# Strongly MDS Convolutional Codes *

Heide Gluesing-Luerssen
Department of Mathematics
University of Oldenburg
P.O. Box 2503
D-26111 Oldenburg, Germany
*e-mail:* gluesing@mathematik.uni-oldenburg.de

Joachim Rosenthal
Department of Mathematics
University of Notre Dame
Notre Dame, Indiana 46556-5683, USA
*e-mail:* Rosenthal.1@nd.edu

Roxana Smarandache
Department of Mathematical Sciences
San Diego State University
5500 Campanile Dr.
San Diego, CA 92182-7720, USA
*e-mail:* rsmarand@sciences.sdsu.edu

March 20, 2003

**Abstract**

MDS convolutional codes have the property that their free distance is maximal among all codes of the same rate and the same degree. In this paper we introduce a class of MDS convolutional codes whose column distances reach the generalized Singleton bound at the earliest possible instant. We call these codes strongly MDS convolutional codes. It is shown that these codes can decode a maximum number of errors per time interval when compared with other convolutional codes of the same rate and degree. These codes have also a maximum or near maximum distance profile. A code has a maximum distance profile if and only if the dual code has this property.

**Keywords:** MDS codes, convolutional codes, column distances, feedback decoding, superregular matrices.

*The research of this paper was presented at the 2002 IEEE International Symposium on Information Theory in Lausanne, Switzerland, June 30–July 5, 2002 and at the International Symposium on the Mathematical Theory of Networks and Systems (MTNS), University of Notre Dame, August, 12–16 2002. The authors were supported in part by NSF grants DMS-00-72383 and CCR-02-05310.

# 1 Introduction

In comparison to the literature on linear block codes there exist only relatively few algebraic constructions of convolutional codes having some good designed distance. There are even fewer algebraic decoding algorithms which are capable of exploiting the algebraic structure of the code.

Convolutional codes are typically decoded via the Viterbi algorithm which has the advantage that soft information can be processed. This algorithm has however the disadvantage that it is too complex for codes with large degree or large memory or when the block length is large. The algorithm is also not practical for convolutional codes defined over large alphabets. There are some alternative sub-optimal algorithms such as sequential decoding and feedback decoding. All these algorithms do not in general exploit the algebraic structure of the convolutional code.

In applications where codes over large alphabets are required the codes of choice are linear block codes with large distance such as Reed-Solomon codes and more general algebraic geometric codes. These codes can be algebraically decoded using e.g. the Berlekamp-Massey algorithm or some of its generalizations.

In this paper we introduce a new class of convolutional codes which we call *strongly MDS convolutional codes*. These codes are particularly suited for applications where large alphabets are involved. The free distance of these codes reaches the generalized Singleton bound. This is the maximal possible distance a convolutional code of a certain rate and degree can have. The number of errors that strongly MDS convolutional codes can correct per time interval is in a certain sense maximal as well. We will make this precise in Section 6.

Let $\mathbb{F}$ be any finite field and denote by $\mathbb{F}[D]$ and $\mathbb{F}(\!(D)\!)$ the polynomial ring respectively the field of all formal Laurent series over $\mathbb{F}$, i. e.

$$\mathbb{F}[D]=\Big\{\sum_{j=0}^{L}a_jD^j\;\Big|\;L\in\mathbb{N}_0,a_j\in\mathbb{F}\Big\}\text{ and }\mathbb{F}(\!(D)\!)=\Big\{\sum_{j=l}^{\infty}a_jD^j\;\Big|\;l\in\mathbb{Z},a_j\in\mathbb{F}\Big\}.$$

For $v=\sum_{j=l}^{\infty}v_jD^j\in\mathbb{F}(\!(D)\!)^n\backslash\{0\}$ we define $\overleftarrow{v}:=\min\{j\in\mathbb{Z}\mid v_j\neq 0\}$ to be the *delay* of the sequence $v$, that is the time instant, at which the sequence actually starts. We put $\overleftarrow{0}:=\infty$.

Let $G\in\mathbb{F}[D]^{k\times n}$ be a $k\times n$ polynomial matrix of rank $k$. We define a convolutional code of rate $k/n$ as the set

$$\mathcal{C}:=\{uG\mid u\in\mathbb{F}(\!(D)\!)^k\}\subseteq\mathbb{F}(\!(D)\!)^n \tag{1.1}$$

and say that $G$ is a generator matrix of the code $\mathcal{C}$. Two generators of $\mathcal{C}$ differ only by a nonsingular left transformation over $\mathbb{F}(\!(D)\!)$. It is well-known that we can assume $G$ to be basic and minimal in the following sense.

**Definition 1.1 (see [2])** A polynomial generator matrix $G\in\mathbb{F}[D]^{k\times n}$ is called *basic* if it has a polynomial right inverse (equivalently, if the $k\times k$-minors are coprime in $\mathbb{F}[D]$). It is called *minimal* if $\sum_{i=1}^k \nu_i$, where $\nu_i$ denotes the $i$th row degree of $G$, attains the minimal value among all generator matrices of $\mathcal{C}$.

Two basic generator matrices differ only by a unimodular left transformation over $\mathbb{F}[D]$. If $G$ is a minimal basic encoder one defines the *degree* [15] of $\mathcal{C}$ as the number $\delta := \sum_{i=1}^k \nu_i$. In the literature the degree $\delta$ is sometimes also called the *total memory* [12] or the *overall constraint length* [9] or the *complexity* [17] of the minimal basic generator matrix $G(D)$. We like to use the term degree as it corresponds to the term *McMillan degree* used in systems theory [3, 21, 23]. We also wish to point out that in algebraic geometry the degree corresponds to the degree of an associated vector bundle (i.e. quotient sheaf), see [13, 20, 22] for more details.

Since the degree depends only on the code itself, but not on the specific choice of the generator matrix $G$, we will call $\delta$ the degree of the code $\mathcal{C}$. Recall also from Forney [3] that the set $\{\nu_1, \dots, \nu_k\}$ of row degrees is the same for all minimal basic encoders of $\mathcal{C}$. Because of this reason McEliece [15] calls these indices the *Forney indices* of the code $\mathcal{C}$. As a consequence, also the number $\nu := \max\{\nu_1, \dots, \nu_k\}$ depends only on the code $\mathcal{C}$ itself and is usually called the *memory* of the code. In the sequel we will adopt the notation used by McEliece [15, p. 1082] and call a convolutional code of rate $k/n$ and degree $\delta$ an $(n, k, \delta)$-code. Every $(n, k, \delta)$-code $\mathcal{C}$ can also be represented in terms of a parity check matrix, i. e. a matrix $H \in \mathbb{F}(\!(D)\!)^{(n-k)\times n}$ such that

$$\mathcal{C} = \{v \in \mathbb{F}(\!(D)\!)^n \mid vH^{\mathsf{T}} = 0\}.$$

It is clear that we can choose $H$ to be polynomial, thus $H \in \mathbb{F}[D]^{(n-k)\times n}$, and basic. Notice also that $GH^{\mathsf{T}} = 0$ for any generator matrix $G$ of $\mathcal{C}$.

For a vector $v \in \mathbb{F}^n$, we define its weight $\mathrm{wt}(v)$ as the number of all its nonzero components. For $v = \sum_{j=l}^{\infty} v_j D^j \in \mathbb{F}(\!(D)\!)^n$ we define

$$\mathrm{wt}(v) := \sum_{j=l}^{\infty} \mathrm{wt}(v_j) \in \mathbb{N}_0 \cup \{\infty\}.$$

Finally, the *free distance* of the convolutional code $\mathcal{C} \subset \mathbb{F}(\!(D)\!)^n$ is defined through

$$d_{\text{free}} := \min\{\mathrm{wt}(v) \mid v \in \mathcal{C}, v \neq 0\}. \tag{1.2}$$

It is an easy, but crucial observation that a basic generator matrix $G$ yields a non-catastrophic and delay-free encoder, i. e., if $v = uG \in \mathbb{F}(\!(D)\!)^n$ for some $u \in \mathbb{F}(\!(D)\!)^k$, then

$$\mathrm{wt}(v) \text{ finite} \implies \mathrm{wt}(u) \text{ finite}$$

and

$$\overleftarrow{v} = \overleftarrow{u}\,.$$

Therefore, in case we are given a basic generator matrix $G$, the free distance can also be obtained as

$$\begin{aligned} d_{\text{free}} &= \min\left\{\text{wt}(v) \,\middle|\, v = uG \text{ for some } u \in \mathbb{F}[D]^k \backslash \{0\}\right\} \\ &= \min\left\{\text{wt}(v) \,\middle|\, v = uG \text{ for some } u \in \mathbb{F}[D]^k \backslash \{0\}, u_0 \neq 0\right\}. \end{aligned}$$

An $(n, k, \delta)$ convolutional code is called MDS if its free distance is maximal among all rate $k/n$ convolutional codes of degree $\delta$, i.e. an $(n, k, \delta)$ convolutional code is MDS if the free distance achieves the generalized Singleton bound [22]:

$$d_{\text{free}} = (n-k)\left(\left\lfloor \frac{\delta}{k} \right\rfloor + 1\right) + \delta + 1.$$

The concept of MDS convolutional codes was introduced by the authors in [22, 26]. Strongly MDS codes are going to be a subclass of MDS codes which have a remarkable decoding capability.

The paper is structured as follows: In Section 2 we review notions from convolutional coding theory such as the column distances, the generalized Singleton bound and we introduce the important concepts for this paper, namely the property of being strongly MDS and having a maximum distance profile. In Section 3 we show the existence of strongly MDS codes in the situation when the rate is $(n-1)/n$. In order to do so we introduce the interesting concept of a *superregular matrix* which might be of independent interest. In Section 4 we illustrate the concepts through a series of examples. In Section 5 we investigate to what extend properties of MDS, strongly MDS and maximum distance profile carry over to the dual code. The main result of this section states that a code has a maximum distance profile if and only if its dual has this property. This allows us then to show that for certain specific parameters a code is strongly MDS if and only if its dual is strongly MDS. Finally in Section 6 we show how strongly MDS convolutional codes can be decoded via feedback decoding. It turns out that the number of errors which can be decoded per time interval compares well to a maximum distance separable block code.

## 2 Strongly MDS Codes and Codes with Maximum Distance Profile

In this section we will recall the column distances of a convolutional code and their relation to the free distance. After showing some upper bounds for these distances we will introduce the notion of strongly MDS codes. It describes codes, for which the column distances attain their maximum value.

Throughout this section let $\mathcal{C} \subseteq \mathbb{F}((D))^n$ be an $(n, k, \delta)$-code with basic generator matrix

$$G = \sum_{j=0}^{\nu} G_j D^j \in \mathbb{F}[D]^{k \times n}, \;\; G_j \in \mathbb{F}^{k \times n}, G_\nu \neq 0 \tag{2.1}$$

and basic parity check matrix

$$H = \sum_{j=0}^{\mu} H_j D^j \in \mathbb{F}[D]^{(n-k)\times n},\ H_j \in \mathbb{F}^{(n-k)\times n}, H_\mu \neq 0. \tag{2.2}$$

Notice that $\nu$ is the memory of the code. For every $j \in \mathbb{N}_0$ we define the truncated sliding generator and parity check matrices

$$\begin{aligned} G_j^c &= \begin{bmatrix} G_0 & G_1 & \dots & G_j \\ & G_0 & \dots & G_{j-1} \\ & & \ddots & \vdots \\ & & & G_0 \end{bmatrix} \in \mathbb{F}^{(j+1)k\times(j+1)n}, \\ H_j^c: &= \begin{bmatrix} H_0 & & & \\ H_1 & H_0 & & \\ \vdots & \vdots & \ddots & \\ H_j & H_{j-1} & \dots & H_0 \end{bmatrix} \in \mathbb{F}^{(j+1)(n-k)\times(j+1)n}, \end{aligned} \tag{2.3}$$

where we let $G_j = 0$ (resp. $H_j = 0$) whenever $j > \nu$ (resp. $j > \mu$), see also [9, p. 110]. The identity $GH^\mathsf{T} = 0$ immediately implies $G_j^c(H_j^c)^\mathsf{T} = 0$ for all $j \in \mathbb{N}_0$. Since $G$ and $H$ are both basic, the matrices $G_j^c$ and $H_j^c$ both have full rank and therefore we even have

$$\{uG_j^c \mid u \in \mathbb{F}^{(j+1)k}\} = \{v \in \mathbb{F}^{(j+1)n} \mid v(H_j^c)^\mathsf{T} = 0\} \text{ for all } j \in \mathbb{N}_0. \tag{2.4}$$

The relevance of these matrices rests on the fact that they single out codeword sequences of length $j$ in the following sense.

**Remark 2.1** For $v := \sum_{j=l}^{\infty} v_j D^j \in \mathbb{F}(\!(D)\!)^n$ and $m, M \in \mathbb{Z}$ with $m \leq M$ define

$$v_{[m,M]} := (v_m, v_{m+1}, \dots, v_M) \in \mathbb{F}^{(M-m+1)n}.$$

Then we have the following:

(a) If $v = uG$ for some $u \in \mathbb{F}(\!(D)\!)^k$ with $\overleftarrow{u} \geq l$, then $v_{[l,l+j]} = u_{[l,l+j]}G_j^c$ and $v_{[l,l+j]}(H_j^c)^\mathsf{T} = 0$ for all $j \in \mathbb{N}_0$.

(b) If $\hat{v} = \hat{u}G_j^c \in \mathbb{F}^{(j+1)n}$ for some $j \in \mathbb{N}_0$ and $\hat{u} \in \mathbb{F}^{(j+1)k}$, then there exists $v \in \mathcal{C}$ such that $\overleftarrow{v} \geq 0$ and $v_{[0,j]} = \hat{v}$.

(c) For all $j \in \mathbb{N}_0$ we have $\{v_{[0,j]} \mid v \in \mathcal{C}, \overleftarrow{v} = 0\} = \{\hat{v} = (\hat{v}_0, \dots, \hat{v}_j) \in \mathbb{F}^{(j+1)n} \mid \hat{v}(H_j^c)^\mathsf{T} = 0,\ \hat{v}_0 \neq 0\}$.

Part (a) follows easily by equating like powers of $D$ in the equation $v = uG$ and by use of (2.4); (b) is obvious by taking $v = uG$ with $u = \sum_{i=0}^{j} \hat{u}_i D^i$, where $\hat{u} = (\hat{u}_0, \dots, \hat{u}_j)$; (c) is a consequence of (a) and (2.4).

Following [9, pp. 110], the $j$th *column distance of the code* $\mathcal{C}$ is defined to be

$$d_j^c := \min\big\{\mathrm{wt}(v_{[0,j]})\,\big|\, v\in\mathcal{C},\, \overleftarrow{v}=0\big\}. \tag{2.5}$$

Using the remark above we obtain the alternative identities

$$\begin{aligned} d_j^c &= \min\big\{\mathrm{wt}(v_{[0,j]})\,\big|\, v=uG, u\in\mathbb{F}[D]^k, u_0\neq 0\big\}\\ &= \min\big\{\mathrm{wt}(v_{[l,l+j]})\,\big|\, v=uG,\, u\in\mathbb{F}(\!(D)\!)^k,\, \overleftarrow{u}=l\big\}\\ &= \min\big\{\mathrm{wt}\big((u_0,\dots,u_j)G_j^c\big)\,\big|\, u_i\in\mathbb{F}^k,\, u_0\neq 0\big\} &(2.6)\\ &= \min\{\mathrm{wt}(\hat{v})\mid \hat{v}=(\hat{v}_0,\dots,\hat{v}_j)\in\mathbb{F}^{(j+1)n},\, \hat{v}(H_j^c)^{\mathsf{T}}=0,\, \hat{v}_0\neq 0\}. &(2.7)\end{aligned}$$

Obviously, $d_j^c\leq d_{\text{free}}$ for all $j\in\mathbb{N}_0$ and one even has [9, pp. 113]

$$d_0^c\leq d_1^c\leq d_2^c\ldots \text{ and } \lim_{j\to\infty} d_j^c = d_{\text{free}}. \tag{2.8}$$

The $(\nu+1)$-tuple of numbers $(d_0^c, d_1^c, d_1^c,\dots,d_\nu^c)$, where $\nu$ is the memory, is is called the *column distance profile* of the code [9, p. 112].

Equation (2.7) immediately implies

**Proposition 2.2** *Let $d\in\mathbb{N}$. Then the following properties are equivalent.*

*(a)* $d_j^c = d$;

*(b) none of the first $n$ columns of $H_j^c$ is contained in the span of any other $d-2$ columns and one of the first $n$ columns of $H_j^c$ is in the span of some other $d-1$ columns of that matrix.*

We leave it to the reader to verify the equivalence of the statements.

**Proposition 2.3** *For every $j\in\mathbb{N}_0$ we have*

$$d_j^c\leq (n-k)(j+1)+1.$$

PROOF: Consider the sliding parity check matrix $H_j^c$ introduced in (2.3). The set of vectors $\hat{v}=(\hat{v}_0,\dots,\hat{v}_j)\in\mathbb{F}^{(j+1)n}$, $\hat{v}(H_j^c)^{\mathsf{T}}=0$, $\hat{v}_0\neq 0$ forms a nonlinear subset of the linear block code defined by the left kernel of $H_j^c$. Since $H_j^c\in\mathbb{F}^{(n-k)(j+1)\times n(j+1)}$ any vector in the left kernel has weight at most $(n-k)(j+1)+1$ by the usual Singleton bound for block codes and this establishes the claim. □

The column distances give information about the error-correcting capabilities of the code. Precisely, $d_j^c$ determines the error-correcting capability of a decoder that estimates the message symbol $u_0$ based on the received symbols $v_{[0,j]}$, see also [9, p. 111]. Therefore, a good performance for sequential decoding requires the column distances as big as possible. The next proposition shows that maximality of $d_j^c$ implies maximality of the preceding column distances.

**Corollary 2.4** *If $d_j^c = (n-k)(j+1)+1$ for some $j \in \mathbb{N}_0$, then $d_i^c = (n-k)(i+1)+1$ for all $i \leq j$.*

PROOF: It suffices to prove the assertion for $i = j-1$. In order to do so notice that

$$H_j^c = \left[\begin{array}{c|c} \begin{array}{c} \\ H_{j-1}^c \\ \\ \end{array} & \begin{array}{c} 0 \\ \vdots \\ 0 \end{array} \\ \hline H_j \; H_{j-1} \; \cdots \; H_1 & H_0 \end{array}\right]$$

and assume that one of the first $n$ columns of $H_{j-1}^c$ is in the span of some other $(n-k)j-1$ columns. Then $\operatorname{rank} H_0 = n-k$ implies that one of the first $n$ columns of $H_j^c$ is in the span of some other $(n-k)j-1+n-k = (n-k)(j+1)-1$ columns of $H_j^c$. But this is a contradiction to the optimality of $d_j^c$ by Proposition 2.2. □

The Singleton-bound for block codes has been generalized to convolutional codes in [22]. Therein the following has been shown.

**Theorem 2.5** *The free distance of an $(n,k,\delta)$-code satisfies*

$$d_{\text{free}} \leq (n-k)\Big(\Big\lfloor \frac{\delta}{k} \Big\rfloor + 1\Big) + \delta + 1. \tag{2.9}$$

The number appearing on the right in (2.9) is called the *generalized Singleton bound*. The code is called an MDS code if it satisfies $d_{\text{free}} = (n-k)\big(\lfloor \frac{\delta}{k} \rfloor + 1\big) + \delta + 1$. It has been shown in [22] that for every set of parameters $(n,k,\delta)$ and every prime number $p$ there exists a suitably large finite field $\mathbb{F}$ of characteristic $p$ and an MDS code with parameters $(n,k,\delta)$ over $\mathbb{F}$.

The generalized Singleton bound reduces to the usual Singleton bound $n-k+1$ when $\delta = 0$, the block code situation.

The proof of the existence of MDS codes given in [22] is based on techniques from algebraic geometry and is non-constructive. In [26] a construction of MDS codes with parameters $(n,k,\delta)$ was given for suitably large fields of characteristic coprime with $n$.

In the sequel we will strengthen the MDS property by requiring that the generalized Singleton bound is attained by the earliest column distance possible. This will lead to the notion of a strongly MDS code.

**Proposition 2.6** *Suppose $\mathcal{C}$ be an MDS code with parameters $(n,k,\delta)$, column distances $d_j^c$, $j \in \mathbb{N}_0$, and free distance $d_{\text{free}}$. Let $M := \min\{j \in \mathbb{N}_0 \mid d_j^c = d_{\text{free}}\}$. Then*

$$M \geq \Big\lfloor \frac{\delta}{k} \Big\rfloor + \Big\lceil \frac{\delta}{n-k} \Big\rceil.$$

PROOF: From Proposition 2.3 we get

$$d_{\text{free}} = (n-k)\Big(\Big\lfloor\frac{\delta}{k}\Big\rfloor + 1\Big) + \delta + 1 = d^c_M \le (n-k)(M+1)+1. \tag{2.10}$$

This yields the assertion. □

The proof also shows that in the case $j > \lfloor\frac{\delta}{k}\rfloor + \lceil\frac{\delta}{n-k}\rceil$ the column distance $d^c_j$ never attains the upper bound $(n-k)(j+1)+1$ of Proposition 2.3, see also (2.8).

**Definition 2.7** An $(n,k,\delta)$-code with column distances $d^c_j$, $j \in \mathbb{N}_0$, is called *strongly* MDS, if

$$d^c_M = (n-k)\Big(\Big\lfloor\frac{\delta}{k}\Big\rfloor + 1\Big) + \delta + 1 \text{ for } M = \Big\lfloor\frac{\delta}{k}\Big\rfloor + \Big\lceil\frac{\delta}{n-k}\Big\rceil.$$

Because of (2.8) the strong MDS property implies the MDS property.

**Remark 2.8** In the case where $(n-k) \mid \delta$, the strong MDS property implies that $d^c_M$ attains the upper bound $(n-k)(M+1)+1$, see Proposition 2.2. Hence Corollary 2.4 implies that in this case *all* column distances attain their optimal value.
If $(n-k) \nmid \delta$, we always have $d^c_M < (n-k)(M+1)+1$ as can be seen from (2.10).

Even when $(n-k) \nmid \delta$ it is very desirable that the column distance profile $d^c_0, d^c_1, d^c_1, \ldots$ has the maximum possible increase at each step. This motivates the following definition.

**Definition 2.9** Let

$$L := \Big\lfloor\frac{\delta}{k}\Big\rfloor + \Big\lfloor\frac{\delta}{n-k}\Big\rfloor. \tag{2.11}$$

An $(n,k,\delta)$-code with column distances $d^c_j$, $j \in \mathbb{N}_0$, is said to have a *maximum distance profile* if

$$d^c_j = (n-k)(j+1)+1, \quad \text{for} \quad j = 1, \ldots, L.$$

Using the notation of Definition 2.7 we have

$$L = \begin{cases} M & \text{if } (n-k) \mid \delta \\ M-1 & \text{otherwise.} \end{cases} \tag{2.12}$$

An immediate consequence of Corollary 2.4 is

**Lemma 2.10** *An $(n,k,\delta)$-code has a maximum distance profile if and only if the $L$th column distance satisfies*

$$d^c_L = (n-k)(L+1)+1.$$

As a consequence we obtain that if $n-k$ divides $\delta$ then an $(n,k,\delta)$-code has maximum distance profile if and only if it is strongly MDS since $(n-k)\big(\lfloor\frac{\delta}{k}\rfloor + \frac{\delta}{n-k} + 1\big) + 1 = (n-k)\big(\lfloor\frac{\delta}{k}\rfloor + 1\big) + \delta + 1$.

**Remark 2.11** The concept is clearly related to the notion of *optimum distance profile* (ODP), see [9, p. 112]. For ODP it is required that the column distances are maximal up to the memory $\nu$. Hence if $\nu \le L$ then a code with maximum distance profile is always ODP. In general one expects a good code to have generic Forney indices, i.e. the indices attain only the two values $\lceil \frac{\delta}{k} \rceil$ and $\lfloor \frac{\delta}{k} \rfloor$. McEliece [15, Corollary 4.3] calls such codes *compact codes*. It has been shown in [22] that an MDS code has always generic indices. Of course if the indices are generic then $\nu = \lceil \frac{\delta}{k} \rceil$ and thus $\nu \le L+1$.

The notion of ODP seems also to be dependent on the base field which is usually assumed to be the binary field. A code with maximum distance profile does in general not exist over the binary field and it can only exist for sufficiently large base fields. This is similar to the situation of MDS block codes. Such codes are known to exist as soon as the field size of $\mathbb{F}$ is larger than the block length $n$.

One of the main results of Section 5 will show that a convolutional code has a maximum distance profile if and only if its dual has this property. The following algebraic criterion which characterizes codes having a maximum distance profile will be very useful.

**Theorem 2.12** *Let $G = \sum_{j=0}^{\nu} G_j D^j$ be the generator matrix of an $(n, k, \delta)$-code. Let $L$ be defined as in (2.11) and let*

$$G_L^c \;=\; \begin{bmatrix} G_0 & G_1 & \dots & G_L \\ & G_0 & \dots & G_{L-1} \\ & & \ddots & \vdots \\ & & & G_0 \end{bmatrix} \in \mathbb{F}^{(L+1)k\times(L+1)n}. \tag{2.13}$$

*Then $G$ represents a maximum distance profile code if and only if every $(L+1)k \times (L+1)k$ full-size minor formed from the columns with indices $1 \le j_1 < \ldots < j_{(L+1)k}$, where $j_{sk+1} > sn$ for $s = 1, \ldots, L$, is nonzero.*

Proof: Assume there are indices $1 \le j_1 < \ldots < j_{(L+1)k}$ satisfying $j_{sk+1} > sn$ for $s = 1, \ldots, L$ whose corresponding minor is zero. It follows that there is a vector $u = (u_0, \ldots, u_L)$ such that $uG_L^c$ has zero coordinates at positions $j_1, \ldots, j_{(L+1)k}$. Let $\ell := \min\{i \mid u_i \neq 0\}$. Consider the vector

$$(u_\ell, \ldots, u_L)\, G_{L-\ell}^c \in \mathbb{F}^{(L-\ell+1)n}.$$

The weight of this vector is at most $(L-\ell+1)(n-k)$ as there are at least $(L-\ell+1)k$ coordinates zero. It follows from (2.6) that $d_{L-\ell}^c \le (L-\ell+1)(n-k)$ and by Corollary 2.4 the code has not a maximum distance profile.

Vice versa assume that $\mathcal{C}$ has not a maximum distance profile. Let $m := \min\{i \mid d_i^c \le (n-k)(i+1)\}$. It follows that there is a vector $u = (u_0, \ldots, u_m)$, $u_0 \neq 0$ such that $uG_m^c$ has at least $k(m+1)$ zeros. As a submatrix inside $G_L^c$ we select the columns corresponding to the first $k(m+1)$ positions where $uG_m^c$ has a zero and we augment it by the last $k(L-m)$

columns of $G_L^c$. We call the indices of the selected columns $j_1, \ldots, j_{(L+1)k}$. This gives an $(L+1)k \times (L+1)k$ full-size minor and we claim that this minor is zero and that the indices $j_1, \ldots, j_{(L+1)k}$ satisfy $j_{sk+1} > sn$ for $s = 1, \ldots, L$. In order to prove the latter note that $d_i^c = (n-k)(i+1)+1$ for $i = 0, \ldots, m-1$. It therefore follows that $(u_0, \ldots, u_i)G_i^c$ has at most $k(i+1)-1$ zeros for $i = 0, \ldots, m-1$. In particular $j_{sk+1} > sn$ for $s = 1, \ldots, m$. Clearly it is also true for $s = m+1, \ldots, L$. It remains to be shown that the minor is zero. For this note that the selected matrix has the form $\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$ where $A$ is an $(m+1)k \times (m+1)k$ submatrix of $G_m^c$ which is singular by construction. The fullsize minor is therefore zero. $\square$

# 3 Existence of Strongly MDS $(n, n-1, \delta)$-Codes

During his investigation of algebraic decoding of convolutional codes B. Allen conjectured in his dissertation [1] the existence of strongly MDS convolutional codes in the situation when $k = 1$ and $n = 2$. In this section we will show the existence of strongly MDS codes with parameters $(n, n-1, \delta)$. It follows from Equation (2.12) and Lemma 2.10 that these codes also have maximum distance profile. By Theorem 2.5 the generalized Singleton bound for these parameters is given by $\lfloor \frac{\delta}{n-1} \rfloor + \delta + 2$. Thus, Definition 2.7 yields that we have to find an $(n, n-1, \delta)$-code such that $d_M^c = M + 2$, where $M = \lfloor \frac{\delta}{n-1} \rfloor + \delta$. In order to do so, let

$$H = [a_1, \ldots, a_n] \in \mathbb{F}[D]^n, \text{ where } a_i = \sum_{j=0}^{\delta} a_{ij} D^j \in \mathbb{F}[D], \tag{3.1}$$

be a basic parity check matrix of the desired code. Without loss of generality we may assume $a_{10} = 1$. The strong MDS property can now be expressed as follows.

**Theorem 3.1** *Let $H \in \mathbb{F}[D]^n$ be as in (3.1), let $a_{10} = 1$ and define $\mathcal{C} := \{v \in \mathbb{F}((D))^n \mid vH^{\mathsf{T}} = 0\}$ be the code with parity check matrix $H$. Furthermore, for $i = 2, \ldots, n$ let*

$$\frac{a_i}{a_1} = \sum_{j=0}^{\infty} h_{ji} D^j \in \mathbb{F}((D)) \tag{3.2}$$

*be the Laurent expansion of $\frac{a_i}{a_1} \in \mathbb{F}(D)$ and for $M = \lfloor \frac{\delta}{n-1} \rfloor + \delta$ define*

$$\hat{H} := \left[\begin{array}{cccc|ccccccccc} 1 & & & & h_{02} & \cdots & h_{0n} & & & & & & \\ & \ddots & & & h_{12} & \cdots & h_{1n} & h_{02} & \cdots & h_{0n} & & & \\ & & \ddots & & \vdots & & \vdots & \vdots & & \vdots & \ddots & & \\ & & & 1 & h_{M2} & \cdots & h_{Mn} & h_{M-1,2} & \cdots & h_{M-1,n} & \cdots\cdots & h_{02} & \cdots & h_{0n} \end{array}\right] \tag{3.3}$$

$$=: \quad [e_1, \ldots, e_{M+1}, H_{12}, \ldots, H_{1n}, \ldots, H_{M+1,2}, \ldots, H_{M+1,n}] \in \mathbb{F}^{(M+1)\times(M+1)n}, \tag{3.4}$$

*where $e_i$ denotes the $i$-th standard basis vector. We call $\hat{H}$ the $M$-th systematic sliding parity check matrix of $\mathcal{C}$. The following conditions are equivalent:*

*(a) $\mathcal{C}$ is strongly MDS, i. e. $d_M^c = M+2$,*

*(b) none of the columns $H_{12}, \ldots, H_{1n}$ of $\hat{H}$ is contained in the span of any other $M$ columns of $\hat{H}$.*

*Notice that (b) automatically implies that $h_{ji} \neq 0$ for all $i$ and $j$ and that also the first column $e_1$ is not in the span of any other $M$ columns.*

PROOF: After a column permutation the sliding parity check matrix $H_M^c \in \mathbb{F}^{(M+1)\times(M+1)n}$ of $\mathcal{C}$ has the form

$$H' := \left[\begin{array}{cccc|ccccccccccc} 1 & & & & a_{20} & \cdots & a_{n0} & & & & & & & & \\ a_{11} & \ddots & & & a_{21} & \cdots & a_{n1} & a_{20} & \cdots & a_{n0} & & & & & \\ \vdots & \ddots & \ddots & & \vdots & & \vdots & \vdots & & \vdots & \ddots & & & & \\ a_{1M} & \ldots & a_{11} & 1 & a_{2M} & \cdots & a_{nM} & a_{2,M-1} & \cdots & a_{n,M-1} & \cdots\cdots & & a_{20} & \cdots & a_{n0} \end{array}\right].$$

It is straightforward to see that left multiplication of $H'$ by the inverse of the first block leads to the matrix $\hat{H}$ of (3.3). After these transformations Proposition 2.2 applied to the case $j = M$ and $d = M+2$ translates into the equivalence: $\mathcal{C}$ is strongly MDS iff neither the first column $e_1$ nor any of the columns $H_{12}, \ldots, H_{1n}$ is in the span of any other $M$ columns of $\hat{H}$. But this in turn is equivalent to (b) above. □

In order to establish the existence of strongly MDS codes we will proceed as follows. Firstly, we will establish the existence of a systematic sliding parity check matrix $\hat{H}$ as in (3.3) with property (b) of the theorem above. Thereafter, we will show that there exist coprime polynomials $a_1, \ldots, a_n$ having maximum degree equal to $\delta$ such that

$$\frac{a_i}{a_1} = \sum_{j=0}^{M} h_{ji} D^j + \text{ higher powers, } i = 2, \ldots, n.$$

Theorem 3.1 then yields that the code with parity check matrix $H = [a_1, \ldots, a_n]$ is a strongly MDS $(n, n-1, \delta)$-code.

As for the first step, let us have a look at the special case of $(2, 1, \delta)$-codes. In this case $M = 2\delta$ and the systematic sliding parity check matrix in (3.3) has the form

$$\hat{H} := \left[\begin{array}{cccc|cccc} 1 & & & & h_0 & 0 & \cdots & 0 \\ & 1 & & & h_1 & h_0 & \ddots & \vdots \\ & & \ddots & & \vdots & \ddots & \ddots & 0 \\ & & & 1 & h_{2\delta} & \cdots & h_1 & h_0 \end{array}\right] =: [I_{2\delta+1}, T] \in \mathbb{F}^{(2\delta+1)\times(4\delta+2)}, \text{ where } h_j \in \mathbb{F}. \tag{3.5}$$

As we will see, the existence of matrices $T$ of any given size and the structure above such that $\hat{H}$ has the column property of Theorem 3.1(b) will be the main tool for the existence of strongly MDS codes even of length $n > 2$. Therefore we will concentrate on these matrices first. The main point is to express the column condition on $\hat{H}$ in terms of the minors of $T$.

**Definition 3.2** Let $R$ be a ring. For a matrix $T \in R^{n\times k}$ denote by $T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s} \in R^{r\times s}$ the $r \times s$-submatrix obtained from $T$ by picking the rows with indices $i_1, \ldots, i_r$ and the columns with indices $j_1, \ldots, j_s$.

In the sequel the following property will play a crucial role.

**Definition 3.3** Let $\mathbb{F}$ be field. A lower triangular matrix $T \in \mathbb{F}^{n\times k}$ is said to be *superregular*[1], if $T^{i_1,\ldots,i_r}_{j_1,\ldots,j_r}$ is nonsingular for all $1 \le r \le \min\{k, n\}$ and all indices $1 \le i_1 < \ldots < i_r \le n$, $1 \le j_1 < \ldots < j_r \le k$ which satisfy $j_\nu \le i_\nu$ for $\nu = 1, \ldots, r$. We call the submatrices obtained by picking such indices the proper submatrices and their determinants the proper minors of $T$.

**Remark 3.4** Observe that the proper submatrices are the only submatrices which can possibly be nonsingular. This can be seen as follows. If $j_\nu > i_\nu$ for some $\nu$, then in the submatrix $\hat{T} := T^{i_1,\ldots,i_r}_{j_1,\ldots,j_r}$ the upper right block consisting of the first $\nu$ rows and the last $r-\nu+1$ columns is identically zero. Hence the first $\nu$ rows of $\hat{T}$ can have at most rank $\nu - 1$. In other words, the improper submatrices of $T$ are trivially singular. For example, for $T = (h_{ij})$ we have

$$T^{1,2,5}_{1,3,4} = \begin{bmatrix} h_{11} & 0 & 0 \\ h_{21} & 0 & 0 \\ h_{51} & h_{53} & h_{54} \end{bmatrix}.$$

Now we can establish the following.

**Theorem 3.5** *Let $\mathbb{F}$ be a field and $T$ be a lower triangular Toeplitz matrix, i. e.*

$$T = [T_1, \ldots, T_l] = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ h_{l-1} & \cdots & h_1 & h_0 \end{bmatrix} \in \mathbb{F}^{l\times l}. \tag{3.6}$$

*Furthermore, put $\hat{H} := [I_l, T] = [e_1, \ldots, e_l, T_1, \ldots, T_l] \in \mathbb{F}^{l\times 2l}$. Then the following are equivalent:*

*(a) $T$ is superregular, i.e. all proper submatrices in the sense of Definition 3.3 are nonsingular.*

*(b) Assume there are indices $1 \le i_1 < \ldots < i_r \le n$, $1 < j_2 < \ldots < j_r \le k$. Then all proper submatrices of $T$ of the form $T^{i_1,i_2,\ldots,i_r}_{1,j_2,\ldots,j_r}$ are nonsingular,*

[1]We adopt this notion from [24], where it has been coined in a slightly different context.

*(c)* $\mathrm{wt}\big(T_1+\sum_{j=1}^{s}\beta_j T_{m_j}\big)\geq l-s$ *for all* $1\leq s\leq l-1$, *all* $1<m_1<\ldots<m_s\leq l$ *and all* $\beta_1,\ldots,\beta_s\in\mathbb{F}$,

*(d)* $T_1\notin\mathrm{span}\{T_{m_1},\ldots,T_{m_s},e_{l_1},\ldots,e_{l_t}\}$ *where* $1<m_1<\ldots<m_s\leq l$ *and* $1\leq l_1<\ldots<l_t\leq l$ *and* $s+t\leq l-1$.

*(e) If* $v\in\mathbb{F}^{2l}$ *satisfies* $v\hat{H}^{\mathsf{T}}=0$ *and* $v_{l+1}\neq 0$, *then* $\mathrm{wt}(v)\geq l+1$.

*(f)* $e_1\notin\mathrm{span}\{T_{m_1},\ldots,T_{m_s},e_{l_1},\ldots,e_{l_t}\}$ *where* $1\leq m_1<\ldots<m_s\leq l$ *and* $1<l_1<\ldots<l_t\leq l$ *and* $s+t\leq l-1$.

*(g) If* $v\in\mathbb{F}^{2l}$ *satisfies* $v\hat{H}^{\mathsf{T}}=0$ *and* $v_1\neq 0$, *then* $\mathrm{wt}(v)\geq l+1$.

PROOF: (a) ⇔ (b) is obvious since in case of properness the Toeplitz structure implies

$$T^{i_1,\ldots,i_r}_{j_1,\ldots,j_r}=T^{i_1-j_1+1,\ldots,i_r-j_1+1}_{j_1-j_1+1,\ldots,j_r-j_1+1}.$$

(b) ⇒ (c): Let $\hat{h}:=T_1+\sum_{j=1}^{s}\beta_j T_{m_j}$ and assume to the contrary $\mathrm{wt}(\hat{h})<l-s$. The assumption implies that $\hat{h}$ consists of at least $s+1$ zero entries, say at the positions $i_1,\ldots,i_{s+1}$. Then

$$T^{i_1,\ldots,i_{s+1}}_{1,m_1,\ldots,m_s}\begin{pmatrix}1\\ \beta_1\\ \vdots\\ \beta_s\end{pmatrix}=\begin{pmatrix}0\\0\\ \vdots\\0\end{pmatrix}. \tag{3.7}$$

The superregularity yields $m_\nu>i_{\nu+1}$ for some $\nu\in\{1,\ldots,s\}$, which we can choose to be minimal with this property. Then the submatrix $T^{i_1,\ldots,i_{\nu+1}}_{m_\nu,\ldots,m_s}$ is identically zero and therefore we obtain from (3.7) the identity $T^{i_1,\ldots,i_\nu}_{1,m_1,\ldots,m_{\nu-1}}(1,\beta_1,\ldots,\beta_{\nu-1})^{\mathsf{T}}=0$, a contradiction to superregularity since by minimality of $\nu$ this coefficient matrix is nonsingular.
(c) ⇒ (b): Assume to the contrary that $\det T^{i_1,\ldots,i_{s+1}}_{1,m_1,\ldots,m_s}=0$ for some indices satisfying $m_\nu\leq i_{\nu+1}$ for $\nu=1,\ldots,s$. We can assume $s$ to be minimal with this property. Then there exists $(\beta_0,\beta_1,\ldots,\beta_s)\in\mathbb{F}^{s+1}\backslash\{0\}$ such that $T^{i_1,\ldots,i_{s+1}}_{1,m_1,\ldots,m_s}(\beta_0,\ldots,\beta_s)^{\mathsf{T}}=0$. Minimality of $s$ and the equivalence of (a) and (b) imply $\beta_0\neq 0$. Hence we can take $\beta_0=1$ and (3.7) is satisfied. Thus $\mathrm{wt}(T_1+\sum_{j=1}^{s}\beta_j T_{m_j})\leq l-(s+1)$, a contradiction.
The properties (d) and (e) are simply reformulations of (c).
The equivalence (d) ⇔ (f) is clear from the structure of $\hat{H}$ (a linear combination of $T_1$ by the other columns of $\hat{H}$ has to involve the column $e_1$ and vice versa).
The property (g) is a reformulation of (f). □

The equivalence of (e) and (g) immediately implies

**Corollary 3.6** *If* $T\in\mathbb{F}^{l\times l}$ *is a superregular lower triangular Toeplitz matrix, then so is* $T^{-1}$.

The following lemma is the main step for establishing the existence of superregular matrices of Toeplitz-structure.

**Lemma 3.7** *Let $\mathbb{F}$ be a field and $X_1, \dots, X_l$ be independent indeterminates over $\mathbb{F}$. Define the matrix*

$$A := \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ X_2 & X_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ X_l & \cdots & X_2 & X_1 \end{bmatrix} \in \mathbb{F}(X_1, \dots, X_l)^{l \times l}.$$

*Then $A$ is superregular.*

Proof: We proceed by contradiction. Assume there exists a singular proper submatrix

$$\hat{A} := A^{i_1,\dots,i_r}_{j_1,\dots,j_r}.$$

We can take the size $r$ to be minimal. Then certainly $r > 1$. By properness we know that $j_\nu \le i_\nu$ for $\nu = 1, \dots, r$.
Notice that for $\mu \le \nu$ the entry of $A$ at the position $(\nu, \mu)$ is given by $A^\nu_\mu = X_{\nu-\mu+1}$. Hence the indeterminate with the largest index appearing in $\hat{A}$ is $X_{i_r - j_1 + 1}$. It appears only once in the matrix and that is in the lower left corner. Thus its coefficient in $\det \hat{A}$ is $\pm \det \tilde{A}$, where

$$\tilde{A} := A^{i_1,\dots,i_{r-1}}_{j_2,\dots,j_r}.$$

Singularity of $A$ now implies $\det \tilde{A} = 0$. By minimality of $r$ this yields that $\tilde{A}$ is an improper submatrix of $A$, i. e. there exists an index $\tau \in \{2, \dots, r\}$ such that $j_\tau > i_{\tau-1}$. Picking $\tau$ minimal we get $i_1 < \ldots < i_{\tau-1} < j_\tau < \ldots < j_r$ and therefore the first $\tau - 1$ rows of $\hat{A}$ have the form

$$\begin{bmatrix} * & \cdots & * & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ * & \cdots & * & 0 & \cdots & 0 \end{bmatrix},$$

where the block of possibly nonzero elements consists of $\tau - 1$ columns. Hence $\hat{A}$ is a blocktriangular matrix and we have

$$0 = \det \hat{A} = \det A^{i_1,\dots,i_{\tau-1}}_{j_1,\dots,j_{\tau-1}} \det A^{i_\tau,\dots,i_r}_{j_\tau,\dots,j_r}.$$

Since both factors are proper minors we get a contradiction to the minimality of the size $r$.
□

The following consequence is standard.

**Theorem 3.8** *For every $l \in \mathbb{N}$ and every prime number $p$ there exists a finite field $\mathbb{F}$ of characteristic $p$ and a superregular matrix $T \in \mathbb{F}^{l \times l}$ having Toeplitz structure.*

Proof: Consider the prime field $\mathbb{F}_p$ and the matrix of the previous lemma with entries in $\mathbb{F}_p(X_1, \dots, X_l)$. All its proper minors are nonzero polynomials in $\mathbb{F}_p[X_1, \dots, X_l]$. Over an algebraic closure $\bar{\mathbb{F}}_p$ a point $a := (a_1, \dots, a_l) \in \bar{\mathbb{F}}_p^l$ can be found such that none of the minors vanishes at $a$. Hence the Toeplitz matrix $T$ having $(a_1, \ldots, a_l)^\mathsf{T}$ as its first column is superregular. Since each $a_i$ is algebraic over $\mathbb{F}_p$, the matrix $T$ has its entries in a finite field extension $\mathbb{F}$ of $\mathbb{F}_p$. □

In particular, for every size $l \in \mathbb{N}$ there exist superregular Toeplitz matrices over a field of characteristic 2. Unfortunately, the theorem above is nonconstructive and it is not at all clear what the minimum field of characteristic 2 is to allow a superregular Toeplitz matrix of given size $l \times l$. We present some examples.

**Example 3.9** (1) Using a computer algebra program one checks that the following matrices are superregular. The first examples are all over prime fields $\mathbb{F}_p$.

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \in \mathbb{F}_2^{2\times 2}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix} \in \mathbb{F}_3^{3\times 3}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \end{bmatrix} \in \mathbb{F}_5^{4\times 4}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 6 & 1 & 2 & 1 & 0 \\ 4 & 6 & 1 & 2 & 1 \end{bmatrix} \in \mathbb{F}_7^{5\times 5},$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 & 0 \\ 3 & 1 & 1 & 2 & 1 & 0 \\ 4 & 3 & 1 & 1 & 2 & 1 \end{bmatrix} \in \mathbb{F}_{11}^{6\times 6}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7 & 1 & 0 & 0 & 0 & 0 & 0 \\ 13 & 7 & 1 & 0 & 0 & 0 & 0 \\ 2 & 13 & 7 & 1 & 0 & 0 & 0 \\ 1 & 2 & 13 & 7 & 1 & 0 & 0 \\ 4 & 1 & 2 & 13 & 7 & 1 & 0 \\ 14 & 4 & 1 & 2 & 13 & 7 & 1 \end{bmatrix} \in \mathbb{F}_{17}^{7\times 7}.$$

The following examples represent superregular matrices over finite fields of characteristic 2. For this assume that $\alpha, \beta$ and $\gamma$ satisfy

$$\alpha^2 + \alpha + 1 = 0, \ \beta^3 + \beta + 1 = 0, \ \text{and } \gamma^4 + \gamma + 1 = 0.$$

Then the following matrices represent superregular matrices over $\mathbb{F}_4$, $\mathbb{F}_8$ and $\mathbb{F}_{16}$ respectively.

$$\begin{bmatrix} 1 & & \\ \alpha & 1 & \\ 1 & \alpha & 1 \end{bmatrix} \in \mathbb{F}_{2^2}^{3\times 3}, \quad \begin{bmatrix} 1 & & & & \\ \beta & 1 & & & \\ \beta^3 & \beta & 1 & & \\ \beta & \beta^3 & \beta & 1 & \\ 1 & \beta & \beta^3 & \beta & 1 \end{bmatrix} \in \mathbb{F}_{2^3}^{5\times 5}, \quad \begin{bmatrix} 1 & & & & & \\ \gamma & 1 & & & & \\ \gamma^5 & \gamma & 1 & & & \\ \gamma^5 & \gamma^5 & \gamma & 1 & & \\ \gamma & \gamma^5 & \gamma^5 & \gamma & 1 & \\ 1 & \gamma & \gamma^5 & \gamma^5 & \gamma & 1 \end{bmatrix} \in \mathbb{F}_{2^4}^{6\times 6}.$$

Assume $\epsilon, \omega$ satisfy

$$\epsilon^5 + \epsilon^2 + 1 = 0 \text{ and } \omega^6 + \omega + 1 = 0.$$

Then the following matrices represent superregular matrices over $\mathbb{F}_{32}$ and $\mathbb{F}_{64}$ respectively.

$$\begin{bmatrix} 1 & & & & & & \\ \epsilon & 1 & & & & & \\ \epsilon^6 & \epsilon & 1 & & & & \\ \epsilon^9 & \epsilon^6 & \epsilon & 1 & & & \\ \epsilon^6 & \epsilon^9 & \epsilon^6 & \epsilon & 1 & & \\ \epsilon & \epsilon^6 & \epsilon^9 & \epsilon^6 & \epsilon & 1 & \\ 1 & \epsilon & \epsilon^6 & \epsilon^9 & \epsilon^6 & \epsilon & 1 \end{bmatrix} \in \mathbb{F}_{2^5}^{7\times 7}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \omega & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \omega^9 & \omega & 1 & 0 & 0 & 0 & 0 & 0 \\ \omega^{33} & \omega^9 & \omega & 1 & 0 & 0 & 0 & 0 \\ \omega^{33} & \omega^{33} & \omega^9 & \omega & 1 & 0 & 0 & 0 \\ \omega^9 & \omega^{33} & \omega^{33} & \omega^9 & \omega & 1 & 0 & 0 \\ \omega & \omega^9 & \omega^{33} & \omega^{33} & \omega^9 & \omega & 1 & 0 \\ 1 & \omega & \omega^9 & \omega^{33} & \omega^{33} & \omega^9 & \omega & 1 \end{bmatrix} \in \mathbb{F}_{2^6}^{8\times 8}.$$

Notice that the matrices above have even more symmetry than required. One can easily show that there is no superregular $4 \times 4$-matrix over $\mathbb{F}_4$ of general Toeplitz structure. However, the above suggests to ask whether one can find for every $l \geq 5$ a superregular $l \times l$-Toeplitz matrix over $\mathbb{F}_{2^{l-2}}$.

(2) In the appendix we prove that for every $n \in \mathbb{N}$ the proper minors of the Toeplitz-matrix

$$T_n := \begin{bmatrix} \binom{n-1}{0} & & & \\ \binom{n-1}{1} & \binom{n-1}{0} & & \\ \vdots & \ddots & \ddots & \\ \binom{n-1}{n-1} & \cdots & \binom{n-1}{1} & \binom{n-1}{0} \end{bmatrix} \in \mathbb{Z}^{n\times n}$$

are all positive. Hence for each $n \in \mathbb{N}$ there exists a smallest prime number $p_n$ such that $T_n$ is superregular over the prime field $\mathbb{F}_{p_n}$. One can check that

$$p_2 = 2,\ p_3 = 5,\ p_4 = 7,\ p_5 = 11,\ p_6 = 23,\ p_7 = 43.$$

Now we can establish the existence of strongly MDS codes in the following sense.

**Theorem 3.10** *For every $n,\ \delta \in \mathbb{N}$ and every prime number $p$ there exists a strongly MDS code with parameters $(n, n-1, \delta)$ over a suitably large field of characteristic $p$.*

The proof of this theorem is rather long and technical and because of this reason it is put into the appendix.

There is of course the natural question if strongly MDS convolutional codes and codes with maximum distance profile exist for all parameters $(n, k, \delta)$. We strongly believe so. The section showed that such codes exist for all parameters $(n, k, \delta)$ with $k = n-1$. For all small values of $(n, k, \delta)$ we have found strongly MDS convolutional codes and codes with maximum distance profile making computer searches. In the next section we present a series of examples of such codes found through computer searches. Based on this wealth of data we conjecture:

**Conjecture 3.11** *For all $n > k > 0$ and for all $\delta \geq 0$ there exists an $(n, k, \delta)$ code over a sufficiently large field which is both strongly MDS and has a maximum distance profile.*

# 4 Examples

In this section we will present some examples of strongly MDS codes with small parameters. The first set of examples is constructed according to the proof of Theorem 3.10 by utilizing the superregular matrices in Example 3.9.

**Example 4.1** Recall the first part of the proof of Theorem 3.10.

(1) We can construct strongly MDS $(2, 1, \delta)$-codes once a $\tau \times \tau$ superregular matrix, where $\tau = 2\delta + 1$, is available. Thus, the $5 \times 5$ and $7 \times 7$ matrices given in Example 3.9(1) lead to the strongly MDS $(2, 1, 2)$-code over $\mathbb{F}_8$ (where $\beta^3 + \beta + 1 = 0$) with parity check matrix

$$H = [a,\, b] = [1 + \beta^2 D + \beta^5 D^2,\, 1 + \beta^4 D + \beta^5 D^2] \in \mathbb{F}_8[D]^2$$

and to the strongly MDS $(2, 1, 3)$-code over $\mathbb{F}_{32}$ (where $\epsilon^5 + \epsilon^2 + 1 = 0$) with parity check matrix

$$H = [a,\, b] = [1 + \epsilon^{18} D + \epsilon^{11} D^2 + \epsilon^{29} D^3,\, 1 + D + \epsilon^{27} D^2 + \epsilon^{18} D^3] \in \mathbb{F}_{32}[D]^2.$$

Indeed, one checks that

$$\frac{1 + \beta^4 D + \beta^5 D^2}{1 + \beta^2 D + \beta^5 D^2} = 1 + \beta D + \beta^3 D^2 + \beta D^3 + D^4 + \text{ higher powers}$$

and

$$\frac{1 + D + \epsilon^{27} D^2 + \epsilon^{18} D^3}{1 + \epsilon^{18} D + \epsilon^{11} D^2 + \epsilon^{29} D^3} = 1 + \epsilon D + \epsilon^6 D^2 + \epsilon^9 D^3 + \epsilon^6 D^4 + \epsilon D^5 + D^6 + \text{higher powers.}$$

Hence the free distance of the two codes above is 6 (resp. 8), and this is also the 4th (resp. 6th) column distance.

(2) Using the $8 \times 8$-superregular matrix of Example 3.9(1), one can construct a strongly MDS $(3, 2, 2)$-code over $\mathbb{F}_{64}$. Hence the code has free distance equal to its 3rd column distance, and this value is 5. Using the construction of the proof of Theorem 3.10 and going through some tedious calculations in the field $\mathbb{F}_{64}$ (where $\omega^6 + \omega + 1 = 0$) one finally arrives at the parity check matrix

$$H = [1 + \omega^{57} D + \omega^{62} D^2,\, \omega + \omega^{44} D + \omega^{54} D^2,\, 1 + \omega^{17} D + \omega^{21} D^2] \in \mathbb{F}_{64}^3.$$

(3) A strongly MDS $(4, 3, 1)$-code has free distance 3 and this is identical with the first column distance. It can be obtained from a $6 \times 6$-superregular matrix using the construction of the proof of Theorem 3.10. Indeed, the matrix

$$\hat{H} = \left[\begin{array}{cc|cccccc} 1 & 0 & \gamma^5 & \gamma & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & \gamma & \gamma^5 & \gamma^5 & \gamma & 1 \end{array}\right]$$

has been obtained from the superregular Toeplitz matrix of Example 3.9(1) and thus it satisfies property (b) of Theorem 3.1. Hence a parity check matrix of a strongly MDS $(4,3,1)$-code over $\mathbb{F}_{16}$ (where $\gamma^4+\gamma+1=0$) is given by

$$H = [1, \gamma^5 + D, \gamma + \gamma D, 1 + \gamma^5 D] \in \mathbb{F}_{16}[D]^4.$$

(4) Of course, not every MDS code is strongly MDS. For instance, the code with parity check matrix $H = [10 + 3D + 2D^2, 4 + 2D + D^2] \in \mathbb{F}_{11}[D]^2$ is an MDS code, but not strongly MDS. In this example, the MDS property follows from the fact, that this code is the result of the construction of MDS codes as presented in [26]. However, a $(2,1,1)$-code is strongly MDS iff it is an MDS code. This can be checked directly by using Theorem 3.1 and the fact that for the (basic) parity check matrix $[a_0 + a_1 D, b_0 + b_1 D]$ of an MDS code all coefficients as well as $a_0 b_1 - a_1 b_0$ are nonzero.

The next series of examples has been found by completely different methods. They are all cyclic convolutional codes in the sense of [4, 5, 16, 19]. In those papers convolutional codes having some additional algebraic structure are being investigated. This additional structure is a generalization of cyclicity of block codes but is a far more complex notion for convolutional codes. In particular cyclicity of convolutional codes does *not* mean invariance under the cyclic shift in $\mathbb{F}((D))^n$. We will not go into the details but rather refer to [4, 5]. However, in order to understand and test the following examples there is no need in understanding the concept of cyclicity for convolutional codes since below we provide all information needed to specify the codes. We present the generator matrices and also provide all column distances; they have been computed with a computer algebra program. All matrices given below are minimal basic in the sense of Definition 1.1. We would like to mention that just like for cyclic block codes, the length of the code and the characteristic of the field have to be coprime. Therefore, only codes with odd length are given below.

One should note that most of the following codes exist over comparatively smaller alphabets than the examples of 4.1. However, we don't know any general construction for strongly MDS cyclic convolutional codes yet. But the abundance of (small) examples suggests that such a construction might be possible and might lead to smaller alphabets for given parameters than the construction of the last section. We will leave this as an open question for future research.

**Example 4.2** (1) A strongly MDS $(3,1,1)$-code over $\mathbb{F}_4$:

$$G = [\alpha + \alpha D,\ \alpha^2 + \alpha D,\ 1 + \alpha D].$$

The column distances are $d_0^c = 3$, $d_1^c = 5$, $d_j^c = 6$ for $j \geq 2$.

(2) A strongly MDS $(3,1,2)$-code over $\mathbb{F}_{16}$ (where $\beta^4 + \beta + 1 = 0$):

$$G = [\beta + \beta D + D^2,\ \beta^6 + \beta D + \beta^{10} D^2,\ \beta^{11} + \beta D + \beta^5 D^2].$$

The column distances are $d_0^c = 3$, $d_1^c = 5$, $d_2^c = 7$, $d_j^c = 9$ for $j \geq 3$.

(3) A strongly MDS $(3,2,2)$-code over $\mathbb{F}_{16}$:

$$G = \begin{bmatrix} \beta^5+\beta^4 D & \beta^3+\beta^8 D & \beta^9+\beta^2 D \\ \beta^9+\beta^{12} D & \beta^5+\beta^{14} D & \beta^3+\beta^3 D \end{bmatrix}.$$

The column distances are $d_0^c = 2$, $d_1^c = 3$, $d_2^c = 4$, $d_j^c = 5$ for $j \geq 3$.

(4) A strongly MDS $(5,1,1)$-code over $\mathbb{F}_{16}$:

$$G = [\beta+\beta D,\ \beta^{13}+\beta^{10} D,\ \beta^{10}+\beta^4 D,\ \beta^7+\beta^{13} D,\ \beta^4+\beta^7 D].$$

The column distances are $d_0^c = 5$, $d_1^c = 9$, $d_j^c = 10$ for $j \geq 2$.

(5) A strongly MDS $(5,1,2)$-code over $\mathbb{F}_{16}$:

$$G = [\beta+\beta^4 D+\beta D^2,\ \beta^7+\beta D+\beta^{10} D^2,\ \beta^{13}+\beta^{13} D+\beta^4 D^2,\ \beta^4+\beta^{10} D+\beta^{13} D^2,\ \beta^{10}+\beta^7 D+\beta^7 D^2].$$

The column distances are $d_0^c = 5$, $d_1^c = 9$, $d_2^c = 13$, $d_j^c = 15$ for $j \geq 3$.

(6) A strongly MDS $(5,2,2)$-code over $\mathbb{F}_{16}$:

$$G = \begin{bmatrix} \beta+\beta D & \beta^{13}+\beta^{10} D & \beta^{10}+\beta^4 D & \beta^7+\beta^{13} D & \beta^4+\beta^7 D \\ 1+\beta^5 D & \beta^3+\beta^{11} D & \beta^6+\beta^2 D & \beta^9+\beta^8 D & \beta^{12}+\beta^{14} D \end{bmatrix}.$$

The column distances are $d_0^c = 4$, $d_1^c = 7$, $d_j^c = 9$ for $j \geq 2$.

(7) A strongly MDS $(7,1,1)$-code over $\mathbb{F}_8$ (where $\gamma^3+\gamma+1=0$):

$$G = [\gamma+\gamma D,\ \gamma^3+D,\ \gamma^5+\gamma^6 D,\ 1+\gamma^5 D,\ \gamma^2+\gamma^4 D,\ \gamma^4+\gamma^3 D,\ \gamma^6+\gamma^2 D].$$

The column distances are $d_0^c = 7$, $d_1^c = 13$, $d_j^c = 14$ for $j \geq 2$.

(8) A strongly MDS $(7,1,2)$-code over $\mathbb{F}_8$:

$$G = [\gamma^2+\gamma D+D^2,\ \gamma^5+\gamma^3 D+\gamma^6 D^2,\ \gamma+\gamma^5 D+\gamma^5 D^2,\ \gamma^4+D+\gamma^4 D^2,\ 1+\gamma^2 D+\gamma^3 D^2,\ \gamma^3+\gamma^4 D+\gamma^2 D^2,\ \gamma^6+\gamma^6 D+\gamma D^2].$$

The column distances are $d_0^c = 7$, $d_1^c = 13$, $d_2^c = 18$, $d_j^c = 21$ for $j \geq 3$.

(9) It is worth being mentioned that there does not exist even an MDS $(7,2,2)$-code over $\mathbb{F}_8$, since the generalized Singleton bound in this case is 13, but due to the Griesmer bound (see [9, p. 133] for the binary case) the parameters of an $(n,k,\delta)$-code over $\mathbb{F}_q$ with memory $m$ and distance $d$ satisfy

$$\sum_{l=0}^{k(m+i)-\delta-1} \left\lceil \frac{d}{q^l} \right\rceil \leq n(m+i) \text{ for all } i \in \mathbb{N}_0.$$

Hence a $(7,2,2)$-code over $\mathbb{F}_8$ with memory 1 has at most distance 12. The inequality applied to $i=1$ shows that the field size has to be at least 13 in order to allow the existence of an MDS $(7,2,2)$-code.

One should notice that the codes in Example 4.2(1) – (7) are not only strongly MDS but also have *all* column distances being optimal in the sense that they reach the upper bound given in Proposition 2.3. In particular they also have a maximum distance profile in the sense of Definition 2.9. For the $(7,1,2)$-code in (8), only the second column distance is not optimal, but rather one less than the upper bound, which is 19 in this case.

# 5 The Dual of a Strongly MDS Code

In this section we will present some results concerning the dual code of a strongly MDS code. The main result shows that a convolutional code has a maximum distance profile if and only if its dual has this property. This then implies for certain parameters that a code is strongly MDS if and only if its dual has this property. These results are very appealing as it generalizes the situation for block codes.

Recall that if

$$\mathcal{C} = \{uG \mid u \in \mathbb{F}(\!(D)\!)^k\} = \{v \in \mathbb{F}(\!(D)\!)^n \mid vH^{\mathsf{T}} = 0\} \subseteq \mathbb{F}(\!(D)\!)^n$$

is an $(n,k,\delta)$-code with generator matrix $G \in \mathbb{F}[D]^{k\times n}$ and parity check matrix $H \in \mathbb{F}[D]^{(n-k)\times n}$, then the dual code, defined as

$$\mathcal{C}^{\perp} = \{w \in \mathbb{F}(\!(D)\!)^n \mid wv^{\mathsf{T}} = 0 \text{ for all } v \in \mathcal{C}\},$$

is given by

$$\mathcal{C}^{\perp} = \{uH \mid u \in \mathbb{F}(\!(D)\!)^{n-k}\} = \{w \in \mathbb{F}(\!(D)\!)^n \mid wG^{\mathsf{T}} = 0\}$$

and thus an $(n, n-k, \delta)$-code. In contrast to the block code situation almost nothing is known about the relation between the distances of a code and its dual. In particular, it has been shown in [25] that no MacWilliams identity relating the weight distributions of $\mathcal{C}$ and $\mathcal{C}^{\perp}$ exists. In block code theory a very simple relation between the distances of a code and its dual is given in the case of MDS codes. In fact, if $\mathcal{C}$ is an MDS $(n,k)$-block-code, then the dual $\mathcal{C}^{\perp}$ is an MDS $(n,n-k)$-code, see [14, Ch. 11, §2] and very specific knowledge on the weight enumerator and its dual is known [14, Ch. 11]. Therefore, it is quite natural to investigate whether the dual of an MDS (or strongly MDS) convolutional code is MDS (or strongly MDS), too. Unfortunately, this is in general not the case.

**Example 5.1** In general the dual of a strongly MDS code is not even an MDS code. This can be seen from the dual of the code given in Example 4.1(3). The dual has generator matrix $G = [1, \gamma^5 + D, \gamma + \gamma D, 1 + \gamma^5 D] \in \mathbb{F}_{16}[D]^4$ which obviously has weight less than the generalized Singleton bound 8 (see Theorem 2.5).

As we will show next the property of maximum distance profile carries over under dualization. In addition, for specific code parameters the strong MDS property carries over to the dual code as well. To this end, recall from Definition 2.7 that an $(n, k, \delta)$-code is strongly MDS if the $M$th column distance attains the generalized Singleton bound where $M = \lfloor \frac{\delta}{k} \rfloor + \lceil \frac{\delta}{n-k} \rceil$. Thus the dual code $\mathcal{C}^\perp$ is MDS if the $\hat{M}$th column distance attains the generalized Singleton bound where $\hat{M} = \lfloor \frac{\delta}{n-k} \rfloor + \lceil \frac{\delta}{k} \rceil$. Obviously, these two numbers differ by one when $k$ divides $\delta$ but $n-k$ does not or vice versa. What remains equal for both the code and its dual is the quantity $L = \left\lfloor \frac{\delta}{k} \right\rfloor + \left\lfloor \frac{\delta}{n-k} \right\rfloor$ used in Definition 2.9 where we introduced the concept of maximum distance profile.

Before we state the main results we need a technical lemma.

**Lemma 5.2** *Let $A \in \mathbb{F}^{k \times n}$ and $B \in \mathbb{F}^{n \times (n-k)}$ such that*

$$AB = 0 \text{ and } \operatorname{rank} A = k,\ \operatorname{rank} B = n-k.$$

*Then the following are equivalent:*

*(a) the $k \times k$-submatrix of $A$ consisting of the columns with indices $1 \leq t_1 < \ldots < t_k \leq n$ is singular,*

*(b) The $(n-k) \times (n-k)$-submatrix of $B$ obtained by taking the rows with indices in $\{1, \ldots, n\} \backslash \{t_1, \ldots, t_k\}$ is singular.*

Proof: Without loss of generality assume $(t_1, \ldots, t_k) = (1, \ldots, k)$ and partition $A = (A_1\ A_2)$, where $A_1$ is the $k \times k$ submatrix under consideration. If $A_1$ is invertible then

$$\ker A = \operatorname{colspan}_{\mathbb{F}} \begin{pmatrix} A_1^{-1} A_2 \\ -I_{n-k} \end{pmatrix} = \operatorname{colspan}_{\mathbb{F}}(B).$$

This shows that the bottom $(n-k) \times (n-k)$-submatrix of $B$ is invertible. □

This lemma, in conjunction with Theorem 2.12 immediately gives an algebraic criterion for maximum distance profile codes in terms of a parity check matrix.

**Theorem 5.3** *Let $H = \sum_{j=0}^{\mu} H_j D^j$ be the parity check matrix of an $(n, k, \delta)$-code. Let $L$ be defined as in (2.11) and let*

$$H_L^c :\ = \begin{bmatrix} H_0 & & & \\ H_1 & H_0 & & \\ \vdots & \vdots & \ddots & \\ H_L & H_{L-1} & \ldots & H_0 \end{bmatrix} \in \mathbb{F}^{(L+1)(n-k) \times (L+1)n}. \tag{5.1}$$

*Then $H$ represents a maximum distance profile code if and only if every $(L+1)(n-k) \times (L+1)(n-k)$ full-size minor formed from the columns with indices $1 \leq i_1 < \ldots < i_{(L+1)(n-k)}$, where $i_{s(n-k)} \leq sn$ for $s = 1, \ldots, L$, is nonzero.*

PROOF: Let the code have generator matrix $G$ as given in (2.1). Recall that $G^c_L(H^c_L)^\mathsf{T} = 0$ and both factors have full rank. By Theorem 2.12 the code has maximum distance profile if and only if every full size minor $G^c_L$ formed from the columns $1 \le j_1 < \ldots < j_{(L+1)k}$, where $j_{sk+1} > sn$ for $s = 1, \ldots, L$, is nonzero. Now the complimentary minors of $H^c_L$ have indices $1 \le i_1 < \ldots < i_{(L+1)(n-k)}$ satisfying $i_{s(n-k)} \le sn$ for $s = 1, \ldots, L$. Thus Lemma 5.2 completes the proof. □

With this we have a nice duality result:

**Theorem 5.4** *An $(n, k, \delta)$-code $\mathcal{C} \subseteq \mathbb{F}((D))^n$ has a maximum distance profile if and only if the dual code $\mathcal{C}^\perp \subseteq \mathbb{F}((D))^n$ has this property.*

PROOF: Let $\mathcal{C}$ have generator matrix $G$ and parity check matrix $H$ as given in (2.1) and (2.2). Assume $\mathcal{C}$ has a maximum distance profile. By Theorem 5.3 every $(L+1)(n-k) \times (L+1)(n-k)$ full-size minor formed from the columns of $H^c_L$ with indices $1 \le i_1 < \ldots < i_{(L+1)(n-k)}$, where $i_{s(n-k)} \le sn$ for $s = 1, \ldots, L$, is nonzero.

Consider now the dual code $\mathcal{C}^\perp$ which is defined as the rowspace of the $(n-k) \times n$ matrix $H$. It follows from (2.6) that the $L$th column distance of the dual code $\mathcal{C}^\perp$ is given by

$$\hat{d}^c_L = \min\big\{\mathrm{wt}\big((u_L, \ldots, u_0)H^c_L\big) \,\big|\, u_i \in \mathbb{F}^{n-k},\, u_0 \ne 0\big\}.$$

Taking the reversed ordering into account we obtain from Theorem 2.12 that the dual code $\mathcal{C}^\perp$ has maximum distance profile as well. □

**Corollary 5.5** *When both $k$ and $n-k$ divide $\delta$ then an $(n, k, \delta)$-code $\mathcal{C} \subseteq \mathbb{F}((D))^n$ is strongly MDS if and only if $\mathcal{C}^\perp \subseteq \mathbb{F}((D))^n$ has this property.*

PROOF: From $k \mid \delta$ and $(n-k) \mid \delta$ it follows that $L = M$ and $d^c_M = (n-k)\big(\frac{\delta}{k}+1\big)+\delta+1$, the generalized Singleton bound of the code $\mathcal{C}$ and $\hat{d}^c_M = k\big(\frac{\delta}{n-k}+1\big)+\delta+1$, the generalized Singleton bound of the dual code $\mathcal{C}^\perp$. □

The result above gives us another class of strongly MDS codes by dualizing Theorem 3.10.

**Corollary 5.6** *For every $n$, $\delta \in \mathbb{N}_0$ such that $(n-1) \mid \delta$ and every prime number $p$ there exists a strongly MDS $(n, 1, \delta)$-code over some suitably large field of characteristic $p$.*

**Example 5.7** (a) Corollary 5.5 tells us that the duals of the $(2, 1, \delta)$-codes given in Example 4.1(1) are strongly MDS. But this is obviously so, since they are — up to ordering — identical to the given codes.

(b) Dualizing the code of Example 4.1(2) gives us a strongly MDS $(3, 1, 2)$-code with generator matrix

$$G = [1+\omega^{57}D+\omega^{62}D^2,\, \omega+\omega^{44}D+\omega^{54}D^2,\, 1+\omega^{17}D+\omega^{21}D^2] \in \mathbb{F}^3_{64}.$$

(c) Dualizing the codes given in Example 4.2(2) and (3) we obtain another two strongly MDS codes with generator matrices

$$H_1 = \begin{bmatrix} 1 & \beta D + \beta^9 & \beta^6 D + \beta^8 \\ \beta^{14} D & \beta^7 D + \beta^6 & \beta^8 D + \beta \end{bmatrix} \in \mathbb{F}_{16}^{2\times 3}$$

and

$$H_2 = [D^2 + D + \beta^2, \beta^{10} D^2 + D + \beta^7, \beta^5 D^2 + D + \beta^{12}] \in \mathbb{F}_{16}^3.$$

It is known that these codes are also cyclic convolutional codes in the sense of [4], see [4, Thm. 7.5].

Finally we would like to mention that even in the case where $k \mid \delta$ and $(n-k) \mid \delta$, the dual of an MDS code is not MDS in general. An example is given by the following code.

**Example 5.8** The $(3, 1, 2)$-code $\mathcal{C} \subseteq \mathbb{F}((D))^3$, where $\mathbb{F} = \mathbb{F}_{16}$, with generator matrix

$$G = [1 + \beta D + \beta^4 D^2,\ \beta^{10} + \beta^2 D + \beta^4 D^2,\ \beta^8 + \beta^5 D + D^2]$$

and parity check matrix

$$H = \begin{bmatrix} 1 & \beta^{14} D + \beta^2 & \beta^3 D + \beta^3 \\ \beta D & \beta^{11} D + \beta^8 & \beta^{10} D + \beta^{10} \end{bmatrix}$$

is an MDS code, but not strongly MDS. It satisfies $d_3^c = 8$ and $d_4^c = 9$. The dual code generated by $H$ is not MDS. Its distance is 4.

# 6 Decoding Strongly MDS Codes

The codes discussed in the previous section have the property that they allow a very good feedback decoding [18] if the error distribution is reasonably mild.

Let us briefly recall the concept of feedback decoding. Assume the codeword $v = \sum_{j\geq 0} v_j D^j \in \mathcal{C}$ has been sent and the word $\hat{v} = \sum_{j\geq 0} \hat{v}_j D^j \in \mathbb{F}((D))^n$ has been received. Write $\hat{v} = v + \epsilon$, where $\epsilon = \sum_{j\geq 0} \epsilon_j D^j$ is the error vector. In the $j$-th cycle of feedback decoding one corrects $\hat{v}_j$ (hence estimates $\epsilon_j$) and then feeds back this information into the decoding algorithm before proceeding with the next decoding step. It should be intuitively clear that the next step will benefit from the resetting $\hat{v} \leftarrow \hat{v} - \epsilon_j D^j$. As for the decoding step itself one estimates $\epsilon_j$ upon the knowledge of the received sequence $\hat{v}_j, \ldots, \hat{v}_{j+l}$. The length $l+1$, of course, depends on the distance properties of the code. This estimate will be correct if not too many errors have occurred on this string.

In the sequel we will show that strongly MDS codes of rate $\frac{n-1}{n}$ have very good error correcting capabilities in terms of the maximum number of errors acceptable on each string without jeopardizing correct decoding. The basis of the feedback decoding algorithm is the following simple reformulation of the distance properties for the parity check matrices.

**Proposition 6.1** *Let $\mathcal{C} \subseteq \mathbb{F}((D))^n$ be a strongly MDS $(n, n-1, \delta)$-code and put $M := \lfloor \frac{\delta}{n-1} \rfloor + \delta$. Let $H^c_M \in \mathbb{F}^{(M+1)\times(M+1)n}$ be the $M$-th parity check matrix of $\mathcal{C}$ and $\epsilon := (\epsilon_0, \ldots, \epsilon_M)$, $\hat{\epsilon} := (\hat{\epsilon}_0, \ldots, \hat{\epsilon}_M) \in \mathbb{F}^{M+1}$. Assume*

$$\epsilon (H^c_M)^{\mathsf{T}} = \hat{\epsilon}(H^c_M)^{\mathsf{T}} \text{ and } \mathrm{wt}(\epsilon),\, \mathrm{wt}(\hat{\epsilon}) \le \frac{M+1}{2}.$$

*Then*

*(a) $\epsilon_0 = \hat{\epsilon}_0$,*

*(b) if $M$ is even, then additionally $\epsilon_1 = \hat{\epsilon}_1$.*

*Notice that $M$ is even for codes with rate 1/2.*

Proof: Put $\eta := (\eta_0, \ldots, \eta_M)$ where $\eta_j = \epsilon_j - \hat{\epsilon}_j$ for all $0 \le j \le M$. Then $\eta (H^c_M)^{\mathsf{T}} = 0$ and $\mathrm{wt}(\eta) \le M+1$. Thus Proposition 2.2 yields $\eta_0 = 0$. As for (b) notice that if $M$ is even, then $\mathrm{wt}(\epsilon), \mathrm{wt}(\hat{\epsilon}) \le \frac{M}{2}$ and therefore $\mathrm{wt}(\eta) = \mathrm{wt}(\eta_1, \ldots, \eta_M) \le M$. Now $\eta_0 = 0$ implies $(\eta_1, \ldots, \eta_M)(H^c_{M-1})^{\mathsf{T}} = 0$ and Proposition 2.2 together with Corollary 2.4 finishes the proof. □

Observe that the proposition above says that the list of $M+1$ consecutive syndromes determines uniquely the error in the first position. This can be iterated and leads to the following algorithm, which at least works reasonably well for small codes. The question how to practically compute the error in the first position from the syndrome vector for large codes will be addressed at the end of this section.

We will make use of the notation in Remark 2.1.

**Theorem 6.2** *Let $\mathcal{C} \subseteq \mathbb{F}((D))^n$ be a strongly MDS $(n, n-1, \delta)$-code with parity check matrix $H \in \mathbb{F}[D]^{1\times n}$ and $H^c_M \in \mathbb{F}^{(M+1)\times(M+1)n}$ as its $M$-th sliding parity check matrix. Assume the codeword $v \in \mathcal{C}$ has been sent and the word $\hat{v} \in \mathbb{F}((D))^n$ has been received. Without loss of generality assume $\overleftarrow{v}, \overleftarrow{\hat{v}} \ge 0$. Put $\hat{v} = v + \epsilon$, thus $\epsilon \in \mathbb{F}((D))^n$ is the error vector and assume that any sliding window of length $(M+1)n$ contains at most $\frac{M+1}{2}$ errors, i. e.*

$$\mathrm{wt}(\epsilon_{[j,j+M]}) \le \frac{M+1}{2} \text{ for all } j \ge 0. \tag{6.1}$$

*Then the following algorithm will decode $\hat{v}$ correctly, i. e. for each $j = 0, 1, 2, \ldots$ we have $\hat{v}_{[0,j]} = v_{[0,j]}$ after the $j$-th cycle:*

*Put $j := -1$.*

**Step 1:** *Put $j := j+1$.*

**Step 2:** *Compute the syndrome vector $S := (\hat{v} H^{\mathsf{T}})_{[j,j+M]}$.*

**Step 3:** *From the syndrome vector $S$ determine the unique $\eta_0 \in \mathbb{F}^n$, such that $S = \eta (H^c_M)^{\mathsf{T}}$ for some $\eta = (\eta_0, \ldots, \eta_M) \in \mathbb{F}^{(M+1)n}$ satisfying $\mathrm{wt}(\eta) \le \frac{M+1}{2}$.*

**Step 4:** *Put $\hat{v} := \hat{v} - \eta_0 D^j$.*

**Step 5:** *Go to Step 1.*

**Remark 6.3** For illustration purposes assume $n = 2$, i.e. the rate is $1/2$. Theorem 6.2 then states that a strongly MDS $(2, 1, \delta)$-code can be correctly decoded as long as there are no more than $\delta$ errors in any sliding window of length $4\delta + 2$. This has to be compared with a MDS block code of rate $k/n$ where $n = 2k = 4\delta + 2$ which is capable of decoding correctly $\delta$ errors in any slotted window of length $n$. Similar comparisons can be made for different values of $n$.

Proof: We first have to show the existence of $\eta$ as required in Step 3 and the uniqueness of $\eta_0$. In order to do so fix some $j \geq 0$. It is easy to see that for all $w \in \mathbb{F}((D))^n$ with $\overleftarrow{w} \geq 0$ one has

$$(wH^\mathsf{T})_{[j,j+M]} = w_{[j,j+M]}(H_M^c)^\mathsf{T} + w_{[0,j-1]}\mathcal{H}_j^\mathsf{T} \tag{6.2}$$

where

$$\mathcal{H}_j = \begin{bmatrix} H_j & \cdots & H_1 \\ H_{j+1} & \cdots & H_2 \\ \vdots & & \vdots \\ H_{j+M} & \cdots & H_{M+1} \end{bmatrix}.$$

Since, due to the previous decoding steps we have $\hat{v}_{[0,j-1]} = v_{[0,j-1]}$, which is the correct codeword sequence, we get

$$0 = (vH^\mathsf{T})_{[j,j+M]} = \hat{v}_{[j,j+M]}(H_M^c)^\mathsf{T} - \epsilon_{[j,j+M]}(H_M^c)^\mathsf{T} + \hat{v}_{[0,j-1]}(\mathcal{H}_j)^\mathsf{T}.$$

Again with (6.2) this yields

$$S = (\hat{v}H^\mathsf{T})_{[j,j+M]} = \epsilon_{[j,j+M]}(H_M^c)^\mathsf{T}$$

and the assumption (6.1) together with Proposition 6.1 establish the existence of $\eta$ as well as the uniqueness of $\eta_0$ as required in Step 3.
It follows directly from the above that $\eta_0 = \epsilon_j$, where $\eta_0$ is computed in Step 3 of the $j$-th cycle. Thus we have $\hat{v}_{[0,j]} = v_{[0,j]}$ after the $j$-th cycle. □

**Remark 6.4** One might wonder how the algorithm above is related to the total error correcting bound $t := \lfloor \frac{d_{\text{free}}-1}{2} \rfloor$ of the code. First notice that $\frac{M+1}{2} = \frac{d_{\text{free}}-1}{2} = \frac{d_M^c-1}{2}$. From this it follows that for each received word $\hat{v}$ there exists at most one codeword $v \in \mathcal{C}$ such that $v - \hat{v}$ satisfies (6.1). This codeword, of course, is then the result of the algorithm above. However, it might happen that there are two codewords $v_1, v_2 \in \mathcal{C}$ such that the total distances satisfy $\text{wt}(\hat{v} - v_1) = \text{wt}(\hat{v} - v_2) = d(\hat{v}, \mathcal{C}) := \min\{\text{wt}(\hat{v} - v) \mid v \in \mathcal{C}\}$. Hence $v_1$ and $v_2$ have equally close distance to $\hat{v}$ when considered over the total length $[0, \infty)$. This of course can happen only if $d(\hat{v}, \mathcal{C}) > \lfloor \frac{d_{\text{free}}-1}{2} \rfloor$. From the above we know that at most one of these codewords can have an error vector satisfying (6.1). In this situation the decoding algorithm will try to successively minimize $\text{wt}\big((\hat{v} - v)_{[j,j+M]}\big)$ over all codewords $v \in \mathcal{C}$ and $j \geq 0$.
This situation arises for instance for the strongly MDS code $\mathcal{C}$ with parity check matrix

$$H = [1 + \beta^2 D + \beta^5 D^2,\ 1 + \beta^4 D + \beta^5 D^2] \in \mathbb{F}_{2^3}[D]^2 \text{ (where } \beta^3 + \beta + 1 = 0)$$

given in Example 4.1(1) and having free distance $d_{\text{free}} = 6$. In this case the received word

$$\hat{v} = (\beta D + \beta^5 D^4,\ \beta^3 D^2 + \beta^2 D^3)$$

satisfies $\text{wt}(\hat{v}) = \text{wt}(\hat{v} - v_1) = 4 = d(\hat{v}, \mathcal{C})$ for the codeword

$$v_1 = (1 + \beta D + \beta^5 D^4 + \beta^2 D^5,\ 1 + \beta^3 D^2 + \beta^2 D^3 + \beta^2 D^5) \in \mathcal{C}.$$

Hence $\hat{v}$ is equally close to $v_1$ and the zero codeword, but only $\hat{v} - v_1$ satisfies the error condition (6.1). Therefore, the decoding algorithm will decode $\hat{v}$ into the codeword $v_1$.

The main step of the algorithm in Theorem 6.2 is, of course, the determination of $\eta_0$ from the syndrome vector in Step 3. For codes with small parameters this can easily be achieved by simply checking (in a smart way) all linear combinations of at most $\frac{M+1}{2}$ columns of $H_M^c$. But for larger codes this is unsatisfactory and one would like to know an algebraic computation of $\eta_0$. Unfortunately, thus far we cannot offer such an algebraic decoding. It will certainly depend on an algebraic construction of superregular matrices along with their algebraic properties.

We close this paper with the following criterion which, in the affirmative case, speeds up Step 3. It makes use of the systematic sliding parity check matrix of the code, see (3.3), which can be used just as well in the decoding algorithm. Notice that there are $n$ different systematic $M$-th sliding parity check matrices for an $(n, n-1, \delta)$ code. Therefore, the following criterion can be tested $n$ times and hopefully leads to an immediate decision on $\eta_0$ at least ones.

**Proposition 6.5** *Let $\mathcal{C} \subseteq \mathbb{F}((D))^n$ be a strongly MDS $(n, n-1, \delta)$-code with systematic $M$-th sliding parity check matrix $\hat{H}$ as in* (3.3)*. Let $\hat{S} = (\hat{S}_0, \ldots, \hat{S}_M) \in \mathbb{F}^{M+1}$ be such that $\hat{S} = \hat{\eta}\hat{H}^\mathsf{T}$ for some*

$$\hat{\eta} = (e_0, \ldots, e_M, f_0, \ldots, f_M) \in \mathbb{F}^{(M+1)+(M+1)(n-1)} \text{ and } \text{wt}(\hat{\eta}) \leq \frac{M+1}{2}.$$

*If $\text{wt}(\hat{S}) \leq \lceil \frac{M+1}{2} \rceil$, then $f_0 = 0$ and $e_0 = S_0$.*

Proof: The assumptions $\text{wt}(\hat{S}) \leq \lceil \frac{M+1}{2} \rceil$ and $\text{wt}(\hat{\eta}) \leq \frac{M+1}{2}$ imply that there exists a linear combination of at most $M+1$ columns of $\hat{H}$ giving the zero vector. But then Theorem 3.1 yields $f_0 = 0$ and $e_0 = S_0$. □

# 7 Conclusion

In this paper we introduced two new classes of convolutional codes called strongly MDS convolutional codes and codes having maximum distance profile. Strongly MDS convolutional codes have the property that the generalized Singleton bound is attained at the earliest possible column distance. Codes with maximum distance profile have a maximal possible increase of the column distances.

It is shown that strongly MDS convolutional codes perform excellent under feedback decoding. The number of errors which can be iteratively decoded per time interval lets these codes compare with MDS linear block codes having a considerable block length. At this point the feedback decoding algorithm we presented is not powerful enough to practically decode strongly MDS convolutional codes when the degree $\delta$ and the block length $n$ are too large for the syndrome decoding step, see Theorem 6.2 and Proposition 6.5 for details. It will be a matter of future research to construct strongly MDS convolutional codes which come equipped with an algebraic structure and an efficient decoding algorithm comparable to the situation of Reed-Solomon block codes. The class of cyclic convolutional codes [5, 4, 16, 19] might hold some promise here.

From an applications point of view strongly MDS convolutional codes are particularly suited in situations where codes over large alphabets are required and in situations where algebraic decoding is desirable. Hadjicostis [7, 6] has recently demonstrated that convolutional codes over large alphabets are very desirable in areas of process control via linear finite state machines where large numbers of non-concurrent errors should be detected and corrected. It seems that strongly MDS convolutional codes have potential for such applications.

# Appendix

We will prove that the proper minors of the matrix $T_n$ given in Example 3.9(2) are all positive. In order to do so consider the matrix

$$X = \begin{bmatrix} 1 & & & & & \\ 1 & 1 & & & & \\ & 1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & 1 & 1 & \\ & & & & 1 & 1 \end{bmatrix} \in \mathbb{Z}^{n\times n}$$

and notice that for all $k \in \{1, \ldots, n-1\}$ we have

$$X^k = \begin{bmatrix} 1 & & & & & & & & & \\ \binom{k}{1} & 1 & & & & & & & & \\ \binom{k}{2} & \binom{k}{1} & 1 & & & & & & & \\ \vdots & \ddots & \ddots & \ddots & & & & & & \\ \vdots & & \ddots & \ddots & \ddots & & & & & \\ 1 & \ldots & \ldots & \binom{k}{2} & \binom{k}{1} & 1 & & & & \\ & 1 & \ldots & \ldots & \binom{k}{2} & \binom{k}{1} & 1 & & & \\ & & \ddots & & & \ddots & \ddots & \ddots & & \\ & & & 1 & \ldots & \ldots & \binom{k}{2} & \binom{k}{1} & 1 \end{bmatrix}. \tag{A.1}$$

In particular, $X^{n-1} = T_n$. Therefore, the positivity of the proper minors is a consequence of the following theorem.

**Theorem A** *Let $k \in \{1, \ldots, n-1\}$ and $1 \leq i_1 < \ldots < i_r \leq n$, $1 \leq j_1 < \ldots < j_r \leq n$ and define $\hat{X} := (X^k)^{i_1,\ldots,i_r}_{j_1,\ldots,j_r}$. Then $\det \hat{X} \geq 0$ and*

$$\det \hat{X} > 0 \Longleftrightarrow j_l \in \{i_l, i_l - 1, \ldots, i_l - k\} \textit{ for all } l = 1, \ldots, r.$$

Proof: 1) We first show that

$$j_l \notin \{i_l, i_l - 1, \ldots, i_l - k\} \text{ for some } l \Longrightarrow \det \hat{X} = 0. \tag{A.2}$$

To this end notice that

$$X_{ij} = 0 \text{ for } j > i \text{ or } j < i - k$$

and thus

$$\hat{X}_{ef} = X_{i_e j_f} = 0 \text{ for } j_f > i_e \text{ or } j_f < i_e - k.$$

Assume now $j_l > i_l$ for some $l$. Then for all $e \leq l$ and $f \geq l$ we have $j_f \geq j_l > i_l \geq i_e$ and thus $\hat{X}_{ef} = 0$. Hence the first $l$ rows of $\hat{X}$ have at most rank $l-1$ and thus $\det \hat{X} = 0$. Similarly, if $j_l < i_l - k$ for some $l$, then we have $\hat{X}_{ef} = 0$ for all $e \geq l$ and $f \leq l$ and the first $l$ columns of $\hat{X}$ have at most rank $l-1$.

2) It remains to prove the implication "$\Longleftarrow$" of the equivalence given in the theorem. We begin with proving the statement for $k=1$, i. e. for the matrix $X$. In order to do so, we proceed by induction on $r$. For $r=1$ we have to consider the submatrices $X^{i_1}_{i_1}$ and $X^{i_1}_{i_1-1}$. They all trivially have determinant 1. Now let $r > 1$. We suppose the statement is true for all $(r-1) \times (r-1)$ proper submatrices with the according restriction on the indices and we have to show that the assertion is also true for $\hat{X} = X^{i_1,\ldots,i_r}_{j_1,\ldots,j_r}$ where $j_l \in \{i_l, i_l - 1\}$ for all $l$. Notice that the first column of $\hat{X}$ has either one or two nonzero entries and they are equal

to 1. If the first column of $\hat{X}$ has one 1 only, then it is on the first row. Applying cofactor expansion along that column we obtain

$$\det \hat{X} = 1 \cdot \det\ X^{i_2,\ldots,i_r}_{j_2,\ldots,j_r}. \tag{A.3}$$

The $(r-1)\times(r-1)$-submatrix satisfies $j_l \in \{i_l, i_l - 1\}$ for all $l = 2,\ldots,r$ and hence by induction has positive determinant. This proves $\det \hat{X} > 0$ in this case. If the first column of $X^{i_1,\ldots,i_r}_{j_1,\ldots,j_r}$ has two entries equal to 1, then they are necessarily on the first two rows, thus $i_2 = i_1 + 1$ and $j_1 = i_1$. Since $j_2 \in \{i_2, i_2 - 1\} = \{i_1 + 1, i_1\}$ and $j_2 > j_1$, we can only have $j_2 = i_1 + 1$. Then the first row will have only one nonzero entry equal to 1 on the first position, and applying cofactor expansion along that row, we obtain again (A.3) and thus $\det \hat{X} > 0$.

We now proceed by induction on $k$ in order to prove the desired result for $X^k$ where $k > 1$. Assume $X^{k-1}$ has the stated property. Using $X^k = X \cdot X^{k-1}$ and the Cauchy-Binet formula for minors we obtain

$$\det \hat{X} = \sum_{\substack{1\le s_1<\ldots<s_r\le n,\\ s_l\in\{i_l,i_{l-1}\}\cap\{j_l,j_l+1,\ldots,j_l+k-1\}}} \det X^{i_1,\ldots,i_r}_{s_1,\ldots,s_r} \cdot \det(X^{k-1})^{s_1,\ldots,s_r}_{j_1,\ldots,j_r}.$$

Due to part 1) of the proof the sum indeed expands only over the given indices. By induction all nonsingular submatrices of both matrices $X$ and $X^{k-1}$ have positive determinant, hence if there are any nonzero terms in the sum, it is necessarily positive. Therefore, the only thing left to be proven is that there is a nonzero term in the above sum. But all products of the form $\det X^{i_1,\ldots,i_r}_{i_1,\ldots,i_r} \cdot \det(X^{k-1})^{i_1,\ldots,i_r}_{j_1,\ldots,j_r}$ with $j_l \in \{i_l, i_l - 1, i_l - 2, \ldots, i_l - (k-1)\}$ for all $l$ are nonzero. Thus $\det \hat{X} > 0$ and the proof is complete. □

Proof of Theorem 3.10: Step 1: We will show the existence of a systematic sliding parity check matrix $\hat{H}$ as in (3.3) satisfying part (b) of Theorem 3.1. This can be accomplished as follows. Let $\tau := (M+1)(n-1)$ and pick a $\tau\times\tau$-superregular matrix in Toeplitz form, say

$$T := \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ t_2 & t_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ t_\tau & \cdots & t_2 & t_1 \end{bmatrix} = \begin{bmatrix} T^1 \\ T^2 \\ \vdots \\ T^\tau \end{bmatrix} = [T_1,\ldots,T_\tau].$$

Theorem 3.8 guarantees the existence of such a matrix over a suitably large field of characteristic $p$. Now define

$$\hat{H} = \left[\begin{array}{cccc|c} 1 & & & & T^{n-1} \\ & 1 & & & T^{2(n-1)} \\ & & \ddots & & \vdots \\ & & & 1 & T^{(M+1)(n-1)} \end{array}\right] \in \mathbb{F}^{(M+1)\times(M+1)n}.$$

Notice that by construction $\hat{H}$ has the form as in (3.3). We will prove by contradiction that this matrix satisfies part (b) of Theorem 3.1. In order to do so, write $\hat{H} = [e_1, \ldots, e_{M+1}, \hat{T}_1, \ldots, \hat{T}_\tau]$ and assume that $i \leq n-1$ is the smallest index such that $\hat{T}_i$ is in the span of $M$ other columns of $\hat{H}$. Hence these other columns do not involve $\hat{T}_1, \ldots, \hat{T}_{i-1}$. This implies that there is a linear combination of $M+1$ columns of the matrix $[I_\tau, T]$ with a nonzero coefficient for the column $T_i$ and having a zero entry at the positions $1, 2, \ldots, i-1, n-1, 2(n-1), \ldots, (M+1)(n-1)$. Since $i \leq n-1$, these positions are indeed different and thus the weight of this linear combination is at most $\tau - i + 1 - (M+1)$. Consider now the matrix

$$Y := [I_{\tau-i+1} \,|\, \tilde{T}_i, \tilde{T}_{i+1}, \ldots, \tilde{T}_\tau] := \left[\begin{array}{ccc|ccc} 1 & & & t_1 & & \\ & \ddots & & \vdots & \ddots & \\ & & 1 & t_{\tau-i+1} & \cdots & t_1 \end{array}\right] \in \mathbb{F}^{(\tau-i+1)\times 2(\tau-i+1)},$$

where $\tilde{T}_j$ denotes vector of the last $\tau - i + 1$ entries of $T_j$. Notice that superregularity of $T$ implies superregularity of the matrix $[\tilde{T}_i, \ldots, \tilde{T}_\tau]$. The linear combination of $M+1$ columns of $[I_\tau, T]$ above now reads as a linear combination of $M+1$ columns of $Y$ with a nonzero coefficient for the column $\tilde{T}_i$ and having weight at most $\tau - i + 1 - (M+1)$. Hence picking a suitable set of (at most) $\tau - i + 1 - (M+1)$ standard basis vectors, we obtain that the column $\tilde{T}_i$ is in the span of $\tau - 1$ other columns of $Y$. But this is a contradiction to Theorem 3.5(d).

Step 2: Having constructed a matrix $\hat{H}$ as in (3.3) with the corresponding column condition, we now establish the existence of an $(n, n-1)$-code having $\hat{H}$ as its $M$-th systematic sliding parity check matrix. In order to simplify notation write

$$\begin{aligned} \hat{H} &= \left[\begin{array}{cccc|cccc} 1 & & & & h_0 & 0 & \cdots & 0 \\ & 1 & & & h_1 & h_0 & \ddots & \vdots \\ & & \ddots & & \vdots & \ddots & \ddots & 0 \\ & & & 1 & h_M & \cdots & h_1 & h_0 \end{array}\right] \\ &= [e_1, \ldots, e_{M+1}, H_{12}, \ldots, H_{1n}, \ldots, H_{M+1,2}, \ldots, H_{M+1,n}] \end{aligned} \tag{A.4}$$

where $h_i = (h_{i2}, \ldots, h_{in}) \in \mathbb{F}^{n-1}$. We have to find polynomials

$$a = 1 + \sum_{i=1}^{\delta} a_i D^i \in \mathbb{F}[D], \quad b = \sum_{i=0}^{\delta} b_i D^i \in \mathbb{F}[D]^{n-1} \tag{A.5}$$

such that

$$\frac{b}{a} = \sum_{j=0}^{M} h_j D^j + \text{ higher powers} \tag{A.6}$$

(see Theorem 3.1). Recall that $M = \lfloor \frac{\delta}{n-1} \rfloor + \delta$. If $M = \delta$ (i. e. $\delta < n-1$), we may simply take $a = 1$ and $b = \sum_{i=0}^{\delta} h_i D^i$. Now let us assume $M > \delta$. Comparing like powers of $D$

in (A.6) shows that the above requires in particular

$$0 = h_l + a_1 h_{l-1} + a_2 h_{l-2} + \ldots + a_\delta h_{l-\delta} \text{ for any } l > \delta \tag{A.7}$$

for suitable $h_{M+1}, h_{M+2}, \ldots \in \mathbb{F}^{n-1}$. For $l = \delta + 1, \ldots, M$ these equations read as

$$(a_\delta, \ldots, a_1) \begin{bmatrix} h_{M-\delta} & h_{M-\delta-1} & \cdots & h_1 \\ h_{M-\delta+1} & h_{M-\delta} & & h_2 \\ \vdots & \vdots & & \vdots \\ h_{M-1} & h_{M-2} & \cdots & h_\delta \end{bmatrix} = -(h_M, \ldots, h_{\delta+1}). \tag{A.8}$$

Notice that $h_1, \ldots, h_M$ are given data. If we can find a solution $(a_\delta, \ldots, a_1)$ of (A.8), then (A.7) can be established for all $l \geq M + 1$ by choosing $h_l$ suitably. Thereafter, the vector polynomial $b \in \mathbb{F}[D]^{n-1}$ can be computed by equating the coefficients of $D^0, \ldots, D^\delta$ in the equation $b = (\sum_{i=0}^{\infty} h_i D^i)(1 + \sum_{i=1}^{\delta} a_i D^i)$. Hence it remains to consider (A.8). This equation is solvable if

$$\operatorname{rank} \mathcal{H} = (n-1)(M-\delta), \quad \text{where } \mathcal{H} := \begin{bmatrix} h_{M-\delta} & \cdots & h_1 \\ \vdots & & \vdots \\ h_{M-1} & \cdots & h_\delta \end{bmatrix} \in \mathbb{F}^{\delta \times (n-1)(M-\delta)}.$$

Notice that $\rho := (n-1)(M-\delta) = (n-1)\lfloor \frac{\delta}{n-1} \rfloor \leq \delta$. We proceed by contradiction and assume $\operatorname{rank} \mathcal{H} < \rho$. Then there is a column of $\mathcal{H}$ that is a linear combination of the other $\rho - 1$ columns. Since $\mathcal{H}$ is a submatrix of $\hat{H}$ (see (A.4)) and because of the specific structure of $\mathcal{H}$, this yields that a column $H_{1j},\ j = 2, \ldots, n$, is a linear combination of $\rho - 1 + M + 1 - \delta$ other columns of $\hat{H}$. But

$$M + \rho - \delta = \Big\lfloor \frac{\delta}{n-1} \Big\rfloor + \delta + (n-1) \Big\lfloor \frac{\delta}{n-1} \Big\rfloor - \delta \leq \Big\lfloor \frac{\delta}{n-1} \Big\rfloor + \delta = M,$$

and thus we arrive at a contradiction to the column property of $\hat{H}$ (see (b) of Theorem 3.1). Hence (A.8) is solvable and the existence of $a$ and $b$ as in (A.5) and (A.6) is established.

Step 3: Put $H = [a, b^{(1)}, \ldots, b^{(n-1)}]$, where $a \in \mathbb{F}[D]$ and $b =: (b^{(1)}, \ldots, b^{(n-1)}) \in \mathbb{F}[D]^{n-1}$ are constructed as in Step 2). Moreover, let $\mathcal{C} = \{v \in \mathbb{F}(\!(D)\!)^n \mid vH^\mathsf{T} = 0\}$. It remains to show that $\mathcal{C}$ has degree $\delta$, which amounts to showing that $a,\ b^{(1)}, \ldots, b^{(n-1)}$ are coprime and

$$\max\{\deg a,\ \deg b^{(1)}, \ldots,\ \deg b^{(n-1)}\} = \delta. \tag{A.9}$$

Coprimeness can be assumed without loss of generality since division by a common factor would lead to another solution of (A.5) and (A.6). Hence $H$ is basic. By construction and Theorem 3.1 the $M$-th column distance of $\mathcal{C}$ is given by $d^c_M = \lfloor \frac{\delta}{n-1} \rfloor + \delta + 2$. Since this number is strictly bigger than the generalized Singleton bound of any $(n, n-1, \hat{\delta})$-code, where $\hat{\delta} < \delta$, Equation (A.9) follows immediately.

Thus $\mathcal{C}$ is a strongly MDS $(n, n-1, \delta)$-code and the proof is complete. $\square$

# References


[1] B.M. Allen. *Linear Systems Analysis and Decoding of Convolutional Codes*. PhD thesis, University of Notre Dame, August 1999. Available at http://www.nd.edu/˜rosen /preprints.html.

[2] G. D. Forney, Jr. Convolutional codes I: Algebraic structure. *IEEE Trans. Inform. Theory*, IT-16(5):720–738, 1970.

[3] G. D. Forney, Jr. Minimal bases of rational vector spaces, with applications to multivariable linear systems. *SIAM J. Control*, 13(3):493–520, 1975.

[4] H. Gluesing-Luerssen and W. Schmale. On cyclic convolutional codes. Preprint: http:// www.mathematik.uni-oldenburg.de/personen/gluesing/publ.html, October 2002.

[5] H. Gluesing-Luerssen, W. Schmale, and M. Striha. Some small cyclic convolutional codes. In D. Gilliam and J. Rosenthal, editors, *Proceedings of the 15-th International Symposium on the Mathematical Theory of Networks and Systems*, University of Notre Dame, August 2002.

[6] C. N. Hadjicostis. Nonconcurrent error detection and correction in fault-tolerant discrete-time lti dynamic systems. *IEEE Trans. Circuits and Systems–I*, 50(1):45–55, 2003.

[7] C. N. Hadjicostis and G. C. Verghese. Encoded dynamics for fault tolerance in linear finite-state machines. *IEEE Trans. Automat. Contr.*, 47(1):189–192, 2002.

[8] R. Johannesson and K. Zigangirov. Distances and distance bounds for convolutional codes – an overview. In *Topics in Coding Theory. In honour of L. H. Zetterberg.*, Lecture Notes in Control and Information Sciences # 128, pages 109–136. Springer Verlag, 1989.

[9] R. Johannesson and K. Sh. Zigangirov. *Fundamentals of Convolutional Coding*. IEEE Press, New York, 1999.

[10] J. Justesen. An algebraic construction of rate $1/\nu$ convolutional codes. *IEEE Trans. Inform. Theory*, IT-21(1):577–580, 1975.

[11] J. Justesen and L.R. Hughes. On maximum-distance-separable convolutional codes. *IEEE Trans. Information Theory*, IT-20:288, 1974.

[12] S. Lin and D. J. Costello Jr. *Error Control Coding: Fundamentals and Applications*. Prentice-Hall, Englewood Cliffs, NJ, 1983.

[13] V. Lomadze. Convolutional codes and coherent sheaves. *Appl. Algebra Engrg. Comm. Comput.*, 12(4):273–326, 2001.

[14] F. J. MacWilliams and N. J.A. Sloane. *The Theory of Error-Correcting Codes*. North Holland, Amsterdam, 1977.

[15] R. J. McEliece. The algebraic theory of convolutional codes. In V. Pless and W.C. Huffman, editors, *Handbook of Coding Theory*, volume 1, pages 1065–1138. Elsevier Science Publishers, Amsterdam, The Netherlands, 1998.

[16] Ph. Piret. Structure and constructions of cyclic convolutional codes. *IEEE Trans. Inform. Theory*, IT-22(2):147–155, 1976.

[17] Ph. Piret. *Convolutional Codes, an Algebraic Approach*. MIT Press, Cambridge, MA, 1988.

[18] J.P. Robinson. Error propogation and definite decoding of convolutional codes. *IEEE Trans. Inform. Theory*, IT-14(1):121–128, January 1968.

[19] C. Roos. On the structure of convolutional and cyclic convolutional codes. *IEEE Trans. Inform. Theory*, 25(6):676–683, 1979.

[20] J. Rosenthal. Connections between linear systems and convolutional codes. In B. Marcus and J. Rosenthal, editors, *Codes, Systems and Graphical Models*, IMA Vol. 123, pages 39–66. Springer-Verlag, 2001.

[21] J. Rosenthal, J. M. Schumacher, and E. V. York. On behaviors and convolutional codes. *IEEE Trans. Inform. Theory*, 42(6, part 1):1881–1891, 1996.

[22] J. Rosenthal and R. Smarandache. Maximum distance separable convolutional codes. *Appl. Algebra Engrg. Comm. Comput.*, 10(1):15–32, 1999.

[23] J. Rosenthal and E. V. York. BCH convolutional codes. *IEEE Trans. Inform. Theory*, 45(6):1833–1844, 1999.

[24] R. M. Roth and A. Lempel. On MDS codes via Cauchy matrices. *IEEE Trans. Inform. Theory*, 35(6):1314–1319, 1989.

[25] J. B. Shearer and R. J. McEliece. There is no MacWilliams identity for convolutional codes. *IEEE Trans. Inform. Theory*, IT-23(6):775–776, November 1977.

[26] R. Smarandache, H. Gluesing-Luerssen, and J. Rosenthal. Constructions for MDS-convolutional codes. *IEEE Trans. Inform. Theory*, 47(5):2045–2049, 2001.