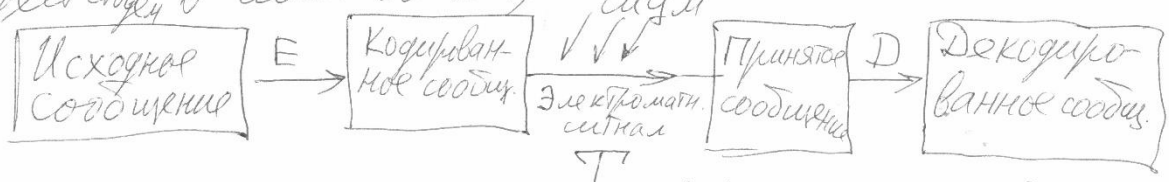
Тема №7. Элементы теор. кодир.

Будем рассм. задачу передачи сообщ. Любое сообщ. можно закодировать послед-ью из 0 и 1, а затем передавать его по каналу связи. Любой кан. св. не идеален, поэтому сообщ. в сообщ. со след. схемой



Здесь E обозначает опер. кодиров., T – "ф-ию ошибок" D – опер. декодиров. E и D выбираются т.о., чтобы композиция $D \circ T \circ E$ была функцией с большой вероятностью близкой к тождественной.

7.1. Блочный двоичный (m,n)-код определяется двумя функциями:

(а) схема кодирования $E: \{0,1\}^m \to \{0,1\}^n$;

(б) схема декодирования $D: \{0,1\}^n \to \{0,1\}^m$.

Функции E и D (как уже было сказано) подбираются так, чтобы ф-ия $D \circ E$ была тожд., а композ. $D \circ T \circ E$ была ф-ией, с большой вер-тью близкой к тожд-ву.

Коды делятся на 2 класса: коды с обнаружением ошибок и коды с исправлением ошибок.

Пример 7.1. Простой код с обнаружением ошибок основан на схеме проверки четности. Это $(m, m+1)$-код, где $m \geq 1$. Пусть $a = a_1 \dots a_m \in \{0,1\}^m$ – сообщ. длины m. Схема кодирования опр-ся т.о.: $E(a) = b = b_1 \dots b_{m+1}$, где $b_i = a_i$, $i = 1, \dots, m$; $b_{m+1} = a_1 + \dots + a_m$ (таким обр., всегда $b_1 + \dots + b_{m+1} = a_1 + \dots + a_m + a_1 + \dots + a_m = 0$).

Схема декодирования отр-ся след. обр.: $D(b) = c_1 \dots c_m$, где $c_i = b_i$ (——), $i = 1, \dots, m$. При этом, если b – принятое сообщ. и при передаче было ошибок, то $\sum_{i=1}^{m+1} b_i$ – четно, если же имелась ошибка в одной позиции, то $\sum_{i=1}^{m+1} b_i$ – нечетно. Т.о., этот код позволяет обнаруживать ошибку в одной позиции.

Напр., при $m=2$ $E(00)=000$, $E(01)=011$, $E(10)=101$, $E(11)=110$, $D(011)=01$ и единичной ош. не было, $D(010)=01$, и была или один. ошибка или тройная ошибка.

Пример 7.2. Простой код с исправлением ошибок дает код с трехкратным повторением. В этом коде каждый символ повторяется трижды, т.е. рассматривается $(1,3)$-код, при этом $E(0)=000$, $E(1)=111$. Функция декодир. такова: в тройке символов выбирается символ, чаще всего встречающийся в этой тройке. Этот код позволяет исправлять единичную ошибку в каждой принятой тройке символов. Это пример очень неэкономного кода.

b_1	b_2	b_3	$D(b)$
1	1	1	1
1	1	0	1
1	0	1	1
1	0	0	0
0	1	1	1
0	1	0	0
0	0	1	0
0	0	0	0

7.2. Расстояние Хемминга.

Весом $w(a)$ слова $a_1 \dots a_m \in \{0,1\}^m$ называется число единиц среди его координат. Расстоянием $d(a,b)$ между словами a и b называют число несовпадающих позиций в этих словах, например, расст. между a. $a = 0111101$ равно 2. $b = 0010111$

Очевидно, что $d(a,b) = w(a+b)$ (сложение покоординатное по $\bmod 2$).

$d(a,b)$ наз-ся **расст. Хемминга**. Оно уд-ет аксиомам расстояний:

(1) $d(a,b) \geq 0$ и $d(a,b) = 0 \Leftrightarrow a = b$;

(2) $d(a,b) = d(b,a)$;

(3) $d(a,b) + d(b,c) \geq d(a,c)$ (нер-во треугольника).

Доказательство (3). Если $a_i \neq c_i$, то $a_i \neq b_i$ или $b_i \neq c_i$ (т.к. в противном случае, если $a_i = b_i$, $b_i = c_i$, то $a_i = c_i$).

Заметим, что для возможности обнаружения ошибки в одной позиции минимальное расстояние между кодовыми словами ($\{E(a) \mid a \in \{0,1\}^m\}$ – мн-во кодовых слов) должно быть ≥ 2. Иначе ошибка в 1-й позиции может превратить одно кодовое слово в другое и она не будет обнаружена.

Теорема 7.1. Для того, чтобы код давал возможность обнаруживать наличие любых ошибок в k (или менее) позициях, н. и д., чт. наим. расст. между кодовыми словами было $\geq k+1$. Док. Необходимость. Пусть код обнаруживает любые ошибки в $\leq k$ позициях, но наименьшее расстояние между двумя различными кодовыми словами $\leq k$. Тогда найдутся два кодовых слова a, b, таких, что $d(a,b) \leq k$ и при этом при передаче слова b без ошибок и слова a с ошибками не более, чем в k позициях, мы можем получить одно и то же слово b. Такая неоднозначность не дает возможность отличить случай с безошибочной передачей от случая с ошибочной передачей с ошибками не более, чем в k позициях, что противоречит исходному предположению.

Достаточность. Пусть наим. расст. между кодовыми словами $\geq k+1$. Предположим, что при передаче некоторого кодового слова a допущено $\leq k$ ошибок. Тогда в случае наличия хотя бы одной ошибки полученное слово не будет являться кодовым! Т. о., критерием наличия в переданном слове $\leq k$ ошибок является непринадлежность этого слова мн-ву кодовых слов.

Теорема 7.2. Для того, чтобы код давал возможность исправлять любые ошибки в $\leq k$ позициях н. и д., чтобы наим. расст. между двумя кодовыми словами было $\geq 2k+1$. Док-во. Необх. Пусть код исправляет любые ошибки в $\leq k$ позициях. Предположим, что мин. расст. между кодовыми словами $\leq 2k$. Тогда найдутся два кодовых слова a и b такие, что $d(a,b) \leq 2k$.

Рассм-м некоторое слово c: (1) совпадающее с a и b во всех позициях в которых a и b совпадают; (2) совпадающее с a в каких-либо k позициях, в которых a не совпадает с b (если число несовпадающих позиций у a и b $\leq k$, то полагаем $c=a$); (3) совпадающее с b в остальных позициях, в которых a не совпадает с b (их число $\leq k$). Тогда $d(a,c) \leq k$, $d(c,b) \leq k$, откуда следует, что при передаче слов a и b с ошибками в $\leq k$ числе позиций кодовые слова a и b могут перейти в одно и то же слово c. Такая неоднозначность не дает возможность установить по принятому слову c передаваемое слово, что противоречит исход. предп.

Дост. Пусть наим. расст. между кодовыми словами $\geq 2k+1$. Предпол., что при передаче какого-либо кодового слова a допущено $\leq k$ ошибок, т.е. получено слово c такое, что $d(a,c) \leq k$. Тогда для любого кодового слова $b \neq a$ имеем $2k+1 \leq d(a,b) \leq d(a,c) + d(c,b)$ откуда $d(c,b) \geq d(a,b) - d(a,c) \geq 2k+1-k \geq k+1 > k \geq d(c,a)$, т след., для исправления ошибок в $\leq k$ позициях достаточно брать ближайшее кодовое слово к принятому слову.

Пример 7.3. Мин. расст. между кодовыми словами в блочном $(m, m+1)$-коде с проверкой четности равно 2, что и позволяет обнаруживать 1 ошибку.

Пример 7.4. Мин. расст. между кодовыми словами в блочном $(1,3)$-коде с трехкратным повторением равно 3 [код.сл.: 000, 111], т.е. и далее возм-ть испр. 1 ош и обнаруж. 2 ош.

7.3. Матричное кодирование. При явном задании схемы кодирования в (n,k)-коде следует указать 2^k кодовых слов, что весьма неэффективно. Одним из экономных способов описания схемы кодирования явл-ся методика мат. к

Пусть $G=[g_{ij}]$ – матр. порядка $k \times n$, где $g_{ij} \in \{0,1\}$ (эл-т g_{ij} – бит)

Определим схему кодирования уравнением

$$b = aG,$$

где $a = a_1 \dots a_k$ – вектор, соотв-ий передаваемому сообщению, $b = b_1 \dots b_n$ – вектор, соотв-ий кодируемому сообщению. При этом G называется порождающей матрицей кода. Умнож. aG осущ-ся обычным образом с той лишь разницей, что сложение заменяется на $+ \pmod 2$.

Код не должен приписывать различным словам-сообщениям одно и то же кодовое слово. Простой способ добиться этого состоит в том, чтобы ранг мат. G равнялся k – числу строк (2^k – число всех двоичных слов длины k $\Rightarrow$ C_2 – поле $\Rightarrow$ все алгебр. понятия сохр-ся)

Заметим, что вместо 2^k кодовых слов достаточно знать k слов, являющихся строками мат. G (т.е. вместо $n \cdot 2^k$ символов $k \cdot n$ символов).

Пример 7.5. Порождающей матрицей $(2,3)$-кода с проверкой чётности явл. мат. $G=\begin{bmatrix}1&0&1\\0&1&1\end{bmatrix}$:

$$a_1 a_2 \begin{bmatrix}1&0&1\\0&1&1\end{bmatrix} = a_1 a_2 (a_1+a_2)$$

Пример 7.6. Порождающей мат. $(1,3)$-кода с повторением явл-ся мат. $G=[1\,1\,1]$: $000 = 0 \cdot G$, $111 = 1 \cdot G$.

7.4. Групповые коды. Двоичный (n,k)-код наз-ся групповым, если его кодовые слова образуют группу.

Заметим, что кол-во всех двоичных слов длины

m образует группу. Это ком. гр. $(Z_2^m, +)$. (Z_2^m = $\{0,1\}^m$ по mod 2; + - покоорд. слож.)
Здесь единичный эл. $0 = 0\ldots0 \in \{0,1\}^m$, и обратный к $a = a_1 \ldots a_m$ элемент есть $-a = a$. След., мн-во слов – собств. ком. гр. ($\forall a^1, a^2 \in Z_2^m$ $a^1 + a^2 \in Z_2^m$; acc. + очевидна).

Пусть G - порождающая матр. кода порядка $m \times n$. Тогда мн-во кодовых слов $B = \{b = aG \mid a \in \{0,1\}^m\}$ эв. гр., т.к. (1) если $b^1 = a^1G$, $b^2 = a^2G$, где $a^1, a^2 \in \{0,1\}^m$, то $b^1 + b^2 = a^1G + a^2G = (a^1 + a^2)G \in B$, т.к. $a^1 + a^2 \in \{0,1\}^m$, т.е. покоорд. слож. (по mod 2) не выводит из мн-ва кодовых слов. Кроме того:
(2) $\underbrace{0\ldots0}_{m}G = \underbrace{0\ldots0}_{n} \in B$; (3) $b \in B \Rightarrow -b = b \in B$.

Теор 7.3. В групповом коде наим. расст. d_{min} между различными код. словами равно наим. весу w_{min} ненулевого кодового слова.

Док-во. (а) Пусть $d_{min} = d(b^1, b^2)$. Тогда $d(b^1, b^2) = w(b^1 + b^2) \geq w_{min}$. (б) Пусть $w_{min} = w(b)$. Тогда $w(b) = d(b, 0) \geq d_{min}$. И (а), (б) влеч., что $w_{min} = d_{min}$.

7.5. Коды Хемминга.

Миним. расст. между кодовыми словами в коде Х. равно 3. Он исправляет любую ошибку в одной позиции и обнар. 2 ош. ($2k + 1 = 3 \Rightarrow k = 1$).

Код Х. можно строить не при всех m и n, а только спец. вида: $m = 2^r - 1 - r$, $n = 2^r - 1$, где $r = 2, 3, \ldots$

Пример 7.7. При $r = 2$ $m = 1$, $n = 3$; при $r = 3$ $m = 4$, $n = 7$; при $r = 4$ $m = 11$, $n = 15$ и т.д.; при этом $\frac{m = 2^r - 1 - r}{n = 2^r - 1} \to 1$ при $r \to \infty$.

Процедура построения кода Х. такова. Опишем сначала схему кодирования.

(1) Выберем $r \geq 2$. Сообщения – слова длины $2^r - 1 - r$, а кодовые слова имеют длину $n = 2^r - 1$.

(2) В каждом кодовом слове $b = b_1 \dots b_{2^r-1}$ символы $b_{2^0}, b_{2^1}, \dots, b_{2^{r-1}}$ являются контрольными, остальные ($n = 2^r - 1 - r = m$) в естественном порядке – символами сообщения. Например, при $r = 3$

$$m = 4, n = 7 \Rightarrow a = a_1 a_2 a_3 a_4 \mapsto b = \underset{\|}{b_1}\ \underset{\|}{b_2}\ \underset{a_1}{b_3}\ \underset{\|}{b_4}\ \underset{a_2}{b_5}\ \underset{a_3}{b_6}\ \underset{a_4}{b_7}$$

(3) Рассмотрим матрицу M порядка $(2^r - 1) \times r$ такую, что в i-й строке этой матрицы стоят символы двоичного разложения числа i. Тогда матр. M при $r = 2, 3$ имеет, соотв-но, вид:

$$M_{3,2} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad M_{7,3} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

(4) Зафиксируем след. ур-ий

$$b M = 0. \tag{7.1}$$

Например, при $r = 3$ эта система имеет вид:

$$b_1 b_2 b_3 b_4 b_5 b_6 b_7 \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{cases} b_4 + b_5 + b_6 + b_7 = 0, \\ b_2 + b_3 + b_6 + b_7 = 0, \\ b_1 + b_3 + b_5 + b_7 = 0, \end{cases} \Leftrightarrow \begin{cases} b_4 = a_2 + a_3 + a_4, \\ b_2 = a_1 + a_3 + a_4, \\ b_1 = a_1 + a_2 + a_4. \end{cases}$$

Заметим, что по построению матр. M в каждом ур-ии сист. (7.1) входит ровно один контрольный символ (в i-е ур. вход. симв. $b_{2^{r-i}}$).

(5) При кодировании сообщения значения контр. символов $b_{2^0}, b_{2^1}, \dots, b_{2^{r-1}}$ находятся из сист. (7.1). Это нетрудно сделать, поск. в каждое из ур-ий сист. (7.1) входит ровно один контр. символ.

Пример 7.8. При $r=3$ $(4,7)$–код) сообщение $a=1011$ кодируется кодовым словом:
$b=(b_1)(b_2)\,b_3\,(b_4)\,b_5\,b_6\,b_7=(b_1)(b_2)\,1\,(b_4)\,0\,1\,1$, где, в силу (7.2),

$$(b_4)+b_5+b_6+b_7=b_4+0+1+1=0\Rightarrow b_4=0$$
($a_2=0$, $a_3=1$, $a_4=1$)

$$(b_2)+b_3+b_6+b_7=b_2+1+1+1=0\Rightarrow b_2=1$$
($a_1=1$, $a_3=1$, $a_4=1$)

$$(b_1)+b_3+b_5+b_7=b_1+1+0+1=0\Rightarrow b_1=0,$$
($a_1=1$, $a_2=0$, $a_4=1$)

а следовательно, $b=0\,1\,1\,0\,0\,1\,1$.

Схема декодирования. Пусть принято слово $c=b+e$, где b – кодовое слово, e – ошибка. Тогда $bM=0$, а следовательно, $(b+e)M=bM+eM=eM$. Если $eM=0$, то считается, что ошибок не было. Это действительно так при $e=0$. Если вектор ошибок e имеет только одну единицу в i-й позиции, то eM есть вектор, совпадающий с i-й строкой матр. M, являющейся двоичным разложением числа i. В этом случае следует прибавить к i-й позиции слова $b+e$.

Пример 7.9. Рассмотрим $(4,7)$–код Хемминга. Декодируем принятое слово $(a)c=1011100$, полученное при передаче некоторого кодового слова b, предполагая, что при передаче произошла ошибка не более чем в одной позиции. Имеем:

$$cM=1\,0\,(1)\,1\,0\,0\;\begin{pmatrix}0&0&1\\0&1&0\\0&1&1\\1&0&0\\1&0&1\\1&1&0\\1&1&1\end{pmatrix}=0\,1\,1\;(=3)$$

Т.о., при передаче кодового слова произошла ошибка в 3-ей позиции, исправляя которую, получаем, что было передано слово $b=(1)(0)\,0\,1\,0\,0$, декодируя кот., нах. сообщ. $a=0100$

Пример 7.10. Доказать, что (4,7)-код Хемминга является матричным (а следовательно, и групповым). Определить порождающую матрицу кода.

Решение. Укажем порождающую матрицу $G = G_{4\times 7}$ такую, что

$$b_1 b_2 b_3 b_4 b_5 b_6 b_7 = a_1 a_2 a_3 a_4 G.$$

Из системы уравнений

$$\begin{cases} b_1 = a_1 + a_2 + a_4, \\ b_2 = a_1 + a_3 + a_4, \\ b_3 = a_1, \\ b_4 = a_2 + a_3 + a_4, \\ b_5 = a_2, b_6 = a_3, b_7 = a_4, \end{cases}$$

следует, что порождающая матрица (4,7)-кода Хемминга имеет вид:

$$G = \begin{bmatrix} 1\,1\,1\,0\,0\,0\,0 \\ 1\,0\,0\,1\,1\,0\,0 \\ 0\,1\,0\,1\,0\,1\,0 \\ 1\,1\,0\,1\,0\,0\,1 \end{bmatrix}.$$

Для **«быстрого» умножения** вектора $a = a_1 a_2 a_3 a_4$ на G складываем (поэлементно, по модулю 2) строки матрицы G с номерами этих строк, соответствующими номерам позиций символа 1 в векторе a. Например,

$$1010G = 1\,1\,1\,0\,0\,0\,0 + 0\,1\,0\,1\,0\,1\,0 = 1\,0\,1\,1\,0\,1\,0.$$

Следует также иметь в виду тривиальные случаи:

$$0000G = 0\,0\,0\,0\,0\,0\,0,\ 1111G = 1\,1\,1\,1\,1\,1\,1.$$

Заметим, что сумма элементов в каждом столбце матрицы G равна 1. Из этого следует простой способ умножения a на G в случае наличия в a ровно одного элемента 0. В этом случае aG - вектор, двойственный к строке матрицы G, номер которой совпадает с номером нулевого элемента в a.

Пример 7.10. Получим кодовое слово с помощью порождающей матрицы $G_{4\times 7}$ для сообщения $a = 1011$ (см. пример 7.8).**Решение.** $1011G = \neg(1\,0\,0\,1\,1\,0\,0) = 0\,1\,1\,0\,0\,1\,1.$

Пример 7.12. Доказать, что миним. расст. между двумя различными кодовыми словами в коде X. равно 3. Реш. Согласно теор. 7.3, миним. расст. между двумя различ. кодовыми словами в групповом коде равняется миним. весу ненулевого кодового слова, и при этом, в силу т. 7.2, миним. вес ненул. кодового слова в коде X. ≥ 3 (т.к. код X. исправляет любую ошибку в одной позиции). Т. о., чтобы убедиться справедливости доказываемого утв-ия, осталось указать кодовое слово веса 3. Им, напр., является кодовое слово 1110...0, соответствующее сообщению $a = 10\ldots0$.

7.6. Коды X. с проверкой на четность.

Добавим к кодовым словам X. еще один контрольный символ b_{2^r}, а к проверочным соотношениям (т.е. к системе уравнений $bM = 0$) еще одно соотн.:

$$b_1 + b_2 + \cdots + b_{2^r} = 0 \quad (\text{т.е. } b_{2^r} = b_1 + \cdots + b_{2^r-1})$$

Полученный (n, n)-код, где $n = 2^r - 1 - r$, $n = 2^r$, $r \geq 2$, называется кодом X. с проверкой на четность. Это код миним. веса 4, способный исправлять ошибку в одной позиции и обнаруживать ошибки в ≤ 3 позициях. Матрица M у этого кода имеет дополнительный столбец из 1 и доп. строку из 0, но с послед. 1. Так в случае $r = 3$ имеем

$$M_{2^r, r+1} = M_{8,4} = \begin{bmatrix} 0&0&1&1 \\ 0&1&0&1 \\ 0&1&1&1 \\ 1&0&0&1 \\ 1&0&1&1 \\ 1&1&0&1 \\ 1&1&1&1 \\ 0&0&0&1 \end{bmatrix}$$

Декодирование. Для простоты будем рассм. (4,8)-код (при $r = 3$). Пусть передано нек. кодовое слово b, а принято

слово $c = b + e$. В случае ошибки в одной позиции добавленное соотн. очевидно, нарушится и в последней позиции слова cM будет стоять 1. При этом первые три символа слова $cM = eM$ дадут двоичное разложение номера позиции, в кот. произошла ошибка (если первые три символа нулевые, то ошибка в 8-й позиции).

Если же произошли ошибки в двух позициях, то добавленное соотношение будет, очевидно, выполнено, т.е. в последней позиции слова cM будет стоять 0. Однако, первые три символа слова cM будут содержать по крайней мере один ненулевой символ (очевидно, что эти символы не укажут нам местонож. ошибок).

В случае $cM = 0$ считаем, что ошибок не было (заметим, что равенство $cM = 0$ невозможно при 1, 2, 3 и любом нечетн. числе ошибок, но возможно при 4, 6, 8 ошибках).

Пример 7.13. Рассматривается $(4, 8)$-код Хэмм.

(1) Пусть принято слово $c = 00011100$. Тогда $cM = 111$. Считаем, что ош. произошла в одной поз. (самой уже 3ош.) и исправляем её: $b = 00011110$, $a = 0111$;

(2) Пусть принято слово $c = 01100111$. Тогда $cM = 001$. Считаем, что ошибка произошла в одной позиции (а именно, в 6-й позиции) и исправляем её: $b = 01100110$, $a = 1011$;

(3) Пусть принято слово $c = 00100001$. Тогда $cM = 010$. Считаем, что произошли ошибки в 2-ух позициях (далее только в 4) и исправить их мы не можем.