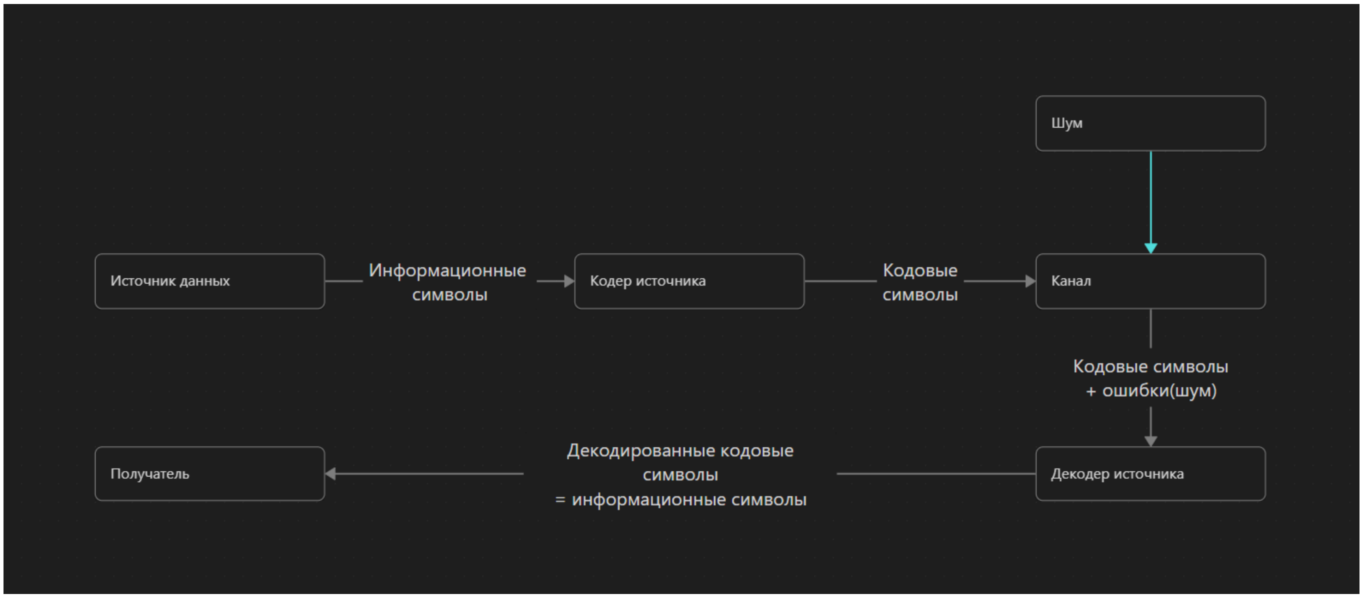
Курс сфокусирован на теории передачи данных через каналы с помехами.

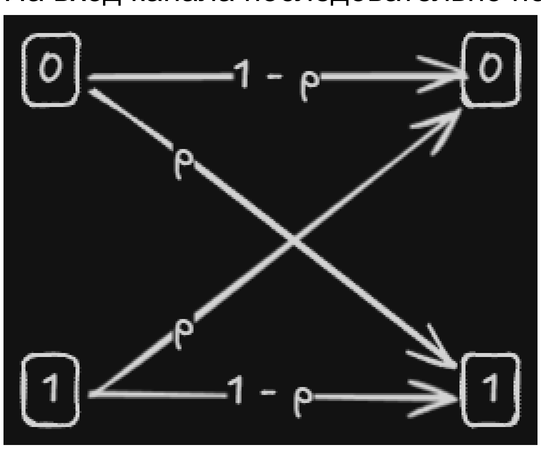
Все данные, для которых существует возможность искажения, кодируются с целью убрать эти искажения.
Упрощенная модель цифровой системы связи:



- Кодер источника - устройство, которое которое сопоставляет информационным символам кодовые символы.
- Кодовые символы обычно имеют в себе некоторую избыточность, которая позволяет им при передаче по каналу с шумом идентифицировать шум и восстановить информационные символы.

Работаем с дискретными последовательностями. Считаем, что информационная последовательность состоит из GF(2) = { 0, 1 }.
Будем рассматривать двоичный симметричный канал.

На вход канала последовательно подаются биты.



p - переходная вероятность(вероятность ошибки одного символа)

Пусть $p = 10^{-3}$. Пусть будем повторять каждый бит трижды. $0011 \to 000\,000\,111\,111$. Обозначим такой способ кодирования (*).
Если кодирования нет, то вероятность ошибки Pe = p. При выбранном выше способе кодирования вероятность ошибки $Pe = 3p^2(1-p) \approx 3 * 10^{-6}$. Это можно найти, рассматривая единственное кодовое слово, состоящее из всех нулей, т.к. код линеен. Наш способ кодирования позволяет исправить одну ошибку на длине 3, скорость R = 1/3.

Рассмотрим иной способ кодирования:

00	**00000**
01	10110
10	01011
11	11101

Такой способ кодирования обозначим (**).
Можно показать, что код линеен. Можно показать, что такой код исправляет, как и ранее, одну ошибку.
Скорость кода $R = k/n$, где k - число информационных символов, n - число символов в кодовом слове.

R(*) = 1/3 < R(**) = 2.5. Итак, код (**), как и (*) исправляет одну ошибку, однако экономичнее.
Изменяя способ кодирование мы можем влиять на скорость и количество ошибок.

Пропускная способность $C = 1 - h(p)$, где $h(x) = -xlog_2(x) - (1-x)log_2(1-x)$ - энтропия двоичного ансамбля.

> **Утверждение:** Для ДСК с переходной вероятностью p, при скорости передачи R, меньшей величины пропускной способности C, может быть обеспечена сколь угодно малая вероятность ошибки декодирования за счет увеличения длины используемых кодов, что ведет к увеличению сложности кодирования и декодирования. Если R > C, надежная передача невозможна.

Вес Хемминга.
Если x - кодовое слово, то $\omega(x)$ - вес Хемминга и определяется как число ненулевых элементов в x. В двоичном случае это число единиц.

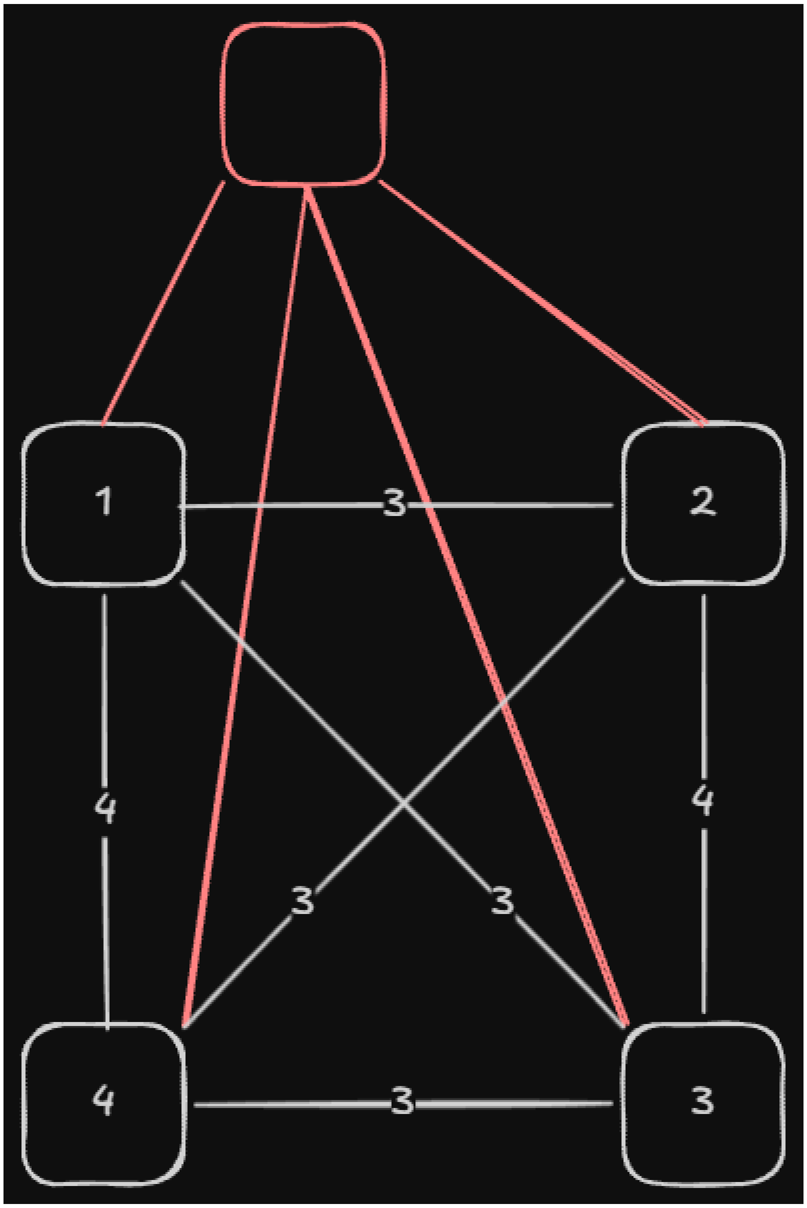
Расстояние Хемминга $d(x, y)$ определяется как количество элементов слова, которые отличаются друг от друга.

> **Example:**
> x = 001101;
> y = 101001;
> $\omega(x) = 3$
> $\omega(y) = 3$
> $d(x, y) = 2$

В двоичном случае $d(x, y) = \omega(x + y)$, где сложение происходит побитово по модулю 2.
Отсюда, $d(x, 0) = \omega(x)$

Вернемся к (**).
Расстояния между словами:

	1	***2***	***3***	***4***
1	0	3	3	4
2	3	0	4	3
3	3	4	0	3
4	4	3	3	0



Делаем выбор в пользу слова, расстояние до которого минимально.
Минимальное расстояние кода $d_{min} = \min_{x \neq y} d(x, y)$.

Линейный код - код, в котором сумма двух любых кодовых слов тоже является кодовым словом. Обозначим С - множество всех кодовых слов.

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

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

В общем случае работаем с линейным q-ичным кодом, т.е. на GF(q).
Линейный q-ичный (n, k) - код - $\forall$ k-мерное подпространство пространства F_q^n

всевозможных векторов длины n.
Мы говорим о линейном подпространстве, значит в нем существует некоторый базис.

Посмотрим на (**). В качестве базиса возьмем
$e_1 = 10110;$
$e_2 = 01011$
Тогда
$c_1 = 0 * e_1 + 0 * e_2;$
$c_3 = 0 * e_1 + 1 * e_2;$
$c_2 = 1 * e_1 + 0 * e_2;$
$c_4 = 1 * e_1 + 1 * e_2.$

Существует соответствие между представлением в виде набора базисных векторов и кодовыми словами. Такое соответствие называется порождающей матрицей.
Порождающей матрицей G (n, k)-кода называется матрица размера k*n, где строки - базисные вектора.

Кодовые слова - линейные комбинации базисных векторов.
Информационное слово $\bar{m} = (m_1, m_2, \ldots, m_k)$.
Кодовое слово $\bar{c} = \bar{m} * G$.

Предположим, что для некоторого вектора $\bar{h} = (h_1, h_2, \ldots, h_n)$ все кодовые слова удовлетворяют $(\bar{c_i}, \bar{h}) = c_1 * h_1 + c_2 * h_2 + \ldots + c_n * h_n$, где $\bar{c_i} = (c_1, c_2, \ldots, c_3)$
Взглянем снова на (**). Возьмем h = 00111. Такой h ортогонален коду. Назовем его проверкой. Т.к. все кодовые слова получаются из порождающей матрицы, можно показать, что $G * h^T = 0$. Таких проверок $n - k$.
Можно построить проверочную матрицу H размера (n-k, n). Тогда $G * H^T = \emptyset$ и $c * H^T = \emptyset$.