Chalmers Publication Library

On Optimal TCM Encoders

(article starts on next page)

On Optimal TCM Encoders

Alex Alvarado, *Member, IEEE,*, Alexandre Graell i Amat, *Senior Member, IEEE*, Fredrik Brännström, *Member, IEEE*, and Erik Agrell, *Senior Member, IEEE*

Abstract—An asymptotically optimal trellis-coded modulation (TCM) encoder requires the joint design of the encoder and the binary labeling of the constellation. Since analytical approaches are unknown, the only available solution is to perform an exhaustive search over the encoder and the labeling. For large constellation sizes and/or many encoder states, however, an exhaustive search is unfeasible. Traditional TCM designs overcome this problem by using a labeling that follows the set-partitioning principle and by performing an exhaustive search over the encoders. In this paper we study binary labelings for TCM and show how they can be grouped into classes, which considerably reduces the search space in a joint design. For 8-ary constellations, the number of different binary labelings that must be tested is reduced from $8! = 40320$ to 240. For the particular case of an 8-ary pulse amplitude modulation constellation, this number is further reduced to 120 and for 8-ary phase shift keying to only 30. An algorithm to generate one labeling in each class is also introduced. Asymptotically optimal TCM encoders are tabulated which are up to 0.3 dB better than the previously best known encoders.

Index Terms—Binary reflected Gray code, bit-interleaved coded modulation, coded modulation, convolutional encoder, performance bounds, set-partitioning, trellis-coded modulation, Viterbi decoding.

I. INTRODUCTION

The first breakthrough in coding for the bandwidth-limited regime came with Ungerboeck's trellis-coded modulation (TCM) [1]–[4] in the early 80s where the concept of labeling by set-partitioning (SP) was introduced. TCM was quickly adopted in the modem standards in the early 90s and is a well studied topic [5], [6, Sec. 8.12], [7, Ch. 18]. Another important discovery in coded modulation (CM) design came in 1992 when Zehavi introduced the so-called bit-interleaved coded modulation (BICM) [8], [9], usually referred to as a pragmatic approach for CM [10].

The design philosophies behind TCM and BICM for the additive white Gaussian noise (AWGN) channel are quite dif-ferent. Ungerboeck's scheme is constructed coupling together a convolutional encoder and a constellation labeled using the SP principle. For constellations having certain symmetries, SP can be achieved by using the natural binary code (NBC) [2, Fig. 4], [4, Fig. 3]. On the other hand, BICM is typically a concatenation of a convolutional encoder and a constellation labeled by the binary reflected Gray code (BRGC) [11], [12] through a bit-level interleaver. The BRGC is often used in BICM because it maximizes the BICM generalized mutual information for medium and high signal-to-noise ratios [9, Sec. III], [13, Sec. IV]. In TCM, the selection of the convolutional encoder is done so that the minimum Euclidean distance (MED) is maximized, while in BICM the encoders are the ones optimized for binary transmission. BICM systems are then based on maximum free Hamming distance codes [7, Sec. 12.3] or on the so-called optimum distance spectrum (ODS) encoders first tabulated in [14, Tables III–V] and [15, Tables II–IV] and later extended in [16].

It was recently shown in [17] that if the interleaver is removed in BICM, its performance over the AWGN channel is greatly improved. This was later analyzed in detail in [18] for a rate $R = 1/2$ encoder and a 4-ary pulse amplitude modulation (PAM) constellation, where the system in [17] was called "BICM with trivial interleavers" (BICM-T) and recognized as a TCM transmitter used with a BICM receiver. Moreover, BICM-T was shown to perform asymptotically as well as TCM (in terms of MED) [2, Table I] if properly chosen convolutional encoders are used [18, Table III]. The transmitters in [2, Table I] and [18, Table III] for the 8-state (memory $\nu = 3$) convolutional encoder[1] are shown in Fig. 1 (a) and Fig. 1 (c), respectively.

The authors in [18] failed to note that in fact the optimal TCM encoder found when analyzing BICM-T is *equivalent*[2] to the one proposed by Ungerboeck 30 years ago [19]. For a 4PAM constellation, one simple (although not unique) way of obtaining Ungerboeck's SP is by using the NBC. Moreover, the NBC can be generated using the BRGC plus one binary addition (which we call *transform*) applied to its inputs, as shown in Fig. 1(b). If the transform is included in the mapper, the encoder in Fig. 1(a) is obtained, while if it is included in the convolutional encoder, the TCM encoder in Fig. 1(c) is obtained. This equivalence also applies to encoders with larger number of states[3] and simply reveals that for 4PAM,

---

[1]Throughout this paper, all polynomial generators are given in octal.

[2]We use "equivalent" to denote two encoders with the same input-output relationship. This is formally defined in Sec. III.

[3]This equivalence does not directly hold because [18, Table III] lists the encoders in lexicographic order and because for some values of ν there are more than one encoder with identical performance.
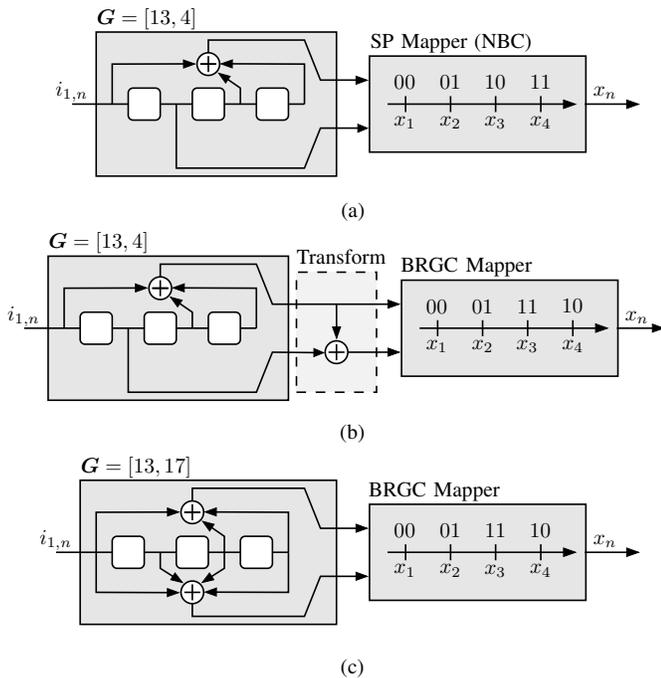
Fig. 1. Three equivalent TCM encoders [19]: (a) convolutional encoder with polynomial generators $\boldsymbol{G} = [13, 4]$ and an SP mapper [2]; (c) convolutional encoder with polynomial generators $\boldsymbol{G} = [13, 17]$ and a BRGC mapper [18]. The encoder in (b) shows how a transformation based on a binary addition can be included in the mapper (to go from (b) to (a)) or in the encoder (to go from (b) to (c)).

a TCM transceiver based on a BRGC mapper will have identical performance to Ungerboeck's TCM if the encoder is properly modified, where the modification is the application of a simple transform. The equivalence between TCM encoders and encoders optimized for the BRGC and the NBC as well as the relationship between the encoders in [18] and [2] were first pointed out to us by R. F. H. Fischer [19]. The idea of applying a linear transformation to the labeling/encoder can be traced back to [20, Fig. 6.5] (see also [21] and [22, Ch. 2]).

TCM designs based on SP are considered heuristic [23, Sec. 3.4], and thus, they do not necessarily lead to an optimal design [24, p. 680].[4] The problem of using non-SP labelings for TCM has been studied in [24, Sec. 13.2.1], [26, Sec. 8.6], and [27]. TCM encoders using the BRGC were designed in [25], by searching over convolutional encoders maximizing the MED. In [20, Ch. 6] and [21], a non-Gray non-SP labeling was used and TCM encoders with optimal spectrum were tabulated.

In a related work, Wesel *et al.* introduced in [28] the concept of the edge profile (EP) of a labeling, and argued that in most cases, the EP can be used to find equivalent TCM encoders in terms of MED. The EP is also claimed to be a good indication of the quality of a labeling for TCM in [28, Sec. I]; however, its optimality is not proven. Consequently, an exhaustive search over labelings with optimal EP does not necessarily lead to an optimal design [29].

In summary, as clearly explained in [28, Sec. I], traditional

TCM designs either optimize the encoder for a constellation labeled using the SP principle, or simply connect a convolutional encoder designed for binary transmission with an ad-hoc binary labeling. It has been known for many years that optimal TCM encoders are obtained only by *jointly* designing the convolutional encoder and the labeling of a TCM encoder [7, p. 966]. However, to the best of our knowledge, there are no works formally addressing this problem, and thus, optimal TCM encoders are yet to be found.

In this paper, we address the joint design of the feedforward convolutional encoder and the labeling for TCM. To this end, we show that binary labelings can be grouped into different classes that lead to equivalent TCM encoders. The classes are closely related to the *Hadamard classes* introduced in [30] in the context of vector quantization. This classification allows us to formally prove that in any TCM encoder, the NBC can be replaced by many other labelings (including the BRGC) without causing any performance degradation, provided that the encoder is properly selected. This explains the asymptotic equivalence between BICM-T and TCM observed in [18]. Moreover, since the classification reduces the number of labelings that must be tested in an exhaustive search, we use it to tabulate optimal TCM encoders for 4-ary and 8-ary constellations.

II. Preliminaries

A. Notation Convention

Throughout this paper, scalars are denoted by italic letters x, row vectors by boldface letters $\boldsymbol{x} = [x_1, \ldots, x_N]$, temporal sequences by underlined boldface letters $\underline{\boldsymbol{x}} = [\boldsymbol{x}[1], \ldots, \boldsymbol{x}[N_\mathrm{s}]]$, and matrices by capital boldface letters $\boldsymbol{X}$ where $x_{i,j}$ represents the entry of $\boldsymbol{X}$ at row i, column j. The transpose of a matrix/vector is denoted by $[\cdot]^\mathsf{T}$. Matrices are sometimes expressed in the compact form $\boldsymbol{X} = [\boldsymbol{x}_1; \boldsymbol{x}_2; \ldots; \boldsymbol{x}_M]$, where $\boldsymbol{x}_i = [x_{i,1}, \ldots, x_{i,N}]$ is the ith row. Sets are denoted using calligraphic letters $\mathcal{C}$ and the binary set is defined as $\mathcal{B} \triangleq \{0, 1\}$. Binary addition is denoted by $a \oplus b$.

The probability mass function (PMF) of the random variable Y is denoted by $P_Y(y)$ and the probability density function (PDF) of the random variable Y by $p_Y(y)$. Conditional PDFs are denoted as $p_{Y|X}(y|x)$. The tail probability of a standard Gaussian random variable is denoted by $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\xi^2/2} \, d\xi$.

B. TCM Encoder

We consider the TCM encoder shown in Fig. 2 where a feedforward convolutional encoder of rate $R = k/m$ is serially connected to a mapper $\Phi_{\boldsymbol{L}}$ and the index $\boldsymbol{L}$ emphasizes the dependency of the mapper on the labeling (defined later). At each discrete time instant n, the information bits $i_{1,n}, \ldots, i_{k,n}$ are fed to the convolutional encoder, which is fully determined by k different ν_p-stage shift registers with $p = 1, \ldots, k$, and the way the input sequences are connected (through the registers) to its outputs. Closely following the notation of [7, Sec. 11.1], we denote the *memory* of the convolutional encoder by $\nu = \sum_{p=1}^k \nu_p$, and the *number of states* by 2^ν. The connection between the input and output bits is defined by the

---

[4]Indeed, the results in [25, Tables 2–3], [20, Ch. 6] and [21] show the suboptimality of the SP principle in terms of the multiplicities associated with the events at MED.
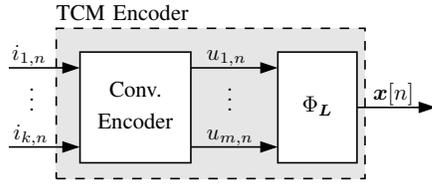
Fig. 2. Generic TCM encoder under consideration: A feedforward convolutional encoder of rate $R = k/m$ with 2^ν states serially concatenated with a memoryless m-bit mapper Φ_L.

binary representation of the *convolutional encoder matrix* [31, eq. (11.6)]

$$
G \triangleq \begin{bmatrix} \boldsymbol{g}_1^{(1)} & \boldsymbol{g}_1^{(2)} & \cdots & \boldsymbol{g}_1^{(m)} \\ \boldsymbol{g}_2^{(1)} & \boldsymbol{g}_2^{(2)} & \cdots & \boldsymbol{g}_2^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{g}_k^{(1)} & \boldsymbol{g}_k^{(2)} & \cdots & \boldsymbol{g}_k^{(m)} \end{bmatrix}, \tag{1}
$$

where $\boldsymbol{g}_p^{(l)} \triangleq [g_{p,1}^{(l)}, \ldots, g_{p,\nu_p+1}^{(l)}]^{\mathsf{T}} \in \mathcal{B}^{\nu_p+1}$ is a column vector representing the connection between the pth input sequence and the lth output sequence with $l = 1, \ldots, m$. The coefficients $g_{p,1}^{(l)}, \ldots, g_{p,\nu_p+1}^{(l)}$ are associated with the input bits $i_{p,n}, \ldots, i_{p,n-\nu_p}$, respectively, and $G \in \mathcal{B}^{(\nu+k)\times m}$. Throughout this paper, we will show the vectors $\boldsymbol{g}_p^{(l)}$ defining G either in binary or octal notation. When shown in octal notation, $g_{p,1}^{(l)}$ represents the most significant bit (see Fig. 1).

The convolutional encoder matrix (1) allows us to express the output of the convolutional encoder at time n, which we define as $\boldsymbol{u}_n \triangleq [u_{1,n}, \ldots, u_{m,n}]$, as a function of $(\nu + k)$ information bits, i.e.,

$$
\boldsymbol{u}_n = \boldsymbol{j}_n G, \tag{2}
$$

where $\boldsymbol{j}_n \triangleq [\boldsymbol{i}_n^{(1)}, \ldots, \boldsymbol{i}_n^{(k)}]$ with $\boldsymbol{i}_n^{(p)} \triangleq [i_{p,n}, \ldots, i_{p,n-\nu_p}]$ are the information bits, and the matrix multiplication is in GF(2).

The coded bits $\boldsymbol{u}_n$ are mapped to real N-dimensional constellation symbols using the mapper $\Phi_L : \mathcal{B}^m \to \mathcal{X}$, where $\mathcal{X} \subset \mathbb{R}^N$ is the constellation used for transmission, with $|\mathcal{X}| = M = 2^m$. We use $\boldsymbol{x}[n] \in \mathcal{X}$ to denote the transmitted symbols at time n and the matrix $X = [\boldsymbol{x}_1; \boldsymbol{x}_2; \ldots; \boldsymbol{x}_M]$ with $\boldsymbol{x}_q \in \mathbb{R}^N$ and $q = 1, \ldots, M$ to denote the ordered constellation points. We assume that the symbols are equally likely and that the constellation $\mathcal{X}$ is normalized to unit energy, i.e., $E_s \triangleq \mathbb{E}_X[\|X\|^2] = 1/M \sum_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{x}\|^2 = 1$. As shown in Fig. 2, each symbol represents k information bits.

The binary labeling of the qth symbol in X is denoted by $\boldsymbol{c}_q = [c_{q,1}, \ldots, c_{q,m}] \in \mathcal{B}^m$, where $c_{q,l}$ is the bit associated with the lth input of the mapper in Fig. 2. The labeling matrix is defined as $L = [\boldsymbol{c}_1; \boldsymbol{c}_2; \ldots; \boldsymbol{c}_M]$, where $\boldsymbol{c}_q$ in L corresponds to the binary label of the symbol $\boldsymbol{x}_q$ in X. Throughout this paper, we will show the vectors $\boldsymbol{c}_q$ in L in either binary or integer notation.

C. Binary Labelings for TCM

The NBC of order m is defined as $\boldsymbol{N}_m \triangleq [\boldsymbol{n}_1; \boldsymbol{n}_2; \ldots; \boldsymbol{n}_M]$ where $\boldsymbol{n}_q = [n_{q,1}, \ldots, n_{q,m}] \in \mathcal{B}^m$ is the base-2 representation of the integer $q-1$ and $n_{q,m}$ is the least significant bit. The

BRGC of order m is defined as $\boldsymbol{B}_m \triangleq [\boldsymbol{b}_1; \boldsymbol{b}_2; \ldots; \boldsymbol{b}_M]$ where $\boldsymbol{b}_q = [b_{q,1}, \ldots, b_{q,m}] \in \mathcal{B}^m$. The bits of the BRGC can be generated from the NBC as $b_{q,1} = n_{q,1}$ and $b_{q,l} = n_{q,l-1} \oplus n_{q,l}$ for $l = 2, \ldots, m$. Alternatively, we have $n_{q,l} = b_{q,1} \oplus \ldots \oplus b_{q,l-1} \oplus b_{q,l}$ for $l = 1, \ldots, m$, or, in matrix notation, $\boldsymbol{B}_m = \boldsymbol{N}_m \boldsymbol{T}$ and $\boldsymbol{N}_m = \boldsymbol{B}_m \boldsymbol{T}^{-1}$, where

$$
\boldsymbol{T} = \begin{bmatrix} 1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 1 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix}, \quad \boldsymbol{T}^{-1} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 & 1 \\ 0 & 1 & 1 & \ldots & 1 & 1 \\ 0 & 0 & 1 & \ldots & 1 & 1 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix}. \tag{3}
$$

Example 1: The NBC and BRGC of order $m = 3$ are

$$
\boldsymbol{N}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 \\ 0 & 1 & 1 \\ \mathbf{1} & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{B}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} \\ 0 & 1 & 1 \\ 0 & \mathbf{1} & 0 \\ \mathbf{1} & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \tag{4}
$$

where the pivots of the labeling matrices (defined in Sec. III-B) are highlighted.

To formally define the SP principle for a given constellation X and labeling L, we define $\mathcal{X}_l([u_{m+1-l}, \ldots, u_m]) \triangleq \{\boldsymbol{x}_q \in \mathcal{X} : [c_{q,m+1-l}, \ldots, c_{q,m}] = [u_{m+1-l}, \ldots, u_m], q = 1, \ldots, M\} \subset \mathcal{X}$ for $l = 1, \ldots, m-1$. Additionally, we define the minimum intra-Euclidean distance (intra-ED) at level l as

$$
\delta_l \triangleq \min_{\substack{\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{X}_l(\boldsymbol{u}) \\ i \neq j, \boldsymbol{u} \in \mathcal{B}^l}} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|, \quad l = 1, \ldots, m-1. \tag{5}
$$

and the MED of the constellation as δ_0.

Definition 1 (Set-partitioning [2]): For a given constellation X, the labeling L is said to follow the SP principle if $\delta_0 < \delta_1 < \ldots < \delta_{m-1}$.

Example 2: Consider an 8PSK constellation (formally defined in Sec. V). It can be easily verified that if this constellation is labeled by the NBC in (4), an SP-labeled constellation is obtained. Although the NBC is the most intuitive form for generating an SP labeling for MPSK constellations, it is not unique. As an example, consider the semi set-partitioning (SSP) labeling proposed in [32, Fig. 2(c)] and the so-called modified set-partitioning (MSP) labeling [33, Fig. 2(b)]:

$$
\boldsymbol{L}_{\text{SSP}} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{L}_{\text{MSP}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}. \tag{6}
$$

It can be shown that both labelings follow the SP principle in Definition 1.

Example 2 shows that there are multiple labelings that follow the SP principle. It can be shown that this is also the case for MPAM constellations, and that in this case, the NBC is also an SP labeling.

D. System Optimization and Search Problems

For a given constellation $\boldsymbol{X}$ and memory ν, a *TCM encoder* is fully defined by the convolutional encoder matrix $\boldsymbol{G}$ and the labeling of the constellation $\boldsymbol{L}$, and thus, a TCM encoder is defined by the pair $\Theta = [\boldsymbol{G}, \boldsymbol{L}]$.

For given integers k, m, and ν, we define the *convolutional encoder universe* as the set $\mathcal{G}_{k,m,\nu}$ of all $(\nu + k) \times m$ binary matrices[5] $\boldsymbol{G}$ which result in a noncatastrophic feedforward encoder and equally likely symbols.[6] We are also interested in the *labeling universe*, defined for a given integer m as the set $\mathcal{L}_m$ of all $M \times m$ binary matrices whose M rows are all distinct.

To the best of our knowledge, there are no works addressing the problem of designing a TCM encoder by *exhaustively* searching over the labeling universe and the convolutional encoder universe. We believe the reason for this is that an exhaustive search over encoders and labelings is unfeasible [34, Sec. I]. For example, for 8-ary constellations, there are in general $8! = 40320$ different binary labelings. In this paper, we show how a joint optimization over all $\boldsymbol{G} \in \mathcal{G}_{k,m,\nu}$ and $\boldsymbol{L} \in \mathcal{L}_m$ can be restricted, without loss of generality, to a joint optimization over all $\boldsymbol{G} \in \mathcal{G}_{k,m,\nu}$ and a subset of $\mathcal{L}_m$.

III. Equivalent Labelings for TCM Encoders

In this section, we show that binary labelings can be grouped into classes, and that all the labelings belonging to the same class lead to equivalent TCM encoders. This analysis is inspired by the one in [30], where the so-called Hadamard classes were used to solve a related search problem in source coding.

A. Equivalent TCM Encoders

The transmitted symbol at time n of a given TCM encoder $\Theta = [\boldsymbol{G}, \boldsymbol{L}]$ can be expressed using (2) as

$$\boldsymbol{x}[n] = \Phi_{\boldsymbol{L}}(\boldsymbol{u}_n) = \Phi_{\boldsymbol{L}}(\boldsymbol{j}_n \boldsymbol{G}). \tag{7}$$

Definition 2: Two TCM encoders $\Theta = [\boldsymbol{G}, \boldsymbol{L}]$ and $\tilde{\Theta} = [\tilde{\boldsymbol{G}}, \tilde{\boldsymbol{L}}]$ are said to be *equivalent* if they give the same output symbol for the same information bit sequence, i.e., if they fulfill $\Phi_{\boldsymbol{L}}(\boldsymbol{j}\boldsymbol{G}) = \Phi_{\tilde{\boldsymbol{L}}}(\boldsymbol{j}\tilde{\boldsymbol{G}})$ for any $\boldsymbol{j} \in \mathcal{B}^{\nu+k}$.

The concept of "equivalent encoders" is more restrictive than the more well-known concept of "equivalent codes". Two equivalent encoders have the same bit error rate (BER) and frame error rate (FER), whereas two equivalent codes have the same FER but in general different BER. In this paper, where BER is an important figure of merit, we are therefore more interested in equivalent encoders.

---

[5]Note that whenever $\boldsymbol{G}$ is given in its binary form, $\nu_1, \ldots, \nu_k$ are also needed to interpret $\boldsymbol{G}$ correctly according to (1).

[6]For some matrices $\boldsymbol{G}$, the symbols $\boldsymbol{x}[n]$ can be nonequally likely. This would induce nonequally likely symbols (signal shaping) which we do not consider in this work.

From now on we use $\mathcal{T}_m$ to denote the set of all binary invertible $m \times m$ matrices.

Lemma 1: $\Phi_{\boldsymbol{L}}(\boldsymbol{c}) = \Phi_{\tilde{\boldsymbol{L}}}(\boldsymbol{c}\boldsymbol{T})$ where $\tilde{\boldsymbol{L}} = \boldsymbol{L}\boldsymbol{T}$, for any two mappers $\Phi_{\boldsymbol{L}}$ and $\Phi_{\tilde{\boldsymbol{L}}}$ that use the same constellation $\boldsymbol{X}$, any $\boldsymbol{T} \in \mathcal{T}_m$, and any $\boldsymbol{c} \in \mathcal{B}^m$.

Proof: Let $\boldsymbol{v}_q \triangleq [0, \ldots, 0, 1, 0, \ldots, 0]$ be a vector of length M, where the one is in position q. From the definition of the labeling matrix $\boldsymbol{L}$, it follows that $\boldsymbol{c}_q = \boldsymbol{v}_q \boldsymbol{L}$ for $q = 1, \ldots, M$. The mapping $\Phi_{\boldsymbol{L}}$ satisfies by definition $\Phi_{\boldsymbol{L}}(\boldsymbol{c}_q) = \boldsymbol{x}_q$ for $q = 1, \ldots, M$, or, making the dependency on $\boldsymbol{L}$ explicit,

$$\Phi_{\boldsymbol{L}}(\boldsymbol{c}) = \boldsymbol{x}_q, \quad \text{if } \boldsymbol{c} = \boldsymbol{v}_q \boldsymbol{L} \tag{8}$$

for any $\boldsymbol{c} \in \mathcal{B}^m$. Similarly, for any $\boldsymbol{c} \in \mathcal{B}^m$,

$$\begin{aligned}\Phi_{\tilde{\boldsymbol{L}}}(\boldsymbol{c}\boldsymbol{T}) &= \boldsymbol{x}_q, \quad \text{if } \boldsymbol{c}\boldsymbol{T} = \boldsymbol{v}_q \tilde{\boldsymbol{L}} \\ &= \boldsymbol{x}_q, \quad \text{if } \boldsymbol{c} = \boldsymbol{v}_q \boldsymbol{L}, \end{aligned} \tag{9}$$

where the last step follows because $\boldsymbol{L} = \tilde{\boldsymbol{L}}\boldsymbol{T}^{-1}$. Since the right-hand sides of (8) and (9) are equal, $\Phi_{\tilde{\boldsymbol{L}}}(\boldsymbol{c}\boldsymbol{T}) = \Phi_{\boldsymbol{L}}(\boldsymbol{c})$ for all $\boldsymbol{c} \in \mathcal{B}^m$. □

The following theorem is the main result of this paper.

Theorem 1: For any $\boldsymbol{G} \in \mathcal{G}_{k,m,\nu}$, $\boldsymbol{L} \in \mathcal{L}_m$, and $\boldsymbol{T} \in \mathcal{T}_m$, the two TCM encoders $\Theta = [\boldsymbol{G}, \boldsymbol{L}]$ and $\tilde{\Theta} = [\tilde{\boldsymbol{G}}, \tilde{\boldsymbol{L}}]$ are equivalent, where $\tilde{\boldsymbol{L}} = \boldsymbol{L}\boldsymbol{T}$ and $\tilde{\boldsymbol{G}} = \boldsymbol{G}\boldsymbol{T}$.

Proof: For any $\boldsymbol{j} \in \mathcal{B}^{\nu+k}$, $\Phi_{\tilde{\boldsymbol{L}}}(\boldsymbol{j}\tilde{\boldsymbol{G}}) = \Phi_{\tilde{\boldsymbol{L}}}(\boldsymbol{j}\boldsymbol{G}\boldsymbol{T}) = \Phi_{\boldsymbol{L}}(\boldsymbol{j}\boldsymbol{G})$, where the last equality follows by Lemma 1. The theorem now follows using Definition 2. □

Theorem 1 shows that a full search over $\mathcal{G}_{k,m,\nu}$ and $\mathcal{L}_m$ will include many pairs of equivalent TCM encoders. Therefore, an optimal TCM encoder with given parameters can be found by searching over a subset of $\mathcal{G}_{k,m,\nu}$ and the whole set $\mathcal{L}_m$ or vice versa. In this paper, we choose the latter approach, searching over a subset of $\mathcal{L}_m$.

B. Matrix Factorization

We briefly summarize here some matrix algebra. The following definition of a *reduced column echelon matrix* comes from [35, pp. 183–184], adapted to the fact that we only consider binary labeling matrices $\boldsymbol{L}$ whose columns are all nonzero. The first nonzero element of the kth column is called the kth *pivot* of $\boldsymbol{L}$. The pivots for $\boldsymbol{N}_3$ and $\boldsymbol{B}_3$ are highlighted in (4).

Definition 3: A matrix $\boldsymbol{L} \in \mathcal{B}^{M \times m}$ is a reduced column echelon matrix if the following two conditions are fulfilled:

1) Every row with a pivot has all its other entries zero.
2) The pivot in column l is located in a row below the pivot in column $l + 1$, for $l = 1, \ldots, m - 1$.

The matrix $\boldsymbol{N}_3$ in Example 1 (or more generally $\boldsymbol{N}_m$) is an example of a reduced column echelon matrix. On the other hand, $\boldsymbol{B}_m$ is not a reduced column echelon matrix because it does not fulfill the first condition in Definition 3.

The following theorem will be used to develop an efficient search algorithm in the next section. We refer the reader to [35, p. 187, Corollary 1] for a proof. From now on we use $\mathcal{R}_m$ to denote the set of all reduced column echelon binary matrices.

TABLE I
NUMBER OF CLASSES ($M_R = |\mathcal{R}_m|$), THEIR CARDINALITY
($M_T = |\mathcal{T}_m|$), AND THE TOTAL NUMBER OF LABELINGS ($2^m!$) FOR
DIFFERENT VALUES OF m.

m	1	2	3	4	5	6
M_R	2	4	240	$1.038 \cdot 10^9$	$2.632 \cdot 10^{28}$	$6.294 \cdot 10^{78}$
M_T	1	6	168	20160	$9.999 \cdot 10^6$	$2.016 \cdot 10^{10}$
$2^m!$	2	24	40320	$2.092 \cdot 10^{13}$	$2.631 \cdot 10^{35}$	$1.269 \cdot 10^{89}$

Theorem 2: Any binary labeling $\boldsymbol{L} \in \mathcal{L}_m$ can be uniquely factorized as

$$\boldsymbol{L} = \boldsymbol{L}_R \boldsymbol{T}, \tag{10}$$

where $\boldsymbol{T} \in \mathcal{T}_m$ and $\boldsymbol{L}_R \in \mathcal{R}_m$.

Theorem 2 shows that all binary labeling matrices $\boldsymbol{L}$ can be uniquely generated by finding all the invertible matrices $\boldsymbol{T}$ (the set $\mathcal{T}_m$) and all reduced column echelon matrices $\boldsymbol{L}_R$ (the set $\mathcal{R}_m$). In particular, we have [36, eq. (1)], [30, eq. (18)]

$$M_T \triangleq |\mathcal{T}_m| = \prod_{l=1}^{m}(2^m - 2^{l-1}), \tag{11}$$

$$M_R \triangleq |\mathcal{R}_m| = \frac{2^m!}{\prod_{l=1}^{m}(2^m - 2^{l-1})}. \tag{12}$$

In Table I, the values for M_R and M_T for $1 \leq m \leq 6$ are shown. In this table we also show the number of binary labelings ($|\mathcal{L}_m| = 2^m! = M_R M_T$), i.e., the number of matrices $\boldsymbol{L}$ in the labeling universe.

The *modified Hadamard class* associated with the reduced column echelon matrix $\boldsymbol{L}_R$ is defined as the set of matrices $\boldsymbol{L}$ that can be generated via (10) by applying all $\boldsymbol{T} \in \mathcal{T}_m$. Note that these modified Hadamard classes are narrower than the regular Hadamard classes defined in [30], each including M reduced column echelon matrices. There are thus M_R modified Hadamard classes, each with cardinality M_T.

As a consequence of Theorems 1 and 2, the two TCM encoders $[\boldsymbol{G}, \boldsymbol{L}]$ and $[\boldsymbol{G}\boldsymbol{T}^{-1}, \boldsymbol{L}_R]$ are equivalent for any $\boldsymbol{G} \in \mathcal{G}_{k,m,\nu}$ and $\boldsymbol{L} \in \mathcal{L}_m$, where $\boldsymbol{L}_R$ and $\boldsymbol{T}$ are given by the factorization (10). In other words, all nonequivalent TCM encoders can be generated using one member of each modified Hadamard class only, and thus, a joint optimization over all $\boldsymbol{G} \in \mathcal{G}_{k,m,\nu}$ and $\boldsymbol{L} \in \mathcal{L}_m$ can be reduced to an optimization over all $\boldsymbol{G} \in \mathcal{G}_{k,m,\nu}$ and $\boldsymbol{L} \in \mathcal{R}_m$ with no loss in performance. This means that the search space is reduced by at least a factor of $M_T = M!/M_R$. For example, for 8-ary constellations ($m = 3$), the total number of different binary labelings that must be tested is reduced from $8! = 40320$ to 240. Moreover, as we will see in Sec. V, this can be reduced even further if the constellation $\boldsymbol{X}$ possesses certain symmetries.

C. Modified Full Linear Search Algorithm

The problem of finding the set $\mathcal{R}_m$ of reduced column echelon matrices for a given m can be solved by using a modified version of the full linear search algorithm (FLSA) introduced in [30, Sec. VIII]. We call this algorithm the modified FLSA (MFLSA). The MFLSA generates one member of

---

Algorithm 1 Modified full linear search algorithm (MFLSA)

Input: The order m
Output: Print the M_R different reduced column echelon vectors $\boldsymbol{r}$
1: $\boldsymbol{r} \leftarrow [0, 1, \ldots, M-1]$
2: **loop**
3:    **print** $\boldsymbol{r}$
4:    $index \leftarrow 0$
5:    **while** $r_M = index$ **do**
6:       $[r_{index+1}, \ldots, r_M] \leftarrow [r_M, r_{index+1}, \ldots, r_{M-1}]$
7:       $index \leftarrow index + 1$
8:       **while** $index$ is a power of 2 **do**
9:          $index \leftarrow index + 1$
10:      **end while**
11:      **if** $index = M - 1$ **then**
12:         Quit
13:      **end if**
14:    **end while**
15:    Find $pointer$ such that $r_{pointer} = index$
16:    Swap $r_{pointer}$ and $r_{pointer+1}$
17: **end loop**

---

each modified Hadamard class, the one that corresponds to a reduced column echelon matrix $\boldsymbol{L}_R$. Its pseudocode is shown in Algorithm 1. In this algorithm, the vector $\boldsymbol{r} = [r_1, \ldots, r_M]$ denotes the integer representation of the rows of the matrix $\boldsymbol{L}_R$ where $r_q = c_{q,m} + 2c_{q,m-1} + \ldots + 2^{m-1}c_{q,1}$ for $q = 1, \ldots, M$. The first labeling generated (line 1) is always the NBC. Then the algorithm proceeds by generating all permutations thereof, under the condition that no power of two $(1, 2, 4, \ldots)$ is preceded by a larger value. By Definition 3, this simple condition assures that only reduced column echelon matrices are generated.

Example 3: For $m = 2$, the MFLSA returns the following reduced column echelon matrices:

$$\mathcal{R}_2 = \left\{ \begin{bmatrix} 0\,0 \\ 0\,1 \\ \mathbf{1}\,0 \\ 1\,1 \end{bmatrix}, \begin{bmatrix} 0\,\mathbf{1} \\ 0\,0 \\ \mathbf{1}\,0 \\ 1\,1 \end{bmatrix}, \begin{bmatrix} 0\,\mathbf{1} \\ \mathbf{1}\,0 \\ 0\,0 \\ 1\,1 \end{bmatrix}, \begin{bmatrix} 0\,\mathbf{1} \\ \mathbf{1}\,0 \\ 1\,1 \\ 0\,0 \end{bmatrix} \right\}, \tag{13}$$

where the first element in $\mathcal{R}_2$ is the NBC defined in Sec. II-C and again we highlighted the pivots of the matrices. The 6 binary invertible matrices for $m = 2$ are

$$\mathcal{T}_2 = \left\{ \begin{bmatrix} 0\,1 \\ 1\,0 \end{bmatrix}, \begin{bmatrix} 0\,1 \\ 1\,1 \end{bmatrix}, \begin{bmatrix} 1\,0 \\ 0\,1 \end{bmatrix}, \begin{bmatrix} 1\,0 \\ 1\,1 \end{bmatrix}, \begin{bmatrix} 1\,1 \\ 0\,1 \end{bmatrix}, \begin{bmatrix} 1\,1 \\ 1\,0 \end{bmatrix} \right\}. \tag{14}$$

Using Theorem 2, all the 24 binary labelings in $\mathcal{L}_2$ (see Table I) can be generated by multiplying the matrices in $\mathcal{R}_2$ and $\mathcal{T}_2$.

Example 4: For $m = 3$, the reduced column echelon matrices generated by the MFLSA are shown in Table II (in integer notation). The MFLSA first generates row number one, then row number two, then row number three, etc., where each row is generated from left to right. The first column in the table corresponds to the output of the FLSA of [30]. Columns two to eight show the additional matrices generated by the MFLSA, which are obtained from the first column by shifting the symbol zero to the right. In this table we also highlight the labelings generated by the MFLSA that at the same time have optimal EP [28] for 8PAM and 8PSK (see Sec. V).

Example 5: If we study the labelings in Example 2, we

TABLE II
REDUCED COLUMN ECHELON MATRICES FOR $m = 3$ GENERATED BY THE MFLSA. THE MFLSA FIRST GENERATES ROW NUMBER ONE, THEN ROW NUMBER TWO, ETC. THE LABELINGS SHOWN IN BOLDFACE HAVE OPTIMAL EP FOR 8PAM (FIRST FOUR COLUMNS) AND FOR 8PSK (FIRST COLUMN).

01234567	10234567	12034567	12304567	12340567	12345067	12345607	12345670
01243567	10243567	12043567	12403567	12430567	12435067	12435607	12435670
01245367	10245367	12045367	12405367	12450367	12453067	12453607	12453670
01245637	10245637	12045637	12405637	12450637	12456037	12456307	12456370
01245673	10245673	12045673	12405673	12450673	12456073	12456703	12456730
01234657	10234657	12034657	12304657	12340657	12346057	12346507	12346570
01243657	10243657	12043657	12403657	12430657	12436057	12436507	12436570
01246357	10246357	12046357	12406357	12460357	12463057	12463507	12463570
01246537	10246537	12046537	**12406537**	12460537	12465037	12465307	12465370
01246573	10246573	**12046573**	12406573	12460573	12465073	12465703	12465730
01234675	10234675	12034675	12304675	12340675	12346075	12346705	12346750
01243675	10243675	12043675	12403675	12430675	12436075	12436705	12436750
01246375	10246375	12046375	12406375	12460375	12463075	12463705	12463750
01246735	10246735	12046735	12406735	12460735	12467035	12467305	12467350
01246753	**10246753**	12046753	12406753	12460753	12467053	12467503	12467530
01234576	**10234576**	12034576	12304576	12340576	12345076	12345706	12345760
01243576	10243576	12043576	12403576	12430576	12435076	12435706	12435760
01245376	10245376	12045376	12405376	12450376	12453076	12453706	12453760
01245736	10245736	12045736	12405736	12450736	12457036	12457306	12457360
01245763	10245763	12045763	12405763	12450763	12457063	12457603	12457630
01234756	10234756	**12034756**	12304756	12340756	12347056	12347506	12347560
01243756	10243756	12043756	12403756	12430756	12437056	12437506	12437560
01247356	10247356	12047356	12407356	12470356	12473056	12473506	12473560
01247536	10247536	12047536	12407536	12470536	12475036	12475306	12475360
01247563	10247563	12047563	12407563	12470563	12475063	12475603	12475630
01234765	10234765	12034765	**12304765**	12340765	12347065	12347605	12347650
01243765	10243765	12043765	12403765	12430765	12437065	12437605	12437650
01247365	10247365	12047365	12407365	12470365	12473065	12473605	12473650
01247635	10247635	12047635	12407635	12470635	12476035	12476305	12476350
01247653	10247653	12047653	12407653	12470653	12476053	12476503	12476530

find that the SSP belongs to the first modified Hadamard class ($\boldsymbol{L}_{\mathrm{R}} = \boldsymbol{N}_3$) while the MSP belongs to a different class, i.e.,

$$\boldsymbol{L}_{\mathrm{SSP}} = \boldsymbol{N}_3 \begin{bmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 1\,0\,1 \end{bmatrix}, \quad \boldsymbol{L}_{\mathrm{MSP}} = \boldsymbol{L}_{\mathrm{R}} \begin{bmatrix} 1\,1\,1 \\ 0\,1\,0 \\ 0\,0\,1 \end{bmatrix}, \qquad (15)$$

where $\boldsymbol{L}_{\mathrm{R}}^{\mathsf{T}} = [0, 1, 2, 4, 7, 6, 5, 3]$ (in integer notation) is the 233th labeling generated by the MFLSA (see Table II). This shows that the NBC does not span all the labelings that follow the SP principle.

D. NBC and BRGC

Another way of interpreting the result in Theorem 1 is that for any TCM encoder $\tilde{\Theta} = [\tilde{\boldsymbol{G}}, \tilde{\boldsymbol{L}}]$, a new equivalent TCM encoder can be generated using an encoder $\boldsymbol{G} = \tilde{\boldsymbol{G}}\boldsymbol{T}^{-1}$ and a labeling $\boldsymbol{L} = \tilde{\boldsymbol{L}}\boldsymbol{T}^{-1}$ that belongs to the same modified Hadamard class as the original labeling $\tilde{\boldsymbol{L}}$. One direct consequence of this result is that any TCM encoder using the NBC labeling $\boldsymbol{N}_m$ and a convolutional encoder $\boldsymbol{G}$ is equivalent to a TCM encoder using the BRGC $\boldsymbol{B}_m$ and a convolutional encoder $\boldsymbol{G}\boldsymbol{T}$ with $\boldsymbol{T}$ given by (3). This is formalized in the following theorem.

Theorem 3: The BRGC and the NBC of any order m belong to the same modified Hadamard class.

Proof: The BRGC and NBC are related via $\boldsymbol{B}_m = \boldsymbol{N}_m\boldsymbol{T}$, with $\boldsymbol{T}$ given by (3). The theorem now follows from Theorem 2 and the definition of a modified Hadamard class. □

Example 6: For the two TCM encoders in Fig. 1, the NBC

and BRGC labelings are related via $\boldsymbol{B}_2 = \boldsymbol{N}_2\boldsymbol{T}$, i.e.,

$$\begin{bmatrix} 0\,0 \\ 0\,1 \\ 1\,1 \\ 1\,0 \end{bmatrix} = \begin{bmatrix} 0\,0 \\ 0\,1 \\ 1\,0 \\ 1\,1 \end{bmatrix} \begin{bmatrix} 1\,1 \\ 0\,1 \end{bmatrix}. \qquad (16)$$

Thus, the BRGC and the NBC of order $m = 2$ belong to the same modified Hadamard class, and convolutional encoders can be chosen to make the two resulting TCM encoders equivalent. This was illustrated in Fig. 1, where the transform block corresponds to the transform matrix $\boldsymbol{T} = [1, 1; 0, 1] = \boldsymbol{T}^{-1}$. Since $\boldsymbol{N}_2 = \boldsymbol{B}_2\boldsymbol{T}^{-1}$, the TCM encoders $[\boldsymbol{G}_{[13,17]}, \boldsymbol{B}_2]$ and $[\boldsymbol{G}_{[13,4]}, \boldsymbol{N}_2]$ are equivalent, where

$$\boldsymbol{G}_{[13,4]} = \begin{bmatrix} 1\,0\,1\,1 \\ 0\,1\,0\,0 \end{bmatrix}^{\mathsf{T}} = \boldsymbol{G}_{[13,17]}\boldsymbol{T}^{-1} = \begin{bmatrix} 1\,0\,1\,1 \\ 1\,1\,1\,1 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} 1\,1 \\ 0\,1 \end{bmatrix}.$$

Example 6 and Theorem 3 explain, in part, the results obtained in [18], where it is shown that the encoders in [18, Table III] used with the BRGC perform asymptotically as well as Ungerboeck's TCM.[7]

IV. ERROR PROBABILITY ANALYSIS

The results in Sec. III are valid for any memoryless channel model and any receiver; however, from now on we focus on the AWGN channel and a maximum likelihood (ML) decoder. In this section, we briefly review bounds on the error probability of TCM encoders under these constraints. These bounds will be used in Sec. IV-B to define optimal TCM encoders. The bounds we develop can be found in standard textbooks, see,

---

[7]The "in part" comes from the fact that the system studied in [18] uses a (suboptimal) BICM receiver.

e.g., [5, Ch. 4] and [23, Ch. 6], and are re-derived here to make the paper self-contained.

Since TCM encoders are in general not linear[8], the probability of error depends on the transmitted sequence, i.e., it is not possible to make the assumption that the all-zero sequence was transmitted [5, p. 101]. This constraint can be lifted if the TCM encoder is "regular" [37, Lemma 2], "superlinear" [34, Sec. II-D], "scrambled" [18], or "uniform" [38], [7, Ch. 18]. However, regularity, superlinearity and uniformity do not hold for all constellation and labelings[9], and thus, we cannot use it in this paper.

We consider a baseband-equivalent discrete-time real-valued multi-dimensional AWGN channel. The transmitted sequence of equally likely symbols is denoted by $\underline{\boldsymbol{x}} = [\boldsymbol{x}[1], \dots, \boldsymbol{x}[N_{\mathrm{s}}]]$ where $\boldsymbol{x}[n] \in \mathcal{X}$ is the N-dimensional symbol transmitted at discrete time n and N_{s} is the block length. The received sequence of symbols is $\underline{\boldsymbol{y}} = [\boldsymbol{y}[1], \dots, \boldsymbol{y}[N_{\mathrm{s}}]]$, where $\boldsymbol{y}[n] = \boldsymbol{x}[n] + \boldsymbol{z}[n] \in \mathbb{R}^N$ is the received vector at time instant n. The channel noise $\boldsymbol{z}[n] \in \mathbb{R}^N$ is an N-dimensional vector with samples of independent and identically distributed (i.i.d.) random variables with zero mean and variance $N_0/2$ per dimension. The signal-to-noise ratio (SNR) is defined as $E_{\mathrm{s}}/N_0 = 1/N_0$. The conditional transition PDF of the channel is given by $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_q) = (N_0\pi)^{-\frac{N}{2}}\exp\left(-N_0^{-1}\|\boldsymbol{y} - \boldsymbol{x}_q\|^2\right)$.

A. Error Bounds

Let $\mathcal{X}_\ell$ be the set of all length-ℓ symbol sequences that start at an arbitrary time instant and encoder state. Let $\hat{\mathcal{X}}_\ell(\underline{\boldsymbol{x}})$ be the set of length-ℓ sequences $\underline{\hat{\boldsymbol{x}}} \neq \underline{\boldsymbol{x}}$ that start and end at the same encoder state as $\underline{\boldsymbol{x}} \in \mathcal{X}_\ell$ and where all the other $\ell - 1$ intermediate states are different. An error event occurs when the decoder chooses a sequence $\underline{\hat{\boldsymbol{x}}} \in \hat{\mathcal{X}}_\ell(\underline{\boldsymbol{x}})$ which is different from the transmitted sequence $\underline{\boldsymbol{x}}$. Using the union bound, the probability of an error event of an ML TCM decoder at a given time instant can be upper-bounded as [5, eq. (4.1)][10]

$$P_e \leq \sum_{\ell=1}^{\infty} \sum_{\underline{\boldsymbol{x}} \in \mathcal{X}_\ell} P_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{x}}) \sum_{\underline{\hat{\boldsymbol{x}}} \in \hat{\mathcal{X}}_\ell(\underline{\boldsymbol{x}})} \mathrm{PEP}(\underline{\boldsymbol{x}}, \underline{\hat{\boldsymbol{x}}}), \qquad (17)$$

where $\mathrm{PEP}(\underline{\boldsymbol{x}}, \underline{\hat{\boldsymbol{x}}})$ is the pairwise error probability (PEP) and $P_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{x}})$ is the probability that the encoder generates the sequence $\underline{\boldsymbol{x}}$.

Assuming i.i.d. information bits, the probability of the sequence starting at a given state is $1/2^\nu$. There are 2^k equally likely branches leaving each state of the trellis at each time instant, and thus,

$$P_{\underline{\boldsymbol{X}}}(\underline{\boldsymbol{x}}) = \frac{1}{2^\nu} \frac{1}{2^{k\ell}}. \qquad (18)$$

The PEP depends only on the accumulated squared ED (SED)

---

[8]Note that the usual definition of linearity applies to codes in $\mathrm{GF}(q)^N$. However, since TCM codes are defined over the real numbers, the usual definition of linearity does not apply.

[9]For 8PSK for example, there is in fact no binary labeling that gives a regular TCM encoder [23, Sec. 3.3].

[10]All the bounds in this section are dependent on the TCM encoder Θ. However, to alleviate the notation, we omit writing out Θ as an explicit argument.

between $\underline{\boldsymbol{x}}$ and $\underline{\hat{\boldsymbol{x}}}$ and can be shown to be

$$\mathrm{PEP}(\underline{\boldsymbol{x}}, \underline{\hat{\boldsymbol{x}}}) = Q\left(\sqrt{\frac{E_{\mathrm{s}}}{2N_0} \sum_{n=1}^{\ell} \|\boldsymbol{x}[n] - \hat{\boldsymbol{x}}[n]\|^2}\right). \qquad (19)$$

Let $A_{d^2,\ell}$ denote the number of pairs $\underline{\boldsymbol{x}} \in \mathcal{X}_\ell$ and $\underline{\hat{\boldsymbol{x}}} \in \hat{\mathcal{X}}_\ell(\underline{\boldsymbol{x}})$ at accumulated SED $d^2 = \sum_{n=1}^{\ell} \|\boldsymbol{x}[n] - \hat{\boldsymbol{x}}[n]\|^2$ and let $A_{w,d^2,\ell}$ denote the number of pairs at accumulated SED d^2 generated by input sequences at Hamming distance w. Using (18)–(19) and the definition of $A_{d^2,\ell}$, (17) can be expressed as

$$P_e \leq \sum_{d^2 \in \mathcal{D}} A_{d^2} Q\left(\sqrt{\frac{d^2 E_{\mathrm{s}}}{2N_0}}\right), \qquad (20)$$

where

$$A_{d^2} \triangleq \sum_{\ell=1}^{\infty} \frac{1}{2^\nu} \frac{1}{2^{k\ell}} A_{d^2,\ell} = \sum_{\ell=1}^{\infty} \frac{1}{2^\nu} \frac{1}{2^{k\ell}} \sum_{w=1}^{\infty} A_{w,d^2,\ell} \qquad (21)$$

is the *distance multiplicity* of the TCM encoder. In (20) $\mathcal{D}$ is the set of all possible accumulated SEDs between any two sequences, i.e., all the values of d^2 for which $A_{d^2} \neq 0$.

To obtain a bound on the BER, each error event must be weighted by the number of bits in error (w out of k), i.e.,

$$\mathrm{BER} \leq \sum_{d^2 \in \mathcal{D}} B_{d^2} Q\left(\sqrt{\frac{d^2 E_{\mathrm{s}}}{2N_0}}\right), \qquad (22)$$

where

$$B_{d^2} \triangleq \sum_{\ell=1}^{\infty} \frac{1}{2^\nu} \frac{1}{2^{k\ell}} \sum_{w=1}^{\infty} \frac{w}{k} A_{w,d^2,\ell} \qquad (23)$$

is the *bit multiplicity* of the TCM encoder.

Finally, to obtain a bound on the FER we generalize the bound presented in [39] for convolutional codes to obtain

$$\mathrm{FER} \leq N_{\mathrm{s}} \sum_{d^2 \in \mathcal{D}} A_{d^2} Q\left(\sqrt{\frac{d^2 E_{\mathrm{s}}}{2N_0}}\right). \qquad (24)$$

B. Optimum Distance Spectrum TCM Encoders

In this section we define TCM encoders that are optimal for asymptotically high SNR. These definitions will be used in Sec. V to tabulate optimized TCM encoders for different configurations.

We call the infinite set of triplets $\{d^2, A_{d^2}, B_{d^2}\}$ the distance spectrum (DS) of a given TCM encoder $\Theta = [\boldsymbol{G}, \boldsymbol{L}]$, where $d^2 \in \mathcal{D}$. We also define the ith SED of a given TCM encoder by d_i^2 with $i = 1, 2, 3, \dots$, where $d_{i+1}^2 > d_i^2$ and d_1^2 is the minimum SED of the TCM encoder. These SEDs correspond to the ordered set of SEDs in $\mathcal{D}$. Based on (22) and (24) we define an optimum DS-TCM (ODS-TCM) as follows.

Definition 4: A TCM encoder $\Theta = [\boldsymbol{G}, \boldsymbol{L}]$ with DS $\{d^2, A_{d^2}, B_{d^2}\}$ is said to have a superior DS to another TCM encoder $\tilde{\Theta} = [\tilde{\boldsymbol{G}}, \tilde{\boldsymbol{L}}]$ with DS $\{\tilde{d}^2, \tilde{A}_{\tilde{d}^2}, \tilde{B}_{\tilde{d}^2}\}$ if one of the following conditions is fulfilled:

1) $d_1^2 > \tilde{d}_1^2$, or
2) $d_1^2 = \tilde{d}_1^2$, $A_{d_1^2} < \tilde{A}_{\tilde{d}_1^2}$ and $B_{d_1^2} < \tilde{B}_{\tilde{d}_1^2}$, or

3) there exist an integer $l > 1$ such that $d_i^2 = \tilde{d}_i^2$, $A_{d_i^2} = \tilde{A}_{\tilde{d}_i^2}$ and $B_{d_i^2} = \tilde{B}_{\tilde{d}_i^2}$ for $i = 1, 2, \ldots, l-1$ and $d_l^2 > \tilde{d}_l^2$ or $d_l^2 = \tilde{d}_l^2$, $A_{d_l^2} < \tilde{A}_{\tilde{d}_l^2}$ and $B_{d_l^2} < \tilde{B}_{\tilde{d}_l^2}$.

Definition 5: For a given constellation $\boldsymbol{X}$ and memory ν, the TCM encoder $\Theta = [\boldsymbol{G}, \boldsymbol{L}]$ is said to be an ODS-TCM encoder if no other TCM encoder $\tilde{\Theta} = [\tilde{\boldsymbol{G}}, \tilde{\boldsymbol{L}}]$, for all $\tilde{\boldsymbol{G}} \in \mathcal{G}_{k,m,\nu}$ and $\tilde{\boldsymbol{L}} \in \mathcal{L}_m$, has a superior DS compared to Θ.

An ODS-TCM encoder in Definition 5 is the asymptotically optimal TCM encoder in terms of BER and FER for a given block length N_s. Unlike the more classical definition of optimal encoders, ODS-TCM encoders are defined as encoders that are optimal in terms of *both* A_{d^2} and B_{d^2}. This implies that in principle, for some combinations of k, m, ν, it is possible that no ODS-TCM encoder exists. As we will see in Sec. V, this is not an uncommon situation. Moreover, by using this somehow nonstandard definition we avoid listing encoders that have optimal BER performance but possibly rather poor FER performance (or vice versa). This situation happens for $R = 1/2$ and 4PAM, as we will show in Sec. V-A.

V. Numerical Results

In this section we study well-structured one- and two-dimensional constellations, i.e., MPAM and MPSK constellations. An MPAM constellation is defined by $\boldsymbol{X} = [x_1, x_2, \ldots, x_M]^\mathsf{T}$ with $x_q = -(M + 1 - 2q)\Delta \in \mathbb{R}$, $q = 1, \ldots, M$, and $\Delta^2 = 3/(M^2 - 1)$ so that $E_s = 1$. An MPSK constellation is defined by $\boldsymbol{X} = [\boldsymbol{x}_1; \boldsymbol{x}_2; \ldots; \boldsymbol{x}_M]$ with $\boldsymbol{x}_q = [\cos(2\pi(q-1)/M), \sin(2\pi(q-1)/M)] \in \mathbb{R}^2$ and $q = 1, \ldots, M$.

In the following sections we show results of exhaustive searches over $\mathcal{G}_{k,m,\nu}$ and $\mathcal{R}_m$, and thus, these results should be understood as a complete answer to the problem of jointly designing the feedforward encoder and the labeling for TCM encoders. The ODS-TCM encoders presented are obtained by comparing the first five nonzero elements in the spectrum, which we numerically calculate using a generalization of the algorithm presented in [31, Sec. 12.4.3].[11] On the other hand, the bounds used to compare with simulation results were calculated using 20 terms. The tabulated results are ordered first in terms of the output of the MFLSA, then in lexicographic order for the memories $\nu_1, \ldots, \nu_k$, and then in lexicographic order for the encoder matrices $\boldsymbol{G}$. This ordering becomes relevant when there are multiple TCM encoder with identical (and optimal) five-term DS.

A. ODS-TCM Encoders for MPAM

MPAM constellations are symmetric around zero. Because of this, two TCM encoders based on an MPAM constellation, the first one using the labeling $\boldsymbol{L} = [\boldsymbol{c}_1; \boldsymbol{c}_2; \ldots; \boldsymbol{c}_{M-1}; \boldsymbol{c}_M]$ and the second one using a "reverse" labeling $\boldsymbol{L}' = [\boldsymbol{c}_M; \boldsymbol{c}_{M-1}; \ldots; \boldsymbol{c}_2; \boldsymbol{c}_1]$, are equivalent for any M. This result implies that the number of binary labelings that give nonequivalent TCM encoders is $M_\mathrm{R}/2$. Specifically, for $m = 2$ and $m = 3$ (4PAM and 8PAM), only 2 and 120 labelings need

---

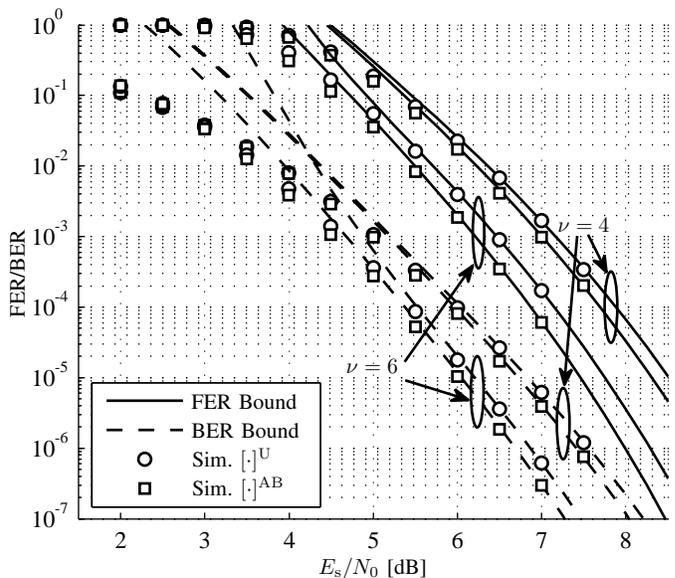[11]Note that if more than five elements are considered different ODS-TCM encoders might be found.

Fig. 3. BER/FER bounds in (22) and (24) and simulations for Ungerboeck's encoders and the ODS-TCM encoders in Table III for $N_s = 1000$, 4PAM, $R = 1/2$ (1 [bit/symbol]), and $\nu = 4, 6$.

to be evaluated, respectively, instead of 24 and 40320 in an exhaustive search, see Table I.

To generate only the $M_\mathrm{R}/2$ nonequivalent labelings for MPAM, the MFLSA in Algorithm 1 can be modified as follows. Replace M on lines 5 and 6 with $e(index)$, where the integer function $e(q)$ is defined as $M/2$ if $q = 0$ and M otherwise. This has the effect of only generating labelings in which the all-zero label is among the first $M/2$ positions (i.e., the first 4 columns of Table II for 8PAM).

1) $R = 1/2$ and 4PAM: The results obtained for $R = 1/2$ and 4PAM and different values of ν are shown in Table III. The table reports the DS as well as the labeling and convolutional encoder for the ODS-TCM encoders (shown as $[\cdot]^{\mathrm{AB}}$). For $\nu = 5$, however, no ODS-TCM encoder was found, i.e., there is no TCM encoder that is optimal in terms of both A_{d^2} and B_{d^2}. Instead, we list the TCM encoder with best A_{d^2} among those with optimal B_{d^2} (shown as $[\cdot]^{\mathrm{B}}$), or vice versa (shown as $[\cdot]^{\mathrm{A}}$). In this table we also include Ungerboeck's encoders[12], which we denote by $[\cdot]^{\mathrm{U}}$. When Ungerboeck's labeling (NBC) or Ungerboeck's convolutional encoder coincide with $[\cdot]^{\mathrm{AB}}$ or $[\cdot]^{\mathrm{B}}$, we use the notation $[\cdot]^{\mathrm{UAB}}$ or $[\cdot]^{\mathrm{UB}}$, respectively. The results in Table III show that no gains in terms of MED are obtained and that the NBC is indeed the optimal labeling for all memories. The key difference between Ungerboeck's design and the ODS-TCM encoders is the better multiplicities obtained. To compare the gains obtained by the ODS-TCM encoders over Ungerboeck's encoders, we show in Fig. 3 their BER/FER for $\nu = 4, 6$. This figure clearly shows the gains obtained by using the ODS-TCM encoders which are visible not only at high SNR, but also for low SNR values (see, e.g., the FER markers for $\nu = 6$).

---

[12]Ungerboeck did not report results for $\nu = 1$, and thus, we do not include them in the Tables, i.e., we only show the ODS-TCM encoder for $\nu = 1$.

TABLE III
DISTANCE SPECTRUM OF ODS-TCM ENCODERS ($[\cdot]^{\mathrm{AB}}$) AND UNGERBOECK'S ENCODERS ($[\cdot]^{\mathrm{U}}$) FOR $k = 1$ [BIT/SYMBOL] AND 4PAM ($m = 2$). THE NOTATION $[\cdot]^{\mathrm{A}}$ AND $[\cdot]^{\mathrm{B}}$ IS USED WHEN NO ODS-TCM ENCODER WAS FOUND.

ν	$\boldsymbol{L}^{\mathrm{T}}$	$\boldsymbol{G}$	Distance Spectrum $\{d^2, A_{d^2}, B_{d^2}\}$				
1	$[0,1,2,3]^{\mathrm{AB}}$	$[3,1]^{\mathrm{AB}}$	$\{4.00, 0.50, 0.50\}$,	$\{4.80, 0.50, 1.00\}$,	$\{5.60, 0.50, 1.50\}$,	$\{6.40, 0.50, 2.00\}$,	$\{7.20, 0.50, 2.50\}$
2	$[0,1,2,3]^{\mathrm{UAB}}$	$[5,2]^{\mathrm{U}}$	$\{7.20, 1.00, 1.00\}$,	$\{8.00, 1.25, 2.50\}$,	$\{8.80, 1.75, 5.25\}$,	$\{9.60, 2.56, 10.25\}$,	$\{10.40, 3.81, 19.06\}$
		$[7,2]^{\mathrm{AB}}$	$\{7.20, 0.50, 0.50\}$,	$\{8.00, 1.25, 2.50\}$,	$\{8.80, 1.63, 4.88\}$,	$\{9.60, 2.56, 10.25\}$,	$\{10.40, 3.78, 18.91\}$
3	$[0,1,2,3]^{\mathrm{UAB}}$	$[13,4]^{\mathrm{UAB}}$	$\{8.00, 0.25, 0.50\}$,	$\{8.80, 1.00, 3.00\}$,	$\{9.60, 1.56, 6.25\}$,	$\{10.40, 2.75, 9.75\}$,	$\{11.20, 3.14, 16.84\}$
4	$[0,1,2,3]^{\mathrm{UAB}}$	$[23,4]^{\mathrm{U}}$	$\{8.80, 0.63, 1.88\}$,	$\{9.60, 0.50, 2.00\}$,	$\{10.40, 2.00, 6.00\}$,	$\{11.20, 2.02, 10.09\}$,	$\{12.00, 2.03, 13.22\}$
		$[23,10]^{\mathrm{AB}}$	$\{8.80, 0.13, 0.38\}$,	$\{9.60, 0.50, 2.00\}$,	$\{10.40, 1.88, 5.38\}$,	$\{11.20, 2.39, 10.34\}$,	$\{12.00, 3.72, 21.03\}$
5	$[0,1,2,3]^{\mathrm{UAB}}$	$[45,10]^{\mathrm{UB}}$	$\{10.40, 1.13, 1.63\}$,	$\{11.20, 1.52, 5.09\}$,	$\{12.00, 2.59, 12.16\}$,	$\{12.80, 3.58, 22.13\}$,	$\{13.60, 5.29, 38.60\}$
		$[55,4]^{\mathrm{A}}$	$\{10.40, 0.75, 1.75\}$,	$\{11.20, 2.13, 8.75\}$,	$\{12.00, 2.14, 10.48\}$,	$\{12.80, 4.47, 24.75\}$,	$\{13.60, 5.45, 37.01\}$
6	$[0,1,2,3]^{\mathrm{UAB}}$	$[103,24]^{\mathrm{U}}$	$\{11.20, 2.34, 5.91\}$,	$\{12.80, 2.82, 22.01\}$,	$\{14.40, 7.60, 57.35\}$,	$\{16.00, 31.39, 268.35\}$,	$\{17.60, 74.37, 779.76\}$
		$[107,32]^{\mathrm{AB}}$	$\{11.20, 0.13, 0.50\}$,	$\{12.00, 1.44, 5.81\}$,	$\{12.80, 1.41, 5.77\}$,	$\{13.60, 1.73, 12.58\}$,	$\{14.40, 4.58, 31.53\}$
7	$[0,1,2,3]^{\mathrm{UAB}}$	$[235,126]^{\mathrm{U}}$	$\{12.80, 2.19, 8.19\}$,	$\{14.40, 3.05, 17.66\}$,	$\{16.00, 10.09, 89.43\}$,	$\{17.60, 25.03, 231.04\}$,	$\{19.20, 90.45, 920.63\}$
		$[313,126]^{\mathrm{AB}}$	$\{12.80, 1.46, 8.02\}$,	$\{14.40, 4.77, 34.60\}$,	$\{16.00, 15.42, 130.51\}$,	$\{17.60, 35.60, 375.08\}$,	$\{19.20, 103.30, 1213.89\}$
8	$[0,1,2,3]^{\mathrm{UAB}}$	$[515,362]^{\mathrm{U}}$	$\{13.60, 0.53, 4.66\}$,	$\{14.40, 1.89, 10.79\}$,	$\{15.20, 1.66, 14.10\}$,	$\{16.00, 3.81, 30.45\}$,	$\{16.80, 6.03, 49.34\}$
		$[677,362]^{\mathrm{AB}}$	$\{13.60, 0.36, 2.05\}$,	$\{14.40, 1.06, 6.41\}$,	$\{15.20, 1.47, 11.09\}$,	$\{16.00, 3.44, 23.69\}$,	$\{16.80, 5.25, 41.32\}$

2) $R = 2/3$ and 8PAM: The results for $R = 2/3$ and 8PAM are shown in Table IV. For $\nu = 1, 2, 3, 4, 6$ the reported encoders are in the form $[\cdot]^{\mathrm{AB}}$, while for $\nu = 5$ no ODS-TCM was found, and we use the same notation as for 4PAM. Unlike for $R = 1/2$, the parity-check matrix reported by Ungerboeck for $R = 2/3$ specifies the code but not the encoder. To have a fair comparison between Ungerboeck's codes with the ODS-TCM encoders, we first listed all the convolutional encoders that give Ungerboeck's parity-check matrix and then pick the one with optimal B_{d^2} (all of them have the same A_{d^2}). These are the encoders reported in Table IV as $[\cdot]^{\mathrm{U}}$. Even though Ungerboeck's encoders in Table IV are the best encoders for that particular parity-check matrix, they coincide with the $[\cdot]^{\mathrm{B}}$ encoders only for one out of six cases ($\nu = 5$). For all the other cases, the ODS-TCM encoders result in a better spectrum. Also, unlike for 4PAM, Table IV shows that the NBC is not the optimal labeling. For example, for $\nu = 4$, the optimal labeling is $\boldsymbol{L}^{\mathrm{T}} = [1, 2, 4, 0, 6, 5, 3, 7]^{\mathrm{AB}}$, which does not follow the SP principle (cf. Definition 1). In Fig. 4, we show the BER/FER results obtained by the ODS-TCM encoders for $R = 2/3$, 8PAM, and $\nu = 4, 6$. This figure shows the tightness of the bounds and again gains over Ungerboeck's encoders.

Fig. 4. BER/FER bounds in (22) and (24) and simulations for Ungerboeck's encoders and the ODS-TCM encoders in Table IV for $N_{\mathrm{s}} = 1000$, 8PAM, $R = 2/3$ (2 [bit/symbol]), and $\nu = 4, 6$.

B. ODS-TCM Encoders for MPSK

A TCM encoder based on an MPSK constellation is not affected by a circular rotation of its labeling, i.e., without loss of generality it can be assumed that the all zero label is assigned to the constellation point $\boldsymbol{x}_1 = [1, 0]$. The consequence of this is that for MPSK constellations, the number of reduced column echelon matrices that give nonequivalent TCM encoders is further reduced by a factor of M. In view of the results in Table I, for 4PSK, there is only one labeling that needs to be tested, e.g., the NBC. For $m \geq 3$, the nonequivalent labelings can be obtained from the MFLSA by setting $index \leftarrow 3$ in line 4, which gives the FLSA of [30]. For example, for $M = 8$, the output corresponds to the first column of Table II, which gives 30 labelings.
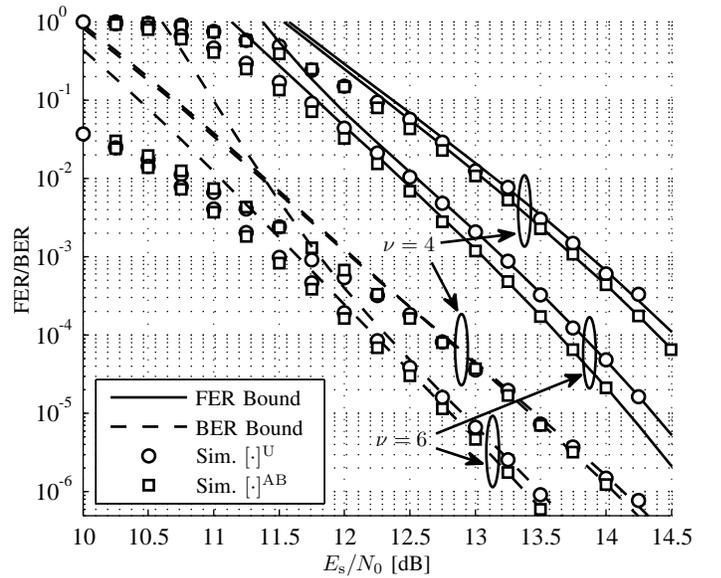
1) $R = 1/2$ and 4PSK: In this case there is only one labeling to be tested (the NBC), and thus, only a search over the encoders needs to be performed. Moreover, without loss of generality, we can use the BRGC instead (because it is in the same Hadamard class as the NBC) and search over encoders for this labeling. Since 4PSK with the BRGC can be considered as two independent 2PAM constellations (one in each dimension), the design of TCM encoders in this case boils down to selecting convolutional encoders with optimal spectrum (in the sense of Definition 5).

We have performed an exhaustive search for convolutional encoders with optimal spectrum up to $\nu = 12$ and found that our results coincide with those reported in [40, Table I]. For $\nu = 1, 2, 3, 4, 5, 6, 11, 12$ the optimal convolutional encoders

TABLE IV

DISTANCE SPECTRUM OF ODS-TCM ENCODERS ($[\cdot]^{AB}$) AND UNGERBOECK'S ENCODERS ($[\cdot]^U$) FOR $k = 2$ [BIT/SYMBOL] AND 8PAM ($m = 3$). THE NOTATION $[\cdot]^A$ AND $[\cdot]^B$ IS USED WHEN NO ODS-TCM ENCODER WAS FOUND.

ν	L^{T}	G	Distance Spectrum $\{d^2, A_{d^2}, B_{d^2}\}$				
1	$[1,2,4,0,6,5,3,7]^{AB}$	$[1,1,1;1,3,0]^{AB}$	$\{0.95, 1.13, 0.84\}$,	$\{1.14, 1.13, 1.69\}$,	$\{1.33, 1.13, 2.53\}$,	$\{1.52, 1.13, 3.38\}$,	$\{1.71, 1.13, 4.22\}$
2	$[0,1,2,3,4,5,6,7]^{UAB}$	$[1,0,0;0,5,2]^{U}$	$\{1.71, 2.25, 1.88\}$,	$\{1.90, 3.52, 5.11\}$,	$\{2.10, 6.05, 12.35\}$,	$\{2.29, 10.56, 27.64\}$,	$\{2.48, 18.47, 58.91\}$
		$[1,0,0;0,7,2]^{AB}$	$\{1.71, 1.69, 1.69\}$,	$\{1.90, 3.52, 5.11\}$,	$\{2.10, 6.01, 12.34\}$,	$\{2.29, 10.56, 27.64\}$,	$\{2.48, 18.46, 58.91\}$
3	$[0,1,2,3,4,5,6,7]^{U}$	$[1,0,0;0,13,4]^{U}$	$\{1.90, 1.27, 2.11\}$,	$\{2.10, 3.38, 6.75\}$,	$\{2.29, 5.49, 14.14\}$,	$\{2.48, 12.45, 32.48\}$,	$\{2.67, 18.59, 64.81\}$
	$[1,2,4,0,6,5,3,7]^{AB}$	$[1,1,1;2,15,0]^{AB}$	$\{1.90, 1.27, 1.90\}$,	$\{2.10, 3.38, 8.44\}$,	$\{2.29, 5.49, 17.25\}$,	$\{2.48, 12.45, 38.50\}$,	$\{2.67, 18.59, 74.81\}$
4	$[0,1,2,3,4,5,6,7]^{U}$	$[1,0,0;0,23,4]^{U}$	$\{2.10, 2.64, 5.59\}$,	$\{2.29, 2.53, 6.75\}$,	$\{2.48, 6.75, 13.50\}$,	$\{2.67, 12.11, 40.55\}$,	$\{2.86, 15.99, 66.51\}$
	$[1,2,4,0,6,5,3,7]^{AB}$	$[1,1,1;2,31,0]^{AB}$	$\{2.10, 0.95, 1.90\}$,	$\{2.29, 2.53, 7.59\}$,	$\{2.48, 7.91, 21.78\}$,	$\{2.67, 13.21, 45.70\}$,	$\{2.86, 19.77, 88.01\}$
5	$[0,1,2,3,4,5,6,7]^{UAB}$	$[1,0,0;0,45,10]^{UB}$	$\{2.48, 4.32, 6.54\}$,	$\{2.67, 7.99, 19.45\}$,	$\{2.86, 14.26, 46.29\}$,	$\{3.05, 27.05, 102.83\}$,	$\{3.24, 44.27, 201.33\}$
		$[1,0,0;0,55,4]^{A}$	$\{2.48, 3.80, 6.96\}$,	$\{2.67, 8.74, 21.63\}$,	$\{2.86, 13.53, 45.10\}$,	$\{3.05, 29.51, 106.50\}$,	$\{3.24, 44.49, 198.08\}$
6	$[0,1,2,3,4,5,6,7]^{UAB}$	$[1,0,0;0,103,24]^{U}$	$\{2.67, 10.74, 22.97\}$,	$\{3.05, 19.91, 86.93\}$,	$\{3.43, 72.68, 343.40\}$,	$\{3.81, 353.99, 1927.40\}$,	$\{4.19, 1137.86, 7442.94\}$
		$[1,0,0;0,107,32]^{AB}$	$\{2.67, 1.42, 4.27\}$,	$\{2.86, 8.46, 24.43\}$,	$\{3.05, 12.94, 40.47\}$,	$\{3.24, 15.68, 74.20\}$,	$\{3.43, 40.61, 182.47\}$

($[\cdot]^{AB}$) are in fact the encoders from [16, Table I] (which were initially optimized only in terms of B_{d^2}). For $\nu = 7, 8, 9, 10$ we found that no optimal encoder exists, i.e., the convolutional encoders optimal in terms of A_{d^2} are not optimal in terms of B_{d^2} and vice versa.[13] These encoders are in fact shown in [40, Table I][14], which extends the results in [14]–[16] because it considers both A_{d^2} and B_{d^2} as optimization criteria.

Based on the discussion above, we conclude that an ODS-TCM encoders can be constructed by concatenating the encoders in [40, Table I] with a 4PSK constellation labeled by the BRGC. Alternatively, ODS-TCM encoders can be obtained by using a 4PSK constellation labeled by the NBC and using the encoders in [40, Table I] after applying the transformation $\boldsymbol{T}^{-1} = [1, 1; 0, 1]$. For example, for $\nu = 8$, we found $\boldsymbol{G}_{[515,677]}$ and $\boldsymbol{G}_{[435,657]}$ to be the optimal encoders in terms of A_{d^2} and B_{d^2}, respectively, and thus, the two pairs of equivalent ODS-TCM encoders are $\Theta = [\boldsymbol{G}_{[515,677]}, \boldsymbol{B}_2]$ and $\tilde{\Theta} = [\boldsymbol{G}_{[515,677]}\boldsymbol{T}^{-1}, \boldsymbol{N}_2]$, and $\Theta = [\boldsymbol{G}_{[435,657]}, \boldsymbol{B}_2]$ and $\tilde{\Theta} = [\boldsymbol{G}_{[435,657]}\boldsymbol{T}^{-1}, \boldsymbol{N}_2]$.

2) $R = 2/3$ and 8PSK: The results obtained for $R = 2/3$ and 8PSK are shown in Table V. Somehow disappointingly, this table shows that the NBC is indeed the optimal labeling in all the cases, and thus, the selection of the labeling for this particular configuration does not provide any gains over Ungerboeck's TCM schemes. The better spectrum obtained by the ODS-TCM encoders in this case then comes only from the selection of the convolutional encoder.

In Fig. 5, we show the DS for the encoders in Table V with $\nu = 4$. It is clear from the figure that an encoder optimal in terms of A_{d^2} can be suboptimal in terms of B_{d^2}, and vice versa. In addition, the figure shows how the set of SEDs $\mathcal{D}$ is in general different for different encoders. It also shows how Ungerboeck's encoder is optimal in terms of A_{d^2} for the term at MED, but in general suboptimal if the whole DS is considered.

We note that depending on ν, the ODS-TCM encoders in Table V have inferior, equivalent, or superior B_{d^2} spectrum to

[13]Convolutional encoders with optimal A_{d^2} and memories up to $\nu = 26$ have been recently published in [41, Table 7.1].

[14]Although the search in [40] was performed only considering events at minimum Hamming distance and not over the whole spectrum.
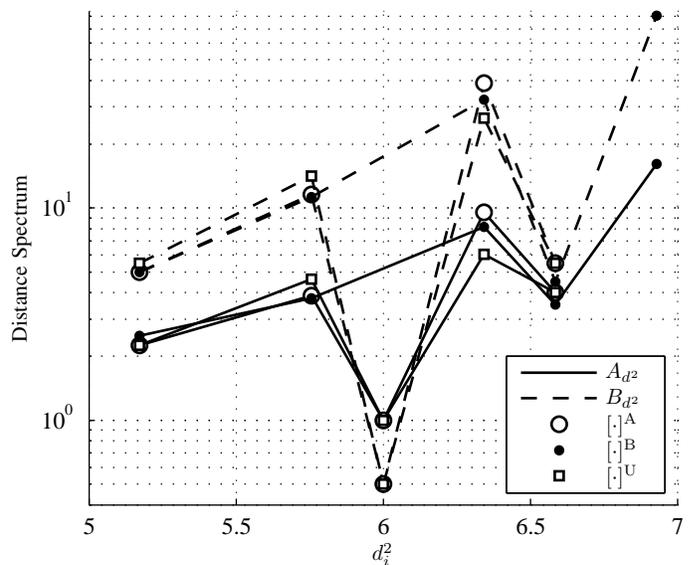

Fig. 5. DS for encoders with $\nu = 4$ for $R = 2/3$ and 8PSK from Table V.

those listed in [23, Table 3.2], [20, Table 6.10].[15] The reason for this is that the codes tabulated in [23, Table 3.2], [20, Table 6.10] are found by searching over parity check matrices and then converted to feedback encoders (in observer canonical form [20, Fig. 2.2]). On the other hand, we search over a different set of encoders, namely, over all the noncatastrophic feedforward encoders.

All labelings we found for the ODS-TCM encoders (i.e., the highlighted labelings in Table II and the optimal ones in Tables IV and V) have optimal EP. This makes us conjecture that good TCM encoders can be found by using the EP of [28] on top of the proposed classification. This approach would indeed reduce the search space (for example, for 8PAM and 8PSK constellations, only eight and two labelings, respectively, would need to be tested). However, it would not allow us to claim optimality in the sense of Definition 5.

[15]To have a fair comparison, the values of B_{d^2} listed in [23, Table 3.2], [20, Table 6.10] should be scaled by a factor $1/k = 1/2$.

TABLE V

Distance spectrum of ODS-TCM encoders ($[\cdot]^{AB}$) and Ungerboeck's encoders ($[\cdot]^{U}$) for $k=2$ [bit/symbol] and 8PSK ($m=3$). The notation $[\cdot]^{A}$ and $[\cdot]^{B}$ is used when no ODS-TCM encoder was found.

ν	$\boldsymbol{L}^{\mathrm{T}}$	$\boldsymbol{G}$	Distance Spectrum $\{d^2, A_{d^2}, B_{d^2}\}$				
1	$[0,1,2,3,4,5,6,7]^{AB}$	$[1,0,0;0,1,2]^{AB}$	$\{2.59,2.00,1.50\}$,	$\{3.17,2.00,3.00\}$,	$\{3.76,2.00,4.50\}$,	$\{4.00,1.00,0.50\}$,	$\{4.34,2.00,6.00\}$
2	$[0,1,2,3,4,5,6,7]^{UAB}$	$[1,0,0;0,5,2]^{UAB}$	$\{4.00,1.00,0.50\}$,	$\{4.59,4.00,4.00\}$,	$\{5.17,8.00,14.00\}$,	$\{5.76,16.00,38.00\}$,	$\{6.34,32.00,96.00\}$
3	$[0,1,2,3,4,5,6,7]^{UAB}$	$[1,2,0;4,1,2]^{U}$	$\{4.59,2.00,2.50\}$,	$\{5.17,4.00,8.50\}$,	$\{5.76,8.00,25.00\}$,	$\{6.00,1.00,0.50\}$,	$\{6.34,16.00,66.00\}$
		$[1,2,0;4,5,2]^{AB}$	$\{4.59,2.00,2.00\}$,	$\{5.17,4.00,8.50\}$,	$\{5.76,8.00,25.00\}$,	$\{6.00,1.00,0.50\}$,	$\{6.34,16.00,66.00\}$
4	$[0,1,2,3,4,5,6,7]^{UAB}$	$[2,7,0;7,3,2]^{U}$	$\{5.17,2.25,5.50\}$,	$\{5.76,4.63,14.13\}$,	$\{6.00,1.00,0.50\}$,	$\{6.34,6.06,26.50\}$,	$\{6.59,4.00,5.50\}$
		$[2,7,0;7,1,2]^{A}$	$\{5.17,2.25,5.00\}$,	$\{5.76,3.88,11.56\}$,	$\{6.00,1.00,0.50\}$,	$\{6.34,9.56,38.81\}$,	$\{6.59,4.00,5.50\}$
		$[1,4,2;6,1,0]^{B}$	$\{5.17,2.50,5.00\}$,	$\{5.76,3.75,11.25\}$,	$\{6.34,8.13,32.44\}$,	$\{6.59,3.50,4.50\}$,	$\{6.93,16.19,80.94\}$
5	$[0,1,2,3,4,5,6,7]^{UAB}$	$[1,2,0;30,25,16]^{U}$	$\{5.76,4.00,10.50\}$,	$\{6.00,1.00,0.50\}$,	$\{6.34,4.00,16.25\}$,	$\{6.93,4.00,24.13\}$,	$\{7.17,3.00,7.50\}$
		$[1,2,0;30,25,10]^{AB}$	$\{5.76,2.00,5.75\}$,	$\{6.00,1.00,0.50\}$,	$\{6.34,3.63,15.56\}$,	$\{6.59,3.00,5.50\}$,	$\{6.93,8.06,40.63\}$
6	$[0,1,2,3,4,5,6,7]^{UAB}$	$[4,11,0;13,4,6]^{U}$	$\{6.34,5.25,22.56\}$,	$\{7.17,10.00,28.88\}$,	$\{7.51,14.53,98.50\}$,	$\{8.00,3.00,3.75\}$,	$\{8.34,38.56,199.78\}$
		$[1,6,0;27,25,12]^{A}$	$\{6.34,3.25,12.00\}$,	$\{7.17,7.25,17.88\}$,	$\{7.51,19.13,119.17\}$,	$\{8.00,3.00,5.00\}$,	$\{8.34,36.69,159.69\}$
		$[1,6,0;35,31,6]^{B}$	$\{6.34,3.56,11.50\}$,	$\{7.17,7.25,16.88\}$,	$\{7.51,16.58,92.05\}$,	$\{8.00,3.50,4.75\}$,	$\{8.34,30.63,150.81\}$

VI. Conclusions

In this paper we analyzed the problem of jointly designing the feedforward convolutional encoder and the labeling of a TCM encoder. It was shown that the number of labelings that need to be checked can be reduced if they are grouped into modified Hadamard classes. This classification allowed us to prove that it is always possible to design a TCM encoder based on the BRGC with identical performance to the one proposed by Ungerboeck in 1982. The numerical results show that in most cases, the NBC is the optimal binary labeling for TCM encoders and that gains up to 0.3 dB over the previously best known TCM schemes can indeed be obtained.

The classification of labelings presented this paper does not make any assumption on the channel nor on the receiver. Because of this, the presented design methodology can be used to design optimal TCM encoders for other channels as well as for suboptimal (BICM) decoders.

The algorithm introduced in this paper to find all the labelings that need to be tested in an exhaustive search becomes impractical for constellations with more than 16 points. In this case, a suboptimal solution based on an algorithm (inspired by the linearity increasing swap algorithm of [30, Sec. IX]) that generates a subset of (good) labelings could be devised. This approach could also be combined with the concept of labelings with optimal EP [28]. The design of such an algorithm is left for further investigation.

Acknowledgement

References

[1] G. Ungerboeck and I. Csajka, "On improving data-link performance by increasing channel alphabet and introducing sequence decoding," in *International Symposium on Information Theory (ISIT)*, Ronneby, Sweden, June 1976, (Book of abstracts).

[2] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[3] ——, "Trellis-coded modulation with redundant signal sets Part I: Introduction," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 5–11, Feb. 1987.

[4] ——, "Trellis-coded modulation with redundant signal sets Part II: State of the art," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 12–21, Feb. 1987.

[5] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. Macmillan, 1991.

[6] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-Hill, 2008.

[7] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 2004.

[8] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.

[9] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[10] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.

[11] F. Gray, "Pulse code communications," U. S. Patent 2 632 058, Mar. 1953.

[12] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.

[13] A. Alvarado, F. Brännström, and E. Agrell, "High SNR bounds for the BICM capacity," in *IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011.

[14] J.-J. Chang, D.-J. Hwang, and M.-C. Lin, "Some extended results on the search for good convolutional codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1682–1697, Sep. 1997.

[15] I. E. Bocharova and B. D. Kudryashov, "Rational rate punctured convolutional codes for soft-decision Viterbi decoding," *IEEE Trans. Inf. Theory*, vol. 43, no. 4, pp. 1305–1313, July 1997.

[16] P. Frenger, P. Orten, and T. Ottosson, "Convolutional codes with optimum distance spectrum," *IEEE Trans. Commun.*, vol. 3, no. 11, pp. 317–319, Nov. 1999.

[17] C. Stierstorfer, R. F. H. Fischer, and J. B. Huber, "Optimizing BICM with convolutional codes for transmission over the AWGN channel," in *International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2010.

[18] A. Alvarado, L. Szczecinski, and E. Agrell, "On BICM receivers for TCM transmission," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2692–2702, Oct. 2011.

[19] R. F. H. Fischer, *private communication*, Jan. 2011.

[20] W. Zhang, "Finite state systems in mobile communications," Ph.D. dissertation, University of South Australia, Adelaide, Australia, Feb. 1996.

[21] W. Zhang, C. Schlegel, and P. Alexander, "The bit error rate reduction for systematic 8PSK trellis codes by a Gray scrambler," in *IEEE International Conference on Universal Wireless Access*, Melbourne, Australia, Apr. 1994.

[22] P. K. Gray, "Serially concatenated trellis coded modulation," Ph.D.

dissertation, University of South Australia, Adelaide, Australia, Mar. 1999.

[23] C. B. Schlegel and L. C. Perez, *Trellis and Turbo Coding*, 1st ed. John Wiley & Sons, 2004.

[24] J. B. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, 3rd ed. Springer, 2004.

[25] J. Du and M. Kasahara, "Improvements of the information-bit error rate of trellis code modulation systems," *The Transactions of the IEICE*, vol. E 72, no. 5, pp. 609–614, May 1989.

[26] G. C. Clark, Jr. and J. B. Cain, *Error-correction coding for digital communications*, 2nd ed. Plenum Press, 1981.

[27] A. J. Viterbi, J. K. Wolf, E. Zehavi, and R. Padovani, "A pragmatic approach to trellis-coded modulation," *IEEE Commun. Mag.*, vol. 27, no. 7, pp. 11–19, July 1989.

[28] R. D. Wesel, X. Liu, J. M. Cioffi, and C. Komninakis, "Constellation labeling for linear encoders," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2417–2431, Sep. 2001.

[29] R. D. Wesel, *private communication*, July 2012.

[30] P. Knagenhjelm and E. Agrell, "The Hadamard transform—a tool for index assignment," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1139–1151, July 1996.

[31] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*. Kluwer Academic, 1999.

[32] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8PSK signaling," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 1250–1257, Aug. 2002.

[33] N. H. Tran and H. H. Nguyen, "Signal mappings of 8-ary constellations for bit interleaved coded modulation with iterative decoding," *IEEE Trans. Broadcast.*, vol. 52, no. 1, pp. 92–99, Mar. 2006.

[34] S. Benedetto, M. A. Marsan, G. Albertengo, and E. Giachin, "Combined coding and modulation: Theory and applications," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 223–236, Mar. 1988.

[35] G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra*, 4th ed. New York: Macmillan, 1977.

[36] P. F. Duvall, Jr. and P. W. Harley, III, "A note on counting matrices," *SIAM Journal on Applied Mathematics*, vol. 20, no. 3, pp. 374–377, May 1971.

[37] A. R. Calderbank and N. J. A. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 2, pp. 177–195, Mar. 1987.

[38] E. Zehavi and J. K. Wolf, "On the performance evaluation of trellis codes," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 2, pp. 196–202, Mar. 1987.

[39] G. Caire and E. Viterbo, "Upper bound on the frame error probability of terminated trellis codes," *IEEE Commun. Lett.*, vol. 2, no. 1, pp. 2–4, Jan. 1998.

[40] N. Sone, M. Mohri, M. Morii, and H. Sasano, "On good convolutional codes with optimal free distance for rates 1/2, 1/3 and 1/4," *IEICE Trans. Commun.*, vol. E84-B, no. 1, pp. 116–119, Jan. 2001.

[41] F. Hug, "Codes on graphs and more," Ph.D. dissertation, Lund University, Lund, Sweden, May 2012.

Alexandre Graell i Amat (S'01-M'05-SM'10) was born in Barcelona, Catalonia, Spain, on January 28, 1976. He received the M.Sc. degree in Telecommunications Engineering from the Universitat Politècnica de Catalunya, Barcelona, Catalonia, Spain, in 2001, and the M.Sc. and the Ph.D. degrees in Electrical Engineering from the Politecnico di Torino, Turin, Italy, in 2000 and 2004, respectively. From September 2001 to April 2002, he was a Visiting Scholar at the Center for Magnetic Recording Research, University of California at San Diego, La Jolla, CA. From September 2002 to May 2003, he held a visiting appointment at Universitat Pompeu Fabra, and at the Telecommunications Technological Center of Catalonia, both in Barcelona. During 2001–2004, he also held a part-time appointment at STMicroelectronics Data Storage Division, Milan, Italy, as consultant on coding for magnetic recording channels. From March 2004 to December 2005, he was a Visiting Professor at Universitat Pompeu Fabra. From January 2006 to December 2010 he was with the Department of Electronics of TELECOM Bretagne (former ENST Bretagne), Brest, France, first as a Postdoctoral Researcher and since September 2008 as an Associate Professor. In January 2011 he joined the Department of Signals and Systems at Chalmers University of Technology, Gothenburg, Sweden, where he is currently an Assistant Professor. His research interests are in the areas of Coding and Information Theory.

Dr. Graell i Amat is currently an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE COMMUNICATIONS LETTERS. He was the General Co-Chair of the 7th International Symposium on Turbo Codes & Iterative Information Processing, Gothenburg, Sweden, August 2012. He received the post-doctoral Juan de la Cierva Fellowship of the Spanish Ministry of Education and Science, and the Marie Curie Intra-European Fellowship of the European Commission. He was the runner-up recipient of the IEEE Communications Society 2010 Europe, Middle East & Africa Region Outstanding Young Researcher Award.

Fredrik Brännström (S'98–M'05) received the M.Sc. degree in Electrical Engineering from Luleå University of Technology (Luleå, Sweden) in 1998, the Lic.Eng. degree in Communication Theory, and the Ph.D. degree in Communication Theory from the Department of Computer Engineering, Chalmers University of Technology (Gothenburg, Sweden) in 2000 and 2004, respectively. In 2012 he received the Docent title in Communication Systems from the Department of Signals and Systems, Chalmers University of Technology.

He has spent periods of 2001, 2002, 2003, and 2005 as a visiting researcher at the Institute for Telecommunications Research, University of South Australia (Adelaide, Australia). From 2004 to 2006, he was a Postdoctoral Researcher in the Communication Systems Group at the Department of Signals and Systems, Chalmers University of Technology. From 2006 to 2010 he was a Senior Algorithm Engineer and Principal Design Engineer at Quantenna Communications (Fremont, CA). At Quantenna he was part of the system team that developed the PHY layer for the world's first WiFi 802.11n chip with 4x4 MIMO using dynamic digital beamforming together with LDPC codes operating at 600 Mbps. In 2010 he joined the Department of Signals and Systems at Chalmers University of Technology, where he is currently an Assistant Professor. His research interests in communication theory and information theory include code design, coded modulation, bit-to-symbol mappings, and efficient iterative processing, as well as MIMO signal processing and algorithm design for vehicular communication systems.

Alex Alvarado (S'06–M'11) was born in 1982 in Quellón, on the island of Chiloé, Chile. He received his Electronics Engineer degree (Ingeniero Civil Electrónico) and his M.Sc. degree (Magíster en Ciencias de la Ingeniería Electrónica) from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2003 and 2005, respectively. He obtained the degree of Licentiate of Engineering (Teknologie Licentiatexamen) in 2008 and his PhD degree in 2011, both of them from Chalmers University of Technology, Gothenburg, Sweden.

In 2008, he was holder of the Merit Scholarship Program for Foreign Students, granted by the Ministère de l'Éducation, du Loisir et du Sports du Québec. From 2011 to 2012, he was a Newton International Fellow at the University of Cambridge, United Kingdom, funded by The British Academy and The Royal Society. He is currently a Marie Curie Intra-European Fellow at the same institution. His general research interests are in the areas of digital communications, coding, and information theory, and in particular, in the design of coded modulation systems.

Erik Agrell (M'99–SM'02) received the Ph.D. degree in information theory in 1997 from Chalmers University of Technology, Sweden.

From 1997 to 1999, he was a Postdoctoral Researcher with the University of California, San Diego and the University of Illinois at Urbana-Champaign. In 1999, he joined the faculty of Chalmers University of Technology, first as an Associate Professor and since 2009 as a Professor in Communication Systems. In 2010, he cofounded the Fiber-Optic Communications Research Center (FORCE) at Chalmers, where he leads the signals and systems research area. His research interests belong to the fields of information theory, coding theory, and digital communications, and his favorite applications are found in optical communications. More specifically, his current research includes polarization-multiplexed coding and modulation, nonlinear channel capacity, optical intensity modulation, cognitive radio, and coded modulation.

Prof. Agrell served as Publications Editor for the IEEE Transactions on Information Theory from 1999 to 2002 and is an Associate Editor for the IEEE Transactions on Communications since 2012. He is a recipient of the 1990 John Ericsson Medal, the 2009 ITW Best Poster Award, the 2011 GlobeCom Best Paper Award, and the 2013 Chalmers Supervisor of the Year Award.