Sparse 0-1-Matrices and Forbidden Hypergraphs[*]

Claudia Bertram-Kretzberg / Thomas Hofmeister / Hanno Lefmann[†]
Lehrstuhl Informatik II, Universität Dortmund, D-44221 Dortmund, Germany
{bertram,hofmeist,lefmann}@Ls2.informatik.uni-dortmund.de

Abstract

We consider the problem of determining the maximal number $N(m,k,r)$ of columns of a 0-1-matrix with m rows and exactly r ones in each column such that every k columns are linearly independent over $\mathbb{Z}_2$. For fixed integers $k \geq 4$ and $r \geq 2$ where k is even and $gcd(k-1,r) = 1$, we shall prove the probabilistic lower bound $N(m,k,r) = \Omega(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}})$. This improves on earlier results from [13] by the factor $\Theta((\ln m)^{\frac{1}{k-1}})$ and extends results from [11] where the case $r = 2$ was considered. Moreover, we give a polynomial time algorithm achieving this new lower bound.

1 Introduction

We consider matrices M with entries 0 and 1 and *k-wise independent* columns, i.e., each k column vectors of M are linearly independent. A *(k, r)-matrix* has the property that the column vectors are k-wise independent and each column contains exactly r ones. Let $N(m,k,r)$ denote the maximal number of columns a (k,r)-matrix with m rows can have. Some special cases of this problem, e.g., estimates on $N(m,k,2)$, and similar questions have been studied in combinatorics and graph theory with respect to forbidden configurations, like cycles in graphs, and there is an extensive literature on this topic, cf. [10]. For example, considering incidence matrices of graphs shows that $N(m,4,2)$ is equal to the maximal number of edges in a graph on m vertices without cycles of length 3 or 4, since in graphs, the minimal dependent configurations correspond to cycles. First estimates on $N(m,k,r)$ for arbitrary values of m,k,r were given in [13]. In this paper we improve on the lower bounds for $N(m,k,r)$ given there. Notice that matrices with k-wise independent columns are *parity check matrices* for linear codes with minimal distance at least $k+1$, cf. [14]. Recall that for a linear code $C \subseteq \mathbb{Z}_2^n$, an $m \times n$ matrix M over $\mathbb{Z}_2 = \{0,1\}$ is a parity check matrix iff

$$\forall\, \underline{x} \in \mathbb{Z}_2^n : \quad \underline{x} \in C \Leftrightarrow M \cdot \underline{x} = \underline{0}\,.$$

If there is no restriction on the number of ones per column, then a lower bound on the maximum number of columns of 0⇔1-matrices with k-wise independent columns is given by the Gilbert-Varšamov bound. Besides of their importance in coding theory, one further application for parity check matrices lies in the area of derandomization. They can be used for constructing small sample spaces preserving some limited independence. Related here are recent results of Sipser

[*] This research was supported by the Deutsche Forschungsgemeinschaft as part of the Collaborative Research Center "Computational Intelligence" (SFB 531).

[†] Part of this research was done during the author's visit at the Institut für Theoretische Informatik, Medizinische Universität Lübeck.

and Spielman [18, 20] who, motivated by the PCP theorem, see [21], constructed asymptotically good codes using expander graphs. The parity check matrices of these codes are also sparse (each column and each row has only a constant number of nonzero entries). Our definition is different in that the number of ones per row is not bounded. A related approach was given by Alon, Bruck, Naor, Naor and Roth [3], who used Ramanujan graphs and Justesen codes for the approximation of probability distributions by small sample spaces.

In [13], the function $N(m, k, r)$ was defined in a slightly different way, namely, each column was allowed to have *at most* r ones, but this does not change the growth rate asymptotically.

The following lower bound on $N(m, k, r)$ was shown in [13] by a probabilistic argument: for fixed integers $k \geq 2$ and $r \geq 2$,

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

Moreover, [13] also provides an algorithm which computes (k, r)-matrices achieving this lower bound. Its running time is $O(m^{kr/2})$, which is polynomial if k and r are fixed.

Kohayakawa, Kreuter and Steger showed in [11] for r=2 the following: for fixed $k \geq 2$, it holds $N(m, 2k, 2) = \Omega(m^{2k/(2k-1)} \cdot (\ln m)^{1/(2k-1)})$. For this special case r=2, there are better constructions known from the work of Lubotzky, Phillips and Sarnak [15] and Margulis [16]. However, the intention in [11] was to study Turán-numbers in the random situation. The constructions from [15] and [16] use algebraic techniques, and yield the so-called Ramanujan graphs, which are graphs on m vertices with at least $\Omega\left(m^{(3k+5)/(3k+3)}\right)$ edges which do not contain any cycle of length smaller than $2k$+1. I.e., $N(m, 2k, 2) = \Omega\left(n^{(3k+5)/(3k+3)}\right)$. Recently Lazebnik, Ustimenko and Woldar [12] showed that $N(m, 2k, 2) = \Omega\left(m^{(3k-1+\beta)/(3k-3+\beta)}\right)$ with $\beta = 0$ if k is odd and $\beta = 1$ otherwise. Here, we do not focus on the case $r = 2$; we consider the case of arbitrary positive integers k and r.

As usual, let $gcd(k, l)$ be the greatest common divisor of positive integers k and l. We improve the general lower bound (1) by the factor $\Theta((\ln m)^{1/(k-1)})$ if k is even and if $gcd(k \Leftrightarrow 1, r) = 1$, i.e.,

$$N(m, k, r) = \Omega(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}}) . \tag{2}$$

The case of odd integers k can easily be reduced to the case of even integers.

It is surprising that our arguments do not work if $gcd(k \Leftrightarrow 1, r) > 1$. This possibly gives a hint that the lower bound (1) might be sometimes sharp in that case. As an example, consider that from the work of Frankl and Füredi [8] on union-free families of sets the following upper bound for $k = 4$ is known:

$$N(m, 4, r) = O(m^{\lceil 4r/3 \rceil / 2}) ,$$

while (1) yields the lower bound $N(m, 4, r) = \Omega(m^{2r/3})$. For $r \equiv 0 \bmod 3$, upper and lower bound match up to constant factors. On the other hand, for, say, $r \equiv 1 \bmod 3$, the upper bound is only of the order $O(m^{2r/3+1/3})$. Our lower bound here shows $N(m, 4, r) = \Omega(m^{2r/3} \cdot (\ln m)^{1/3})$.

The proof of our lower bound first transforms the problem into a problem on hypergraphs and then applies some probabilistic arguments. We also show that derandomization and recent results from [9] and [6] can be used to obtain a polynomial time algorithm.

Upper Bounds

For the existing bounds concerning the special case $r = 2$, we refer to [10]. We remark, that only for $k = 4, 6, 10$ matching lower and upper bounds exist, i.e., for these values of k it is known that $N(m, k, 2) = \Theta(m^{(k+2)/k})$.
For arbitrary values of k and r, the only known upper bounds on $N(m, k, r)$ are given in [13], namely: for fixed integers $k \geq 4$ and $r \geq 2$, where k is even, and $0 \leq s < r$ it holds that

$$\begin{aligned} N(m,k,r) &= O\left(N(m,k/2,2r \Leftrightarrow 2s) + m^s\right) \text{ , and} \\ N(m,k,r) &= O\left(\sqrt{m^s \cdot N(m,k/2,2r \Leftrightarrow 2s)} + m^s\right) \text{ .} \end{aligned}$$

In particular, one knows for k a power of 2:

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

2 Basic Definitions

A *hypergraph* $\mathcal{G} = (V, \mathcal{E})$ consists of a set V of vertices and a set $\mathcal{E}$ of edges where every edge $E \in \mathcal{E}$ satisfies $E \subseteq V$. A hypergraph $\mathcal{G} = (V, \mathcal{E})$ is called k*-uniform* if every edge $E \in \mathcal{E}$ has cardinality k. In our arguments 2*-cycles* are crucial, i.e., pairs $\{E, E'\}$ of distinct edges $E, E' \in \mathcal{E}$ which satisfy $|E \cap E'| \geq 2$. A 2-cycle $\{E, E'\}$ is called $(2, j)$*-cycle* if $|E \cap E'| = j$. The *average degree*, denoted by t^{k-1}, of a k-uniform hypergraph $\mathcal{G} = (V, \mathcal{E})$ is defined by $t^{k-1} = k \cdot |\mathcal{E}|/|V|$.
Given a subset $V_1 \subseteq V$ of the vertices, the *induced* subhypergraph on V_1 has edge set $\{E \mid E \in \mathcal{E} \text{ and } E \subseteq V_1\}$. An *independent set* is a subset V_1 of the vertices such that the induced subhypergraph on V_1 has no edge. The *independence number* $\alpha(\mathcal{G})$ of a hypergraph $\mathcal{G}$ is the size of a largest independent set.

3 Lower Bounds on N(m,k,r)

The trivial bounds are $N(m, 2, r) = \binom{m}{r}$ as no column can occur more than once and $N(m, k, 1) = m$. In [13] it was shown by a probabilistic argument that asymptotically it suffices to consider even dependencies only. Using also monotonicity we have for $k, r \geq 2$ that

$$N(m, 2k+1, r) \begin{cases} \leq & N(m, 2k, r) \\ \geq \; 1/2 \cdot & N(m, 2k, r)\,. \end{cases} \tag{3}$$

It is easy to see that every $0 \Leftrightarrow 1$-matrix M of dimension $m \times n$ contains a $m \times n/2$-submatrix M' such that no odd number of columns of M' adds to zero. Namely, for m even, partition at random the row positions $1, \ldots, m$ into disjoint sets S_1, S_2 of equal size $m/2$. Consider the random set M' of all column vectors of M which contain in S_1 an odd number of ones. The expected number of columns of M' is equal to $n/2$. Hence, there exists such a submatrix M' with at least this number of columns. No odd number of columns of M' add to zero.
We will show for arbitrary, but fixed values of k and r the following new lower bound on $N(m, k, r)$.

Theorem 3.1 *Let $k \geq 4$ and $r \geq 2$ be fixed integers where k is even. If $gcd(k \Leftrightarrow 1, r) = 1$, then*

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

From this theorem and inequality (3), we obtain:

Corollary 3.2 *Let $k \geq 5$ and $r \geq 2$ be fixed integers where k is odd. If $gcd(k\Leftrightarrow 2, r) = 1$, then*

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

In the proof of Theorem 3.1 we will use the following extension (see [7]) of a result of Ajtai, Komlós, Pintz, Spencer and Szemerédi [1]:

Theorem 3.3 *Let $k \geq 3$ be a fixed integer. Let $\mathcal{G} = (V, \mathcal{E})$ be a k-uniform hypergraph on n vertices with average degree at most t^{k-1}. If $\mathcal{G}$ does not contain any 2-cycles, then*

$$\alpha(\mathcal{G}) = \Omega\left(\frac{n}{t} \cdot (\ln t)^{\frac{1}{k-1}}\right) .$$

This result (as well as the result from [1]) has been applied various times in recent years, for example in [2], [17], [4], [19], cf. [6] for further references. Here, we obtain another application of Theorem 3.3, where we have to use fine-tuned arguments in order to keep the theorem applicable. Such a careful counting was not necessary in the case $r = 2$, where the minimal dependent configurations correspond to cycles in graphs.

We will use the following notation: Let C_m^r be the set of all $\binom{m}{r}$ possible $0\Leftrightarrow 1$-column vectors of length m with exactly r ones.

Proof. (of Theorem 3.1) Let $\oplus$ be the componentwise addition in $\mathbb{Z}_2^m$. We form a hypergraph $\mathcal{G} = (V, \mathcal{E}_3 \cup \ldots \cup \mathcal{E}_k)$ with vertex set $V = C_m^r$ and j-element edges $\{a_1, \ldots, a_j\} \in \mathcal{E}_j$ iff $a_1 \oplus \cdots \oplus a_j = 0$ where $j = 3, \ldots, r$. By this definition, an independent set in $\mathcal{G}$ corresponds to a set of column vectors which are k-wise linear independent, and it is our aim to find a large independent set.

First we give an upper bound on $|\mathcal{E}_j|$, $j = 3, \ldots, k$. If the column vectors $a_1, \ldots, a_j \in C_m^r$ satisfy $a_1 \oplus \ldots \oplus a_j = 0$, then each entry 1 in some column a_i, $i = 1, \ldots, j$, needs a matching entry 1 in the same row in another column $a_{i'}$. Hence, all ones occurring in $a_1, \ldots, a_j$ are contained in at most $\lfloor jr/2 \rfloor$ rows. Choosing these rows can be done in at most $\sum_{i=r+1}^{\lfloor jr/2 \rfloor} \binom{m}{i}$ many ways. Then the jr ones can be chosen in at most $\binom{\lfloor jr/2 \rfloor}{r}^j$ many ways. Thus, for $j = 3, \ldots, k$ we have for some positive constant $c_j = c_j(k, r)$ that

$$|\mathcal{E}_j| \leq c_j \cdot m^{\lfloor jr/2 \rfloor} . \qquad (4)$$

To obtain the improvement over (1) by a logarithmic factor, we will count the number $s_{2,j}(\mathcal{G}_k)$ of $(2, j)$-cycles, $j = 2, \ldots, k\Leftrightarrow 1$, in the k-uniform hypergraph $\mathcal{G}_k = (V, \mathcal{E}_k)$, i.e., we consider only the k-element edges of $\mathcal{G}$ and - with foresight - neglect for the moment the i-element edges for $i < k$. To do so, consider a fixed j-element set $J = \{a_1, \ldots, a_j\}$ of column vectors, viewed as an $m \times j$-matrix $M(J)$. Let $p(J)$ be the number of rows of $M(J)$ which contain at least one entry 1. Let $p_{odd}(J)$, ($p_{even}(J)$, respectively) be the number of rows of $M(J)$ containing an odd (even, respectively) nonzero number of ones. Let $a(J)$ be the number of $(k\Leftrightarrow j)$-element subsets $\{b_1, \ldots, b_{k-j}\}$ from $C_m^r \setminus \{a_1, \ldots, a_j\}$ such that $\{a_1, \ldots, a_j, b_1, \ldots, b_{k-j}\} \in \mathcal{E}_k$, i.e., $a_1 \oplus \cdots \oplus a_j \oplus b_1 \oplus \cdots \oplus b_{k-j} = 0$. If some row of $a_1 \oplus \cdots \oplus a_j$ contains a 1, then some column $b_{i'}$ also must contain a 1 in the same row. Hence, the $p_{odd}(J)$ rows containing an odd number of ones need matching ones among the column vectors $b_1, \ldots, b_{k-j}$, and we are free to choose, respecting dependence, the other at most

$(\lfloor (k \Leftrightarrow j)r \Leftrightarrow p_{odd}(J))/2\rfloor$ ones within the columns $b_1, \ldots, b_{k-j}$. Thus, for some positive constant $C_0 = C_0(k,r)$ we have

$$a(J) \leq C_0 \cdot m^{\lfloor \frac{(k-j)r - p_{odd}(J)}{2} \rfloor} .$$

We now count quite carefully the number of $(2,j)$-cycles in $\mathcal{G}_k$, $j = 2, \ldots, k \Leftrightarrow 1$. The reason is that the number of 2-cycles $\{E_1, E_2\}$, where $p(E_1 \cap E_2)$ is small, is too large for our purposes. We will handle these 2-cycles in a different way.

For $u = r+1, \ldots, jr$, let $s_{2,j:u}(\mathcal{G}_k)$ be the number of pairs $\{E_1, E_2\}$ of distinct edges with $E_1, E_2 \in \mathcal{E}_k$ and $|E_1 \cap E_2| = j$ and $p(E_1 \cap E_2) = u$. Clearly, we have $s_{2,j}(\mathcal{G}_k) = \sum_{u=r+1}^{jr} s_{2,j:u}(\mathcal{G}_k)$.

For a j-element subset $J \in [C_m^r]^j$, there are at most $\binom{a(J)}{2}$ pairs $\{J_1, J_2\}$ of sets such that $E_1 = J \cup J_1$ and $E_2 = J \cup J_2$ form a $(2,j)$-cycle in $\mathcal{G}_k$ with $E_1 \cap E_2 = J$. Hence, we have for some positive constant $C_1 = C_1(k,r)$:

$$\begin{aligned} s_{2,j:u}(\mathcal{G}_k) &\leq \sum_{J \in [C_m^r]^j : p(J) = u} (a(J))^2 \\ &\leq C_0^2 \cdot \sum_{J \in [C_m^r]^j : p(J) = u} m^{2 \cdot \lfloor \frac{(k-j)r - p_{odd}(J)}{2} \rfloor} \\ &\leq C_1 \cdot \sum_{J \in [C_m^r]^j : p(J) = u} m^{(k-j)r - p_{odd}(J)} . \end{aligned} \tag{5}$$

To evaluate (5) more precisely, observe that the rows counted by $p_{even}(J)$ contain at least two ones, and those counted by $p_{odd}(J)$ contain at least one 1. With $p_{even}(J) + p_{odd}(J) = p(J)$ we infer

$$2 \cdot p(J) \Leftrightarrow p_{odd}(J) = 2 \cdot p_{even}(J) + p_{odd}(J) \leq j \cdot r . \tag{6}$$

Let $p_{j,u}(V)$ denote the number of j-element subsets $J = \{a_1, \ldots, a_j\} \in [C_m^r]^j$ of column vectors with $p(J) = u$. Then, for $j < k$, $p_{j,u}(V)$ is bounded from above by $C_2' \cdot m^u$ for some positive constant $C_2' = C_2'(k,r)$. With (6), inequality (5) becomes:

$$\begin{aligned} s_{2,j:u}(\mathcal{G}_k) &\leq C_1 \cdot \sum_{J \in [C_m^r]^j : p(J) = u} m^{(k-j)r + jr - 2p(J)} \\ &= C_1 \cdot \sum_{J \in [C_m^r]^j : p(J) = u} m^{kr - 2u} \\ &\leq C_1' \cdot m^{kr - u} . \end{aligned} \tag{7}$$

We have to take care of the j-element subsets $J \in [C_r^m]^j$ of columns which, viewed as a matrix $M(J)$, have only a few rows containing at least one entry 1.

The average degree t^{k-1} of $\mathcal{G}_k = (V, \mathcal{E}_k)$ satisfies $t^{k-1} \leq k \cdot c_k \cdot m^{kr/2} / \binom{m}{r}$, hence, for some positive constant $C_k' = C_k'(k,r)$ we have

$$t \leq t_+ := C_k' \cdot m^{\frac{(k-2)r}{2(k-1)}} .$$

For applying Theorem 3.3, we have to find a subhypergraph without 2-cycles among the k-element edges and without i-element edges, $i = 3, \ldots, k \Leftrightarrow 1$. For obtaining such a subhypergraph, we choose in $\mathcal{G}$ at random vertices (column vectors) independently of each other with probability $p = t_+^{-1+\delta}$, where δ is a small positive constant which will be specified later. Vertices not chosen are removed from the hypergraph. We remark that the choice $p = t_+^{-1}$ would lead to the lower bound (1).

Let V_1 be the arising random subset of V and let $\mathcal{G}_1 = (V_1, \mathcal{E}_3^1 \cup \ldots \cup \mathcal{E}_k^1)$ be the induced random subhypergraph of $\mathcal{G}$. We consider the expected values $\mathbb{E}[\,\cdot\,]$ of the number of vertices, edges and j-element subsets $J \in [V_1]^j$ of columns (vertices) with $p(J) = s$ in $\mathcal{G}_1$ and also the number of 2-cycles among the k-element edges.

For the number of vertices we have for some positive constant $c = c(k, r)$ that

$$\mathbb{E}[|V_1|] = p \cdot \binom{m}{r} \geq c \cdot m^{\frac{kr}{2(k-1)} + \delta \cdot \frac{(k-2)r}{2(k-1)}} .$$

The expected number of i-element edges, $i = 3, \ldots, k$, satisfies by (4) for some positive constants $D_i = D_i(k, r)$ that

$$\begin{aligned} \mathbb{E}[|\mathcal{E}_i^1|] &\leq p^i \cdot c_i \cdot m^{\lfloor ir/2 \rfloor} && (8) \\ &\leq D_i \cdot m^{-\frac{(k-2)ir}{2(k-1)} + \lfloor ir/2 \rfloor + \delta \cdot \frac{(k-2)ir}{2(k-1)}} . \end{aligned}$$

For $j = 2, \ldots, k \Leftrightarrow 1$, let $p_{j,s}(V_1)$ count the number of j-element subsets $J \in [V_1]^j$ of column vectors with $p(J) = s$. Since $p_{j,s}(V) \leq C_2' \cdot m^s$, we have for some positive constant $C' = C'(k, r)$ that

$$\begin{aligned} \mathbb{E}[p_{j,s}(V_1)] &= p^j \cdot p_{j,s}(V) \\ &\leq C_2' \cdot p^j \cdot m^s && (9) \\ &\leq C' \cdot m^{s - \frac{j(k-2)r}{2(k-1)} + \delta \cdot \frac{(k-2)jr}{2(k-1)}} . \end{aligned}$$

For $j = 2, \ldots, k \Leftrightarrow 1$, the expected number of $(2, j)$-cycles $\{E_1, E_2\}$ in $\mathcal{G}_1^k = (V_1, \mathcal{E}_k^1)$ with $p(E_1 \cap E_2) = u$ satisfies by (7) for $u = r + 1, \ldots, jr$:

$$\begin{aligned} \mathbb{E}[s_{2,j:u}(\mathcal{G}_1^k)] &= p^{2k-j} \cdot s_{2,j:u}(\mathcal{G}_k) \\ &\leq C_1' \cdot p^{2k-j} \cdot m^{kr-u} && (10) \\ &\leq C_2 \cdot m^{\frac{(k-2)j+2k}{2(k-1)} \cdot r - u + \delta \cdot \frac{(k-2)(2k-j)r}{2(k-1)}} . \end{aligned}$$

By Chernoff's and Markov's inequality there exists an induced subhypergraph $\mathcal{G}_1 = (V_1, \mathcal{E}_3^1 \cup \ldots \cup \mathcal{E}_k^1)$ of $\mathcal{G}$, such that simultaneously for $i = 3, \ldots, k$, and $j = 2, \ldots, k \Leftrightarrow 1$ and $s = r + 1, \ldots, jr$ and $u = r + 1, \ldots, jr$ we have

$$\begin{aligned} |V_1| &\geq \frac{1}{2} \cdot c \cdot m^{\frac{kr}{2(k-1)} + \delta \cdot \frac{(k-2)r}{2(k-1)}} \\ |\mathcal{E}_i^1| &\leq E_i \cdot m^{-\frac{(k-2)ir}{2(k-1)} + \lfloor ir/2 \rfloor + \delta \cdot \frac{(k-2)ir}{2(k-1)}} \\ p_{j,s}(V_1) &\leq E' \cdot m^{s - \frac{j(k-2)r}{2(k-1)} + \delta \cdot \frac{(k-2)jr}{2(k-1)}} \\ s_{2,j:u}(\mathcal{G}_1^k) &\leq F \cdot m^{\frac{(k-2)j+2k}{2(k-1)} \cdot r - u + \delta \cdot \frac{(k-2)(2k-j)r}{2(k-1)}} \end{aligned}$$

for some appropriate positive constants E_i, E', E, F depending only on k and r.

We first show that some properties in $\mathcal{G}_1$ hold:

Claim 3.4 *Let $0 < \delta < 1/(k \Leftrightarrow 2)^2$. Then $|\mathcal{E}_i^1| = o(|V_1|)$ for $i = 3, \ldots, k \Leftrightarrow 1$.*

Proof. We have $|\mathcal{E}_i^1| = o(|V_1|)$ iff

$$\begin{aligned}
& p^i \cdot |\mathcal{E}_i| = o\,(p \cdot m^r) \\
\Leftrightarrow\quad & p^{i-1} \cdot m^{\lfloor ir/2 \rfloor - r} = o(1) \\
\Leftrightarrow\quad & t_+^{(-1+\delta)(i-1)} \cdot m^{\lfloor ir/2 \rfloor - r} = o(1) \\
\Leftrightarrow\quad & \Leftrightarrow \frac{(k \Leftrightarrow 2)ir + kr}{2(k \Leftrightarrow 1)} + \lfloor ir/2 \rfloor + \delta \cdot \frac{(i \Leftrightarrow 1)(k \Leftrightarrow 2)r}{2(k \Leftrightarrow 1)} < 0\ .
\end{aligned} \qquad (11)$$

It suffices to verify (11) for the case ir even. Then (11) becomes

$$\frac{\Leftrightarrow kr + ir}{2(k \Leftrightarrow 1)} + \delta \cdot \frac{(i \Leftrightarrow 1)(k \Leftrightarrow 2)r}{2(k \Leftrightarrow 1)} < 0$$

which holds for $0 < \delta < 1/(k \Leftrightarrow 2)^2$, since $3 \leq i \leq k \Leftrightarrow 1$. □

For $j = 2, \ldots, k \Leftrightarrow 1$, set

$$T(j) = \left\lfloor \frac{(j \Leftrightarrow 1)kr}{2(k \Leftrightarrow 1)} + \delta \cdot \frac{(j \Leftrightarrow 1)(k \Leftrightarrow 2)r}{2(k \Leftrightarrow 1)} \right\rfloor\ . \qquad (12)$$

Now, $gcd(k, k \Leftrightarrow 1) = 1$ and $gcd(j \Leftrightarrow 1, k \Leftrightarrow 1) < k$ for $j = 2, \ldots, k \Leftrightarrow 1$. As we assumed that $gcd(k \Leftrightarrow 1, r) = 1$, we have that $\frac{(j-1)kr}{2(k-1)}$ is not an integer for $j = 2, \ldots, k \Leftrightarrow 1$. Hence, there is a range of $\delta > 0$ such that $\frac{(j-1)kr}{2(k-1)} \pm \delta \cdot \frac{(j-1)(k-2)r}{2(k-1)}$ is for no $j = 2, \ldots, k \Leftrightarrow 1$ integer-valued.

Claim 3.5 *Let $0 < \delta < 1/(k \Leftrightarrow 2)^2 r$. For $j = 2, \ldots, k \Leftrightarrow 1$ and $s < jr \Leftrightarrow T(j)$ it holds that*

$$p_{j,s}(V_1) = o(|V_1|)\ .$$

Proof. To see this, notice that

$$\begin{aligned}
& p^j \cdot p_{j,s}(V) = o\,(p \cdot m^r) \\
\Leftrightarrow\quad & p^{j-1} \cdot m^{s-r} = o(1) \\
\Leftrightarrow\quad & m^{s - \frac{(k-2)j+k}{2(k-1)} \cdot r + \delta \cdot \frac{(j-1)(k-2)r}{2(k-1)}} = o(1) \\
\Leftrightarrow\quad & jr \Leftrightarrow s > \frac{(j \Leftrightarrow 1)k}{2(k \Leftrightarrow 1)} \cdot r + \delta \cdot \frac{(j \Leftrightarrow 1)(k \Leftrightarrow 2)r}{2(k \Leftrightarrow 1)} \\
\Leftrightarrow\quad & jr \Leftrightarrow s > T(j)
\end{aligned}$$

by definition of $T(j)$, cf. (12), since $\delta < 1/(k \Leftrightarrow 2)^2 r$ and $\frac{(j-1)kr}{2(k-1)}$ is not an integer. □

Now we want to delete all j-element subsets $J \in [V_1]^j$ of column vectors with small values of $p(J)$ and all edges $E \in \mathcal{E}_3^1 \cup \ldots \cup \mathcal{E}_{k-1}^1$ in the hypergraph $\mathcal{G}_1$. We fix some $\delta > 0$ with $\delta < 1/(k \Leftrightarrow 2)(2k \Leftrightarrow 3)r$. In the subhypergraph $\mathcal{G}_1$, we remove one vertex (column) from each j-element subset J $p(J) < jr \Leftrightarrow T(j)$ and thus destroy all these j-element sets of columns. Moreover, for $i = 3, \ldots, k \Leftrightarrow 1$ we also omit one vertex from each i-element edge $E \in \mathcal{E}_i^1$, therefore, destroying all i-element edges. We are left with at least $(c/2 \Leftrightarrow o(1)) \cdot m^{\frac{kr}{2(k-1)}+\delta\cdot\frac{(k-2)r}{2(k-1)}}$ column vectors (vertices), and we obtain a k-uniform induced subhypergraph $\mathcal{G}_2 = (V_2, \mathcal{E}_k^1 \cap [V_2]^k) = (V_2, \mathcal{E}^2)$ with

$$\begin{aligned} |V_2| &\geq (c/2 \Leftrightarrow o(1)) \cdot m^{\frac{kr}{2(k-1)}+\delta\cdot\frac{(k-2)r}{2(k-1)}} \\ |\mathcal{E}^2| &\leq E_k \cdot m^{\frac{kr}{2(k-1)}+\delta\cdot\frac{(k-2)kr}{2(k-1)}} , \end{aligned}$$

and each j-element subset $J \in [V_2]^j$ of columns satisfies $p(J) \geq jr \Leftrightarrow T(j)$. Observe that now we have $s_{2,j:u}(\mathcal{G}_2) = 0$ for $u < jr \Leftrightarrow T(j)$. Recall that $p_{j,u}(V) = |\{J \in [C_m^r]^j \mid p(J) = u\}| \leq C_2' \cdot m^u$. Then, for $j = 2, \ldots, k \Leftrightarrow 1$ and by (10) the number of $(2,j)$-cycles in $\mathcal{G}_2$ satisfies for some positive constants $C_j^* = C_j^*(k,r)$:

$$\begin{aligned} s_{2,j}(\mathcal{G}_2) &\leq C \cdot \sum_{u=jr-T(j)}^{jr} s_{2,j:u}(\mathcal{G}_1^k) \\ &\leq C'' \cdot p^{2k-j} \cdot \sum_{u=jr-T(j)}^{jr} m^{kr-u} \\ &\leq C_j^* \cdot p^{2k-j} \cdot m^{kr-jr+T(j)} . \end{aligned} \qquad (13)$$

Claim 3.6 *Let* $0 < \delta < 1/(k \Leftrightarrow 2)(2k \Leftrightarrow 3)r$. *For* $j = 2, \ldots, k \Leftrightarrow 1$ *it holds that*

$$s_{2,j}(\mathcal{G}_2) = o(|V_2|) .$$

Proof. By (13), this holds iff

$$\begin{aligned} & p^{2k-j} \cdot m^{kr-jr+T(j)} = o\left(p \cdot m^r\right) \\ \Leftrightarrow\quad & m^{(\delta-1)\frac{(k-2)r(2k-j-1)}{2(k-1)}+r(k-j-1)+T(j)} = o(1) \\ \Leftrightarrow\quad & T(j) < \frac{(j \Leftrightarrow 1)kr}{2(k \Leftrightarrow 1)} \Leftrightarrow \delta \cdot \frac{(k \Leftrightarrow 2)(2k \Leftrightarrow j \Leftrightarrow 1)r}{2(k \Leftrightarrow 1)} \end{aligned}$$

which holds by (12) since $\delta < 1/(k \Leftrightarrow 2)(2k \Leftrightarrow 3)r$ and $\frac{(j-1)kr}{2(k-1)}$ is not an integer. □

Then, for $j = 2, \ldots, k \Leftrightarrow 1$ from each $(2,j)$-cycle in $\mathcal{G}_2 = (V_2, \mathcal{E}^2)$ we omit one vertex, hence destroying all 2-cycles in $\mathcal{G}_2$ and we obtain a k-uniform subhypergraph $\mathcal{G}_3 = (V_3, \mathcal{E}^3)$ of $\mathcal{G}_2$ with $|V_3| \geq (c/2 \Leftrightarrow o(1)) \cdot m^{\frac{kr}{2(k-1)}+\delta\cdot\frac{(k-2)r}{2(k-1)}}$ and $|\mathcal{E}^3| \leq E_k \cdot m^{\frac{kr}{2(k-1)}+\delta\cdot\frac{(k-2)kr}{2(k-1)}}$. Notice the following: If we had not destroyed the j-element subsets J of columns (vertices) with $p(J) < jr \Leftrightarrow T(j)$, then we could not have argued that the number of 2-cycles in $\mathcal{G}_2$ is much less than the number of vertices, and the argument would not work that way.

Now we apply Theorem 3.3 and we obtain an independent set in $\mathcal{G}_2$, and hence in $\mathcal{G}$, of size

$$\Omega\left(\frac{m^{\frac{kr}{2(k-1)}+\delta\cdot\frac{(k-2)r}{2(k-1)}}}{m^{\delta\cdot\frac{(k-2)r}{2(k-1)}}} \cdot (\ln m^{\frac{\delta(k-2)r}{2(k-1)}})^{\frac{1}{k-1}}\right) = \Omega\left(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}}\right) .$$

The vertices of this independent set yield the desired $0 \Leftrightarrow 1$-matrix with k-wise independent columns.

□

4 An Algorithm

So far we proved only an existence result. Here we show that this existence proof can be derandomized which yields (k, r)-matrices which satisfy the lower bound (2). We use the method of conditional probabilities, cf. [5], for proving the following:

Theorem 4.1 *Let $k \geq 4$ and $r \geq 2$ be fixed integers where k is even. If $gcd(k \Leftrightarrow 1, r) = 1$, then one can find in polynomial time a $0 \Leftrightarrow 1$-parity check matrix with m rows and $\Omega\left(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}}\right)$ columns such that each column contains exactly r ones and each set of k columns is linearly independent.*

The corresponding result for odd k (like in Corollary 3.2) can be obtained analogously. Namely, for a fixed partition of the row indices into sets S_1, S_2 of equal size, consider all column vectors with exactly r ones, which have an odd number of ones in S_1. No odd number of these column vectors add to zero. Then, we apply essentially Theorem 4.1.
For proving Theorem 4.1, we will use the following result from [6], which extends former investigations of Fundia [9] and which is the algorithmic analogue of Theorem 3.3.

Theorem 4.2 [6] *Let $k \geq 3$ be a fixed integer. Let $\mathcal{G} = (V, \mathcal{E})$ be a k-uniform hypergraph on n vertices with average degree at most t^{k-1} where $\mathcal{G}$ does not contain any 2-cycles. Then one can find for any fixed $\delta > 0$ in time $O(|V| + |\mathcal{E}| + n^3/t^{3-\delta})$ an independent set of size at least $c(k, \delta) \cdot n/t \cdot (\ln t)^{1/(k-1)}$.*

Proof. (of Theorem 4.1) Let $C_m^r = \{a_1, \ldots, a_M\}$. Associate with each column a_i a weight $p_i \in [0, 1]$. For $i = 3, \ldots, k$, let $\mathcal{E}_i$ be the set of all i-element subsets $\{a_1, \ldots, a_i\} \subseteq [C_m^r]^i$ such that $a_1 \oplus \ldots \oplus a_i = 0$. Moreover, for $j = 2, \ldots, k \Leftrightarrow 1$ and $s = r + 1, \ldots, jr$, let $P(j; s)$ be the set of all j-element subsets $J = \{a_1, \ldots, a_j\} \subseteq [C_m^r]^j$ with $p(J) = s$. Let $T(j; u)$ be the set of all pairs $\{E_1, E_2\} \in [\mathcal{E}_k]^2$, with $|E_1 \cap E_2| = j$ and $p(E_1 \cap E_2) = u$, where $j = 2, \ldots, k \Leftrightarrow 1$ and $u = r + 1, \ldots, jr$. We define the following potential:

$$\begin{aligned} V(p_1, \ldots, p_M) \quad := \quad & 3^{pM/3} \cdot \prod_{i=1}^{M} (1 \Leftrightarrow \frac{2}{3} \cdot p_i) + \sum_{i=3}^{k} \frac{\sum_{E \in \mathcal{E}_i} \prod_{a_j \in E} p_j}{6k \cdot c_i \cdot m^{\lfloor ir/2 \rfloor} \cdot p^i} + \\ & + \sum_{j=2}^{k-1} \sum_{s=r+1}^{jr-T(j)-1} \frac{\sum_{J \in P(j;s)} \prod_{a_i \in J} p_i}{6k^2 \cdot r \cdot C_2' \cdot m^s \cdot p^j} + \\ & + \sum_{j=2}^{k-1} \sum_{u=jr-T(j)}^{jr} \frac{\sum_{\{E_1, E_2\} \in T(j;u)} \prod_{a_i \in E_1 \cup E_2} p_i}{6k^2 \cdot r \cdot C_1' \cdot p^{2k-j} \cdot m^{kr-u}} \end{aligned}$$

where $T(j)$ is defined as in (12). Notice that in the three sum terms the denominators contain upper bounds on the sizes of the objects which we want to control, cf. (8), (9) and (10).
We claim that for $pM/3 \geq 1$ we have $V(p, \ldots, p) < 1$. Namely, using $1 \Leftrightarrow x \leq e^{-x}$, we infer

$$\begin{aligned} V(p, \ldots, p) \leq \quad & 3^{pM/3} \cdot \left(1 \Leftrightarrow \frac{2}{3} \cdot p\right)^M + \\ & + \sum_{i=3}^{k} \frac{1}{6k} + \sum_{j=2}^{k-1} \sum_{s=r+1}^{jr-T(j)-1} \frac{1}{6k^2 r} + \end{aligned}$$

$$\begin{aligned} &+\sum_{j=2}^{k-1} \sum_{u=jr-T(j)}^{jr} \frac{1}{6k^2 r} \\ \le\ & \left(\frac{3}{e^2}\right)^{pM/3} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} < 1\ . \end{aligned}$$

We fix the value of p to $p := m^{-\frac{(k-2)r}{2(k-1)}+\varepsilon}$, where $0 < \varepsilon < 1/(k \Leftrightarrow 2)(2k \Leftrightarrow 3)r$. Then, $pM/3 \ge 1$ for m large enough, and we have $V(p,\ldots,p) < 1$. In each step $i = 1,\ldots,M$ the algorithm chooses $p_i = 0$ or $p_i = 1$ to minimize the value of the current potential. Finally, $p_1,\ldots,p_M \in \{0,1\}$. As the potential $V(p_1,\ldots,p_M)$ is linear in each p_i, we have $V(p_1,\ldots,p_M) \le V(p,\ldots,p) < 1$. Let $S = \{a_i \in C_m^r \mid p_i = 1\}$ be the set of the chosen column vectors.
By definition of the potential V,

$$|S| \ge p/3 \cdot M = c_1 \cdot m^{\frac{kr}{2(k-1)}+\varepsilon}\ .$$

For $i = 3,\ldots,k$ the number of i-element subsets $\{a_1,\ldots,a_i\} \subseteq [S]^i \cap \mathcal{E}_i$, i.e., $a_1 \oplus \ldots \oplus a_i = 0$, is at most

$$\begin{aligned} 6k \cdot c_i \cdot m^{\lfloor \frac{ir}{2} \rfloor} \cdot p^i &\le c_2 \cdot m^{\frac{ir}{2(k-1)}+i\varepsilon} \\ &= o\left(m^{\frac{kr}{2(k-1)}+\varepsilon}\right)\ , \end{aligned}$$

as otherwise $V(p_1,\ldots,p_M) > 1$. For $j = 2,\ldots,k \Leftrightarrow 1$ and $s = r+1,\ldots,jr \Leftrightarrow T(j) \Leftrightarrow 1$, the number of all j-element subsets $J = \{a_1,\ldots,a_j\} \subseteq [S]^j$ with $p(J) = s$ is at most

$$\begin{aligned} 6k^2 \cdot r \cdot C_2' \cdot m^s \cdot p^j &\le c_3 \cdot m^{s-\frac{(k-2)jr}{2(k-1)}+j\varepsilon} \\ &= o\left(m^{\frac{kr}{2(k-1)}+\varepsilon}\right)\ . \end{aligned}$$

Moreover, for $j = 2,\ldots,k \Leftrightarrow 1$ and $u = jr \Leftrightarrow T(j),\ldots,jr$, the number of all pairs $\{E_1,E_2\}$ with $E_1, E_2 \in [S]^k \cap \mathcal{E}_k$, with $|E_1 \cap E_2| = j$ and $p(E_1 \cap E_2) = u$ is at most

$$\begin{aligned} 6k^2 \cdot r \cdot C'' \cdot p^{2k-j} \cdot m^{kr-u} &\le c_4 \cdot m^{kr-u-\frac{(k-2)(2k-j)r}{2(k-1)}+(2k-j)\varepsilon} \\ &= o\left(m^{\frac{kr}{2(k-1)}+\varepsilon}\right)\ . \end{aligned}$$

Now, for $i = 3,\ldots,k$, from each i-element subset $\{a_1,\ldots,a_i\} \subseteq [S]^i \cap \mathcal{E}_i$ we omit one column. Also, for $j = 2,\ldots,k \Leftrightarrow 1$ and $s = r+1,\ldots,jr \Leftrightarrow T(j) \Leftrightarrow 1$ and each j-element subset $J = \{a_1,\ldots,a_j\} \subseteq [S]^j$ with $p(J) = s$ we omit one column. Moreover, for $j = 2,\ldots,k \Leftrightarrow 1$ and $u = jr \Leftrightarrow T(j),\ldots,jr$, from each pair $E_1 = \{a_1,\ldots,a_k\}, E_2 = \{b_1,\ldots,b_k\} \in [S]^k \cap \mathcal{E}_k$ of k-element subsets satisfying $|E_1 \cap E_2| = j$ and $p(E_1 \cap E_2) = u$ we omit one column. Thus, we remove the few disturbing groups by omitting one column from each such group. We obtain a subset $S_1 \subseteq S$ of column vectors with

$$|S_1| \ge |S|/2 = c_1/2 \cdot m^{\frac{kr}{2(k-1)}+\varepsilon}\ ,$$

and S_1 contains at most $c_2 \cdot m^{\frac{kr}{2(k-1)}+k\varepsilon}$ dependencies of sets of k columns, i.e., the corresponding hypergraph has at least $c_1/2 \cdot m^{\frac{kr}{2(k-1)}+\varepsilon}$ vertices, is k-uniform with at most $c_2 \cdot m^{\frac{kr}{2(k-1)}+k\varepsilon}$ edges, and contains no 2-cycles anymore.

Now we apply the algorithm from Theorem 4.2 and we obtain in polynomial time an independent set of size at least

$$\Omega\left(\frac{m^{\frac{kr}{2(k-1)}+\varepsilon}}{m^{\varepsilon}}\cdot(\ln m^{\varepsilon})^{\frac{1}{k-1}}\right)=\Omega\left(m^{\frac{kr}{2(k-1)}}\cdot(\ln m)^{\frac{1}{k-1}}\right)\ .$$

The vertices of this independent set yield the columns of the desired matrix with k-wise independent columns. The running time is mainly given by the number of groups we have to control and is of the order $O(m^{kr})$. □

5 Concluding Remarks

It might be interesting to investigate whether analogues of the results of Lubotzky, Phillips and Sarnak [15] and Margulis [16] might be applied here, to obtain better bounds on $N(m,k,r)$. This might also lead to some (more) explicit constructions. It might also be possible that using algebraic constructions like in the case $k=4$ and $r=2$, i.e., graphs without triangles and cycles of length four, where one obtains $N(m,4,2)=\Omega(m^{3/2})$, lead to better lower bounds for $N(m,k,r)$. However, the situation seems to be a little tricky, as the improvement by a logarithmic factor is (with these techniques) only possible if certain divisibility conditions hold. On the other hand, the results of Frankl and Füredi [8] and those in [13] show that the lower bounds (1) sometimes match the upper bounds.

References

[1] M. Ajtai, J. Komlós, J. Pintz, J. Spencer and E. Szemerédi, *Uncrowded Hypergraphs,* Journal of Combinatorial Theory Ser. A 32, 1982, 321-335.

[2] M. Ajtai, J. Komlós and E. Szemerédi, *A Dense Infinite Sidon Sequence,* European Journal of Combinatorics 2, 1981, 2-11.

[3] N. Alon, J. Bruck, J. Naor, M. Naor and R. Roth, *Construction of Asymptotically Good Low-Rate Error Correcting Codes Through Pseudo-Random Graphs,* IEEE Transactions on Information Theory 38, 1992, 509-516.

[4] N. Alon, H. Lefmann and V. Rödl, *On an Anti-Ramsey Type Result,* Colloquia Mathematica Societatis János Bolyai, 60. Sets, Graphs and Numbers, 1991, 9-22.

[5] N. Alon and J. Spencer, *The Probabilistic Method,* Wiley & Sons, New York, 1992.

[6] C. Bertram-Kretzberg and H. Lefmann, *The Algorithmic Aspects of Uncrowded Hypergraphs,* Proc. 8th ACM-SIAM Symposium on Discrete Algorithms SODA, 1997, 296-304.

[7] R. A. Duke, H. Lefmann and V. Rödl, *On Uncrowded Hypergraphs,* Random Structures & Algorithms 6, 1995, 209-212.

[8] P. Frankl and Z. Füredi, *Union-Free Families of Sets and Equations over Fields,* Journal of Number Theory 23, 1986, 210-218.

[9] A. Fundia, *Derandomizing Chebychev's Inequality to find Independent Sets in Uncrowded Hypergraphs,* Random Structures & Algorithms 8, 1996, 131-147.

[10] Z. Füredi, *Turán Type Problems,* Surveys in Combinatorics, London Math. Soc., LNS 166, Cambridge University Press, (1991), 253-300.

[11] Y. Kohayakawa, B. Kreuter and A. Steger, *An Extremal Problem for Random Graphs and the Number of Graphs with Large Even-Girth*, preprint, 1995.

[12] 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.

[13] H. Lefmann, P. Pudlák and P. Savický, *On Sparse Parity Check Matrices,* Designs, Codes and Cryptography 12, 1997, 107-130.

[14] R. Lidl and H. Niederreiter, *Introduction to Finite Fields and their Applications,* Cambridge University Press, 1994.

[15] A. Lubotzky, R. Phillips and P. Sarnak, *Ramanujan Graphs,* Combinatorica 8, 1988, 261-277.

[16] 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.

[17] P. Pudlak and V. Rödl, *Modified Ranks of Tensors and the Size of Circuits,* Proc. 25th ACM Symposium on the Theory of Computing STOC, 1993, 523-531.

[18] M. Sipser and D. A. Spielman, *Expander Codes,* Proc. 35th IEEE Symposium on Foundations of Computer Science FOCS, 1994, 566-576.

[19] V. Rödl and E. Siňajová, *Note on Independent Sets in Steiner Systems,* Random Structures & Algorithms 5, 1994, 183-190.

[20] D. A. Spielman, *Linear-time Encodable and Decodable Error-correcting Codes,* Proc. 27th ACM Symposium on the Theory of Computing STOC, 1995, 388-397.

[21] D. A. Spielman, *Computationally Efficient Error-Correcting Codes and Holographic Proofs,* PhD thesis, MIT, 1995.