Лекция 1

Оформила: Балакшина А.Д. 5030102/20101

9 февраля 2026 г.

Все данные, для которых существует возможность потери или искажения, всегда кодируются с целью эти искажения брать.

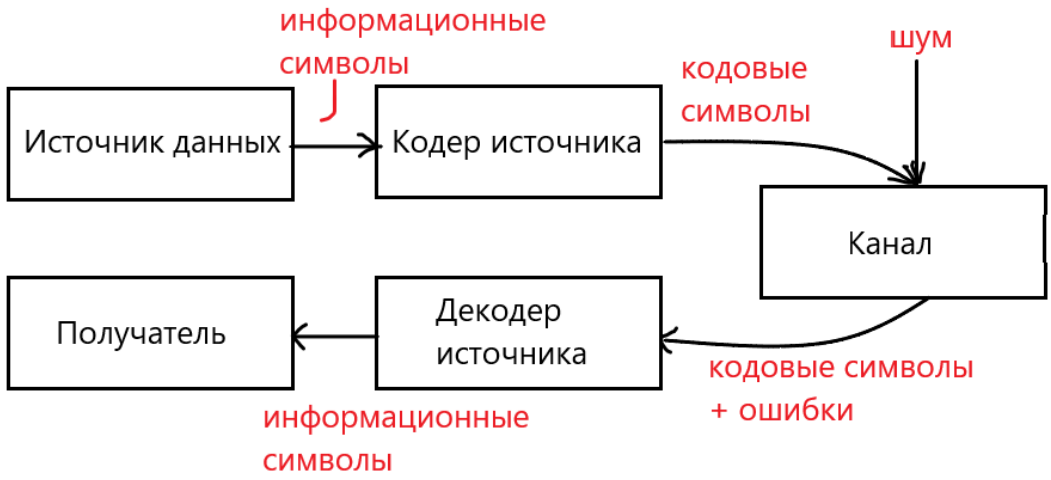
1 Упрощенная модель цифровой системы связи

Есть источник данных (флешка, компакт диск, человек)
Выход этого источника данных - информационные символы
Кодер источника - устройство, которое ставит в соответствие информационным символам кодовые

1. Мы работаем с дискретными последовательностями
2. Будем считать, что выходящая информац. посл-ть будет состоять из двух элементов $GD(2) = 0, 1$.

Источник генерирует посл-ть нулей и единиц, после кодера выходят тоже битовые символы, но (возможно) другие. Их обычно больше.

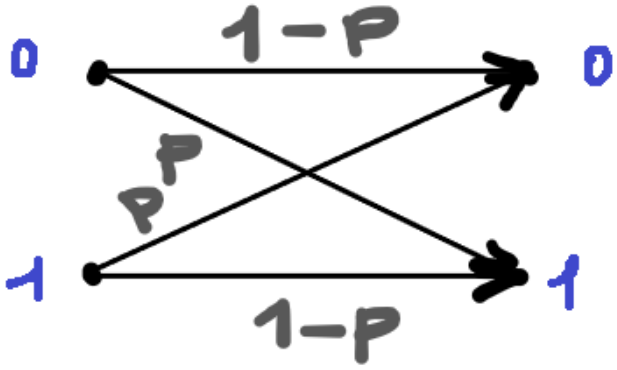
Мы будем рассматривать следующую математическую модель канала: **двоичный симметричный канал**. На вход его последовательно подаются 0 и 1.



2 Переходная вероятность

Поданный на канал сигнал может быть интерпретирован неверно с вероятностью p - **переходной вероятностью**.

Двоичный симметричный канал



Процедура перевода из информационных к кодовым символам - **кодирование**
Если кодируем одним символом - вероятность ошибки - p
Что мы можем сделать, чтобы улучшить качество передачи данных?

- Дублировать данные:
  Дублируем по три (если бы было два - мы бы смогли сказать, что ошибка произошла, но не было бы понятно - какая)

$$0 \to 000$$

$$1 \to 111$$

  Скорость кода $1/3$
  Таким образом, улучшили дублированием вероятность, но получили уменьшение в скорости
  Может исправить 1 ошибку.

- Выходящие из источника данные склеим парами

$$00 \to 00000$$
$$01 \to 10110$$
$$10 \to 01011$$
$$11 \to 11101$$

  ДЗ: Посчитать вероятность ошибки при использовании такого кода при $p = 10^{-3}$

  Этот код линеен (сумма любых двух даёт код оттуда же)

  Исправляет одну ошибку

  Но: он чуть лучше из-за скорости кода

  Скорость кода: $R = \frac{k}{n}$. Здесь он равен $2/5$

1948 Клод Шеннон
Теория информации делится на два раздела: **кодирование источников** (убираем избыточноть) и **канальное кодирование** (избыточность добавляем)
Для ДСК с переходной вероятностью p вводится понятие **пропускная способность** $c = 1 - h(p)$
$h(p) = -x \log_2 x - (1-x) \log_2 (1-x)$
При скорости передачи R меньшей величины пропускной способности C может быть обеспечено сколь угодно малая вероятность ошибки декодирования за счёт увеличения длины используемых кодов, что ведёт к увеличению сложности кодирования/декодирования
Если $R > C$, то надежная передача невозможна

3 Как сравнить близость кодовых слов?

Если x - кодовое слово (двоичный вектор), то $\omega(x)$ - **вес Хэмминга**, определяется как число ненулевых элементов в x (в двоичном случае - число единиц)
Расстояние Хэмминга между двумя кодовыми словами x и y - d(x, y) - количество элементов слова, которые отличаются друг от друга

Пример:

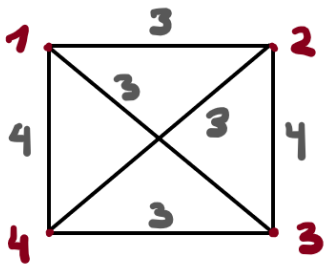
$$x = 001101, \omega = 3$$
$$y = 101001, \omega = 3$$
$$d(x, y) = 2$$

ДЗ Показать, что расстояние Хэмминга удовлетворяет аксиомам рассстояния и может быть использовано как метрика)

$$d(x, y) = \omega(x + y) (\text{побитовое сложение})$$
$$d(x, 0) = \omega(x)$$

1. $00 \to 00000$
2. $01 \to 10110$
3. $10 \to 01011$
4. $11 \to 11101$

	$x=1$	$x=2$	$x=3$	$x=4$
$y=1$	0	3	3	4
$y=2$	3	0	4	3
$y=3$	3	4	0	3
$y=4$	4	3	3	0



То расстояние, которое окажется меньшим, и будет выбрано
$d_{min} = mind(x, y), x \neq y$ - минимальное расстояние кода
ДЗ: пусть некоторый код исправляет ошибки кратности t
Показать, что $t <= [\frac{d_{min}-1}{2}]$
[x] - наибольшее целое, не превышающее x

4 Линейный код

Линейный код - это код, в котором сумма двух любых кодовых слов тоже является кодовым словом
код C - множество кодовых слов

$$\forall x, y \in C : (x + y) \in C$$

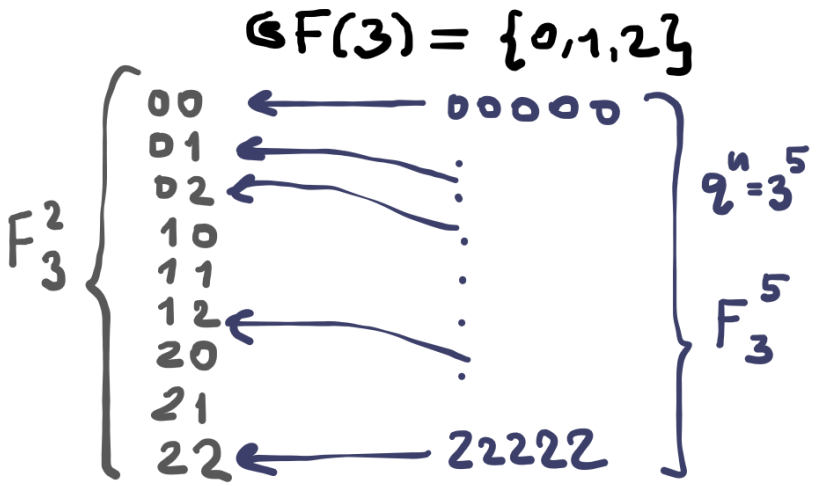
ДЗ: доказать

$$d(x, y) = \omega(x + y) = \omega(z) = \omega(z + 0) = d(z, 0)$$

$$d_{min} = \min_{\text{x,y}\in\text{C,x}\neq\text{y}} d(x, y) = \min_{\text{z}\in\text{C,z}\neq 0} \omega(z)$$

Линейный q-ичный (n, k)-код (поле GF(q)) (где n - число кодовых символов, k - информационных) - это любое k-мерное подпространство пространства F_q^n - всевозможных векторов длины n.

Пример: пусть q=3, k=2, n=5
Информационных слов $q^k = 9$, мы их отображаем на n=5: Из векторов длины 5 выбираем подпространство из 9-ти элементов.



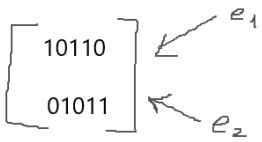
В пространстве существует базис.
Пример: возьмём в качестве базисных векторов два ненулевых

c_1 00000

$e_1 = c_2$ 10110

$e_2 = c_3$ 01011

c_4 11101



Тогда:

$$0 \cdot e_1 + 0 \cdot e_2 = c_1$$
$$0 \cdot e_1 + 1 \cdot e_2 = c_1$$
$$1 \cdot e_1 + 0 \cdot e_2 = c_1$$
$$1 \cdot e_1 + 1 \cdot e_2 = c_1$$

Существует соответствие между этим представлением и набором кодовых слов.

Порождащей матрицей (n, k) кода называется матрица размера $k \times n$, где строки - базисные вектора.

Кодовые слова - линейные комбинации базисных векторов.

Обозначается: G - порождающая матрица

m - информационное слово, $m = (m_1, ..., m_k)$ 2^kштук

c - кодовое слово, $c = m \cdot G$

Предположим, что для некоторого вектора $h = (h_1, ..., h_n)$ все кодовые слова удовлетворяют $(c_i, h) = c_1 h_1 + ...c_n h_n = 0$ - сумма произведений элементов, стоящих на одинаковых позиций

c_1 00000

c_2 10110

c_3 01011

c_4 11101

h = 00111 - ортогонален коду

называется "проверкой"

$$G \cdot h^T = 0$$

всего n-k проверок

Построим матрицу H размера $(n - k, n)$ - **проверочная матрица**

$$G \cdot H^T = 0$$