# Two Families of Optimal Identifying Codes in Binary Hamming Spaces


## Sanna Ranto

Turku Centre for Computer Science
Lemminkäisenkatu 14 A
FIN-20520 Turku, Finland

## Iiro Honkala

Department of Mathematics
University of Turku
FIN-20014 Turku, Finland

## Tero Laihonen

Department of Mathematics
University of Turku
FIN-20014 Turku, Finland


The research of this author was supported by the Academy of Finland under grant #46186.



**Turku Centre for Computer Science**
**TUCS Technical Report No 349**
**May 2000**

**ISBN 952-12-0668-3**
**ISSN 1239-1891**

**Abstract**

A motivation for identifying codes comes from quality control in multiprocessor systems, that is, we are able to find faulty processors in such a system with the aid of these codes. In this paper we give a construction of two infinite families of optimal codes, which identify up to two malfunctioning processors in Hamming spaces.

**Keywords:** identifying codes, code construction, Hamming space, binary codes, optimal identifying codes

**TUCS Research Group**
Coding Theory Group

# 1 Introduction

The following problem of locating faulty processors in a multiprocessor system was introduced by Karpovsky, Chakrabarty and Levitin in [12].

Suppose that there are $2^n$ processors with interconnections given by the Hamming space $F_2^n$. Some of the processors are chosen to check themselves and their neighbours at Hamming distance one, and report if there is something wrong. The idea is to choose as few checking processors as possible in such a way that when we see the reports we know immediately where the malfunctioning processors lie.

In a Hamming space this problem has been further examined in [1, 2, 10, 11] and it is also studied in the case where the architecture of the processors is defined by other graphs (for instance, by the square and triangular lattices and the hexagonal grid) in [12, 3, 5, 6, 7, 8]. A closely related problem of locating-dominating sets in graphs is discussed for example in [9, Section 7.3].

Let us now introduce some notations and definitions. As usual we write

$$B_1(x) = \{y \in F_2^n | d(x,y) \leq 1\},$$

where $d(\cdot,\cdot)$ denotes the Hamming distance. We say that $x$ *covers* $y$ if $d(x,y) \leq 1$ (and vice versa). For a code $C$, the codeword neighbourhood of the words $x_1, \ldots, x_k \in F_2^n$ is denoted by

$$I(x_1, \ldots, x_k) = \left( \bigcup_{i=1}^{k} B_1(x_i) \right) \cap C.$$

**Definition 1.** *A code $C \subseteq F_2^n$ is called a $(1, \leq 1)$-identifying code if for all words $x, y \in F_2^n$ $(x \neq y)$, the sets $I(x)$ and $I(y)$ are nonempty and different.*

*If further $I(x,y)$ is nonempty and*

$$I(x,y) \neq I(z,w) \text{ and } I(x) \neq I(z,w)$$

*for all $x, y, z, w \in F_2^n$ such that $\{x,y\} \neq \{z,w\}$, then $C$ is called a $(1, \leq 2)$-identifying code.*

A more general notion of identifying codes can be found from [11], but this definition is enough for our purposes. The smallest cardinalities of $(1, \leq 1)$- and $(1, \leq 2)$-identifying codes of length $n$ are denoted by $M_1^{(\leq 1)}(n)$ and $M_1^{(\leq 2)}(n)$, respectively. A code attaining the smallest cardinality is called *optimal.*

In this paper we construct two infinite sequences of optimal $(1, \leq 2)$-identifying codes. No such optimal families are previously known in Hamming spaces — not even in the case of $(1, \leq 1)$-identifying codes.

## 2 Preliminaries and short optimal codes

Let $C \subseteq F_2^n$ be a code and $x \in F_2^n$ a word such that $|I(x)| \leq 2$. Then there exists a word $y \in F_2^n (y \neq x)$ such that $I(x) \subset I(y)$ and hence $I(x, y) = I(y)$. Thus we have the theorem below. The proof of this theorem can be found in a more general form in [11].

**Theorem 1.** *Assume that* $n \geq 3$. *If* $C$ *is a* $(1, \leq 2)$*-identifying code, then*

$$\forall x \in F_2^n : \quad |I(x)| \geq 3.$$

Let us recall a definition from [4, Chapter 14]. A code $C \subseteq F_2^n$ is a $\mu$-fold 1-covering if all the points in the space $F_2^n$ are covered by at least $\mu$ distinct codewords of $C$. The smallest cardinality of such a code is denoted by $K(n, 1, \mu)$. If every point is covered by exactly $\mu$ codewords we call $C$ a *perfect* $\mu$-fold 1-covering.

By Theorem 1 we get the first part of the next theorem. The following three theorems and Corollary 1 are from [11].

**Theorem 2.** *We have*

$$M_1^{(\leq 2)}(n) \geq K(n, 1, 3) \geq \left\lceil 3\frac{2^n}{n+1} \right\rceil.$$

*Moreover, if* $n \geq 4$ *is even, then*

$$M_1^{(\leq 2)}(n) \geq \left\lceil \frac{(3n-8)2^n}{n^2 - 2n - 2} \right\rceil.$$

**Theorem 3.** $M_1^{(\leq 2)}(n) \leq 2M_1^{(\leq 2)}(n-1)$.

In the sequel we shall need two initial codes in order to provide the two infinite sequences of optimal codes. These initial codes are given next. The optimal code for length five is from [11], but the corresponding code of length seven is new.

**Theorem 4.** *i) There are exactly two nonequivalent perfect* 3*-fold* 1*-coverings of length five.*

*ii) One of these perfect* 3*-fold* 1*-coverings is* $(1, \leq 2)$*-identifying, namely,*

$$\{00100, 00010, 00001, 11000, 10100, 10010, 01100, 01001,$$
$$00011, 11010, 11001, 10101, 01110, 10111, 01111, 11111\}$$

*and the other one is not, namely,*

$$\{00100, 00010, 00001, 10100, 10010, 10001, 01100, 01010,$$
$$01001, 11100, 11010, 11001, 00111, 10111, 01111, 11111\}.$$

**Corollary 1.** $M_1^{(\leq 2)}(5) = 16$.

Denote by $\mathcal{H}_3$ the Hamming code of length seven with the parity check matrix

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Let $C_1 = \mathcal{H}_3 + 1011001$ and $C_2 = \mathcal{H}_3 + 0000100$ be two cosets of $\mathcal{H}_3$. Let further $P_1$ and $P_3$ be the codes obtained by permuting $C_1$ using $(7,3)(4,2)$ (the notation $(i,j)$ means that $i$th and $j$th columns are interchanged) and $(6,3)(4,1)$, respectively. By $P_2$ we denote the code obtained from $C_2$ using the permutation $(1,2)(3,5)$. It is easy to check (with computer) that $U = P_1 \cup P_2 \cup P_3$ is $(1, \leq 2)$-identifying. The following result now follows from Theorem 2.

**Theorem 5.** *The code $U$ is a $(1, \leq 2)$-identifying code of length seven and hence*

$$M_1^{(\leq 2)}(7) = 48.$$

# 3 The construction

In the next theorem we describe a construction which together with the initial codes gives the two infinite families of optimal $(1, \leq 2)$-identifying codes. The construction does not work in the same way for $(1, \leq 1)$-identifying codes as observed in Example 1. Before the theorem it is worth noticing that in a Hamming space there are no cycles of odd length.

**Theorem 6.** *If $C \subseteq F_2^n$ is a $(1, \leq 2)$-identifying code, then also*

$$C' = \{(\pi(u), u, u+v) \mid u \in F_2^n, v \in C\},$$

*where $\pi(u)$ denotes a parity check bit on $u$, is a $(1, \leq 2)$-identifying code of length $2n+1$.*

*Proof.* By Theorem 1 the code $C$ covers each word at least three times, and thus by [4, Theorems 3.4.3 and 14.4.3] the code $C'$ also does. In Hamming spaces it is always true that for all words $x, y$ $(x \neq y)$ : $|B_1(x) \cap B_1(y)| \leq 2$. Since $C'$ covers each word at least three times, and the intersection of three different spheres of radius one contains at most one element, $C'$ is $(1, \leq 1)$-identifying. Moreover, all single words and pairs of words are distinguishable. Indeed, consider the sets $\{x\}$ and $\{y, z\}$, $y \neq z$. Then without loss of generality $y \neq x$, and because $|B_1(y) \cap B_1(x)| \leq 2$, we know that $B_1(y) \cap C$ contains at least one codeword which is not in $B_1(x)$.

Thus we only need to check that all pairs are identified from one another.

Let us divide the words of $F_2^{2n+1}$ into two classes by their first bit and consider the codewords which cover a word in each class. Let $x = (a, u, u + v) \in F_2^{2n+1}$.

I If $a = \pi(u)$ then $I(x) = \{(\pi(u), u, u + c) \mid c \in C, d(c, v) \leq 1\}$.

II If $a \neq \pi(u)$ then $I(x) = A \cup \{(a, u', u + v) \mid d(u', u) = 1, \exists c \in C : u + v = u' + c\}$. Here $A = \{(\pi(u), u, u + v)\}$ if $v \in C$, and $A = \emptyset$ if $v \notin C$.

So in both classes we are interested in codewords $c \in C$ such that $d(c, v) \leq 1$. Namely in the class II the properties $d(u', u) \leq 1$ and $u + v = u' + c$ imply that also $d(v, c) \leq 1$. If $I(x) = \{(b_i, s_i, t_i) \mid i = 1, 2, \ldots, k\}$, then in both cases $I(v) = B_1(v) \cap C = \{s_i + t_i \mid i = 1, 2, \ldots, k\}$.

Suppose there were words $x, y, z$ and $w$ in $F_2^{2n+1}$ such that

$$I(x, y) = I(z, w) \text{ and } \{x, y\} \neq \{z, w\}, x \neq y, z \neq w. \tag{1}$$

If $v_1$, $v_2$, $v_3$ and $v_4$ are $v$'s of $x, y, z$ and $w$ respectively, then in $F_2^n$ by the previous discussion $I(v_1, v_2) = I(v_3, v_4)$. Since $C$ is a $(1, \leq 2)$-identifying code we must have $\{v_1, v_2\} = \{v_3, v_4\}$. We will show that (1) cannot hold. Assume to the contrary that (1) holds.

Because $|I(x)| \geq 3$, we know by (1) that at most one of the sets $I(x) \cap I(z)$ and $I(x) \cap I(y)$ has cardinality one or less. A similar remark applies to $I(y)$, $I(z)$ and $I(w)$. Hence we can without loss of generality assume that $|I(x) \cap I(z)| \geq 2$ and $|I(y) \cap I(w)| \geq 2$. Depending on which class $x$ belongs to, also $z$ belongs to the same class. Indeed, if $x$ is in the class I, then only last $n$ bits are changing in the codewords of $I(x)$, and thus those are the only bits which can change in the codewords of $I(z)$. This means that $z$ belongs to the class I. Also, if $x$ belongs to the class II, then the last $n$ bits do not change at all in $I(x)$ and so it is true also for $I(z)$, and hence $z$ belongs to the class II. Similarly $y$ and $w$ belong to the same class.

If $x = (\pi(u_1), u_1, u_1 + v_1)$ and $y = (\pi(u_2), u_2, u_2 + v_2)$ are words in the class I, then also $z = (\pi(u_3), u_3, u_3 + v_3)$ and $w = (\pi(u_4), u_4, u_4 + v_4)$ are. Since in $I(x)$ and $I(z)$ the codewords begin with the same $n + 1$ bits, we get $u_1 = u_3$. Similarly $u_2 = u_4$. We can assume that $|I(z) \cap I(y)| \geq 1$ (or that $|I(x) \cap I(w)| \geq 1$, which is a symmetric case): otherwise $I(z) = I(x)$ and $I(w) = I(y)$ and hence $z = x$ and $w = y$. Since $z$ and $y$ are both words in the class I, we get $u_2 = u_1$. The fact that $\{v_1, v_2\} = \{v_3, v_4\}$ now implies that $\{x, y\} = \{z, w\}$.

Assume $x = (\pi(u_1) + 1, u_1, u_1 + v_1)$ and $y = (\pi(u_2) + 1, u_2, u_2 + v_2)$ and so also $z = (\pi(u_3) + 1, u_3, u_3 + v_3)$ and $w = (\pi(u_4) + 1, u_4, u_4 + v_4)$ are words in the class II. Since in $I(x)$ and $I(z)$ the codewords end with the same $n$ bits as $x$ and $z$ we get $u_1 + v_1 = u_3 + v_3$, and similarly $u_2 + v_2 = u_4 + v_4$. If now $v_1 = v_3$ and $v_2 = v_4$ we are done, since then $x = z$ and $y = w$. Suppose therefore that $v_2 = v_3$ and $v_1 = v_4$. As in the previous case, we can assume that $|I(z) \cap I(y)| \geq 1$. Now the last $n$ bits must be the same in $I(z)$ and $I(y)$, and thus $u_3 + v_3 = u_2 + v_2$ and we get $u_3 = u_2$, i.e., $y = z$. The word $z$ cannot cover the whole $I(x)$, otherwise $z = x$, since $|I(x)| \geq 3$ and the intersection of three different spheres of radius one has cardinality at most one. This would imply that $x = z = y$. So $w$ must cover at least one word from $I(x)$ which implies $u_1 + v_1 = u_4 + v_4$ and now $u_1 = u_4$, i.e., $x = w$. Therefore $\{x, y\} = \{z, w\}$.

Suppose finally that $x = (\pi(u_1), u_1, u_1 + v_1)$ is a word in the class I and $y = (\pi(u_2)+1, u_2, u_2+v_2)$ is a word in the class II. Now $z = (\pi(u_3), u_3, u_3+v_3)$ is a word in the class I and $w = (\pi(u_4) + 1, u_4, u_4 + v_4)$ is a word in the class II. Since the first $n + 1$ bits are the same in the codewords of $I(x)$ and $I(z)$ we must have $u_1 = u_3$. In the codewords of $I(y)$ and $I(w)$ the last $n$ bits are the same, so $u_2 + v_2 = u_4 + v_4$. If now $v_1 = v_3$ and $v_2 = v_4$ we are done since then $x = z$ and $y = w$. Suppose therefore that $v_1 = v_4$ and $v_2 = v_3$. As in the previous case, $|I(z) \cap I(x)| = 2$: otherwise $z = x$, $v_1 = v_3$ and thus $v_1 = v_2 = v_3 = v_4$ which together with the equality $u_2 + v_2 = u_4 + v_4$ implies that $u_2 = u_4$ and hence $\{x, y\} = \{z, w\}$. Also $y \neq w$, since otherwise $v_2 = v_4 = v_1 = v_3$ and again $\{x, y\} = \{z, w\}$; consequently $d(y, w) = 2$ and $|I(y) \cap I(w)| = 2$. Now $|I(x) \cap I(w)| = 1$ and $|I(z) \cap I(y)| = 1$ (these cannot be greater than one, since $x$ and $w$, $y$ and $z$ belong to different classes). From this we also get $|I(x)| = |I(y)| = |I(z)| = |I(w)| = 3$.

Let $c_1 = (\pi(s_1), s_1, s_1 + t_1)$ be the unique codeword in $I(x) \cap I(w)$, and $c_2 = (\pi(s_2), s_2, s_2 + t_2)$ be the unique codeword in $I(y) \cap I(z)$. Since all the codewords in $I(x)$ begin with $(\pi(u_1), u_1, \dots)$ and all the codewords in $I(w)$ end in $(\dots, u_4 + v_4)$, we know that $c_1 = (\pi(u_1), u_1, u_4 + v_4)$. Similarly, $c_2 = (\pi(u_3), u_3, u_2 + v_2)$. Hence $c_1 = c_2$, and $I(y) = I(w)$ and $I(x) = I(z)$. Therefore $x = z$ and $y = w$, completing the proof.

□

**Corollary 2.** $M_1^{(\leq 2)}(2n + 1) \leq 2^n M_1^{(\leq 2)}(n)$.

**Corollary 3.**

$$\text{For } n \geq 1: \quad M_1^{(\leq 2)}(3 \cdot 2^n - 1) = 2^{3 \cdot 2^n - n - 1}.$$
$$\text{For } n \geq 3: \quad M_1^{(\leq 2)}(2^n - 1) = 3 \cdot 2^{2^n - n - 1}.$$

*Proof.* By Theorem 4 we know that $M_1^{(\leq 2)}(5) = 16$. Using Corollary 2 recursively and the lower bound from Theorem 2 we get the first equation. Similarly by Theorem 5 we get the second claim. □

| $n$ | |
|---|---|
| 4 | a 11 b |
| 5 | a 16 c |
| 6 | a 30–32 d |
| 7 | a 48 e |
| 8 | a 90–96 d |
| 9 | a 154–176 f |
| 10 | a 289–352 d |
| 11 | a 512 f |
| 12 | a 972-1024 d |
| 13 | a 1756–2048 d |
| 14 | a 3356–4096 d |
| 15 | a 6144 f |

Table 1: Bounds on $M_1^{(\leq 2)}(n)$.

| | |
|---|---|
| a | Theorem 2 |
| b | [11, Theorem 8] |
| c | Corollary 1 |
| d | Theorem 3 |
| e | Theorem 5 |
| f | Corollaries 2 and 3 |

By Corollary 3 we have an optimal $(1, \leq 2)$-identifying code for all lengths $n \geq 5$ for which there also exists a perfect 3-fold 1-covering [4, Theorem 14.2.4].

**Example 1.** In this example we show that in the case of $(1, \leq 1)$-identifying codes the construction mentioned above does not directly apply. Let us take the $(1, \leq 1)$-identifying code $C = \{000, 110, 101, 011\}$ and construct the code $C'$ of length seven as above. Then $I(0000011) = I(1000011)$ and therefore $C'$ is not $(1, \leq 1)$-identifying.

**Example 2.** The code $C = \{00, 01, 10, 11\}$ of length two is a perfect 3-fold 1-covering, but not $(1, \leq 2)$-identifying, since $I(00, 11) = I(00, 10)$. By the construction of Theorem 6 we get a perfect 3-fold 1-covering $C'$ of length 5. This code is not $(1, \leq 2)$-identifying, because $c_1 = 00000$, $c_2 = 00010$, $c_3 = 00011$, $c_4 = 00010$ are codewords in $C'$ and $I(c_1, c_2) = I(c_3, c_4) = \{c_1, c_2, c_3, c_4\}$.

**Example 3.** Let $\mathcal{H}$ be the Hamming code of length $n = 2^m - 1$, and $c \in \mathcal{H}$ such that $w(c) = 3$. Let the coordinates of nonzero components of $c$ be $i, j$ and $k$. Denote $e_s = (0, \dots, 0, 1, 0, \dots, 0) \in F_2^n$ where the only 1 is in the $s$-th coordinate. Now the code $C = \mathcal{H} + e_i \cup \mathcal{H} + e_j \cup \mathcal{H} + e_k$ is a perfect 3-fold 1-covering, but $I(\mathbf{0}, c) = I(e_i, c + e_i)$. Thus $C$ is not a $(1, \leq 2)$-identifying code. So at the lengths of Hamming codes we get a sequence of perfect 3-fold 1-coverings which are not $(1, \leq 2)$-identifying codes.

It is an obvious research problem to try to find optimal families of codes identifying up to $l \geq 3$ words, but in view of the proof of Theorem 6 the arguments probably become much harder.

# References


[1] U. Blass, I. Honkala, S. Litsyn: Bounds on identifying codes, Discrete Mathematics, to appear.

[2] U. Blass, I. Honkala, S. Litsyn: On binary codes for identification, Journal of Combinatorial Designs 8, 151-156, 2000.

[3] G. Cohen, S. Gravier, I. Honkala, A. Lobstein, M. Mollard, C. Payan, G. Zémor: Improved identifying codes for the grid, Electronic Journal of Combinatorics, Comments to 6(1), R19, 1999.

[4] G. Cohen, I. Honkala, S. Litsyn and A. Lobstein, *Covering Codes*, Amsterdam: Elsevier, 1997.

[5] G. Cohen, I. Honkala, A. Lobstein, G. Zémor: New bounds for codes identifying vertices in graphs, Electronic Journal of Combinatorics 6(1), R19, 1999.

[6] G. Cohen, I. Honkala, A. Lobstein, G. Zémor: On codes identifying vertices in the two-dimensional square lattice with diagonals, IEEE Transactions on Computers, to appear.

[7] G. Cohen, I. Honkala, A. Lobstein, G. Zémor: Bounds for codes identifying vertices in the hexagonal grid, SIAM J. Discr. Math., to appear.

[8] G. Cohen, I. Honkala, A. Lobstein, G. Zémor: On identifying codes, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Proceedings of the DIMACS Workshop on Codes and Association Schemes, November 9–12, 1999, submitted.

[9] T. W. Haynes, S. T. Hedetniemi, P. J. Slater: *Fundamentals of Domination in Graphs*. Marcel Dekker, 1998.

[10] I. Honkala, On the identifying radius of codes, *Proceedings of the Seventh Nordic Combinatorial Conference*, Turku, pp. 39–43, 1999.

[11] I. Honkala, T. Laihonen, S. Ranto, On codes identifying sets of vertices in Hamming spaces, Designs, Codes and Cryptography, submitted.

[12] M.G. Karpovsky, K. Chakrabarty, L. B. Levitin: On a new class of codes for identifying vertices in graphs, IEEE Trans. Inform. Theory 44:599–611, 1998.

**Turku Centre for Computer Science**
**Lemminkäisenkatu 14**
**FIN-20520 Turku**
**Finland**

http://www.tucs.abo.fi

**University of Turku**
- **Department of Mathematical Sciences**

**Åbo Akademi University**
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

**Turku School of Economics and Business Administration**
- **Institute of Information Systems Science**