Jak wykrywać i poprawiać błędy

MiNI Akademia Matematyki

MiNI PW
Agata Pilitowska

18.11.2017

UPC - Universal Product Code

Każdy produkt opisany jest 13-to cyfrowym kodem: 7656778768791

UPC - Universal Product Code

Kod EAN-13

$$X_1X_2X_3X_4X_5X_6X_7X_8X_9X_{10}X_{11}X_{12}X_{13}$$

Funkcja kontrolna:

$$X_1+3X_2+X_3+3X_4+X_5+3X_6+X_7+3X_8+X_9+3X_{10}+X_{11}+3X_{12}+X_{13} = 10\cdot k$$

Przykład: 7656778768791

$$7+3\cdot 6+5+3\cdot 6+7+3\cdot 7+8+3\cdot 7+6+3\cdot 8+7+3\cdot 9+1 = 10\cdot 17$$

Kody wagowe z jednym symbolem kontrolnym

Kod długości n ze współczynnikami wagowymi $w_1, w_2, \ldots, w_{n-1}, w_n$

$$X_1X_2X_3X_4X_5 \ldots X_{n-1}X_n$$

Funkcja kontrolna:

$$w_1X_1 + w_2X_2 + w_3X_3 + w_4X_4 + \ldots + w_{n-1}X_{n-1} + w_nX_n = r + p \cdot k \equiv_p r$$

Przykład: Kod EAN-13

$$7+3\cdot 6+5+3\cdot 6+7+3\cdot 7+8+3\cdot 7+6+3\cdot 8+7+3\cdot 9+1 = 10\cdot 17 \equiv_{10} 0$$

$n = 13, p = 10, r = 0$
$w_1 = w_3 = w_5 = w_7 = w_9 = w_{11} = w_{13} = 1$
$w_2 = w_4 = w_6 = w_8 = w_{10} = w_{12} = 3$

Kod wagowy z jednym symbolem kontrolnym

Numer PESEL

$$X_1X_2X_3X_4X_5X_6X_7X_8X_9X_{10}X_{11}$$

$X_1X_2X_3X_4X_5X_6$	–	*data*	*urodzenia*
$X_7X_8X_9$	–	*liczba*	*porządkowa*
X_{10}	–	K – *parzysta*	M – *nieparzysta*
X_{11}	–	*cyfra*	*kontrolna*

Funkcja kontrolna:

$$1\cdot X_1+3\cdot X_2+7\cdot X_3+9\cdot X_4+1\cdot X_5+3\cdot X_6+7\cdot X_7+9\cdot X_8+1\cdot X_9+3\cdot X_{10}+1\cdot X_{11} \equiv_{10} 0$$

$n = 11, p = 10, r = 0$
$w_1 = 1, w_2 = 3, w_3 = 7, w_4 = 9, w_5 = 1, w_6 = 3, w_7 = 7,$
$w_8 = 9, w_9 = 1, w_{10} = 3, w_{11} = 1$

Numer PESEL

Funkcja kontrolna:

$$1 \cdot X_1 + 3 \cdot X_2 + 7 \cdot X_3 + 9 \cdot X_4 + 1 \cdot X_5 + 3 \cdot X_6 + 7 \cdot X_7 + 9 \cdot X_8 + 1 \cdot X_9 + 3 \cdot X_{10} + 1 \cdot X_{11} \equiv_{10} 0$$

Przykład: 02070803628

$$1 \cdot 0 + 3 \cdot 2 + 7 \cdot 0 + 9 \cdot 7 + 1 \cdot 0 + 3 \cdot 8 + 7 \cdot 0 + 9 \cdot 3 + 1 \cdot 6 + 3 \cdot 2 + 1 \cdot 8 = 140 = 10 \cdot 14 \equiv_{10} 0$$

Kody wagowe z jednym symbolem kontrolnym

Przykłady

- ISBN - International Standard Book Number
- ISSN - International Standard Serial Number
- NIP - Numer Identyfikacji Podatkowej
- REGON - Rejestr Gospodarki Narodowej

KODY BINARNE

Kod długości $n \in \mathbb{N}$

$X_1X_2X_3X_4X_5 \ldots X_{n-1}X_n$

$$X_1, X_2, \ldots, X_{n-1}, X_n \in \{0, 1\}$$

Dla $n = 22$

$$1011010101010101010101$$

Kod kontroli parzystości $X_1X_2X_3X_4X_5 \ldots X_{n-1}X_n$

Kod wagowy z jednym symbolem kontrolnym

Funkcja kontrolna:

$$X_1 + X_2 + \ldots + X_{n-1} + X_n = 2 \cdot k \equiv_2 0$$

$n \in \mathbb{N}, p = 2, r = 0$
$w_i = 1$, dla $i = 1, 2, \ldots, n$

$n = 3$

$X_1X_2X_3 = 000$	$0 + 0 + 0 \equiv_2 0$
$X_1X_2X_3 = 101$	$1 + 0 + 1 \equiv_2 0$
$X_1X_2X_3 = 011$	$0 + 1 + 1 \equiv_2 0$
$X_1X_2X_3 = 110$	$1 + 1 + 0 \equiv_2 0$

Kody binarne z powtórzeniami

Powtórzenia dwukrotne

$$X_1X_2X_3X_4X_5\ldots X_{n-1}X_n X_1X_2X_3X_4X_5\ldots X_{n-1}X_n$$

Dla	$n=1$	$0\,0$
		$1\,1$
Dla	$n=2$	$00\,00$
		$10\,10$
		$01\,01$
		$11\,11$

Kody binarne z powtórzeniami

Powtórzenia trzykrotne

$$X_1X_2X_3X_4X_5\ldots X_{n-1}X_nX_1X_2X_3X_4X_5\ldots X_{n-1}X_nX_1X_2X_3X_4X_5\ldots X_{n-1}X_n$$

Dla	$n=1$	000
		111
Dla	$n=2$	000000
		101010
		010101
		111111

Transmisja dwóch symboli binarnych

Bez symboli kontrolnych

$$01 \overset{Transmisja}{\Longrightarrow} 11$$ jeden symbol niepoprawny niewykryty

Z symbolami kontrolnymi

011	$\overset{Transmisja}{\Longrightarrow}$	111	jeden symbol niepoprawny wykryty
0101	$\Longrightarrow$	1101	jeden symbol niepoprawny wykryty
010101	$\Longrightarrow$	110101	jeden symbol niepoprawny wykryty i popraw.
010101	$\Longrightarrow$	111101	dwa symbole niepoprawne wykryte

Cyfrowa transmisja danych

Schemat kodowania

$$\underset{u_1u_2\ldots u_k}{\textit{Informacja}} \xrightarrow{\textit{Kodowanie}} \underset{c_1c_2\ldots c_kc_{k+1}\ldots c_n}{\textit{Inf.zakodowana}} \xrightarrow{\textit{Transmisja}}$$

$$\underset{f_1f_2\ldots f_n}{\textit{Inf.odebrana}} \xrightarrow{\textit{Dekodowanie}} \underset{v_1v_2\ldots v_k \overset{???}{=} u_1u_2\ldots u_k}{\textit{Inf.otrzymana}}$$

Informacja zakodowana = *Słowo kodowe*

$$c_1\ldots c_kc_{k+1}\ldots c_n$$

n = długość słowa kodowego
k = liczba symboli informacji (wiadomości)
$n-k$ = liczba symboli kontrolnych

Pomysłodawca teorii kodowania

Richard Hamming

1915 - ur. w Chicago
1937 - dyplom University of Nebraska
1942 - doktorat na University of Illinois
II Wojna Światowa - Profesor na Universytecie w Louisville
1945 - Projekt Manhattan
1946-76 - Bell Telephone Laboratories
1950 - "Error detecting and error correcting codes"
1968 - Nagroda Turinga (Association for Computing Machinery)
1976-97 - Naval Postgraduate School
1998 - zm. w Monterey (Kalifornia)

Zastosowania

Kody korekcyjne stosowane są

- w telekomunikacji
- w sieciach komputerowych
- w pamięciach półprzewodnikowych
- w systemach zapisu danych
- w kryptografii

Kody binarne

$$\underset{u_1 u_2 \ldots u_k}{\textit{Informacja}} \quad \overset{\textit{Kodowanie}}{\longrightarrow} \quad \underset{c_1 \ldots c_k c_{k+1} \ldots c_n}{\textit{Informacja zakodowana}}$$

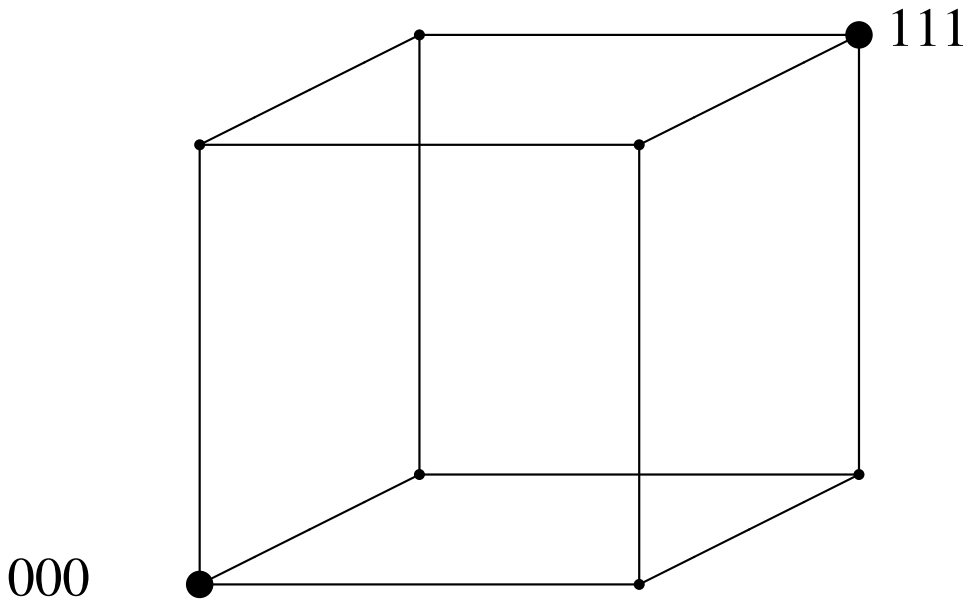
KOD

Kodem (binarnym) $\mathcal{C}$ długości n nazywamy zbiór słów kodowych długości n.

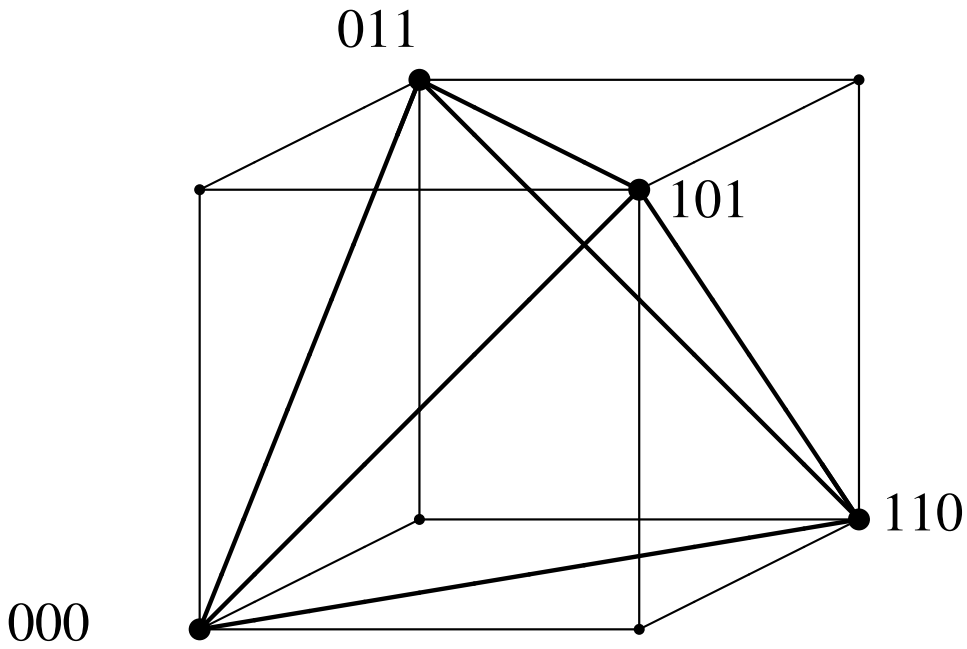
- $\mathcal{C}_0 = \{000, 111\}$ - kod długości $n = 3 \;\; (k = 1)$
- $\mathcal{C}_1 = \{000, 101, 011, 110\}$ - kod długości $n = 3 \;\; (k = 2)$
- $\mathcal{C}_2 = \{0000, 1010, 0101, 1111\}$ - kod długości $n = 4 \;\; (k = 2)$
- $\mathcal{C}_3 = \{000000, 101010, 010101, 111111\}$ - kod długości $n = 6 \;\; (k = 2)$

Jeżeli liczba symboli w przesyłanych wiadomościach równa jest k to liczba (binarnych) słów kodowych = 2^k

Kod binarny $\mathcal{C}_0 = \{000, 111\}$



Kod kontroli parzystości $\mathcal{C}_1 = \{000, 101, 011, 110\}$



"Możliwości" kodów

- Co to znaczy, że kod może wykryć lub poprawić błąd w przesyłanej informacji?
- Dlaczego jedne kody tylko wykrywają błędy a inne mogą je również poprawić?
- Dlaczego różne kody wykrywają lub poprawiają różne ilości błędów?

Odległość Hamminga

Odległość między ciągami $\boldsymbol{c} = c_1 \ldots c_n$ i $\boldsymbol{f} = f_1 \ldots f_n$

Liczba miejsc, na których ciągi $\boldsymbol{c} = c_1 \ldots c_n$ i $\boldsymbol{f} = f_1 \ldots f_n$ się różnią.
Oznaczamy $d(\boldsymbol{c}, \boldsymbol{f})$.

Przykład

$$d(1110, 1100) = 1$$

$$d(1010, 0000) = 2$$

Odległość kodu

Odległością kodu $\mathcal{C}$ nazywamy liczbę

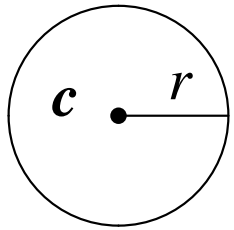
$$d(\mathcal{C}) := min\{d(\boldsymbol{c},\boldsymbol{f}) \mid \boldsymbol{c},\boldsymbol{f} \in \mathcal{C}, \boldsymbol{c} \neq \boldsymbol{f}\}.$$

Przykład

$\mathcal{C}_0 = \{000, 111\}$, $d(\mathcal{C}_0) = 3$
$\mathcal{C}_1 = \{000, 101, 011, 110\}$, $d(\mathcal{C}_1) = 2$
$\mathcal{C}_2 = \{0000, 1010, 0101, 1111\}$, $d(\mathcal{C}_2) = 2$
$\mathcal{C}_3 = \{000000, 101010, 010101, 111111\}$, $d(\mathcal{C}_3) = 3$

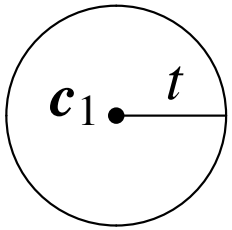
Kula $K(\boldsymbol{c}, r)$ o promieniu t wokół słowa $\boldsymbol{c} \in \mathcal{C}$

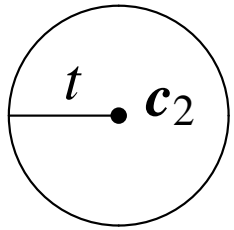
$$K(\boldsymbol{c}, r) = \{\boldsymbol{v} = v_1 v_2 \ldots v_n \mid d(\boldsymbol{c}, \boldsymbol{v}) \leq r\}$$



$$d = d(\mathcal{C}) = 2t + 1 \implies$$

- Dowolne dwa słowa kodowe różnią się na co najmniej d miejscach
- Kule $K(\boldsymbol{c}, t)$ o promieniu t wokół słów kodowych $\boldsymbol{c} \in \mathcal{C}$ są rozłączne



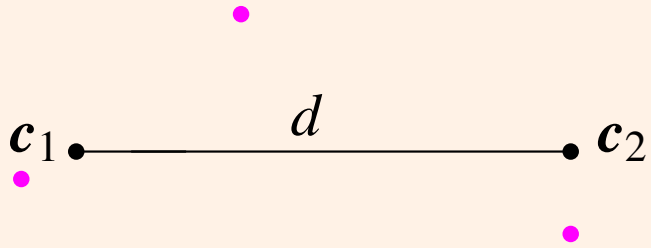


Błędy podczas transmisji

$\mathcal{C}$ - kod
$d(\mathcal{C}) = d$
$\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathcal{C}$

$$\boldsymbol{c}_1 \overset{\textit{Transmisja}}{\Longrightarrow} \boldsymbol{c}_1$$

$$\boldsymbol{c}_1 \overset{\textit{Transmisja}}{\Longrightarrow} \boldsymbol{c}_2!!!$$



Własności detekcyjne kodów

Twierdzenie

Warunkiem koniecznym i dostatecznym na to, aby kod $\mathcal{C}$ umożliwiał wykrycie $d-1$ (lub mniej) błędów jest, aby odległość kodu była równa co najmniej $d(\mathcal{C}) = d$.

Przykład

$\mathcal{C}_0 = \{000, 111\}$, $d(\mathcal{C}_0) = 3$, wykrywa błędy pojedyncze i podwójne
$\mathcal{C}_1 = \{0000, 1010, 0101, 1111\}$, $d(\mathcal{C}_1) = 2$, wykrywa błędy pojedyncze
$\mathcal{C}_2 = \{000, 101, 011, 110\}$, $d(\mathcal{C}_2) = 2$, wykrywa błędy pojedyncze
$\mathcal{C}_3 = \{000000, 101010, 010101, 111111\}$, $d(\mathcal{C}_3) = 3$, wykrywa błędy pojedyncze i podwójne

Własności korekcyjne kodów

$\mathcal{C}$ - kod

$d(\mathcal{C}) = d = 2t + 1$

$\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathcal{C}$

$\boldsymbol{c}_1 \overset{Transmisja}{\Longrightarrow}$????

Własności korekcyjne kodów

Twierdzenie

Kod $\mathcal{C}$ może poprawić co najwyżej t błędów wtedy i tylko wtedy, gdy jego odległość wynosi co najmniej $d = 2t + 1$.

Przykład

$\mathcal{C}_0 = \{000, 111\}$, $d(\mathcal{C}_0) = 3$, poprawia błędy pojedyncze
$\mathcal{C}_1 = \{0000, 1010, 0101, 1111\}$, $d(\mathcal{C}_1) = 2$, nie poprawia błędów
$\mathcal{C}_2 = \{000, 101, 011, 110\}$, $d(\mathcal{C}_2) = 2$, nie poprawia błędów
$\mathcal{C}_3 = \{000000, 101010, 010101, 111111\}$, $d(\mathcal{C}_3) = 3$, poprawia błędy pojedyncze

Poprawianie błędów

Kod $\mathcal{C}_3 = \{000000, 101010, 010101, 111111\}$

$d(\mathcal{C}_3) = 3$ - koryguje błędy pojedyncze

$$010101 \overset{Transmisja}{\Longrightarrow} 110101$$

$$d(110101, 000000) = 4, \quad d(110101, 101010) = 5$$

$$d(110101, 010101) = 1, \quad d(110101, 111111) = 3$$

Kody praktycznie stosowane - Kody BCH

R.C. Bose, D.K. Ray-Chaudhuri, A. Hocquenghema
Rodzina kodów poprawiających błędy wielokrotne

Jednym z pierwszych praktycznie realizowanych ważnych kodów był kodem BCH o długości $n = 127$ z 35 symbolami kontrolnymi. Składał się z 2^{92} słów kodowych. Miał odległość $d = 11$, czyli poprawiał do 5 błędów.

Powszechnie stosowanym w europejskich systemach przesyłania danych jest binarny kod BCH długości $n = 255$, o 24 symbolach kontrolnych. Ma 2^{231} słów kodowych a jego odległość wynosi $d = 7$, tzn. poprawia błędy potrójne.

Kody praktycznie stosowane - Kod Reeda - Mullera

1972 - Sonda Mariner: zdjęcia z Marsa

Zdjęcia czarno-białe: każdy "punkt" zdjęcia był opisany przez jeden z 64 odcieni szarości.
Każdy odcień szarości zapisano jako ciąg binarny długości 6. ($64 = 2^6$)
Kodowanie: do wiadomości długości 6 dopisano 26 symboli kontrolnych. ($n = 32$)
Kod poprawia 7 błędów ($d = 15$).

Kody praktycznie stosowane - Kod Goley'a

1979 - Sonda Voyager: zdjęcia z Jupitera

Zdjęcia kolorowe: każdy "punkt" zdjęcia był opisany przez jeden z 4096 kolorów.
Każdy kolor zapisano jako ciąg binarny długości 12. ($4096 = 2^{12}$)
Kodowanie: do wiadomości długości 12 dopisano kolejne 12 symboli kontrolnych. ($n = 24$)
Kod poprawia 3 błędy ($d = 7$).

Binarna reprezentacja liczby n

$m \in \mathbb{N}$

Każdą liczbę naturalną $0 \leq n \leq 2^m - 1$ można przedstawić w postci:

$$n = w_1 \cdot 2^0 + w_2 \cdot 2^1 + \ldots + w_m \cdot 2^{m-1}$$

Binarny ciąg długości m

$$w_1 w_2 \ldots w_m$$

nazywamy *binarną reprezentacją liczby* n (długości m)

Binarna reprezentacja liczby $6 \leq 2^3 - 1$ długości $m = 3$

$$6 = 0 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2$$

Binarna reprezentacja: $w_1 w_2 w_3 = 011$

Kod Hamminga $\mathcal{H}_2$

Binarne reprezentacje długości $m = 2$ liczb:
$1 \rightsquigarrow 10, 2 \rightsquigarrow 01, 3 \rightsquigarrow 11$

$$H_2 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$n = 2^m - 1 = 2^2 - 1 = 3$ - długość kodu
$c_1c_2c_3 \in \mathcal{H}_2$ wtedy i tylko wtedy, gdy

$$\begin{array}{ccccccc} 0 \cdot c_1 & + & 1 \cdot c_2 & + & 1 \cdot c_3 & = & 0 \\ 1 \cdot c_1 & + & 0 \cdot c_2 & + & 1 \cdot c_3 & = & 0 \end{array}$$

c_3 - symbol wiadomości; c_1, c_2 - symbole kontrolne

$$\begin{array}{ccccc} c_2 & + & c_3 & = & 0 \\ c_1 & + & c_3 & = & 0 \end{array} \Leftrightarrow c_1 = c_2 = c_3$$

Kod Hamminga $\mathcal{H}_2$ jest taki sam jak kod $\mathcal{C}_0 = \{000, 111\}$

Kod Hamminga $\mathcal{H}_m$

$m \in \mathbb{N}$, $n = 2^m - 1$, $1 \leq i \leq 2^m - 1$

$$H_m = \begin{bmatrix} w_{m1} & \dots & w_{mi} & \dots & w_{mn} \\ w_{(m-1)1} & \dots & w_{(m-1)i} & \dots & w_{(m-1)n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{j1} & \dots & w_{ji} & \dots & w_{jn} \\ \dots & \dots & \dots & \dots & \dots \\ w_{11} & \dots & w_{1i} & \dots & w_{1n} \\ & & \Uparrow & & \end{bmatrix}$$

$w_{1i}w_{2i}\dots w_{mi}$ - binarna reprezantacja liczby n (długości m)

$$\begin{aligned} w_{m1}c_1 + w_{m2}c_2 + \dots + w_{mi}c_i + \dots + w_{mn}c_n &= 0 \\ \dots\dots\dots\dots &= 0 \\ w_{j1}c_1 + w_{j2}c_2 + \dots + w_{ji}c_i + \dots + w_{jn}c_n &= 0 \\ \dots\dots\dots\dots &= 0 \\ w_{11}c_1 + w_{12}c_2 + \dots + w_{1i}c_i + \dots + w_{1n}c_n &= 0 \end{aligned}$$

Kod Hamminga $\mathcal{H}_m$, $m \in \mathbb{N}$

Definicja

Kodem Hamminga $\mathcal{H}_m$ nazywamy kod długości $n = 2^m - 1$ o m symbolach sprawdzających: $c_1, c_2, c_4, \ldots, c_{2^{m-1}}$, dla którego $\boldsymbol{c} = c_1 \ldots c_n \in \mathcal{H}_m$ wtedy i tylko wtedy, gdy $c_1 \ldots c_n$ spełniają każde z równań

$$\begin{aligned} w_{m1}c_1 + w_{m2}c_2 + \ldots + w_{mi}c_i + \ldots + w_{mn}c_n &= 0 \\ \ldots\ldots\ldots\ldots &= 0 \\ w_{j1}c_1 + w_{j2}c_2 + \ldots + w_{ji}c_i + \ldots + w_{jn}c_n &= 0 \\ \ldots\ldots\ldots\ldots &= 0 \\ w_{11}c_1 + w_{12}c_2 + \ldots + w_{1i}c_i + \ldots + w_{1n}c_n &= 0 \end{aligned} \tag{1}$$

.

Kod $\mathcal{H}_m$ wykrywa błędy podwójne oraz poprawia błędy pojedyncze ($d = 3$).

Kody Hamminga są doskonałe

Kod Hamminga $\mathcal{H}_3$

- Liczba wszystkich binarnych ciągów długości 7 równa jest $2^7 = 128$
- Kod Hamminga $\mathcal{H}_3$ ($m = 3$) jest długości $n = 2^3 - 1 = 7$ i ma $2^{7-3} = 2^4 = 16$ słów kodowych.
- Każda kula $K(\boldsymbol{c}, 1)$ o środku w słowie kodowym $\boldsymbol{c} \in \mathcal{H}_3$ zawiera 8 ciągów binarnych długości 7.
- We wszystkich takich kulach mieści się $8 \cdot 16 = 128 = 2^7$ ciągów binarnych długości 7, czyli wszystkie!

Wszystkie binarne ciągi długości $n = 3$ mieszczą się w kulach $K(\boldsymbol{c}, 1)$, dla $\boldsymbol{c} \in \mathcal{H}_3$. Kod Hamminga $\mathcal{H}_3$ jest **kodem doskonałym**.

Wszystkie binarne kody Hamminga $\mathcal{H}_m$ są doskonałe.