

On the Minimum Distance of Subspace Codes Generated by Linear Cellular Automata

Luca Mariot¹ and Federico Mazzone¹

¹Semantics, Cybersecurity and Services Group, University of Twente,
Drienerlolaan 5, 7511GG Enschede, The Netherlands
{l.mariot, f.mazzone}@utwente.nl

May 10, 2023

Abstract

Motivated by applications to noncoherent network coding, we study subspace codes defined by sets of linear cellular automata (CA). As a first remark, we show that a family of linear CA where the local rules have the same diameter—and thus the associated polynomials have the same degree—induces a Grassmannian code. Then, we prove that the minimum distance of such a code is determined by the maximum degree occurring among the pairwise greatest common divisors (GCD) of the polynomials in the family. Finally, we consider the setting where all such polynomials have the same GCD, and determine the cardinality of the corresponding Grassmannian code. As a particular case, we show that if all polynomials in the family are pairwise coprime, the resulting Grassmannian code has the highest minimum distance possible.

Keywords cellular automata, network coding, finite fields, Grassmannian, greatest common divisor, Sylvester matrix

1 Introduction

The conventional way of routing packets from source to sink nodes frequently fails to exploit a network’s full potential, which is a common issue in networking. The *butterfly network* [9] serves as a classic example of this problem. The field of network coding emerged around two decades ago, and seeks to solve this problem by exploiting a simple idea: instead of simply routing packets, intermediate nodes in the network can *combine* them, usually by employing linear operators [15]. In this way, more packets can be multiplexed over a single channel usage. In the *noncoherent* network coding strategy, the messages transmitted between nodes are subspaces of an ambient vector space [8]. In this scenario, the need to encode and decode subspaces in

a reliable way for transmission over networks spawned a branch of coding theory that deals with *subspace codes* [7]. These codes can be seen as a generalization of classic linear error correcting codes, where the codewords are subspaces rather than vectors. By embedding the projective space of a vector space with a suitable metric, it is possible to define the minimum distance between any two codewords in a subspace code. Similarly to the usual case of linear error correcting codes, it is desirable to define codes containing a large number of subspaces (to maximize the network's capacity) such that they are at the highest possible distance from each other (to correct as many errors and erasures as possible).

The aim of this paper is to explore the idea of using cellular automata (CA) to construct subspace codes. We consider the specific case of linear CA, motivated by the fact that the body of literature concerning them is quite extensive. Previous work [11] focused on a construction of maximal sets of mutually orthogonal Latin squares (MOLS) based on linear bijective CA. Such a construction is equivalent to finding a maximal family of pairwise coprime polynomials over a finite field, all having the same degree and a nonzero constant term. Here, we investigate another research question stemming from this construction: *what kind of subspace codes can be obtained by families of linear CA, if the underlying polynomials that define their local rules are not pairwise coprime?*

The main contributions of this work are listed below:

- We show that a family of linear CA with local rules of the same diameter generates a *constant dimension code*, also known as a Grassmannian code [1].
- We characterize the minimum distance of a Grassmannian code generated by a family of linear CA.
- We observe that the minimum distance of such codes is optimal when the defining polynomials are pairwise coprime. This is the case considered for MOLS and bent functions [2, 3], with the resulting Grassmannian codes being a particular breed of the partial spread codes introduced in [6].
- We study the specific case where the rules of a family of linear CA are defined by polynomials that have the same GCD, and determine the number of codewords in the resulting Grassmannian code.

The remainder of this paper is organized as follows. Section 2 recalls all background notions related to cellular automata and subspace codes that are necessary to introduce our results. Section 3 formally defines the subspace code generated by a family of linear CA, and remarks that if the underlying local rules have all the same diameter, the resulting code has constant dimension. Section 4 proves the main result of the paper, namely the relationship between the minimum distance of a Grassmannian code generated by linear CA and the maximum degree occurring among the pairwise GCDs of the associated polynomials. Section 5 analyzes the cardinality of the Grassman-

nian codes induced by linear CA whose underlying polynomials have the same pairwise GCD. Finally, Section 6 summarizes the key contributions of the paper, and elaborates on several directions and open problems for future research on the topic.

2 Basic Definitions

In this section, we cover all the basic definitions and results related to cellular automata and subspace codes used throughout the paper. As a general notation, given $q \in \mathbb{N}$ a power of a prime, we denote by \mathbb{F}_q the finite field of order q . For all $n \in \mathbb{N}$, the set of all n -tuples over \mathbb{F}_q is denoted by \mathbb{F}_q^n , and we endow it with the structure of a vector space, where vector sum and multiplication by a scalar are inherited in the usual way from the sum and product operations of \mathbb{F}_q .

2.1 Cellular Automata

A Cellular Automaton (CA) is a type of discrete dynamical system that consists of a regular lattice of cells, which can be either finite or infinite. Each cell updates its state based on a local rule that is applied to its own state and the states of its neighboring cells. This updating process occurs simultaneously for all cells in the lattice, and it is repeated over multiple time steps, giving rise to the dynamic behavior of the system. If the lattice is finite, periodic boundary conditions are typically assumed. This ensures that each cell always has enough neighbors to evaluate the local rule.

While most research on CA focuses on their long-term dynamical behavior, in this work we consider CA as algebraic systems. Specifically, the local rule is applied only once, and only by cells that have enough neighbors to evaluate it. This leads to a CA model that can be viewed as a particular type of vectorial functions over finite fields, which we formally define below:

Definition 1. *Let $d, n \in \mathbb{N}$ such that $d \leq n$, and set $k = d - 1$. Further, let $f : \mathbb{F}_q^d \rightarrow \mathbb{F}_q$ be a d -variable function over the finite field \mathbb{F}_q . A cellular automaton of length n , diameter d , and local rule f is a vectorial mapping $F : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^{n-k}$ whose i -th output coordinate is defined as:*

$$F(x_0, \dots, x_{n-1})_i = f(x_i, \dots, x_{i+k}) \quad (1)$$

for all $i \in \{0, \dots, n - k - 1\}$ and $x \in \mathbb{F}_q^n$.

Intuitively, the output coordinate F_i consists in the application of the local rule f over the neighborhood formed by the i -th input coordinate and the k coordinates on its right. This is the reason why the function maps the vector space \mathbb{F}_q^n to the smaller subspace \mathbb{F}_q^{n-k} : the local rule is applied as long as we have enough right neighbors, i.e., up to the $(n - k)$ -th coordinate.

Thus, the cellular lattice size is reduced by k coordinates after evaluating F . As we mentioned above, this is not a problem, since we consider only the one-shot application of F and we are not interested in iterating the CA over multiple time steps.

In this work we focus on *linear* CA, where the local rule is a linear combination of the input coordinates, that is, for all $x \in \mathbb{F}_q^d$ we have:

$$f(x_0, \dots, x_k) = a_0x_0 + a_1x_1 + \dots + a_kx_k \quad (2)$$

for some $a_0, \dots, a_k \in \mathbb{F}_q$. Further, one can associate to each linear rule of the form (2) a polynomial $P_f \in \mathbb{F}_q[X]$ in a natural way as follows:

$$P_f(X) = a_0 + a_1X + \dots + a_kX^k \quad (3)$$

In other words, we use the coefficients of the vector (a_0, \dots, a_k) that define the local rule as the coefficients of the monomials X^i , in increasing order of powers. In what follows, we will assume that $a_0, a_k \neq 0$, and in particular that $a_k = 1$. This implies that the local rule is *bipermutive*, since any restriction of f obtained by fixing either the first or the last k input variables induces a permutation of \mathbb{F}_q respectively on the last or on the first variable [11]. Moreover, the polynomial associated to f is monic of degree k and has a nonzero constant term.

2.2 Subspace Codes

In this section we cover only the basic notions related to subspace codes. We refer the reader to [8] for a more comprehensive treatment of the subject.

We start by considering the vector space \mathbb{F}_q^n . We denote by $\mathcal{P}(\mathbb{F}_q^n)$ its *projective space*, i.e., the family of all subspaces of \mathbb{F}_q^n . Usually, in the context of network coding, the projective space is interpreted as a metric space under the following distance: for all $A, B \in \mathcal{P}(\mathbb{F}_q^n)$, we have

$$d(A, B) = \dim(A) + \dim(B) - 2\dim(A \cap B) \quad (4)$$

We can now introduce the definition of subspace code:

Definition 2. Let $n \in \mathbb{N}$. A subspace code \mathcal{C} of parameters $[n, \ell(\mathcal{C}), \log_q |\mathcal{C}|, D(\mathcal{C})]$ is a subset of $\mathcal{P}(\mathbb{F}_q^n)$ where $\ell(\mathcal{C}) = \max_{V \in \mathcal{C}} \{\dim(V)\}$ and $D(\mathcal{C})$ is the minimum distance of \mathcal{C} , defined as:

$$D(\mathcal{C}) = \min_{U, V \in \mathcal{C}} \{d(U, V)\} \quad (5)$$

where $d(\cdot, \cdot)$ is computed as in (4).

This definition generalizes the concept of error-correcting codes by considering codewords that are subspaces rather than vectors. In other words,

the elements of the code are not individual vectors, but sets of vectors that span a subspace of the underlying vector space.

The set of all subspaces of dimension k , for a given $0 \leq k \leq n$, is also called the *Grassmannian*, and it is denoted by $Gr(\mathbb{F}_q^n, k)$. Accordingly, a subspace code $\mathcal{C} \subseteq Gr(\mathbb{F}_q^n, k)$ is known as a *Grassmannian code*, or equivalently a *constant dimension code*, since each subspace in \mathcal{C} has dimension k , i.e. $\ell(\mathcal{C}) = k$.

The main problem studied for subspace codes is analogous to the one studied for classic error-correcting codes: for a fixed minimum distance δ , what is the maximum cardinality achievable by a subspace code \mathcal{C} with $D(\mathcal{C}) = \delta$? Intuitively, the lower the allowed minimum distance δ is, the more subspaces we can pack together in a code—and therefore, in the context of network coding, the more messages we can transmit over a network. On the other hand, one also wants that any two subspaces are as far as possible from each other for error-correction purposes, or equivalently a subspace code with the highest minimum distance possible. In the following sections we explore this trade-off for subspace codes defined by families of linear CA.

3 Subspaces Codes from Families of Linear CA

We now describe our method to construct subspace codes using sets of linear CA. Suppose that $F : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^{n-k}$ is a linear CA defined by a local rule $f : \mathbb{F}_q^d \rightarrow \mathbb{F}_q$ of diameter d with associated polynomial $P_f(X) = a_0 + a_1X + \dots + a_kX^k$ where $k = d - 1$. This CA is a linear mapping of the form $F(x) = M_F \cdot x^\top$ for all $x \in \mathbb{F}_q^n$, where M_F is a $k \times n$ matrix over \mathbb{F}_q of the following form:

$$M_F = \begin{pmatrix} a_0 & \dots & a_k & 0 & 0 & \dots & 0 \\ 0 & a_0 & \dots & a_k & 0 & \dots & 0 \\ \vdots & \dots & \ddots & \ddots & \ddots & \dots & \vdots \\ 0 & \dots & 0 & 0 & a_0 & \dots & a_k \end{pmatrix}. \quad (6)$$

The matrix M_F is called the *transition matrix* of F , and it is obtained by shifting the coefficients of P_f one place to the right per each subsequent row. As we discussed in Section 2.1, we assume that $a_0 \neq 0$ and $a_k = 1$. In this way, the polynomial P_f is monic of degree k with a nonzero constant term, and all the columns of M_F are nonzero.

Now, let us consider the *kernel* $\ker(f)$ of the linear CA F . By definition, this is the subspace of input vectors $x \in \mathbb{F}_q^n$ such that $F(x) = \underline{0}$, and it is equivalent to the nullspace of the matrix M_F . A method to construct the kernel of F is by using the following *preimage computation* procedure:

- Set the output configuration of the CA F to the null vector $\underline{0}$.
- For all $\tilde{x} = (\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_{k-1}) \in \mathbb{F}_q^k$, do:

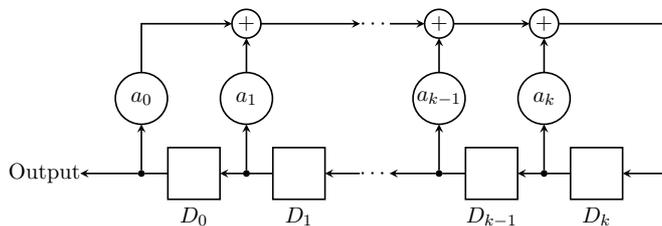


Figure 1: Example of linear feedback shift register of order k .

1. Set the first k coordinates of the CA input $x \in \mathbb{F}_q^n$ to \tilde{x} , that is, $x_0 = \tilde{x}_0, x_1 = \tilde{x}_1, \dots, x_{k-1} = \tilde{x}_{k-1}$.
2. For all $i \in \{k, \dots, n-1\}$, compute the i -th input coordinate x_i as:

$$x_i = -(a_0x_{i-k} + a_1x_{i-k+1} + \dots + a_{k-1}x_{i-1}) . \quad (7)$$

Equation (7) stems from the fact that the i -th output coordinate of the CA must be zero (since the whole output configuration is the null vector $\underline{0}$). Thus, one can recover x_i from the equation of the local rule, moving all the terms $a_0x_{i-k}, \dots, a_{k-1}x_{i-1}$ to the left hand side and changing their sign. Since we assume $a_k = 1$, one then obtains Equation (7).

This preimage computation procedure is equivalent to the computation of a k -th order homogeneous *Linear Recurring Sequence* (LRS) [10]. In particular, the kernel of F corresponds to all infinite sequences x_0, x_1, \dots of elements in \mathbb{F}_q that satisfy the following recurrence equation:

$$a_0x_i + a_1x_{i+1} + \dots + a_kx_{i+k} = 0 , \quad (8)$$

and truncating such sequences to the n -th element. Thus, one can compute the kernel of F by using a *Linear Feedback Shift Register* (LFSR) of order k and with feedback polynomial P_f (see Figure 1). The idea is to initialize the registers D_0, \dots, D_{k-1} with the starting block $\tilde{x} \in \mathbb{F}_q^k$ of the preimage x , and then run the LFSR for n clock steps. At each step $i \in \{0, \dots, n-1\}$, the rightmost register D_k is updated with the feedback of the linear recurrence equation (8), while the leftmost register D_0 outputs the value of x_i . We remark that this approach has been adopted in [14] to study the period of spatially periodic preimages in linear bipermutive CA and in [12] to construct cyclic codes from linear CA.

From the discussion above, we can conclude the following result:

Lemma 1. *Let $F : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^{n-k}$ be a linear CA defined by a local rule $f : \mathbb{F}_q^d \rightarrow \mathbb{F}_q$ of diameter d , where $k = d-1$, and denote by M_F the transition matrix of F . Then, $\dim(\ker(F)) = k$ and $\text{rank}(M_F) = n - k$.*

Proof. Using the preimage computation procedure outlined above, the number of preimages of the null vector $\underline{0}$ under F is q^k , since each of them

is uniquely determined by a vector $\tilde{x} \in \mathbb{F}_q^k$. Hence, $|\ker(F)| = q^k$, and $\dim(\ker(F)) = k$. The rank of M_F now follows from the fact that $\ker(F)$ is also the nullspace of M_F , and from the rank-nullity theorem: the number of columns of M_F equals the sum of the rank of M_F and the dimension of its nullspace. \square \square

We are now ready to define a subspace code generated by a set of linear CA.

Definition 3. *Let $n, d \in \mathbb{N}$ with $d \leq n$ and $k = d - 1$. The subspace code generated by a family \mathcal{F} of t linear CA $F_1, \dots, F_t : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^{n-k}$, respectively defined by bijective local rules $f_1, \dots, f_t : \mathbb{F}_q^d \rightarrow \mathbb{F}_q$ of diameter d , is the set*

$$\mathcal{C}_{\mathcal{F}} = \{\ker(F_i) : 1 \leq i \leq t\} . \quad (9)$$

In other words, the subspace code consists of the kernels of all the t linear CA in the family \mathcal{F} . The reader might wonder why we choose specifically the kernels of the CA instead of, for instance, their images. This will become clearer in the next section where we exploit this fact to characterize the minimum distance of the code. Moreover, from Lemma 1 it holds that each kernel in $\mathcal{C}_{\mathcal{F}}$ has dimension k . Thus, we have the following result:

Lemma 2. *The subspace code $\mathcal{C}_{\mathcal{F}}$ defined in Equation (9) is a Grassmannian code, i.e. $\mathcal{C}_{\mathcal{F}} \subseteq \text{Gr}(\mathbb{F}_q^n, k)$.*

4 Relation between Minimum Distance and GCD

Lemma 2 prompts us with the following natural question: is it possible to characterize the minimum distance of a Grassmannian code generated by a family \mathcal{F} of linear CA, possibly linking it with the properties of the polynomials associated to the local rules? In this section, we analyze this issue.

In the following discussion, we make the assumption that $n = 2k$. Hence, a subspace code is generated by a family of linear CA $F_1, \dots, F_t : \mathbb{F}_q^{2k} \rightarrow \mathbb{F}_q^k$. The codewords of the Grassmannian code $\mathcal{C}_{\mathcal{F}}$ are the kernels $\ker(F_i)$ for $1 \leq i \leq t$. By applying (4), and Lemma 2, the distance between any two kernels in $\mathcal{C}_{\mathcal{F}}$ equals:

$$\begin{aligned} d(\ker(F), \ker(G)) &= \dim(\ker(F)) + \dim(\ker(G)) - 2\dim(\ker(F) \cap \ker(G)) = \\ &= 2k - 2\dim(\ker(F) \cap \ker(G)) . \end{aligned} \quad (10)$$

Thus, this distance is inversely proportional to the size of the intersection of the kernels. We can then characterize the minimum distance $D(\mathcal{C}_{\mathcal{F}})$ in terms of the largest intersection between any two kernels in the subspace code. To this end, we first need some further results. Given any two CA

$F, G \in \mathcal{F}$, with local rules f, g respectively, we can define their concatenation $H : \mathbb{F}_q^{2k} \rightarrow \mathbb{F}_q^{2k}$ as the map

$$H(x) := F(x) \parallel G(x) . \quad (11)$$

Remark 1. *We can easily see that H is still a linear application, and $H(x) = \underline{0}$ if and only if $F(x) = \underline{0}$ and $G(x) = \underline{0}$. So we have that the kernel of H is nothing else than the intersection $\ker(F) \cap \ker(G)$.*

The matrix associated to H is the vertical concatenation of M_F and M_G :

$$M_H = \begin{pmatrix} a_0 & \dots & a_k & 0 & 0 & \dots & 0 \\ 0 & a_0 & \dots & a_k & 0 & \dots & 0 \\ \vdots & \dots & \ddots & \ddots & \ddots & \dots & \vdots \\ 0 & \dots & 0 & 0 & a_0 & \dots & a_k \\ b_0 & \dots & b_k & 0 & 0 & \dots & 0 \\ 0 & b_0 & \dots & b_k & 0 & \dots & 0 \\ \vdots & \dots & \ddots & \ddots & \ddots & \dots & \vdots \\ 0 & \dots & 0 & 0 & b_0 & \dots & b_k \end{pmatrix} . \quad (12)$$

We can recognize such matrix as the *Sylvester matrix* associated to the polynomials P_f, P_g corresponding to the local rules f, g . Notably, the determinant of this matrix is called the *resultant* of P_f and P_g , denoted by $Res(P_f, P_g)$, and it is known that $Res(P_f, P_g) \neq 0 \Leftrightarrow \gcd(P_f, P_g) = 1$ [5]. In other words, the Sylvester matrix M_H is invertible if and only if the two polynomials P_f, P_g defining the local rules of F and G are relatively prime. This fact was used by the authors of [11] to construct orthogonal Latin squares from linear CA.

In our setting of Grassmannian codes, we are interested in the more general situation where the Sylvester matrix associated to F and G is not necessarily invertible. To determine the dimension of the intersection of $\ker(F)$ and $\ker(G)$ we need the following result that links the dimension of the null space of the Sylvester matrix to the degree of the GCD of the two polynomials¹:

Lemma 3. *Let $f, g \in \mathbb{F}_q[X]$ be two polynomials, and denote by $S_{f,g}$ their Sylvester matrix. Then,*

$$\dim(\text{null}(S_{f,g})) = \deg(\gcd(f, g)) . \quad (13)$$

Proof. Notice that $S_{f,g}$ has size $m \times m$, where $m = \deg(f) + \deg(g)$. The idea is to compute the null space $\text{null}(S_{f,g}^\top) = \{z \in \mathbb{F}_q^m : S_{f,g}^\top z^\top = \underline{0}\}$ of the transposed Sylvester matrix. For any $z \in \text{null}(S_{f,g}^\top)$ we write $z = (w \parallel v)$

¹This seems to be a widely known result, but we could not find any reference in the literature that proves it. Hence, we report a proof here for our convenience.

as the concatenation of the vectors $w \in \mathbb{F}_q^{\deg(g)}$ and $v \in \mathbb{F}_q^{\deg(f)}$. Next, we associate to w and v two polynomials s, t respectively defined as:

$$s(X) = w_0 + w_1X + w_2X^2 + \cdots + w_{\deg(g)}X^{\deg(g)} , \quad (14)$$

$$t(X) = v_0 + v_1X + v_2X^2 + \cdots + v_{\deg(f)}X^{\deg(f)} . \quad (15)$$

where clearly $\deg(s) \leq \deg(g)$ and $\deg(t) \leq \deg(f)$. Then we have that $S_{f,g}^\top z$ can be written in polynomial form as:

$$f(X)s(X) + g(X)t(X) = \gcd(f, g)(X) (f_0(X)s(X) + g_0(X)t(X)) , \quad (16)$$

for suitable $f_0, g_0 \in \mathbb{F}_q[X]$ that are relatively prime. Therefore, z belongs to the null space of $S_{f,g}^\top$ if and only if

$$f_0(X)s(X) + g_0(X)t(X) = 0 . \quad (17)$$

By taking this identity modulo g_0 , and omitting from now on the (X) notation, we obtain

$$f_0s \equiv 0 \pmod{g_0} . \quad (18)$$

Since f_0 and g_0 are coprime, we have $g_0 \mid s$, thus $s = g_0p$ for some $p \in \mathbb{F}_q[X]$. Further, note that $\deg(p) = \deg(s) - \deg(g_0) \leq \deg(g) - \deg(g_0) = \deg(\gcd(f, g))$. By replacing this in (17) we get

$$f_0g_0p + g_0t = g_0(f_0p + t) = 0 , \quad (19)$$

hence $t = -f_0p$. Thus, z belongs to the null space if and only if (s, t) is of the form $(g_0p, -f_0p)$ for some p with degree at most $\deg(\gcd(f, g))$. The dimension of the nullspace of (the transpose of) $S_{f,g}$ is then $\deg(\gcd(f, g))$. \square \square

We can now prove our main result: the minimum distance of a Grassmannian code $\mathcal{C}_{\mathcal{F}}$ generated by a family \mathcal{F} of linear CA of diameter d is determined by the largest degree of the pairwise GCD computed over the polynomials that define the local rules.

Theorem 1. *Let \mathcal{F} be a family of linear CA of length $2k$, each defined by a linear local rule of diameter d where $k = d - 1$. Then, the minimum distance of the Grassmannian code $\mathcal{C}_{\mathcal{F}}$ generated by \mathcal{F} is equal to:*

$$D(\mathcal{C}_{\mathcal{F}}) = 2k - 2 \cdot \max_{F, G \in \mathcal{F}} \{\deg(\gcd(P_f, P_g))\} , \quad (20)$$

where P_f, P_g are the polynomials associated to the local rules of F and G .

Proof. By Equation (10), the distance between any two kernels $\ker(F), \ker(G)$ in $\mathcal{C}_{\mathcal{F}}$ is equal to $2k - 2\dim(\ker(F) \cap \ker(G))$. Hence, to determine $D(\mathcal{C}_{\mathcal{F}})$, we need to compute

$$\max_{F, G \in \mathcal{F}} \{\dim(\ker(F) \cap \ker(G))\} . \quad (21)$$

Recall that, by Remark 1, the nullspace of the Sylvester matrix M_H defined by F and G is the intersection of $\ker(F)$ and $\ker(G)$. Therefore, we have

$$\dim(\ker(F) \cap \ker(G)) = \text{null}(M_H) . \quad (22)$$

Now, by Lemma 3, we have that $\text{null}(M_H) = \deg(\gcd(f, g))$. We can thus rewrite (21) as:

$$\max_{F, G \in \mathcal{F}} \{\dim(\ker(F) \cap \ker(G))\} = \max_{F, G \in \mathcal{F}} \{\deg(\gcd(f, g))\} , \quad (23)$$

which proves our theorem. \square \square

5 Equidistant Constant Dimension Codes from Linear CA

In the previous section we proved that the minimum distance of a Grassmannian code generated by a family of linear CA depends on the maximum degree of the pairwise GCDs of their associated polynomials. We now analyze how large such a code can be by considering some specific cases.

For a given minimum distance δ , one ideally wants to define a subspace code in such a way that it contains as many codewords as possible. To phrase it differently, we want to find the maximum number of degree k polynomials in $\mathbb{F}_q[X]$, such that their pairwise GCD has degree at most $t = k - \delta/2$.

The optimal case of the highest minimum distance occurs when $t = 0$. As a matter of fact, this happens when all polynomials that define the linear CA in the family \mathcal{F} are pairwise coprime, as shown in the next result:

Lemma 4. *Let $\mathcal{C}_{\mathcal{F}}$ be a Grassmannian code generated by a set \mathcal{F} of linear CA $F_1, \dots, F_r : \mathbb{F}_q^{2k} \rightarrow \mathbb{F}_q^k$, defined by the local rules $f_1, \dots, f_r : \mathbb{F}_q^d \rightarrow \mathbb{F}_q$ of diameter d where $k = d - 1$. Suppose that for each $F_i, F_j \in \mathcal{F}$ with $i \neq j$ the polynomials P_{f_i}, P_{f_j} associated to the local rules respectively of F_i and F_j are coprime, that is $\gcd(P_{f_i}, P_{f_j}) = 1$. Then, the minimum distance of the code is:*

$$D(\mathcal{C}_{\mathcal{F}}) = 2k . \quad (24)$$

Notice that the code in Lemma 4 is also *equidistant*: every pair of codewords in $\mathcal{C}_{\mathcal{F}}$ has distance $2k$. The maximum cardinality achievable by a subspace code of this kind corresponds to the size N_k of the largest family of pairwise coprime polynomials with degree k and nonzero constant term. This problem has already been addressed in [11], where the authors algorithmically build such sets of polynomials and prove their maximality. Specifically, N_k is equal to:

$$N_k = I_k + \sum_{j=1}^{\lfloor \frac{k}{2} \rfloor} I_j , \quad (25)$$

where, for all $n \in \mathbb{N}$, I_n is the cardinality of the set \mathcal{I}_n of irreducible polynomials of degree n , which can be computed through *Gauss's formula* [4]:

$$I_n := |\mathcal{I}_n| = \frac{1}{n} \sum_{d|n} \mu(d) q^{n/d} , \quad (26)$$

with $\mu(\cdot)$ denoting the *Möbius function* [10].

If we relax the assumption on the minimum distance, allowing it for being non-optimal, we get into the generic case, where we allow the pairwise GCDs to have degree at most $t > 0$. In what follows, let us define the set of all monic polynomials of degree k with nonzero constant term as:

$$\text{Poly}_k(\mathbb{F}_q) := \{f \in \mathbb{F}_q[X] : f \text{ monic}, f(0) \neq 0, \deg(f) = k\} . \quad (27)$$

Further, let $\text{CD}_{k,t}(\mathbb{F}_q)$ be the family of subsets of $\text{Poly}_k(\mathbb{F}_q)$ such that the degree of the pairwise GCDs is at most t :

$$\text{CD}_{k,t}(\mathbb{F}_q) := \{S \subseteq \text{Poly}_k(\mathbb{F}_q) : \forall f_1, f_2 \in S, \deg(\gcd(f_1, f_2)) \leq t\} . \quad (28)$$

The goal is to find a maximal element of $\text{CD}_{k,t}(\mathbb{F}_q)$ and its cardinality, that is $\max_{S \in \text{CD}_{k,t}(\mathbb{F}_q)} |S|$. This general case is quite tricky to handle. For this reason, here we address an intermediate problem, where we assume that *all pairs of polynomials have exactly the same GCD $g \in \mathbb{F}_q[X]$ with degree t* . This corresponds to finding the largest set in:

$$\text{CF}_{k,g}(\mathbb{F}_q) := \{S \subseteq \text{Poly}_k(\mathbb{F}_q) : \forall f_1, f_2 \in S, \gcd(f_1, f_2) = g\} . \quad (29)$$

Remark that the resulting Grasmannian code is again equidistant in this case, with minimum distance $2k - 2t$.

The fixed polynomial g is a common divisor of all the polynomials in the set S . So, for any polynomial $f \in \text{Poly}_k(\mathbb{F}_q)$, we can find f' such that $f = gf'$, with $\deg(f') = k - t$. To build our maximal set S and compute its size, we can therefore use the same approach as in [11] applied to $\text{Poly}_{k-t}(\mathbb{F}_q)$. In particular, we can build a set $S \in \text{CF}_{k,g}(\mathbb{F}_q)$ by adopting a straightforward variation of the algorithm CONSTRUCTION-IRREDUCIBLE. The modified pseudocode is reported below:

CONSTRUCTION-UNIFORM-GCD(k, g)

Initialization: Initialize set T to \mathcal{I}_{k-t} , where $t = \deg(g)$

Loop: For all $1 \leq i \leq \lfloor \frac{k-t}{2} \rfloor$ do:

1. Build set T_i by multiplying each polynomial in \mathcal{I}_i with a distinct polynomial in \mathcal{I}_{k-t-i}
2. Add set T_i to T

Final step: If $k - t$ is odd, build set $T_{(k-t-1)/2}$ by multiplying each polynomial in $\mathcal{I}_{(k-t-1)/2}$ with a distinct polynomial in $\mathcal{I}_{(k-t+1)/2}$, and add $T_{(k-t-1)/2}$ to T . If $k - t$ is even, build set $T_{(k-t)/2}$ by squaring each irreducible polynomial in $\mathcal{I}_{(k-t)/2}$, and add $T_{(k-t)/2}$ to T . Finally, define the set $S := \{gf' : f' \in T\}$.

Output: return S

It is easy to see that the set built by the above algorithm belongs to $\text{CF}_{k,g}(\mathbb{F}_q)$: every element of S is monic since the product of monic polynomials, it has constant coefficient non-zero since both factors do as well, and it has degree k . Moreover, since the intermediate set T belongs to $\text{CF}_{k,1}(\mathbb{F}_q)$ thanks to [11], it follows that for all $f'_1, f'_2 \in T$ we have $\text{gcd}(f'_1, f'_2) = 1$ and thus $\text{gcd}(gf'_1, gf'_2) = g$.

Therefore, by following the same arguments in [11], we can see that the cardinality of such set is:

$$|S| = I_{k-t} + \sum_{i=1}^{\lfloor \frac{k-t}{2} \rfloor} I_i . \quad (30)$$

Finally, regarding the maximality, we pick a maximal element $A \in \text{CF}_{k,g}(\mathbb{F}_q)$ and define $A' := \{f/g : f \in A\}$. Then, the proof can just follow the argument of [11] by applying it to the set A' .

6 Conclusions and Future Works

In this paper, we started to investigate subspace codes generated by families of linear CA. We first remarked that the subspaces codes generated by CA with uniform diameter are Grassmannian. Then, we proved that the minimum distance of such codes is determined by the maximum degree of the pairwise GCDs of the polynomials associated to the local rules. Finally, we analyzed the maximal cardinality achievable by these subspace codes, considering two particular cases. The first one corresponds to the problem of counting how many pairwise coprime monic polynomials of fixed degree and nonzero constant term over a finite field exist, already addressed in [11], and we remarked that the resulting Grassmannian codes achieve the highest possible minimum distance $2k$. Next, we focused on the case where the polynomials have the same pairwise GCD. We presented a modified version of the algorithm in [11] to construct such a set of polynomials, and we showed that it is maximal.

There are several interesting directions to explore for future research. The most straightforward generalization would be to build Grassmannian codes from sets of linear CA where the underlying polynomials do not have the same pairwise GCD, but the degree is still fixed. The next step would

then be to build and count the codes by setting an upper bound on the degree of the GCD. In this way, the cardinality of the optimal code can be determined exactly. Further, a comparison with the Grassmannian codes obtained with our method against those already published in the literature is in order, since the optimal case of our construction is a specific instance of the partial spreads codes introduced in [6]. Finally, we would like to investigate the *decoding* aspect of our subspace codes, and study if it is possible to exploit the parallel nature of the CA to build an efficient decoder. We believe that the inversion algorithm for mutually orthogonal CA presented in [13] represents a viable starting point to investigate this direction.

References

- [1] T. Etzion and H. Zhang. Grassmannian codes with new distance measures for network coding. *IEEE Trans. Inf. Theory*, 65(7):4131–4142, 2019.
- [2] M. Gadouleau, L. Mariot, and S. Picek. Bent functions from cellular automata. *IACR Cryptol. ePrint Arch.*, page 1272, 2020.
- [3] M. Gadouleau, L. Mariot, and S. Picek. Bent functions in the partial spread class generated by linear recurring sequences. *Des. Codes Cryptogr.*, 91(1):63–82, 2023.
- [4] C. F. Gauß. *Disquisitiones arithmeticae*. Humboldt-Universität zu Berlin, 1801.
- [5] I. M. Gelfand, M. Kapranov, and A. Zelevinsky. *Discriminants, resultants, and multidimensional determinants*. Springer Science & Business Media, 2008.
- [6] E. Gorla and A. Ravagnani. Partial spreads in random network coding. *Finite Fields Their Appl.*, 26:104–115, 2014.
- [7] A. Khaleghi, D. Silva, and F. R. Kschischang. Subspace codes. In M. G. Parker, editor, *Cryptography and Coding, 12th IMA International Conference, Cryptography and Coding 2009, Cirencester, UK, December 15-17, 2009. Proceedings*, volume 5921 of *Lecture Notes in Computer Science*, pages 1–21. Springer, 2009.
- [8] R. Koetter and F. R. Kschischang. Coding for errors and erasures in random network coding. *IEEE Trans. Inf. Theory*, 54(8):3579–3591, 2008.
- [9] F. R. Kschischang. An introduction to network coding. In *Network Coding*, pages 1–37. Elsevier, 2012.

- [10] R. Lidl and H. Niederreiter. *Finite fields*. Cambridge university press, 1997.
- [11] L. Mariot, M. Gadouleau, E. Formenti, and A. Leporati. Mutually orthogonal latin squares based on cellular automata. *Des. Codes Cryptogr.*, 88(2):391–411, 2020.
- [12] L. Mariot and A. Leporati. A cryptographic and coding-theoretic perspective on the global rules of cellular automata. *Nat. Comput.*, 17(3):487–498, 2018.
- [13] L. Mariot and A. Leporati. Inversion of mutually orthogonal cellular automata. In G. Mauri, S. E. Yacoubi, A. Dennunzio, K. Nishinari, and L. Manzoni, editors, *Cellular Automata - 13th International Conference on Cellular Automata for Research and Industry, ACRI 2018, Como, Italy, September 17-21, 2018, Proceedings*, volume 11115 of *Lecture Notes in Computer Science*, pages 364–376. Springer, 2018.
- [14] L. Mariot, A. Leporati, A. Dennunzio, and E. Formenti. Computing the periods of preimages in surjective cellular automata. *Nat. Comput.*, 16(3):367–381, 2017.
- [15] M. Médard and A. Sprintson. *Network coding: Fundamentals and applications*. Academic Press, 2011.