N_R} \sum_{l=1}^{N_t} \gamma_{k,l} = w \). Let \( d_{k,t} \) denote the BSK distance associated with \( LLR_{k,t} \). We can use the factorization \( LLR_{k,t} = \sum_{i=1}^{N_t} LLR_{k,t,i} \) of all the LLR variables filtered with \( S H_k \)

\[
LLR_{k,t} \sim \mathcal{N}\left( \frac{R_{k,t}}{2 N_0}, \frac{R_{k,t}}{N_0} \right)
\]  

where \( R_{k,t} = |V_{k,t} H_k|^2 \), \( V_{k,t} \), and \( \gamma_{k,t} = \sum_{i=1}^{N_t} R_{k,t,i} \). The variable \( R_{k,t} \) is a generalized chi-square random variable with \( 2 N_t \) correlated centered Gaussian components. The random variable \( LLR_{k} = \sum_{t=1}^{N_t} LLR_{k,t} \) satisfies

\[
LLR_{k} \sim \mathcal{N}\left( \frac{\sum_{t=1}^{N_t} R_{k,t}}{2 N_0}, \frac{\sum_{t=1}^{N_t} R_{k,t}}{N_0} \right)
\]  

From Appendix A, we get the following characteristic function:

\[
E_{H_k}[\Psi_{LLR_k}(j\nu)] = \prod_{l=1}^{N_t} \prod_{u=1}^{N_r} \left( 1 - \frac{\nu (j - \nu)}{2 N_0} \phi_{k,u} \right)^{-n_r}
\]  

where \( \phi_{k,u} \) is the \( u \)th eigenvalue of

\[
\Sigma_{k}^{[\ell]} = \sum_{l=1}^{N_t} \gamma_{k,t}^2 S_l^{[\ell]s} S_l^{[\ell]s} = \sum_{l=1}^{N_t} \gamma_{k,t}^2 \sum_{s=1}^{s/n_r} S_l^{[\ell]s} S_l^{[\ell]s} = M_k^{[\ell]s} M_k^{[\ell]s}
\]  

\( M_k^{[\ell]s} \) is as an \( n_t \times n_t \) Hermitian square root matrix of \( \Sigma_{k}^{[\ell]} \) and row vectors \( S_l^{[\ell]s} \) of size \( n_t \) and \( s/n_r \times n_t \) matrices \( S_l^{[\ell]s} \) are defined from \( S \) as shown in (20) at the bottom of the page, and

\[
S_l^{[\ell]s} = \begin{bmatrix}
S_l^{[\ell]s} [1 \times N_t/n_r] \\
S_l^{[\ell]s} [2 \times N_t/n_r] \\
\vdots \\
S_l^{[\ell]s} [N_t \times N_t]
\end{bmatrix}
\]

\( N_t/s = n_t \) coefficients

\[
\psi_{\Delta, \Lambda}(j \nu) = \prod_{l=1}^{N_t} \prod_{u=1}^{N_r} \left( 1 - \frac{\nu (j - \nu)}{2 N_0} \phi_{k,u} \right)^{-n_r}
\]  

The set of eigenvalues \( \phi_{k,u} \) is the function of the precoding matrix \( \psi_{\Delta, \Lambda}(j \nu) \) and the BSK distances set \( D^w \). Thanks to the independence of channel realizations for different \( k \) values, we can multiply the characteristic functions

\[
\psi_{\Delta, \Lambda}(j \nu) = \psi_{\Delta, \Lambda}(j \nu) \prod_{k=1}^{N_k} \psi_{\Delta, \Lambda}(j \nu) \prod_{l=1}^{N_t} \prod_{u=1}^{N_r} \left( 1 - \frac{\nu (j - \nu)}{2 N_0} \phi_{k,u} \right)^{-n_r}
\]  

We denote \( \Delta = \{ \delta_k \} \) the set of \( n_k \) square roots of nonnull eigenvalues extracted from the sequence defined by the \( \phi_{k,u} \) values. Each eigenvalue \( \phi_{k,u} \) is repeated \( \lambda_{\nu,v} \) times. We observe that \( n_k \leq n_t \). Finally, using the partial fraction expansion of \( \psi_{\Delta, \Lambda}(j \nu) \), we obtain the exact pairwise error probability \( P_{\psi}(\Delta, \Lambda) \) conditioned on \( d_{k,t} = c \) as:

\[
P_{\psi}(\Delta, \Lambda) = \psi_{\Delta, \Lambda}(j \nu) \prod_{k=1}^{N_k} \psi_{\Delta, \Lambda}(j \nu) \prod_{l=1}^{N_t} \prod_{u=1}^{N_r} \left( 1 - \frac{\nu (j - \nu)}{2 N_0} \phi_{k,u} \right)^{-n_r}
\]  

The asymptotic expression of \( P_{\psi}(\Delta, \Lambda) \) is

\[
P_{\psi}(\Delta, \Lambda) \sim \prod_{k=1}^{N_k} \prod_{l=1}^{N_t} \prod_{u=1}^{N_r} \left( 1 - \frac{\nu (j - \nu)}{2 N_0} \phi_{k,u} \right)^{-n_r}
\]  

where \( N_8 = \sum_{k=1}^{N_k} \lambda_{\nu,v} \) is the total number of nonnull eigenvalues.
The diversity associated with the considered pairs of Hamming weight $w$ is the exponent equal to $\sum_{i=1}^{n_t} \lambda_i n_r = N_h n_r$. The coding gain is given by

$$G_{s,n_r}(\Delta, \Lambda) = \left( \prod_{i=1}^{n_t} \left( \frac{2^{2w}}{2} \right)^{\frac{1}{N_h n_r}} \right) G_{s,n_r}(\Delta, \Lambda). \quad (25)$$

We have derived for any signal-to-noise ratio an exact expression of the pairwise error probabilities of a BICM with linear precoding, which is useful for a tight BER and FER estimation. The asymptotic expression leads to the well-known rank and determinant criteria [19], [40] for space–time code optimization over MIMO block-fading channels, where the considered space–time code is the whole BICM structure. As a remark, the asymptotic design criterion is usually derived by first upper-bounding the $Q(x)$ function by $\exp(-x^2/2)/2$ and then averaging over the channel realizations. The obtained asymptotic expression has a multiplicative coefficient different from $\left( \frac{2^{2n_r} n_r}{n_r N_h n_r} \right)$, which is inexact but provides the same design criterion.

Moreover, we notice that applying the Tarokh criterion [40] on the rank and determinant to the precoder alone does not lead to the whole BICM optimization. Quasi-optimal linear precoders will be designed to achieve full diversity and approach optimal coding gain in Section V.

E. Evaluation of the FER

For ergodic channels, the FER is easily computed via a union bound. Indeed, only error events with minimum Hamming distance impact the error rate for a high signal-to-noise ratio and the observed diversity is equal to $n_r d_{\text{min}}$. For block-fading channels, the FER computation is much more tricky since each pairwise error probability is supposed to have the full-diversity order $n_r n_t n_r n_t$. Due to the random nature of each eigenvalue in (25), it is difficult to know the impact of each distance configuration on the final FER.

However, one may assume that for a sufficiently high signal-to-noise ratio, the FER satisfies the following expression:

$$\text{FER} \approx \sum_{w} A_w E_{(\Delta, \Lambda)}[P_w(\Delta, \Lambda)] \quad (26)$$

where $A_w$ is weighting the impact of pairwise error probabilities with Hamming weight $w$ in the global error probability and the expectation on $(\Delta, \Lambda)$ is allowed by the interleaver random structure. Let us define $G$ the global coding gain. Since each pairwise error probability is supposed to have full diversity, we write

$$\text{FER} \approx \left( \frac{2^{n_r n_t n_r n_t - 1}}{n_r n_t n_r n_t} \right) \left( \frac{2N_h}{G} \right)^{n_r n_t n_r n_t} \quad (27)$$

and

$$G^{n_r n_t n_r n_t} = \sum_{w} A_w E_{(\Delta, \Lambda)}[G(\Delta, \Lambda)^{-n_r n_t n_r n_t}] \quad (28)$$

where $G(\Delta, \Lambda)$ is the coding gain associated with one pair of codewords. We note that optimizing independently all pairwise error probabilities, which will be done in the following, enhances the global performance. Moreover, we observe that the number of receive antennas does not affect the coding gain of a single pairwise error probability. The effect of the receive diversity appears in the expression of the global coding gain (28). As $n_r n_t n_r n_t$ grows, the smallest coding gains have more impact on the final performance. Asymptotically, if $n_r n_t n_r n_t \to +\infty$, only the nearest neighbors in the Euclidean code have an influence on the FER, as for AWGN channels.

We will see in Section V-A that the best coding gain is achieved when all eigenvalues $\lambda_k$ are equal. In this ideal configuration, the coding gain is shown to be the same as with the same coded modulation transmitted on a $1 \times n_r n_t n_r n_t$ quasi-static SIMO channel. Simulating this latter case is less complex: the performance curve is semi-analytically computed using a reference curve on an AWGN channel. Alternatively, performance may be obtained by computing the tangential sphere bound for spherical modulations [27]. In the following, ideal BICM will refer to the performance of the ideal configuration, which will be drawn on simulation results. This lower bound has the advantage to take the modulation and error-correcting code into account and will be useful to evaluate the optimality of both the linear precoder and the channel interleaver.

IV. THE SINGLETON BOUND WITH LINEAR PRECODER

Definition 1 ensures that any pair of codewords benefits from a full diversity order. In this section, we derive a condition on the existence of a practical interleaver that could achieve the conditions of Definition 1. Let us first make the following assumption.

Assumption 1: The detector perfectly converts the $N_t \times N_c$ correlated MIMO $N_c$-block-fading channel $S H_k$ with QAM input into a $1 \times n_r n_t n_r n_t$ SIMO channel with $n_r n_t n_r n_t$-block-fading channel with BSK input, assuming that $s$ is a divisor of $n_r n_t n_r n_t$.

We will present in Section V linear precoders that satisfy Assumption 1. Under this condition, the detector collects an amount of diversity equal to $sn_r n_t n_r n_t$. The full diversity $n_r n_t n_r n_t$ is collected by the detector when $s = n_r n_t n_r n_t$, unfortunately, the APP signal detection has an exponential complexity in $s$. On the other hand, the BICM channel decoder is also capable of collecting a large amount of diversity, but the latter is still limited by the Singleton bound [29], [30], [34]. Hence, the lowest complexity solution that reaches full diversity is to draw advantage of the whole channel code diversity and recover the remaining diversity by linear precoding. The best way to choose the spreading factor $s$ is given by the Singleton bound described hereafter.

The studied ST-BICM is a serial concatenation of a rate $R_c$ binary convolutional code $C$, an interleaver of size $N_c L_c$ bits, and a QAM mapper followed by the precoder as described in Section II. When $S$ is the identity matrix, the ST-BICM diversity order is upper-bounded by [30]

$$Y \leq n_r \left( \left[ n_r n_t (1 - R_c) \right] + 1 \right). \quad (29)$$

The maximal diversity given by the outage limit under a finite size QAM alphabet also achieves the above Singleton bound [25]. With a vanishing coding rate, i.e., $R_c \leq 1/(n_r n_t)$, it is possible to
attain the overall system diversity order \( n_T n_c n_T \) produced by the receive antennas, the transmit antennas, and the distinct channel states. Unfortunately, this is unacceptable due to the vanishing transmitted information rate. Precoding is one means to achieve maximum diversity with a nonvanishing coding rate.

The integer \( N_H = n_T n_T / s \) is the best diversity multiplication factor to be collected by \( C \). The length of a \( C \) codeword is \( L_C N_C \) binary elements. Let us group \( L_C N_C / N_H \) bits into one nonbinary symbol creating a nonbinary code \( C' \). Now, \( C' \) is a length-\( N_H \) code built on an alphabet of size \( 2^{L_C N_C / N_H} \). The Singleton bound on the minimum Hamming distance of the nonbinary \( C' \) becomes \( D_H \leq N_H - \lfloor N_H R_C \rfloor + 1 \). Multiplying the previous inequality with the Nakagami law order \( sn_T \) yields the maximum achievable diversity order after decoding [22]

\[
\Upsilon \leq sn_T \left[ \frac{n_T n_T}{s} (1 - R_C) + 1 \right].
\]

Finally, since \( \Upsilon \) is upper-bounded by the channel intrinsic diversity and the minimum Hamming distance \( d_{H_{\text{min}}} \) of the binary code, we can write

\[
\Upsilon \leq \min \left( \text{sn_T} \left[ \frac{n_T n_T}{s} (1 - R_C) + 1 \right] ; n_T n_c n_T ; s n_T d_{H_{\text{min}}} \right) = \Upsilon_{\text{max}}
\]

(31)

If \( d_{H_{\text{min}}} \) is not a limiting factor (we choose \( C \) accordingly), we can select the value of \( s \) that leads to a modified Singleton bound greater than or equal to \( n_T n_T n_T \).

**Proposition 1:** Considering a BICM with a rate \( R_C \) binary error-correcting code on an \( n_T \times n_T \) MIMO channel with \( n_c \) distinct channel states per codeword, the spreading factor \( s \) of a linear precoder must be a divisor of \( n_T n_T n_T \) and must satisfy \( s \geq R_C n_T n_T n_T \) in order to achieve the full diversity \( n_T n_T n_T \) for any pair of codewords. In this case, the ideal interleaving conditions can be achieved with an optimized interleaver.

The smallest integer \( s_{\text{opt}} \) satisfying the above proposition minimizes the detector’s complexity. If \( R_C > 1/2 \), then \( s_{\text{opt}} = n_T n_T n_T \) which involves the highest complexity. If \( R_C \leq 1/(n_T n_T) \), linear precoding is not required.

Tables I and II show the diversity order derived from the Singleton bound versus \( s \) and \( n_T \) for \( n_c = 1 \) and \( n_c = 2 \), respectively. The values in bold indicate full diversity configurations. For example, in Table I, for \( n_T = 4 \), \( s = 2 \) is a better choice than \( s = 4 \) since it leads to an identical diversity order with a lower complexity.

### V. Linear Precoder Optimization

Many studies have been published on space–time spreading matrices introducing some redundancy, a technique known as space–time block coding. On one hand, some of them are decoded by a low-complexity ML decoder, but they sacrifice transmission data rate for the sake of high performance. Among them, the Alamouti scheme [1] is the most famous, but is only optimal for a \( 2 \times 1 \) MIMO channel. The other designs allowing for low-ML decoding complexity are based on an extension of the Alamouti principle (e.g., double space–time transmit diversity [42]) but also sacrifice the data rate. On the other hand, full rate space–time codes have recently been proposed [4], [11]–[14], [18], [35]. However, their optimization does not take into account their concatenation with an error-correcting code. In this section, we describe a near-ideal solution for linear precoding in BICMs under iterative decoding process. Our strategy is to separate the coding step and the geometry properties in order to express some criteria allowing for the construction of a space–time spreading matrix for given channel parameters \( n_T \), \( n_r \), and \( n_c \). The inclusion of rotations to enhance the BICM performance over single-antenna channels has been proposed in [32]. Our solution uses this concept for designing a space–time code including a powerful error-correcting code.

When the channel is quasi-static or block-fading with parameter \( n_c \), the diversity is upper-bounded by \( n_T n_T n_T \) which may be more limiting than \( n_T d_{H_{\text{min}}} \) (e.g., \( n_T = 2 \), \( n_r = 1 \), \( n_c = 1 \)). We introduce a new design criterion of space–time spreading matrices that guarantees a diversity proportional to the spreading factor, within the upper bound, and a maximal coding gain at the last iteration of an iterative joint detection and decoding.

#### A. Coding Gain Under Both Ideal Interleaving and Precoding

First, we look for the best achievable coding gain for fixed parameters \( n_T \), \( n_T \), \( n_c \), \( R_C \), and the appropriate way to choose the error-correcting code, the binary mapping, the linear precoder and its parameters \( s \) and \( n_T \) to achieve the ideal coding gain.

We want to achieve full diversity under ML decoding or iterative joint detection and decoding, this induces that there are \( n_T n_T \) nonnull eigenvalues \( \psi_{k, u}^{[1]} \) (see (24))

\[
G_{s,n_T} (\Delta, \Lambda) = \left( \prod_{k=1}^{n_T} \prod_{n=1}^{n_T} \prod_{u=1}^{n_T} \psi_{k,u}^{[1]} \right)^{1/(n_T n_T)}.
\]

### Table I

| \( n_T \times s \) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------------------|---|---|---|---|---|---|---|---|
| 1                 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

### Table II

| \( n_T \times s \) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------------------|---|---|---|---|---|---|---|---|
| 1                 | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 5 |
| 2                 | 3 | 4 | 4 | 5 | 6 | 6 | 10 | 10 |
| 3                 | 4 | 4 | 6 | 6 | 8 | 8 | 12 | 12 |
| 4                 | 5 | 6 | 6 | 8 | 8 | 14 | 16 | 16 |
Furthermore, we want to maximize the $G_{s, n_c} (\Delta, \Lambda) \expres-
ion$. Assuming that each row $S_l$ is normalized to 1, we get

$$G_{s, n_c} (\Delta, \Lambda) = \sum_{k=1}^{N_c} \sum_{l=1}^{N_l} \frac{\gamma_{k,l}^2}{n_t n_c} = \sum_{k=1}^{N_c} \sum_{l=1}^{N_l} \sum_{u=1}^{r_u} d_{k,l,u}^2 .$$  \hspace{1cm} (33)

Under this condition, the ideal coding gain is achieved when all eigenvalues are equal

$$\gamma_{k,l}^2 = \sum_{k'=1}^{N_c} \sum_{l'=1}^{N_l} \frac{\gamma_{k',l'}^2}{n_t n_c} \forall (k, l, u)$$  \hspace{1cm} (34)

which leads to

$$G_{\text{ideal}} (\Delta, \Lambda) = \sum_{k=1}^{N_c} \sum_{l=1}^{N_l} \frac{\gamma_{k,l}^2}{n_t n_c} = \frac{w}{n_t n_c} .$$  \hspace{1cm} (35)

The exact pairwise error probability expression simplifies to the classical expression of the performance of a binary phase-shift keying (BPSK) with distance $\sum_{j=1}^{w} d_j^2$ over a diversity channel with order $n_t n_c n_r$. [37]

$$P_{w, \text{ideal}} (\Delta, \Lambda) = \left( 1 - \frac{1}{\sum_{j=1}^{w} d_j^2} \right)^{n_r n_t n_r} \frac{n_r n_t n_r}{2} \times \left( 1 + \frac{1}{\sum_{j=1}^{w} d_j^2} \right)^{-1} .$$  \hspace{1cm} (36)

As stated in the Introduction, in an ST-BICM, precoded modulation symbols quantify the Shannon sphere and the best quantization is obtained by uniformly distributing them on the sphere. After transmission on a fading channel, vectors belong to an ellipsoid obtained by applying an homothety on the sphere. From (34) and (35), we see that the ideal coding gain is obtained by equally distributing the Euclidean distance between two codewords among the $n_t n_c$ channel states. Hence, the Euclidean distance varies as an $n_t n_c n_r$ Nakagami distribution, according to the square norm of the ellipsoid axes.

Thus, an ideal ST-BICM aims at uniformly distributing the precoded modulation symbols, whatever the channel realization, i.e., whatever the homothety. The ideal coding gain is a fundamental limit which cannot be outperformed. It is useful to evaluate how optimal the practical design of a BICM is. We aim at finding the best design, corresponding to eigenvalues which are as close to each other as possible. The more different from each other the eigenvalues are, the lower the product in (32) and the coding gain are. From (35), we see that the ideal coding gain is the same as for the same coded modulation transmitted on a $1 \times n_t n_c n_r$ SIMO channel, applying the appropriate $E_b/N_0$ normalization.

Without linear precoding, the ideal coding gain is only achieved if all $\gamma_{k,l}$ are equal. Remember that each $\gamma_{k,l}$ is a sum of $n_t n_c$ distances $d_{k,l,u}$. Thanks to the second point in Definition 1, the $n_t n_c$ values are close to $w/(n_t n_c)$ and their variance decreases when $w$ increases. Thus, with a powerful error-correcting code having minimum Hamming distance much greater than $n_t n_c$ and $|D|$, each $\gamma_{k,l}$ value is almost equal to the average $\sum_{k=1}^{N_c} \sum_{l=1}^{N_l} \sum_{u=1}^{r_u} d_{k,l,u}^2 / (n_t n_c)$ of $d_{k,l,u}$ values and quasi-ideal coding gain is observed.

If the error-correcting code is not powerful enough to achieve the ideal coding gain, i.e., the $\gamma_{k,l}$ values are very different, the linear precoder provides an additional coding gain by averaging the $\gamma_{k,l}$ values, as we will see in the following. First, we derive the optimal coding gain which can be achieved using an ideal linear precoder for a given binary labeling and error-correcting code. Variables $\gamma_{k,l}$ for different $k$ values correspond to independent channel realizations $H_k$ which are not linked by the linear precoder. Thus, random variables $\sum_{k=1}^{N_c} \sum_{l=1}^{N_l} \gamma_{k,l}$ are independent for distinct values of $k$. The optimal coding gain with linear precoding is

$$G_{s, n_c, \text{opt}} (\Delta, \Lambda) = \prod_{k=1}^{N_c} \left( \sum_{l=1}^{N_l} \frac{\gamma_{k,l}^2}{n_t n_c} \right) .$$  \hspace{1cm} (37)

Equation (37) means that an optimal linear precoder is capable of making eigenvalues equal for a same $k$. However, for different values of $k$, eigenvalues $\gamma_{k,l}$ are different, which induces a coding gain loss. When the mapping and error-correcting code are given and the interleaving is ideal, the choice of linear precoding parameters impacts on optimal coding gain. Let us consider codewords that are equidistant from the transmitted codeword, i.e., a set of distance configurations corresponding to a same value of $\sum_{k=1}^{N_c} \gamma_{k,l}^2$. The variance of $\sum_{k=1}^{N_c} \gamma_{k,l}^2 / (n_t n_c)$ over this set decreases when $n_c$ increases, as the number of distances building each $\gamma_{k,l}$ is higher. The lower the variance of eigenvalues, the higher the coding gain. Thus, $G_{s, n_c, \text{opt}} (\Delta, \Lambda)$ is an increasing function of $n_c$ and, for a given $s$, we should choose $n_c = \min (s, n_r)$. The optimal coding gain $G_{s, n_r, \text{opt}} (\Delta, \Lambda)$ is an increasing function of $s$. If $n_r = n_c$, which implies $s = n_t n_r$, the ideal coding gain is achieved by the optimal precoder. Finally, we can surround the coding gain at full diversity as follows:

$$\forall s \quad G_{s, n_r, \text{opt}} (\Delta, \Lambda) \geq G_{s, n_r, \text{opt}} (\Delta, \Lambda) \geq G_{s, \text{ideal}} (\Delta, \Lambda) \geq G_{s, \text{ideal}} (\Delta, \Lambda) .$$  \hspace{1cm} (38)

If, for any pairwise error probability, $G_{s, \text{opt}} (\Delta, \Lambda) \geq G_{\text{ideal}} (\Delta, \Lambda)$, the linear precoder optimization is useless from a coding gain point-of-view. However, obtaining near-ideal coding gain without precoding requires an optimization of the error-correcting code and mapping for any pairwise error probability, which is intractable for nontrivial modulations and codes. Furthermore, the first objective of linear precoding is the diversity control, which has a high influence on the performance even at medium FER (10$^{-2}$ $\sim$ 10$^{-3}$), especially for low-diversity orders. Therefore, precoding is often useful in the BICM structure.

After the impact of the linear precoding for a given pairwise error probability, let us consider the behavior of the global performance under linear precoding. As stated in Section III-E, if $n_t n_c n_r$ grows, the pairs of codewords providing the smallest coding gains have more impact on the final performance. Since the linear precoder provides a more substantial gain for the low
Hamming weight configurations, the coding gain of the linear precoder will be magnified as the diversity grows.

1) Example of Ideal Coding Gain: In order to illustrate the role of the linear precoding in the coding gain optimization, we consider a $2 \times 1$ quasi-static MIMO channel and a pairwise error probability between two codewords separated by a Hamming distance of $w$ bits. Fig. 3 represents the distribution of the two $\gamma_1$ and $\gamma_2$ values over the two transmit antennas without linear precoding. Bits transmitted on antennas 1 and 2 are transmitted on the sets of time periods $T_1$ and $T_2$, respectively. Thanks to ideal interleaving, $T_1 \cap T_2 = \emptyset$. This illustrates the factorization of the distances into the $\gamma$ values. The instantaneous coding gain is equal to $\sqrt{\frac{\gamma_1^2}{2}}$. Now let us consider a specific linear precoder, which spreads the values $\gamma_1$ and $\gamma_2$ as presented in Fig. 4 over two time periods and two transmit antennas dividing the squared distance in two equal parts $\gamma_1^2/2$ and $\gamma_2^2/2$, respectively. The average value $(\gamma_1^2 + \gamma_2^2)/2$ is transmitted on each antenna, the coding gain is optimal and equal to $(\gamma_1^2 + \gamma_2^2)/2$. For example, consider a BPSK modulation and a pairwise error probability with Hamming weight 3. With optimal linear precoding, the ideal interleaving provides for example $\gamma_1^2 = 2 \times 2^2$ and $\gamma_2^2 = 1 \times 2^2$. With optimal linear precoding, we have a distance $(2 \times 2^2 + 1 \times 2^2)/2$ associated with each antenna. The ratio between the two averaged coding gains is equal to $\sqrt{9}/8$, i.e., we expect a gain of 0.26 dB when using linear precoding. With $w = 5$ and $w = 11$, the coding gain becomes $10 \log_{10}(\sqrt{24}/25) \approx 0.09$ dB and $10 \log_{10}(\sqrt{120}/121) \approx 0.02$ dB, respectively. The higher the Hamming weight involved in the pairwise error probability is, the less the coding gain provided by linear precoding is.

We see in Table III the best gain to be provided by linear precoding for a quasi-static channel with BPSK input with respect to a full diversity unprecoded scheme. These gains are particularly low because the error-correcting code aims at recovering a large amount of coding gain. This illustrates that BICMs are very efficient transmission schemes. As a remark, if a modulation with higher spectral efficiency is used with Gray mapping, the nearest neighbor in the Euclidean code has the same distance configuration as if a BPSK modulation was used. Moreover, for high-diversity orders, the global error rate for high $E_b/N_0$ will be dominated by the neighbors and the gain provided by linear precoding will be very close to the ones shown in Table III. However, if the diversity is low, the gains provided by linear precoding may be much more important. Let us assume that a 16-QAM modulation with Gray mapping is transmitted on a $n_t = 2$ quasi-static channel. For instance, if $w = 5$, there exists a neighbor with distance configuration $(2A,3A,3A,A,A)$ (e.g., see [23]), and $\gamma_1^2 = 9A^2 + 9A^2 + 9A^2 = A^2 + A^2$. The gain to be provided by linear precoding is equal to $10 \log_{10}(29)/2/\sqrt{54} = 2.95$ dB. As already stated, the final coding gain is equal to a weighted sum of all coding gains, where the weighting coefficients cannot be easily computed in the case of low-diversity orders.

Even if linear precoding does not always provide a substantial coding gain, its prior aim is the diversity order control. Thus, we will focus on the design of linear precoders aiming at reaching full diversity and maximizing the coding gain for any set of parameters $(n_t, s, n_s)$.

B. A New Class of Linear Precoders

Under linear precoding, the optimal coding gain is achieved if all $\varphi_{k,a}^{[l]}$ variables are equal for a same $k$. Let us first consider

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
& $w = 2$ & $w = 3$ & $w = 4$ & $w = 5$ & $w = 6$ & $w = 7$ & $w = 8$ \\
\hline
$n_t = 2$ & 0.00 & 0.26 & 0.00 & 0.09 & 0.00 & 0.05 & 0.00 \\
$n_t = 3$ & / & 0.00 & 0.25 & 0.21 & 0.00 & 0.08 & 0.08 \\
$n_t = 4$ & / & / & 0.00 & 0.22 & 0.26 & 0.17 & 0.00 \\
$n_t = 5$ & / & / & / & 0.00 & 0.19 & 0.26 & 0.24 \\
$n_t = 6$ & / & / & / & / & 0.00 & 0.17 & 0.25 \\
$n_t = 7$ & / & / & / & / & / & 0.00 & 0.15 \\
$n_t = 8$ & / & / & / & / & / & / & 0.00 \\
\hline
\end{tabular}
\caption{Best Gain in Decibels to be Provided by Linear Precoding With Respect to an Unprecoded System, With Ideal Interleaving and for a Given Pair of Codewords With Hamming Distance $w$ and BPSK Input}
\end{table}
the eigenvalues associated with the independent realizations in the spreading matrix, indexed by $t$. First, two matrices $M_{k}^{[s]}$ and $M_{k}^{[s']}$, as introduced in (19), should have the same eigenvalues, which is satisfied if \( \forall (t_1, t_2), M_{k}^{[s]} = R^{t_1, t_2} M_{k}^{[s]} R^{t_1, t_2} \), where $R^{t_1, t_2}$ is a unitary matrix, for example a rotation. Hence, \( \forall (t_1, t_2), S_{l}^{[s]} = S_{l}^{[s']} R^{t_1, t_2} \). The precoding subpart $S_{l}^{[s]}$, with spreading factor $s' = s/n_k$, experiences a quasi-static channel. We assume that $s'$ is an integer, divisor of $n_k$. It is sufficient to design the first part of the precoder matrix rows for a quasi-static channel and rotate it to compute the other subparts. Furthermore, any choice of $R^{t_1, t_2}$ leads to the same performance because eigenvalues remain unchanged. The condition simplifies to $\|S_{l}^{[s]}\| = \|S_{l}^{[s']}\|$.

Let us now optimize for a given index $t$ the equivalent precoder over the quasi-static channel $\text{diag} \left( H_{k}^{[s]}, \ldots, H_{k}^{[s']} \right)$, in which $H_{k}^{[t]}$ is repeated $s'$ times. If all eigenvalues of $M_{k}^{[s]} M_{k}^{[s']}$ are equal, $M_{k}^{[s]}$ and $M_{k}^{[s']}$ are weighted unitary matrices and

\[
M_{k}^{[s]} M_{k}^{[s']} = M_{k}^{[s]} M_{k}^{[s']} = \sum_{l=1}^{N_k} \gamma_{k,l}^{2} \sum_{i=1}^{s'} S_{l}^{[i]} S_{l}^{[i]}.
\]

Matrix $S_{l}^{[i]} S_{l}^{[i]}$ has rank one and matrix $\sum_{i=1}^{s'} S_{l}^{[i]} S_{l}^{[i]}$ has maximum rank $s'$. If $s' < n_k$, it can be shown that it is impossible to get all eigenvalues equal to $\sum_{l=1}^{N_k} \gamma_{k,l}^{2}/n_k n_{\text{c}}$ as required in order to achieve the optimal coding gain. However, in order to insure that $M_{k}^{[s]} M_{k}^{[s']}$ has a rank $n_k$ and that the eigenvalues are as equal as possible, we group $s'$ together and associate them with one of the $s'$ groups of $s'$ eigenvalues: we denote $S_{k_{l_2}, 1}$ the $l_2$th part of size $s'$ of $S_{l_2, 1}$ and $\{l_2, 1\}$ the index of the $(l_2 - 1)s' + 1$th row of $S$, where $l_2 \in [1, n_k/s']$ and $l_1 \in [1, s']$. Let us assume that $S_{l_2, 1}$ has only one non-null subpart in position $l_2$, i.e.,

\[
\forall j \neq l_2, S_{l_2, 1}^{[j]} = [0, \ldots, 0].
\]

Considering such a structure is equivalent to considering separate precoding on $n_{k}/s'$ distinct groups of $s'$ transmit antennas. We have

\[
\sum_{l=1}^{N_k} \gamma_{k,l}^{2} S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i]} = \sum_{l_2=1}^{N_k} \sum_{l_1=1}^{s'} \gamma_{k_{l_2}, l_1}^{2} S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i]}
\]

\[
= \sum_{l_2=1}^{N_k} \sum_{l_1=1}^{s'} \gamma_{k_{l_2}, l_1}^{2} \mathcal{D}_{l_2} \left( \sum_{i=1}^{s'} S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i]} \right)
\]

where $\mathcal{D}_{l_2} (A)$ is a block-diagonal matrix with only one non-null block $A$ in position $l_2$. We choose $S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i]}$ proportional to the $l_2$th row of a $s' \times s'$ unitary matrix, such as $\|S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i]}\|^2 = 1/s$

\[
\sum_{l_2=1}^{N_k} \gamma_{k_{l_2}, l_1}^{2} S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i]} = \frac{n_{k}}{s'} \mathcal{D}_{l_2} \left( \frac{1}{s} I_{s'} \right)
\]

\[
= \sum_{l_2=1}^{N_k} \sum_{l_1=1}^{s'} \gamma_{k_{l_2}, l_1}^{2} \mathcal{D}_{l_2} \left( \frac{1}{s} I_{s'} \right)
\]

\[
= \frac{1}{s} \sum_{n_{k} l_1=1}^{s'} \mathcal{D}_{l_2} \left( \frac{1}{s} \right)
\]

which leads to

\[
l_2 \leq n_{k}/s', u \leq d', \quad S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i]} = \frac{1}{s} \sum_{l_1=1}^{s'} \gamma_{k_{l_2}, l_1}^{2} I_{s'}.
\]

The random variables $\gamma_{k_{l_2}, l_1}^{2}$ are independent and identically distributed for different values of $l_1$ and $l_2$, the coding gain is

\[
\mathcal{G}_{s, n_{k}}(\Delta, \Lambda) = \prod_{k_{l_2}, l_1=1}^{N_k} \prod_{s' l_1=1}^{s'} \left( \sum_{l_1=1}^{s'} \gamma_{k_{l_2}, l_1}^{2} \right)^{s'/s}.
\]

For any value of $n_{\text{c}}$, the gain expressed in (46) is a geometric mean of order $n_{\text{c}}/s$. For a given realization $\{d_1, \ldots, d_{n_{\text{c}}}\}$, given $s$ and for any $n_{\text{c}}$

\[
\sum_{l_2=1}^{N_k} \sum_{l_1=1}^{s'} \gamma_{k_{l_2}, l_1}^{2} I_{s'}
\]

and thus, $\sum_{l_1=1}^{s'} \gamma_{k_{l_2}, l_1}^{2} I_{s'}$ are constant, ensuring the same coding gain. However, such a precoder does not achieve the optimal coding gain for any value of $s'$. The summation is made over $s'$ different values whereas the optimal coding gain in (37) necessitates a summation over $s n_{\text{c}}$ values. Only if $s' = n_{k}$, the complete spatial transmit diversity is collected by the detector and the optimal coding gain is achieved.

**Proposition 2: Dispersive Nucleo Algebraic (DNA) Precoder** Let $S$ be the $N_k \times N_k$ precoding matrix of a BICM over a $n_\text{c} \times n_\text{c}$ MIMO $n_\text{c}$-block-fading channel. Let us assume that $S$ precodes a channel block-diagonal matrix with $s$ blocks and $n_\text{c}$ channel realizations. We denote $s$ the spreading factor, $n_\text{c}$ is the spreading factor, $s' = s/n_\text{c}$. Let $S_{l_2, 1}$ be the $l_2$th part of size $N_k/n_\text{c}$ of the $l_2$th row of $S$, $S_{l_2, 1}^{[i]}$ be the $l_2$th part of size $n_\text{c}$ of $S_{l_2, 1}$, and $S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i]}$ be the $l_2$th part of size $s'$ of $S_{l_2, 1}^{[i]}$. The subpart $S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i]}$ is called nucleotide. The linear precoder guarantees full diversity and quasi-optimal coding gain at the decoder output under ML decoding of the BICM if it satisfies the two conditions of null nucleotides and orthogonal nucleotides for all $t \in [1, n_\text{c}]$, $i \in [1, s']$, $l_1 \in [1, s's']$, $l_2 \in [1, n_\text{c}/s']$, and $\{l_2, l_1\} = (l_2 - 1)s' + 1$ as shown in (47) at the bottom of the page, where $\mathcal{D}(0) = 1$ and $\mathcal{D}(x \neq 0) = 0$.

\[
\left\{\begin{array}{l}
\forall j \neq l_2, j \in [1, n_{k}/s'], \quad S_{l_2, 1}^{[j]} S_{l_2, 1}^{[j]} = 0_{1 \times s'} \quad \text{Null Nucleotide condition} \\
\forall i' \neq i, i' \in [1, s'], \quad S_{l_2, 1}^{[i]} S_{l_2, 1}^{[i']} = \frac{1}{s} \mathcal{D}(i - i') \quad \text{Orthogonal Nucleotide condition}
\end{array}\right.
\]
Let us take for example \( n_t = 4, \ n_s = 1, \) and \( s = 2. \) A DNA matrix would have the following structure:

\[
\text{DNA}(n_t = 4, \ n_s = 1, \ s = 2) = \begin{bmatrix}
S^{[1]}_{[1][1]} & 0 & S^{[1]}_{[2][1]} & 0 \\
S^{[1]}_{[2][1]} & 0 & S^{[1]}_{[2][2]} & 0 \\
S^{[1]}_{[3][1]} & 0 & S^{[1]}_{[2][1]} & 0 \\
S^{[1]}_{[4][1]} & 0 & S^{[1]}_{[2][2]} & 0 \\
0 & S^{[2]}_{[1][2]} & 0 & S^{[2]}_{[2][2]} \\
0 & S^{[2]}_{[2][2]} & 0 & S^{[2]}_{[2][2]} \\
0 & S^{[2]}_{[3][2]} & 0 & S^{[2]}_{[2][3]} \\
0 & S^{[2]}_{[4][2]} & 0 & S^{[2]}_{[2][4]}
\end{bmatrix}.
\tag{48}
\]

Now, let us consider a linear precoder matrix \( \mathbf{S} \) that satisfies Proposition 2. We build a \( s' \times N_r \) matrix \( \mathbf{H}^{(i)}_k \) from the rows of \( \mathbf{H}_k \) corresponding to the \( i \)th group of \( s' \) transmit antennas. Elements of \( \mathbf{H}^{(i)}_k \) are defined as follows:

\[
\forall \ i = 0 \ldots n_t / s', \ \forall \ j = 0 \ldots s' - 1, \\
\forall \ u = 0 \ldots s - 1, \ \forall \ v = 0 \ldots N_r - 1, \\
H^{(i)}_k (j + us', v) = H_k (j + u n_t + is', v).
\tag{49}
\]

Likewise, \( \mathbf{S}^{(i)} \) is the \( s' \times s' \) matrix obtained from the \( i \)th block of \( s' \) rows of \( \mathbf{S} \) and every \( n_t / s' \)th block of \( s' \) columns beginning with the \( i \)th block. We easily show that

\[
\mathbf{S} \mathbf{H}_k = \begin{bmatrix}
\mathbf{S}^{(1)} \mathbf{H}^{(1)}_k \\
\mathbf{S}^{(2)} \mathbf{H}^{(2)}_k \\
\vdots \\
\mathbf{S}^{(n_t / s')} \mathbf{H}^{(n_t / s')}_k
\end{bmatrix}
\tag{50}
\]

which means that the matrix \( \mathbf{S} \) independently precodes the \( n_t / s' \) groups of transmit antennas.

Thus, the optimization may be split into \( n_t / s' \) independent optimizations of linear precoders for \( s' \times N_r \) MIMO channel with \( s \) transmission blocks and \( s' \) antennas, full space–time spreading of the \( s' \times n_s \) block-fading channel is performed, i.e., the maximum diversity order \( s n_r \) is collected by the detector.

From (15) and (46), we notice that, at the decoder input and under ideal interleaving conditions, the linear precoder at the transmitter end and the detector at the receiver end allow the conversion of the \( n_t \times n_s \) MIMO channel with \( n_t \) independent blocks into a \( 1 \times s n_r \) SIMO channel with \( n_t / s' \) independent blocks with BSK input. The independence of the blocks is provided by the structure of the linear precoding matrix.

1) The null nucleotides dispatch the transmitted symbols on \( n_t / s' \) different blocks of \( s' \) antennas.

2) The orthogonal nucleotides provide full diversity and a coding gain increasing with the spreading factor.

For instance, if a rate–1/2 BICM is transmitted on a quasi-static \( 4 \times 2 \) MIMO channel, linear precoding with \( s = 2 \) is required to achieve full diversity: a full-rate space–time block code with spreading factor \( s = s' = 2 \) may independently be applied on two separate groups of two transmit antennas. Good \( 2 \times 2 \) space–time block codes are, for instance, the Threaded Algebraic Space–Time (TAST) architecture [12] and the Golden code [4].

Let us assume that \( n_t = 1, \ n_r = 1, \ n_r = 2, \) and \( s = s' = 2. \) The Golden code is the best space–time code for uncoded \( 2 \times 2 \) quasi-static MIMO channels. However, it does not satisfy the equal norm property of orthogonal nucleotides in Proposition 2. Indeed, one row of the Golden linear precoder contains two non-null coefficients of square norm \( \alpha_1 = 0.277 \) and \( \alpha_2 = 0.723, \) respectively. Thus, (46), which assumes equality between the eigenvalues of \( \mathbf{M}^{[1]} \mathbf{M}^{[1\ast]} \), does not hold. The following can be shown (let \( \gamma^2 = \beta^2 \gamma_{1,1}^2 \) at (51) at the bottom of the page, where \( \gamma^2 = \gamma_{2,2}^2 \). As \( d_{\text{H, min}} \) increases, \( (\gamma^2 + \gamma_{1,1}^2) / (\gamma^2 + \gamma_{2,2}^2) \) tends to 1 for any pairwise error probability and \( \xi_{\text{Golden}}(\Delta, \Lambda) \rightarrow \xi_{\text{Golden}}(\Delta, \Lambda): \) The error-correcting code limits the coding loss due to the nonequal norm of the subparts of the Golden code. As a result, if \( \gamma_{1,1}^2 = \gamma_{2,2}^2 = 0, \) which is the worst case, the coding loss is \( 10 \log_{10}(\sqrt{\alpha_1} \alpha_2/4) = 0.5 \) dB.

With DNA precoder and ideal interleaving, Assumption 1 is satisfied and the modified Singleton bound on the diversity order can apply. All results from the field of error-correction coding over block-fading channels directly apply without any modification to the new \( 1 \times s n_r \) SIMO channel with \( n_t n_c / s \) independent blocks.

C. The Genie Method Design Criterion for Full Spreading Linear Precoders (\( s' = n_t \))

A linear precoding design criterion based on the genie performance optimization at the detector output has been proposed in [8]. When a genie gives a perfect information feedback on the \( m n_t \) coded bits required in the APP detector computation, the performance is computed by averaging all pairwise error probabilities obtained when changing only one bit out of \( m n_t \). Let us denote \( d \) the distance of the BSK and assume that the BSK is transmitted on antenna \( l \). The asymptotic expression of the error probability with genie is

\[
P_{\text{genie}}(\Delta, \Lambda) \sim \left( \frac{2 n_r N_\delta - 1}{n_r N_\delta} \right)^{n_t} \left( \frac{\gamma^2}{2 N_0} \right)^{-n_r \lambda_0}
\tag{52}
\]

where \( \Delta = \{ d \} \) is the set of square roots of distinct non-null eigenvalues of \( [\mathbf{P} \mathbf{s}_i] \mathbf{s}_i^\dagger \) for all \( t, \lambda_0 \), their frequency, and \( N_\delta \) their number. In the best case, there are \( s \) non-null eigenvalues and the coding gain is maximized if they are equal. First, a sufficient condition for equality between the eigenvalues of \( \mathbf{s}_i \mathbf{s}_i^\dagger \) and \( \mathbf{s}_i \mathbf{s}_i^\dagger \) is \( \| \mathbf{s}_i^\dagger \|^2 = \| \mathbf{s}_i \|^2 \). Then, all

\[
\xi_{\text{Golden}}(\Delta, \Lambda) = \sqrt{(\alpha_1 (\gamma^2 + \gamma_{1,1}^2) + \alpha_2 (\gamma^2 + \gamma_{2,2}^2)) (\alpha_1 (\gamma_1^2 + \gamma_{2,2}^2) + \alpha_2 (\gamma_2^2 + \gamma_{1,1}^2))}
\tag{51}
\]
are equal if \( S' \) is a unitary matrix, which leads to the following proposition:

**Proposition 3:** A linear precoder achieving a diversity order \( sn_t \) with maximum coding gain at the detector output must satisfy the following conditions under perfect iterative APP decoding of the space–time BICM.

1) The \( n_s \) subparts of the rows in the \( sn_t \times sn_t \) precoding matrix have the same Euclidean norm.
2) In each of the \( n_s \) subparts, the \( s \) subparts (nucleotides) are orthogonal and have the same Euclidean norm.

Proposition 3, which is more intuitive, is equivalent to Proposition 2 only if \( s' = n_t \), i.e., in case of full spreading.

### D. Non-Full Spreading Quasi-Optimal Linear Precoder: DNA Cyclotomics

If \( s' \neq n_t \), Proposition 3 is not optimal in terms of ML performance. However, we can split the optimization of an \( N_t \times N_t \) linear precoder with spreading factor \( s \) into \( n_t/s' \) optimizations of full spreading \( N_t' \times N_t' \) linear precoders with \( N_t' = s'/s \). The optimization of \( S \) is now done in two steps.

1) Apply the genie method to design a full spreading \( N_t' \times N_t' \) linear precoder for an \( s' \times n_t \) MIMO channel with \( n_s \) blocks, satisfying Proposition 3.
2) Place the non-null subparts in \( S \) as described in Proposition 2.

Cyclotomic rotations [6] provide good performance on ergodic Rayleigh single-input single-output (SISO) channels and have the great advantage of existing for any number of complex dimensions. Moreover, any coefficient has a unity norm which implies that the norm condition of Proposition 3 is naturally satisfied. We modified the cyclotomic matrices to satisfy the orthogonality condition in the case of full spreading \( s = ntn_s \). The coefficients of \( S \) are equal to (53), shown at the bottom of the page. We denote \( S(n_t, n_s, n_t n_s) \) the modified cyclotomic rotation designed for an \( n_t \times n_t \) MIMO block-fading channel, assuming that the precoder experiences \( n_s \) channel realizations.

The last parameter in \( S(n_t, n_s, n_t n_s) \) denotes the spreading factor.

In order to satisfy Proposition 2, which gives the design criterion for non-full spreading quasi-optimal linear precoders, we follow the two steps described above. Following (53), we first construct \( S(s', n_s, s) \) designed for full spreading of an \( s' \times n_t \) MIMO block-fading channel with \( n_s \) channel states in each precoded matrix. Then, we place \( n_t/s' \) times each subpart of \( S(s', n_s, s) \) in the precoding matrix in order to satisfy Proposition 2 and construct the quasi-optimal linear precoder \( S(n_t, n_s, s) \) for any set of parameters \( n_t, n_s \), and \( s \). Its coefficients are shown in (54), also at the bottom of the page.

### E. Performance of the Quasi-Optimal Precoder With Iterative Receiver

We have presented quasi-optimal linear precoders providing good coding gain and full diversity ML performance under ideal interleaving. However, the ML decoder of the global Euclidean code does not exist and we process iterative joint detection and decoding. Proposition 2 is satisfied by an infinity of matrices, all providing the same ML performance. Let us consider the performance behavior after the first iteration. As no a priori information is available at the detector, errors before decoding are numerous and not necessarily transmitted on different precoding time periods. Let us consider one precoding time period and assume that we observe two erroneous bits. If the bits are transmitted on the same modulation symbol, the Euclidean distance \( d_{BC} \) changes but this does not affect the linear precoder optimization. However, if the two bits are placed onto two different rows of \( S \), the average performance might be modified and interference inside a block and between blocks should be considered. An optimization of the precoder following the Tarokh criterion should be done, under the conditions presented in Proposition 2. Simulation results show that the modified cyclotomic rotation has good uncoded ML performance, close to algebraic full rate space–time block codes. Thus, we expect good performance at the first iteration of a joint detection and decoding process, which is desirable to reduce the number of iterations needed to
achieve the near-ML performance and to provide good performance with noniterative receivers. The optimization of the first iteration is not addressed in this paper, but first answers are given in [31].

VI. PRACTICAL INTERLEAVER DESIGN FOR CONVOLUTIONAL CODES

The maximum diversity to be gathered is limited by the channel characteristics, the linear precoding spreading factor, and the minimum Hamming distance of the binary code, all summarized in (31). Let us assume that the linear precoder spreading factor $s$ is chosen in such a way that diversity order is maximized: $\gamma_{\text{max}} = n_c n_t n_r$. Thus, there exists an interleaver that allows ML performance with full diversity. We present a new BICM interleaver design which satisfies Definition 1 and leads to the concept of full diversity BICM since the system exhibits a predetermined diversity whatever the parameters of the considered block-fading channel.

We first build an interleaver that enables to achieve maximum diversity on an $n_t \times n_r$ quasi-static MIMO channel ($n_c = 1$) with BPSK input. Then, we generalize the interleaver construction in order to apply it to higher spectral efficiency modulations, linear precoding and finally block-fading channels ($n_c > 1$).

A. Interleaver Design for Quasi-Static MIMO Channels With BPSK Input

On quasi-static channels, a codeword undergoes only one channel realization. Let us consider an error event in the code trellis for which $w$ coded bits differ from the transmitted codeword. As all error events are supposed to have a nonzero probability, the interleaver should be designed for any of them. Let us ensure the equidistribution property that $L_f N_c$ successive coded bits, $L_f N_c$ being the length of an error path with $L_f$ branches, are transmitted by all $n_t$ transmit antennas in the same proportion. In order to optimize performance, we must also ensure the noninterference of erroneous bits within the same time period. In the ML sense, two interfering erroneous bits may either degrade the diversity or the coding gain. When considering the graph representation of our system model in Fig. 1, a time period corresponds to one channel node. Probabilistic messages on bits should be independent. Practically, bits inside a channel node should be connected to distant positions in the code trellis. These conditions lead to a design criterion for quasi-static channels, well known in the algebraic space–time coding theory as the rank criterion [40] and applied here to the BICM interleaver.

In order to design an interleaver with size $L_c N_c$ ensuring that consecutive bits are mapped on different symbol time periods over all transmit antennas, we demultiplex the $L_c N_c$ coded bits into $n_t$ vectors of length $L_c N_c/n_t$. Each of these $n_t$ subframes is separately interleaved and transmitted on a predetermined transmit antenna. However, the demultiplexing step is not simply processed via the periodic selection of every $n_t$ bits. Indeed, some error patterns of convolutional codes have periodic structure. This may result in nonequally distributed erroneous bits on the $n_t$ transmit antennas and bad coding gain for these error patterns [24]. In order to break periodic structures, we apply the following demultiplexing:

$$0 \leq i < n_t, \quad 0 \leq j < L_c N_c/n_t, \quad V_i(j) = V((i + j)n_t + jn_t)$$

(55)

where $V$ is the codeword to be demultiplexed, $V_i$ is the $i$th demultiplexed frame. This ensures the uniform distribution of erroneous bits over $n_t$ transmit antennas along the transmitted frame. Let us now limit the interference of erroneous bits during the same time period. We assume that only simple error events occur. If the same interleaver is used for all subframes, $n_t$ consecutive bits are in the same position of interleaved subframes and we can limit the interference by sliding each subframe by one bit position and transmit all frames serially on their associated antennas. Yet, this does not guarantee that $L_f N_c$ successive bits are transmitted over distinct time periods. In order to satisfy this strong condition, we use a particular $S$-random interleaver [16] with a sliding input separation which guarantees that any $L_f$ successive bits in the interleaved frames are not transmitted during the same block of $n_t$ time periods. If we consider that bit position $i$ is placed at position $\Pi_k(i)$ by the interleaver $\Pi$, we should have

$$0 \leq j < L_c N_c/n_t - L_f, \quad 0 \leq i < L_f, \quad \left[ \frac{\Pi_k(j)}{n_t} \right] \neq \left[ \frac{\Pi_k(j + i)}{n_t} \right].$$

(56)

Each of the $n_t$ subframes $V_i$ is interleaved into $V_i'$:

$$0 \leq i < n_t, \quad 0 \leq j < L_c N_c/n_t, \quad V_i(\Pi_k(j)) = V_i'(j).$$

(57)

Then, a new subframe $V_i'$ is built from $V_i$ as follows:

$$0 \leq i < n_t, \quad 0 \leq j_1 < L_c N_c/n_t^2, \quad 0 \leq j_2 < n_r, \quad V_i'((i + j_2)n_t + j_2 n_r) = V_i'(j_2 + j_1 n_r).$$

(58)

The above construction keeps blocks of $n_t$ bits of $V_i$ in positions corresponding to the same $n_t$ time periods in $V_i'$, but with a cyclic shift of $i$ positions in a block of size $n_t$.

B. Basic Interleaver Construction

Let us generalize the interleaver construction to design a basic interleaver $T_{N_f, S_f, L_f}$ for $N_f$ channel inputs, a frame size $S_f$ bits, and a separation $L_f$. We described $T_{n_t, L_c N_c, L_f}$ in the previous section. For more general system configurations, the basic interleaver $T_{N_f, S_f, L_f}$ will be used in the sequel.

In Fig. 5, we present the basic interleaver for $N_f = 4$ channel inputs. Codeword bits are distinguished by four different patterns, each pattern corresponding to a specific channel input. In step 1, the codeword is demultiplexed into $N_f$ subframes $V_i$, $i = 0, \ldots, N_f - 1$, of length $S_f/N_f$ each, as presented in the previous section. In step 2, each vector $V_i$ of size $S_f/N_f$ is interleaved by the $S$-random-like interleaver into a vector $V_i'$. In step 3, we build an $N_f \times S_f/N_f$ matrix as the concatenation of $S_f/N_f$ matrices of size $N_f \times N_f$. The latter are circulant matrices where the first row contains the $N_f$ first values of $V_0$, and the second row contains the first $N_f$ values of $V_1$. Rows 3 and 4 are built from $V_2$ and $V_3$ similarly.
Finally, the $N_t \times S_I/N_t$ matrix is transmitted over the space–time channel by distributing its rows on channel inputs and its columns on time periods.

This interleaving guarantees that $(L_I - 1)N_t + 1$ consecutive codeword bits are not transmitted during the same time period. The value of $L_I$ of the S-random-like interleaver should be chosen as large as possible in order to take into account long error events. An upper bound for $L_I$ can be found based on the interleaver separation as for classical S-random [16]. The interleaver has a sliding input separation equal to $L_I$ and an output block separation equal to $N_f$ within a subframe. Hence, drawing a simple two-level tree representation would lead to $(2L_I - 1) \times N_f \leq S_I/N_t$, rewritten as

$$L_I \leq \frac{1}{2} \left( \frac{S_I}{N_t} + 1 \right).$$

(59)

C. Interleaver Design for Quasi-Static MIMO Channels With M-ary Input

In Section VI-A, we have presented an interleaver for MIMO quasi-static channels and BPSK modulation. For a modulation with higher spectral efficiency, erroneous bits in an error path should be dispatched on different time periods and equally transmitted over all transmit antennas and bit positions. Repartition on different bit positions is required as different bits of a modulation scheme are not equally protected. These conditions are satisfied by the $\mathcal{I}_{s\eta_t, L_C, N_C, L_I}$ interleaver.

Increasing the diversity by transmitting erroneous bits on all antennas is more important than increasing the coding gain by transmitting them on all modulation bits. The $n_t$ first subframes should be transmitted on the $n_t$ transmit antennas and on the first mapping bit. The second block of $n_t$ subframes should be transmitted on the second mapping bit, and so on.

D. Application to Linear Precoding

When a linear precoder is used to recover a part of the transmit diversity, the new channel matrix $SH$ has $s_{n_t} \times s_{n_t}$ rows and columns. Linear precoders have been optimized in Section V when at most one erroneous bit is observed on each precoding time period. We have shown that the precoded channel output is divided into independent blocks. We modify the order of the rows ($s' = s/n_t$ and $N' = s')$ as shown in (60) at the bottom of the page. Now, the $n_t/s'$ consecutive rows of $S$ lead to independent row vectors $S_i'H_k$, that look like a true multiple antenna channel. In this case, the interleaver $\mathcal{I}_{s\eta_t, L_C, N_C, L_I}$ is designed

$$\forall l_2 \in [1, n_t/s'], \quad \forall l_1 \in [1, ss'], \quad \forall t \in [1, n_s], \quad \forall i \in [1, s'], \quad \forall v \in [1, s'], \quad$$

$$S(l_1-1)n_t/s + l_2, v+(l_2-1)n_t/s' + (i-1)n_t + (t-1)s' n_t$$

$$= \frac{1}{\sqrt{N'_t}} \exp \left( 2j\pi \left( (l_1-1) \left( \frac{1}{\phi^{-1}(2N'_t)} + \frac{v - 1 + (i-1)s' + (t-1)s'^2}{N'_t} \right) + (i-1) \left( \frac{1}{\phi^{-1}(2s')} + \frac{v - 1}{s'} \right) \right) \right)$$

and 0 elsewhere.  

(60)
for diversity and gain exploitation. As presented in the previous subsection, the first rows of the last interleaver matrix will be transmitted on the first mapping bit, and so on.

E. Interleaver Design for Block-Fading MIMO Channels

For block-fading channels, different channel realizations occur during the codeword. In order to take advantage of the transmission and time diversity given by the linear precoding and the different realizations of a block-fading MIMO channel, the interleaver of a BICM should place consecutive bits on different precoding time periods and equally distribute them among all linear precoding rows and all channel realizations.

We extract subframes from the codeword, each subframe will be transmitted on one of the blocks, and only experience one channel realization. We interleave each subframe with the interleaver optimized for MIMO quasi-static channel for exploiting the linear precoding diversity.

The demultiplexing into subframes is done in the same manner as for the channel inputs in step 1 of Fig. 5

\[ 0 \leq i_{nc} < n_c, \quad 0 \leq j < L_c N_c/(n_c n_t), \]
\[ V_{i_{nc}}^{n_c}(j) = V(\text{mod}(i_{nc} + j)n_c + jn_c). \quad (61) \]

This demultiplexing/interleaving is sufficient to exploit the time diversity. Indeed, there is no interference between the symbols experiencing the different channel realizations unlike symbols transmitted on different linear precoding rows and bit positions.

F. Application to Turbo Codes

The BICM precoder and interleaver have been designed to provide full diversity and optimal coding gain for any pairwise error probability. However, the error rate is given by the probability to leave the Voronoi region. With convolutional codes, the number of neighbors increases with the frame length whereas the minimum Hamming distance \(d_{H\text{min}}\) remains constant. Thus, the FER increases with frame length. In order to obtain the opposite behavior, the Euclidean distance must increase with frame length and provide a performance gain higher than the performance degradation due to the increased number of neighbors. It has been shown in [9], [25] that turbo-like codes can fulfill such a condition over block-fading channels. As proposed in [24], we modify the classical parallel turbo code with two encoders RSC1 and RSC2 and an interleaver \(\Pi_t\) by adding a de-interleaving \(\Pi_t^{-1}\) of coded bits at the output of RSC2. Thanks to this de-interleaving, error events are localized and the optimized channel interleaver can be applied.
NRNSC code and linear precoders

BSK distances

is constant during the block. This results in a complexity reduction for the marginalization, which now requires around $L_C N_C/(msn_t)2^{mN_t}$ operations per iteration if $s \ll L_C N_C$. For large values of $mN_t$, the complexity of the exhaustive search becomes prohibitive. In order to cope with complexity issues, quasi-optimal or suboptimal MIMO detectors may also be used, e.g., a SISO list sphere decoder [28], [38], [3], [7], a SISO-MMSE detector [17], [41], or a detector using sequential Monte Carlo method [15].

Let us consider a $2 \times 1$ quasi-static ($n_c = 1$) MIMO channel and quaternary phase-shift keying (QPSK) modulation. We use (7, 5)$_8$ NRNSC or (3, 2)$_8$ NRNSC codes with rate 1/2 and a block length of 1024 coded bits. From the Singleton bound we know that full diversity can be achieved without linear precoding. We compare in Fig. 6 the performance obtained with a classical PR interleaver and the performance obtained with the optimized interleaver described in Section VI. Full diversity order is only achieved with the optimized interleaver, for which the performance slope is equal to the one of the outage probability. The optimized interleaver provides performance improvement without any increase of complexity neither at the transmitter nor at the receiver. In most cases, the PR interleaver only provides a diversity $n_t$, i.e., it does not allow for any transmit diversity order recovery. The (7, 5)$_8$ NRNSC code achieves a higher coding gain than the (3, 2)$_8$ NRNSC code. It achieves performance within only 2.5 dB from the outage capacity with Gaussian input and within 1.5 dB from the outage capacity with QPSK input. The performance lower bound corresponding to ideally precoded BICM is also drawn. It is obtained from the performance of the same coded modulation transmitted on a $1 \times n_t$ SIMO channel, as explained in Section V-A. There is a 1-dB gap between ideal and actual performances with the (3, 2)$_8$ NRNSC code and a 0.75-dB gap with the more powerful (7, 5)$_8$ NRNSC code. This confirms the analytical result of Section V-A obtained for ML performance: The higher the Hamming weight is, the closer to the ideal performance the actual iterative receiver can perform. However, a better code does not always provide better FER. Indeed, we have seen that, when $w \geq n_t n_c$, the full diversity of the considered pairwise error probability can be achieved with an ideal interleaver. The remaining $w - n_t n_c$ BSK distances are uniformly distributed among all the channel states. A better error-correcting code with greater Hamming weights $w$ does not enhance the diversity but the coding gain per pairwise error probability. However, the degradation induced by the increased number of neighbors may be higher than the improvement brought by increased coding gains. How to handle this tradeoff is left for further study.

In Fig. 7, we show the performance of a rate-1/2 (7, 5)$_8$ NRNSC code over a $2 \times 2$ MIMO block-fading channel with

VII. SIMULATION RESULTS

In this section, we evaluate the performance of actual iterative joint detection and decoding of the ST-BICM. The APP detector is performed by exhaustive marginalization. The set of $2^{mN_t}$ noiseless received precoded symbols $zSH$ is computed once per channel block realization since the channel matrix $SH$ is constant during the block. The optimized interleaver provides performance improvement with rate-1/2 (7, 5)$_8$ NRNSC code and linear precoders—QPSK, 2 × 2 MIMO channel, $n_c = 2$, five iterations, $L_C N_C = 2^{16}$: No linear precoder, DNA cyclotomic precoder ($s = 2$, $n_s = 1$), Golden code ($s = 2$, $n_s = 1$).
nrns = 2 and QPSK input. The frame length is 256 coded bits. With a PR interleaver, a diversity order \( n_r = 2 \) is achieved, as transmit diversity is not collected. Even with the optimized interleaver, full diversity is not obtained at the last iteration. Indeed, the Singleton bound is equal to 6 without linear precoding. Two different linear precoders, the Golden code and the DNA code, both with \( s = 2 \), are used to achieve the full diversity order 8. The slope difference between diversity orders 6 and 8 is not significant. However, linear precoding provides an additional coding gain which allows to perform within 2 dB from the outage capacity with Gaussian input using a four-state convolutional code and a small frame length. The Golden code does not satisfy the equal norm condition, which induces a slight loss in coding gain. Nevertheless, this loss is fully compensated by the averaging of the \( d_{K,L} \) into equal \( \gamma_{K,L} \) values provided by the error-correcting code as explained in Section V-A. For a higher frame length, the performance with convolutional codes is degraded. Therefore, we will also investigate performance with turbo codes.

In Fig. 8, we compare two strategies for achieving full diversity with BICM: linear precoding and constellation expansion [26]. Constellation expansion consists in increasing \( m \) while decreasing the coding rate, in order to achieve the full diversity without precoding and with the same spectral efficiency. A MIMO 2 \( \times \) 2 channel with \( n_r = 2 \) is considered. The frame length is 1024 coded bits. Using QPSK modulation and rate-1/2 coding, full diversity is not achieved. Using a precoded QPSK with \( s = 2 \) and a 16-state rate-1/2 \( (23,35)\) NRNS code having minimal Hamming distance 7, we get the same spectral efficiency, 2 bits per channel use, and the Singleton bound is equal to 8, the full diversity order. We compare this full-diversity scheme using linear precoding with a scheme using constellation expansion from QPSK to 16-QAM with a 64-state rate-1/4 NRSC code having generator polynomials \( (135,135,147,163)\) and minimal Hamming distance 20. With the latter scheme, we get the same spectral efficiency and the Singleton bound is also equal to 8. The linear precoder provides a greater diversity order at the first iteration. At the last iteration, both schemes have same diversity and the precoded scheme slightly outperforms the scheme with constellation expansion. Since the detector complexity is around \( L_cN_c/(msn_t)2^{mN_t} \) operations per iteration if \( s \ll L_cN_c \), the detection of the precoded system is as complex as the detection of the one with constellation expansion. However, channel decoding of the 64-state \( (135,135,147,163)\) NRNS code is more complex than the decoding of the 16-state \( (23,35)\) NRNS code. Thus,
Fig. 9. Optimized interleaver with rate-1/2 RSC (7,5)₆ turbo code and DNA cyclotomic precoder: 16-QAM, 1 × 1 MIMO channel, nₑ = 4, 15 iterations, LₑNₑ = 2048. Parity-check bits of the second constituent are multiplexed via the inverse turbo interleaver.

Fig. 10 illustrates the performance of a (7,5)₆ RSC turbo code over a 1 × 1 channel with nₑ = 4, 16-QAM input, and either a PR or an optimized interleaver. Two different frame lengths (256 and 2048 coded bits) are tested. With the PR interleaver and without precoding, the full diversity order 4 is not achieved. If the optimized interleaver is used, the full diversity order is not achieved neither, but the smaller slope is not visible down to a FER equal to 10⁻³. A similar behavior is obtained with PR interleaver and precoding s = 2. Finally, the DNA precoded modulation with optimized interleaver achieves full diversity performance within less than 2 dB from the outage capacity with Gaussian input.

Fig. 11 represents the performance of a (7,5)₆ RSC turbo code over a 4 × 1 quasi-static channel with BPSK input and either a PR or an optimized interleaver. Two different frame lengths (256 and 2048 coded bits) are tested. Without linear precoder and using a PR interleaver, the full diversity gain is not achieved. Asymptotically, the observed diversity is nₑ = 1, but, for low Eb/N₀, the performance is close to the performance obtained with the optimized interleaver. Indeed, the turbo code generates a large amount of errors for low Eb/N₀ and the probability of satisfying the ideal interleaving condition with a PR interleaver is high. However, when Eb/N₀ is high, only neighbors have influence on the error rate and it is crucial to place the few erroneous bits on all the channel states. This behavior is stressed with increased frame length. In order to achieve maximum diversity, according to the Singleton bound, a precoding with at least s = 2 is needed. This is confirmed by the simulation results and again the error rate decreases when the frame length increases. With the 4 × 1 MIMO channel, a large amount of interference exists between the transmit antennas. Neverthe-
Fig. 10. Impact of frame size with a rate-1/2 RSC \( \{7, 5\}_R \) turbo code: QPSK, \( 2 \times 2 \) MIMO channel, \( n_c = 1 \), 15 iterations. Parity-check bits of the second constituent are multiplexed via the inverse turbo interleaver.

Fig. 11. Impact of frame length with a rate-1/2 RSC \( \{7, 5\}_R \) turbo code: BPSK, \( 4 \times 1 \) MIMO channel, \( n_c = 1 \), 15 iterations. Parity-check bits of the second constituent are multiplexed via the inverse turbo interleaver.
less, performance is within 2.5 dB from the outage probability with Gaussian input. Performance will be even closer to the outage probability with a higher number of receive antennas or channel realizations.

In Fig. 12, performances of NRNSC codes and parallel turbo codes with RSC constituent codes over a $2 \times 1$ quasi-static MIMO channel are drawn versus frame size for $E_b/N_0 = 15$ dB. Performance of the Alamouti scheme [1] having same spectral efficiency without channel coding is also drawn as a reference. The FER increases with the frame size when using Alamouti scheme or NRNSC codes whereas it remains constant when using turbo codes. This strong property may be partly explained by the interleaving gain of the turbo code but further research is required on this point.

VIII. CONCLUSION

In this paper, we have analyzed the ideal behavior of an ST-BICM using full-rate linear precoding on a MIMO block-fading channel. Ideal performance has been derived analytically using exact pairwise error probabilities under ideal interleaving conditions. Using a bound on the diversity order, we have shown how to set the time dimension of the linear precoder. Then, we have presented how to design the linear precoder and the interleaver to obtain an ST-BICM achieving full-diversity and performing close to the ideal performance and the outage probability. Fig. 13 summarizes the optimization steps followed in this paper. The proposed DNA precoder slightly outperforms the algebraic Golden code. Furthermore, the design of DNA precoders holds for any parameter set $(n_t, s, n_s)$, whereas algebraic codes have to be specifically designed for each pair $(n_t, s)$. We have also shown that, for a same performance, using linear precoding is less complex than using constellation expansion. Finally, using turbo codes with the optimized interleaver, we have obtained a FER which does not increase with the frame length.

APPENDIX

DERIVATION OF $E_{b}H_{k} [\Psi_{L_{\text{LR}},j\nu}^T]$ FOR BLOCK-FADING CHANNELS

We first consider $n_s = 1$ and extend the result to any value of $n_s$.

A. Precoding Matrix Experiences One Channel Realization ($n_s = 1$)

For $n_s = 1$, the quasi-static channel matrix $H_k$ is defined as

$$H_k = \text{diag} \left( H_k^{[1]}, \ldots, H_k^{[s]} \right);$$

$H_k^{[i]}$ being repeated $s$ times. From $S_l$, we construct the $s \times n_t$ matrix

$$S_l^{[i]} = \left( S_l^{[i]}T, S_l^{[i]}[2]T, \ldots, S_l^{[i]}[s]T \right)^T.$$
The row vector $S_l^{[i]}$ of size $n_t$ denotes the $i$th subpart of $S_l$. The $n_r$ columns of $H_k^{[i]}$ are independent realizations of a $n_t \times 1$ multiple-input single-output channel. Let us define $M_k^{[i]}$ as an $n_r \times n_r$ Hermitian square-root of $\Sigma_k^{[i]}$. Thus

$$M_k^{[i]} = M_k^{[i]*} = U^* \sqrt{\Phi_{\Sigma_k^{[i]}}} U$$

where $\Phi_{\Sigma_k^{[i]}} = \text{diag} (\phi_{k,1}^{[i]}, \ldots, \phi_{k,n_r}^{[i]})$, $\phi_{k,i}$ being the $i$th real eigenvalue of $\Sigma_k^{[i]}$, and $U$ is a unitary matrix. We write

$$\sum_{l=1}^{N_t} R_{k,l} = \sum_{i=1}^{n_r} \sum_{l=1}^{N_t} \phi_{k,l}^r S_l^{[i]} h_i^r S_l^{[i]*} h_i^*$$

$$= \text{Tr} \left( \left( \sum_{i=1}^{n_r} M_k^{[i]} \phi_{k,i}^r M_k^{[i]*} \right) \right).$$

The random variable $\sum_{i=1}^{n_r} M_k^{[i]} \phi_{k,i}^r M_k^{[i]*}$ has a Wishart distribution with $n_r$ degrees of freedom and parameter matrix $\Sigma_k$. The characteristic function of the trace of $\sum_{i=1}^{n_r} M_k^{[i]} \phi_{k,i}^r M_k^{[i]*}$ is given in [33]. Finally

$$E_{H_k}[\Psi_{\text{LLR}}(j\nu)] = E_{H_k} \left[ \exp \left( \frac{j(\nu - \nu) \text{Tr} \left( \sum_{i=1}^{n_r} M_k^{[i]} \phi_{k,i}^r M_k^{[i]*} \right) }{2N_0} \right) \right]$$

$$= \left( \text{det} \left( \frac{\Phi_{\Sigma_k^{[i]}}} \right) \right)^{-\frac{n_r}{2}} \left( \text{det} \left( \frac{\Phi_{\Sigma_k^{[i]}}} \right) \right)^{-\frac{n_r}{2}} \left( \frac{\nu^2 - \nu}{2N_0} \text{Tr} \right).$$

$$= \sum_{i=1}^{n_r} \left( 1 - \frac{\nu^2 - \nu}{2N_0} \phi_{k,i}^r \right)^{-\frac{n_r}{2}}.$$
[15] B. Dong, X. Wang, and A. Doucet, “A new class of soft MIMO demodulation algorithms,” *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2752–2763, Nov. 2003.

[16] D. Divsalar and F. Pollara, “Turbo codes for pcs applications,” in *Proc. IEEE Int. Conf. Communications*, Seattle, WA, Jun. 1995, vol. 1, pp. 54–59.

[17] H. El Gamal and A. R. Hammons, Jr., “New approach for space-time transmitter/receiver design,” in *Proc. 37th Allerton Conf. Communication and Control*, Monticello, IL, Sep. 1999, pp. 186–195.

[18] H. El Gamal and M. O. Damen, “Universal space-time coding,” *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1097–1119, May 2003.

[19] H. El Gamal and A. R. Hammons, Jr., “On the design of algebraic space-time codes for MIMO block fading channels,” *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 151–163, Jan. 2003.

[20] M. Fozunbal, S. W. McLaughlin, and R. W. Schafer, “On performance limits of space-time codes: a sphere-packing bound approach,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2681–2687, Oct. 2003.

[21] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multiple antennas,” *AT&T Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[22] N. Gresset, J. J. Boutros, and L. Brunel, “Optimal linear precoding for BICM over MIMO channels,” in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun. 2004, p. 66.

[23] N. Gresset, J. Boutros, and L. Brunel, “Multidimensional mappings for iteratively decoded BICM on multiple antenna channel,” *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3337–3346, Sep. 2005.

[24] N. Gresset, “New Space-Time Coding Techniques With Bit-Interleaved Coded Modulations,” Ph.D. dissertation, ENST, Paris, France, 2004.

[25] A. Guillén-Fábregas and G. Caire, “Turbo-like codes are good in the block-fading channel,” in *Proc. IEEE 2004 Int. Symp. Information Theory and Applications* (ISITA), Parma, Italy, Oct. 2004.

[26] A. Guillén-Fábregas and G. Caire, “Impact of signal constellation expansion on the achievable diversity of pragmatic bit-interleaved space-time codes,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2032–2037, Aug. 2006.

[27] H. Herzberg and G. Poltyrev, “The error probability of $M$-ary PSK block coded modulation schemes,” *IEEE Trans. Commun.*, vol. 44, no. 4, pp. 427–433, Apr. 1996.

[28] B. M. Hochwald and S. teen Brink, “Achieving near-capacity on a multiple-antenna channel,” *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[29] R. Knopp and P. Humblet, “Maximizing diversity on block fading channels,” in *Proc. IEEE Int. Conf. Communications*, Montreal, QC, Canada, Jun. 1997, vol. 2, pp. 647–651.

[30] R. Knopp and P. Humblet, “On coding for block fading channels,” *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.

[31] G. Kaisidy, N. Gresset, and J. Boutros, “Information theoretical versus algebraic constructions of linear unitary precoders for non-ergodic multiple antenna channels,” in *Proc. Canadian Workshop on Information Theory*, Montréal, QC, Canada, Jun. 2005.

[32] C. Lamy, “Communications à Grande Efficacité Spectrale Sur Le Canal à évanouissement,” Ph.D. dissertation, ENST, Paris, France, 2001.

[33] D. Maiwald and D. Kraus, “Calculation of the moments of complex Wishart and complex inverse Wishart distributed matrices,” *Proc. IEE—Radar, Sonar Navig.*, vol. 147, no. 4, Aug. 2000.

[34] E. Malkamaki and H. Leib, “Coded diversity on block-fading channel,” *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 771–781, Mar. 1999.

[35] F. Oggeri, G. Rekaya, I.-C. Belfiore, and E. Viterbo, “Perfect space-time block codes,” *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3885–3902, Sep. 2006.

[36] G. Poltyrev, “Bounds on decoding error probability of binary linear codes via their spectra,” *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1284–1292, Jul. 1996.

[37] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA: McGraw-Hill, 2001.

[38] A. B. Reid, A. J. Grant, and A. P. Kind, “Low-complexity list-detection for high-rate multiple-antenna channels,” in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun. 2003, pp. 273–273.

[39] C. E. Shannon, “Probability of error for optimal codes in a Gaussian channel,” *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, May 1959.

[40] V. Tarokh, N. Seshadri, and A. R. Calderbank, “Space-time codes for high data rate wireless communication: Performance criterion and code construction,” *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[41] M. Tüchler, A. Singer, and R. Koetter, “Minimum mean squared error equalization using a-priori information,” *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.

[42] “Improved Double-STTD Schemes Using Asymmetric Modulation and Antenna Shuffling,” Texas Instruments, Busan, Korea, 2001, 3GPP RAN1 contribution.

[43] V. V. Veeravalli, “On performance analysis for signaling on correlated fading channels,” *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1879–85, Nov. 2001.