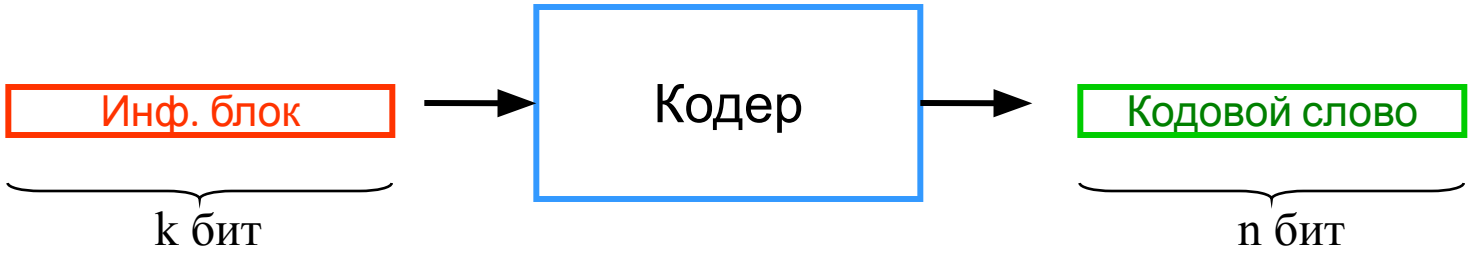
Линейные блочные коды

ОТК 2018

Линейный блоковый код

- Информационный поток бит (символов) разбивается на блоки по k бит (символов).
- Каждый блок кодируется кодовым словом из n бит (символов).



$r = n\text{-}k$ избыточные, проверочные биты

$R_c = \frac{k}{n}$ Техническая скорость кода

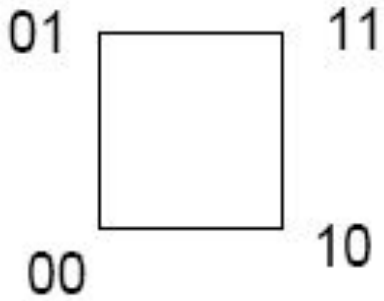
Линейный блоковый код (ЛБК) Геометрический подход

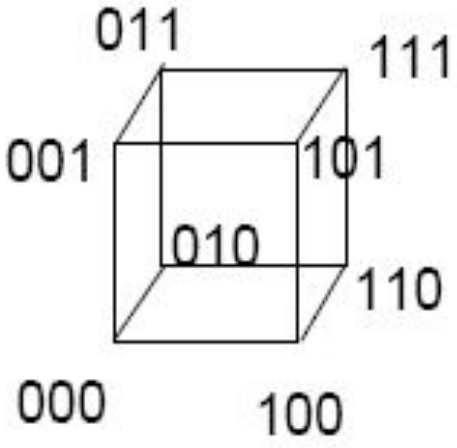
- Линейный блоковый код (*n*,*k*)
  - Множество $C \subset V_n$ с мощностью 2^k называется ЛБК если и только если существует подпространство векторного пространства V_n .

$$V_k \to C \subset V_n$$

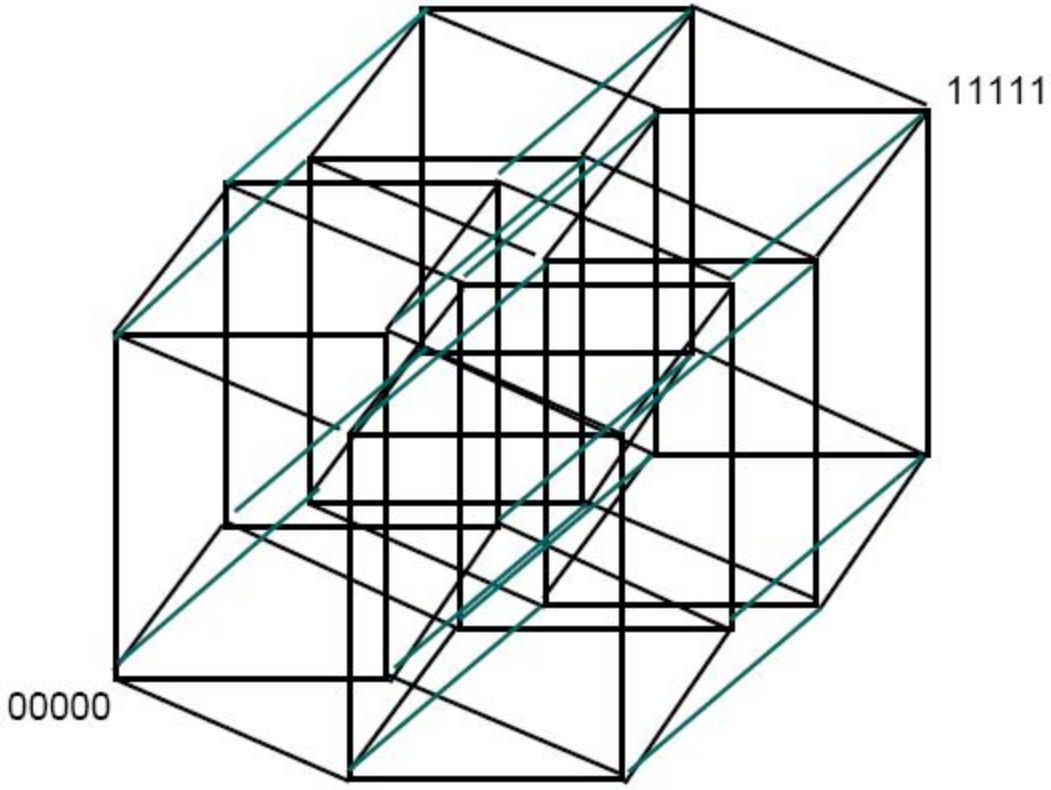
  - Элементы С называются кодовыми словами.
  - Нулевое слово – является словом кода.
  - Любая линейная комбинация кодовых слов есть слово того же кода справедливо для линейного группового кода (свойство замкнутости).

Геометрия пространства
Пример





Пятимерный куб Хэмминга



Алгебраический подход

- Кодирование (n,k) блочного кода

$$\mathbf{C} = \mathbf{aG}$$

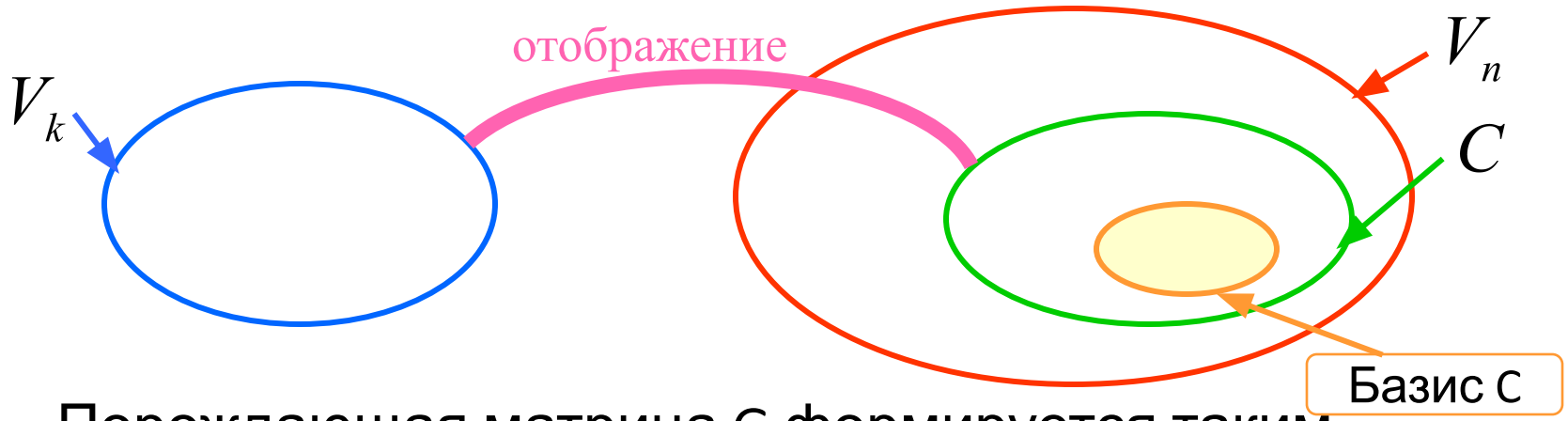
G –порождающая, генераторная матрица

$$(c_0, c_1, \mathrm{K}\ , c_{n-1}) = (a_0, a_1, \mathrm{K}\ , a_{k-1}) \begin{bmatrix} \mathbf{V}_0 \\ \mathbf{V}_1 \\ \mathrm{M} \\ \mathbf{V}_{k-1} \end{bmatrix}$$

$$(u_1, u_2, \mathrm{K}\ , u_n) = a_0 \cdot \mathbf{V}_0 + a_1 \cdot \mathbf{V}_1 + \mathrm{K} + a_{k-1} \cdot \mathbf{V}_{k-1}$$

  – Строки **G** - линейно независимы.

Порождающая матрица



– Порождающая матрица G формируется таким образом, чтобы её строки
– были векторами базиса,

$$\{\mathbf{V}_0, \mathbf{V}_1, \boxtimes, \mathbf{V}_{k-1}\}$$

.

$$\mathbf{G} = \begin{bmatrix} \mathbf{V}_0 \\ \boxtimes \\ \mathbf{V}_{k-1} \end{bmatrix} = \begin{bmatrix} v_{0,0} & v_{0,1} & \boxtimes & v_{0,n-1} \\ v_{1,0} & v_{1,1} & \boxtimes & v_{1,n-1} \\ \boxtimes & & \boxtimes & \boxtimes \\ v_{k-1,0} & v_{k-1,1} & \boxtimes & v_{k-1,n-1} \end{bmatrix}$$

Пример

- Блочный код (6,3)

$$\mathbf{G} = \begin{bmatrix} \mathbf{V}_0 \\ \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix} = \begin{bmatrix} 1\,1\,0\,1\,0\,0 \\ 0\,1\,1\,0\,1\,0 \\ 1\,0\,1\,0\,0\,1 \end{bmatrix}$$

Вектор сообщения	Слова кода
000	000000
100	110100
010	011010
110	101110
001	101001
101	011101
011	110011
111	000111

Пример 2

Код повторений **C** = {00000,11111}

G=[11111]

Код с проверкой на четность (один избыточный символ

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

Систематический блоковый код

- Систематичный блоковый код (*n*,*k*)
  - Для систематического кода, первые (или последние) *k* символов являются информационными.

$$\mathbf{G} = [\mathbf{P} \vdots \mathbf{I}_k] \quad \text{или} \quad \mathbf{G} = [\mathbf{I}_k \mid \mathbf{P}]$$

$\mathbf{I}_k = k \times k$ единична — матрица

$\mathbf{P}_k = k \times (n - k)$ — подматрица

$$\mathbf{C} = (c_0, c_1, \ldots, c_{n-1}) =$$

$$= (\ p_0, p_1, \ldots, p_{n-k-1}\ ,\ a_0, a_1, \ldots, a_{k-1}\ )$$

проверочные символы — информационные символы

Проверочная матрица

- Для любого линейного кода можно найти матрицу $\mathbf{H}_{(n-k)\times n}$, такую, что её строки будут ортогональны строкам **G** :

$$\mathbf{G}\mathbf{H}^T = \mathbf{0}$$

- **H** называется проверочной матрицей, её строки линейно независимы *t*.
- Для систематического блочного кода :

$$\mathbf{H} = [\mathbf{I}_{n-k} \vdots -\mathbf{P}^T] \quad \text{или} \quad \mathbf{H} = [-\mathbf{P}^T \vdots \mathbf{I}_{n-k}]$$

$$\mathbf{H}\mathbf{G}^T = [-\mathbf{P}^T \mid \mathrm{I}_{n-k}][\mathrm{I}_k \mid \mathbf{P}]^T = -\mathbf{P}^T\mathrm{I}_k + \mathrm{I}_{n-k}\mathbf{P}^T = 0$$

Определение кода через проверочные уравнения:

- $C = \{\boldsymbol{x} \in F\colon H\,\boldsymbol{x}^T = 0\}$
- Рассмотрим код длины 6: $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5, x_6)$
- Положим, что

$$x_1 + x_2 + x_3 + x_4 = 0$$
$$x_2 + x_3 + x_5 = 0$$
$$x_1 + x_3 + x_6 = 0$$

- Система уравнений имеет решения x_1, x_2, x_3, для заданных x_4, x_5, x_6
- Проверочные уравнения задает проверочная матрица

$$H\,\boldsymbol{x}^T = 0$$

Систематический код

- Пусть $\mathbf{G} = [\mathrm{I}_k \mid \mathbf{P}]$,
- $\mathbf{P} = [\, p_{i,j}]$ - подматрица размером $k \times (n - k)$,
- $\mathbf{p}_1, \ldots, \mathbf{p}_k$ – строки $\mathbf{P}$
- тогда
- $\mathbf{c} = \mathbf{aG} = (a_0, \ldots, a_{k-1}, \mathbf{p})$,
- где $\mathbf{p} = \Sigma_j \, a_j \mathbf{p}_j$
- Проверочная матрица
- $\mathbf{H} = [-\mathbf{P}^T \mid \mathrm{I}_{n-k}] = [\, h_{i,j}]$.
- Элементы $h_{i,j}$ удовлетворяют системе уравнений
  - $\mathbf{Ha}^T = 0$,
- где $\mathbf{a} = (a_0, \ldots, a_{k-1}, p_0, p_1, \ldots, p_{n-k-1})$

Ортогональное подпространство H

- Пусть код **C={c**$_i$**}** -> **G**

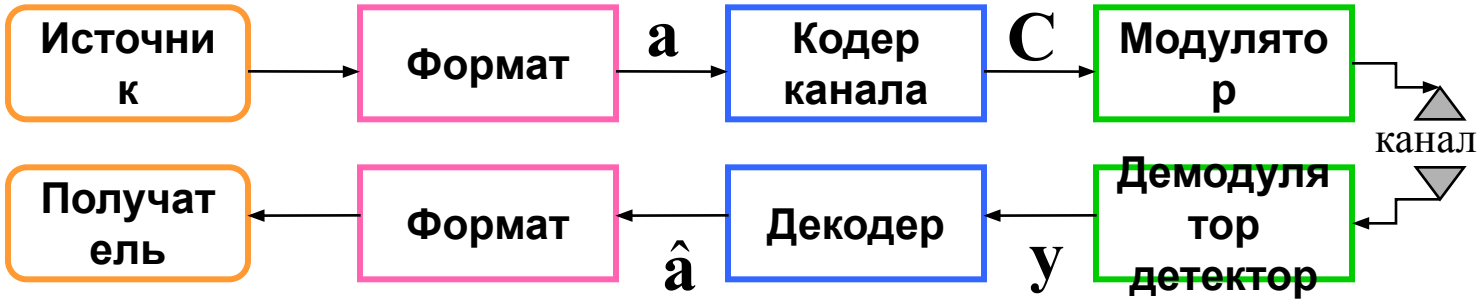
- Определим код через нуль-пространство **H**

- $\mathbf{C} = \{\mathbf{c} \in F: \mathbf{H}\mathbf{c}^T = 0\}$

- где *F* – поле (множество) элементов кода
- Матрица **H** аннулирует кодовые слова

- Образуется **ортогональное подпространство** относительно кода **C**

Синдром ошибки



$$\mathbf{y} = \mathbf{c} + \mathbf{e}$$

$\mathbf{y} = (y_0, y_1, ..., y_{n-1})$ принятое кодовое слово или вектор

$\mathbf{e} = (e_0, e_1, ..., e_{n-1})$ вектор ошибок

- Синдром (отображение ошибок):
  - **S** является синдромом **y**, соответствующий отображению вектора ошибок **e** в **H**.

$$\mathbf{S} = \mathbf{y}\mathbf{H}^T = \mathbf{e}\mathbf{H}^T$$

Двойственный (дуальный) код

- Если ***C*** – линейный код размером *k*, то *ортогональным или двойственным* ему будет код $C^{\perp}$, определяемый как

$$C^{\perp} = \left\{ \mathbf{x} \in R^n \middle| \forall \mathbf{y} \in C, (\mathbf{x}, \mathbf{y}) = 0 \right\}$$

где $(\mathbf{x}, \mathbf{y}) = \sum_i x_i \cdot y_i$

- Двойственный код является (*n*, *n* - *k*)-кодом. Если **H** – это порождающая матрица двойственного кода, то матрицу **H** называют проверочной матрицей кода **C**

Двойственные (дуальные) коды
Пример. Троичные коды

Троичный алфавит $Q=\{0,1,2\}$ Операции по mod 3

Пространство Q^n $n = 4$

Полный код

000,001,002,010,011,012,020,021,022,100,200,101,….

Код (4, 2)

$$G=\begin{pmatrix}0&1&2&1\\1&1&1&1\end{pmatrix}$$

$$H=\begin{pmatrix}1&1&1&0\\0&2&0&1\end{pmatrix}$$

Комбинаторный подход
Пространство Хэмминга

- X– конечное q-элементное множество
- X^n состоит из всех n-ок (n – мерных векторов) с координатами из множества X
- *Пусть* $\boldsymbol{a} = (a_0, \ldots, a_{n-1})$; $\boldsymbol{b} = (b_0, \ldots, b_{n-1})$ – векторы множества X^n
- ***Метрика Хемминга*** $d(\boldsymbol{a},\boldsymbol{b})$

- $d(\mathbf{a}, \mathbf{b}) =$ число позиций j, для которых $a_j \neq b_j$

Справедливо неравенство треугольника

$$d(\mathbf{a}, \mathbf{b}) \leq d(\mathbf{a}, \mathbf{c}) + d(\mathbf{c}, \mathbf{b})$$

Пространство X^n – вместе с метрикой d называется метрическим пространством Хэмминга

Вес Хэмминга

- **Вес Хэмминга** вектора **U**, обозначается как *wt*(**U**), определяется как число ненулевых элементов в **U**

Комбинаторные соотношения

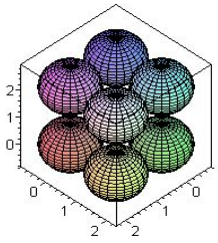
- . Число различных кодовых слов длины *n* и веса *t*
- $$|\{\mathbf{x}\in\mathbf{F};wt(\mathbf{x})=t\}|=\binom{n}{t}=C_n^t=\frac{n!}{t!(n-t)!}$$
- Если 0 ≤*t* ≤*n* и **c** – кодовое слово длины *n*, тогда число различных слов длины *n* и расстоянием по отношению к **c** не более *t* равно
- $$\binom{n}{0}+\binom{n}{1}+\boxtimes\quad+\binom{n}{t}$$

Комбинаторно-геометрический подход
Сферическая модель

- Определим множество $S(\mathbf{c}_i, \rho) = \{\mathbf{c}_j \in R^n \mid d(\mathbf{c}_i, \mathbf{c}_j) \le \rho\}$
- которое описывает геометрическую фигуру шара радиусом ρ с центром в $\mathbf{c}_i$
- Сфера шара описывается множеством

$$S_{=}(\mathbf{c}_i, \rho) = \{\mathbf{c}_j \in R^n \mid d(\mathbf{c}_i, \mathbf{c}_j) = \rho\}$$

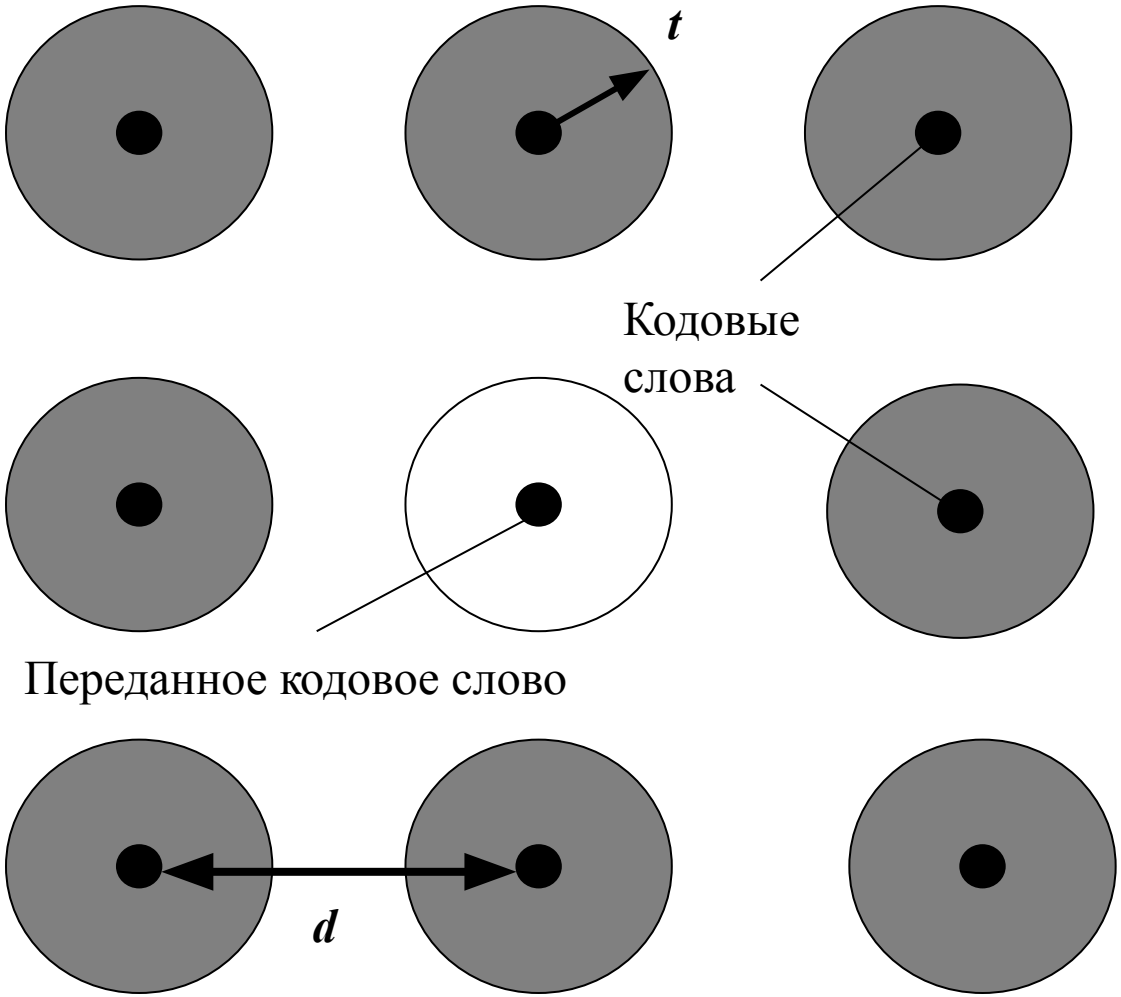
- Количество элементов шара или мощность шара
- *Vol*(**c**,ρ) = $\left|S(\mathbf{c}, \rho)\right| = \sum_{i=0}^{\rho} (q-1)^i C_n^i$



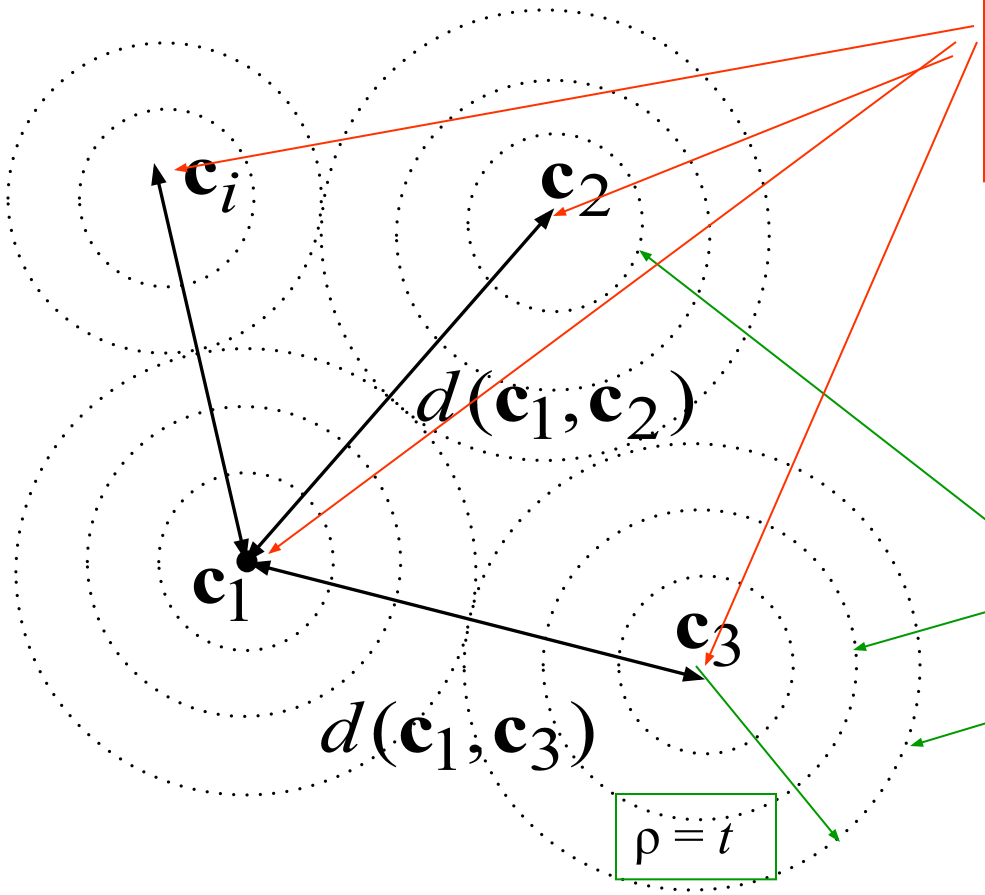
Сферическая модель

Слова корректирующего кода задают координаты центров пространственно расположенных шаров, имеющих радиус *t*.

Внутри шара располагаются слова, получающие из кодового слова в результате искажения *t* символов (так называемые запрещенные комбинации).



Сферическая модель



$\{\mathbf{c}_j\}$ – кодовые слова «Разрешенные комбинации»

Искаженные кодовые слова

⇒ Кодовое слово + t ошибок

⇒ Модель

сфера радиусом $\rho = t$

⇒ «Запрещенные комбинации»

Корректирующая способность

- ***Число обнаруживаемых ошибок*** . Потребуем, чтобы сферы не захватывали соседние центры.
- В этом случае центры шаров однозначно различимы, но при этом кодовое расстояние должно удовлетворять соотношению

$$d(C) = d_{\min} \geq t_{обн} + 1 \;\Rightarrow\; t_{обн} = d_{\min} - 1$$

$$t_{обн} = d_{\min} - 1$$

Гарантированное число исправляемых ошибок

- Возьмём кодовые вектора и образуем вокруг них сферы радиусом *t* и потребуем, чтобы эти сферы не пересекались, что гарантирует исправление ошибок.
- Условия для требуемого кодового расстояния:

$$d_{\min} = 2t_{испр} + 1\ , \qquad t_{испр} = \frac{d_{\min} - 1}{2}$$

$$t = \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor$$

Обнаружение и исправление ошибок

Это случай является композицией первых двух:

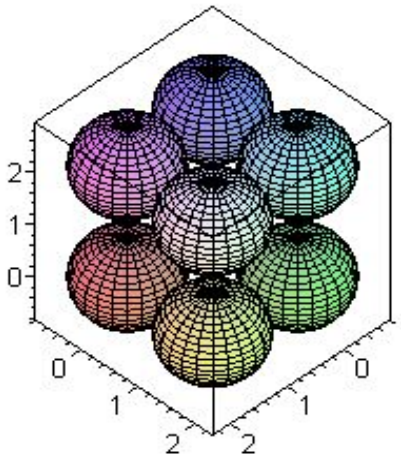
$$d_{\min} \geq t_{испр} + t_{обн} + 1$$

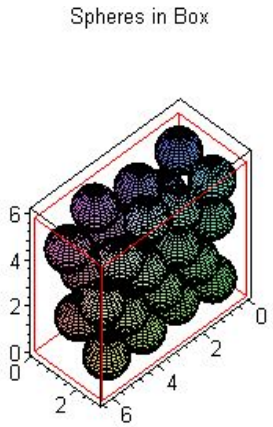
- *Примеры.*
- 1. Пусть $n = 15$, $d_{\min} = 5$, то $t_{испр} = 2$.
- 2. Если следует исправить 2 ошибки и обнаружить 4, то требуется $d_{\min} = 7$.

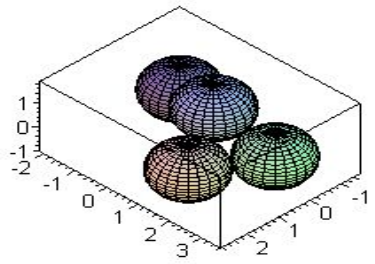
Задача оптимальной упаковки:

максимальное заполнение объема для заданного расстояния $d_{\min}$

Совершенный код







Код $\mathbf{C} \subset \boldsymbol{R}^{n}$, исправляющий ошибки, называются *совершенным*, если

Коды Хэмминга

Рида–Соломона

$$\bigcap_{\mathbf{c}\in C} S(\mathbf{c}, e) = R^n$$

Подход на основе теории графов

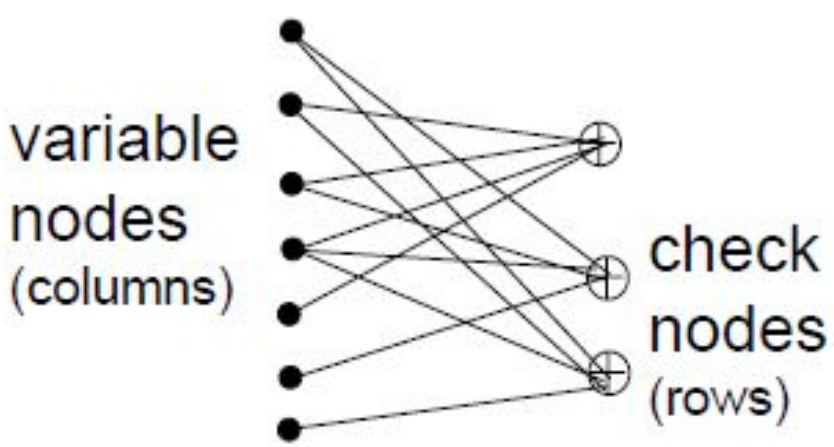
- Код Хэмминга (7,4,3)

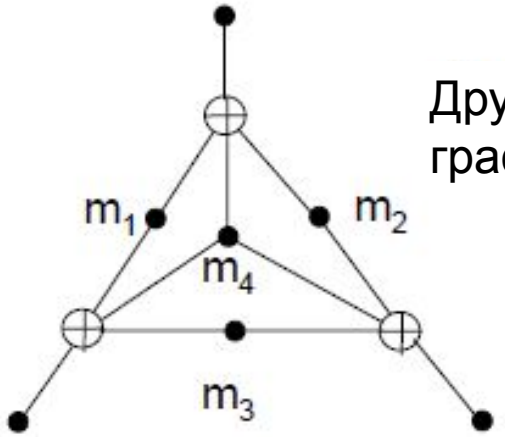
$$H = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad H = \begin{bmatrix} m_1 & m_2 & m_3 & m_4 & & & \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Несистемати-ческий
Систематический

- Граф Таннера





Другая форма графа

Алгебро-геометрический подход

- Определение кода через функцию отображения полинома
  - $eval\big(f(x)\big)$.
- Функция отображения полинома $f(x)$ в вектор.

$$c \in \left(\mathbb{F}_q\right)^n,$$

$$f \longmapsto c = (c_1, c_2, \dots, c_n),$$

- где $c_i = f(\alpha_i), i = 1,2, \dots, n$
- $\mathbb{F}_q = \left(\alpha_0 = 0, \alpha_1, \dots, \alpha_{q-1}\right)$ – конечное поле

Пример

- Задается множество (поле). В полином подставляются вместо x значения элементов поля. Вычисляется вектор.
- *Пример*. Пусть $\mathbb{F}_7 = (0,1,2,3,4,5,6)$, примитивный элемент $\alpha = 3$
- $\mathcal{F} = \{1, \alpha, \alpha^2, \dots, \alpha^5\} = \{1,3,2,6,4,5\}$

- $f(x) = 2x + 1 \qquad c = eval(f) = (3,0,5,6,2,4)$
-
- $f(x) = 3x^2 + x + 2 \qquad c = eval(f) = (6,4,2,4,5,5)$
-

Математика
Векторное пространство

- Пусть V будет множеством векторов и F поле (множество) элементов называемых скалярами. V формирует векторное пространство над полем F если выполняются :

1. Закон коммутативности: $\forall \mathbf{u}, \mathbf{v} \in V \Rightarrow \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} \in F$

2. $\forall a \in F, \forall \mathbf{v} \in \mathbf{V} \Rightarrow a \cdot \mathbf{v} = \mathbf{u} \in \mathbf{V}$

3. Закон дистрибутивности:

$$(a+b)\cdot \mathbf{v} = a\cdot \mathbf{v} + b\cdot \mathbf{v} \quad \text{и} \quad a\cdot(\mathbf{u}+\mathbf{v}) = a\cdot \mathbf{u} + a\cdot \mathbf{v}$$

4. Ассоциативность:

5. $\forall a, b \in F, \forall \mathbf{v} \in V \Rightarrow (a\cdot b)\cdot \mathbf{v} = a\cdot(b\cdot \mathbf{v})$

$$\forall \mathbf{v} \in \mathbf{V},\ 1\cdot \mathbf{v} = \mathbf{v}$$

Пространство и подпространство

- Пример векторного пространства
  - Множество бинарных *n*-ок, образует пространство V_n

$$V_4 = \{(0000),(0001),(0010),(0011),(0100),(0101),(0110),(0111),(1000),(1001),(1010),(1011),(1100),(1101),(1110),(1111)\}$$

- **Векторное подпространство:**
  - Подмножество *S векторного пространства* V_n называется подпространством, если :
    - Нулевой вектор находится в *S*.
    - Сумма двух векторов в *S* также располагается в *S*.

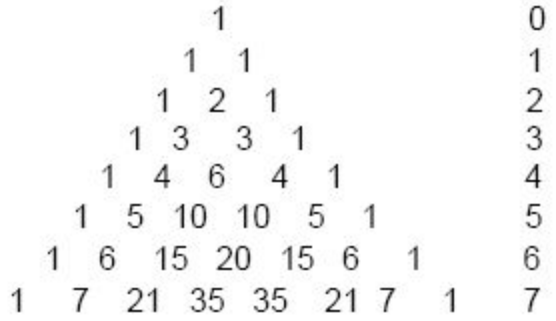
  Пример:

$\{(0000),(0101),(1010),(1111)\}$ есть подпространство V_4.

Комбинаторные соотношения

- **Перестановка** : (abc, acb, bac, bca, cba, cab)
- $$n(n-1)(n-2)\ldots 2\cdot 1 = n!$$
- ❤️ число вариантов выбрать упорядоченный кортеж из k символов в множестве из n символов
- $n(n-1)(n-2)\ldots(n-k+2)(n-k+1) = (n)_k$
- ❤️ число вариантов выбрать неупорядоченный кортеж из k символов в множестве из n символов

$$\binom{n}{k} = C_n^k = \frac{(n)_k}{k!}$$



Вопросы

- ?