

# Кодирование информации

- Один из видов преобразования ин-и
  - Кодирование источника
  - Канальное кодирование (в основном – избыточное)
- Основные методы преобразования ин-и:
  - **Избыточное кодирование** (для увеличения надежности)
  - **Сжатие** (компрессия, архивация) (для снижения объема файлов)
  - **Криптографическое и стеганографическое** преобразования (для повышения уровня конфиденциальности)

# Методы структурной, информационной и временной избыточности в ИВС

1

## Основные цели применения

- обеспечение качества и надежности ИВС (программных, аппаратных, аппарат-прогр средств) - организация работ на основе международных стандартов серии **ISO 9000** или *TQM (Total Quality Management)* - Международная организация по стандартизации (*International Standard Organization, ISO*)

- используемые методы

1) увеличение наработки, 2) снижение интенсивности отказов, 3) улучшение восстанавливаемости, 4) резервирование - **третью и четвертую группы можно объединить под единым названием – избыточных методов**

- различают: **структурную, временную, информационную избыточность** (redundancy) либо их комбинации

- простая структурная избыточн - **структурное резервирование**

- системы с **временной избыточностью** – системы с повторениями (передачи)

Классика: Реализуется введением в структуру ТС накопительного звена, позволяющего в течение определенного времени выполнять основную функцию при отказе элемента за накопительным звеном. Если избыточное время будет выше времени восстановления отказавшего элемента, то функция по назначению будет выполняться непрерывно и даже при отказе.

- **информационная избыточность** – обуславливает возможность применения функций **сжатия информации**

Информ избыточность сообщений  $R$  определяется по формуле

$$R = 1 - H / \log_2 k, \quad (1)$$

где  $k$  — число букв алфавита, а  $H$  — энтропия источника на букву сообщения

- **Комбинированные методы - на основе использования специальных кодов (избыточных кодов или помехоустойчивых кодов)**

# МЕТОДЫ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ ДАННЫХ

3

## Классификация кодов

- **Блочные коды** — каждому сообщению из  $k$  ( $X_k$ ) символов (бит) сопоставляется блок из  $n$  символов (кодированный вектор  $X_n$  длиной  $n=k+r$ ).
- **Непрерывные (рекуррентные, цепные, свёрточные) коды** - непрерывная последовательность символов, не разделяемая на блоки. Передаваемая последовательность образуется путём размещения в определённом порядке проверочных символов между информационными символами исходной послед-ти.
- **Систематические коды** характеризуются тем, что сумма по модулю 2 двух разрешённых кодовых комбинаций кодов снова даёт разрешённую кодовую комбинацию.

Проверочные символы вычисляются как линейная комбинация информационных, откуда и возникло другое наименование этих кодов — **линейные**. Для кодов принимается обозначение  $[n,k]$  - код,

- **Несистематические коды** не обладают отмеченными выше свойствами (к ним относятся **итеративные коды**).
- **Циклические коды** – относятся к линейным систематическим.

Основное свойство, давшее им название, состоит в том, что каждый вектор, получаемый из исходного кодового вектора путём **циклической перестановки** его символов, также является разрешённым кодовым вектором. Принято описывать циклические коды при помощи **порождающих полиномов  $G(X)$**  степени  $r$

Среди ЦК особое место занимает класс кодов, предложенных Боузом и Чоудхури и независимо от них Хоквингемом. Коды Боуза-Чоудхури-Хоквингема получили сокращённое наименование **БЧХ- коды : CRC-32 (CRC-16)**.

БЧХ- коды являются обобщением **кодов Хемминга** на случай исправления нескольких независимых ошибок ( $t_n > 1$ ).

- **Определение 1.** Сообщение  $X_k$  ( $k$  – длина сообщения, *символов* или *бит*), называется **информационным словом**.
- **Определение 2.** Избыточные символы длиной  $r$  символов (бит) составляющие **избыточное слово**  $X_r$ .
- **Определение 3.** Слово  $X_n$  длиной  $n=k+r$  символов  $X_n = X_k X_r$  называется **кодovým словом**.

Информацию содержит только информационное слово. Назначение избыточности  $X_r$  – обнаружение и исправление ошибок.

- **Определение 4.** **Вес по Хеммингу** произвольного двоичного слова  $X$  ( $w(X)$ ) равен количеству ненулевых символов в слове.

**Пример 1.**  $X=1101$ . Тогда  $w(X=1101) = 3$ .

- **Определение 5.** **Расстояние по Хеммингу** или кодовое расстояние ( $d$ ) между двумя произвольными словами ( $X, Y$ ) одинаковой длины равно количеству позиций, в которых  $X$  и  $Y$  отличаются между собой.

Кодовое расстояние можно вычислить как вес от суммы по модулю 2 этих двух слов:  $d(X, Y) = w(X \oplus Y)$ .

Пример 2.  $X=101, Y=111$ ). Очевидно, что  $d(X, Y) = 1$ .

Пример 3.  $X=1011, Y=0000$ ;  $d(X, Y)=3$ :

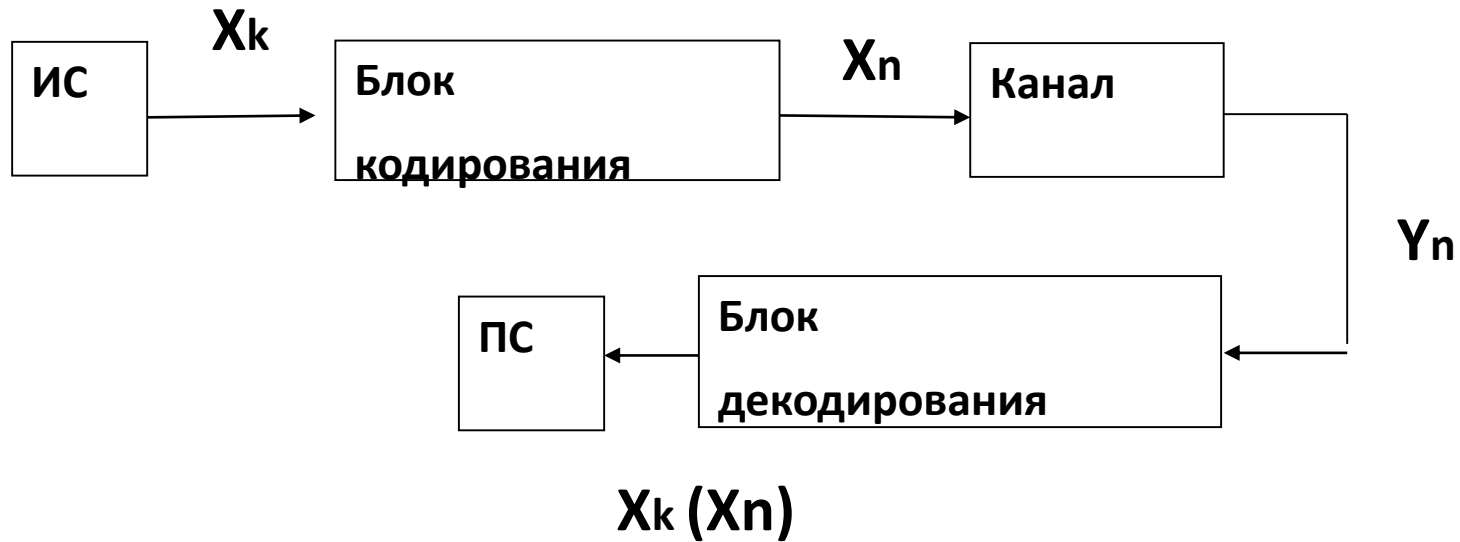


Рис.1.

где  $X_n = y_1, y_2, \dots, y_n$ ;  $Y_n = y_1, y_2, \dots, y_n$ .

# Теоретические основы линейных блочных кодов

7

- Для формирования проверочных символов ( $Xr$ ) используется **порождающая матрица**: совокупность базисных векторов будем записывать в виде матрицы  $G$  размерностью  $k \times n$  с единичной подматрицей ( $I$ ) в первых  $k$  строках и столбцах:

$$G = [ P | I ] \quad (2)$$

- Кодовые слова являются линейными комбинациями строк матрицы  $G$  (кроме слова, состоящего из нулевых символов).
  - Кодирование заключается в умножении вектора сообщения  $X_k$  длиной  $k$  на  $G$  по правилам матричного умножения (все операции выполняются по модулю два);
- при этом последние  $r$  символов ( $Xr$ ) образуются как линейные комбинации первых ( $X_k$ )



- Для всякой порождающей матрицы  $\mathbf{G}$  существует матрица  $\mathbf{H}$  размерности  $r \times n$ , удовлетворяющая равенству

$$\mathbf{G} * \mathbf{H}^T = \mathbf{0} \quad (3)$$

- Матрица  $\mathbf{H}$  называется *проверочной* и записывается как

$$\mathbf{H} = [\mathbf{P}^T | \mathbf{I}] \quad (4)$$

**Определение 6.** Результат умножения вектора сообщения ( $\mathbf{Y}_n$ ) на транспонированную проверочную матрицу ( $\mathbf{H}$ ) называется *синдромом* (вектором ошибки)  $\mathbf{S}$ :

$$\mathbf{S} = \mathbf{H}^T * \mathbf{Y}_n = \mathbf{H} * (\mathbf{Y}_n)^T \quad (5)$$

Если  $\mathbf{S}=\mathbf{0}$  (все  $r$  символов синдрома – нулевые) – в принятом сообщ ошибок нет; в противном случае -  $t_n > 1$ .

**Определение 7.** *Проверочным (избыточным) словом*  $\mathbf{X}_r$  кодового слова  $\mathbf{X}_n$  является вектор-строка, удовлетворяющая тождеству:

$$\mathbf{H}^T * \mathbf{X}_n = \mathbf{H} * (\mathbf{X}_n)^T = \mathbf{0} \quad (6)$$

**Определение 8.**  **$n$ -разрядным вектором ошибки** называем последовательность вида:

$$\mathbf{E}_n = \mathbf{X}_n + \mathbf{Y}_n \quad (7)$$

В (6) суммирование – по мод 2

**Определение 9.** **Корректирующие способности кода** определяются исключительно **минимальным кодовым расстоянием** ( $d_{\min}$ ) между двумя произвольными кодовыми словами, принадлежащими коду:

$$t_u = \begin{cases} \frac{d-1}{2}, & d - \text{нечетное}, \\ \frac{d-2}{2}, & d - \text{четное}. \end{cases} \quad (8)$$

$$t_o = d/2, \quad d - \text{четное}$$

$$t_o = (d-1)/2, \quad d - \text{нечетное}$$

(9)

## На стороне ИС

1. Построить проверочную матрицу  $H_{n,k}$  для заданного  $k$  ( $X_k$ )
2. Вычислить символы избыточного слова  $X_r$  (на основе (6) )
3. Сформировать кодовое слово  $X_n = x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_{k+r}$  и осуществить его передачу

## На стороне ПС

1. Получение сообщения ( $Y_n = y_1, y_2, \dots, y_k, y_{k+1}, \dots, y_{k+r}$  )
2. Вычисление синдрома (на основе (5) ), **используя ту же  $H_{n,k}$  :**

$$S = H^T * Y_n = H * (Y_n)^T = H^T * (X_n + E_n) = H^T * X_n + H^T * E_n = H^T * E_n \quad (10)$$

для этого вычисляем  $Y_r' = y_{ri}' : y_{ri}' = y_{k+i}' = \sum h_{ij} * y_j$  (11)

и далее:  $S = s_1, s_2, \dots, s_r$  , где  $s_i = y_{k+i} + y_{ri}'$  (12)

4. Анализ (декодирование синдрома) – определение местоположения ошибочного бита ( $E_n$  посредством  $S \rightarrow h_m$  )

5. Исправление ошибки:  $X_n = Y_n + E_n$  (13)

**Теорема 1.** Минимальное кодовое расстояние линейного кода равно минимальному весу ненулевых кодовых слов

- $R_n = r/k$  — относительная избыточность кода
- $R_c = k/n$  - скорость кода
- **Выбор кода** определяется вероятностью ошибки в канале,  $p$  (чем больше  $p$ , тем с большим  $d$ , т.е. с большим  $r$  следует выбирать код, однако это снижает  $R_c$ )

**Определение 10.** Пропускная способность ДСК с вероятностью ошибки  $p$  равна

$$C(p) = 1 + p \log_2 p + q \log_2 q \quad (14)$$

**Теорема 2 (Шеннона).** Для любого ДСК и любого  $\varepsilon > 0$  существует  $(n, k)$ - двоичный код со скоростью  $R_c$ , если  $R_c < C(p)$ ,  $n$  достаточно велико и  $p < \varepsilon$ .

**ОСНОВНАЯ ПРОБЛЕМА ТЕОРИИ КОДИРОВАНИЯ: НАЙТИ КОДЫ С БОЛЬШИМИ  $d$  И  $R_c$  (задача оптимизационная)**

- Простейший избыточный код;
- Основан на **контроле четности** (либо **нечетности**) единичных символов в сообщении.
- Количество избыточных символов  $r$  всегда равно **1** и не зависит от  $k$ .
- Значение этого символа будет **нулевым**, если сумма всех символов кодового слова по модулю 2 равна нулю – при контроле четности.

**Пример 4.** Пусть  $X_k = 1010$

Построить проверочную матрицу  $H$  кода и вычислить значение проверочного символа.

# Код Хемминга

13

- код характеризуется минимальным кодовым расстоянием  $d_{\min}=3$ ,  
используется расширенный контроль четности групп символов информационного слова,
- вес столбцов подматрицы  $P$  (см (4) ) должен быть больше либо равен 2,
- позволяет не только обнаруживать, но и исправлять **одиночную ошибку в кодовом слове** ( $t_o=1$ ,  $t_n=1$ ),
- параметры кода:  
$$n = 2^r - 1$$
$$k = 2^r - r - 1 \quad (14)$$
- Для упрощенного вычисления  $r$  можно воспользоваться следующим простым соотношением:

$$r = \log_2 k + 1 \quad (15)$$

**Пример 5.** Построить матрицу  $H$ , если  $X_n = 010110$