Linear codes for b-symbol read channels attaining the Griesmer bound

Sascha Kurz
University of Bayreuth
sascha.kurz@uni-bayreuth.de

July 10, 2025

Abstract Reading channels where b-tuples of adjacent symbols are read at every step have e.g. applications in storage. Corresponding bounds and constructions of codes for the b-symbol metric, especially the pair-symbol metric where $b = 2$, were intensively studied in the last fifteen years. Here we determine the optimal code parameters of linear codes in the b-symbol metric assuming that the minimum distance is sufficiently large. We also determine the optimal parameters of linear binary codes in the pair-symbol metric for small dimensions.

Keywords pair-symbol codes; b-symbol metric; optimal linear codes; Griesmer bound

1 Introduction

In storage applications the reading device is sometimes insufficient to isolate adjacent symbols, which makes it necessary to adjust the standard coding-theoretic error model. Cassuto and Blaum studied a model where pairs of adjacent symbols are read in every step and introduced the so-called symbol-pair metric for codes [3]. This notion was generalized to the b-symbol metric where b-tuples of adjacent symbols are read at every step, see e.g. [22]. While general codes where studied, see e.g. [5], for representation and decoding purposes it is beneficial to assume more structured codes. A rather general and important subclass of codes are linear codes i.e. subspaces of some vector space $\mathbb{F}_q^k$ over a finite field. For small minimum distances a Singleton type bound, introduced by Chee et al. for the symbol-pair metric [4], turned out to be very effective. Like for the Hamming metric codes attaining this bound are called maximum distance separable (MDS) codes and quite some constructions were studied for the symbol-pair metric, see e.g. [6, 13, 14, 17, 18, 19, 20]. Here we are interested in optimal linear codes w.r.t. the b-symbol metric in those situations where the minimum distance is *large*. For the Hamming metric Solomon and Stiffler [21] showed that the so-called *Griesmer bound* [10] can always be attained with equality if the minimum distance is sufficiently large. The main objective of this paper is to show the analogous result for the b-symbol metric for the Griesmer type bound recently introduced in [12, 16]. In order to get a more complete picture of the intermediate situation we determine the parameters of optimal binary linear codes in the symbol-pair metric for small dimensions.

The remaining part of the paper is structured as follows. In Section 2 we introduce the necessary preliminaries. Our main result on the minimum length of a linear code in the b-symbol metric for large minimum distances in stated in Theorem 3.16 in Section 3. To complement this asymptotic result we determine the exact minimum lengths of binary linear codes in the b-symbol metric for small dimensions in Section 4. The paper is closed by a brief conclusion in Section 5.

2 Preliminaries

For some prime power q let $\mathbb{F}_q$ denote the finite field with q elements and $\mathbb{F}_q^n$ the n-dimensional vector space over $\mathbb{F}_q$. An $[n,k]_q$ *code* C is a k-dimensional linear subspace of $\mathbb{F}_q^n$. If C is given as the row span of a $k \times n$ matrix G, then G is called a *generator matrix* of C. A generator matrix G is called systematic if it starts with a $k \times k$ identity matrix. The elements of an $[n,k]_q$ code C are called *codewords*. The *Hamming weight* $\mathrm{wt_H}(c)$ of a codeword $c = (c_0, \cdots, c_{n-1})$ is the number of non-zero entries $|\{c_i \,:\, c_i \neq 0, 0 \le i \le n-1\}|$. With this, the *weight enumerator* of C is the homogeneous polynomial $\mathrm{w}_C^{\mathrm{H}}(x,y) := \sum_{c \in C} x^{\mathrm{wt_H}(c)} y^{n-\mathrm{wt_H}(c)}$ and the *Hamming distance* between two codewords $c, c' \in C$ is given by $\mathrm{d_H}(c,c') := \mathrm{wt_H}(c-c')$.

The Hamming distance is an appropriate measure for error detection and correction in certain channels and is indeed a metric. In [3] the authors introduced another channel where a different kind of metric is more suitable. Instead of single symbols one assumes that neighboring pairs of symbols are read, which was subsequently generalized to neighboring b-tuples of symbols for $b \ge 2$, see e.g. [22] for details. To this end let

$$\pi_b(c) := \big((c_0, \ldots, c_{b-1}), (c_1, \ldots, c_b), \ldots, (c_{n-1}, c_0, \ldots, c_{b-2})\big) \tag{1}$$

be the b-*symbol read vector* of a codeword $c = (c_0, \ldots, c_{n-1}) \in C \subseteq \mathbb{F}_q^n$ and

$$\mathrm{d}_b(c,c') := \Big|\big\{0 \le i \le n-1 \,:\, (c_i, c_{i+1}, \ldots, c_{i+b-1}) \neq \big(c'_i, c'_{i+1}, \ldots, c'_{i+b-1}\big)\big\}\Big| \tag{2}$$

be the b-*symbol distance* between codewords $c = (c_0, \ldots, c_{n-1})$ and $c' = \left(c'_0, \ldots, c'_{n-1}\right)$, where the indices are read modulo n. Similar to the definition of $\mathrm{wt}_{\mathrm{H}}(c)$ we can define the b-*weight* $\mathrm{wt}_b(c) := \mathrm{d}_b(c, 0)$ of a codeword $c \in C$ and the corresponding weight enumerator $\mathrm{w}_C^b(x, y) := \sum_{c \in C} x^{\mathrm{wt}_b(c)} y^{n - \mathrm{wt}_b(c)}$ of C (w.r.t. the b-symbol distance). We remark that d_1 is equal to the Hamming distance and d_2 denotes the pair-symbol distance.

The *minimum Hamming distance* $\mathrm{d}_{\mathrm{H}}(C)$ of an $[n, k]_q$ code C is defined as the minimum Hamming distance between two different codewords, i.e.

$$\mathrm{d}_{\mathrm{H}}(C) := \min\{\mathrm{d}_{\mathrm{H}}(c, c') \,:\, c, c' \in C, c \neq c'\} = \min\{\mathrm{wt}_{\mathrm{H}}(c) \,:\, c \in C, c \neq 0\}\,. \quad (3)$$

If $d = \mathrm{d}_{\mathrm{H}}(C)$ then we also speak of an $[n, k, d]_q$ *code*. Similarly,

$$\mathrm{d}_b(C) := \min\{\mathrm{d}_b(c, c') \,:\, c, c' \in C, c \neq c'\} = \min\{\mathrm{wt}_b(c) \,:\, c \in C, c \neq 0\} \quad (4)$$

and we speak of an $[n, k, d]_q^b$ *code* if $d = \mathrm{d}_b(C)$. We call an $[n, k, d]_q$ code *optimal* (w.r.t. the Hamming metric) if no $[n-1, k, d]_q$ code exists. Similarly, we call an $[n, k, d]_q^b$ code *optimal* (w.r.t. the b-symbol metric) if no $[n-1, k, d]_q^b$ code exists. Note that there are several notions of optimality for linear codes and here we choose length-optimality, i.e. the smallest possible length for given parameters k, d, q, and b, which is justified by the following observation.

Lemma 2.1 *The existence of an* $[n, k, d]_q^b$ *code implies the existence of an* $[n+1, k, \geq d]_q^b$ *code.*

PROOF. If G is the generator matrix of an $[n, k, d]_q^b$ code, then appending a zero vector to G yields a generator matrix of an $[n+1, k, \geq d]_q^b$ code. □

So, by $n_q(k, d)$ we denote the smallest integer n such that an $[n, k, d]_q$ code exists and by $n_q^b(k, d)$ we denote the smallest integer n such that an $[n, k, d]_q^b$ code exists. While the determination of $n_q(k, d)$, for certain parameters, is a classical problem in coding theory, besides some general upper and lower bounds for $n_q^b(k, d)$, not many exact values of $n_q^b(k, d)$ are known. So, we aim to partially close this gap to determining $n_2^2(k, d)$ for small values of the dimension k, see Section 4. For linear codes w.r.t. the Hamming metric e.g. the so-called *Griesmer bound* [10]

$$n \geq \sum_{i=0}^{k-1} \left\lceil \frac{d}{q^i} \right\rceil =: g_q(k, d) \quad (5)$$

relates the parameters of an $[n, k, d]_q$ code. Interestingly enough, this bound can always be attained with equality if the minimum distance d is sufficiently large and a nice geometric construction was given by Solomon and Stiffler [21].[1] In other words, we have $n_q(k,d) = g_q(k,d)$ for all sufficiently large d given k and q. While this solves the asymptotic case, the full determination of the function $n_q(k,\cdot)$ is still a rather challenging problem that is solved in the binary case $q=2$ for dimensions $k \le 8$ only [2]. It is known that $n_q(k,d) = g_q(k,d)$ for $k \le 4$ when $q=2$ and for $k \le 2$ where q is an arbitrary prime power. In all other cases we only know $n_q(k,d) \ge g_q(k,d)$ for all $d \in \mathbb{N}$ and that there exists at least one, but only finitely many, $d \in \mathbb{N}$ such that $n_q(k,d) > g_q(k,d)$.

From the many bounds for $[n,k,d]_q^b$ codes from the literature we would like to single out

$$\frac{q^b-1}{q-1} \cdot n \ge g_q\big(k, q^{b-1} \cdot d\big)\,, \tag{6}$$

see [16, Theorem 1]. More precisely, starting from an $[n,k]_q^b$ code C with minimum distance d and a generator matrix G of a $\left[\frac{q^b-1}{q-1}, b, q^{b-1}\right]_q$ simplex code an $\left[\frac{q^b-1}{q-1} \cdot n, k\right]_q$ code C' is constructed as follows. For each codeword $c \in C$ multiply G with all elements of $\pi_b(c)$ and concatenate the results to a codeword $c' \in C$, so that $\mathrm{wt}_{\mathrm{H}}(c') = \mathrm{wt}_b(c) \cdot q^{b-1}$, since all non-zero codewords of a b-dimensional simplex code over $\mathbb{F}_q$ have Hamming weight q^{b-1}. Applying the Griesmer bound to C' gives Inequality (6). Given an $[n,k,d]_q^b$ code C we call C' the *associated Hamming code* and will state some of its basic properties. To this end let us call an $[n,k]_q$ code or an $[n,k]_q^b$ code Δ-*divisible* if the weight of every codeword c, i.e. $\mathrm{wt}_{\mathrm{H}}(c)$ or $\mathrm{wt}_b(c)$, is divisible by Δ.

Lemma 2.2 *Let C' be the Hamming code associated to an $[n,k,d]_q^b$ code C. Then, C is a q^{b-1}-divisible $\left[\frac{q^b-1}{q-1} \cdot n, k, q^{b-1} \cdot d\right]_q$ code with maximum weight at most $q^{b-1} \cdot n$.*

PROOF. Given the above reasoning it suffices to observe that the maximum weight of C is at most n. □

Example 2.3 *For $n=5$, $q=2$, and $b=2$ let $c_1 = (10101)$, $c_2 = (11110)$, and $c_3 = (11100)$ so that the read vectors are given by $\pi_2(c_1) = \big((1{,}0),(0{,}1),(1{,}0),(0{,}1),(1{,}1)\big)$, $\pi_2(c_2) = \big((1{,}1),(1{,}1),(1{,}1),(1{,}0),(0{,}1)\big)$, and $\pi_2(c_3) = \big((1{,}1),(1{,}1),(1{,}0),(0{,}0),(0{,}1)\big)$. Using the generator matrix $\left(\begin{smallmatrix}1&1&0\\0&1&1\end{smallmatrix}\right)$ for the $[3,2,2]_2$ simplex code, c_1 is mapped to $c_1' = (110\,011\,110\,011\,101)$, c_2 is mapped to $c_2' = (101\,101\,101\,110\,011)$, and c_3 is mapped to $c_3' = (101\,101\,110\,000\,011)$. Setting $C_1 := \langle c_1\rangle$, $C_2 := \langle c_2\rangle$, and $C_3 := \langle c_3\rangle$ we have that the C_i are $[5,1]_2$ codes with weight enumerators $y^5x^0 + x^3y^2$, $y^5x^0 + x^4y^1$, and $y^5x^0 + x^3y^2$, respectively. Moreover, we have $\mathrm{d}_2(C_1) = \mathrm{wt}_H(c_1')/2 = 5$, $\mathrm{d}_2(C_2) = \mathrm{wt}_H(c_2')/2 = 5$, and $\mathrm{d}_2(C_3) = \mathrm{wt}_H(c_3')/2 = 4$.*

[1] More precisely, the cited papers show the statements for field size $q=2$, while they were generalized by other authors slightly later.

Clearly the minimum b-symbol distance $\mathrm{d}_b(C)$ of an $[n,k]_q$ code C can be bounded by its minimum Hamming distance as follows:

$$\min\{\mathrm{d_H}(C)+b-1,n\} \le \mathrm{d}_b(C) \le \min\{b\cdot \mathrm{d_H}(C),n\}. \tag{7}$$

For linear cyclic codes and $b=2$ the lower bound was improved to $\mathrm{d}_2(C) \ge \left\lceil \frac{3\cdot \mathrm{d_H}(C)}{2} \right\rceil$ [22, Lemma 2].

Remark 2.4 *Given an $[n,k,d]_q^b$ code C we can construct a $\left[\frac{q^b-1}{q-1}\cdot n, k, q^{b-1}\cdot d\right]_q$ code C'. The other direction is not always possible. As an example consider the generator matrix*

$$\begin{pmatrix} 101 & 000 & 101 & 011 & 110 \\ 011 & 000 & 011 & 110 & 101 \\ 000 & 101 & 101 & 101 & 101 \\ 000 & 011 & 011 & 011 & 011 \end{pmatrix}$$

that generates a $[15,4,8]_2$ code C'. Here we have grouped the columns in pairs of three, which geometrically corresponds to a so called line spread of $\mathrm{PG}(3,2)$. Indeed we can compute corresponding 2-symbol read vectors from the four generating codewords

$$\begin{pmatrix} 10 & 00 & 10 & 01 & 11 \\ 01 & 00 & 01 & 11 & 10 \\ 00 & 10 & 10 & 10 & 10 \\ 00 & 01 & 01 & 01 & 01 \end{pmatrix}.$$

However, in none of the rows we can find an element $c \in \mathbb{F}_2^5$ such that $\pi_2(c)$ would equal the corresponding row. While we have some freedom in permuting the columns of a generator matrix of a linear code without changing its weight enumerator w.r.t. the Hamming metric we cannot end up with a $[5,4,4]_2^2$ code since such a code does not exist. To this end we note that computing the reduced echelon form of the generator matrix of a code does not change the code and that the possible generator matrices in reduced form echelon are given by

$$\begin{pmatrix} 1000* \\ 0100* \\ 0010* \\ 0001* \end{pmatrix}, \begin{pmatrix} 100{*}0 \\ 010{*}0 \\ 001{*}0 \\ 00001 \end{pmatrix}, \begin{pmatrix} 10{*}00 \\ 01{*}00 \\ 00010 \\ 00001 \end{pmatrix}, \begin{pmatrix} 1{*}000 \\ 00100 \\ 00010 \\ 00001 \end{pmatrix}, \begin{pmatrix} 01000 \\ 00100 \\ 00010 \\ 00001 \end{pmatrix}.$$

Since the fourth row always starts with at least three consecutive zeroes the minimum symbol-pair distance is at most $5-2=3$ and not 4.

It is well known that linear codes correspond to multisets of points in projective geometries [8]. The set of all subspaces of $\mathbb{F}_q^r$, ordered by the incidence relation $\subseteq$, is called

$(r-1)$-*dimensional projective geometry over* $\mathbb{F}_q$ and denoted by $\mathrm{PG}(r-1,q)$. Employing this algebraic notion of dimension instead of the geometric one, we will use the term i-space to denote an i-dimensional subspace of $\mathbb{F}_q^r$. To highlight the important geometric interpretation of subspaces we will call 1-, $2-$, and $(r-1)$-spaces points, lines, and hyperplanes, respectively. For two subspaces S and S' we write $S \le S'$ if S is contained in S'. Moreover, we say that S and S' are *incident* iff $S \le S'$ or $S \ge S'$. Let $[i]_q := \frac{q^i-1}{q-1}$ denote the number of points of an arbitrary i-space in $\mathrm{PG}(r-1,q)$ where $r \ge i$. Here we describe a multiset of points by a mapping $\mathcal{M}$ from the set of points of $\mathrm{PG}(k-1,q)$ to $\mathbb{N}$ and call $\mathcal{M}(P)$ the multiplicity of point P. The *cardinality* of $\mathcal{M}$ is given by $\#\mathcal{M} = \sum_P \mathcal{M}(P)$. For some subspace S we define the multiplicity of S by $\mathcal{M}(S) = \sum_{P \le S} \mathcal{M}(P)$ and let χ_S denote the characteristic function of S, i.e., $\chi_S(P) = 1$ if $P \le S$ and $\chi_S(P) = 0$ otherwise. A multiset of points is called *spanning* if the set of points with positive multiplicity span the entire space. With this, we can state more precisely that each $[n,k,d]_q$ code with dual minimum distance at least two, i.e. for each coordinate there exists a codeword with non-zero entry at this position, is in one-to-one correspondence to a spanning multiset $\mathcal{M}$ in $\mathrm{PG}(k-1,q)$ with cardinality n and $\mathcal{M}(H) \le n-d$ for every hyperplane H, where equality occurs at least once.

Multisets of points can be generalized as follows, see [1, Definition 4] and [15, Definition 1].

Definition 2.5 *A* projective $h-(n,r,s)_q$ system *is a multiset* $\mathcal{S}$ *of* n *subspaces of* $\mathrm{PG}(r-1,q)$ *of dimension at most* h *such that each hyperplane contains at most* s *elements of* $\mathcal{S}$ *and some hyperplane contains exactly* s *elements of* $\mathcal{S}$. *We say that* $\mathcal{S}$ *is* faithful *if all of its elements have dimension* h. *A projective* $h-(n,r,s)_q$ *system* $\mathcal{S}$ *is a projective* $h-(n,r,s,\mu)_q$ system *if each point is contained in at most* μ *elements from* $\mathcal{S}$ *and there is some point that is contained in exactly* μ *elements from* $\mathcal{S}$.

Allowing 0-spaces, corresponding to zero-columns in the generator matrix of a linear code, one can say that $[n,k,d]_q$ codes are in one-to-one correspondence to projective $1-(n,k,n-d)_q$ systems. In general, projective $h-(n,r,s)_q$ systems (with $s<n$) are in one-to-one correspondence to additive codes, see [1] or [15] for details. Here, we call a subset of $\mathbb{F}_q^n$ an *additive code* iff the sum of any two codewords is also a codeword, i.e. additive codes are a super class of linear codes.

Lemma 2.6 *To each* $[n,k,d]_q^b$ *code* C, *where* $k \ge b$, *we can associate a projective* $b-(n,k,n-d)_q$ *system* $\mathcal{S} = \{S_0,\dots,S_{n-1}\}$. *Moreover, we have*

$$\dim(S_i \cap S_{i+1}) \ge \max\{\dim(S_i), dim(S_{i+1})\} - 1 \tag{8}$$

for each index $0 \le i \le n-1$, *where the indices are read modulo* n.

PROOF. Let G be a generator matrix of C and let $g_0,\dots,g_{n-1}$ denote the ordered list of columns of G. For each index $0 \le n-1$ let S_i denote the subspace spanned by $g_i, g_{i+1},\dots,g_{i+b-1}$, where the indices are read modulo n. With this we let $\mathcal{S}$ consist of the S_i, which have dimension at most b and are contained in $\operatorname{PG}(k-1,q)$. Consider an arbitrary non-zero codeword $c=(c_0,\dots,c_{n-1})\in C$. There exists a unique row vector $h\in\mathbb{F}_q^k\backslash\{0\}$ such that $c=hG$. To h we can assign the set of all points P such that $hP=0$, which is a hyperplane H that is equal for all non-zero multiples of h. Note that we have $(c_i,c_{i+1},\dots,c_{i+b-1})=0$ iff $S_i\le H$, where the indices are read modulo n. Thus, every hyperplane H in $\operatorname{PG}(k-1,q)$ contains at most $n-d$ elements from $\mathcal{S}$ and equality indeed occurs. Thus, $\mathcal{S}$ is a a projective $b-(n,k,n-d)_q$ system. Inequality (8) directly follows from the construction since $S_i\cap S_{i+1}$ is generated by $g_{i+1},g_{i+2},\dots,g_{i+b-1}$. □

Example 2.7 *In* $\operatorname{PG}(2,2)$ *consider the ordered list of subspaces*

$$S_0=\left\langle \begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}1\\0\\0\end{pmatrix}\right\rangle, S_1=\left\langle \begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}\right\rangle, S_2=\left\langle \begin{pmatrix}0\\0\\1\end{pmatrix},\begin{pmatrix}1\\0\\1\end{pmatrix}\right\rangle, S_3=\left\langle \begin{pmatrix}1\\0\\1\end{pmatrix},\begin{pmatrix}1\\1\\1\end{pmatrix}\right\rangle, S_4=\left\langle \begin{pmatrix}1\\1\\1\end{pmatrix}\right\rangle$$

In $\operatorname{PG}(2,2)$ *the* $[3]_2=7$ *hyperplanes are indeed lines. Since the point* S_4 *is contained in the three hyperplanes* S_0*,* S_3*,* $\left\langle(1,1,1)^\top,(0,0,1)^\top\right\rangle$ *and* $S_1=S_2$ *there are exactly three hyperplanes that contain two elements from the multiset* $\mathcal{S}:=\{S_0,\dots,S_4\}$*, one hyperplane that contains one element from* $\mathcal{S}$*, and three hyperplanes that contain no element from* $\mathcal{S}$*. So,* $\mathcal{S}$ *is a projective* $2-(5,3,2)_2$ *system. Noting that Inequality (8) is satisfied let us try to reverse Lemma 2.6 and build up a generator matrix G with columns* $g_0,\dots,g_4$ *as in the corresponding proof. Except for* $S_1\cap S_2$*, the intersections* $S_i\cap S_{i+1}$ *determines a unique point, which leaves the three choices*

$$G_1=\begin{pmatrix}11111\\00011\\00111\end{pmatrix}, G_2=\begin{pmatrix}11111\\00011\\01111\end{pmatrix}, G_3=\begin{pmatrix}10111\\00011\\01111\end{pmatrix}$$

having weight enumerators $x^0y^5+x^2y^3+2x^3y^2+2x^4y^1+2x^5y^0$*,* $x^0y^5+x^2y^3+2x^3y^2+2x^4y^1+2x^5y^0$*, and* $x^0y^5+3x^3y^2+x^4y^1+3x^5y^0$*, respectively. Here, applying Lemma 2.6 to* G_i *only gives* $\mathcal{S}$ *for* $i=3$*.*

Now let $\mathcal{S}'$ *arise from* $\mathcal{S}$ *via replacing* S_4 *by* $\left\langle(1,1,1)^\top,(0,1,0)^\top\right\rangle$*. If we could find a generator matrix* G' *such that the application of the construction in the proof of Lemma 2.6 gives* $\mathcal{S}'$*, then* G' *would generate a* $[5,3,2]_2^2$ *code with weight enumerator* $x^0y^5+x^3y^2+3x^4y^1+3x^5y^0$*. However, such a generator matrix does not exist.*

So, by Lemma 2.6 upper bounds for the cardinality n of projective $b-(n,k,n-d)_q$ systems or the corresponding additive codes, see e.g. [15] and the references therein, imply lower bounds for $n_q^b(k,d)$. It is interesting to note that additive codes also have to satisfy a Griesmer type inequality like Inequality (6), which can always be attained with equality for sufficiently larger minimum distance d, see [15] for details. In our situation we additionally need to ensure that the projective system $b-(n,k,n-d)_q$ system can be written as a list of subspaces satisfying Inequality (8), which is even not sufficient as shown in Exercise 2.7. For parameters k, d, q, and b we call

$$n \geq \left\lceil \frac{g_q\big(k, q^{b-1}\cdot d\big)}{[b]_q} \right\rceil \tag{9}$$

the *Griesmer bound* for an $[n,k,d]_q^b$ code, c.f. [16, Theorem 1]. The aim of the subsequent Section 3 is to show that the Griesmer bound can always be attained with equality if the minimum distance d is assumed to be sufficiently large.

Lemma 2.8 *(C.f. [12, Theorem 3]) We have*

$$\left\lceil \frac{g_q\big(k, q^{b-1}\cdot d\big)}{[b]_q} \right\rceil = d + \left\lceil \frac{g_q(k-b+1,d)-d}{[b]_q} \right\rceil. \tag{10}$$

PROOF.

$$\left\lceil \frac{g_q\big(k, q^{b-1}\cdot d\big)}{[b]_q} \right\rceil = \left\lceil \frac{\sum_{i=0}^{k-1} \left\lceil \frac{d\cdot q^{b-1}}{q^i} \right\rceil}{[b]_q} \right\rceil = \left\lceil \frac{\sum_{i=0}^{b-1} \left\lceil d\cdot q^{b-i-1} \right\rceil + \sum_{i=b}^{k-1} \left\lceil d\cdot q^{b-i-1} \right\rceil}{[b]_q} \right\rceil$$

$$= \left\lceil \frac{d\cdot \sum_{i=0}^{b-1} q^i + \sum_{i=1}^{k-b} \left\lceil \frac{d}{q^i} \right\rceil}{[b]_q} \right\rceil = \left\lceil \frac{d[b]_q - d + \sum_{i=0}^{k-b} \left\lceil \frac{d}{q^i} \right\rceil}{[b]_q} \right\rceil = d + \left\lceil \frac{g_q(k-b+1,d)-d}{[b]_q} \right\rceil$$

□

Using a specific parameterization of the minimum distance d of a linear code in the Hamming metric, the corresponding Griesmer bound in Inequality (5) can be written more explicitly:

Lemma 2.9 *Let k and d be positive integers. Write d as*

$$d = \sigma q^{k-1} - \sum_{i=1}^{k-1} \varepsilon_i q^{i-1}, \tag{11}$$

where $\sigma \in \mathbb{N}_0$ and the $0 \le \varepsilon_i < q$ are integers for all $1 \le i \le k-1$. Then, Inequality (5) is satisfied with equality iff

$$n = \sigma[k]_q - \sum_{i=1}^{k-1} \varepsilon_i [i]_q, \tag{12}$$

which is equivalent to

$$n - d = \sigma[k-1]_q - \sum_{i=1}^{k-1} \varepsilon_i [i-1]_q. \tag{13}$$

Remark 2.10 *Given k and d Equation (11) always determines σ and the ε_i uniquely. This is different for Equation (13) given k and $n - d = s$. Here it may happen that no solution with $0 \le \varepsilon_i \le q-1$ exists. By relaxing to $0 \le \varepsilon_i \le q$ we can ensure existence and uniqueness is enforced by additionally requiring $\varepsilon_j = 0$ for all $j < i$ where $\varepsilon_i = q$ for some i. The same is true for Equation (12) given k and n. For more details we refer to [9, Chapter 2] which also gives pointers to Hamada's work on minihypers.*

Lemma 2.11

(a) $g_q\big(k, \lambda \cdot q^{k-1}\big) = \lambda [k]_q$ *for each* $\lambda \in \mathbb{N}$.

(b) $\left\lceil \frac{g_q\left(k, q^{b-1} \cdot d\right)}{[b]_q} \right\rceil = [k]_q$ *for* $d = q^{k-b} \cdot [b]_q$.

(c) For each $\lambda, d' \in \mathbb{N}$ *we have*

$$\left\lceil \frac{g_q\Big(k, q^{b-1} \cdot \big(\lambda \cdot q^{k-b} \cdot [b]_q + d'\big)\Big)}{[b]_q} \right\rceil = \lambda \cdot [k]_q + \left\lceil \frac{g_q\big(k, q^{b-1} \cdot d'\big)}{[b]_q} \right\rceil. \tag{14}$$

PROOF. Parts (a), (b) directly follow from Lemma 2.9, which then imply (c). □

3 Linear codes attaining the Griesmer bound

In this section we want to construct optimal $[n, k, d]_q^b$ codes. One possible construction is to start with a cyclic group G in $\mathrm{GL}(k, q)$ generated by some element $g \in G$. Denoting the action of g on a point P by P^g, we can partition the set of points of $\mathrm{PG}(k-1, q)$ into orbits of the form $P, P^g, P^{g^2}, P^{g^3}, \dots, P^{g^{l-1}}$, where the length l of the orbit can be different for different starting points P. We consider the sequence of points $P, P^g, P^{g^2}, P^{g^3}, \dots, P^{g^{l-1}}$ as a generator matrix of an $[l, k', d]_q^b$ code C with $k' \le k$. In some cases we have $k' = k$ and d is suitably large w.r.t. the other parameters.

Example 3.1 *In* $\mathrm{GL}(5,2)$ *there exist six cyclic groups of order* $2^5-1=31$. *Choosing a generator matrix arising from the orbit of all points with respect to a generator of the cyclic group yields a* $[31,5,24]_2^2$ *code in all cases.*

- $\begin{pmatrix} 01000 \\ 00100 \\ 00010 \\ 00001 \\ 10111 \end{pmatrix} : \begin{pmatrix} 0110010011111011100010101101000 \\ 0011001001111101110001010110100 \\ 0111110111000101011010000110010 \\ 0101101000011001001111101110001 \\ 1100100111110111000101011010000 \end{pmatrix},$

- $\begin{pmatrix} 01000 \\ 00100 \\ 00010 \\ 00001 \\ 11110 \end{pmatrix} : \begin{pmatrix} 0101101010001110111110010011000 \\ 0111011111001001100001011010100 \\ 0110000101101010001110111110010 \\ 0110101000111011111001001100001 \\ 1011010100011101111100100110000 \end{pmatrix},$

- $\begin{pmatrix} 01000 \\ 00100 \\ 00010 \\ 00001 \\ 11101 \end{pmatrix} : \begin{pmatrix} 0111001101111101000100101011000 \\ 0100101011000011100110111110100 \\ 0101011000011100110111110100010 \\ 0010101100001110011011111010001 \\ 1110011011111010001001010110000 \end{pmatrix},$

- $\begin{pmatrix} 01000 \\ 00100 \\ 00010 \\ 00001 \\ 11011 \end{pmatrix} : \begin{pmatrix} 0110101001000101111101100111000 \\ 0101111101100111000011010100100 \\ 0010111110110011100001101010010 \\ 0111110110011100001101010010001 \\ 1101010010001011111011001110000 \end{pmatrix},$

- $\begin{pmatrix} 01000 \\ 00100 \\ 00010 \\ 00001 \\ 10010 \end{pmatrix} : \begin{pmatrix} 0101011101100011111001101001000 \\ 0010101110110001111100110100100 \\ 0001010111011000111110011010010 \\ 0101110110001111100110100100001 \\ 1010111011000111110011010010000 \end{pmatrix},$

- $\begin{pmatrix} 01000 \\ 00100 \\ 00010 \\ 00001 \\ 10100 \end{pmatrix} : \begin{pmatrix} 0100101100111110001101110101000 \\ 0010010110011111000110111010100 \\ 0101100111110001101110101000010 \\ 0010110011111000110111010100001 \\ 1001011001111100011011101010000 \end{pmatrix}.$

Interestingly enough, deleting one arbitrary column from the generator matrices yields an $[30,5,22]_2^2$ *code in all cases, i.e. the minimum symbol pair distance is decreased by two while the length is just increased by one.*

In $\mathrm{GL}(k,q)$ the maximum order of a cyclic group is $[k]_q$ and it acts transitively on the set of points in $\mathrm{PG}(k-1,q)$. Those groups exist for all parameters and are called *Singer groups*, see e.g. [11].

Proposition 3.2 *For* $k \ge b \ge 2$ *there exists an* $[n,k,d]_q^b$ *code* C *with* $n=[k]_q$ *and* $d=[b]_q\cdot q^{k-b}$.

PROOF. Let $M\in\mathrm{GL}(k,q)$ be a $k\times k$ matrix over $\mathbb{F}_q$ that generates a cyclic group of order $n=[k]_q$, i.e. a *Singer cycle*. Denoting the first unit vector by e_1 we construct a generator matrix G for C by choosing the sequence of columns $M^0e_1, M^1e_1, \dots, M^{n-1}e_1$. Note that M acts transitively on the set of n points of $\mathrm{PG}(k-1,q)$. Set

$$S_i := \left\langle M^ie_1, M^{i+1}e_1, \dots, M^{i+b-1}e_1\right\rangle \tag{15}$$

for all $0 \le i \le n-1$, where the indices are read modulo n. Let $P := M^i e_1$ and suppose that there exist $\lambda_0, \dots, \lambda_{b-2} \in \mathbb{F}_q$ such that $M^{b-1}P = \sum_{i=0}^{b-2} M^i P$. Then, $M^0 P, M^1 P, \dots$ is contained in $\left\langle M^0 P, M^1 P, \dots, M^{b-2} P \right\rangle$, which is impossible since M acts transitively on the set of points and $k \ge b$. Thus, we have $\dim(S_i) = b$ for all $0 \le i \le n-1$.

Let $\mathcal{M} := \sum_{i=0}^{n-1} \chi_{S_i}$, so that $\#\mathcal{M} = n[b]_q = [k]_q[b]_q$. Note that each point P in S_i can be written as $\left\langle \sum_{j=0}^{b-1} \lambda_j M^{i+j} e_1 \right\rangle = M^i \cdot \left\langle \sum_{j=0}^{b-1} \lambda_j M^j e_1 \right\rangle$, where $(\lambda_0, \dots, \lambda_{b-1}) \in \mathbb{F}_q^b \backslash \{0\}$ is uniquely determined. Thus, the transitivity of M on the set of points implies $\mathcal{M}(P) = \mathcal{M}(P')$ for all pairs of points P, P'. Counting points then yields $\mathcal{M}(P) = [b]_q$ for every point P. Since each of the b-spaces S_i intersects a given hyperplane H in either $[b]_q$ or $[b-1]_q$ points and $[b]_q = q^{b-1} + [b-1]_q$ we have

$$[k-1]_q[b]_q = \mathcal{M}(H) = [k]_q[b-1]_q + s \cdot q^{b-1}, \tag{16}$$

where s denotes the number of indices $0 \le i \le n-1$ with $S_i \le H$. Since

$$\begin{aligned} [k-1]_q[b]_q - [k]_q[b-1]_q &= \frac{\left(q^{k-1}-1\right) \cdot \left(q^b-1\right) - \left(q^k-1\right) \cdot \left(q^{b-1}-1\right)}{(q-1)^2} \\ &= \frac{q^k + q^{b-1} - q^{k-1} - q^b}{(q-1)^2} = \frac{q^{k-1} - q^{b-1}}{q-1} = q^{b-1} \cdot [k-b]_q, \end{aligned}$$

we conclude $s = [k-b]_q$. Since

$$n - s = [k]_q - [k-b]_q = \frac{q^k - q^{k-b}}{q-1} = q^{k-b} \cdot \frac{q^b - 1}{q-1} = q^{k-b} \cdot [b]_q,$$

the constructed code has minimum distance $\mathrm{d}_b(C) = [b]_q \cdot q^{k-b}$. □

In the Hamming metric the constructed code C is just a simplex code. The minimum distance of those codes w.r.t. the b-symbol metric was studied in [18] for the special case $b \le q-1$. Due to Lemma 2.11.(b) the codes from Proposition 3.2 attain the Griesmer bound of Inequality (9). Appending λ copies of the corresponding generator matrix yields $[n,k,d]_q^b$ codes with $n = \lambda \cdot [k]_q$ and $d = \lambda \cdot [b]_q \cdot q^{k-b}$, so that Lemma 2.11.(c) gives

$$\lim_{d \to \infty} n_q^b(k,d) / \left\lceil \frac{g_q\left(k, q^{b-1} \cdot d\right)}{[b]_q} \right\rceil = 1. \tag{17}$$

Lemma 3.3 *Let G_1 be a generator matrix of an $[n_1, k, d_1]_q^b$ code and G_2 be a generator matrix of an $[n_2, k, d_2]_q^b$ code. If the first $b-1$ columns of G_1 and G_2 coincide, then appending G_2 to G_1 is the generator matrix of an $[n_1+n_2, k, d_1+d_2]_q^b$ code.*

Definition 3.4 *We call an $[n,k]_q^b$ code C* faithful *if the restriction of the codewords to the coordinates $i, i+1, \dots i+b-1$ (read modulo n) form an $[b,b]_q^b$ code for each index $0 \le i \le n-1$.*

In other words, an $[n,k,d]_q^b$ code is faithful iff the projective $b-(n,k,n-d)_q$ system is faithful, where the generator matrix can be chosen arbitrarily. If $\mathcal{S}$ is an arbitrary $h-(n,k,s)_q$ system, then we can obtain a faithful $h-(n,k,\le s)_q$ system $\mathcal{S}'$ by replacing each element $S \in \mathcal{S}$ with $\dim(S) < h$ by a b-space S' with $S \le S'$. A similar results also holds for $[n,k]_q^b$ codes.

Lemma 3.5 *Let C be an $[n,k,d]_q^b$ code with $k \ge b$. Then, there exists a faithful $[n,k,\ge d]_q^b$ code C'.*

PROOF. We iteratively construct a generator matrix G_i of an $[n,k,\ge d]_q^b$ code such that i subsequent columns span an i-space for all $0 \le i \le b$, where the column indices are read modulo n. Going from G_i to G_{i+1} we modify just one column, so that the length remain n during our entire construction. The improvement property we maintain during the construction is the following. On the way from G_i to G_{i+1} we assume that we have already constructed a generator matrix U of an an $[n,k,\ge d]_q^b$ code such that i subsequent columns span an i-space. In the next step we construct a generator matrix U' of an an $[n,k,\ge d]_q^b$ code such that i subsequent columns span an i-space, $\langle u_j, u_{j+1}, \dots, u_{j+i}\rangle \le \left\langle u'_j, u'_{j+1}, \dots, u'_{j+i}\right\rangle$ for all $0 \le j \le n-1$ and that the dimension increases at least once, where $u_0, \dots, u_{n-1}$ are the columns of U, $u'_0, \dots, u'_{n-1}$ are the columns of U', and the indices are read modulo n. Let us denote the index where we change a column by h, i.e., we have $u_j = u'_j$ for all $0 \le j \le n-1$ with $j \ne h$. Let us first observe that U' generates an $[n,k]_q^b$ code if U does. To this end we consider

$$\begin{aligned}
& \langle u'_0, u'_1, \dots, u'_{n-1}\rangle \\
= \;& \left\langle \langle u'_h, u'_{h+1}, \dots u'_{h+i},\rangle, \langle u'_j | 0 \le j \le n-1, j \notin [h, h+i], j \notin [h-n, h+i-n]\rangle \right\rangle \\
= \;& \left\langle \langle u'_h, u'_{h+1}, \dots u'_{h+i},\rangle, \langle u_j | 0 \le j \le n-1, j \notin [h, h+i], j \notin [h-n, h+i-n]\rangle \right\rangle \\
\ge \;& \left\langle \langle u_h, u_{h+1}, \dots u_{h+i},\rangle, \langle u_j | 0 \le j \le n-1, j \notin [h, h+i], j \notin [h-n, h+i-n]\rangle \right\rangle \\
= \;& \langle u_0, u_1, \dots, u_{n-1}\rangle.
\end{aligned}$$

We can perform the same computation for the span of b subsequent columns of U and U' to conclude that the minimum b-symbol distance does not decrease.

As G_0 we can take an arbitrary generator matrix G of C. For the construction of G_1 we iteratively replace each occurring zero vector by an arbitrary non-zero vector. Clearly, we have $\langle u_j\rangle \le \left\langle u'_j\right\rangle$ for all $0 \le j \le n-1$ since $u_j = u'_j$ if $j \ne h$ and u'_h is the zero vector. Moreover, we have $\dim(\langle u'_h \rangle) = 1 > 0 = dim(\langle u_h \rangle)$. Here we assume that we directly set $G_{i+1} = G_i$, if G_i already has the desired property of G_{i+1}, and do not consider modifications from U to U'.

Now we assume that we have a generator matrix U such that all i subsequent columns span an i-space and that there exists a index $0 \le j \le n-1$ such that the columns $u_j, \dots, u_{j+i}$ also span an i-space and indeed not span an $(i+1)$-span. So, we have $\langle u_j, \dots, u_{j+i-1} \rangle = \langle u_{j+1}, \dots, u_{j+i} \rangle$. Since the span of all n columns has dimension k, we can assume $\langle u_{j+1}, \dots, u_{j+i} \rangle \neq \langle u_{j+2}, \dots, u_{j+i+1} \rangle$ for $i < b \le k$ by possibly increasing the initial index j. With this we choose $h = j + i$ (modulo n) and set $u'_h = u_h + u_{h+1}$. For brevity we set $\mathcal{S}_l := \langle u_l, \dots, u_{l+i-1} \rangle$, $\mathcal{S}'_l := \langle u'_l, \dots, u'_{l+i-1} \rangle$, $\mathcal{T}_l := \langle u_l, \dots, u_{l+i} \rangle$, and $\mathcal{T}'_l := \langle u'_l, \dots, u'_{l+i} \rangle$ for all $0 \le l \le n-1$, so that e.g. $\mathcal{S}_j = \mathcal{S}_{j+1} \neq \mathcal{S}_{j+2}$ and $\dim(\mathcal{T}_j) = i$. Since $\langle u_h, u_{h+1} \rangle = \langle u'_h, u'_{h+1} \rangle$ we have $\mathcal{S}_l = \mathcal{S}'_l$ for $l \notin \{j+1, j+1-n\}$ and $\mathcal{T}_l = \mathcal{T}'_l$ $l \notin \{j, j-n\}$. Since $\mathcal{S}_{j+1} \neq \mathcal{S}_{j+2}$ we have $\mathcal{T}'_j > \mathcal{T}_j$, $\dim(\mathcal{S}'_{j+1}) = i$, $\mathcal{S}'_j \neq \mathcal{S}'_{j+1}$, and $\mathcal{S}'_{j+1} \neq \mathcal{S}'_{j+2}$. Thus, U' satisfies the required property. □

We remark that the code constructed in Proposition 3.2 is faithful.

Lemma 3.6

$$n_q^b(k, d_1 + d_2) \le n_q^b(k, d_1) + n_q^b(k, d_2)$$

PROOF. Let C_i be faithful $[n_i, k, d_i]_q^b$ codes with $n_i = n_q^b(k, d_i)$ and G_i be corresponding generator matrices for $i \in \{1, 2\}$, see Lemma 3.5. Multiplying G_i by a suitable matrix in $\mathrm{GL}(k, q)$ we obtain a generator matrix G'_i of a faithful $[n_i, k, d_i]_q^b$ code starting with a $k \times k$ unit matrix I_k, where $i \in \{1, 2\}$. Applying Lemma 3.3 to G'_1 and G'_2 yields a generator matrix of an $[n_1 + n_2, k, d_1 + d_2]_q^b$ code. □

For the 2-symbol metric we can directly state the resulting minimum distance for the concatenation of two arbitrary generator matrices. For simplicity we state the observation in terms of codewords.

Lemma 3.7 *Let $a \in \mathbb{F}_q^{n_1}$ and $b \in \mathbb{F}_q^{n_2}$, where $n_1, n_2 \ge 2$, such that $\mathrm{wt}_2(a) = d_1$ and $\mathrm{wt}_2(b) = d_2$. Construct $c \in \mathbb{F}_q^{n_1+n_2}$ as the concatenation of a and b – written $c = a|b$. Writing $\star$ for an arbitrary element in $\mathbb{F}_q \backslash \{0\}$ we have the following four cases for a:*

(a) $a = (0, a_1, \dots, a_{n-2}, 0)$;

(b) $a = (0, a_1, \dots, a_{n-2}, \star)$;

(c) $a = (\star, a_1, \dots, a_{n-2}, 0)$;

(d) $a = (\star, a_1, \dots, a_{n-2}, \star)$.

Similar for b:

(i) $b = (0, b_1, \dots, b_{n-2}, 0)$*;*

(ii) $b = (0, b_1, \dots, b_{n-2}, \star)$*;*

(iii) $b = (\star, b_1, \dots, b_{n-2}, 0)$*;*

(iv) $b = (\star, b_1, \dots, b_{n-2}, \star)$*.*

With this, we have $\mathrm{wt}_2(c) = \mathrm{wt}_2(a) + \mathrm{wt}_2(b) - 1$ *if we are in case (b).(iii) or case (c).(ii) and* $\mathrm{wt}_2(c) = \mathrm{wt}_2(a) + \mathrm{wt}_2(b)$ *in all other cases.*

Corollary 3.8 *Let $n \geq 2$, $k \geq 2$, and $t \geq 1$ be integers. For each $c \in \mathbb{F}_q^n$ with $n \geq 2$ we have $\mathrm{wt}_2(\overbrace{c|\cdots|c}^{t \text{ times}}) = t \cdot \mathrm{wt}_2(c)$. For each full rank matrix $G \in \mathbb{F}_q^{k \times n}$ we have $\mathrm{d}_2\big(\mathrm{rowspan}\big(\overbrace{G|\dots|G}^{t \text{ times}}\big)\big) = t \cdot \mathrm{d}_2\big(\mathrm{rowspan}(G)\big)$.*

Our next aim is to show

$$\lim_{d\to\infty} n_q^b(k,d) - \left\lceil \frac{g_q\big(k, q^{b-1}\cdot d\big)}{[b]_q} \right\rceil = 0. \tag{18}$$

Lemma 3.9 *Let G_i be generator matrices of faithful $[n_i, k, d_i]_q^b$ codes for $0 \leq i < l$ such that the last $b-1$ columns of G_i coincide with the first $b-1$ columns of G_{i+1} for all $0 \leq i < l$, where the indices are read modulo l. Then, the concatenation of the matrices $G_0, G_1, \dots, G_{l-1}$ is the generator matrix of a faithful $\left[\sum_{i=0}^{l-1} n_i, k, \sum_{i=0}^{l-1} d_i\right]_q^b$ code.*

Example 3.10 *The matrix*

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \end{pmatrix} \in \mathrm{GL}(5,2)$$

generates the orbit of points

$$G=\begin{pmatrix}1000011001001111101110001010110\\0000110010011111011100010101101\\0001100100111110111000101011010\\0011001001111101110001010110100\\0110010011111011100010101101000\end{pmatrix}$$

when starting from the first unit vector. We write v_j for the vector $v_j \in \mathbb{F}_2^5 \backslash \{0\}$ which coincides with the binary expansion of $j \le i \le 31$. So for the subspaces $\mathcal{S}_i$ according to Lemma 2.6 we e.g. have $\mathcal{S}_0 = \langle v_{16}, v_1, v_3 \rangle$, $\mathcal{S}_{29} = \langle v_{20}, v_8, v_{16} \rangle$, and $\mathcal{S}_{30} = \langle v_8, v_{16}, v_1 \rangle$. More precisely, the first two columns of G equal v_{16}, v_1 and the last column of G equals v_8. Setting $\mathcal{S}'_{29} = \langle v_{20}, v_8, v_{16} + v_{20} \rangle$, $\mathcal{S}'_{30} = \langle v_8, v_{16} + v_{20}, v_1 \rangle$, $\mathcal{S}''_{29} = \langle v_{20}, v_8, v_{16} \rangle$, and $\mathcal{S}''_{30} = \langle v_8, v_{16}, v_1 + v_{20} \rangle$, we note $\mathcal{S}_{29} = \mathcal{S}'_{29} = \mathcal{S}''_{29}$ and $\mathcal{S}_{30} = \mathcal{S}'_{30} = \mathcal{S}''_{30}$. So, we may either append the vectors v_{16}, v_1 to G or the vectors $v_{16} + v_{20}, v_1$ or the vectors $v_{16}, v_1 + v_{20}$ so that the first 31 planes coincide.

Definition 3.11 *A* b-chain of length n over $\mathbb{F}_q^k$ *is a list of vectors $v_0, \ldots, v_{n+b-2} \in \mathbb{F}_q^k$. The sublist $v_0, v_1, \ldots, v_{b-2}$ is called the* start *and the sublist $v_n, v_{n+1}, \ldots, v_{n+b-2}$ is called the* end. *The* associated projective $b-(n,k,s)_q$ system $\mathcal{M}$ *consists of the spaces $\langle v_i, \ldots v_{i+b-1} \rangle$ for $0 \le i \le n-1$ and s is the maximum number of those subspaces that are contained in some hyperplane of $\mathrm{PG}(k-1,q)$. Two b-chains of length n over $\mathbb{F}_q^k$ are called* equivalent *if their associated projective $b-(n,k,s)_q$ systems coincide.*

Directly from the definition we verify:

Lemma 3.12 *Let $v_0, \ldots, v_{n+b-2} \in \mathbb{F}_q^k$ be a b-chain of length n with associated projective $n-(n,k,s)_q$ system $\mathcal{M}$ and $v'_0, \ldots, v'_{n'+b-2} \in \mathbb{F}_q^k$ be a b-chain of length n' with associated projective $n-(n',k,s')_q$ system $\mathcal{M}'$. If $v_n, \ldots v_{n+b-2} = v'_0, \ldots, v'_{b-2}$, i.e. if the end of the first chain equals the end of the second chain, then $v_0, \ldots, v_{n-1}, v'_0, \ldots, v_{n'+b-2}$ is a b-chain of length $n+n'$ with associated projective $b-(n+n',k,s+s')_q$ system $\mathcal{M}+\mathcal{M}'$.*

Lemma 3.13 *Let $v_0, \ldots, v_{b-2} \in \mathbb{F}_q^k$ span a $(b-1)$-space, $h \in \{0, \ldots, b-2\}$, $v'_0, \ldots, v'_{b-2} \in \mathbb{F}_q^k$ with $v'_i = v_i$ for $i \neq h$, $\dim(\langle v'_0, \ldots, v'_{b-2} \rangle) = b-1$, and $\langle v_0, \ldots, v_{b-2} \rangle \neq \langle v'_0, \ldots, v_{b-2} \rangle$. Then, there exist a b-chain of length $[k]_q$ with start $v_0, \ldots, v_{b-2}$, end $v'_0, \ldots, v'_{b-2}$, and associated projective $b-\left([k]_q, k, [k-b]_q\right)_q$ system.*

PROOF. First we note $\dim(\langle v_0, \ldots, v_{b-2}, v'_h \rangle) = b$. From Proposition 3.2 we conclude the existence of a faithful $\left([k]_q, k, [b]_q \cdot q^{k-b}\right)_q^b$ code C. Let G denote a generator matrix of G

and $g_0, \dots, g_{n-1}$ be its columns, where $n := [k]_q$. Noting that $\dim(\langle g_0, \dots, g_{b-2}, g_{n-1}\rangle) = b$ and $\dim(\langle v_0, \dots, v_{b-2}, v_h + v'_h\rangle) = b$ we conclude the existence of a matrix $M \in \mathrm{GL}(k, q)$ that maps g_i to v_i for $0 \le i \le b-2$ and g_{n-1} to $v_h + v'_h$. With this, $G' = M \cdot G$ starts with $v_0, \dots, v_{b-2}$, ends with $v_h + v'_h$, and generates a faithful $\left([k]_q, k, [b]_q \cdot q^{k-b}\right)_q^b$ code. From this we obtain a b-chain with length n over $\mathbb{F}_q^k$ with start $v_0, \dots, v_{b-2}$, end $v_0, \dots, v_{b-2}$, and associated projective $b-(n, k, s)_q$ system, where $s = [k]_q - [b]_q \cdot q^{k-b} = [k-b]_q$. Since $v_{n-1} = v_h + v'_h$ this chain is equivalent to a b-chain with length n over $\mathbb{F}_q^k$ with start $v_0, \dots, v_{b-2}$, end $v'_0, \dots, v'_{b-2}$, and associated projective $b-(n, k, s)_q$ system, where $s = [k-b]_q$. □

Lemma 3.14 *Let $v_0, \dots, v_{b-2} \in \mathbb{F}_q^k$ and $v'_0, \dots, v'_{b-2} \in \mathbb{F}_q^k$ both span a $(b-1)$-space. Then, there exists a b-chain of length $\lambda \cdot [k]_q$ with start $v_0, \dots, v_{b-2}$, end $v'_0, \dots, v'_{b-2}$, and associated projective $b - \left(\lambda \cdot [k]_q, k, \lambda \cdot [k-b]_q\right)_q$ system for some positive integer λ.*

PROOF. If $v'_0 \notin \langle v_0, \dots, v_{b-2}\rangle$ (or $v_0 = v'_0$), then Lemma 3.13 yields a b-chain with length $[k]_q$ over $\mathbb{F}_q^k$ that has start $v_0, \dots, v_{b-2}$, end $v'_0, v_1, \dots, v_{b-2}$, and an associated projective $b-\left([k]_q, k, [k-b]_q\right)_q$ system. Now assume $v'_0 \in \langle v_0, \dots, v_{b-2}\rangle$ and choose an index $0 \le i \le b-2$ such that $v'_0 \notin \langle v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_{b-2}\rangle$ as well as a vector $x \in \mathbb{F}_q\backslash\{0\}$ that is not contained in $\langle v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_{b-2}, v'_0\rangle$. By Lemma 3.13 there exists a b-chain C_1 with length $[k]_q$ over $\mathbb{F}_q^k$ that has start $v_0, \dots, v_{b-2}$, end $v_0, \dots, v_{i-1}, x, v_{i+1}, \dots, v_{b-2}$, and an associated projective $b-\left([k]_q, k, [k-b]_q\right)_q$ system. Now let C_2 be a b-chain of length $[k]_q$ over $\mathbb{F}_q^k$ with start $v_0, \dots, v_{i-1}, x, v_{i+1}, \dots, v_{b-2}$, end $v'_0, v_1, \dots, v_{i-1}, x, v_{i+1}, \dots, v_{b-2}$, and an associated projective $b-\left([k]_q, k, [k-b]_q\right)_q$ system. Applying Lemma 3.12 to the chains C_1 and C_2 gives a b-chain with length $2 \cdot [k]_q$ over $\mathbb{F}_q^k$ that has start $v_0, \dots, v_{b-2}$, end $v'_0, v_1 \dots, v_{i-1}, x, v_{i+1}, \dots, v_{b-2}$, and an associated projective $b-\left(2 \cdot [k]_q, k, 2 \cdot [k-b]_q\right)_q$ system.

Suppose we have already constructed a b-chain C_1 with length $\lambda' \cdot [k]_q$ over $\mathbb{F}_q^k$ that has start $v_0, \dots, v_{b-2}$, end $v'_0, \dots, v'_i, u_{i+1}, \dots, u_{b-2}$, where $u_{i+1} \ne v'_{i+1}$, and an associated projective $b-\left(\lambda' \cdot [k]_q, k, \lambda' \cdot [k-b]_q\right)_q$ system for some positive integer λ'. If $v'_i \notin \langle v'_0, \dots, v'_i, u_{i+1}, \dots, u_{b-2}\rangle$, then Lemma 3.13 yields a b-chain C_2 with length $[k]_q$ over $\mathbb{F}_q^k$ that has start $v'_0, \dots, v'_i, u_{i+1}, \dots, u_{b-2}$, end $v'_0, \dots, v'_{i+1}, u_{i+2}, \dots, u_{b-2}$, and an associated projective $b-\left([k]_q, k, [k-b]_q\right)_q$ system. Applying Lemma 3.12 to the chains C_1 and C_2 gives a b-chain with length $(\lambda'+1) \cdot [k]_q$ and the same properties as the initial chain while the value of i is increased. Now assume $v'_{i+1} \in \langle v'_0, \dots, v'_i, u_{i+1}, \dots, u_{b-2}\rangle$ and choose an index $0 \le j \le b-2$ such that $v'_{i+1} \notin \langle v'_0, \dots, v_i, u_{i+1}, \dots, u_{j-1}, u_{j+1}, \dots, u_{b-2}\rangle$ as well as a vector $x \in \mathbb{F}_q\backslash\{0\}$ that is not contained in $\left\langle v'_0, \dots, v'_{i+1}, u_{i+1}, \dots, u_{j-1}, u_{j+1}, \dots, u_{b-2}\right\rangle$. Using Lemma 3.13 we can construct a b-chain C'_2 with length $[k]_q$ over $\mathbb{F}_q^k$ that has start $v'_0, \dots, v'_i, u_{i+1}, \dots, u_{b-2}$, end $v'_0, \dots, v'_i, u_{i+1}, \dots, u_{j-1}, x, u_{j+1}, \dots, u_{b-2}$, and an associated projective $b-\left([k]_q, k, [k-b]_q\right)_q$ system. Using Lemma 3.13 we construct a b-chain

C_3' with length $[k]_q$ over $\mathbb{F}_q^k$ that has start $v_0', \ldots, v_i', u_{i+1}, \ldots, u_{j-1}, x, u_{j+1}, \ldots, u_{b-2}$, end $v_0', \ldots, v_{i+1}', u_{i+2}, \ldots, u_{j-1}, x, u_{j+1}, \ldots, u_{b-2}$, and associated proj. $b - \left([k]_q, k, [k-b]_q\right)_q$ system. Applying Lemma 3.12 to the chains C_1 and C_2' and then applying Lemma 3.12 another time to the previous result and C_3' gives a b-chain with length $(\lambda' + 2) \cdot [k]_q$ and the same properties as the initial chain while the value of i is increased. After an iterative application of this construction we end up with $i = b - 2$ and have obtained the final desired chain. □

Example 3.15 *We continue Example 3.10. Using the introduced notation and $v_{20} + v_{16} = v_4$ we start from a 3-chain C_1 of length 31 with start v_{16}, v_1, end v_4, v_1, and associated projective $3 - (31, 5, 3)_2$ system $\mathcal{M}_1$. From Lemma 3.13 we can obtain a 3-chain C_2 with length 31 having start v_4, v_1, end v_4, v_{16}, and associated projective $3 - (31, 5, 3)_2$ system $\mathcal{M}_2$. Applying Lemma 3.12 to C_1 and C_2 gives a 3-chain C of length 62 with start v_{16}, v_1, end v_4, v_{16}, and associated projective $3 - (62, 5, 6)_2$ system $\mathcal{M}$. Note that Lemma 3.14 guarantees the existence of a 3-chain of length 31λ with start v_{16}, v_1, end v_4, v_{16}, and associated projective $3 - (31\lambda, 5, 3\lambda)_2$ system for some positive integer λ. Now let $S = \langle v_4, v_{16}, v_1 \rangle$ be a 3-space. Reordering the stated basis of S we can interprete S as a 3-chain of length 1 with start v_4, v_{16}, v_{16}, v_1,, and associated $3 - (1, 5, 1)_2$ system. Appending these two chains yields a 3-chain with length 63 over $\mathbb{F}_2^5$ with equal start and end, so that it can be interpreted as an $[63, 5]_2^3$ code.*

Now we are ready to prove that the Griesmer bound can be attained for $[n, k, d]_q^b$ codes if the minimum distance d is sufficiently large. The main idea is to start with a matching additive code that attains the Griesmer bound and to consider its corresponding projective system of b-spaces. Those b-spaces then are linked to together with suitable chains, constructed in Lemma 3.14, to a large chain having the same start and end that can then be interpreted as $[n, k, d]_q^b$ code. On the technical side we have to deal with the analysis of the minimum distance and the periodicity pattern of the Griesmer bound, see Lemma 2.11(c).

Theorem 3.16 *Given parameters k, q, and b we have*

$$n_q^b(k, d) = \left\lceil \frac{g_q\big(k, q^{b-1} \cdot d\big)}{[b]_q} \right\rceil \tag{19}$$

for all sufficiently large d.

PROOF. Consider $1 \le d' \le q^{k-b} \cdot [b]_q$ separately. From [15, Theorem 4] we conclude he existence of a constant $\lambda \in \mathbb{N}$ such that there exists a faithful projective $h-(n_\lambda, k, n_\lambda - d_\lambda)_q$

system $\mathcal{S}_\lambda$ with $d_\lambda = \lambda \cdot q^{k-b} \cdot [b]_q + d'$ and $n_\lambda = \left\lceil \frac{g_q\left(k, q^{b-1} \cdot d_\lambda\right)}{[b]_q} \right\rceil$. Now we interprete the n_λ b-spaces of $\mathcal{S}_\lambda$ as b-chains of length 1 over $\mathbb{F}_q^k$ and link them via the chains from Lemma 3.14 to a b-chain with length $n_\lambda + \lambda' \cdot [k]_q$ having the same start and end and being associated with a faithful projective $b - \left(n_\lambda + \lambda' \cdot [k]_q, k, (n_\lambda - d_\lambda) + \lambda' \cdot [k-b]_q\right)_q$ system that corresponds to a $\left(n_\lambda + \lambda' \cdot [k]_q, k, d' + (\lambda + \lambda') \cdot q^{k-b} \cdot [b]_q\right)_q^b$ code. Using Lemma 2.11 we conclude that the Griesmer bound is attained, i.e. the validity of the stated equation, for $d = d' + (\lambda + \lambda') \cdot q^{k-b} \cdot [b]_q$. From Proposition 3.2, Lemma 3.6, and Lemma 2.11 we then conclude that the Griesmer bound is attained for all $d = d' + \lambda'' \cdot q^{k-b} \cdot [b]_q$ where $\lambda'' \geq \lambda + \lambda'$. □

Instead of decomposing $\mathcal{S}_\lambda$ into chains of length 1 we can also used decompositions into larger chains, that usually exist if the cardinality of $\mathcal{S}_\lambda$ is not too small. We remark that the periodicity property of the Griesmer bound stated in Lemma 2.11 is accompanied by the upper bound

$$n \leq \frac{[k]_q \cdot s}{[k-b]_q} \tag{20}$$

for a projective $b - (n, k, s)_q$ system, that has an easy counting explanation, see e.g. [15, Lemma 15]. We remark that the Griesmer bound is always at least as good as this bound and attained with equality for the construction in Proposition 3.2.

The proof of Theorem 3.16 in general gives constructions for rather large values of the minimum distance d only. So, in order to find $[n, k, d]_q^b$ codes we can utilize ILP (integer linear programming) formulations. To this end let $\mathcal{P}$ denote the set of points and $\mathcal{H}$ denote the set of hyperplanes in $\mathrm{PG}(k-1, q)$. With this, let $\mathcal{T}(b-1)$ be the elements in $\mathcal{P}^{b-1}$ that span a $(b-1)$-space and $\mathcal{T}(b)$ be the elements in $\mathcal{P}^b$ that span a b-space.We use indicator variables $x_S^i \in \{0, 1\}$ for all $S \in \mathcal{T}(b)$ and all $0 \leq i \leq n-1$. The interpretation is that $x^i_{(P_0, \ldots, P_{b-1})} = 1$ iff the $(i+j)$th column of a generator matrix G contains a representant of P_j for all $0 \leq j \leq b-1$. In order to ensure a unique choice we require

$$\sum_{S \in \mathcal{T}(b)} x_S^i = 1 \quad \forall 0 \leq i \leq n-1. \tag{21}$$

The *chain property* is ensured via

$$\sum_{P_0 \in \mathcal{P} \,:\, (P_0, \ldots, P_{b-1}) \in \mathcal{T}(b)} x^i_{(P_0, \ldots, P_{b-1})} = \sum_{P_b \in \mathcal{P} \,:\, (P_1, \ldots, P_b) \in \mathcal{T}(b)} x^{i+1}_{(P_1, \ldots, P_b)} \quad \forall \, (P_1, \ldots, P_{b-1}) \in \mathcal{T}(b-1) \, \forall 0 \leq i \leq n-2 \tag{22}$$

and

$$\sum_{P_0\in\mathcal{P}\,:\,(P_0,\dots,P_{b-1})\in\mathcal{T}(b)} x^{n-1}_{(P_0,\dots,P_{b-1})}$$

$$= \sum_{P_b\in\mathcal{P}\,:\,(P_1,\dots,P_b)\in\mathcal{T}(b)} x^{0}_{(P_1,\dots,P_b)} \quad \forall\,(P_1,\dots,P_{b-1})\in\mathcal{T}(b-1). \tag{23}$$

In order to count the maximum number of b-spaces per hyperplane we use

$$\sum_{i=0}^{n-1} \sum_{(P_0,\dots,P_{b-1})\in\mathcal{T}(b)\,:\,\langle P_0,\dots,P_{b-1}\rangle\le H} x^{i}_{(P_0,\dots,P_{b-1})} \le s \quad \forall H\in\mathcal{H}. \tag{24}$$

With this we either minimize s and compute $d = n - s$ or we directly set $s = n - d$ and search for a feasible solution. We can add additional constraints mimicking our knowledge on the multiset of points covered by the b-spaces. I.e. we may e.g. prescribe precise point multiplicities or upper bounds. We may also assume that the generator matrix G starts with an identity matrix.[2] We may also prescribe some automorphism g and assume $x^i_{(P_0,\dots,P_{b-1})} = x^{i+1}_{\left(P_0^g,\dots,P_{b-1}^g\right)}$ for all $(P_0,\dots,P_{b-1})\in\mathcal{T}(b)$ and all $0\le i\le n-1$.

Remark 3.17 *Interpreting the sequence of columns of a generator matrix as a tour allows alternative formulations or relaxations. I.e. we may just use counting variables* $z_{(P_0,\dots,P_{b-1})} := \sum_{i=0}^{n-1} x^i_{(P_0,\dots,P_{b-1})} \in \mathbb{N}$ *instead of the* $x^i_{(P_0,\dots,P_{b-1})} \in \{0,1\}$*. Clearly we need*

$$\sum_{P_0\in\mathcal{P}\,:\,(P_0,\dots,P_{b-1})\in\mathcal{T}(b)} z_{(P_0,\dots,P_{b-1})} = \sum_{P_b\in\mathcal{P}\,:\,(P_1,\dots,P_b)\in\mathcal{T}(b)} z_{(P_1,\dots,P_b)}$$

for all $(P_1,\dots,P_{b-1})\in\mathcal{T}(b-1)$*. However, this formulation does note exclude the possibility that the desired tour is composed of several subtours. So, we can use subtour elimination constraints as done by Dantzig, Fulkerson, and Johnson for the traveling salesperson problem. In the context of the latter optimization problem our formulation is similar to the idea used by Miller, Tucker, and Zemlin.*

4 The functions $\mathbf{n_2^2(k,\cdot)}$ for $\mathbf{k\le 5}$

The aim of this section is to completely determine the function $n_2^2(k,\cdot)$ for the minimum possible length of an $[n,k,d]_2^2$ code for small dimensions k. Many of the presented methods are in principle also applicable for $[n,k,d]_q^b$ codes. However, our asymptotic result

[2]By our definition of $\mathcal{T}(b)$ and $\mathcal{T}(b-1)$ we search for faithful codes only. Of course we can replace $\mathcal{T}(b-1)$ by $\mathcal{P}^{b-1}$ and $\mathcal{T}(b)$ by $\mathcal{P}^b$ for the more general situation. Note that we can assume that G is obtained after an application of the Gauss-Jordan elimination algorithm, i.e., that G is in reduced row echelon form.

in Theorem 3.16 usually applies for rather large values of d only, so that many values of $n_q^b(k,d)$ would need to be determined to fully determine the function $n_q^b(k,\cdot)$ for other parameters.

For each $c \in \mathbb{F}_q^n$ we obviously have $0 \le \mathrm{wt}_2(c) \le n$ and $\mathrm{wt}_2(c) \in \mathbb{N}$. For small minimum distances d the Singleton-type bound $\mathrm{d}_b(C) \le n+b-k$ [7] is rather effective. More generally, the Griesmer-type bound, see Inequality (6), can often by improved for small d by applying Lemma 2.2, which is the source for many improved lower bounds for additive codes, see e.g. [15]. Even for $q=b=2$ and $k \le 5$ there exist minimum distances d where additive codes or the corresponding projective systems exist for some length n, but codes in the b-symbol metric require larger lengths, see e.g. Lemma 4.8 and Lemma 4.9. For constructive upper bounds we mostly utilize the ILP formulation from the end of Section 3.

For dimensions $k \le 2$ the determination of $n_q^b(k,\cdot)$ can still be solved completely:

Theorem 4.1 *For each integer $d \ge 2$ we have $n_q^2(1,d) = n_q^2(2,d) = d$.*

PROOF. Clearly we have $n_q^2(k,d) \ge d$ for all positive integers k and d. For dimension $k=1$ we can consider a codeword $c \in \mathbb{F}_q^d$ consisting of d ones and check $\mathrm{wt}_2(\lambda \cdot c) = d$ for all $\lambda \in \mathbb{F}_q \backslash \{0\}$, so that $n_q^2(1,d) \le d$ for all $d \ge 2$. For dimension $k=2$ we consider

$$A = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{F}_q^{2\times 2} \quad \text{and} \quad B = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \in \mathbb{F}_q^{2\times 3}.$$

We easily check $\mathrm{d}_2(\mathrm{rowspan}(A)) = 2$ and $\mathrm{d}_2(\mathrm{rowspan}(B)) = 3$, so that Corollary 3.8 implies $\mathrm{d}_2(\mathrm{rowspan}(\overbrace{A|\dots|A}^{t\text{ times}})) = 2t$ for all integers $t \ge 1$. In order to check

$$\mathrm{d}_2(\mathrm{rowspan}(B|\overbrace{A|\dots|A}^{t\text{ times}})) = \mathrm{d}_2(\mathrm{rowspan}(\overbrace{A|\dots|A}^{t\text{ times}}|B)) = 2t+3$$

we apply Lemma 3.7 and Corollary 3.8. More precisely, for some fixed integer $t \ge 1$ let $c_1 = (\overbrace{a_1|\dots|a_1}^{t\text{ times}})$ and $c_2 = (\overbrace{a_2|\dots|a_2}^{t\text{ times}})$. For each $\lambda \in \mathbb{F}_q \backslash \{0\}$ both $(\lambda \cdot b_1|\lambda \cdot c_1)$ and $(\lambda \cdot c_1|\lambda \cdot b_1)$ are of type (c).(iii) while $(\lambda \cdot b_2|\lambda \cdot c_2)$ and $(\lambda \cdot c_2|\lambda \cdot b_2)$ are of type (b).(ii). For $\lambda_1, \lambda_2 \in \mathbb{F}_q \backslash \{0\}$ we have that both $(\lambda_1 b_1 + \lambda_2 b_2|\lambda_1 c_1 + \lambda_2 c_2)$ and $(\lambda_1 c_1 + \lambda_2 c_2|\lambda_1 b_1 + \lambda_2 b_2)$ are of type (d).(iv). Thus we have constructed examples showing $n_q^2(2,d) \le d$ for all integers $d \ge 2$. □

Theorem 4.2 *For all integers $t \ge 0$ and $1 \le i \le 6$ with $6t+i \ge 2$ we have $n_2^2(3, 6t+i) = 7t+i+1$.*

PROOF. We choose

$$G_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ G_4 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix},\ G_5 = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix},$$

$$G_6 = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix},\ G_7 = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$$

and verify $\mathrm{d}_2(\mathrm{rowspan}(G_i)) = i - 1$ for all $3 \le i \le 7$ by exhaustively considering all corresponding seven non-zero codewords and computing the weights w.r.t. the symbol-pair metric. From Lemma 3.6 we conclude $n_2^2(3,6) \le n_2^2(3,3) + n_2^2(3,3) \le 8$ and $n_2^2(3,7) \le n_2^2(3,4) + n_2^2(3,3) \le 9$. For brevity, we denote corresponding generator matrices by G_8 and G_9, respectively. Using Proposition 3.2, or Corollary 3.8 applied to G_7. we conclude the existence of $[7t,3,6t]_2^2$ codes for all positive integers t. Combining these with the codes generated by $G_3, \dots, G_9$ via Lemma 3.6 yields the proposed lower bounds. For the other direction we apply the Griesmer bound in Inequality (6) and compute

i	$g_2(3, 12t + 2i)$	$\lceil g_2(3, 12t + 2i)/3 \rceil$
1	$21t + 4$	$7t + 2$
2	$21t + 7$	$7t + 3$
3	$21t + 11$	$7t + 4$
4	$21t + 14$	$7t + 5$
5	$21t + 18$	$7t + 6$
6	$21t + 21$	$7t + 7$

□

We remark that examples of $[6,3,5]_2^2$ and $[7,3,6]_2^2$ codes can also be found in [4, Table I], cf. [7, Table 1]. Up to isomorphism there exists a unique $[n,3,n-1]_2^2$ code for all $n \in \{3,\dots,7\}$ while there exist 248 non-isomorphic $[8,3,6]_2^2$ codes.

Corollary 4.3 *We have* $n_2^2(3,d) = \left\lceil \frac{g_2(3,2d)}{3} \right\rceil$ *for all* $d \ge 2$.

Theorem 4.4 *For all integers* $t \ge 0$ *and* $1 \le i \le 4$ *with* $4t+i \ge 5$ *we have* $n_2^2(4,4t+i) = 5t+i+1$. *Moreover, we have* $n_2^2(4,2) = 4$, $n_2^2(4,3) = 5$, *and* $n_2^2(4,4) = 6$.

PROOF. Examples of $[4,4,2]_2^2$, $[5,4,3]_2^2$, $[6,4,4]_2^2$, $[7,4,5]_2^2$, $[8,4,6]_2^2$, $[9,4,7]_2^2$, $[10,4,8]_2^2$, $[12,4,9]_2^2$, $[13,4,10]_2^2$, $[14,4,11]_2^2$, and $[15,4,12]_2^2$ codes are given by the generator matrices

$$\begin{pmatrix} 1000 \\ 0100 \\ 0010 \\ 0001 \end{pmatrix}, \begin{pmatrix} 10001 \\ 01001 \\ 00101 \\ 00011 \end{pmatrix}, \begin{pmatrix} 100010 \\ 010001 \\ 001010 \\ 000101 \end{pmatrix}, \begin{pmatrix} 1000110 \\ 0100011 \\ 0010111 \\ 0001101 \end{pmatrix}, \begin{pmatrix} 10001110 \\ 01001011 \\ 00100111 \\ 00011101 \end{pmatrix}, \begin{pmatrix} 100011111 \\ 010001011 \\ 001011010 \\ 000101101 \end{pmatrix}, \begin{pmatrix} 1000111111 \\ 0100101010 \\ 0010011011 \\ 0001101101 \end{pmatrix},$$

$$\begin{pmatrix} 101001111111 \\ 011001010101 \\ 000101101001 \\ 000010110111 \end{pmatrix}, \begin{pmatrix} 1010011111111 \\ 0110001101010 \\ 0001010110011 \\ 0000111101101 \end{pmatrix}, \begin{pmatrix} 10100111111000 \\ 01100101001111 \\ 00010111010110 \\ 00001011101101 \end{pmatrix}, \begin{pmatrix} 101000110011110 \\ 011000011110101 \\ 000101011101110 \\ 000011101011101 \end{pmatrix}.$$

By recursively applying Lemma 3.6 we conclude $n_2^2(4, i+8t) \le i+2+10t$ for $5 \le i \le 8$ as well as $n_2^2(4, i+8t) \le i+3+10t$ for $9 \le i \le 12$, where $t \in \mathbb{N}$ is arbitrary. Thus we have $n_2^2(4, 4t+i) \le 5t+i+1$ for all integers $t \ge 0$ and $1 \le i \le 4$ with $4t+i \ge 5$.

For the other direction we apply the Griesmer bound from Inequality (6) and compute

i	$g_2(4, 8t+2i)$	$\lceil g_2(8, 8t+2i)/3 \rceil$
1	$15t+5$	$5t+2$
2	$15t+8$	$5t+3$
3	$15t+12$	$5t+4$
4	$15t+15$	$5t+5$

The improved lower bounds $n_2^2(4,2) \ge 4$, $n_2^2(4,3) \le 5$, and $n_2^2(4,4) \le 6$ are given by the Singleton-type bound $d \le n+b-k$ [7]. □

Corollary 4.5 *We have* $n_2^2(4,d) = \left\lceil \frac{g_2(4,2d)}{3} \right\rceil$ *for all* $d \ge 5$.

d	2	3	4	5	6	7	8	9	10	11	12
$n_2^2(4,d)$	4	5	6	7	8	9	10	12	13	14	15
#	1	1	4	1	4	1	1	21030	13772	755	33

Table 1: Number of non-isomorphic optimal $[n_2^2(4,d), 4, d]_2^2$ codes

Theorem 4.6 *We have* $n_2^2(5,d) = \left\lceil \frac{g_2(5,2d)}{3} \right\rceil$ *for all* $d \ge 9$*. Moreover, we have* $n_2^2(5,d) = d+3$ *for* $d \in \{2,3,4,6,7\}$*,* $n_2^2(5,5) = 9$*, and* $n_2^2(5,8) = 12$.

PROOF. Examples of $[5,5,2]_2^2$, $[6,5,3]_2^2$, $[7,5,4]_2^2$, $[9,5,5]_2^2$, $[9,5,6]_2^2$, $[10,5,7]_2^2$, $[12,5,8]_2^2$, $[13,5,9]_2^2$, $[14,5,10]_2^2$, $[15,5,11]_2^2$, $[16,5,12]_2^2$, $[18,5,13]_2^2$, $[19,5,14]_2^2$, $[20,5,15]_2^2$, $[21,5,16]_2^2$, $[23,5,17]_2^2$, $[24,5,18]_2^2$, $[25,5,19]_2^2$, $[26,5,20]_2^2$, $[28,5,21]_2^2$, $[29,5,22]_2^2$, $[30,5,23]_2^2$, $[31,5,24]_2^2$, $[33,5,25]_2^2$, $[34,5,26]_2^2$, $[36,5,27]_2^2$, $[37,5,28]_2^2$, $[38,5,29]_2^2$, $[39,5,30]_2^2$, $[41,5,31]_2^2$, and $[42,5,32]_2^2$ codes are given by the generator matrices

$$\begin{pmatrix} 10000 \\ 01000 \\ 00100 \\ 00010 \\ 00001 \end{pmatrix}, \begin{pmatrix} 100001 \\ 010001 \\ 001001 \\ 000101 \\ 000011 \end{pmatrix}, \begin{pmatrix} 1000010 \\ 0100011 \\ 0010001 \\ 0001011 \\ 0000101 \end{pmatrix}, \begin{pmatrix} 110000111 \\ 001000110 \\ 000100111 \\ 000010101 \\ 000001011 \end{pmatrix}, \begin{pmatrix} 100001111 \\ 010001001 \\ 001001010 \\ 000101101 \\ 000011011 \end{pmatrix}, \begin{pmatrix} 1000011110 \\ 0100001011 \\ 0010011001 \\ 0001010111 \\ 0000101101 \end{pmatrix}, \begin{pmatrix} 110000111110 \\ 001000011101 \\ 000100101010 \\ 000010111001 \\ 000001101111 \end{pmatrix},$$

$$
\begin{pmatrix} 1001001111111 \\ 0101001010101 \\ 0011000110011 \\ 0000101101101 \\ 0000011011011 \end{pmatrix},
\begin{pmatrix} 10010011111111 \\ 01010010100001 \\ 00110001001011 \\ 00001011011001 \\ 00000110101111 \end{pmatrix},
\begin{pmatrix} 111000011011110 \\ 111011100001011 \\ 111100110110001 \\ 110001101110010 \\ 011001000110111 \end{pmatrix},
\begin{pmatrix} 0101010101101101 \\ 0000010110111011 \\ 0000101001010101 \\ 0011001100011111 \\ 1001001000101010 \end{pmatrix},
\begin{pmatrix} 00010110101010110101 \\ 00100000110110111111 \\ 01000010010110 1101 \\ 00001000110111 1011 \\ 10000101011111 0001 \end{pmatrix},
$$

$$
\begin{pmatrix} 000011011101101 1011 \\ 001101000110110 1101 \\ 000111101000010 1111 \\ 110000011110011 1101 \\ 010010111001010 0001 \end{pmatrix},
\begin{pmatrix} 000001101110110 10101 \\ 000110110111011 00110 \\ 011000010011010 11011 \\ 001010101000100 11010 \\ 100000100101111 01011 \end{pmatrix},
\begin{pmatrix} 0000011010111011 01101 \\ 0001100011011101 10110 \\ 0010101100101110 00101 \\ 1100000101010111 00111 \\ 0100101011010000 11001 \end{pmatrix},
\begin{pmatrix} 0000001110111011 0111011 \\ 0000110011011101 1001101 \\ 0011010111101100 0011010 \\ 0101000001111001 1101010 \\ 1001000110110110 0001101 \end{pmatrix},
$$

$$
\begin{pmatrix} 10100010100010110111 1011 \\ 01100001101101100010 0111 \\ 00010010010101011100 1011 \\ 00001011101010011011 0001 \\ 00000111010110101011 100110 \end{pmatrix},
\begin{pmatrix} 01111101101010000110 11000 \\ 01011110010011101001 10100 \\ 00111110101101111100 00010 \\ 01001111111101000011 10001 \\ 11101010100110110011 00000 \end{pmatrix},
\begin{pmatrix} 01101111111010100110 001000 \\ 01001101000101110010 110100 \\ 01011110101100000101 010010 \\ 00111100111000111011 110001 \\ 11111001001111011110 010000 \end{pmatrix},
$$

$$
\begin{pmatrix} 0111101101101010100100001000 \\ 0101111011000011100011110100 \\ 0000011101010010111110110010 \\ 0110101000100101111111100001 \\ 1000100111111100101101110000 \end{pmatrix},
\begin{pmatrix} 01111100001110111101010011000 \\ 01010011010110010101111100100 \\ 01001001111111001011101010010 \\ 00010100111011101001110100001 \\ 10111010110100101111001000000 \end{pmatrix},
\begin{pmatrix} 011100100110010110110010111000 \\ 010001110111101011010100100100 \\ 010111100010011101001101010010 \\ 001111001011110010110101000001 \\ 111101011100101010001011010000 \end{pmatrix},
$$

$$
\begin{pmatrix} 0110010011111011100010101101000 \\ 0011001001111101110001010110100 \\ 0111110111000101011010000110010 \\ 0101101000011001001111101110001 \\ 1100100111110111000101011010000 \end{pmatrix},
\begin{pmatrix} 100001110010110101100011001111001 \\ 010001011110101100010010100010011 \\ 001001010111111000101100110101100 \\ 000100110111011101111100010010001 \\ 000010100011001011010001010111111 \end{pmatrix},
$$

$$
\begin{pmatrix} 10000100101011001111011100101011 10 \\ 01000011110101010011001010110001 10 \\ 00100000011100001010110111010111 011 \\ 00010110000101111110010101101110 010 \\ 00001001001010101101111010010000 101 \end{pmatrix},
\begin{pmatrix} 100001100000110101010101101101110001 \\ 010000111011101101111001000101101000 \\ 001001110001101110001111101001001011 \\ 000100100100011110111100011100101111 \\ 000011011111001111110001110001010 1101 \end{pmatrix},
$$

$$
\begin{pmatrix} 100001001111101101010101010010011110000 \\ 0100011001001010000001110101101011011 \\ 00100101101111001010000010111001010010 \\ 0001011100100111001100011101100010101 \\ 000011100111010111101010010001010100 \end{pmatrix},
\begin{pmatrix} 10000100110111101111001100111000011011 \\ 01000100111010010000111011111101101110 \\ 001001010111111110101100001011111111001 \\ 00010111001010011111010010001111010011 \\ 00001010101110110100110101110010000011 \end{pmatrix},
$$

$$
\begin{pmatrix} 100000100100100010101111011011110010110 \\ 010001010100010111101101101110001100100 \\ 001001111111000000100100101111110111101 \\ 000100011001111101101111001101010100001 \\ 000010110011111110010001111011101101101 \end{pmatrix},
\begin{pmatrix} 10000100100011111000010101011110101100101 \\ 01000111101110010100000101001001011011110 \\ 00100010011001010011001001010010111101010 \\ 00010100001100011011101010010100101011111 \\ 00001001001000001110110101101010111110111 \end{pmatrix},
$$

$$
\begin{pmatrix} 1000000001011101101010110001001011110101 10 \\ 0100010011100100100001110101110111101011 00 \\ 0010010001010110111010011000011110101000 11 \\ 0001011010100111000100101010101010101100 00 \\ 0000101101010100101100001101100100111110 01 \end{pmatrix}.
$$

(A $[41,5,31]_2^2$ code can also be constructed from a $[31,5,24]_2^2$ and a $[10,5,7]_2^2$ code. For a general construction of a $[31,5,24]_2^2$ code we refer to Proposition 3.2.) Combining these codes with a suitable number of $[31,5,24]_2^2$ codes via Lemma 3.6 gives the proposed upper bounds for $n_2^2(5,d)$ for all $d \ge 33$.

The proposed lower bounds for $n_2^2(5,d)$ are given by the Griesmer bound from Inequality (6) for all $d \ge 9$ and by the Singleton-type bound $d \le n+b-k$ [7] for all $d \in \{2,3,4,6,7\}$. Using `LinCode` we have enumerated all 10358 even $[24,5,10]_2$ codes. While several of the corresponding multisets of points can be partitioned into lines, no partition yields a $[8,5,5]_2^2$ code, so that $n_2^2(5,5) \ge 9$. Alternatively, we can directly use Lemma 4.8. Using `LinCode` we have enumerated all 21 even $[33,5,16]_2$ codes. While several of the corresponding multisets of points can be partitioned into lines, no partition yields a $[11,5,8]_2^2$ code, so that $n_2^2(5,8) \ge 12$. Alternatively, we can directly use Lemma 4.9. $\square$

We remark that in [4, Table I] generator matrices for $[7,4,5]_2^2$, $[8,4,6]_2^2$, $[9,4,7]_2^2$, and $[10,4,8]_2^2$ codes are stated. The length of the list of explicitly constructed generator matrices is partially due to the fact that the existence of an $[n,k,d]_q^b$ code does not necessarily imply the existence of an $[n-1,k,d-1]_q^b$ code, see Example 3.1.

Our next aim is to replace the computer enumerations for $d \in \{5,8\}$ in the proof of Theorem 4.6 by theoretical arguments.

Lemma 4.7 *Given an $[n,k,d]_q^2$ code C with $n = n_q^2(k,d)$, there exists a spanning multiset of points in $\mathrm{PG}(k-1,q)$ of cardinality $(q+1)n$ that can be written as $\mathcal{M} = \sum_{i=0}^{n-1} \chi_{L_i}$ for n lines $L_0, \dots, L_{n-1}$ such that each hyperplane H contains at most $(n-d)$ of the lines and $\sum_{P \le H} \mathcal{M}(P) \le q(q+1) - dq$. If $q = 2$ there additionally exist n points $P_0, \dots P_{n-1}$ such that $\mathcal{M} = \sum_0^{n-1} (2 \cdot \chi_{P_i} + \chi_{Q_i})$, where $Q_i = P_i + P_{i+1}$ (reading indices modulo n).*

PROOF. Due to Lemma 3.5 we can assume that C is faithful. Given a generator matrix G of C we denote the spans of the columns of G by $P_0, \dots, P_{n-1}$ and set $L_i := \langle P_i, P_{i+1} \rangle$, where the indices are read modulo n. Setting $\mathcal{M} := \sum_{i=0}^{n-1} \chi_{L_i}$ gives the multiset of points associated to C. From Lemma 2.6 we conclude that at most $n-d$ lines can be fully contained in a hyperplane. Since each hyperplane either fully contains a line or intersects it in a point we conclude the upper bound $\sum_{P \le H} \mathcal{M}(P) \le q(q+1) - dq$ for all hyperplanes H. For $q = 2$ we observe $L_i = \{P_i, Q_i, P_{i+1}\}$ for all $0 \le i \le n-1$. □

Lemma 4.8 *No $[8,5,5]_2^2$ code exists.*

PROOF. Assuming that such a code exist we use Lemma 4.7 to construct a multiset of points $\mathcal{M} = \sum_{i=0}^{7} (2\chi_{P_i} + \chi_{Q_i})$, where $Q_i = P_i + P_{i+1}$ with indices read modulo 8, that corresponds to a $[24,5,10]_2$ code using Lemma 4.7. Since the points $P_0, \dots, P_7$ span $\mathrm{PG}(4,2)$ and each $[8,5,d]_2$ code satisfies $d \le 2$ we conclude the existence of a hyperplane H that contains at least six of the points $P_0, \dots, P_7$. Thus, H contains at least four of the points $Q_0, \dots, Q_7$, so that the multiplicity of H is at least $2 \cdot 6 + 4 = 16$. However, the maximum possible multiplicity of a hyperplane in a multiset of points corresponding to a $[24,5,10]_2$ code is $24 - 10 = 14$ – contradiction. □

In order to show the non-existence of a $[11,5,8]_2^2$ code we need to refine the argument a bit. (We may deduce the existence of a hyperplane H that contains a least seven of the eleven double points, say $\{P_0, P_1, P_3, P_4, P_6, P_7, P_9\}$ while the points in $\{P_2, P_5, P_8, P_{10}\}$ are not contained. In this example only three of the eleven points Q_i are contained in H, which gives a multiplicity of $2 \cdot 7 + 3 = 17$ for H, which goes in line with a $[33,5,16]_2$ code.)

Lemma 4.9 *No $[11,5,8]_2^2$ code exists.*

PROOF. Assuming that such a code exist we construct a multiset of points $\mathcal{M} = \sum_{i=0}^{10} \chi_{L_i}$ as in Lemma 4.7 corresponding to a $[33, 5, 16]_2$ code, where the L_i are lines. By construction we have $\mathcal{M}(L_i) \geq 2 \cdot 2 + 1 = 5$. Since no $[27, 3, 16]_2$ code exists we have indeed $\mathcal{M}(L_i) = 5$ for all $0 \leq i \leq 10$. Projection through L_i yields a multiset of points $\mathcal{M}'$ that corresponds to a $[28, 3, 16]_2$ code. Since we have $\mathcal{M}'(L') \leq 12$ for every line L' in $\mathrm{PG}(4, 2)/L_i \cong \mathrm{PG}(2, 2)$ we conclude $\mathcal{M}'(P') \leq 4$ for every point P' in the factor space. Since there are only seven points and $\mathcal{M}'$ has cardinality 28 we have $\mathcal{M}'(P') = 4$ for all points. Thus, we have $\mathcal{M}(H) = 5 + 3 \cdot 4 = 17$ for every hyperplane $H \geq L_i$. Counting gives that those hyperplanes contain exactly three lines. Since we can choose $0 \leq i \leq 10$ arbitrarily we can state that each hyperplane that contains at least one of the eleven lines contains exactly three of them. Denoting the number of hyperplanes that contain at least one line by x and double counting lines gives

$$3 \cdot x + 0 \cdot (31 - x) = 7 \cdot 11,$$

which does not have an integral solution – contradiction. □

5 Conclusion

In Theorem 3.16 we have shown that the minimum possible length $n_q^b(k, d)$ of an $[n, k, d]_q^b$ code is attained by a Griesmer-type bound $n_q^b(k, d) \geq \left\lceil \frac{g_q\left(k, q^{b-1} \cdot d\right)}{[b]_q} \right\rceil$ if d is sufficiently large. With this the problem of the determination of $n_q^b(k, \cdot)$ becomes a finite problem. In Section 4 we have solved this problem for $q = b = 2$ and $k \leq 5$. Besides the general construction in Proposition 3.2 and Lemma 3.6 for the combination of codes, we only used explicit codes found by ILP searches. In order to determine the functions $n_q^b(k, \cdot)$ for more parameters, more general constructions are desired. Although we have described $[n, k, d]_q^b$ codes from the geometric point of view as projective $b - (n, k, n - d)_q$ systems we are still very far from a one-to-one correspondence.

References

[1] S. Ball, M. Lavrauw, and T. Popatia. Griesmer type bounds for additive codes over finite fields, integral and fractional MDS codes. *Designs, Codes and Cryptography*, 93(1):175–196, 2025.

[2] I. Bouyukhev, D. B. Jaffe, and V. Vavrek. The smallest length of eight-dimensional binary linear codes with prescribed minimum distance. *IEEE Transactions on Information Theory*, 46(4):1539–1544, 2000.

[3] Y. Cassuto and M. Blaum. Codes for symbol-pair read channels. *IEEE Transactions on Information Theory*, 57(12):8011–8020, 2011.

[4] Y. M. Chee, L. Ji, H. M. Kiah, C. Wang, and J. Yin. Maximum distance separable codes for symbol-pair read channels. *IEEE Transactions on Information Theory*, 59(11):7259–7267, 2013.

[5] B. Chen and H. Liu. New bounds on the code size of symbol-pair codes. *IEEE Transactions on Information Theory*, 69(2):941–950, 2023.

[6] B. Ding, G. Ge, J. Zhang, T. Zhang, and Y. Zhang. New constructions of MDS symbol-pair codes. *Designs, Codes and Cryptography*, 86:841–859, 2018.

[7] B. Ding, T. Zhang, and G. Ge. Maximum distance separable codes for b-symbol read channels. *Finite Fields and Their Applications*, 49:180–197, 2018.

[8] S. Dodunekov and J. Simonis. Codes and projective multisets. *The Electronic Journal of Combinatorics*, 5:1–23, 1998.

[9] P. Govaerts. *Classifications of blocking set related structures in Galois geometries*. PhD thesis, Ghent University, 2003.

[10] J. H. Griesmer. A bound for error-correcting codes. *IBM Journal of Research and Development*, 4(5):532–542, 1960.

[11] M. D. Hestenes. Singer groups. *Canadian Journal of Mathematics*, 22(3):492–513, 1970.

[12] D. Huang, Q. Liao, G. Tang, and S. Zhu. On the b-symbol weights of linear codes for large b. *Finite Fields and Their Applications*, 107:102647, 2025.

[13] X. Kai, S. Zhu, and P. Li. A construction of new MDS symbol-pair codes. *IEEE Transactions on Information Theory*, 61(11):5828–5834, 2015.

[14] X. Kai, S. Zhu, Y. Zhao, H. Luo, and Z. Chen. New MDS symbol-pair codes from repeated-root codes. *IEEE Communications Letters*, 22(3):462–465, 2018.

[15] S. Kurz. Additive codes attaining the Griesmer bound. *arXiv preprint 2412.14615*, 2024.

[16] G. Luo, M. F. Ezerman, C. Güneri, S. Ling, and F. Özbudak. Griesmer bound and constructions of linear codes in b-symbol metric. *IEEE Transactions on Information Theory*, 70(11):7840–7847, 2024.

[17] G. Luo, M. F. Ezerman, S. Ling, and X. Pan. New families of MDS symbol-pair codes from matrix-product codes. *IEEE Transactions on Information Theory*, 69(3):1567–1587, 2022.

[18] J. Ma and J. Luo. On symbol-pair weight distribution of MDS codes and simplex codes over finite fields. *Cryptography and Communications*, 13:101–115, 2021.

[19] J. Ma and J. Luo. Constructions of MDS symbol-pair codes with minimum distance seven or eight. *Designs, Codes and Cryptography*, 90(10):2337–2359, 2022.

[20] J. Ma and J. Luo. MDS symbol-pair codes from repeated-root cyclic codes. *Designs, Codes and Cryptography*, pages 1–17, 2022.

[21] G. Solomon and J. J. Stiffler. Algebraically punctured cyclic codes. *Information and Control*, 8(2):170–179, 1965.

[22] E. Yaakobi, J. Bruck, and P. H. Siegel. Constructions and decoding of cyclic codes over b-symbol read channels. *IEEE Transactions on Information Theory*, 62(4):1541–1551, 2016.