# On Codes Identifying Sets of Vertices in Hamming Spaces

**Iiro Honkala**
**Tero Laihonen**
**Sanna Ranto**

Turku Centre for Computer Science

TUCS Technical Reports

No 331, February 2000

# On Codes Identifying Sets of Vertices in Hamming Spaces


## 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.

## Sanna Ranto

Turku Centre for Computer Science
Lemminkäisenkatu 14 A, 4th floor
FIN-20520 Turku, Finland



**Turku Centre for Computer Science**
**TUCS Technical Report No 331**
**February 2000**

**ISBN 952-12-0617-9**
**ISSN 1239-1891**

**Abstract**

A code $C \subseteq F_2^n$ is called $(t, \leq 2)$-identifying if for all the words $x$, $y$ ($x \neq y$) and $z$ in a Hamming space the sets $(B_t(x) \cup B_t(y)) \cap C$ and $B_t(z) \cap C$ are nonempty and different. Constructions of such codes and a lower bound on the cardinality of these codes are given. The lower bound is shown to be sharp in some cases. We also discuss a more general notion of $(t, \mathcal{F})$-identifying codes and introduce weakly identifying codes.

**Keywords:** identifying code, Hamming space, binary codes, sets of vertices, hypercube

**TUCS Research Group**
Coding Theory Group

# 1 Introduction

Let $F_2$ denote the binary field. Any subset $C$ of the Cartesian product $F_2^n$ with at least two elements is called a binary (block) code of length $n$. The Hamming distance $d(x,y)$ between any two distinct vectors $x$ and $y$ is the number of coordinates in which they differ. We denote

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

and

$$S_t(x) = \{y \in F_2^n | d(x,y) = t\}.$$

The Hamming weight $w(x)$ of $x$ is defined as $d(x,0)$.

The following problem was introduced in [11]. Assume that $2^n$ processors are arranged in the nodes of an $n$-dimensional hypercube. A processor can check itself and all the neighbours in the hypercube within Hamming distance $t \geq 1$, and reports YES to a central controller if everything is all right and NO if there is something wrong in the processor itself or one of these neighbours. We wish to find a code $C$ consisting of some of the nodes in the hypercube such that if there are problems in at most $l$ processors, the central controller immediately knows where the problems are when it sees the reports of the codewords.

The problem where we know that at most one processor is malfunctioning in a binary Hamming space has been studied in [11, 1, 2]. A more general setting for faulty processors in a *graph* was introduced in [11] and has been studied in, e.g., [3, 5, 6, 7]. For a closely related problem, see also [9, Section 7.3].

Let us now state the problem rigorously. The cardinality of a set $X$ is denoted by $|X|$. Let $\mathcal{P}(X)$ be the set of all subsets (the power set) of $X$. For any element $X \in \mathcal{P}(F_2^n)$ define its neighbourhood in $F_2^n$ by

$$N_t(X) = \bigcup_{x \in X} B_t(x).$$

Let $\mathcal{F} \subseteq \mathcal{P}(F_2^n)$ and $C \subseteq F_2^n$. If for all $X \in \mathcal{F}$ the sets

$$I_t(X) = N_t(X) \cap C$$

are different (where $I_t(\emptyset) = \emptyset$), then the code $C$ is called a $(t, \mathcal{F})$-*identifying code.*

If $t = 1$ we usually omit the subscript $t$ and furthermore we write $I(\{x_1, \ldots, x_k\}) = I(x_1, \ldots, x_k)$. In the case when $\mathcal{F} = \mathcal{F}_{\leq l} := \{X \in \mathcal{P}(F_2^n) | |X| \leq l\}$, we call a $(t, \mathcal{F})$-identifying code a $(t, \leq l)$-identifying code.

If $\mathcal{F} = \mathcal{F}_s := \{X \in \mathcal{P}(F_2^n) || X| = s\}$, we call the code $(t,s)$-identifying. In this paper we mainly focus on these two types of identifying codes. The smallest cardinalities of $(t,s)$-identifying and $(t, \leq l)$-identifying codes are denoted by $M_t^{(s)}(n)$ and $M_t^{(\leq l)}(n)$, respectively. A code realizing the smallest cardinality is called *optimal*.

If $x, y \in F_2^n$ and $d(x,y) \leq t$, we say that $x$ $t$-covers $y$ (and vice versa); when $t$ is clear from the context, it can be omitted.

The sets of vertices $X \in \mathcal{F}$ that we try to identify are called *fault patterns*.

# 2 Lower bounds

Because a code $C$ has only $2^{|C|}$ different subsets and there are $\sum_{i=0}^{l} \binom{2^n}{i}$ distinct fault patterns to be identified by a $(t, \leq l)$-identifying code, we can deduce the following result.

**Theorem 1.** *[11] Let $C$ be a $(t, \leq l)$-identifying code. Then*

$$|C| \geq \left\lceil \log_2 \sum_{i=0}^{l} \binom{2^n}{i} \right\rceil .$$

In the next theorem we improve for $(t, \leq l)$-identifying codes this lower bound. The derived lower bound is shown to be optimal in some cases.

First we prove some general properties of certain identifying codes.

**Lemma 1.** *Let $C$ be a $(t, \leq l)$-identifying code where $l \geq 2$. For any two different fault patterns $X = \{x_1, \ldots, x_k\} \in \mathcal{F}_{\leq l}$ and $Y = \{y_1, \ldots, y_v\} \in \mathcal{F}_{\leq l}$ where $|X \cup Y| \leq l$, we have $I_t(X) \not\subseteq I_t(Y)$ and $I_t(Y) \not\subseteq I_t(X)$. In particular, if $\mathrm{B}_t(x) \cap C \subseteq \mathrm{B}_t(y) \cap C$, for some $x, y \in F_2^n$, then $x = y$.*

*Proof.* If for some $X$ and $Y$, $I_t(X) \subseteq I_t(Y)$, then

$$I_t(X \cup Y) = I_t(Y),$$

which leads to a contradiction, since $|X \cup Y| \leq l$. □

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

$$\forall x \in F_2^n : \quad |I_t(x)| = |\mathrm{B}_t(x) \cap C| \ \geq 3.$$

*Proof.* Every $x \in F_2^n$ is covered by at least one codeword. If $x$ is covered only by $c \in C$, then for any $y \neq x$ such that $d(y,c) \leq t$ yields a contradiction with

Lemma 1, because $I_t(x) \subseteq I_t(y)$. Suppose therefore that $x$ is only covered by $c_1 \in C$ and $c_2 \in C$, $c_1 \neq c_2$. Obviously $1 \leq d(c_1, c_2) \leq 2t$. By [4, Theorem 2.4.8.], $|B_t(c_1) \cap B_t(c_2)| \geq \binom{2t}{t} \geq 2$. Take $y \in B_t(c_1) \cap B_t(c_2)$, $y \neq x$. Then $I_t(x) \subseteq I_t(y)$ contradicting Lemma 1. □

**Corollary 1.** *If $C$ is a $(1, \leq 2)$-identifying code of length $n \geq 3$, then*

$$\forall x, y \in F_2^n, x \neq y : \ |I(x,y)| \geq 4.$$

*Proof.* By Theorem 2, $|I_t(x)| \geq 3$ and $|I_t(y)| \geq 3$. Because $C$ is $(1,1)$-identifying, $I_t(x) \neq I_t(y)$ and therefore $|I_t(x,y)| = |I_t(x) \cup I_t(y)| \geq 4$. □

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

The smallest cardinality of such a code is denoted by $K(n, \mu, t)$.

**Theorem 3.** *We have*

$$M_t^{(\leq 2)}(n) \geq K(n, 3, t).$$

*In particular,*

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

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

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

*Proof.* Let $C$ be a $(t, \leq 2)$-identifying code. By Theorem 2 we know that all the words of the ambient space are $t$-covered by at least three codewords of $C$. Hence $M_t^{(\leq 2)}(n) \geq K(n, 3, t)$. Obviously $3 \cdot 2^n/(n+1)$ is a lower bound on $K(n, 3, 1)$. The last assertion follows from [8, Corollary 2]:

$$K(n, 3, 1) \geq \frac{(3(n-3)+1)2^n}{(n-3)(n+1)+1}$$

for even $n$ where $n \geq 4$. □

Evidently, the result $M_t^{(\leq l)}(n) \geq K(n, 3, t)$ also holds for $l > 2$, because a $(t, \leq l)$-identifying code is a $(t, \leq 2)$-identifying code as well. Next we provide a nonexistence result. The covering radius of a code $C$ of length $n$ is defined by $R = \max_{x \in F_2^n} \min_{c \in C} d(x, c)$.

**Theorem 4.** *Let $r(n, K)$ denote the smallest covering radius among all binary codes of length $n$ and cardinality $K$. Let $l < 2^n$. If $t \geq r(n, l)$, then there does not exist a $(t, \leq l)$-identifying code of length $n$.*

*Proof.* Let $C(l)$ be a code realizing the value $r(n, l)$. Thus the Hamming spheres of radius $t$ centered at the codewords of radius at least $r(n, l)$ cover the entire space. Take $x \in F_2^n$ such that $C(l) \neq x + C(l)$ (this is always possible when $C(l) \neq F_2^n$). Consequently, the sets $C(l)$ and $x + C(l)$ of cardinality $l$ cannot be identified by any code $C \subseteq F_2^n$. $\square$

Values of $r(n, K)$ can be found for instance from [4].

**Corollary 2.** *There is no $(1, \leq 2)$-identifying code of length less than four and no $(2, \leq 2)$-identifying code of length less than six.*

# 3 Constructions

## 3.1 Using a union of cosets of a linear code

In this section we state results concerning code constructions. For results concerning $(1, \leq 1)$-identifying and $(2, \leq 1)$-identifying codes consult [1, 2].

Denote by $k(n, R)$ the smallest dimension of a binary linear code of length $n$ and covering radius at most $R$.

**Theorem 5.** *If $l \leq (2^{n-k(n,1)} - 1)/2$, then*

$$M_1^{(\leq l)}(n) \leq (2l + 1)2^{k(n,1)}.$$

*Proof.* Let $C_1 \subseteq F_2^n$ be a code attaining the dimension $k(n, 1)$ and $C$ the code consisting of $2l + 1$ distinct cosets of $C_1$. Consequently, every word $x \in F_2^n$ is 1-covered by at least $2l + 1$ different codewords of $C$.

Take any two distinct fault patterns $X$ and $Y$ ($X \neq Y$) such that the cardinalities of $X$ and $Y$ are at most $l$. Without loss of generality we may assume that there exists $x \in X$ such that $x \notin Y$. The word $x$ is 1-covered by at least $2l + 1$ codewords. It is not difficult to check that if a word $z$ 1-covers at least three of the words in $I(x)$, then $x = z$ (this also holds in general — not only for $I(x)$). Therefore, any of the words in $Y$ can only cover at most two of the words in $I(x)$. Since $x \notin Y$ and $|I(x)| \geq 2l + 1 > 2|Y|$, the two different fault patterns $X$ and $Y$ are distinguishable from each other. $\square$

**Corollary 3.** *If $n = 2^r - 1 \geq 7$ and $l \leq n/2$, then*

$$M_1^{(\leq l)}(n) \leq \frac{(2l + 1)2^n}{n + 1}.$$

*Proof.* Let $C_1$ (in the proof above) be the $[2^r-1, 2^r-r-1]$ Hamming code ($r > 2$) which has covering radius one. $\square$

**Example 1.** Taking $n = 15$ and $l = 2$ we get the upper bound $M_1^{(\leq 2)}(15) \leq 10240$.

## 3.2 Direct Sum

The *direct sum* of $C_1 \subseteq F_2^{n_1}$ and $C_2 \subseteq F_2^{n_2}$ is

$$C_1 \oplus C_2 = \{(c_1, c_2) | c_1 \in C_1, c_2 \in C_2\} \subseteq F_2^{n_1+n_2}.$$

We prove that $M_1^{(\leq 2)}(n) \leq 2M_1^{(\leq 2)}(n-1)$. The analogous result is not known for $(1, \leq 1)$-identifying codes, see [1].

**Theorem 6.** *[1] Assume that $C$ is a $(1, \leq 1)$-identifying code. The direct sum $C \oplus \{0,1\}$ is $(1, \leq 1)$-identifying if and only if $d(c, C \setminus \{c\}) = 1$ for all $c \in C$.*

**Theorem 7.** *If $C$ is a $(1, \leq 2)$-identifying code then the direct sum $D = C \oplus \{0,1\}$ is also a $(1, \leq 2)$-identifying code.*

*Proof.* By Theorem 2 and Theorem 6 we know that $D$ is $(1, \leq 1)$-identifying.

We can think that the words of $F_2^{n+1} = F_2^n \oplus \{0,1\}$ belong to 0- or 1-layers depending on their last bit. Because $C$ is a $(1, \leq 2)$-identifying code, words and the pairs of words in the same layer are identified from each other.

If $c = c'0 \in D$ is a codeword on the 0-layer, then $c'1 \in D$, $c'1 \in \mathrm{B}_1(c)$ and no other codeword from the 1-layer belongs to $\mathrm{B}_1(c)$. If $c \notin D$ is a word on the 0-layer then in $\mathrm{B}_1(c)$ there is no codeword on the 1-layer. A similar remark holds for codewords on the 1-layer.

A pair on the 1-layer (resp. 0-layer) is covered by at least four codewords by Corollary 1. A pair on the 0(1)-layer can cover at most two codewords from the 1(0)-layer. So pairs from different layers are identified from each other. A single word on the 0(1)-layer covers at most one word on the 1(0)-layer so a word and a pair on different layers are also identified from each other.

Let $a, b, e$ and $f$ be different arbitrary words in $F_2^{n+1}$ such that $a \neq b$ are on the 1-layer and $e \neq f$ are on the 0-layer. Let us consider the pairs $\{a, e\}$ and $\{b, f\}$. Assume that these pairs were covered by the same codewords. By Lemma 1 for $C$, there must be exactly one codeword $x_1$ on the 1-layer which is in $\mathrm{B}_1(a) \cap D$ but not in $\mathrm{B}_1(b) \cap D$. Similarly, there must be exactly one codeword $x_2$ in the 1-layer which is in $\mathrm{B}_1(b) \cap D$ but not in $\mathrm{B}_1(a) \cap D$.

Furthermore, the words $e$ and $f$ must be codewords and $e$ must cover $x_2$ and $f$ cover $x_1$. In particular, $d(a,b) = 1$ or 2. By symmetry, also $a$ and $b$ are codewords, and $d(e,f) = 1$ or 2 and consequently, $d(x_1,x_2) = 1$ or 2. For every $z$ we denote by $z'$ the word obtained by puncturing the last coordinate of $z$. Suppose first that $d(a,b) = 1$. Then $d(x_1,x_2) = 1$, and the only codewords of $C$ that cover $a'$ are $a'$, $b'$ and $x'_1$, and the only codewords of $C$ that cover $b'$ are $a'$, $b'$, $x'_2$. But then $C$ would not be $(1,\leq 2)$-identifying, because the pairs $\{a',x'_2\}$ and $\{b',x'_2\}$ would be covered by exactly the same set of codewords. Suppose then that $d(a,b) = 2$. Then the intersection $B_1(a) \cap B_1(b)$ consists of exactly two points, say $z_1$ and $z_2$, and hence $x_1 = a$ and $x_2 = b$. Now the only codewords of $C$ that cover $a'$ are $a'$, $z'_1$ and $z'_2$ and the only codewords of $C$ that cover $b'$ are $b'$, $z'_1$ and $z'_2$. Again, $C$ would not be $(1,\leq 2)$-identifying, because the pairs $\{a',z'_1\}$ and $\{b',z'_1\}$ would be covered by exactly the same set of codewords.

Clearly $\{a,e\}$ and $\{a,f\}$ cannot be covered by the same codewords of $D$, either. The same is true for $\{a,e\}$ and $\{b,e\}$.

Let now $g \neq f$ be a word on the 0-layer. Let us consider pairs $\{a,e\}$ and $\{f,g\}$. By Theorem 2 the word $a$ is covered by at least three codewords on the 1-layer. The pair $\{f,g\}$ can cover only two words on the 1-layer. So these pairs cannot cover the same set of codewords. The situation does not change if we consider $\{a,e\}$ and any single point from the 0-layer. The same holds for $\{a,e\}$ and any single point or pair from the 1-layer. $\square$

**Corollary 4.**

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

### 3.3 Short codes

In the next theorems we shall obtain optimal $(1,\leq 2)$-identifying codes of lengths four and five.

**Theorem 8.** $M_1^{(\leq 2)}(4) = 11$.

*Proof.* The code

$$C = \{0000, 0010, 0001, 1100, 1010, 1001, 0101, 0011, 1110, 1101, 0111\}$$

gives the upper bound $M_1^{(\leq 2)}(4) \leq 11$. One way of verifying this is by looking at the following matrix $A$ where the first row corresponds to the codeword 0000 and the second to 0010 etc., and the columns correspond to all words $x = (x_1,x_2,x_3,x_4)$ of the space $F_2^4$ in the order of integers $I$ $(0 \leq I \leq 15)$ when they are expressed as $I = \sum_{i=1}^{4} x_i 2^{(i-1)}$. The entry of the matrix is one if the codeword corresponding to that row 1-covers the word of the corresponding column, otherwise the entry is zero. Thus

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

By comparing single columns and the logical OR of all pairs of columns we get the claim.

By Theorem 3 we get $M_1^{(\leq 2)}(4) \geq 11$. □

In the following proof $\overline{x}$ denotes the complement of $x \in F_2^n$, i.e., $\overline{x} = \mathbf{1}+x$.

**Theorem 9.** *i) If $C$ is a perfect 3-fold 1-covering of length five, then for every $x \in F_2^5$, either $x \in C$ or $\overline{x} \in C$ but not both.*

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

*iii) One of these perfect 3-fold 1-coverings is $(1, \leq 2)$-identifying and the other one is not.*

*Proof.* Without loss of generality, assume that $\mathbf{0} \notin C$, 00100, 00010, 00001 $\in C$, 01000, 10000 $\notin C$. Because all the vectors of weight one are covered by exactly three codewords, the number of codewords of weight two must be six.

To prove i), it suffices to notice that if $\mathbf{1} \notin C$, then the same argument shows that there are three codewords of weight four and six of weight three, and altogether the number of codewords would be more than sixteen, which is the cardinality of the perfect 3-fold 1-covering of length five.

Having proved i), it suffices to consider which vectors of weight two are in $C$. Because 10000 is covered by three codewords, three of the six codewords of weight two have 1 in the first coordinate, and the same is true for the second coordinate.

If none of the codewords of weight two have 1 in both the first two coordinates, there is a unique choice for $C$:

$$\begin{array}{l} 00100 \\ 00010 \\ 00001 \\ 10100 \\ 10010 \\ 10001 \\ 01100 \\ 01010 \\ 01001 \end{array}.$$

The seven other codewords are the complements of the non-codewords of weight at most two. A routine verification shows that it is indeed a perfect 3-fold 1-covering. We notice that $C_1 = \{00100, 10100, 01100, 11100\} \subseteq C$. Clearly $C$ is not $(1, \leq 2)$-identifying, because $I(00100, 10100) = C_1 = I(01100, 11100)$.

The other possibility is that $11000 \in C$. Apart from the order of the coordinates, there is again a unique way of choosing the codewords of weight two in such a way that every vector of weight one is covered by exactly three codewords:

$$\begin{array}{l} 00100 \\ 00010 \\ 00001 \\ 11000 \\ 10100 \\ 10010 \\ 01100 \\ 01001 \\ 00011 \end{array}.$$

A routine verification shows that the resulting code $C = \{00100, 00010, 00001, 11000, 10100, 10010, 01100, 01001, 00011, 11010, 11001, 10101, 01110, 10111, 01111, 11111\}$ is a perfect 3-fold 1-covering and $(1, \leq 2)$-identifying, and hence, in particular, inequivalent to the code constructed earlier. □

**Corollary 5.** *There is a unique* $(1, \leq 2)$*-identifying code (up to equivalence) attaining the bound* $M_1^{(\leq 2)}(5) = 16$.

If $\mu \geq 3$, a perfect $\mu$-fold 1-covering of length $n$ can find one faulty processor [1] and *detect* a larger number of malfunctioning processors, because then $|I(X)| \geq 2\mu - 2 > \mu = |I(z)|$ for any $z \in F_2^n$ and $X \in \mathcal{P}(F_2^n)$ where $|X| \geq 2$. This property is clearly not shared by all codes, for example, the $(1, \leq 1)$-identifying code $C = \{x \in F_2^4 | w(x) \in \{0, 2, 3\}\}$ gives $I(1000) = I(0000, 1000)$.

| $n$ | $M_1^{(\le 2)}(n)$ |
|---|---|
| 4 | 11 |
| 5 | 16 |
| 6 | 30–32 |
| 7 | 48–64 |
| 15 | 6144–10240 |

Table 1: Bounds on $(1, \le 2)$-identifying codes

**Corollary 6.** $30 \le M_1^{(\le 2)}(6) \le 32$.

*Proof.* By Theorem 7 we conclude that $D := C \oplus \{0, 1\}$, where $C$ is the identifying code of the proof of Theorem 9, is a $(1, \le 2)$-identifying code. Theorem 3 implies the lower bound. □

The code $D$ can be shown (by computer) to be also $(2, \le 2)$-identifying. In fact, the subcode

| | | | | |
|---|---|---|---|---|
| 000100 | 000010 | 101000 | 100100 | 011000 |
| 010010 | 000110 | 110100 | 110010 | 101010 |
| 011100 | 101110 | 011110 | 001001 | 000101 |
| 000011 | 110001 | 101001 | 100101 | 011001 |
| 010011 | 000111 | 110101 | 110011 | 101011 |
| 111111 | | | | |

of 26 codewords turns out to be sufficient for the $(2, \le 2)$-identification.

Results concerning $M_1^{(\le 2)}(n)$ when $n$ is small are summarized in Table 1.

# 4 On $(1, 2)$-identifying codes

## 4.1 Basics

Next we study the $(1, 2)$-identifying codes. Naturally, every $(1, \le 2)$-identifying code is $(1, 2)$-identifying as well. After proving some basic properties of such codes we show in the next subsection that the *optimal* $(1, 2)$-identifying code of length four has only ten codewords (compare with Theorem 8).

**Theorem 10.** *Let $C$ be a $(1, 2)$-identifying code. If there exists $w \in F_2^n$ such that $|I(w)| = 0$, then for all $x \in F_2^n, x \ne w : |I(x)| \ge 3$.*

*Proof.* If $|I(x)| \leq 2$ for some $x \in F_2^n$, then there is $z \in F_2^n, z \neq x$, such that $I(x) \subseteq I(z)$ and thus $I(z,w) = I(z,x)$ This is a contradiction. □

**Theorem 11.** *Let $C$ be a $(1,2)$-identifying code. Then*

$$\forall x, y,\ x \neq y :\ |I(x,y)| \geq 2. \tag{1}$$

*In* (1) *the equality can hold only if there are some words $x, y$, such that $|I(x)| = |I(y)| = 1$ and $d(I(x), I(y)) \geq 3$.*

*Proof.* If for some word $w$, $|I(w)| = 0$ then by Theorem 10 $|I(x,y)| \geq 3$ for all $x, y$, $x \neq y$. For all words $x, y \in F_2^n, x \neq y$ we have $I(x) \neq I(y)$, otherwise for any word $z$ we get $I(z,x) = I(z,y)$. Hence $|I(x,y)| \geq 2$ for all $x, y \in F_2^n$, $x \neq y$.

Suppose now that $I(y) \subset I(x)$, and $y$ is covered by one codeword and $x$ by two. Then there exists $z$ ($x \neq z \neq y$) such that $I(x) \subset I(z)$ and thus $I(x,z) = I(y,z)$.

Suppose then that $I(x) = \{a\}$ and $I(y) = \{b\}$, where $a \neq b$. If $d(a,b) = 1$, then $I(x,a) = I(y,a)$. If $d(a,b) = 2$, then there exists $z$ such that $\{a,b\} \subset \mathrm{B}_1(z)$, and then $I(z,x) = I(z,y)$. This completes the proof. □

## 4.2 An optimality proof

**Lemma 2.** *Assume that $C$ is a $(1,2)$-identifying code of length four and* $1000, 0100, 0000 \notin C$. *Then* $0011, 0111, 1011 \in C$.

*Proof.* If $0011 \notin C$, then $I(0011, 1111) = I(0000, 1111)$ and $C$ is not $(1,2)$-identifying. Similarly if $0111 \notin C$ then $I(0011, 1011) = I(0000, 1011)$, and if $1011 \notin C$ then $I(0011, 0111) = I(0000, 0111)$. □

**Lemma 3.** *If $C \subseteq F_2^4$ is $(1,2)$-identifying, then no code equivalent to $D_1 = \{1000,\ 0100, 1010, 0110\}$ or $D_2 = \{1000, 0100, 0111, 1011\}$ is contained in $F_2^4 \setminus C$.*

*Proof.* If $D_1 \subseteq F_2^4 \setminus C$, then $I(1001, 0111) = I(0101, 1011)$: both contain all the codewords ending in 1, but none ending in 0.

If $D_2 \subseteq F_2^4 \setminus C$, then $I(1000, 0111) = I(0100, 1011)$: namely both contain all the codewords of even weight but none of odd weight. □

**Lemma 4.** *If $C \subseteq F_2^4$ is a $(1,2)$-identifying code of cardinality nine, then no code equivalent to $\{1000, 0100, 0010\}$ is contained in $F_2^4 \setminus C$.*

*Proof.* Assume that 1000, 0100, 0010 $\notin C$. Let $N_i$ denote the number of non-codewords of weight $i$.

**Case 1:** $0001 \notin C$: Then $0000 \in C$, otherwise $N_2 = N_3 = 0$ by Lemma 2, and $|C| \geq 10$. If $x \notin C$ has weight two then Lemma 2 for $x + C$ implies $1111 \in C$. By Lemma 3, $N_2 \leq 2$ and $N_3 \leq 1$. If $N_2 = 2$ and $N_3 = 1$, then 1100, 0011, 0111 $\notin C$, say, but this conflicts Lemma 2 for $1100 + C$. Hence $N_2 + N_3 + N_4 \leq 2$ and $|C| \geq 10$.

**Case 2:** $0001 \in C$, but $0000 \notin C$: By Lemma 2, $0111, 1011, 1101 \in C$ and hence $N_3 \leq 1$, and $1001, 0101, 0011 \in C$ and then $N_2 \leq 1$ by Lemma 3. As in Case 1, $N_2 = 1$ implies $N_4 = 0$. Hence $N_2 + N_3 + N_4 \leq 2$ and $|C| \geq 10$.

**Case 3:** $0001 \in C$ and $0000 \in C$: By Lemma 3, $N_2 \leq 2$ and $N_3 \leq 2$. If $N_3 = 2$, then $1111 \in C$; otherwise Lemma 2 for $1111 + C$ would give $N_1 \leq 2$. Hence $N_3 + N_4 \leq 2$ and $N_2 = 2$. One of the non-codewords of weight two must end in 0; say $1100 \notin C$. Lemma 2 for $1100 + C$ shows that $0111, 1011 \in C$, and hence $x \notin C$ where $x = 1110$ or $x = 1101$. Then $\{1100, 1000, 0100, x\} \subseteq F_2^4 \setminus C$, but we already showed in Cases 1 and 2 that this is impossible (translate by 1100). $\square$

**Theorem 12.** $M_1^{(2)}(4) = 10$.

*Proof.* The code

$$C = \{a \in F_2^4 | w(a) = 2 \text{ or } w(a) = 3\}$$

gives the upper bound $M_1^{(2)}(4) \leq 10$.

Let us assume that $C$ is a $(1,2)$-identifying code with only nine codewords. In average $(S_1(x) \cup S_3(x)) \setminus C$ (for $x \in F_2^4$) contains $56/16 > 3$ elements. Choose $x$ so that $|(S_1(x) \cup S_3(x)) \setminus C| \geq 4$. Then by Lemma 4, $|S_1(x) \setminus C| = |S_3(x) \setminus C| = 2$; say $1000 \notin C$ and $0100 \notin C$. Because $|S_1(1100) \setminus C| \leq 2$ (again by Lemma 4), we know that $1011 \notin C$ and $0111 \notin C$, contradicting Lemma 3. $\square$

**Example 2.** The useful property of $(1, \leq 2)$-identifying codes which was given in Theorem 7 does not work for $(1,2)$-identifying codes. This can be seen for example using the code $C$ from the proof of Theorem 12. Denote $D = C \oplus \{0,1\}$. Now $|\mathrm{B}_1(00000) \cap D| = |\mathrm{B}_1(00001) \cap D| = 0$ and $D$ is not a $(1,2)$-identifying code by Theorem 10.

# 5 Weakly identifying codes

We can utilize some $(1, \leq 1)$-identifying codes to identify *iteratively* the fault pattern. This is possible if the code fulfills the requirement of the following definition.

**Definition 1.** *A* $(1, \leq 1)$*-identifying code* $C$ *is called* weakly $(1, \leq l)$-identifying *if for all* $k \geq 1$, $I(X_1) = I(X_2) = \ldots = I(X_k)$, $X_1, X_2, \ldots, X_k \in \mathcal{F}_{\leq l}$, *implies* $X_1 \cap X_2 \cap \ldots \cap X_k \neq \emptyset$.

Thus if, for example, the fault patterns $\{x, y\}$ and $x$ are 1-covered by the same codewords, i.e., $I(x) = I(x, y)$ and no other fault pattern gives that particular configuration among the codewords, then we can say that $x$ must be in the fault pattern and after repairing the faulty processor $x$ we can then use $C$ to find $y$.

**Theorem 13.** *Let* $C$ *be the code given in the proof of Theorem 8. The code* $C \setminus \{1110\}$ *is weakly* $(1, \leq 2)$*-identifying.*

*Proof.* Denote $q = 1110$ and $D = C \setminus \{q\}$. The word $q$ and words at distance one from $q$ are covered by exactly two codewords of $D$. Thus for every one of those words, say $x$, there is a unique word $z \neq x$ such that $I(x) \subseteq I(z)$. The five $z$-words are 1000, 0010, 0101, 0011 and 1101: in particular, different $x$-words have different $z$-words.

First of all, given $S \subseteq C$, there can be at most two fault patterns $F \in \mathcal{F}_{\leq 2}$ for which $I(F) = S$. Indeed, the sets $I(F)$ are obtained from the corresponding sets $I^*(F)$ for $C$ simply by dropping $q$ if it is in $I^*(F)$. If we have three sets $F$ with the same $I(F)$, at least two of the sets $I^*(F)$ are equal, contradicting the fact that $C$ is $(1, \leq 2)$-identifying.

If $|F| = |F'| = 1$ and $F \neq F'$, then $I(F) \neq I(F')$. Namely, by the previous discussion, exactly one of $I^*(F)$ and $I^*(F')$ contains $q$, say $q \in I^*(F)$, but then $I(F)$ has size only two and cannot be equal to $I(F')$.

Assume that $F = \{z\}$, $F' = \{x, y\}$ and $I(z) = I(x, y)$. By the previous discussion, $q \notin I^*(z)$ and $q \in I^*(x, y)$, say $q \in I^*(x)$. Then $I(x) \subseteq I(z)$ and we have seen that $z$ has to be the unique $z$-word associated with $x$. Moreover, $I(y) \subseteq I(z)$, which is possible only if $y = z$ or $y \in \mathrm{B}_1(q)$ and $z$ is the unique $z$-word associated with $y$. However, $x$ and $y$ can have the same $z$-word only if $x = y$, which is not the case. Hence $y = z$ and $F \cap F' \neq \emptyset$.

Suppose finally that there were pairs $\{x, u\}$ and $\{y, v\}$ of different words such that $I(x, u) = I(y, v)$. We can assume that $x \in \mathrm{B}_1(q)$. Moreover, we can assume that $d(y, q) \geq 2$ and $d(v, q) \geq 2$; otherwise the sets $I(x, u)$ and $I(y, v)$ are both again obtained by dropping out $q$ from the corresponding sets for $C$, and cannot be the same. We prove that there does not exist such words $x, y, v$ and $u$ by considering the distance $d(u, q) = d$. All the words at distance three from $q$ are codewords. The cases $d = 0$ and $d = 1$ are clear: $I(y, v)$ contains codewords at distance three from $q$ whereas $I(x, u)$ does not. If $d = 2$, then $I(x, u)$ does not contain $q + \mathbf{1} = 0001$, and hence $d(y, q) = d(v, q) = 2$. However, all the words at distance two cover exactly two codewords from distance three, and no two other words at distance two from $q$ can cover exactly those two codewords at distance three from $q$ that $u$ covers. The word 0001 itself is covered by five codewords and no pair not containing 0001 can cover all of them; hence $d = 4$ is impossible. In the case $d = 3$ we notice that $u$ covers 0001 and thus either $y$ or $v$ must cover it too, suppose $y$ does. If $y$ is at distance three form $q$, $y$ is a codeword and the pair $\{u, x\}$ cannot cover it. Now $y$ must be 0001, but this we have seen to be impossible.

This concludes the proof, because then if $|F| = |F'| = 2$ and $I(F) = I(F')$, then $F \cap F' \neq \emptyset$. $\square$

By computer it can be checked that no such code of cardinality nine exists. Notice also that Lemma 3 is valid for the current codes. From the identifying code $C$ of Theorem 9 we may drop the word 00100 and obtain a weakly $(1, \leq 2)$-identifying code of length five. Let $C$ be again the identifying code of Theorem 9. The code $C \oplus \{0,1\} \setminus \{000110, 111110, 101001, 110101\}$ can be seen using computer to be a weakly $(1, \leq 2)$-identifying code of length 6 and cardinality 28.

However, as will be seen in the next example, the direct sum method does not directly yield from one weakly identifying code another weakly identifying code of higher length.

**Example 3.** Let $C$ be the identifying code of Theorem 9 from which we have excluded the word 00100. Now the direct sum $C \oplus \{0,1\}$ cannot distinguish between the pairs $\{000000, 000111\}$ and $\{000001, 000110\}$.

# 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, to appear.

[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, submitted.

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

[8] H. Hämäläinen, I. Honkala, M. Kaikkonen and S. Litsyn, Bounds for binary multiple covering codes, Designs, Codes and Cryptography, 3:251–275, 1993.

[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 7th Nordic Combinatorial Conference*, Turku, pp. 39–43, 1999.

[11] 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.

[12] F. J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, Amsterdam: North-Holland, 1977.

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

http://www.tucs.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**