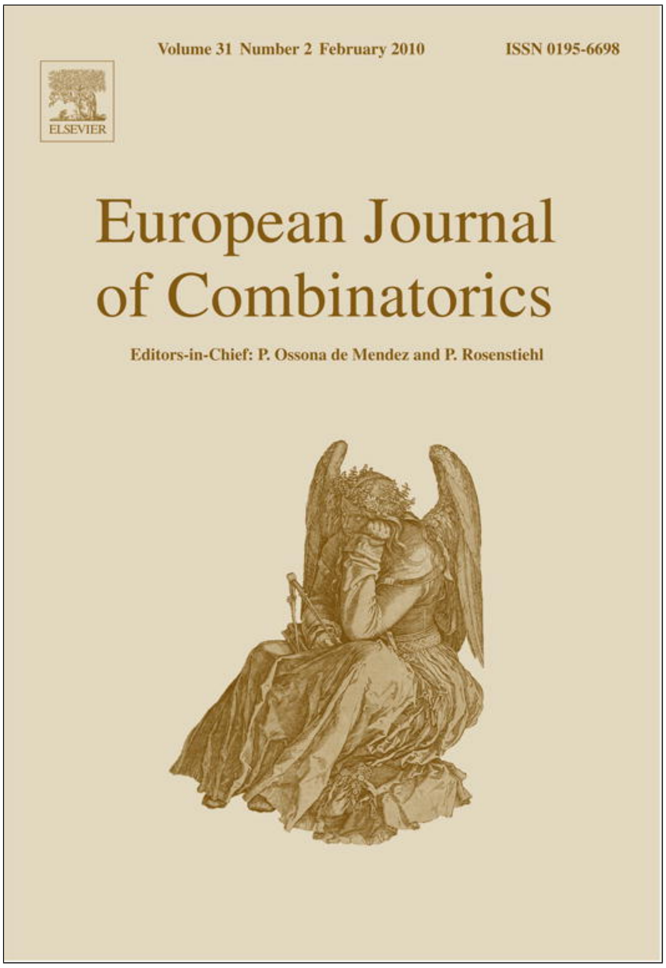
Volume 31 Number 2 February 2010
ISSN 0195-6698
ELSEVIER
European Journal of Combinatorics
Editors-in-Chief: P. Ossona de Mendez and P. Rosenstiehl

European Journal of Combinatorics 31 (2010) 491–501

Contents lists available at ScienceDirect

European Journal of Combinatorics

journal homepage: www.elsevier.com/locate/ejc

New identifying codes in the binary Hamming space

Irène Charon, Gérard Cohen, Olivier Hudry [1], Antoine Lobstein

GET - Télécom Paris & CNRS - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13, France

ARTICLE INFO

Article history:
Available online 25 April 2009

ABSTRACT

Let F^n be the binary n-cube, or binary Hamming space of dimension n, endowed with the Hamming distance. For $r \geq 1$ and $x \in F^n$, we denote by $B_r(x)$ the ball of radius r and centre x. A set $C \subseteq F^n$ is said to be an r-identifying code if the sets $B_r(x) \cap C$, $x \in F^n$, are all nonempty and distinct. We give new constructive upper bounds for the minimum cardinalities of r-identifying codes in the Hamming space.

1. Introduction

We define identifying codes in a connected, undirected graph $G = (V, E)$, in which a *code* is simply a nonempty subset of vertices. This definition can help to unambiguously determine a vertex, and the motivations may come from processor networks where we wish to locate a faulty vertex under certain conditions.

In G we define the usual distance $d(v_1, v_2)$ between two vertices $v_1, v_2 \in V$ as the smallest possible number of edges in any path between them. For an integer $r \geq 0$ and a vertex $v \in V$, we define $B_r(v)$ (resp., $S_r(v)$), the *ball* (resp., *sphere*) of radius r centred at v, as the set of vertices within distance (resp., at distance exactly) r from v. Whenever two vertices v_1 and v_2 are such that $v_1 \in B_r(v_2)$ (or, equivalently, $v_2 \in B_r(v_1)$), we say that they r-*cover* each other. A set $X \subseteq V$ r-covers a set $Y \subseteq V$ if every vertex in Y is r-covered by at least one vertex in X.

The elements of a code $C \subseteq V$ are called *codewords*. For each vertex $v \in V$, we denote by

$$K_{C,r}(v) = C \cap B_r(v)$$

the set of codewords r-covering v. Two vertices v_1 and v_2 with $K_{C,r}(v_1) \neq K_{C,r}(v_2)$ are said to be r-*separated* by code C, and any codeword belonging to exactly one of the two sets $B_r(v_1)$ and $B_r(v_2)$ is said to r-*separate* v_1 and v_2.

E-mail addresses: irene.charon@enst.fr (I. Charon), gerard.cohen@enst.fr (G. Cohen), olivier.hudry@enst.fr (O. Hudry), antoine.lobstein@enst.fr (A. Lobstein).

[1] Fax: +33 1 45 81 31 19.

0195-6698/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejc.2009.03.032

A code $C \subseteq V$ is called *r-identifying* [13] if all the sets $K_{C,r}(v)$, $v \in V$, are nonempty and distinct. In other words, every vertex is r-covered by at least one codeword, and every pair of vertices is r-separated by at least one codeword. Such codes are also sometimes called *differentiating dominating sets* [9].

In the following, we drop the general case and turn to the binary Hamming space of dimension n, also called the binary n-cube. First we need to give some specific definitions and notation.

We consider the n-cube as the set of binary row-vectors of length n, and in view of this, we denote it by $G = (F^n, E)$ with $F = \{0, 1\}$ and $E = \{\{x, y\} : d(x, y) = 1\}$, the usual graph distance $d(x, y)$ between two vectors x and y being called here the *Hamming distance* — it simply consists of the number of coordinates where x and y differ. The *Hamming weight* of a vector x is its distance to the all-zero vector, i.e., the number of its nonzero coordinates. Additions are always carried out coordinatewise and modulo 2.

We denote by 0^n (resp., 1^n) the all-zero (resp., all-one) vector of length n. For two sets $X \subseteq F^{n_1}$, $Y \subseteq F^{n_2}$, the *direct sum* of X and Y, denoted by $X \oplus Y$, is defined by $X \oplus Y = \{x|y \in F^{n_1+n_2} : x \in X, y \in Y\}$, where | stands for concatenation of vectors. We use the notation (r, n) or $(r, n)K$ for a code in F^n which is r-identifying and has K elements. Finally, we denote by $M_r(n)$ the smallest possible cardinality of an (r, n) code.

In Section 2, we give various methods for constructing identifying codes, thus obtaining, in Section 3, upper bounds on $M_r(n)$, of which several are new. These bounds are summarized in Tables at the end of the paper.

For previous works, we refer the reader to, e.g., [1–3,6–8,10–13,15–17]. In the recent [7], tables for exact values or bounds on $M_1(n)$, $2 \le n \le 19$, and $M_2(n)$, $3 \le n \le 21$, were given.

See also [18] for a bibliography on identifying codes and related concepts.

2. Constructing identifying codes

Our constructions will use Theorem 2 below, as well as various heuristics.

2.1. Extending an identifying code

In the constructions of Theorems 2 and 3 below, we use a new definition: a code is called *r-separating* if every pair of vertices is r-separated by at least one codeword [2, Sec. 3] (we do not require any longer that every vertex be r-covered by at least one codeword). The following remark and lemma are easy.

Remark 1. (i) For $0 \le r \le n - 1$, a code $C \subseteq F^n$ is r-separating if, and only if, it is also $(n - r - 1)$-separating, because $B_r(x) = F^n \setminus B_{n-r-1}(x + 1^n)$ for all $x \in F^n$.

(ii) (cf. [10]) Since a separating code is such that at most one vertex can be covered by no codeword, the size of an optimum r-separating code in F^n is $M_r(n)$ or $M_r(n) - 1$, and we have

$$M_{\max\{r,n-r-1\}}(n) \le M_{\min\{r,n-r-1\}}(n) \le M_{\max\{r,n-r-1\}}(n) + 1, \tag{1}$$

i.e., the symmetry, with respect to $\lfloor (n-1)/2 \rfloor$, observed for separating codes, still holds, within 1, for identifying codes. □

Lemma 1. *For all $p \ge 1$ and $\Delta \in \{0, 1, \ldots, p - 1\}$, the set $F^p \setminus \{0^p\}$ is Δ-separating.* □

The following theorem is inspired by [13, Th. 9] and [7, Ex. 2 and Th. 4]. Starting with an (r, n) code C, we intend to see how the direct sum $C \oplus F^p$ can be used for constructing an $(r, n + p)$ code. In construction $\mathcal{C}2$ below, k is an additional parameter on which we can act.

More comments on how to understand and use Theorem 2 are given after its statement.

Theorem 2. *Let $r \ge 1$, $p \ge 1$, and $k \in \{0, 1, \ldots, p - 1\}$; let C be an (r, n) code and*

$$X_p = \{x \in F^n : \forall c \in C, d(x, c) \le r - p \text{ or } d(x, c) > r\}.$$

Construction $\mathcal{C}1$: *Let* $Y_p \subseteq F^n$ *be a set such that for every* $x \in X_p$ *there exists* $y \in Y_p$ *with* $r-p+1 \le d(x,y) \le r$. *Then*

$$C' = \left(C \oplus F^p\right) \cup \left(Y_p \oplus (F^p \setminus \{0^p\})\right) \tag{2}$$

is $(r, n+p)$.

Construction $\mathcal{C}2$: *Let* $Y_{p,k} \subseteq F^n$ *be a set such that for every* $x \in X_p$ *there exists* $y \in Y_{p,k}$ *with* $d(x,y) = r-k$, *and let* $C_{p,k}$ *be a* k*-separating code in* F^p. *Then*

$$C' = (C \oplus F^p) \cup (Y_{p,k} \oplus C_{p,k}) \tag{3}$$

is $(r, n+p)$.

Proof. See the proof of Theorem 3, which contains Theorem 2 as a particular case. □

Theorem 2 calls for several remarks, in order to make its dry technicality more friendly.

Remark 2. Obviously, it is best to choose Y_p (for construction $\mathcal{C}1$) and $Y_{p,k}$, $C_{p,k}$ (for construction $\mathcal{C}2$) with the smallest possible cardinalities. □

Remark 3. Ideally, $X_p = \emptyset$; then $Y_p = Y_{p,k} = \emptyset$ and $C \oplus F^p$ is $(r, n+p)$. This is Th. 4 in [7] (Th. 1 in [3] for $r = 1$). *This is the case as soon as* $p \ge r+1$; cf. Cor. 3 in [7] (Th. 2 in [3] for $r = 1$). Therefore we can limit our investigations to

$$p \le r.$$

On the other hand, we have

$$X_1 \supseteq X_2 \supseteq \cdots \supseteq X_r,$$

so the smaller the number p, probably the more difficult to jump to length $n+p$ without having a large set Y_p or $Y_{p,k}$. □

Remark 4. In construction $\mathcal{C}1$, we build a minimum set Y_p using the union of p spheres of radii ranging from $r-p+1$ to r, whereas in construction $\mathcal{C}2$, for $Y_{p,k}$ we use only one sphere of radius $r-k$. We can therefore hope for a set Y_p (much) smaller than each set $Y_{p,k}$. The price to pay is that $|Y_p|$ has to be multiplied by $2^p - 1$, whereas $|Y_{p,k}|$ has a (much) smaller factor in (3).

When $k = 0$ or $k = p-1$, the smallest k-separating codes in F^p have size $2^p - 1$, and construction $\mathcal{C}2$ is not better than construction $\mathcal{C}1$; therefore, for construction $\mathcal{C}2$ we can limit ourselves to the cases

$$1 \le k \le p-2, \qquad 3 \le p \le r.$$

For different values of p and k, it seems very difficult to compare constructions $\mathcal{C}1$ and $\mathcal{C}2$, or constructions $\mathcal{C}2$ between themselves. For a fixed p, k varies from 1 to $p-2$. When k increases, up to $\lfloor (p-1)/2 \rfloor$, it may be that $|Y_{p,k}|$ increases and $|C_{p,k}|$ decreases (and, by Remark 1(i) before Theorem 2, in this case $|C_{p,k}|$ would increase when k ranges from $\lfloor (p-1)/2 \rfloor + 1$ to $p-2$); but actually the former hypothesis highly depends on particular situations (see Example 1), and the latter, more general, one remains to be proved. □

Example 1. We use the notation of Theorem 2. In F^{10}, consider the five vectors $x_1 = 1^2|0^8$, $x_2 = 0^2|1^2|0^6$, $x_3 = 0^4|1^2|0^4$, $x_4 = 0^6|1^2|0^2$, $x_5 = 0^8|1^2$. Then 0^{10} is at distance 2 from each of them, but it is easy to see that it is impossible to find a vector which is at distance 1 from each of them or a vector which is at distance 3 from each of them. So, if $X_p = \{x_1, x_2, x_3, x_4, x_5\}$, then we have $|Y_{p,r}| = 5$, $|Y_{p,r-1}| > 1$, $|Y_{p,r-2}| = 1$ and $|Y_{p,r-3}| > 1$. □

This could indicate that, in the absence of information on $|Y_{p,k}|$, a reasonable bet is to take $k = \lfloor (p-1)/2 \rfloor$, assuming that $|C_{p,k}|$ is minimum for this k. Let us give two small examples.

Example 2. - The case of $p = 3$; $r \ge 3$, $k = 1$.

Y_3 is such that $d(x,y) = r-2$, $r-1$ or r, and $|Y_3|$ is multiplied by 7.

$Y_{3,1}$ is such that $d(x,y) = r-1$, and $|Y_{3,1}|$ is multiplied by $M_1(3) - 1 = 3$: $C_{3,1} = \{000, 001, 100\}$ is 1-separating in F^3 (but not 1-identifying: 111 is not 1-covered by $C_{3,1}$).

- The case of $p = 5$; $r \geq 5, k \in \{1, 2, 3\}$.
  $Y_5 : d(x, y) \in \{r-4, r-3, r-2, r-1, r\}$, and $|Y_5|$ multiplied by 31.
  $Y_{5,1} : d(x, y) = r-1$, $|Y_{5,1}|$ multiplied by $M_1(5) = 10$ (the method of [8, Th. 2] can also be used to give a general lower bound on separating codes showing that there is no 1-separating code of size 9 in F^5).
  $Y_{5,2} : d(x, y) = r-2$, $|Y_{5,2}|$ multiplied by $M_2(5) = 6$ (it is not very difficult to prove that there is no 2-separating code of size 5 in F^5). □

Remark 5. The definition of C' shows that $|C|$ will have a factor 2^p, so it seems best, in general, to take a code C as small as possible. However, it may be that a larger C, together with a (smaller) X_p inducing a smaller Y_p or $Y_{p,k}$, gives better results. In practice, since one cannot try everything, we were led to use the best identifying codes at our disposal. □

Remark 6. If in (2) we replace $F^p \setminus \{0^p\}$ by F^p, we obtain a new code $C'' = (C \oplus F^p) \cup (Y_p \oplus F^p) = (C \cup Y_p) \oplus F^p$ which is also $(r, n+p)$ and has $X''_p = \{x \in F^{n+p} : \forall c \in C'', d(x, c) \leq r-p \text{ or } d(x, c) > r\} = \emptyset$ (indeed, for all $x_1 \in F^n$, there is, by construction, $v_1 \in C \cup Y_p$ such that $r-p+1 \leq d(x_1, v_1) \leq r$, so for all $x_1|x_2 \in F^{n+p}$, we have $r-p+1 \leq d(x_1|x_2, v_1|x_2) \leq r$, with $v_1|x_2 \in C''$). Therefore, we can apply [7, Th. 4], mentioned in Remark 3, to C'' and reach lengths higher than just $n+p$. □

Open problem. Among all (r, n) codes C with $|C| = M_r(n)$, is there at least one such that the set X_r defined in Theorem 2 is empty? If the answer is YES, then $M_r(n+r) \leq 2^r M_r(n)$; in particular, we would have $M_1(n+1) \leq 2M_1(n)$. Could this be true for X_p for any $p \in \{1, \ldots, r\}$, so that we would have $M_r(n+p) \leq 2^p M_r(n)$? □
It is possible to generalize the previous construction, changing both length (from n to $n+p$) and radius (from r_1 to r_1+r_2), the case $r_2 = 0$ being exactly Theorem 2. Similarly, it will be best to choose Y_{p,r_2}, $Y_{p,r_2,k}$, $C_{p,k}$ with the smallest possible sizes.

Theorem 3. *Let $r_1 \geq p \geq r_2 \geq 0$, and $k \in \{0, 1, \ldots, p-1\}$; let C be an (r_1, n) code and*

$$X_{p,r_2} = \{x \in F^n : \forall c \in C, d(x, c) \leq r_1 - p + r_2 \text{ or } d(x, c) > r_1 + r_2\}.$$

Construction $\mathcal{C}1$: Let $Y_{p,r_2} \subseteq F^n$ be a set such that for every $x \in X_{p,r_2}$ there exists $y \in Y_{p,r_2}$ with $r_1 - p + r_2 + 1 \leq d(x, y) \leq r_1 + r_2$. Then

$$C' = \left(C \oplus F^p\right) \cup \left(Y_{p,r_2} \oplus (F^p \setminus \{0^p\})\right)$$

is $(r_1 + r_2, n+p)$.

Construction $\mathcal{C}2$: Let $Y_{p,r_2,k} \subseteq F^n$ be a set such that for every $x \in X_{p,r_2}$ there exists $y \in Y_{p,r_2,k}$ with $d(x, y) = r_1 + r_2 - k$, and let $C_{p,k}$ be a k-separating code in F^p. Then

$$C' = (C \oplus F^p) \cup (Y_{p,r_2,k} \oplus C_{p,k})$$

is $(r_1 + r_2, n+p)$.

Proof. First, we prove, in both constructions, $\mathcal{C}1$ and $\mathcal{C}2$, that any $x \in F^{n+p}$ is $(r_1 + r_2)$-covered by a codeword in C'. We write $x = x_1|x_2$ with $x_1 \in F^n$, $x_2 \in F^p$. Because C is r_1-identifying in F^n, there is a codeword $c \in C$ such that $d(c, x_1) \leq r_1$. Therefore, $d(c|x_2, x_1|x_2) \leq r_1 \leq r_1 + r_2$, with $c|x_2 \in C \oplus F^p \subseteq C'$.

Next, we prove that, given any two vectors $x, y \in F^{n+p}$ $(x \neq y)$, there is a codeword in C' which $(r_1 + r_2)$-separates them. We write $x = x_1|x_2$, $y = y_1|y_2$, with $x_1, y_1 \in F^n$, $x_2, y_2 \in F^p$. We distinguish between four cases. The first three cases, (i)–(iii), work for both constructions $\mathcal{C}1$ and $\mathcal{C}2$, because only $C \oplus F^p$ is needed.

(i) $x_1 \neq y_1$, $x_2 \neq y_2$. Then there is a codeword $c \in C$ such that, say, $d(c, x_1) \leq r_1$ and $d(c, y_1) > r_1$. If $r_2 \leq p-1$, then two spheres with radius r_2 and distinct centres are different in F^p, and one is not included in the other. So there is a vector $v \in F^p$ which is within distance r_2 from x_2 and not from y_2. If $r_2 = p$, we take $v = y_2 + 1^p$, so that $d(v, y_2) = r_2$ and $d(v, x_2) \leq r_2$.

In both cases, $d(c|v, x_1|x_2) \leq r_1 + r_2$ and $d(c|v, y_1|y_2) > r_1 + r_2$, with $c|v \in C \oplus F^p \subseteq C'$.

(ii) $x_1 \neq y_1, x_2 = y_2$. Apply the argument in (i) with $v = x_2 + 1^{r_2}|0^{p-r_2}$.

(iii) $x_2 \neq y_2$ and $x_1 = y_1 \notin X_{p,r_2}$. Then there is a codeword $c \in C$ such that $r_1 - p + r_2 + 1 \leq d(c, x_1) \leq r_1 + r_2$. If we set $\Delta = r_1 + r_2 - d(c, x_1)$, we see that $0 \leq \Delta \leq p - 1$. Therefore, as in case (i), we can find a vector $v \in F^p$ which is within distance Δ from x_2 and not from y_2. Now $d(c|v, x_1|x_2) \leq d(c, x_1) + \Delta = r_1 + r_2$ and $d(c|v, x_1|y_2) > d(c, x_1) + \Delta = r_1 + r_2$, with $c|v \in C \oplus F^p \subseteq C'$.

(iv) $x_2 \neq y_2$ and $x_1 = y_1 \in X_{p,r_2}$.

In construction $\mathcal{C}1$, there is a vector $z \in Y_{p,r_2}$ such that $r_1 - p + r_2 + 1 \leq d(z, x_1) \leq r_1 + r_2$. Then if we set $\Delta = r_1 + r_2 - d(z, x_1)$, we see that $0 \leq \Delta \leq p - 1$, and by Lemma 1, there is a vector $v \in F^p \setminus \{0^p\}$ which is within distance Δ from x_2 and not from y_2, or the other way round. Then $d(z|v, x_1|x_2) \leq d(z, x_1) + \Delta = r_1 + r_2$ and $d(z|v, x_1|y_2) > d(z, x_1) + \Delta = r_1 + r_2$, or the other way round, with $z|v \in Y_{p,r_2} \oplus (F^p \setminus \{0^p\}) \subseteq C'$, and we have proved that x and y are $(r_1 + r_2)$-separated by C'.

In construction $\mathcal{C}2$, there is a vector $z \in Y_{p,r_2,k}$ such that $d(z, x_1) = r_1 + r_2 - k$ and a codeword $c \in C_{p,k}$ such that, say, $d(c, x_2) \leq k$ and $d(c, y_2) > k$. Then $d(z|c, x_1|x_2) \leq r_1 + r_2$ and $d(z|c, x_1|y_2) > r_1 + r_2$, with $z|c \in Y_{p,r_2,k} \oplus C_{p,k} \subseteq C'$. $\square$

Remark 7. The set X_{p,r_2} is empty whenever $1 \leq r_2 \leq p - 1$ [14]. Indeed, if C is (r_1, n), then for any vertex $x \in F^n$, there is a codeword c such that $d(x, c) = r_1$ or $r_1 + 1$; otherwise, if e denotes a vector of weight 1, then x and $x + e$ could not be r_1-separated by C. With the conditions $-p + r_2 \leq -1$ and $r_2 \geq 1$, a vertex x is in X_{p,r_2} if for all $c \in C$, $d(x, c) \leq r_1 - 1$ or $d(x, c) > r_1 + 1$, and we have just seen that this is impossible. $\square$

Together with [7, Cor. 3] and Remark 7, Theorems 2 and 3 immediately yield the following corollary on codes sizes.

Corollary 4. (1) *Let $r \geq 1$, $p \geq 1$, and $k \in \{0, 1, \dots, p - 1\}$. We have*

$$M_r(n+p) \leq \begin{cases} 2^p M_r(n) \text{ if } p \geq r+1 & \text{[7, Cor. 3]} \\ 2^p M_r(n) + (2^p - 1)|Y_p| & (\mathcal{C}1 \text{ of Theorem 2}) \\ 2^p M_r(n) + |C_{p,k}||Y_{p,k}| & (\mathcal{C}2 \text{ of Theorem 2}), \end{cases}$$

where Y_p, $Y_{p,k}$ and $C_{p,k}$ are as in Theorem 2.

(2) *Let $r_1 \geq p \geq r_2 > 0$, and $k \in \{0, 1, \dots, p - 1\}$. We have*

$$M_{r_1+r_2}(n+p) \leq \begin{cases} 2^p M_{r_1}(n) \text{ if } 0 < r_2 < p & \text{(Remark 7)} \\ 2^p M_{r_1}(n) + (2^p - 1)|Y_{p,p}| \text{ if } r_2 = p & (\mathcal{C}1 \text{ of Theorem 3}) \\ 2^p M_{r_1}(n) + |C_{p,k}||Y_{p,p,k}| \text{ if } r_2 = p & (\mathcal{C}2 \text{ of Theorem 3}), \end{cases}$$

where $Y_{p,p}$, $Y_{p,p,k}$ and $C_{p,k}$ are as in Theorem 3. $\square$

2.2. Heuristics: Noising and greedy

It is known [12] that deciding whether a given code $C \subseteq F^n$ is r-identifying is co-NP-complete. This suggests that constructing good identifying codes in the Hamming space might be hard.

Here, we use two different heuristic methods in order to build good identifying codes, *noising* and *greedy*, as well as combinations of the two.

Noising algorithms have already been used in [5] for the construction of identifying codes in various grids; they constitute a family of metaheuristics, of which one is a generalization of simulated annealing [4]. Another of these consists of the following. Once r, n and a number of codewords, M, have been fixed, we consider codes $C \subseteq F^n$ with M codewords, and we define $NC(C)$ as the number of vectors which are not r-covered by C, $NS(C)$ as the number of pairs of vectors not r-separated by C, and the *evaluation function*

$$f(C) = NC(C) + NS(C),$$

which we try to make equal to zero. A starting code is chosen, which will be the current code C. We iteratively modify the current code, using an *elementary transformation* which consists in replacing a codeword by a noncodeword, thus keeping $|C| = M$.

Now when do we accept an elementary transformation? We cyclically go through all codewords: after looking into the last codeword, we start again with the first one. Looking into a codeword m means that we go through all vectors s in $F^n \setminus C$, we note $C_{m,s} = C \setminus \{m\} \cup \{s\}$, and we compute

$$\Delta(C, m, s) = f(C_{m,s}) - f(C).$$

For each s, we also compute a noised value

$$\Delta_{\text{noise}}(C, m, s) = \Delta(C, m, s) + (\rho \times \ln(R)),$$

where ρ is a tuning parameter which we make decrease, and R is a number which is randomly chosen for each new elementary transformation (see below for more details).

If there is a vector s for which $\Delta(C, m, s) < 0$, then we keep a vector s_0 which minimizes $\Delta(C, m, s)$.

If for all vectors s, we have $\Delta(C, m, s) \geq 0$, then we look for a vector s_0 which minimizes $\Delta_{\text{noise}}(C, m, s)$, and we keep s_0 only if $\Delta_{\text{noise}}(C, m, s_0) < 0$.

If a vector s_0 has been found in one of the two cases above, then we apply the elementary transformation with C, m and s_0, so that C becomes $C \setminus \{m\} \cup \{s_0\}$. Otherwise, the current code is not modified after looking into m. After each accepted elementary transformation, we check the evaluation function of the current code: if $f(C) = 0$, then C is r-identifying.

If we have found an identifying code, we reinitialize the process by removing from the current code C a codeword m which minimizes $f(C \setminus \{m\})$, and we cyclically go through the remaining codewords.

The parameter R is a real number, randomly chosen, in a uniform way, between zero and one; the noising rate ρ is a positive real number which we decrease arithmetically from an initial value down to zero, and for each value of ρ, we cyclically go through the codewords a certain number of times.

How do we choose the starting codes? We observed that, for given r, n and M, it was more efficient to use an r-identifying code of size larger than M, from which we deleted codewords until it had size M, rather than simply take a random code of size M. These starting identifying codes were very often obtained by the constructions of Theorem 2.

Greedy algorithms are based on the following simple idea: starting from an empty code C, at each step we choose to add in C a codeword m which will maximize $f(C) - f(C \cup \{m\})$. In the case of a tie, the choice is made at random.

3. Results

We give tables of lower and upper bounds on $M_r(n)$ for $1 \leq r \leq 5$, $1 \leq n \leq 21$. There are boldface figures when the exact value is known. Up to now, the most extensive tables ($r = 1, n \leq 19$, and $r = 2, n \leq 21$) had been given in [7].

3.1. *Using heuristics*

The upper bounds which are marked by a star in our tables were obtained by heuristic methods. For instance (see Table 1), the code consisting of the length-9 binary expressions of the following 112 integers:

0, 13, 14, 27, 31, 32, 35, 39, 43, 44, 53, 54, 56, 58, 65, 67, 68, 79, 81, 82, 84, 86, 110, 115, 120, 121, 125, 130, 133, 134, 136, 137, 144, 149, 155, 162, 169, 177, 181, 190, 200, 204, 211, 215, 218, 220, 221, 222, 225, 226, 235, 239, 246, 247, 248, 253, 256, 263, 266, 275, 276, 278, 281, 284, 300, 301, 313, 319, 328, 330, 331, 341, 343, 344, 351, 354, 357, 358, 365, 366, 368, 370, 371, 373, 382, 391, 398, 399, 400, 402, 405, 409, 417, 420, 423, 426, 434, 444, 446, 447, 449, 452, 454, 459, 461, 468, 481, 484, 488, 491, 498, 509,

is a $(1, 9)112$ code obtained by noising. All of our best codes can be found, in the same form, at http://www.infres.enst.fr/~charon/newIdentifyingNcube.html

3.2. *Applying* Theorem 2

The codes obtained by noising and greedy methods are used in this section in order to obtain codes of greater lengths, thanks to the constructions of Theorem 2. Since most of these results will be further improved, we do not give many details here.

Table 1
Lower and upper bounds, $r = 1$.

n	Lower bound	Upper bound	Previous known upper bound
2	a **3**		**3** B
3	b **4**		**4** A
4	d **7**		**7** C
5	b **10**		**10** A
6	n **19**		**19** D
7	e **32**		**32** E
8	c 56	61*	62 F
9	c 101	112*	114 J
10	c 183	208*	214 J
11	c 337		352 F
12	c 623	684*	696 F
13	c 1158	1280*	1344 F
14	c 2164	2550*	2784 F
15	c 4063	4787*	5120 F
16	c 7654	9494*	10240 F
17	c 14469	18558*	20480 F
18	c 27434	35604*	40960 F
19	c 52155		65536 F
20	c 99392	131072 (5)	
21	c 189829	262144 (4)	

a [13, Th. 1(iii)].
A [13].
b [13, Th. 2].
B $M_{n-1}(n) = 2^n - 1$ [2, Th. 5].
c [13, Th. 3].
C [3, Th. 4].
d [3, Th. 4].
D [3, Th. 5].
e [3, Th. 11].
E [3, Th. 6].
F [7, Tables 3 and 4].
J [8].
∗ *Heuristics*.
n [8, Th. 11].

First, using [7, Cor. 3] ([3, Th. 2] for $r = 1$), mentioned in Remark 3,

$$M_1(21) \le 4M_1(19) \le 262144; \qquad M_2(21) \le 8M_2(18) \le 51440. \tag{4}$$

(i) Because the $(1, 19)65536$ code from [7] is such that every vector is 1-covered by at least two codewords, we have $X_1 = \emptyset$ and

$$M_1(20) \le 2 \cdot 65536 = 131072. \tag{5}$$

(ii) Using a $(2, 18)$ code which has since been improved, we obtained

$$M_2(19) \le 13458; \qquad M_2(20) \le 26710. \tag{6}$$

(iii) Using a $(3, 18)1628$ code, we obtained

$$M_3(19) \le 3330; \qquad M_3(20) \le 6569; \qquad M_3(21) \le 13030. \tag{7}$$

(iv) We have a $(4, 18)511$ code which leads to

$$M_4(19) \le 1047; \qquad M_4(20) \le 2056; \qquad M_4(21) \le 4094. \tag{8}$$

(v) We have a $(5, 18)210$ code yielding

$$M_5(19) \le 428; \qquad M_5(20) \le 840; \qquad M_5(21) \le 1680. \tag{9}$$

Table 2
Lower and upper bounds, $r = 2$.

n	Lower bound	1st upper bound	2nd upper bound	Previous known upper bound
3	f **7**			**7** B
4	g **6**			**6** G
5	a **6**			**6** G
6	a **8**			**8** G
7	h **14**			**14** F
8	a 17			21 F
9	a 26	32*		34 J
10	i 41	57*		62 J
11	i 67	100*		109 J
12	i 112	177*		191 J
13	i 190	318*		496 J
14	i 326	566*		872 J
15	i 567	1020*		1528 J
16	i 995	1844*		3056 J
17	i 1761	3476*		6112 J
18	i 3141	6430*		11264 F
19	i 5638	13458 (6)	12458**	21824 F
20	i 10179	26710 (6)	25401 (10)	40480 F
21	i 18471	51440 (4)	50342 (10)	80040 F

a [13, Th. 1(iii)].
i [15, Cor. 4].
h [7, Table 4].
f $M_{n-1}(n) = 2^n - 1$ [2, Th. 5].
g [2, Th. 6].
G [2, Th. 6].
B $M_{n-1}(n) = 2^n - 1$ [2, Th. 5].
F [7, Tables 3 and 4].
J [8].
* *Heuristics*.
** *Removing* codewords.

3.3. *Further improvements: Removing codewords*

Perhaps Theorems 2 and 3 can be sharpened, since in practice we observe (with the help of a computer) that the sizes of several codes obtained by Theorem 2 can be reduced by simply *removing* some of their codewords, which are "useless".

As a consequence, we have new upper bounds for some values of n and r, marked by a double star in the tables. The corresponding codes can be found at http://www.infres.enst.fr/~charon/newIdentifyingNcube.html

3.4. *Re-applying* Theorem 2

We can again use Theorem 2 with the newly improved codes obtained in Section 3.3. All the details can be found at http://www.infres.enst.fr/~charon/newIdentifyingNcube.html, and we only develop here case (d), to serve as an example. Note that the various sets Y_i needed in the constructions are obtained via a greedy-type algorithm, and are subject to small improvements.

(a) There is now a $(2, 19)12458$ code, which gives

$$M_2(20) \le 25401; \qquad M_2(21) \le 50342. \tag{10}$$

(b) (b) We have now a $(3, 19)2846$ code, with which we obtain

$$M_3(20) \le 5813; \qquad M_3(21) \le 11477. \tag{11}$$

Table 3
Lower and upper bounds, $r = 3$.

n	Lower bound	1st upper bound	2nd upper bound
4	f **15**	**15** B	
5	ℓ 9	10 H	
6	a **7**	**7***	
7	a **8**	**8***	
8	a 10	13*	
9	a 13	17*	
10	a 18	25*	
11	a 25	36*	
12	a 39	67*	
13	a 61	109*	
14	a 95	180*	
15	a 151	305*	
16	a 241	530*	
17	a 383	901*	
18	a 608	1628*	
19	a 959	3330 (7)	2846**
20	k 1593	6569 (7)	5813 (11)
21	j 2722	13030 (7)	11477 (11)

a [13, Th. 1(iii)].
ℓ Using (1).
f $M_{n-1}(n) = 2^n - 1$ [2, Th. 5].
k [15, Cor. 7].
j [15, Cor. 5].
B $M_{n-1}(n) = 2^n - 1$ [2, Th. 5].
H Using (1).
* *Heuristics*.
** *Removing* codewords.

(c) There is a $(4, 19)835$ code which leads to

$$M_4(20) \leq 1710; \qquad M_4(21) \leq 3358. \tag{12}$$

(d) There is a $(5, 19)326$ code with

$$X_1 = \{27295, 32440, 34030, 72402, 82154, 83370, 86526, 88505, 94930, 95882, 116692, 118724, 120796, 128603, 134214, 142236, 143019, 145498, 147063, 165018, 181588, 191949, 210357, 214527, 221493, 223979, 225622, 226623, 228669, 242104, 245245, 258906, 262456, 265258, 271128, 272468, 274143, 295216, 311330, 312668, 324512, 330041, 336184, 343344, 348668, 349632, 366688, 379552, 379649, 381044, 382699, 390156, 420002, 425365, 426068, 427361, 434532, 449090, 460995, 461154, 462763, 480905, 483322, 486134, 490226, 495812, 497976\},$$

$$Y_1 = \{346531, 489983, 165316, 381224, 163547, 250140, 350922, 207280, 264722, 415128, 429534\} \quad (11 \text{ elements}),$$

$$X_2 = \{88505, 134214, 181588, 480905\},$$

$$Y_2 = \{151364, 228505\} \quad (2 \text{ elements}), \text{ and so}$$

$$M_5(20) \leq 2 \cdot 326 + 11 = 663; \qquad M_5(21) \leq 4 \cdot 326 + 2 \cdot 3 = 1310. \tag{13}$$

Due to time and space limitations, we could not try to remove codewords from these new codes.

3.5. Tables

We give our results for $1 \leq r \leq 5$, $r + 1 \leq n \leq 21$. For some values of r and n, we give two upper bounds, the first one from Section 3.2, the second one from Section 3.3 or 3.4, so that one can see how we used Theorem 2, then possibly removed codewords and possibly reused Theorem 2.

Table 4
Lower and upper bounds, $r = 4$.

n	Lower bound	1st upper bound	2nd upper bound
5	f **31**	**31** B	
6	ℓ **18**	**18***	
7	ℓ 13	14 H	
8	a 9	13 H	
9	a 10	14*	
10	a 12	16*	
11	a 15	20*	
12	a 19	33*	
13	a 27	47*	
14	a 38	76*	
15	a 54	123*	
16	a 77	192*	
17	a 121	305*	
18	a 190	511*	
19	a 304	1047 (8)	835**
20	a 489	2056 (8)	1710 (12)
21	a 792	4094 (8)	3358 (12)

f $M_{n-1}(n) = 2^n - 1$ [2, Th. 5].
ℓ Using (1).
a [13, Th. 1(iii)].
B $M_{n-1}(n) = 2^n - 1$ [2, Th. 5].
H Using (1).
* *Heuristics*.
** *Removing* codewords.

Table 5
Lower and upper bounds, $r = 5$.

n	Lower bound	1st upper bound	2nd upper bound
6	f **63**	**63** B	
7	ℓ 31	32 H	
8	ℓ 19	21 H	
9	ℓ 12	17 H	
10	a 11	16 H	
11	a 12	17*	
12	a 14	22*	
13	a 17	26*	
14	a 21	43*	
15	a 28	64*	
16	a 37	94*	
17	a 53	136*	
18	a 77	210*	
19	a 112	428 (9)	326**
20	a 161	840 (9)	663 (13)
21	a 229	1680 (9)	1310 (13)

f $M_{n-1}(n) = 2^n - 1$ [2, Th. 5].
ℓ Using (1).
a [13, Th. 1(iii)].
B $M_{n-1}(n) = 2^n - 1$ [2, Th. 5].
H Using (1).
* *Heuristics*.
** *Removing* codewords.

We think that there is still room for ameliorations, and that this is a nice field for investigations, where different heuristics (such as tabu search, genetic algorithms, ...) could also be applied and tested.

3.6. Conclusion

By mixing both heuristic and theoretical constructing arguments, we were able to present numerous upper bounds on $M_r(n)$, the smallest possible cardinality of an r-identifying code in F^n: we first used heuristics for constructions of codes, and we then used some of these codes to build new codes with the help of Theorem 2; after that, the computer possibly removed codewords from these codes, and eventually we reapplied Theorem 2. We stopped to apply heuristics when the time/space constraints were too demanding.

There still remains a large, challenging gap between the lower and upper bounds for most of the values of r, n in Tables 1–5.

Acknowledgment

Our warmest thanks to Geoffrey Exoo, Tero Laihonen and Sanna Ranto, who let us use [7,15] at a time when they were only submitted papers. We also wish to thank Tero and Sanna for very helpful remarks.

References

[1] U. Blass, I. Honkala, S. Litsyn, On the size of identifying codes, in: Lecture Notes in Computer Science, vol. 1719, Springer-Verlag, 1999, pp. 142–147.
[2] U. Blass, I. Honkala, S. Litsyn, On binary codes for identification, Journal of Combinatorial Designs 8 (2000) 151–156.
[3] U. Blass, I. Honkala, S. Litsyn, Bounds on identifying codes, Discrete Mathematics 241 (2001) 119–128.
[4] I. Charon, O. Hudry, The noising methods: A generalization of some metaheuristics, European Journal of Operational Research 135 (1) (2001) 86–101.
[5] I. Charon, O. Hudry, A. Lobstein, Identifying codes with small radius in some infinite regular graphs, Electronic Journal of Combinatorics 9 (1) (2002) R11.
[6] G. Exoo, Computational results on identifying t-codes, Preprint, 1999.
[7] G. Exoo, T. Laihonen, S. Ranto, Improved upper bounds on binary identifying codes, IEEE Transactions on Information Theory 53 (2007) 4255–4260.
[8] G. Exoo, T. Laihonen, S. Ranto, New bounds on binary identifying codes, Discrete Applied Mathematics 156 (2008) 2250–2263.
[9] J. Gimbel, B.D. Van Gorden, M. Nicolescu, C. Umstead, N. Vaiana, Location with dominating sets, Congressus Numerantium 151 (2001) 129–144.
[10] I. Honkala, On the identifying radius of codes, in: Proc. Seventh Nordic Combinatorial Conference, Turku, Finland, 1999, pp. 39–43.
[11] I. Honkala, A. Lobstein, On identifying codes in binary Hamming spaces, Journal of Combinatorial Theory, Ser. A 99 (2002) 232–243.
[12] I. Honkala, A. Lobstein, On the complexity of the identification problem in Hamming spaces, Acta Informatica 38 (2002) 839–845.
[13] M.G. Karpovsky, K. Chakrabarty, L.B. Levitin, On a new class of codes for identifying vertices in graphs, IEEE Transactions on Information Theory 44 (1998) 599–611.
[14] T. Laihonen, S. Ranto, Personal communication, 2007.
[15] T. Laihonen, S. Ranto, Codes identifying sets of binary words with large radii, in: Proc. Workshop on Coding and Cryptography 2007, Versailles, France, 2007, pp. 215–224.
[16] J. Moncel, Monotonicity of the minimum cardinality of an identifying code in the hypercube, Discrete Applied Mathematics 154 (2006) 898–899.
[17] S. Ranto, Identifying and locating-dominating codes in binary Hamming spaces, Ph. D Thesis, University of Turku, 2007, pp. 95.
[18] http://www.infres.enst.fr/~lobstein/bibLOCDOMetID.html.