Sparse Parity-Check Matrices over GF(q)

Dedicated to the 60th Birthday of Walter Deuber

Hanno Lefmann
Fakultät für Informatik, TU Chemnitz
D-09107 Chemnitz, Germany
lefmann@informatik.tu-chemnitz.de

Abstract

For fixed positive integers k, q, r with q a prime power and large m, we investigate matrices with m rows and a maximum number $N_q(m, k, r)$ of columns, such that each column contains at most r nonzero entries from the finite field $GF(q)$ and each k columns are linearly independent over $GF(q)$. For even integers $k \geq 2$ we obtain the lower bounds $N_q(m, k, r) = \Omega(m^{kr/(2(k-1))})$, and $N_q(m, k, r) = \Omega(m^{((k-1)r)/(2(k-2))})$ for odd $k \geq 3$. For $k = 2^i$ we show that $N_q(m, k, r) = \Theta(m^{kr/(2(k-1))})$ if $\gcd(k-1, r) = k-1$, while for arbitrary even $k \geq 4$ with $\gcd(k-1, r) = 1$ we have $N_q(m, k, r) = \Omega(m^{kr/(2(k-1))} \cdot (\log m)^{1/(k-1)})$. Matrices, which fulfill these lower bounds, can be found in polynomial time. Moreover, for char $(GF(q)) > 2$ we obtain $N_q(m, 4, r) = \Theta(m^{\lceil 4r/3 \rceil/2})$, while for char $(GF(q)) = 2$ we can only show that $N_q(m, 4, r) = O(m^{\lceil 4r/3 \rceil/2})$. Our results extend and complement earlier results from [5, 18], where the case $q = 2$ was considered.

1 Introduction

For a prime power q, let $GF(q)$ be the finite field with q elements. We consider matrices over $GF(q)$ with *k-wise independent columns,* i.e. each k columns are linearly independent over $GF(q)$. Moreover, each column contains at most r nonzero entries from $GF(q) \setminus \{0\}$. For such matrices we use the notion of *(k, r)-matrices.* Given a number m of rows, let $N_q(m, k, r)$ denote the maximum number of columns such a matrix can have. Recall that matrices with k-wise independent columns are just parity-check matrices for linear codes with minimum distance at least $k + 1$, hence we investigate here the sizes of sparse parity-check matrices over $GF(q)$.

By monotonicity, we have $N_q(m, k + 1, r) \leq N_q(m, k, r)$ for $k = 2, 3, \ldots$. Throughout this paper, k, r, q are fixed positive integers and m is large.

For $q = 2$, i.e. we are working in $GF(2) = \{0, 1\}$, it has been shown by a probabilistic argument that $N_2(m, 2k + 1, r) \geq 1/2 \cdot N_2(m, 2k, r)$, see [18], hence it suffices in this case to consider even independences. Moreover, for $q = 2$ and $r = 2$ the values of $N_2(m, k, 2)$ are asymptotically equal (up to an additive term of $O(m)$ for the number of columns with exactly one entry 1) to the maximum number of edges in a graph on m vertices, which does not contain any cycle of length at most k. The growth of $N_2(m, k, 2)$ has been studied a lot in the past, however not that much is known on the exact asymptotic growth rate for arbitrary fixed integers $k \geq 2$. Known are only the values $N_2(m, 4, 2) = \Theta(m^{3/2})$,

see [9, 11, 12], and $N_2(m,6,2) = \Theta(m^{4/3})$ and $N_2(m,10,2) = \Theta(m^{6/5})$, see [4, 26]. In general, for fixed integers $k \geq 1$ a simple probabilistic argument yields $N_2(m,2k,2) = \Omega(m^{1+1/(2k-1)})$. By constructions of Margulis [22], and Phillips, Lubotzky and Sarnak [21] this lower bound was improved to $N_2(m,2k,2) = \Omega(n^{1+2/(3k+3)})$, which was further improved by Lazebnik, Ustimenko and Woldar [17] to $N_2(m,2k,2) = \Omega(m^{1+2/(3k-3+\varepsilon)})$ with $\varepsilon \in \{0,1\}$ and $\varepsilon = 0$ if and only if k is odd. However, concerning upper bounds we only know that $N_2(m,2k,2) = O(m^{1+1/k})$ for fixed integers $k \geq 1$ by the work of Bondy and Simonovits [8].
For $q = 2$ and arbitrary fixed integers $r \geq 1$, the following lower and upper bounds on $N_2(m,k,r)$ were given by Pudlák, Savický and this author [18].

Theorem 1.1 *Let $k \geq 2$ even and $r \geq 1$ be fixed integers. Then for positive integers m,*

$$N_2(m,k,r) = \Omega\left(m^{\frac{kr}{2(k-1)}}\right) \tag{1}$$

and for $k = 2^i$,

$$N_2(m,k,r) = O\left(m^{\lceil k \cdot r/(k-1) \rceil/2}\right) . \tag{2}$$

Thus, for $\gcd(k-1,r) = k-1$ and k a power of 2, the lower bound (1) and the upper bound (2) match. However, for k even and $\gcd(k-1,r) = 1$, the lower bound (1) was improved by Bertram-Kretzberg, Hofmeister and this author [5] to

$$N_2(m,k,r) = \Omega\left(m^{\frac{kr}{2(k-1)}} \cdot (\log m)^{\frac{1}{k-1}}\right) .$$

Here we generalize and extend some of these earlier results on the growth of $N_2(m,k,r)$ to the case of arbitrary finite fields $GF(q)$: we infer the lower bounds $N_q(m,k,r) = \Omega(m^{kr/(2(k-1))})$ for even integers $k \geq 2$, and $N_q(m,k,r) = \Omega(m^{(k-1)r/(2(k-2))})$ for odd integers $k \geq 3$. For $k = 2^i$ we show that $N_q(m,k,r) = \Theta(m^{kr/(2(k-1))})$ for $\gcd(k-1,r) = k-1$, while for every even integer $k \geq 4$ with $\gcd(k-1,r) = 1$ we have $N_q(m,k,r) = \Omega(m^{kr/(2(k-1))} \cdot (\log m)^{1/(k-1)})$. Also, for $k = 4$ and char $(GF(q)) > 2$ we prove that $N_q(m,4,r) = \Theta(m^{\lceil 4r/3 \rceil/2})$, while so far for $q = 2^l$ we can only show that $N_q(m,4,r) = O(m^{\lceil 4r/3 \rceil/2})$. The corresponding matrices can be found deterministically in polynomial time. Possible applications for such sparse matrices are that quite often algorithms run fast on such matrices. In Section 5 we discuss some applications.
Related here, but different, are the results from Sipser and Spielman, see [24, 25], where in connection with the PCP-theorem low-density $0,1$-matrices have been investigated, which yield linear-time encodable error-correcting codes, see also [19, 20, 23]. These low-density matrices contain in each row and in each column only a constant number of nonzero entries. Here, however, we do not restrict the number of nonzero entries in each row.

2 Preliminaries

From now on we will assume that in every matrix M under consideration all columns are pairwise distinct, in each column the first nonzero entry is equal to 1 and M does not contain the all zeros column. This is no restriction, since $k \geq 2$ and we only care about independencies among the columns. Obviously, we have $N_q(m,k,1) = m$ for $k \geq 2$ and $N_q(m,2,r) = \sum_{i=1}^{r} \binom{m}{i} \cdot (q-1)^{i-1} = \Theta(m^r)$, where the last can be seen by taking all column vectors of length m with at most r nonzero vectors, where the first nonzero entry is 1, and M does not contain the all zeros column. The following lemma will be crucial in our further arguments.

Lemma 2.1 *Let $r \geq 1$ be an integer. Let M be an $m \times n$-matrix over $GF(q)$ with at most r nonzero entries in each column and with pairwise distinct columns, where M does not contain the all zeros column.*
Then the matrix M contains an $m \times n'$-submatrix M' with the following properties:

(*i*) *$n' \geq n \cdot r!/(r^r \cdot q^r)$, and*

(*ii*) *there is a partition $\{1,\ldots,m\} = R_1 \cup \ldots \cup R_r$ of the set of row-indices of M' and a sequence $(e_1, e_2, \ldots, e_r)$ of elements from $GF(q)$ such that each column of M' contains at most one nonzero entry e_j within the rows in R_j, $j = 1,\ldots,r$, ($e_j = 0$ means that in each column every entry within the rows of R_j is equal to zero, and $e_j \neq 0$ means that there is exactly one entry e_j within the rows of R_j and the other entries within R_j are zero), and*

(*iii*) *the columns of M' are 3-wise independent.*

Proof. Uniformly and independently of the others assign at random $1,\ldots,r$ to the row-indices $1,\ldots,m$ of the matrix M. Let R_j, $j = 1,\ldots,r$, be the random set of row-indices with assignment j. The probability $Prob$, that a fixed column c in M with $i \leq r$ nonzero entries contains in every row-set R_j at most one nonzero entry, can be bounded from below as follows

$$Prob \;=\; \frac{[r]_i}{r^i} \geq \frac{r!}{r^r} \,.$$

Thus for such a random partition $\{1,\ldots,m\} = R_1 \cup \ldots \cup R_r$ the expected number of columns in M with at most one nonzero entry in each row-set R_j, $j = 1,\ldots,r$, is at least $n \cdot r!/r^r$. Take such a subset of columns of M with corresponding partition $\{1,\ldots,m\} = R_1 \cup \ldots \cup R_r$ and call the resulting matrix M^*. For each column in the matrix M^* record for $j = 1,\ldots,r$ as a sequence of length r, the possibly occurring nonzero entries e_j, and set $e_j = 0$ if all entries within R_j are zero. Since there are at most $(q^r - 1) < q^r$ such sequences there are at least $n' \geq n \cdot r!/(r^r \cdot q^r)$ columns in M^* with the same pattern $(e_1,\ldots,e_r)$. Take these columns and call the resulting matrix M', thus (i) and (ii) are fulfilled.
Assume that three columns a_1, a_2, a_3 of the matrix M' are linearly dependent over $GF(q)$. If $e_j \neq 0$ for some $j = 1,\ldots,r$, then within the rows in R_j each column a_i contains exactly one entry e_j. Since the columns in M and hence in M' are pairwise distinct and since a_1, a_2, a_3 are linearly dependent, each entry $e_j \neq 0$, $j = 1,\ldots,r$, is contained in the same row of a_1, a_2, a_3. But then $a_1 = a_2 = a_3$, contradicting our assumption, hence the matrix M' satisfies (iii). □

Lemma 2.1 can be made constructive in polynomial time if one applies one of the known derandomization techniques for the MAXCUT-problem, compare for example [15].

As mentioned in the introduction, we have $N_2(m, 2k+1, r) \geq 1/2 \cdot N_2(m, 2k, r)$. While for $q = 2$ it was easy to reduce asymptotically the case of odd dependencies to the case of even dependencies, for arbitrary prime powers $q > 2$ this does not seem to be the case anymore.

Corollary 2.2 *Let $r \geq 1$ and a prime power q be fixed integers. Then, for positive integers m,*

$$N_q(m, 3, r) = \Theta(m^r) \; .$$

Proof. The upper bound $N_q(m, 3, r) \leq N_q(m, 2, r) = \Theta(m^r)$ follows by monotonicity. For the lower bound, partition the set $\{1, \ldots, m\}$ of row-indices into subsets $R_1, \ldots, R_r$ of nearly equal size $\lfloor m/r \rfloor$ or $\lceil m/r \rceil$. Fix any sequence $(e_1, e_2, \ldots, e_r) \in (GF(q) \setminus \{0\})^r$ of nonzero entries. Define an $m \times n$-matrix M over $GF(q)$ without repeated columns by taking all possible columns of length m with exactly one entry e_j within the row-set R_j for $j = 1, \ldots, r$. Then $n \geq (\lfloor m/r \rfloor)^r$ and the columns are 3-wise independent by the proof of Lemma 2.1 (iii). $\square$

Corollary 2.3 *Let q be a fixed prime power. Then there exists a constant $c > 0$ such that for positive integers m,*

$$N_q(m, 5, 2) \geq c \cdot N_q(m, 4, 2) \; .$$

Proof. Let M be an $m \times n$-matrix, $n = N_q(m, 4, 2)$, with entries from $GF(q)$, where each column contains at most two nonzero entries and the columns are 4-wise independent. By Lemma 2.1, the matrix M contains an $m \times n'$-submatrix M' satisfying assertions (i), (ii) there, hence $n' \geq c \cdot n$ for some constant $c > 0$. Assume that some columns $a_1, \ldots, a_5$ from M' are linearly dependent over $GF(q)$. Consider the occurrence of the first nonzero entry e_1 in the columns $a_1, \ldots, a_5$. Since the columns $a_1, \ldots, a_5$ are linearly dependent, either all five entries e_1 must occur in the same row, or three entries e_1 occur in the same row and the two others in some other row. The same holds for the possibly next occurring nonzero entry e_2. In any case, whether $e_2 = 0$ or $e_2 \neq 0$, at least two of the columns $a_1, \ldots, a_5$ are identical, a contradiction, hence $N_q(m, 5, 2) \geq c \cdot N_q(m, 4, 2)$. $\square$

A more general result than stated in Corollary 2.3 can be found in Corollary 4.4.

3 Upper Bounds

In this section we will show some general upper bounds on the growth rate of $N_q(m, k, r)$.

Theorem 3.1 *Let $k \geq 4$ with k even, $r \geq 1$ and q a prime power be fixed integers. Then, for some positive constant $c \leq q^r \cdot r^r / r!$ and for $s = 0, \ldots, r-1$ the following holds*

$$N_q(m, k, r) \leq 2c \cdot N_q(m, k/2, 2r-2s) + c \cdot \sum_{i=1}^{s} \binom{m}{i} \tag{3}$$

and

$$\begin{aligned} N_q(m, k, r) \quad \leq \quad & c \cdot \sqrt{2 \cdot \binom{m}{s} \cdot \binom{r-1}{s} \cdot N_q(m, k/2, 2r-2s)} + \\ & + c \cdot \left(\binom{m}{s} + \sum_{i=1}^{s} \binom{m}{i} \right) \; , \end{aligned} \tag{4}$$

thus $N_q(m,k,r) = O(m^{s/2} \cdot N_q(m,k/2,2r-2s)^{1/2} + m^s)$ *for fixed* k, r, q.

The proof is similar, but different, to that by Pudlák, Savický and this author [18], where analogous results for the case $q = 2$ were proved.

Proof. Let M be an $m \times n$-matrix, $n = N_q(m,k,r)$, where each column of M contains at most r nonzero entries from $GF(q)$ and the columns are k-wise independent. By Lemma 2.1, the matrix M contains an $m \times n'$-submatrix M' with $n' \geq c^* \cdot n$ and $c^* = r!/(r^r \cdot q^r)$ and M' satisfies assertion (ii) there.

We begin by proving inequality (3). We collect as long as possible pairs of distinct columns in M', say $c_1, c_2, \ldots, c_{n_1}$ with n_1 even, such that c_{2i-1} and c_{2i}, $i = 1, 2, \ldots, n_1/2$, have in at least s positions the same nonzero entries. Then for any two distinct of the remaining $n_2 := n' - n_1$ columns, the number of positions with the same nonzero entries is at most $s-1$. By Lemma 2.1 (ii), the positions of the nonzero entries determine also these nonzero entries. Hence, each of these n_2 columns with at least s nonzero entries is determined by a subset of size s of the set of row-indices with nonzero entries, and the other columns have less than s nonzero entries, thus $n_2 \leq \sum_{i=1}^{s} \binom{m}{i}$.

From the columns $c_1, c_2, \ldots, c_{n_1}$ we form a new matrix M^* of dimension $m \times n_1/2$ with columns $c_1 - c_2, c_3 - c_4, \ldots, c_{n_1-1} - c_{n_1}$, where $-c_j$ is the additive inverse of c_j in $(GF(q))^m$. These $n_1/2$ columns are pairwise distinct (and not equal to the all zeros column), as otherwise $c_{2i-1} - c_{2i} = c_{2j-1} - c_{2j}$ for some $i \neq j$ implies dependence of these four columns which contradicts the assumption that the columns of M are k-wise independent with $k \geq 4$. Each column in M^* contains at most $2r - 2s$ nonzero entries and the columns are $k/2$-wise independent as k is even, hence $n_1/2 \leq N_q(m, k/2, 2r-2s)$. Summing up, we infer

$$c^* \cdot n \leq n' = n_1 + n_2 \leq 2 \cdot N_q(m, k/2, 2r-2s) + \sum_{i=1}^{s} \binom{m}{i}$$

and inequality (3) follows with $c := r^r \cdot q^r / r!$.

Next we will prove inequality (4). We partition the set of columns of M' into two parts and put these into two matrices M_1 and M_2 of dimensions $m \times n_1$ and $m \times n_2$, respectively, with $n' = n_1 + n_2$. In M_1 we put those columns in M' which have with some other column from M' at least s nonzero entries at the same positions. In matrix M_2 we put the remaining columns, i.e. those, which have with any other column from M' less than s nonzero entries at the same positions. Clearly, $n_2 \leq \sum_{i=1}^{s} \binom{m}{i}$ as above.

Set $[m] := \{1, 2, \ldots, m\}$ and for a column c, let $|c|$ denote the number of nonzero entries in c. Consider the matrix M_1. For any s-element subset $S \in [[m]]^s$ of row-indices, let $n(S)$ denote the number of columns in M_1 which have a nonzero entry at each position $s \in S$ and set

$$L := \sum_{S \in [[m]]^s} n(S) = \sum_{c \in M_1} \binom{|c|}{s} . \tag{5}$$

Clearly, we have $n_1 \leq L$ since each column in M_1 contains at least s nonzero entries. By the Cauchy-Schwartz inequality, we infer

$$\sum_{S \in [[m]]^s} (n(S))^2 \geq \frac{L^2}{\binom{m}{s}} ,$$

and with (5) we obtain

$$\sum_{S \in [[m]]^s} \binom{n(S)}{2} \geq \frac{1}{2} \cdot \frac{L \cdot (L - \binom{m}{s})}{\binom{m}{s}} \; . \tag{6}$$

Consider the matrix M_1^* obtained from M_1 by taking all differences $c_i - c_j$, $i < j$, of those columns, which share at least at s positions the same nonzero entries. Since in the matrix M the columns are 4-wise independent over $GF(q)$, the columns in M_1^* are pairwise distinct. Each column in M_1^* contains at most $2r - 2s$ nonzero entries and the columns in M_1^* are $k/2$-wise independent, hence the number of columns in M_1^* is at most $N_q(m, k/2, 2r - 2s)$. In the sum $\sum_{S \in [[m]]^s} \binom{n(S)}{2}$ every pair of distinct columns is counted at most $\binom{r-1}{s}$ times, since two distinct columns have at most $r - 1$ common positions with the same nonzero entry, hence

$$\sum_{S \in [[m]]^s} \binom{n(S)}{2} \leq \binom{r-1}{s} \cdot N_q(m, k/2, 2r - 2s) \; . \tag{7}$$

It follows from (6) and (7) that

$$\frac{1}{2} \cdot \frac{L \cdot (L - \binom{m}{s})}{\binom{m}{s}} \leq \binom{r-1}{s} \cdot N_q(m, k/2, 2r - 2s) \; ,$$

hence we infer

$$n_1 \leq L \leq \sqrt{2 \cdot \binom{m}{s} \cdot \binom{r-1}{s} \cdot N_q(m, k/2, 2r - 2s)} + \binom{m}{s} \; .$$

With $n_1 + n_2 = n' \geq c^* \cdot n$ and $n_2 \leq \sum_{i=1}^{s} \binom{m}{i}$ and $c := q^r \cdot r^r / r!$ the upper bound (4) follows. $\square$

Next we will give some consequences of Theorem 3.1.
From (3) we infer for fixed integers $k = 2^j$, $j \geq 1$, and $r \geq 1$ with $\gcd(k-1, r) = k - 1$ that

$$N_q(m, k, r) = O\left(m^{kr/(2(k-1))}\right) \; . \tag{8}$$

To see this, we use induction on j. For $j = 1$, the upper bound (8) holds. Let $k = 2^j$ and $\gcd(k-1, r) = k - 1$. By (3) with $s := kr/(2(k-1))$ it suffices to show that $\gcd(k/2 - 1, 2r - 2s) = k/2 - 1$, which holds as $2r - 2s = (k-2)r/(k-1)$, and that

$$\frac{k/2 \cdot (2r - 2s)}{2(k/2 - 1)} \leq \frac{kr}{2(k-1)} \iff \frac{kr}{2(k-1)} \leq s \; ,$$

which holds by choice of s.
Without any divisibility conditions, we infer for fixed integers $k = 2^l$ and $r \geq 1$ that

$$N_q(m, k, r) = O\left(m^{\lceil kr/(2(k-1)) \rceil}\right) \; , \tag{9}$$

which implies (8) for $\gcd(k-1,r) = k-1$. Clearly, (9) holds for $l = 1$. Using induction on l, it suffices by (3) with $s := \lceil kr/(2(k-1)) \rceil$ to show that

$$
\begin{aligned}
& \left\lceil \frac{k/2 \cdot (2r-2s)}{2(k/2-1)} \right\rceil \leq \left\lceil \frac{kr}{2(k-1)} \right\rceil \\
\Longleftarrow \quad & \frac{k(r-s)}{k-2} \leq \left\lceil \frac{kr}{2(k-1)} \right\rceil \qquad\qquad \left(\text{since } \left\lceil \tfrac{kr}{2(k-1)} \right\rceil \text{ is an integer}\right) \\
\Longleftrightarrow \quad & \frac{kr}{k-2} - \frac{k \cdot \left\lceil \frac{kr}{2(k-1)} \right\rceil}{k-2} \leq \left\lceil \frac{kr}{2(k-1)} \right\rceil \\
\Longleftarrow \quad & \frac{kr}{2(k-1)} \leq \left\lceil \frac{kr}{2(k-1)} \right\rceil ,
\end{aligned}
$$

which obviously holds, and hence (9) is shown, compare also [18].
Inequality (4) gives in some cases better estimates than (3), namely:

Corollary 3.1 *Let $k = 2^j$, $j \geq 1$, $r \geq 1$ and q a prime power be fixed integers. Then, for positive integers m,*

$$N_q(m,k,r) = O\left(m^{\lceil kr/(k-1) \rceil /2}\right) . \tag{10}$$

Proof. For the proof we use induction on j, compare Corollary 3 in [18].
For $j = 1$ we have that $N_q(m,2,r) = \Theta(m^r)$. For $k = 2^j$, let $s := \lfloor \lceil kr/(k-1) \rceil /2 \rfloor$. Since $s \leq \lceil kr/(k-1) \rceil /2$ it suffices by (4) to prove

$$\frac{1}{2} \cdot \left(s + \frac{1}{2} \cdot \left\lceil \frac{k/2 \cdot (2r-2s)}{k/2-1} \right\rceil \right) \leq \frac{\lceil kr/(k-1) \rceil}{2} ,$$

which is equivalent to

$$\left\lceil \frac{k(r-s)}{k/2-1} \right\rceil \leq 2 \cdot \left(\lceil kr/(k-1) \rceil - s \right) . \tag{11}$$

Since the right hand side of (11) is an integer, it suffices to prove

$$
\begin{aligned}
& \frac{k(r-s)}{k/2-1} \leq 2 \cdot \left(\lceil kr/(k-1) \rceil - s \right) \\
\Longleftrightarrow \quad & \lceil kr/(k-1) \rceil - 2s \leq (k-1) \cdot \lceil kr/(k-1) \rceil - kr . \qquad (12)
\end{aligned}
$$

The right hand side of (12) is at least 0 and its left hand side is at most 1. If $\lceil kr/(k-1) \rceil$ is even, (12) holds, since its left hand side is equal to 0. If $\lceil kr/(k-1) \rceil$ is odd, then (12) also holds, since the right hand side is odd, thus at least 1, hence (10) holds. $\square$

The next two lemmas show that asymptotically it suffices to consider the growth rate of $N_q(m,k,r)$ for q a prime.

Lemma 3.2 *Let $k \geq 2$, $l \geq 1$, $r \geq 1$, and a prime p be fixed integers. Then there exists a constant $d > 0$ such that for positive integers m,*

$$N_{p^l}(m,k,r) \leq d \cdot N_p(m,k,r) . \tag{13}$$

Proof. Let M be a (k,r)-matrix over $GF(p^l)$ of dimension $m \times n$, where $n = N_{p^l}(m,k,r)$. By Lemma 2.1, the matrix M contains an $m \times n'$-submatrix M' satisfying (i) – (iii) there, hence $n' \geq c \cdot n$ for some constant $c \geq r!/(r^r \cdot p^{lr})$. We form a new $m \times n'$-matrix M^* from M' by identifying every nonzero entry in M' by $1 \in GF(p)$. By Lemma 2.1 (ii), the columns in M^* are pairwise distinct and each column contains at most r nonzero entries. If $n' > N_p(m,k,r)$, then some $j \leq k$ columns in M^*, say $a_1^*, \ldots, a_j^*$, are linearly dependent over $GF(p)$, but then the corresponding columns $a_1', \ldots, a_j'$ in M' are also linearly dependent over $GF(p^l)$, which contradicts the assumption that M' is a (k,r)-matrix over $GF(p^l)$, hence (13) follows with $d \leq (p^{lr} \cdot r^r)/r!$. □

Lemma 3.3 *Let $k \geq 2$, $r \geq 1$ and p a prime be fixed integers. Then there exists a constant $c > 0$ such that for positive integers m,*

$$N_{p^l}(m,k,r) \geq c \cdot N_p(m,k,r) \ . \tag{14}$$

Proof. Let M be a (k,r)-matrix over $GF(p)$ of dimension $m \times n$, where $n = N_p(m,k,r)$. By Lemma 2.1, the matrix M contains an $m \times n'$-submatrix M' with entries $a'_{h,i}$ satisfying (i) – (iii) there, hence $n' \geq c \cdot N_p(m,k,r)$ for some constant $c \geq r!/(r^r \cdot p^{lr})$. All nonzero entries in row h have some value $e_h \in GF(p) \setminus \{0\}$.
We claim that the columns of M' are also linearly independent over $GF(p^l)$. To see this, consider the entries of the matrix M' as from $GF(p^l)$. Suppose for contradiction that some $j \leq k$ columns $a_1', \ldots, a_j'$ of M' are linearly dependent over $GF(p^l)$, hence for some $\lambda_i \in GF(p^l)$ we have $\sum_{i=1}^{j} \lambda_i \cdot a_i' = 0$. For row h in M', $h = 1, \ldots, m$, let $I_h = \{i \in \{1, \ldots, j\} \mid a'_{h,i} \neq 0\}$. For every $h = 1, \ldots, m$ with $I_h \neq \emptyset$ and for some nonzero element $e_h \in GF(p) \setminus \{0\}$ we have

$$0 \quad = \quad \sum_{i \in I_h} \lambda_i \cdot a'_{h,i} = \sum_{i \in I_h} \lambda_i \cdot e_h \ ,$$

hence $\sum_{i \in I_h} \lambda_i = 0$. However, since $a_1, \ldots, a_j$ are linearly independent over $GF(p)$ we infer in $GF(p^l)$ that $\lambda_1 = \ldots = \lambda_j = 0$ and (14) follows. □

Corollary 3.4 *Let $k \geq 2$, $r \geq 1$ and a prime p be fixed integers. Then, for positive integers m,*

$$N_{p^l}(m,k,r) = \Theta(N_p(m,k,r)) \ . \tag{15}$$

4 Graphs without Short Cycles, the Case $r = 2$

Using our previous considerations, in this section we will show some consequences on the growth of $N_q(m,k,r)$ for $r = 2$, i.e. each column contains at most two nonzero entries.

Corollary 4.1 *Let $k \geq 2$ and a prime power q be fixed integers. Then, for some constant $c > 0$ and for every positive integer m,*

$$N_q(m,k,2) \leq c \cdot m^{1+2/2^{\lfloor \log k \rfloor}} \ . \tag{16}$$

Proof. We use induction on $\lfloor \log_2 k \rfloor$. Inequality (16) holds for $k = 2, 3$ by Corollary 2.2. Assume it holds for all $k' < 2^{\lfloor \log k \rfloor}$. Let $k = 2^{\lfloor \log k \rfloor} + j$, $k \geq 4$, with $0 \leq j < 2^{\lfloor \log k \rfloor}$. By (4) for $s := 1$ and for even $k \geq 4$ we infer that $N_q(m, k, 2) \leq c' \cdot m^{1/2} \cdot N_q(m, k/2, 2)^{1/2} + c' \cdot m$ for some constant $c' > 0$ and (16) follows by the induction assumption. For odd $k \geq 5$, we have by monotonicity and by (4) that $N_q(m, k, 2) \leq N_q(m, k-1, 2) \leq c' \cdot m^{1/2} \cdot N_q(m, (k-1)/2, 2)^{1/2} + c' \cdot m$ and again (16) follows by the induction assumption. $\square$

Corollary 4.2 *Let q be a fixed prime power. Then, for positive integers m,*

$$\begin{aligned} N_q(m, 4, 2) &= \Theta(m^{3/2}) \\ N_q(m, 5, 2) &= \Theta(m^{3/2}) \,. \end{aligned}$$

Proof. The upper bound for $N_q(m, 4, 2)$ follows from (16). The lower bound can be shown similarly as in [18]. Let s be the largest prime power with $2 \cdot (s^2 - 1) \leq m$. Partition the set $\{1, \ldots, 2s^2 - 2\}$ of row-indices into two sets A and B of equal size $s^2 - 1$. Identify the elements of both A and B with the elements of $(GF(s))^2 \setminus \{(0, 0)\}$, i.e. $A = B = (GF(s))^2 \setminus \{(0, 0)\}$. We define an $m \times n$-matrix M with exactly two nonzero entries in each column by putting in each column always within row-set A a 1 at some position $g \in (GF(s))^2 \setminus \{(0, 0)\}$ and within row-set B some fixed nonzero element $e \in GF(q) \setminus \{0\}$ at some position $h \in (GF(s))^2 \setminus \{(0, 0)\}$ if and only if $< g, h >= 1$, where $<, >$ denotes the usual component-wise scalar product. All other entries within the row-sets A and B and the entries in rows $l \notin A \cup B$ are equal to 0.
By construction no three columns in M are linearly dependent over $GF(q)$. If four distinct columns $a_1, \ldots, a_4$ would be linearly dependent over $GF(q)$, then for some nonzero row-positions $g_i, h_i \in (GF(s))^2 \setminus \{(0, 0)\}$, $i = 1, 2$, we infer $< g_1, h_1 >=< g_2, h_2 >=< g_1, h_2 >=< g_2, h_1 >= 1$. The row-positions g_1, g_2, h_1, h_2 are pairwise distinct, as otherwise we have two identical columns. Hence $< g_1, h_1 - h_2 >= 0$ and $< g_2, h_1 - h_2 >= 0$, thus g_1 and g_2 are collinear, i.e. $g_1 = \lambda \cdot g_2$ for some $\lambda \in GF(s)$. But then $< g_1, h_1 >= \lambda \cdot < g_2, h_1 >= 1$ and $< g_2, h_1 >= 1$ implies $\lambda = 1$, hence $g_1 = g_2$, a contradiction.
The matrix M has $m = \Theta(s^2)$ rows and $n = \Theta(s^3)$ columns and, since the prime powers are sufficiently dense, the lower bound $N_q(m, 4, 2) = \Omega(m^{3/2})$ follows.
With Corollary 2.3 and by monotonicity we infer $N_q(m, 5, 2) = \Theta(m^{3/2})$. $\square$

Indeed, for a proof of Corollary 4.2 we can also identify the set $\{1, \ldots, m\}$ of row-indices of a matrix M with the vertex set of a graph on m vertices, which has n edges and contains no cycles of length at most 4 or 5, respectively. We construct an $m \times n$-matrix, where the columns in M have exactly two entries 1 and correspond in a natural way to the edges of the graph. Then the result follows also from the known results for graphs. This leads to the following observation:

Corollary 4.3 *Let $k \geq 3$ and a prime power q be fixed integers. Then for positive integers m,*

$$N_q(m, k, 2) \geq (1 - o(1)) \cdot N_2(m, k, 2) \,. \tag{17}$$

Proof. The number $N_2(m, k, 2)$ is asymptotically equal to the number of edges in a graph on m vertices without any cycle of length at most k.

Let $G = (V, E)$ be a graph on m vertices and with n edges without any cycle of length at most k. We construct an $m \times n$-matrix M with two entries 1 and $e \in GF(q) \setminus \{0\}$ in each column. The row-indices of M correspond to the vertices of the graph and the column-indices correspond to the edges in the graph G and for an edge $\{u, v\} \in E$ with $u < v$ we put the entries 1 and e at row-positions u and v in the column.
Suppose that $j \leq k$ columns of the matrix M are linearly dependent over $GF(q)$, where j is minimal with this property. The $2j$ nonzero entries in these j columns are contained in at most $2 \cdot \lfloor j/2 \rfloor \leq j$ rows due to the linear dependence. In terms of the graph we have j edges which cover at most j vertices. Among these edges there must be a cycle of length i, $i \leq j \leq k$, but the graph G was supposed to contain no cycles of length at most k. $\square$

From (17) and $N_2(m, 2k+1, 2) \geq 1/2 \cdot N_2(m, 2k, 2)$ we immediately obtain

Corollary 4.4 *Let $k \geq 2$ and a prime power q be fixed integers. Then, for positive integers m,*

$$N_q(m, 2k+1, 2) \geq (1/2 - o(1)) \cdot N_2(m, 2k, 2) .$$

Also from (17) we have the following lower bounds from the case of graphs, see [4, 17, 26]:

Corollary 4.5 *Let $k \geq 1$ and a prime power q be fixed integers. Then, for positive integers m,*

$$\begin{aligned} N_q(m, 6, 2) &= \Omega(m^{4/3}) \\ N_q(m, 10, 2) &= \Omega(m^{6/5}) \\ N_q(m, 2k, 2) &= \Omega(m^{1+2/(3k-3+\varepsilon)}) \end{aligned}$$

with $\varepsilon \in \{0, 1\}$ and $\varepsilon = 1$ if and only if k is odd.

Moreover, with Lemmas 3.2 and 3.3 we have the following bounds from the case of graphs, see [4, 26]:

Corollary 4.6 *Let $q = 2^l$ be fixed. Then, for positive integers m,*

$$\begin{aligned} N_q(m, 6, 2) &= \Theta(m^{4/3}) \\ N_q(m, 10, 2) &= \Theta(m^{6/5}) . \end{aligned}$$

From the results of Bondy and Simonovits [8] for the case of graphs and by Lemma 3.2 we obtain the following, compare also Corollary 4.1:

Corollary 4.7 *Let $q = 2^l$ and $k \geq 1$ be fixed integers. Then, for positive integers m,*

$$N_q(m, 2k, 2) = O(m^{1+1/k}) .$$

5 4-wise Independent Columns

Now we consider the case of matrices with 4-wise independent columns over $GF(q)$ and with at most r nonzero entries in each column.

Lemma 5.1 *Let $r \geq 1$ and a prime power q be fixed integers, where char $(GF(q)) > 2$. Let M' be an $m \times n$-matrix over $GF(q)$ with exactly r nonzero entries in each column, such that the assertions (ii) and (iii) in Lemma 2.1 are satisfied. Let $F_1', \ldots, F_n'$ be the sets of positions of the nonzero entries in the n columns of M'. If for no four sets both $F_g' \cup F_h' = F_i' \cup F_j'$ and $F_g' \cap F_h' = F_i' \cap F_j'$ are fulfilled, then the columns of the matrix M' are 4-wise independent.*

Proof. Suppose for contradiction that four columns $a_1, \ldots, a_4$ in M' are linearly dependent over $GF(q)$. Then, there exist nonzero elements $\lambda_1, \ldots, \lambda_4 \in GF(q) \setminus \{0\}$ such that $\sum_{i=1}^4 \lambda_i \cdot a_i = 0$. Let $F_1', \ldots, F_4'$ be defined as in the lemma. Let $S := F_1' \cap \ldots \cap F_4'$ and set $F_i := F_i' \setminus S$ for $i = 1, \ldots, 4$. Then the sets $F_1, \ldots, F_4$ are pairwise distinct.

Fact 5.2 *For any $1 \leq i < j < k \leq 4$ it is*

$$F_i \cap F_j \cap F_k = \emptyset \; .$$

Proof. Consider the $m \times 4$ matrix $M(a_1, \ldots, a_4)$. By assumption its columns $a_1, \ldots, a_4$ are linearly dependent but 3-wise independent over $GF(q)$.
Suppose first that each row in $M(a_1, \ldots, a_4)$ with at least one nonzero entry contains exactly three such entries. There are two distinct sets with nonempty intersection, say $F_1 \cap F_2 \neq \emptyset$, and let $C := F_1 \cap F_2$. Then for some subset $G \subseteq C$ we have $F_3 = (F_1 \Delta F_2) \cup G$ and $F_4 = (F_1 \Delta F_2) \cup (C \setminus G)$. However, the set $F_1 \Delta F_2$ cannot be contained in any set F_i by Lemma 2.1 (ii).
Hence there is some row in $M(a_1, \ldots, a_4)$, which contains exactly two nonzero entries, say row $i \in F_1 \cap F_2$, which implies $\lambda_2 = -\lambda_1$. Then every row $j \in F_1 \cap F_2$ contains also exactly two nonzero entries, otherwise, say $j \in F_3 \cap F_1 \cap F_2$ for $j \neq i$ implies $\lambda_3 = 0$, a contradiction, thus $F_1 \cap F_2 \cap F_i = \emptyset$ for $i = 3, 4$. By symmetry assume that $F_2 \cap F_3 \cap F_4 = H \neq \emptyset$. Then $\lambda_2 + \lambda_3 + \lambda_4 = 0$. With $\lambda_2 = -\lambda_1$ this implies with char $(GF(q)) > 2$ that $F_1 \cap F_3 \cap F_4 = \emptyset$. Moreover, we have $H = F_2 \setminus (F_1 \cap F_2)$ since $\lambda_i \neq 0$ for $i = 1, \ldots, 4$. But then the matrix M' does not satisfy Lemma 2.1 (ii), a contradiction. □

Two of the sets $F_1, \ldots, F_4$ have nonempty intersection, say $F_1 \cap F_2 \neq \emptyset$, hence $\lambda_2 = -\lambda_1$ by Fact 5.2. If $F_1 \cap F_3 \neq \emptyset$ and $F_2 \cap F_3 \neq \emptyset$, then $\lambda_3 = -\lambda_1$ and $\lambda_2 = -\lambda_3$ by Fact 5.2, thus $\lambda_1 = 0$ with char $(GF(q)) > 2$, a contradiction. Hence, $F_3 \cap (F_1 \setminus F_2) = \emptyset$ or $F_4 \cap (F_1 \setminus F_2) = \emptyset$.
Therefore we have $F_3 \setminus (F_1 \cup F_2) = F \neq \emptyset$. Due to the dependence of $a_1, \ldots, a_4$ we obtain $F_4 \setminus (F_1 \cup F_2) = F$ thus $\lambda_3 = -\lambda_4$. But then either $F_3 = F \cup (F_2 \setminus F_1)$ and $F_4 = F \cup (F_1 \setminus F_2)$ or $F_3 = F \cup (F_1 \setminus F_2)$ and $F_4 = F \cup (F_2 \setminus F_1)$. In the first case we have $F_1 \cup F_3 = F_2 \cup F_4$ and $F_1 \cap F_3 = F_2 \cap F_4$ and similarly in the second case, contradicting the assumption. □

In [14] Frankl and Füredi proved that there exists a family $\mathcal{F}$ of r-element subsets of an m-element set containing no four sets $F_1, \ldots, F_4$ with $F_1 \cup F_2 = F_3 \cup F_4$ and $F_1 \cap F_2 = F_3 \cap F_4$ where $|\mathcal{F}| = \Omega(m^{\lceil 4r/3 \rceil / 2})$. Their construction is based on symmetric polynomials over finite fields: Let $r \equiv 1 \bmod 3$, say $r = 3t + 1$. (For other values of $(r \bmod 3)$ the construction is similar.) For given positive integers m let K be any field with $m/2 \leq |K| \leq m$. For a subset $X = \{x_1, \ldots, x_g\} \subseteq K$ and an integer i let

$$s_i(X) := \sum_{I \in [[g]]^i} \prod_{j \in I} x_j$$

be the ith elementary symmetric polynomial in the variables $x_1, \ldots, x_g$, where $s_i(X) = 0$ for $i < 0$ or $i > |X|$. For given integers $h \geq 1$ define an $h \times h$-matrix $M_h(X)$ with entries $m_{i,j} = s_{2i-j}(X)$. Then for suitable elements $c_2, c_4, \ldots, c_{2t} \in K$ the family $\mathcal{F}$ of r-element subsets of K is defined as follows:
$X = \{x_1, \ldots, x_r\} \in \mathcal{F}$ if $s_{2i}(X) = c_{2i}$ for $i = 1, \ldots, t$ and $\det(M_h(S)) \neq 0$ for every subset $S \subseteq X$ and $h = 1, \ldots, |S| - 1$.
This yields a polynomial time (semi-) construction and we conclude:

Corollary 5.3 *Let $r \geq 1$ and a prime power q be fixed integers, where char $(GF(q)) > 2$. Then, for positive integers m,*

$$N_q(m, 4, r) = \Theta\left(m^{\lceil 4r/3 \rceil/2}\right) .$$

Proof. The upper bound follows immediately from Corollary 3.1. For the lower bound, let $\mathcal{F} = \{F_1, \ldots, F_n\}$ be a maximum family of r-element subsets of $\{1, \ldots, m\}$ with $n = \Theta(m^{\lceil 4r/3 \rceil/2})$, such that for no four sets $F_i, F_j, F_k, F_l \in \mathcal{F}$ it is $F_i \cup F_j = F_k \cup F_l$ and $F_i \cap F_j = F_k \cap F_l$. This family exists by the above mentioned result of Frankl and Füredi. Define an $m \times n$-matrix M with entries 0 and 1, which has columns $c_1, \ldots, c_n$. In column c_i there is an entry 1 in position f if and only if $f \in F_i$, $i = 1, \ldots, n$. By Lemma 2.1 we obtain an $m \times n'$-submatrix M' of M with $n' \geq c \cdot n$ for some constant $c > 0$ such that (ii) (in each row-set $R_1, \ldots, R_r$ there is exactly one entry 1) and (iii) there are satisfied. By Lemma 5.1, the columns of M' are 4-wise independent and the lower bound follows. □

Corollary 5.4 *Let $r \geq 1$ and $q = 2^l$ be fixed integers. Then, for positive integers m,*

$$N_q(m, 4, r) = O\left(m^{\lceil 4r/3 \rceil/2}\right) .$$

Proof. The upper bound follows immediately from Corollary 3.1, or alternatively from Lemma 3.2 and Corollary 3 in [18]. □

Notice, that from Corollary 6.2, which is stated in the next section, we have the lower bound $N_q(m, 4, r) = \Omega(m^{2r/3})$. To avoid four dependent columns over $GF(q)$, more configurations than mentioned in Lemma 5.1 have to be forbidden in the case char $(GF(q)) = 2$.

6 Lower Bounds

For proving our lower bounds on $N_q(m, k, r)$ we will use hypergraphs. A *hypergraph* $\mathcal{G} = (V, \mathcal{E})$ has vertex set V and edge set $\mathcal{E}$ with $E \subseteq V$ for every edge $E \in \mathcal{E}$. A hypergraph $\mathcal{G} = (V, \mathcal{E})$ is called *l-uniform*, if the edge set $\mathcal{E}$ contains only l-element edges, i.e. $\mathcal{E} \subseteq [V]^l$. An *independent set* in a hypergraph $\mathcal{G} = (V, \mathcal{E})$ is a subset $I \subseteq V$ which contains no edges from $\mathcal{E}$. A *2-cycle* in an l-uniform hypergraph $\mathcal{G} = (V, \mathcal{E})$ is a pair $\{E, E'\}$ of distinct edges $E, E' \in \mathcal{E}$ with $|E \cap E'| \geq 2$.
For proving our lower bounds on the dimensions of large (k, r)-matrices over $GF(q)$, we will reformulate our problem in terms of finding in a suitably defined hypergraph a large independent set.

Theorem 6.1 *Let $k \geq 4$, $r \geq 1$ and a prime power q be fixed integers. Then, for positive integers m,*

$$N_q(m,k,r) = \Omega\left(m^{\frac{kr}{2(k-1)}} \cdot (\log m)^{\frac{1}{k-1}}\right) \quad \textit{for } k \textit{ even and } \gcd(k-1,r)=1 \tag{18}$$

and

$$N_q(m,k,r) = \Omega\left(m^{\frac{(k-1)r}{2(k-2)}} \cdot (\log m)^{\frac{1}{k-2}}\right) \quad \textit{for } k \textit{ odd and } \gcd(k-2,r)=1. \tag{19}$$

As a by-product the proof of Theorem 6.1 yields lower bounds on $N_q(m,k,r)$ for arbitrary fixed pairs (k,r), see Corollary 6.2. The case $q=2$ was considered in [5], hence with Lemma 3.3 inequalities (18) and (19) hold for $q = 2^l$. However, in the proof of Theorem 6.1 we cannot make use of the fact that it suffices by Lemma 3.3 to consider primes q only.
Proof. We partition the set $\{1,\ldots,m\}$ of row-indices into r subsets $R_1,\ldots,R_r$ of nearly equal size $\lfloor m/r \rfloor$ or $\lceil m/r \rceil$. According to some choice of a sequence $(e_1,\ldots,e_r) \in (GF(q)\setminus\{0\})^r$ of nonzero elements, let $C_q(m,r)$ consist of all column vectors of length m, which contain within each row-set R_j exactly one nonzero entry $e_j \in GF(q)\setminus\{0\}$, $j=1,\ldots,r$. Hence $|C_q(m,r)| \geq (\lfloor m/r \rfloor)^r$, say $|C_q(m,r)| = c \cdot m^r$ for some constant $c > 0$. By the proof of Lemma 2.1 (iii) the columns of $C_q(m,r)$ are 3-wise independent.
We form a hypergraph $\mathcal{G} = (V, \mathcal{E}_3 \cup \ldots \cup \mathcal{E}_k)$ with vertex set $V = C_q(m,r)$. An i-element subset $\{a_1,\ldots,a_i\}$ of V, $i=4,\ldots,k$, is an edge in this hypergraph $\mathcal{G}$, that is $\{a_1,\ldots,a_i\} \in \mathcal{E}_i$, if and only if $a_1,\ldots,a_i$ are linearly dependent but any $h < i$ of these columns are linearly independent over $GF(q)$. Then, an independent set in this hypergraph $\mathcal{G}$ yields a set of k-wise independent column vectors. In the following we will prove a lower bound on the maximum size of an independent set in $\mathcal{G}$.
First we will bound from above the numbers $|\mathcal{E}_i|$, $i=4,\ldots,k$, of i-element edges in $\mathcal{G}$. For a subset E of i column vectors $a_1,\ldots,a_i \in C_q(m,r)$ consider the corresponding $m \times i$-matrix $M(E)$. This matrix $M(E)$ contains exactly $i \cdot r$ nonzero entries. If $a_1,\ldots,a_i$ are linearly dependent over $GF(q)$, but not any $h < i$ of these, then in each row of $M(E)$ there are either at least two nonzero entries or all entries are zero. Since every column contains within each row-set R_j exactly one nonzero entry $e_j \in GF(q)\setminus\{0\}$, within each row-set R_j, $j=1,\ldots,r$, the i nonzero entries e_j of $M(E)$ are contained in at most $\lfloor i/2 \rfloor$ rows. Therefore, in $M(E)$ all the nonzero entries are contained in at most $\lfloor i/2 \rfloor \cdot r$ rows. By construction, the choice of the rows determines also the nonzero entries in these rows. Thus, for some constants $c_i > 0$, $i=4,\ldots,k$, the number of i-element edges in the hypergraph $\mathcal{G}$ satisfies

$$|\mathcal{E}_i| \leq \binom{m}{\lfloor i/2 \rfloor \cdot r} \cdot \binom{i \cdot \lfloor i/2 \rfloor \cdot r}{ir} \leq c_i \cdot m^{\lfloor i/2 \rfloor \cdot r} . \tag{20}$$

For some value $l \geq 3$, which will be fixed later and only depends on the parity of k, we consider for the moment only the l-element edges in $\mathcal{G}$, i.e. edges in $\mathcal{E}_l$.
We will now take care of the 2-cycles arising from the edges in $\mathcal{E}_l$. Recall that a 2-*cycle* is a pair $\{E,E'\}$ of distinct edges $E,E' \in \mathcal{E}_l$ with $|E \cap E'| \geq 2$. A 2-*cycle* $\{E,E'\}$ is called $(2,j)$-*cycle* if $|E \cap E'| = j$, where $j = 2,\ldots,l-1$.
We will apply a result of Ajtai, Komlós, Pintz, Spencer and Szemerédi [1], originally an existence result, see also [10], in the sequel extended and turned into a deterministic polynomial time algorithm in [13]. Here we will use it in its algorithmic version from [6]:

Theorem 6.2 *Let $l \geq 3$ be a fixed integer. Let $\mathcal{G} = (V, \mathcal{E})$ be an l-uniform hypergraph on $|V| = N$ vertices and with average degree $t^{l-1} := l \cdot |\mathcal{E}|/|V|$.*
If the hypergraph $\mathcal{G} = (V, \mathcal{E})$ contains no 2-cycles, then one can find for any fixed $\delta > 0$ in $\mathcal{G}$ in time $O(N \cdot t^{l-1} + N^3/t^{3-\delta})$ an independent set of size at least $\Omega(N/t \cdot (\log t)^{1/(l-1)})$. The assertion also holds, if the parameter t^{l-1} is an upper bound on the average degree.

To apply Theorem 6.2 we will show in the following that there are not too many 2-cycles arising from $\mathcal{E}_l$ and these will be discarded randomly. For a j-element subset $J = \{a_1, \ldots, a_j\} \subseteq C_q(m, r)$ of column vectors, $j = 2, \ldots, l-1$, let $p(J)$ be the number of rows in the corresponding matrix $M(J)$ which contain at least one nonzero entry. Moreover, let $p_1(J)$ be the number of rows in $M(J)$ with exactly one nonzero entry.
Let $b(J)$ be the number of $(l-j)$-element subsets $S = \{b_1, \ldots, b_{l-j}\} \subseteq C_q(m, r)$ such that $\{a_1, \ldots, a_j, b_1, \ldots, b_{l-j}\} \in \mathcal{E}_l$, that is, the column vectors $a_1, \ldots, a_j, b_1, \ldots, b_{l-j}$ are linearly dependent but any $h < l$ of these are linearly independent over $GF(q)$. If $J \cup S \in \mathcal{E}_l$, then for every row in $M(J)$ with exactly one nonzero entry e there must be in the same row of $M(S)$ at least one nonzero entry e and all these nonzero entries are identical. There are at most $(l-j)^{p_1(J)}$ possibilities to choose the positions of these *matching* nonzero entries in $M(S)$.
Let $M(J)$ contain the $p(J)$ nonzero rows $1, \ldots, p(J)$, say. If $M(S)$ contains in row $s > p(J)$ at least one nonzero entry, then there must be in $M(S)$ in this row at least two nonzero entries, since the columns $a_1 \ldots, a_j, b_1 \ldots, b_{l-j}$ are linearly dependent over $GF(q)$, but not any $h < l$ of these. Therefore, we have at most $\lfloor ((l-j)r - p_1(J))/2 \rfloor$ rows $s > p(J)$ in $M(S)$ with nonzero entries. To choose these rows there are at most

$$\binom{m - p(J)}{\lfloor \frac{(l-j)r - p_1(J)}{2} \rfloor}$$

possibilities. Having fixed these rows, to choose the positions of the at most $((l-j)r - p_1(J))$ remaining nonzero entries, we have at most $((\lfloor ((l-j)r - p_1(J))/2 \rfloor + p(J)) \cdot (l-j))^{(l-j)r - p_1(J)}$ choices, thus for some constant $c_p > 0$ we obtain

$$\begin{aligned} b(J) &\leq \binom{m}{\lfloor \frac{(l-j)r - p_1(J)}{2} \rfloor} \cdot ((\lfloor \frac{(l-j)r - p_1(J)}{2} \rfloor + p(J)) \cdot (l-j))^{(l-j)r - p_1(J)} \cdot (l-j)^{p_1(J)} \\ &\leq c_p \cdot m^{\lfloor \frac{(l-j)r - p_1(J)}{2} \rfloor} . \end{aligned} \tag{21}$$

Next, we consider $(2, j)$-cycles arising from the l-element edges, i.e. pairs $\{E, E'\}$ of distinct l-element edges from $\mathcal{E}_l$ with $|E \cap E'| = j \geq 2$.
For $j = 2, \ldots, l-1$ and $u = 0, \ldots, jr$, let $s_{2,j}(u; \mathcal{G}_l)$ be the number of $(2, j)$-cycles $\{E, E'\}$ in $\mathcal{G}_l = (V, \mathcal{E}_l)$ with $p_1(E \cap E') = u$ and of course $|E \cap E'| = j$. Clearly, the total number $s_{2,j}(\mathcal{G}_l)$ of $(2, j)$-cycles among the l-element edges satisfies

$$s_{2,j}(\mathcal{G}_l) = \sum_{u=0}^{j \cdot r} s_{2,j}(u; \mathcal{G}_l) . \tag{22}$$

Indeed, the summation in (22) only runs up to $\min \{jr, (l-j)r\}$ (but this we cannot use in the following), as for a j-element subset $J \subseteq C_q(m, r)$ we have $p_1(J) \leq jr$, and if this set J is contained in an l-element edge $E \in \mathcal{E}_l$, then $p_1(J) \leq (l-j)r$.

The number $p_{j,u}(V)$ of j-element subsets $J \in [V]^j$ of column vectors with $p_1(J) = u$ can be bounded from above for some constant $c_{j,u} > 0$ as follows:

$$
\begin{aligned}
p_{j,u}(V) &\leq \binom{m}{u} \cdot \binom{m-u}{\lfloor (jr-u)/2 \rfloor} \cdot j^u \cdot (\lfloor (jr-u)/2 \rfloor \cdot j)^{jr-u} \\
&\leq c_{j,u} \cdot m^{u + \lfloor \frac{jr-u}{2} \rfloor} , \qquad (23)
\end{aligned}
$$

since the matrix $M(J)$ has u rows with exactly one nonzero entry and the remaining $jr-u$ nonzero entries are contained in rows with at least two nonzero entries.
The number of $(2,j)$-cycles $\{E, E'\}$ in $\mathcal{G}_l = (V, \mathcal{E}_l)$ with $E \cap E' = J$ is at most $\binom{b(J)}{2}$, thus by (21) and (23) we infer for some constant $C_1 > 0$:

$$
\begin{aligned}
s_{2,j}(u; \mathcal{G}_l) &\leq \sum_{J \in [C_q(m,r)]^j;\, p_1(J) = u} \binom{b(J)}{2} \\
&\leq \frac{c_p^2}{2} \cdot \sum_{J \in [C_q(m,r)]^j;\, p_1(J) = u} m^{2 \cdot \lfloor \frac{(l-j)r-u}{2} \rfloor} \\
&= \frac{c_p^2}{2} \cdot p_{j,u}(V) \cdot m^{2 \cdot \lfloor \frac{(l-j)r-u}{2} \rfloor} \\
&\leq C_1 \cdot m^{2 \cdot \lfloor \frac{(l-j)r-u}{2} \rfloor + u + \lfloor \frac{jr-u}{2} \rfloor} . \qquad (24)
\end{aligned}
$$

By (20) the *average degree* t^{l-1} of the l-uniform hypergraph $\mathcal{G}_l = (V, \mathcal{E}_l)$ satisfies

$$
t^{l-1} = \frac{l \cdot |\mathcal{E}_l|}{|V|} \leq \frac{l \cdot c_l \cdot m^{\lfloor l/2 \rfloor \cdot r}}{c \cdot m^r} ,
$$

hence for some constant $C_2 > 0$ we have

$$
t \leq t_0 := C_2 \cdot m^{(\lfloor l/2 \rfloor \cdot r - r)/(l-1)} .
$$

To apply Theorem 6.2, we choose a random subset $V^* \subseteq V$ by picking vertices at random from V, independently of each other and each with probability $p := t_0^{-1} \cdot m^{\varepsilon}$ for some small constant $\varepsilon > 0$ to get a uniform hypergraph without any 2-cycles. We will estimate the expected values $E(\cdot)$ of certain parameters of the induced random hypergraph $\mathcal{G}^* = (V^*, \mathcal{E}_3^* \cup \ldots \cup \mathcal{E}_k^*)$ with $\mathcal{E}_i^* := \mathcal{E}_i \cap [V^*]^i$, $i = 4, \ldots, k$.
The expected number $E(|V^*|)$ of vertices in $\mathcal{G}^*$ satisfies for some constant $c^* > 0$:

$$
\begin{aligned}
E(|V^*|) &= p \cdot |C_q(m,r)| = t_0^{-1} \cdot m^{\varepsilon} \cdot c \cdot m^r \\
&\geq c^* \cdot m^{r - \frac{\lfloor l/2 \rfloor \cdot r - r}{l-1} + \varepsilon} . \qquad (25)
\end{aligned}
$$

By (20) the expected numbers $E(|\mathcal{E}_i^*|)$ of i-element edges, $i = 4, \ldots, k$, satisfy for some constants $c_i^* > 0$:

$$
E(|\mathcal{E}_i^*|) \leq p^i \cdot c_i \cdot m^{\lfloor i/2 \rfloor \cdot r} \leq c_i^* \cdot m^{\lfloor i/2 \rfloor \cdot r - \frac{\lfloor l/2 \rfloor \cdot r - r}{l-1} \cdot i + i \cdot \varepsilon} . \qquad (26)
$$

Let $p_{j,u}(V^*)$ be the numbers of j-element subsets $J \in [V^*]^j$ with $p_1(J) = u$ and let $E(p_{j,u}(V^*))$ be their expected values. With (23) we infer for $j = 2, \ldots, l-1$ and $u = 0, \ldots, j \cdot r$ and some constants $c^*_{j,u} > 0$:

$$\begin{aligned} E(p_{j,u}(V^*)) &= p^j \cdot p_{j,u}(V) \leq c_{j,u} \cdot p^j \cdot m^{u+\lfloor \frac{jr-u}{2} \rfloor} \\ &\leq c^*_{j,u} \cdot m^{u+\lfloor \frac{jr-u}{2} \rfloor - \frac{\lfloor l/2 \rfloor \cdot r - r}{l-1} \cdot j + j \cdot \varepsilon} . \end{aligned} \tag{27}$$

Let $s_{2,j}(u; \mathcal{G}^*_l)$ denote the numbers of pairs $\{E, E'\} \in [\mathcal{E}^*_l]^2$ of distinct edges with $p_1(E \cap E') = u$ and $|E \cap E'| = j$ in the random hypergraph $\mathcal{G}^*_l = (V^*, \mathcal{E}^*_l)$. By (24) the expected numbers $E(s_{2,j}(u; \mathcal{G}^*_l))$ satisfy for $u = 0, \ldots, jr$ and $j = 2, \ldots, l-1$ for some constant $C^*_1 > 0$:

$$\begin{aligned} & E(s_{2,j}(u; \mathcal{G}^*_l)) = p^{2l-j} \cdot s_{2,j}(u; \mathcal{G}_l) \leq \\ \leq \quad & C^*_1 \cdot m^{2 \cdot \lfloor \frac{(l-j)r-u}{2} \rfloor + u + \lfloor \frac{jr-u}{2} \rfloor - \frac{\lfloor l/2 \rfloor \cdot r - r}{l-1} \cdot (2l-j) + (2l-j) \cdot \varepsilon} . \end{aligned} \tag{28}$$

With (25) – (28) and using Markov's resp. Chebychev's inequality, we know that there exists a subhypergraph $\mathcal{G}^* = (V^*, \mathcal{E}^*_3 \cup \ldots \cup \mathcal{E}^*_k)$ of $\mathcal{G}$ with the following properties

$$|V^*| \geq c^* \cdot m^{r - \frac{\lfloor l/2 \rfloor \cdot r - r}{l-1} + \varepsilon} \tag{29}$$

$$|\mathcal{E}^*_i| \leq c^*_i \cdot m^{\lfloor i/2 \rfloor \cdot r - \frac{\lfloor l/2 \rfloor \cdot r - r}{l-1} \cdot i + i \cdot \varepsilon} \tag{30}$$

$$p_{j,u}(V^*) \leq c^*_{j,u} \cdot m^{u + \lfloor \frac{jr-u}{2} \rfloor - \frac{\lfloor l/2 \rfloor \cdot r - r}{l-1} \cdot j + j \cdot \varepsilon} \tag{31}$$

$$s_{2,j}(u; \mathcal{G}^*_l) \leq C^*_1 \cdot m^{2 \cdot \lfloor \frac{(l-j)r-u}{2} \rfloor + u + \lfloor \frac{jr-u}{2} \rfloor - \frac{\lfloor l/2 \rfloor \cdot r - r}{l-1} \cdot (2l-j) + (2l-j) \cdot \varepsilon} , \tag{32}$$

where we used for simplicity the same notation for the constant factors, although they differ from those above by a constant factor dependent only on k, r, q, but this will not change our asymptotic considerations.
Now we fix the value of l to $l := k$ if k is even and to $l := k-1$, if k is odd, hence l is always even.

Lemma 6.1 *For $k \geq 4$ and $0 < \varepsilon < r/(2(k-1)(k-2))$ it holds:*

$$|\mathcal{E}^*_i| = o(|V^*|) \qquad \textit{for every } i \neq l. \tag{33}$$

Proof. Since l is even, by (29) and (30), we have for $i = 4, \ldots, k$

$$\begin{aligned} |V^*| &\geq c^* \cdot m^{r - \frac{lr/2 - r}{l-1} + \varepsilon} \\ |\mathcal{E}^*_i| &\leq c^*_i \cdot m^{\lfloor i/2 \rfloor \cdot r - \frac{lr/2 - r}{l-1} \cdot i + i \cdot \varepsilon} , \end{aligned}$$

hence it is $|\mathcal{E}^*_i| = o(|V^*|)$ if

$$r - \left\lfloor \frac{i}{2} \right\rfloor \cdot r + (i-1) \cdot \frac{(l-2)r}{2(l-1)} - (i-1) \cdot \varepsilon > 0 . \tag{34}$$

Inequality (34) holds if

$$\frac{(l-i)r}{2(l-1)} - (i-1) \cdot \varepsilon > 0$$

which is fulfilled for $i = 4, \ldots, l-1$ and $\varepsilon < r/(2(l-1)(l-2))$.
For $i > l$, which is only possible for $i = k$ odd and $l = k-1$ inequality (34) is equivalent to

$$\frac{(k-3)r}{2(k-2)} - (k-1) \cdot \varepsilon > 0 \ ,$$

which holds for $0 < \varepsilon < ((k-3)r)/(2(k-1)(k-2))$, hence (33) holds for $0 < \varepsilon < r/(2(k-1)(k-2))$. $\square$

From Lemma 6.1 we infer:

Corollary 6.2 *Let q be a prime power and let $k \geq 4$ and $r \geq 1$ be fixed positive integers. Then, for positive integers m,*

$$N_q(m,k,r) = \Omega\left(m^{\frac{kr}{2(k-1)}}\right) \qquad \text{if } k \text{ is even} \tag{35}$$

and

$$N_q(m,k,r) = \Omega\left(m^{\frac{(k-1)r}{2(k-2)}}\right) \qquad \text{if } k \text{ is odd.} \tag{36}$$

Thus, for $k = 2^i$ and $\gcd(k-1, r) = k-1$ lower (35) and upper bound (10) match (and similarly for $k = 2^i + 1$ and $\gcd(k-2, r) = k-2$), while for even k and $\gcd(k-1, r) = 1$ as well as for odd k and $\gcd(k-2, r) = 1$ the lower bounds (35) resp. (36) can be improved, see (18) and (19).
Proof. From Lemma 6.1 we know that for all values $i \neq l$ we have $|\mathcal{E}_i^*| = o(|V^*|)$. We remove one vertex from each of the *bad edges*, i.e. i-element edges with $i \neq l$, and we obtain a subset $V^{**} \subseteq V^*$ with $|V^{**}| \geq (c^* - o(1)) \cdot m^{lr/(2(l-1))+\varepsilon} \geq (c^*/2) \cdot m^{lr/(2(l-1))+\varepsilon}$, where the induced subhypergraph $\mathcal{G}^{**}$ of $\mathcal{G}^*$ is l-uniform with $|[V^{**}]^l \cap \mathcal{E}_l^*| \leq |\mathcal{E}_l^*| \leq c_l^* \cdot m^{lr/(2(l-1))+l\cdot\varepsilon}$, thus $\mathcal{G}^{**} = (V^{**}, [V^{**}]^l \cap \mathcal{E}_l^*)$.
Again we pick vertices from V^{**} at random, independently of each other with probability $p := c_h \cdot m^{-\varepsilon}$ for the constant $c_h := (c^*/(4c_l^*))^{1/(l-1)}$.
Then for the random subset $V^{***} \subseteq V^{**}$ we obtain for the expected values

$$E(|V^{***}|) = p \cdot |V^{**}| \geq (c_h \cdot c^*/2) \cdot m^{lr/(2(l-1))} \ ,$$

and

$$E(|[V^{***}]^l \cap \mathcal{E}_l^*|) \leq p^l \cdot |\mathcal{E}_l^*| \leq c_h^l \cdot c_l^* \cdot m^{lr/(2(l-1))} \ .$$

Using linearity of expectation, there exists a subset $V^{***} \subseteq V^{**}$ such that

$$|V^{***}| - |[V^{***}]^l \cap \mathcal{E}_l^*| \geq c_h \cdot (c^*/2 - c_l^* \cdot c_h^{l-1}) \cdot m^{lr/(2(l-1))} \geq (c_h \cdot c^*/4) \cdot m^{lr/(2(l-1))} \ .$$

By deleting from V^{***} one vertex from every edge in $[V^{***}]^l \cap \mathcal{E}_l^*$ we obtain an independent set I in $\mathcal{G}$ with

$$|I| = \Omega\left(m^{lr/(2(l-1))}\right) \ ,$$

and the lower bounds (35) and (36) follow by inserting $l := k$ for k even, and $l := k-1$ for k odd. $\square$

Notice, that we could have derived Corollary 6.2 already from (20), using similar computations as above, by picking right away from the set V vertices at random, independently from each other, each with probability $p := c_h' \cdot t_0^{-1}$ with $c_h' = (c/(4c_l))^{1/(l-1)}$. Hence, matrices satisfying (35) or (36) respectively can be constructed in polynomial time by using the method of conditional probabilities.

Lemma 6.3 *For $j = 2, \ldots, l-1$ and $\varepsilon > 0$ and $u > ((l-j)r)/(l-1) + 2 \cdot (2l-j-1) \cdot \varepsilon$ it holds*

$$s_{2,j}(u; \mathcal{G}_l^*) = o(|V^*|) . \tag{37}$$

Proof. Using (29) and (32) with l even we have $s_{2,j}(u; \mathcal{G}_l^*) = o(|V^*|)$ for $j = 2, \ldots, l-1$ if

$$\begin{aligned}
& 0 > 2 \cdot \left\lfloor \frac{(l-j)r - u}{2} \right\rfloor + u + \left\lfloor \frac{jr-u}{2} \right\rfloor \\
& \quad - \frac{(l-2)r}{2(l-1)} \cdot (2l-j-1) - r + (2l-j-1) \cdot \varepsilon \\
\iff \quad & 0 > (l-1) \cdot r - 2 \cdot \left\lceil \frac{jr+u}{2} \right\rceil + \left\lfloor \frac{jr-u}{2} \right\rfloor + u \\
& \quad - \frac{(l-2)r}{2(l-1)} \cdot (2l-j-1) + (2l-j-1) \cdot \varepsilon \\
\Longleftarrow \quad & u/2 > (l-1) \cdot r - \frac{jr}{2} - \frac{(l-2)r}{2(l-1)} \cdot (2l-j-1) + (2l-j-1) \cdot \varepsilon \\
\iff \quad & u > \frac{(l-j)r}{l-1} + 2 \cdot (2l-j-1) \cdot \varepsilon
\end{aligned}$$

and (37) follows. $\square$

Lemma 6.4 *For $j = 2, \ldots, l-1$ and $\varepsilon > 0$ and for $u < ((l-j)r)/(l-1) - 2 \cdot (j-1) \cdot \varepsilon$ it is*

$$p_{j,u}(V^*) = o(|V^*|) . \tag{38}$$

Proof. With l even we have by (29) and (31) that $p_{j,u}(V^*) = o(|V^*|)$ if

$$\begin{aligned}
& u + \left\lfloor \frac{jr-u}{2} \right\rfloor - \frac{(l-2)r}{2(l-1)} \cdot j + j \cdot \varepsilon < r - \frac{(l-2)r}{2(l-1)} + \varepsilon \\
\iff \quad & u + \left\lfloor \frac{jr-u}{2} \right\rfloor < \frac{(l-2)r}{2(l-1)} \cdot (j-1) + r - (j-1) \cdot \varepsilon \\
\Longleftarrow \quad & u < \frac{(l-j)r}{l-1} - 2 \cdot (j-1) \cdot \varepsilon
\end{aligned}$$

and inequality (38) follows. $\square$

Consider the values $((l-j)r)/(l-1)$ for $j = 2, \ldots, l-1$. If $\gcd(l-1, r) = 1$, these are never integers. Moreover, $((l-j)r)/(l-1)$ is at least $1/(l-1)$ apart from the next integer. Using Lemmas 6.3 and 6.4, we choose $\varepsilon > 0$ so small such that both $2 \cdot (2l-j-1) \cdot \varepsilon < 1/(l-1)$

and $2 \cdot (j-1) \cdot \varepsilon < 1/(l-1)$ are fulfilled for $j = 2, \ldots, l-1$, say $\varepsilon := 1/((2k-2)(2k-3))$. Then, '$u > ((l-j)r)/(l-1) + 2 \cdot (2l-j-1) \cdot \varepsilon$ or $u < ((l-j)r)/(l-1) - 2 \cdot (j-1) \cdot \varepsilon$' is satisfied for $u = 0, \ldots, jr$ and $j = 2, \ldots, l-1$. We summarize Lemmas 6.3 and 6.4 as follows:

Corollary 6.5 *For $\varepsilon = 1/((2k-2)(2k-3))$ and $j = 2, \ldots, l-1$ and $u = 0, \ldots, jr$ and $\gcd(l-1, r) = 1$ it is valid*

$$min\ \{p_{j,u}(V^*), s_{2,j}(u; \mathcal{G}_l^*)\} = o(|V^*|)\ .$$

Now, from V^* we delete one vertex from each *bad edge* $E \in \mathcal{E}_i^*$ for $i \neq l$ and by Lemma 6.1, we obtain a subset $V^{**} \subseteq V^*$ with $|V^{**}| = (1 - o(1)) \cdot |V^*|$. The resulting induced subhypergraph on the vertex set V^{**} is l-uniform. Then we proceed for $j = 2, \ldots, l-1$ as follows. For $u > ((l-j)r)/(l-1) + 2 \cdot (2l-j-1) \cdot \varepsilon$ we delete one vertex from each $(2, j)$-cycle $\{E, E'\}$ with $E, E' \in \mathcal{E}_l^* \cap [V^{**}]^l$ where $p_1(E \cap E') = u$ and $|E \cap E'| = j$, and for $u < ((l-j)r)/(l-1) - 2 \cdot (j-1) \cdot \varepsilon$ we remove from V^{**} one vertex from each j-element subset $J \in [V^{**}]^j$ with $p_1(J) = u$.
We end up with a subset $V^{***} \subseteq V^{**}$, which does not contain any 2-cycles anymore and satisfies $|V^{***}| = (1 - o(1)) \cdot |V^*|$ by Corollary 6.5. Hence, we can apply Theorem 6.2 to our l-uniform hypergraph $\mathcal{G}^{***} = (V^{***}, [V^{***}]^l \cap \mathcal{E}_l^*)$, which has average degree $t^{l-1} \leq l \cdot |\mathcal{E}_l^*|/|V^{***}| \leq c_0 \cdot p^{l-1} \cdot t_0^{l-1}$ for some constant $c_0 > 0$, and we obtain in polynomial time an independent set of size at least

$$\Omega\left(\frac{|V^{***}|}{p \cdot t_0} \cdot (\log(p \cdot t_0))^{\frac{1}{l-1}}\right) = \Omega\left(m^{\frac{lr}{2(l-1)}} \cdot (\log m)^{\frac{1}{l-1}}\right)\ ,$$

which yields the desired lower bounds (18) and (19) by inserting the appropriate value of l, i.e. $l := k$ for k even, and $l := k-1$ for k odd.
Using the method of conditional probabilities in the same fashion as in [5], the running time is essentially dominated by the number $|\mathcal{E}_k| = O(m^{\lfloor k/2 \rfloor \cdot r})$ of k-element edges and, by (23), the numbers $p_{j,u}(V) = O(m^{(jr+u)/2})$ of u-element subsets $J \in [V]^j$ with $p_1(J) = u$ for $u \leq \lfloor (l-j)r/(l-1) \rfloor$ and, by (24), the numbers $s_{2,j}(\mathcal{G}_l, u) = O(m^{lr-(jr+u)/2})$ of pairs of edges $\{E, E'\} \in [\mathcal{E}_l]^2$ with $|E \cap E'| = j$ and $p_1(E \cap E') = u$ for $\lceil (l-j)r/(l-1) \rceil \leq u \leq\ \min\{jr, (l-j)r\}$. The dominating term here is $O(m^{lr-(jr+u)/2})$ for small values of u, r, which is at most $O(m^{r(k-3/2+1/(2k-2))} = O(m^{(k-4/3)r})$, and this, see Theorem 6.2, we have to compare with the term $N^3/t^{3-3\delta}$ where $N = \Theta(m^{\frac{lr}{2(l-1)}+\varepsilon})$ and $t_0 = \Theta(m^\varepsilon)$ (as otherwise, for $t_0 = o(m^\varepsilon)$, we can improve (18) and (19)), i.e. $N^3/t^{3-3\delta} = \Theta(m^{3r/2-\frac{3lr}{2(l-1)}+3\delta\varepsilon})$, thus the running time is at most $O(m^{(k-4/3)r})$. □

Remark: All calculations in the proof of Theorem 6.1 remain valid, if we pick in our arguments the columns at random according to a $(2l-2)$-wise independent distribution, compare [2]. For simulating a $(2l-2)$-wise independent distribution, it suffices to consider a sample space of size $O(m^{r(4l-4)})$, see [16], hence with these observations we also obtain polynomial running time.

7 Concluding Remarks

Some of the following possible applications have been stated already in [18] for the case $q = 2$.

Proposition 7.1 *Let A be an $l \times m$-matrix over $GF(q)$ with kr-wise independent columns, and let B be a (k,r)-matrix with dimension $m \times n$. Then the matrix-product $A \times B$ has k-wise independent columns.*

This observation can be used to extend the length of a linear code, but at the same time we reduce its minimum distance.
Also we can use sparse matrices, which are only approximately k-wise independent (k-wise ε-independent), for the construction of small probability spaces as follows, see also [3].

Definition 7.2 *The random variables $X_1, \ldots, X_m$ over $GF(q)$ are* k-wise ε-biased*, if for every choice of $\beta_1, \ldots, \beta_m \in GF(q)$, where at most k are nonzero but not all of them, and for each $c \in GF(q)$ it is*

$$\left|(q-1) \cdot \; Prob \; \left(\sum_{i=1}^{m} \beta_i \cdot X_i = c\right) - Prob \; \left(\sum_{i=1}^{m} \beta_i \cdot X_i \neq c\right)\right| \leq \varepsilon \; .$$

A sample space $S \subseteq (GF(q))^m$ is called k-wise ε-biased*, if the following holds: if a sequence $(x_1, \ldots, x_m)$ is chosen uniformly at random from S according to the uniform distribution, then $x_1, \ldots, x_m$ as random variables, are k-wise ε-biased.*
A sample space $S \subseteq (GF(q))^m$ is called (ε, k)-independent *(with respect to the uniform distribution in $(GF(q))^m$), if for each k positions $1 \leq i_1 < \ldots < i_k \leq n$ and for every sequence $\alpha = (\alpha_1, \ldots, \alpha_k) \in (GF(q))^k$ and any uniformly at random chosen sequence $X = (x_1, \ldots, x_m) \in S$, it is*

$$\left|Prob \; ((x_{i_1}, \ldots, x_{i_k}) = \alpha) - 1/q^k\right| \leq \varepsilon \; .$$

We remark that one can show along the lines in [7] that a k-wise ε-biased sample space $S \subseteq (GF(q))^m$ is also $(2 \cdot \varepsilon \cdot (1 - q^{-k})/q, k)$-independent.

Proposition 7.3 *Let $X = (X_1, \ldots, X_m)$ be a kr-wise ε-biased random vector over $GF(q)$, and let M be a (k,r)-matrix of dimension $m \times n$. Then the vector $Y = (Y_1, \ldots, Y_n) = X \times M$ is k-wise ε-biased over $GF(q)$.*

Proposition 7.4 *Let $S \subseteq (GF(q))^m$ be a kr-wise ε-biased sample space, and let M be a (k,r)-matrix of dimension $m \times n$ over $GF(q)$. Then the sample space $T = \{s \times M \mid s \in S\} \subseteq (GF(q))^n$ is k-wise ε-biased, thus also $(2 \cdot \varepsilon \cdot (1 - q^{-k})/q, k)$-independent.*

It would be interesting to find explicite constructions of (k,r)-matrices, the dimensions of which match at least the lower bounds proven in this paper. However, so far this proved to be hard already for the case $q = r = 2$ and larger values of k, i.e. $k \geq 12$, compare [17].

References

[1] M. Ajtai, J. Kómlos, J. Pintz, J. Spencer and E. Szemerédi, Extremal uncrowded hypergraphs, Journal of Combinatorial Theory A 32, 1982, 321-335.

[2] N. Alon, L. Babai and A. Itai, A fast and simple randomized parallel algorithm for the maximal independent set problem, Journal of Algorithms 7, 1986, 567-583.

[3] N. Alon, O. Goldreich, J. Håstad and R. Peralta, Simple constructions of almost k-wise independent random variables, Rand. Struct. & Algorithms 3, 1992, 289-304, and 4, 1993, 119-120.

[4] C. T. Benson, Minimal regular graphs of girth eight and twelve, Canadian Journal of Mathematics 18, 1966, 1091-1094.

[5] C. Bertram-Kretzberg, T. Hofmeister and H. Lefmann, Sparse 0-1-matrices and forbidden hypergraphs, Combinatorics, Probability and Computing 8, 1999, 417-427.

[6] C. Bertram-Kretzberg and H. Lefmann, The algorithmic aspects of uncrowded hypergraphs, SIAM Journal on Computing 29, 1999, 201-230.

[7] C. Bertram-Kretzberg and H. Lefmann, MOD_p-tests, almost independence and small probability spaces, Random Structures & Algorithms 16, 2000, 293-313.

[8] A. Bondy and M. Simonovits, Cycles of even length in graphs, Journal of Combinatorial Theory Ser. B 16, 1974, 97-105.

[9] W. G. Brown, On graphs that do not contain a Thomsen graph, Canadian Mathematical Bulletin 9, 1966, 281-289.

[10] R. Duke, H. Lefmann and V. Rödl, On uncrowded hypergraphs, Random Structures & Algorithms 6, 1995, 209-212.

[11] P. Erdös, On sequences of integers no one of which divides the product of two others and some related problems, Izvestiya Nauchno-Issl. Inst. Mat. i Meh. Tomsk 2, 1938, 74-82; see also Zentralblatt 20, 5.

[12] P. Erdös, A. Rényi and V. T. Sós, On a problem of graph theory, Studia Scientiarum Mathematicarum Hungarica 1, 1966, 213-235.

[13] A. Fundia, Derandomizing Chebychev's inequality to find independent sets in uncrowded hypergraphs, Random Structures & Algorithms 8, 1996, 131-147.

[14] P. Frankl and Z. Füredi, Union-free families of sets and equations over fields, Journal of Number Theory 23, 1986, 210-218.

[15] T. Hofmeister and H. Lefmann, A combinatorial design approach to MAXCUT, Random Structures & Algorithms 9, 1996, 163-175.

[16] H. Karloff and Y. Mansour, On construction of k-wise independent random variables, Proc. 26th Ann. ACM Symposium on Theory of Computing (STOC), 1994, 564-573.

[17] F. Lazebnik, V. A. Ustimenko and A. J. Woldar, A new series of dense graphs of high girth, Bulletin (New Series) of the American Mathematical Society 32, 1995, 73-79.

[18] H. Lefmann, P. Pudlák and P. Savický, On sparse parity-check matrices, Designs, Codes and Cryptography 12, 1997, 107-130.

[19] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman and V. Stemann, Practical loss-resilient codes, Proc. 29th Ann. ACM Symposium on Theory of Computing (STOC), 1997, 150-159.

[20] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, Analysis of low-density codes and improved designs using irregular graphs, Proc. 30th Ann. ACM Symposium on Theory of Computing (STOC), 1998, 249-258.

[21] A. Lubotzky, R. Phillips and P. Sarnak, Ramanujan graphs, Combinatorica 8, 1988, 261-277.

[22] G. A. Margulis, Explicit group theoretical construction of combinatorial schemes and their application to the design of expanders and concentrators, J. Probl. Inform. Transmission 24, 1988, 39-46.

[23] T. J. Richardson and R. L. Urbanke, Efficient encoding of low-density parity-check codes, IEEE Trans. Inform. Theory 47, 2001, 638-656.

[24] D. A. Spielman, Linear-time encodable and decodable error-correcting codes, Proc. 27th Ann. ACM Symposium on the Theory of Computing (STOC), 1995, 388-397.

[25] M. Sipser and D. A. Spielman, Expander codes, Proc. 35th Ann. Symposium on Foundations of Computer Science (FOCS), 1994, 566-576.

[26] R. Wenger, Extremal graphs with no C_4's, C_6's or C_{10}'s, Journal of Combinatorial Theory Ser. B 52, 1991, 113-116.