Europ. J. Combinatorics (1983) **4**, 215–220

On Subsets with Cardinalities of Intersections Divisible by a Fixed Integer

P. FRANKL AND A. M. ODLYZKO

If $m(n, l)$ denotes the maximum number of subsets of an n-element set such that the intersection of any two of them has cardinality divisible by l, then a trivial construction shows that

$$m(n, l) \geq 2^{[n/l]}.$$

For $l = 2$, this was known to be essentially best possible. For $l \geq 3$, we show by construction that $m(n, l)2^{-[n/l]}$ grows exponentially in n, and we provide upper bounds.

1. INTRODUCTION

We consider the problem of estimating $m(n, l)$ which is the maximum number of subsets $A_1, \ldots, A_m$ of an n-element set such that

$$|A_i \cap A_j| \equiv 0 \pmod{l}, \quad 1 \leq i < j \leq m.$$

Suppose $B_1, \ldots, B_{[n/l]}$ are pairwise disjoint l-element subsets of $\{1, 2, \ldots, n\}$. Then the sets formed by the union of any collection of the B_i have the desired property, and so

$$m(n, l) \geq 2^{[n/l]}. \tag{1.1}$$

P. Erdös conjectured that this is essentially best possible for $l = 2$. This was proved by Berlekamp [1] and Graver [5] by different methods. They showed that if $n = 8$ or $n \geq 10$, then

$$m(n, 2) = 2^{n/2}, \quad \text{if } n \text{ is even},$$
$$m(n, 2) = 2^{(n-1)/2} + 1, \quad \text{if } n \text{ is odd}.$$

It turns out that for $l > 2$, the natural generalization, namely that

$$m(n, l) = O(2^{n/l}), \tag{1.2}$$

is false. We prove the following bounds for $m(n, l)$.

THEOREM 1. *If a Hadamard matrix of order $4l$ exists (which is known to be true for $1 \leq l \leq 66$, and is conjectured to be true for all l), then*

$$m(n, l) \geq (8l)^{[n/(4l)]}. \tag{1.3}$$

In any event, for $l \geq 67$

$$m(n, l) \geq 256^{[n/(4l)]} = 2^{8[n/(4l)]}. \tag{1.4}$$

THEOREM 2. *If $\Omega(l)$ is the number of prime-power divisors of l, then*

$$m(n, l) \leq 2^{[n/2]} + \Omega(l)n, \tag{1.5}$$

and

$$m(n, l) \leq 2 \sum_{i=0}^{[n/(2l)]} \binom{n}{i} + \Omega(l)n. \tag{1.6}$$

0195-6698/83/030215+06 \$02.00/0

For $l=2,3,4$ the bound (1.5) is better than (1.6), but for larger values of l, (1.6) is sharper, and it is markedly so for large l. It is not hard to show that

$$c(l)=\lim_{n\to\infty} m(n,l)^{1/n}$$

exists, and the two theorems imply that

$$c(l)\geqslant \exp\left(\frac{1}{4l}\ln 8l\right) \tag{1.7}$$

if a Hadamard matrix of order $4l$ exists, and that

$$c(l)\leqslant \min(2^{1/2}, \exp(h((2l)^{-1}))), \tag{1.8}$$

where $h(x)=-x\ln x-(1-x)\ln(1-x)$ is the entropy function. For $l\to\infty$, (1.7) gives

$$c(l)\geqslant 1+\frac{1+o(1)}{4l}\ln l,$$

while (1.8) yields

$$c(l)\leqslant 1+\frac{1+o(1)}{2l}\ln l.$$

It would be very interesting to know whether one has equality in (1.7). Some other open questions are discussed in Section 4.

2. Constructions

Let $m_1(n,l)$ denote the maximum size of a collection of subsets $A_1,\dots,A_m$ of $\{1,\dots,m\}$ such that

$$|A_i\cap A_j|\equiv 0 \pmod{l}, \quad 1\leqslant i,j\leqslant m,$$

(i.e. we omit the condition $i\neq j$).

Lemma 1. *We have*

$$m_1(n,l)\leqslant m(n,l)\leqslant m_1(n,l)+\Omega(l)n,$$

where $\Omega(l)$ denotes the total number of prime factors of 1, *multiple factors counted according to their multiplicity.*

Proof. The first inequality of the lemma is trivial. To prove the second suppose that $|A_i|\not\equiv 0 \pmod{l}$ for $1\leqslant i\leqslant k\leqslant m$ and let $\boldsymbol{B}=[b_{ij}]$ be the incidence matrix of the collection $A_1,\dots,A_k$; i.e.,

$$b_{ij}=\begin{cases}1, & \text{if } i\in A_j,\\ 0, & \text{if } i\notin A_j,\end{cases}$$

where $|A_i\cap A_j|\equiv 0 \pmod{l}$ for $1\leqslant i<j\leqslant k$. To prove the lemma, it is sufficient to prove $k\leqslant \Omega(l)n$. $\boldsymbol{B}$ has n rows, so $\operatorname{rank}(\boldsymbol{B})\leqslant n$. Next set

$$\boldsymbol{C}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}.$$

If $\boldsymbol{C}=[c_{ij}]$, then

$$c_{ij}=|A_i\cap A_j|.$$

Let $l=\prod_{i=1}^r p_i^{\alpha_i}$, where the p_i are distinct primes. As $|A_j|\not\equiv 0 \pmod{l}$, $|A_j|\not\equiv 0 \pmod{p_i^{\alpha_i}}$ for some i, $1\leqslant i\leqslant r$.

For a fixed i, $1\leqslant i\leqslant r$, and for a fixed β, $1\leqslant\beta\leqslant\alpha_i$, let $A_{i_1},\ldots,A_{i_s}$ be the sets for which

$$|A_{i_j}|\equiv 0 \pmod{p_i^{\beta-1}},$$
$$|A_{i_j}|\not\equiv 0 \pmod{p_i^{\beta}}.$$

The submatrix of $\boldsymbol{C}$ formed by taking rows and columns numbered $i_1,\ldots,i_s$ becomes, when divided by $p_i^{\beta-1}$, a diagonal matrix $\bmod p_i$ with non-zero entries on the diagonal. This implies $s\leqslant n$, since $\operatorname{rank}(\boldsymbol{C})\leqslant\operatorname{rank}(\boldsymbol{B})\leqslant n$. Summing over i and β, we obtain the claim of the lemma.

LEMMA 2. *For* $1\leqslant r\leqslant n$,

$$m_1(n,l)\geqslant m_1(r,l)m_1(n-r,l).$$

PROOF. Suppose $A_1,\ldots,A_s$ are subsets of $\{1,2,\ldots,r\}$ such that

$$|A_i\cap A_j|\equiv 0 \pmod{l},\quad 1\leqslant i,j\leqslant s,$$

and $B_1,\ldots,B_t$ are subsets of $\{r+1,\ldots,n\}$ such that

$$|B_i\cap B_j|\equiv 0 \pmod{l},\quad 1\leqslant i,j\leqslant t.$$

Define

$$C_{i,j}=A_i\cup B_j,\quad 1\leqslant i\leqslant s,\quad 1\leqslant j\leqslant t.$$

Then the $C_{i,j}$ are all distinct, and

$$|C_{i,j}\cap C_{p,q}|=|A_i\cap A_p|+|B_j\cap B_q|\equiv 0 \pmod{l},$$

which proves the lemma.

We now proceed to our constructions of large collections of subsets $A_1,\ldots,A_m$ of $\{1,\ldots,n\}$ such that

$$|A_i\cap A_j|\equiv 0 \pmod{l},\quad 1\leqslant i,j\leqslant m.$$

These constructions are based on Hadamard matrices. Recall that a Hadamard matrix $\boldsymbol{M}$ of order $4t$ is a $4t$ by $4t$ matrix with ± 1 entries such that the scalar product of any two distinct rows is zero. One can always assume that the first row is of the form $(1,1,\ldots,1)$.

It is conjectured that Hadamard matrices of order $4t$ exist for every $t\in\mathbb{Z}^+$ and this is known to be true for $t\leqslant 66$, as well as for several infinite families of values of t, including $t\equiv 3 \pmod 4$, t a prime power—cf. [4].

Assume first that a Hadamard matrix $\boldsymbol{M}=[m_{ij}]$ of order $4l$ exists. Define subsets $S_1,\ldots,S_{4l}$, $T_1,\ldots,T_{4l}$ of $\{1,\ldots,4l\}$ by

$$S_i=\{j\colon 1\leqslant j\leqslant 4l,\ m_{ij}=1\}$$
$$T_i=\{j\colon 1\leqslant j\leqslant 4l,\ m_{ij}=-1\}.$$

Of course $T_i=\{1,\ldots,4l\}-S_i$, $T_1=\varnothing$. The orthogonality of the rows implies

(a) $|T_i|=|S_i|=2l$, $\quad 2\leqslant i\leqslant 4l$,
(b) $|S_i\cap S_j|=|T_i\cap T_j|=l$, $\quad 2\leqslant i<j\leqslant 4l$,
(c) $|T_i\cap S_j|=l$, $\quad 2\leqslant i,j\leqslant 4l$, $\quad i\neq j$.

Setting $\mathrm{F}=\{T_1, T_2, \ldots, T_{4l}, S_1, \ldots, S_{4l}\}$, we deduce that $|F\cap F'|\equiv 0 \pmod l$ holds for $F, F'\in \mathrm{F}$. Thus

$$m_1(4l, l)\geq 8l,$$

and so, by Lemmas 1 and 2

$$m(n, l)\geq m_1(n, l)\geq (8l)^{[n/(4l)]}.$$

Now consider the hypothetical case that there is no Hadamard matrix of order $4l$. Suppose $l=l_1+\cdots+l_q$, where $l_1\leq l_2\leq\cdots\leq l_q$, and $l_i\in\mathbb{Z}^+$ are such that Hadamard matrices M_i of order $4l_i$ exist.

Let $S_j(i)$, $T_j(i)$, $1\leq j\leq 4l_i$, $1\leq i\leq q$ be the sets obtained from the matrices M_i by our construction above, where we can assume that $S_j(i)$ and $T_j(i)$ are subsets of $\{4l_1+\cdots+4l_{i-1}+1, \ldots, 4l_1+\cdots+4l_i\}$. Now define, for $1\leq j\leq 4l_1$,

$$S_j=\bigcup_{i=1}^{q} S_j(i),$$

$$T_j=\bigcup_{i=1}^{q} T_j(i).$$

6It is straightforward to verify that these sets have pairwise intersections of cardinality divisible by l, and so

$$m_1(4l, l)\geq 8l_1.$$

Since every integer $l\geq 67$ can be written as the sum of integers from $\{33, 34, \ldots, 66\}$, an application of Lemmas 1 and 2 yields the desired lower bound.

The bound (1.4) can be improved for large n and l even without assuming unproved hypotheses about existence of Hadamard matrices. It can be shown that l has a representation $l=l_1+\cdots+l_q$ with $l_i\geq\epsilon l$, $\epsilon>0$ a fixed constant, such that Hadamard matrices of order $4l_i$ exist, which enables one to replace 256 by $8\epsilon l$.

3. Upper Bounds

First we derive the upper bound

$$m_1(n, l)\leq 2^{[n/2]}. \tag{3.1}$$

Suppose $A_1, \ldots, A_m$ are subsets of $\{1, \ldots, n\}$ such that

$$|A_i\cap A_j|\equiv 0 \pmod l, \qquad 1\leq i, j\leq m.$$

Let c_i be a vector of length n defined by

$$(c_i)_j=\begin{cases}1, & \text{if } j\in A_i,\\ 0, & \text{if } j\notin A_i.\end{cases}$$

Let p be a prime divisor of l. Consider the vector space C over $GF(p)$ spanned by the c_i. Then C is self-orthogonal, since

$$c_i\cdot c_j=0, \quad 1\leq i, j\leq m.$$

Therefore, by basic linear algebra ([6]),

$$\dim C\leq [n/2].$$

Now each c_i is a 0–1 vector in C, thus (3.1) follows from the following result.

THEOREM 3 ([7]). *Suppose that U is a k-dimensional subspace of a vector space V over some field. Then, in any coordinate system for V, U has at most 2^k 0–1 vectors.*

Combining (3.1) and Lemma 1 we obtain (1.5).

In order to prove (1.6) we need the following result (a somewhat weaker bound follows from results in [8]).

THEOREM 4 ([3; Theorem 11]). *Suppose* F *is a collection of subsets of* $\{1, 2, \ldots, n\}$ *such that for* $F \neq F'$, $F, F' \in \mathrm{F}$, $|F \cap F'|$ *takes only s values. Then*

$$|\mathrm{F}| \leqslant \sum_{i=0}^{s} \binom{n}{i}.$$

In view of Lemma 1 it is sufficient to prove

$$m_1(n, l) \leqslant 2 \sum_{i=0}^{[n/(2l)]} \binom{n}{i}. \tag{3.2}$$

Suppose without loss of generality that $A_1, \ldots, A_k$ have cardinalities $\leqslant n/2$, and $A_{k+1}, \ldots, A_m$ have cardinalities $> n/2$.

Then $|A_i \cap A_j| \in \{0, l, \ldots, [n/(2l)]l\}$, $1 \leqslant i, j \leqslant k$, and moreover $|A_i \cap A_j| = [n/(2l)]l$ implies $i = j$. Thus, by Theorem 4, we have

$$k \leqslant \sum_{i=0}^{[n/(2l)]} \binom{n}{i}. \tag{3.3}$$

Next, define $B_i = \{1, 2, \ldots, n\} - A_i$, $k+1 \leqslant i \leqslant n$. Then

$$|B_i \cap B_j| = n - |A_i| - |A_j| + |A_i \cap A_j| \equiv n \pmod{l},$$

moreover for $i \neq j$ we deduce $|B_i \cap B_j| \leqslant n/2 - l = (n-2l)/2$. Thus $|B_i \cap B_j|$ for $i \neq j$ takes at most $[n/(2l)]$ different values. Again from Theorem 4 we obtain

$$m - k \leqslant \sum_{i=0}^{[n/(2l)]} \binom{n}{i}. \tag{3.4}$$

From (3.3) and (3.4) the bound (3.2) and thus (1.6) follows.

4. RELATED PROBLEMS

Our paper leaves a number of questions open. The main problem, as stated in the introduction, is to determine $c(l)$. Barring that, it would be interesting to decide whether $c(l)$ is monotone decreasing. (At this point we only know that $c(2) \geqslant c(l)$ for $l = 3, 4$, and $c(2) > c(l)$ for $l \geqslant 5$.)

One can also ask similar questions about collections of equal-sized sets. Let k be a positive integer and I a subset of $\{0, 1, \ldots, k-1\}$. Denote by $m(n, k, I)$ the maximum number of k-subsets of an n-set such that the intersection of any two distinct sets has cardinality belonging to I. It was proved in [2] that for $n > n_0(k, I)$,

$$m(n, k, I) \leqslant \prod_{i \in I} (n-i)/(k-i). \tag{4.1}$$

In particular, if $n = bl$, $k = al$, and $I = \{0, l, \ldots, (a-1)l\}$, then (4.1) gives

$$m(n, k, I) \leqslant \binom{b}{a}. \tag{4.2}$$

It would be nice to know given a and l what is the least value of b_0 such that for $b \geqslant b_0$ and $n = bl$, $k = al$, (4.2) holds. Binary self-dual codes show that in general the bound $\binom{b}{a}$ does not hold even for $l = 2$.

One can generalize our problem by asking for $m(n, l, s)$, the maximum number of subsets of an n-set, such that the intersection of any s distinct ones has cardinality divisible by l.

Obviously $m(n, l, s) \geq 2^{[n/l]}$. It can be shown that

$$c(l, s) = \lim_{n \to \infty} m(n, l, s)^{1/n}$$

exists, and that $c(l, s)$ is monotone nonincreasing in s. It seems reasonable to conjecture that for $s > s(l)$, $c(l, s) = 2^{1/l}$.

References

1. E. R. Berlekamp, On subsets with intersections of even cardinality. *Canad. Math. Bull.* **12** (1969), 363–366.
2. M. Deza, P. Erdös, and P. Frankl, Intersection properties of systems of finite sets. *Proc. London Math. Soc.* **36** (1978), 369–384.
3. P. Frankl and R. M. Wilson, Intersection theorems with geometric consequences. *Combinatorica* **1** (1981), 357–368.
4. A. V. Geramita and J. Seberry, *Orthogonal designs.* Lecture notes in pure and applied mathematics, Vol. 45, Marcel Dekker, New York, 1979.
5. J. E. Graver, Boolean designs and self-dual matroids, *Lin. Alg. Appl.* **10** (1975) 111–128.
6. F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-correcting codes*, North-Holland 1978
7. A. M. Odlyzko, On the ranks of some (0, 1)-matrices with constant row sums. *J. Austral. Math. Soc.* **31** (1981), 193–201.
8. D. K. Ray-Chaudhuri and R. M. Wilson, On t-designs. *Osaka J. Math.* **12** (1975), 735–744.

Received 2 July 1982

P. Frankl
Bell Laboratories, Murray Hill,
New Jersey 07974, U.S.A.
and CNRS, Paris, France

A. M. Odlyzko
Bell Laboratories, Murray Hill,
New Jersey 07974, U.S.A.