

On the minimum size of 4-uniform hypergraphs without property B

Patric R. J. Östergård¹

*Department of Communications and Networking
Aalto University School of Electrical Engineering
P.O. Box 13000, 00076 Aalto, Finland*

Abstract

A hypergraph is said to have property B if it is 2-colorable. Let $m(k)$ denote the minimum number of edges in a k -uniform hypergraph that does not have property B . Erdős and Hajnal introduced the problem of determining $m(k)$ in the early 1960s. The smallest cases, $m(2) = 3$ and $m(3) = 7$, are rather straightforward, but the next case has so far withstood all attacks; the possible values have been narrowed down to $21 \leq m(4) \leq 23$. By an exhaustive computer search, it is here shown that $m(4) = 23$.

Key words: graph coloring, hypergraph, property B

1. Introduction

A hypergraph \mathcal{H} is said to have property B if there exists a subset S of vertices such that every edge of \mathcal{H} intersects both S and \bar{S} , in other words, if the hypergraph has a proper 2-coloring. Erdős and Hajnal [7] proposed the problem of determining the minimum number $m(k)$ of edges in a k -uniform hypergraph that does not have property B , that is, the minimum size of a 3-chromatic such hypergraph. For short surveys of this problem, including links to other related problems, see [4, Sect. 6.6] and [11, Sect. 15.1].

*Corresponding author, Tel. +358 9 470 22341, Fax. +358 9 470 22474.

Email address: `patric.ostergard@tkk.fi` (Patric R. J. Östergård)

¹Supported by the Academy of Finland under Grants No. 107493, 110196, 130142, and 132122.

It is not difficult to prove that $m(2) = 3$ and $m(3) = 7$, attained by the complete graph K_3 and the Fano plane, respectively. These values were known to Erdős and Hajnal [5, 7], but the problem of determining $m(4)$ has remained unsolved for almost fifty years. Erdős was indeed right when writing in [5] that “...it does not seem easy to determine $m(p)$ even for $p = 4$.”

Upper bounds on $m(k)$ can be proved by constructing corresponding hypergraphs, whereas lower bounds are nonconstructive. In the early 1970s, Seymour [20] and Toft [22] independently provided (isomorphic) hypergraphs showing that $m(4) \leq 23$, improving the earlier bound $m(4) \leq 24$ by Abbott and Hanson [1]. The best known lower bound, $m(4) \geq 21$, was obtained by Manning in his dissertation [15] in 1997; earlier results include $m(4) \geq 17$ by Goldberg and Russell [9].

The aim of this work is to determine $m(4)$. In an exhaustive computer search, it is shown that all 4-uniform hypergraphs with 22 edges have property B , and thereby $m(4) = 23$. The approach relies on a study of subcases related to $m_v(k)$, a concept introduced by Erdős in [6]. Slightly different forms of the definition of $m_v(k)$ can be found in the literature; the established definition—which supports the study of $m(k)$ —coincides with that of $m(k)$ but with the additional requirement that the hypergraphs have v vertices and that every pair of vertices occur in some edge. New results obtained specifically for $m_v(4)$ include classifications of hypergraphs attaining $m_9(4) = 26$, $m_{10}(4) = 25$, and $m_{11}(4) = 23$.

The paper is organized as follows. In Section 2 some basic results and observations needed in the computer search are presented. The search is discussed in Section 3, and the results are presented in Section 4.

2. Preliminaries

Throughout the rest of the work, we view hypergraphs as combinatorial designs, with points and blocks instead of vertices and edges. (Note, however, that the objects obtained will not in general satisfy the regularity properties of 2-designs.) For clarity, design-theoretic notations are also adopted for the parameters: the number of points is denoted by v , the number of blocks by b , and the block size by k .

The following basic observation (see, for example, [15, Theorem 2.1]) has a central position in this and earlier work.

Lemma 2.1. *If there is a pair of points not occurring in any blocks of a*

design that does not have property B , then the design obtained by contracting the two points does not have property B .

Consequently, in the determination of $m(4)$, one may restrict the search for designs without property B to designs that cover all pairs of points (such designs are known as *covering designs*). By the Schönheim bound [10] it then follows that the central parameters must fulfill

$$b \geq \left\lceil \frac{v}{k} \left\lceil \frac{v-1}{k-1} \right\rceil \right\rceil. \quad (1)$$

To show that there are no designs with $b = 22$ and $k = 4$ and without property B , by (1) it suffices to consider $v \leq 16$. When bounding v from below, it is obvious that $v > 2k - 2$, since for $v \leq 2k - 2$ any partitioning of the vertices into two sets of size $\lfloor v/2 \rfloor$ and $\lceil v/2 \rceil$ forms a proper 2-coloring. Moreover, Erdős [6] showed that

$$m_{2k-1}(k) = m_{2k}(k) = \binom{2k-1}{k}.$$

Since it is also known from [2, 15] that $m_9(4) = 26$, and from [8, 9] that $23 \leq m_{10}(4) \leq 25$, it suffices to study the cases $m_v(4)$ with $11 \leq v \leq 16$ in order to determine $m(4)$. (But we will use the developed algorithm also to determine $m_{10}(4)$ and to classify the objects attaining $m_9(4)$ and $m_{10}(4)$.)

We shall next see how the problem of finding designs without property B can be viewed as a covering problem. Consider an arbitrary coloring—that is, partition P into two subsets—of the points of a design without property B . Then there exists a block D that is unicolor, and we say that D *covers* P (a block that covers a coloring is not necessarily unique). In this setting, a design without property B is a design whose blocks cover all possible colorings.

We call a coloring with the two subsets of size $\lfloor v/2 \rfloor$ and $\lceil v/2 \rceil$ *balanced*. To speed up the computations, we modify the question “Does there exist a design that covers all colorings P ?” to the form “Does there exist a design that covers all *balanced* colorings P ?” A negative answer to the modified question implies a negative answer to the original question. On the other hand, if the answer to the modified question is positive, we can find the answer to the original question by classifying all objects that give such a positive answer and check whether they are 2-colorable. The discussion of the results in Section 4 will shed further light on this issue.

3. The search

To search for a desired covering of balanced colorings with given parameters of v , k , and b , we carry out an exhaustive search for incidence matrices, proceeding point by point. On a high level, this approach is similar to that of Manning [15], but the details of the two approaches differ.

The number of 1s in a particular row of the $v \times b$ incidence matrix is not known in advance. Manning precomputed possible sets of row weights—that is, numbers of 1s—and searched for matrices of decreasing row weights. In the current work, we require increasing row weights. Denote the weight of row i of an incidence matrix by r_i . Given $r_1, r_2, \dots, r_{v'}$ for a partial $v' \times b$ matrix with $v' \geq 1$, we get that

$$r_{v'} \leq r_{v'+1} \leq \frac{bk - \sum_{i=1}^{v'} r_i}{v - v'}. \quad (2)$$

The upper bound of (2) applies also when $v' = 0$; in that case we can get a lower bound by focusing on the criterion that the real inner product of any two rows be at least 1:

$$r_1 \geq \frac{v - 1}{k - 1}. \quad (3)$$

The idea used to obtain (3) can be used for further pruning in the following way. Consider a $v' \times b$ partial incidence matrix and one of its rows of weight r ; denote the sum of the real inner products of the other rows of the partial matrix with the particularized row by S . A necessary condition for the matrix to be extendable to a $v \times b$ matrix is then that $kr - r - S \geq v - v'$, that is,

$$v - v' \leq (k - 1)r - S. \quad (4)$$

Manning [15] used a database of almost 500 rules to prune the search. Here, we instead adopt an idea that has been successful in studying a covering problem in Hamming space. The idea was first used already in the 1960s [12, 21]; more recent studies utilizing (a refined version of) it include [3, 14, 18, 19]. See also [13, Sect. 7.2.2].

Assume that we have a partial incidence matrix with v' completed rows. Now consider all possible colorings of these v' rows (w.l.o.g., we may fix the color of the first row), and for each such coloring all possible distributions of the number of the respective colors in the undetermined rows. Since we

restrict to balanced colorings, there will be 0, 1, or 2 such distributions. Finally, for each coloring and accompanying distribution, we determine how many corresponding colorings of all v rows there are.

Example. Assume that $k = 4$, $v = 11$, $v' = 4$, and call the colors R and G. For a coloring RGGG of the first four rows, the possible distributions of Rs and Gs in the last 7 rows are $4 + 3$ and $5 + 2$, and their numbers are $\binom{7}{4} = 35$ and $\binom{7}{5} = 21$, respectively.

Next, for each column of the partial incidence matrix, we can calculate how many colorings of the respective type are covered.

Example (cont). If a column of the partial matrix is 0110, one gets that the number of colorings RGGG that it covers with the Rs and Gs distributed $4 + 3$ in the last seven rows is 5. To see this, notice that the two 1s in last seven rows must have color G; the count is then obtained as there are $7 - 2$ possible positions for the final G. Similarly, for RGGG with the Rs and Gs distributed $5 + 2$, the column covers only one coloring.

Having summed up the coverings for all columns, we can check whether the thresholds (discussed in the beginning of the example) for all (balanced) colorings are reached. We call this the *covering criterion*. If the covering criterion is not fulfilled, the search is pruned.

An essential part of most combinatorial searches is isomorph rejection of partial objects. Manning [15] utilized an incomplete invariant, whereas we rely on *nauty* [16] to get a complete invariant. Moreover, for efficiency reasons, we use the approach of canonical augmentation, developed by McKay [17]. More precisely, we let *nauty* individualize one orbit of points (rows) within the rows that have the highest weight (recall that an added row is among those of highest weight). The added row is rejected if it is not in the individualized orbit. Isomorph rejection among partial solutions can then be restricted to those solutions that origin from the same parent. For more details about classification in general and isomorph rejection approaches in particular, see [13].

The algorithm for producing the set $\mathcal{M}_{v'+1}$ of partial incidence matrices with $v' + 1$ rows from those in the set $\mathcal{M}_{v'}$ with v' rows can then be summarized as follows. For each matrix in $\mathcal{M}_{v'}$, consider all possible candidate rows that have a non-zero real inner product with earlier rows and that fulfill (2), (3), and (4)—the last inequality should be fulfilled with respect to all previous rows. For each such row that leads to a matrix fulfilling the covering criterion, carry out isomorph rejection as described above; those matrices that pass the isomorph rejection phase are included in the set $\mathcal{M}_{v'+1}$.

4. The results

Although the approach is rather straightforward on a general level, efficient implementation of the algorithms and data structures are essential for its practical performance. However, since all subtasks are rather standard tasks in combinatorial computing, further details about the implementation are not discussed here. The CPU times reported here are for a Linux PC with a 2-GHz AMD Athlon 64 processor.

For $9 \leq v \leq 11$, a classification of hypergraphs attaining the value of $m_v(4)$ turned out to be feasible. The results are presented in Table 1, where the number of $v' \times m_v(4)$ partial matrices obtained in the search are shown for $1 \leq v' \leq v$. It is important to notice that we indeed consider only covering of balanced colorings in the search, so we need to test the constructed objects for 2-colorability. It turns out that 2-colorable final objects occur only for $v = 10$. Some of the incidence matrices for $v = 10$ also have a column sum different from 4, so out of the 54 constructed objects, exactly 5 are proper.

Experiments carried out with a modified version of the overall algorithm, considering arbitrary rather than balanced colorings, showed only minor differences compared with the intermediate results of the current algorithm. For example, the first column of Table 1 is then identical, except for an entry of 4551 in row 6.

INSERT TABLE 1 ABOUT HERE

The classification results in Table 1 are interesting not only for their own sake but as a partial correctness validation of the algorithm used. The designs that have been published earlier [2, 8, 20, 22] to obtain upper bounds are all among the classified objects. We list the designs attaining $m_v(4)$ for $9 \leq v \leq 11$ in the Appendix.

Theorem 4.1. *$m_9(4) = 26$ with 9 nonisomorphic solutions, $m_{10}(4) = 25$ with 5 nonisomorphic solutions, and $m_{11}(4) = 23$ with a unique solution.*

Classification can be a useful tool in the study of properties of combinatorial objects. One common property among known designs attaining $m(k)$ and $m_v(k)$ is that any pair of blocks intersect in at least one point. One might therefore conjecture—see, for example, [20]—that all designs attaining $m(k)$ have this property. However, by studying the designs in the Appendix, it

turns out that one of the nine designs attaining $m_9(4) = 26$ as well as three of the five designs attaining $m_{10}(4) = 25$ do not have this property.

Intermediate results from the search for a design with 22 blocks for $12 \leq v \leq 16$ are presented in Table 2. From the results in Table 2, we get that $m_v(4) \geq 23$ for $12 \leq v \leq 16$. The main result of this work then follows.

Theorem 4.2. $m(4) = 23$.

INSERT TABLE 2 ABOUT HERE

The total amount of CPU time consumed to obtain the presented classification and nonexistence results was about 5 months, more than half of which was consumed for one single case, $v = 13$. The amount of CPU time needed for a classification of the designs attaining $m(4) = 23$ with the current approach would even be orders of magnitude greater. It is plausible that the unique design attaining $m_{11}(4) = 23$ is the unique design attaining $m(4) = 23$. The results of the current work are in line with earlier conjectures, but an independent verification would be needed to gain full confidence in the tabulated data.

The case $k = 5$ is already out of reach for the current method, but best known upper bounds for $m(k)$ with $k \geq 5$ could perhaps be challenged by other computational methods.

Acknowledgments

The author is grateful to Helge Tverberg for introducing this problem to him.

Appendix

The nine objects attaining $m_9(4) = 26$:

11111111000000000000000000	11111111000000000000000000
11100000111111110000000000	11100000111111110000000000
11010000110000001111110000	11010000110000001111110000
10110000001100001100001111	00001110101110001110001100
00001110101011001011001100	00001101101001101001100011
00001101100110100110100011	00001011010110010001111010
00001011011001010100111010	00000111010001110110010101
00000111010100111001010101	10110000001101000101011111
01110000000011110011111111	01110000000010111010101111
11111111100000000000000000	11111111100000000000000000
11100000011111100000000000	11100000011111100000000000
11100000000000011111110000	00011100011100011111000000
00011100011100011110000100	00011100000011110000111100
00011100000011110001110010	00000011111100010000111010
00000011111100010001110001	00000011100011111110000001
10010011010011001101001111	11010010010010001101110111
01001010101010101010101111	10101001001001001011101111
00100101100101100110011111	01100100100100100111011111
11111111100000000000000000	11111111100000000000000000
11100000011111110000000000	11100000011111110000000000
00011100011100001111000000	00011100011110000111100000
00000011100011111111000000	00000011111110000000011110
11100000000010001111011110	11100000010000000111011101
00011100000011111000011101	00011100001001110000011101
10010011011001000100111011	10010011000001101110110010
01001010110100100010110111	01001010100101011101101011
00100101101100010001101111	00100101100010111011100111
11111111110000000000000000	11111111111000000000000000
11110000001111111000000000	11110000000111111100000000
11001100001100000111110000	11110000000000000001111110
10101000000011000110001111	10001110000111000011100001
00010011101010100101101100	01001110000000111000011101
00010011011001010011010011	00100001110111000000011101
00000110110110101010011010	00001001101100110111010010
00000101110101011100100101	00000101011010101110101010
01101000000000111001111111	00010010111001011101100111

11111111110000000000000000
 11110000000111111100000000
 10001110000111000011110000
 01001001111100110011000000
 00100101100010101010101110
 11001000000100001100111111
 00100011010001100101011101
 00010100101001010110011011
 00010010011010011001100111

The five objects attaining $m_{10}(4) = 25$:

11111000000000000000000000	11111100000000000000000000
10000111111000000000000000	10000011111111000000000000
01100110000111111000000000	01000011100000111110000000
00011110000110000111100000	00110010011000110001110000
10000100000001100110011111	0000111001100000110100110
10000010000000011001111111	0011000100011000110010101
0101000111010101010101100	0000110100011011000001011
0010100110101101001011010	1010100010000110101011110
0101000101110010101010011	0101010011010100001110011
0010100011101010110100101	1100000000101101011101101
11111100000000000000000000	11111111000000000000000000
10000011111111000000000000	11000000111111000000000000
01000011100000111110000000	00111000110000111000000000
00110010011000110001110000	0010010000110010011110000
0000110101100000110100110	0010010000001110000001111
0011001000011000110010101	0001001000110001010001110
0000110100011011000001011	0001001000001101001110001
1010100010000110101011110	1000100110101000111101101
0101010011010100001110011	0100100101010100111011011
1100000000101101011101101	1100011111000011100110110

```

11111111100000000000000000
110000001111111000000000
10110000100000111110000
100010000110000100001111
0010011000101100011001100
0010010100010011010100011
0001001010011010001011010
0001000110100101000110101
0100110011000110110010110
0100101101001001101101001

```

The unique object attaining $m_{11}(4) = 23$:

```

11111111100000000000000000
110000001111111000000000
10100000100000111110000
10010000010000100001111
00001100001100011001100
00001010001010010101010
00000110000110001100110
01100001100001000011110
01010001010001011100001
00110001001110100010001
00001110110001100010001

```

References

- [1] H. L. Abbott and D. Hanson, On a combinatorial problem of Erdős, *Canad. Math. Bull.* **12** (1969), 823–829.
- [2] H. L. Abbott and A. C. Liu, On property B of families of sets, *Canad. Math. Bull.* **23** (1980), 429–435.
- [3] U. Blass and S. Litsyn, Several new lower bounds on the size of codes with covering radius one, *IEEE Trans. Inform. Theory* **44** (1998), 1998–2002.
- [4] F. Chung and R. Graham, *Erdős on Graphs; His Legacy of Unsolved Problems*, A K Peters, Natick, Mass., 1999.

- [5] P. Erdős, On a combinatorial problem, *Nordisk Mat. Tidskr.* **11** (1963), 5–10.
- [6] P. Erdős, On a combinatorial problem III, *Canad. Math. Bull.* **12** (1969), 413–416.
- [7] P. Erdős and A. Hajnal, On a property of families of sets, *Acta Math. Acad. Sci. Hungar.* **12** (1961), 87–123.
- [8] G. Exoo, On constructing hypergraphs without property B , *Ars Combin.* **30** (1990), 3–12.
- [9] M. K. Goldberg and H. C. Russell, Toward computing $m(4)$, *Ars Combin.* **39** (1995), 139–148.
- [10] D. M. Gordon and D. R. Stinson, Coverings, in: *Handbook of Combinatorial Designs* (C. J. Colbourn and J. H. Dinitz, editors), 2nd ed., Chapman & Hall/CRC, Boca Raton, 2007, pp. 365–373.
- [11] T. R. Jensen and B. Toft, *Graph Coloring Problems*, Wiley, New York, 1995.
- [12] H. J. L. Kamps and J. H. van Lint, The football pool problem for 5 matches, *J. Combin. Theory* **3** (1967), 315–325.
- [13] P. Kaski and P. R. J. Östergård, *Classification Algorithms for Codes and Designs*, Springer, Berlin, 2006.
- [14] J. Linderoth, F. Margot, and G. Thain, Improving bounds on the football pool problem by integer programming and high-throughput computing, *INFORMS J. Comput.* **21** (2009), 445–457.
- [15] G. Manning, The $M(4)$ problem of Erdős and Hajnal, Ph.D. dissertation, Northern Illinois University, 1997.
- [16] B. D. McKay, *nauty* user’s guide (version 1.5), Tech. Rep. TR-CS-90-02, Computer Science Department, Australian National University, 1990.
- [17] B. D. McKay, Isomorph-free exhaustive generation, *J. Algorithms* **26** (1998), 306–324.

- [18] P. R. J. Östergård and U. Blass, On the size of optimal binary codes of length 9 and covering radius 1, *IEEE Trans. Inform. Theory* **47** (2001), 2556–2557.
- [19] P. R. J. Östergård and A. Wassermann, A new lower bound for the football pool problem for 6 matches, *J. Combin. Theory Ser. A* **99** (2002), 175–179.
- [20] P. D. Seymour, A note on a combinatorial problem of Erdős and Hajnal, *J. London Math. Soc. (2)* **8** (1974), 681–682.
- [21] R. G. Stanton and J. G. Kalbfleisch, Intersection inequalities for the covering problem, *SIAM J. Appl. Math.* **17** (1969), 1311–1316.
- [22] B. Toft, On colour-critical hypergraphs, in: *Infinite and Finite Sets*, Vol. III (A. Hajnal, R. Rado and V. T. Sós, editors), Colloq. Math. Soc. János Bolyai 10, North-Holland, Amsterdam, 1975, pp. 1445–1457.

Table 1: Classification results

$v \setminus v, b$	9,26	10,25	11,23
1	9	8	5
2	27	65	54
3	319	1137	1130
4	4623	27744	32224
5	29021	429348	1047141
6	4553	1482550	6912371
7	316	168514	2630418
8	10	1319	5405
9	9	57	3252
10		54	1
11			1

Table 2: Nonexistence results for $b = 22$

$v' \setminus v$	12	13	14	15	16
1	4	3	2	1	1
2	30	17	11	2	1
3	529	332	110	12	2
4	10908	14785	1749	166	3
5	252039	809775	42589	3809	4
6	4306867	21812049	867709	144040	8
7	4603896	183756381	10359177	4228538	12
8	276533	48207126	73303128	46411646	16
9	148	5613104	22235178	55733068	106
10	1	59	63869	1467230	574
11	0	0	82	3244	281
12			0	0	1
13					0