3. ОСНОВНЫЕ ЗАДАЧИ КОДИРОВАНИЯ ИНФОРМАЦИИ

Прежде, чем формулировать основные задачи кодирования информации, рассмотрим фазы *процесса преобразования информации (сообщения) в сигнал* (ППИС) на передающей стороне, задав его в виде макровектора

$$\text{ППИС}=\{\text{ИИ}, \text{ППИ}, \text{УД}, \text{КИ}, \text{УКШ}, \text{КК}, \text{СУ}, \text{М}, \text{КС}\}, \quad (3.1)$$

где ИИ – источник информации (сообщения), представляемой в произвольной форме;

ППИ – первичный преобразователь информации, формирующий *электрический аналог* информации, представленный в *непрерывной* или *дискретной* (цифровой) форме;

УД – устройство дискретизации электрического аналога информации (УД отсутствует в структуре, если ППИ формирует электрический аналог в дискретной форме), декомпозирующее информационный массив на *символы* или *кванты*;

КИ – «кодер источника», который осуществляет кодирования с учетом вероятностей $p(x_i)$ появления символов x_i на выходе ИИ (или УД в случае сплошной природы первоначальной информации);

УКШ – устройство криптографического шифрования, осуществляющего защиту информации от несанкционированного доступа к ней;

КК – «кодер канала», который осуществляет помехозащитное кодирование на основе вероятностных характеристик шумовой среды в канале связи;

СУ – скремблирующее устройство, обеспечивающее перемешивание элементарных сигналов кода с целью обеспечения условий синхронизации работы генераторов передающего и приемного полукомплектов аппаратуры, путем минимизации длин пачек «нулей» и «единиц» суммированием помехозащищенного двоичного кода с «псевдослучайной» двоичной последовательностью;

М – модулятор, осуществляющий согласование закодированного сигнала с предоставленным частотным каналом связи путем трансформации амплитудного частотного спектра кодового сигнала вдоль оси частот вправо на величину частоты модулирующего сигнала;

КС – предоставленный канал связи.

Элементы макровектора ППИС (3.1) размещены в порядке выполнения процедур преобразования информации (сообщения) в сигнал.

Процесс преобразования сигнала в информацию (ППСИ) на приемной стороне также может быть задан в виде макровектора

$$\text{ППСИ}=\{\text{КС}, \text{ДМ}, \text{ДСУ}, \text{ДКК}, \text{УКДШ}, \text{ДКИ}, \text{УДД}, \text{ДПИ}, \text{ПИ}\ \}, \quad (3.2)$$

где КС – предоставленный канал связи;

ДМ – демодулятор, осуществляющий восстановление амплитудного частотного спектра закодированного сигнала путем трансформации частотного спектра модулированного кодового сигнала вдоль оси частот влево на величину частоты модулирующего сигнала;

ДСУ – дескремблирующее устройство, осуществляющее процедуру, обратную перемешиванию элементарных сигналов кода, с целью обеспечения условий синхронизации работы генераторов передающего и приемного полукомплектов аппаратуры, путем минимизации длин пачек «нулей» и «единиц», суммированием по модулю два восстановленной скремблированной двоичной кодовой последовательности со скремблирующей последовательностью;

ДКК – декодер канала, осуществляющий обнаружение искажений принятого из канала связи передаваемой помехозащищенной двоичной кодовой комбинации и их исправление;

УКДШ – устройство, осуществляющее процедуру преобразования двоичной кодовой последовательности, обратную УКШ, то есть снятие защиты информации от несанкционированного доступа к ней;

ДКИ – декодер источника, представляет собой устройство, реализующее процедуру преобразования обратного выполняемому устройством КИ – кодером источника, направленному на восстановление символов сообщения или дискретов передаваемого информационного массива;

УДД – устройство, выполняющее процедуру преобразования, обратного процедуре УД – устройство дискретизации электрического аналога информации, то есть восстановление электрического аналога исходного информационного массива сплошной природы;

ДПИ – устройство, выполняющее процедуру преобразования, обратного процедуре ППИ – первичного преобразователя информации, формирующего *электрический аналог* информации (сообщения), восстанавливая информацию (сообщения) в исходной форме;

ПИ – получатель (приемник) информации (сообщения) в исходной форме.

Таким образом, процесс «передачи – приема» информации (ППИ), представленный макровектором

$$\text{ППИ} = \{ППИС, ППСИ\}, \qquad (3.3)$$

содержит ***пять фаз*** преобразования информации (сообщения) в форме «кодирование – декодирование»:

– первичное кодирование при первичном преобразовании информации (сообщения) в электрический сигнал чаще всего цифровой (дискретный), представляемый в виде числового (цифрового) кода;

– кодирование в соответствии с вероятностными свойствами источника дискретной информации;

– шифрование информации (сообщения) с целью обеспечения защиты информации от несанкционированного доступа к ней;

– помехозащитное кодирование, гарантирующее заданную достоверность передаваемой информации (информационную надежность), в условиях помеховой среды в канале связи, заданной вероятностями искажения бита информации в виде вероятностей трансформаций p_{01} и p_{10};

–скремблирование кодовых сигналов передаваемой информации (сообщения) с целью обеспечения условий синхронизации работы генераторов передающего и приемного полукомплектов аппаратуры

В разделе «Основные понятия и определения» настоящего пособия даны следующие определения.

Определение 0.37(О.037). *Кодированием* называется процесс присвоения элементам q_j исходного информационного массива Q их кодов $K\{q_j\} = row\{x_i; 1 \le i \le n\}$ с элементами $x_i \in GF(p)$, осуществляемого по правилам построения используемого кода. □

Определение 0.38(О.038). *Декодированием* называется процесс восстановления элементов q_j исходного информационного массива Q по их кодам $K\{q_j\} = row\{x_i; 1 \le i \le n\}$. □

Таким образом, при построении числового кода $K\{q_j\} = row\{x_i; 1 \le i \le n\}$ как кодового аналога элемента q_j исходного информационного массива Q должны быть решены следующие задачи:

– формирование правила построения числового кода $K\{q_j\} = row\{x_i; 1 \le i \le n\}$;

– аргументированный выбор основания p кода $K\{q_j\} = row\{x_i; 1 \le i \le n\}$;

– определение размерности n кода $K\{q_j\} = row\{x_i; 1 \le i \le n\}$.

Проблемы ***первой фазы*** преобразования информации (сообщения) в форме «кодирование – декодирование» при первичном преобразовании информации (сообщения) в электрический сигнал обеспечиваются решением перечисленных задач в следующих формах.

Построение числового кода в классе *равномерных кодов* на все сочетания, при этом возможно удовлетворение дополнительному требованию построения кодов методом *соседнего кодирования*, при котором кодовое расстояние между соседними кодами равняется единице так, что выполняется равенство

$$d\{K(q_{j-1}), K(q_j)\} = d\{K(q_j), K(q_{j+1})\} = 1. \qquad (3.4)$$

Аргументированный выбор основания p кода $K\{q_j\}=row\{x_i;1\le i\le n\}$ и определение его размерности n осуществляется на основе минимизации функционал размещения, задаваемого в форме

$$J(p,n)=p\cdot n. \qquad (3.5)$$

Тогда основание p кода $K\{q_j\}=row\{x_i;1\le i\le n\}$ при фиксированном информационном объеме V_Q информационного массива Q определится выражением

$$p=\arg\min_{p}\left\{J(p,n)=p\cdot n\ \&\ p^n\ge V_Q\right\}. \qquad (3.6)$$

При выборе основания p кода $K\{q_j\}=row\{x_i;1\le i\le n\}$ следует учитывать простоту «кодовой» арифметики, сложность реализации сигнальных уровней, задающих число p, а также простоту коррекции искажений при передаче и хранении кодов.

В качестве примера для информационного объема V_Q=10000 рассчитаны значения функционалов размещения и оценены размерности n кодов $K\{q_j\}=row\{x_i;1\le i\le n\}$, полученные результаты приведены в таблице 3.1.

Таблица 3.1

Информационный объем V_Q=10000 информационного массива Q									
Основание кода p	1	2	3	4	5	10	100	1000	10000
Размерность кода n	10000	14	9	7	6	4	2	2	1
Функционал размещения $J(p,n)=p\cdot n$	10000	28	27	28	30	40	200	2000	10000

Из таблицы 3.1 видно, что оптимальным по критерию минимума значения функционала размещения является основание кода $p=3$. Незначительно по этому показателю ему проигрывают основания кода $p=2$ и $p=4$, но основание кода $p=2$ выигрывает у $p=3$ и $p=4$ по простоте «кодовой» арифметики и по сложности реализации сигнальных уровней, задающих число p, а также простоте коррекции искажений кода.

По совокупности факторов в прикладных задачах кодирования оптимальным основанием кода является $p=2$.

Проблемы ***второй фазы*** преобразования информации (сообщения) в форме «кодирование – декодирование» в соответствии с

вероятностными свойствами источника дискретной информации обеспечиваются решением перечисленных задач в следующих формах.

Код строится по правилу «чем больше вероятность появления символа сообщения на выходе ИДИ, тем короче его код», с использованием основания кода $p = 2$ в классе неравномерных кодов, то есть с переменной размерностью n.

Проблемы ***третьей фазы*** преобразования информации (сообщения) в форме «кодирование – декодирование» гарантирующее заданную достоверность передаваемой информации (информационную надежность), в условиях помеховой среды в канале связи, заданной вероятностями искажения бита информации в виде вероятностей трансформаций p_{01} и p_{10} обеспечиваются решением перечисленных задач в следующих формах.

Помехозащищенный (n,k) – код (ПЗК), в котором $n,k,n-k=m-$ соответственно полное число разрядов кода, число разрядов его информационной части (кода сообщения), число проверочных разрядов кода, наличие которых гарантирует помехозащиту передаваемого сообщения, формируется в силу правила: при фиксированном значении k число m проверочных разрядов ПЗК тем больше, чем больше вероятность $p = \max\{p_{01}, p_{10}\}$ искажения одного бита кода, с основанием два и фиксированной размерности n для всех (n,k) – ПЗК с фиксированным числом k. При этом систематические ПЗК обладают свойством делимости их модулярных многочленов (ММ) $y(x)$ без остатка на неприводимый образующий ММ ПЗК $g(x)$ степени m. Декодирование принятого из канала связи ПЗК состоит в проверке сохранения указанного свойства делимости так, что ненулевые остатки $E(x)$от деления используются как синдромы (опознаватели) ошибок (искажений) при передаче.

Проблемы ***четвертой фазы*** преобразования информации (сообщения) в форме шифрование информации (сообщения) с целью обеспечения защиты информации от несанкционированного доступа к ней обеспечиваются решением перечисленных задач в следующих формах. Шифрование осуществляется над простым двоичным полем Галуа $GF(2)$ по схеме близкой схеме построения ПЗК с той разницей, что для шифрования используются образующие шифрованное представление передаваемой информации (сообщении) ММ очень высокой степени m.

Проблемы ***пятой фазы*** преобразования информации (сообщения), состоящей в скремблировании кодовых сигналов передаваемой информации (сообщения), опираются на определения

скремблирования и *дескремблирования*, данные в разделе «Основные понятия и определения» настоящего пособия.

Определение 0.39(О.039). *Скремблированием* называется процесс, осуществляемый на передающей стороне, преобразования передаваемой кодовой посылки для сокращения длительности пачек нулевых и единичных символов путем суммирования по модулю два передаваемой посылки со *псевдослучайной* двоичной кодовой последовательностью для целей максимизации числа передних фронтов переходов от нуля к единице, что повышает синхронизирующие работу генераторов передающей и приемной сторон свойства кода. □

Определение 040(О.040). *Дескремблированием* называется процесс, осуществляемый на приемной стороне, преобразования принятой *скремблированной* кодовой посылки для восстановления передаваемой путем суммирования по модулю два скремблированной посылки с той же, что и на передающей стороне, *псевдослучайной* двоичной кодовой последовательностью. □

Примеры и задачи

3.1. Информационный массив Q после его дискретизации характеризуется информационным объемом $V_Q = 500$. Вычислить функционалы $J = p \cdot n$ размещения при кодировании элементов этого массива равномерными кодами с основаниями $p = 10; p = 5; p = 4; p = 3; p = 2.$

3.2. Трехмерный информационный массива Q характеризуется информационным объемом $V_Q = N_X \cdot N_Y \cdot N_Z$ с компонентами $N_X = 25; N_Y = 50; N_Z = 100$. Вычислить функционалы $J = p \cdot n$ размещения при кодировании элементов этого массива равномерными кодами с основаниями $p = 25; p = 10; p = 5; p = 3; p = 2.$

3.3. Четырехмерный информационный массива Q характеризуется информационным объемом $V_Q = N_X \cdot N_Y \cdot N_Z \cdot N_T$ с компонентами $N_X = 20; N_Y = 40; N_Z = 80; N_T = 160..$ Вычислить функционалы $J = p \cdot n$ размещения при кодировании элементов этого массива равномерными кодами с основаниями $p = 100; p = 50; p = 25; p = 10; p = 2.$

3.4. ИДИ генерирует символы, образующие алфавит $X = \{x_1, x_2, x_3, x_4 : p(x_1) = 0.5; p(x_2) = 0.25; p(x_3) = 0.15; p(x_4) = 0.1\}.$ Символы кодируются двоичными неравномерными кодами с учетом вероятности их появления на выходе ИДИ. Какой из символов

алфавита $X=\{x_1,x_2,x_3,x_4\}$ будет иметь самый короткий код и какой – самый длинный?

3.5. ИДИ генерирует периодическое сообщение: XYWXZXYUXXYXZXWYXVXYXZXYXUXWYXZXYX, состоящее из символов X,Y,Z,W,U,V. Символы кодируются двоичными неравномерными кодами, код какого из символов будет самым коротким, а какого – самым длинным?

3.6. ИДИ генерирует символы, образующие алфавит $X=\{x_j : p(x_j)=1/n; n=8; j=\overline{0,n-1};\}$. Закодируйте символы x_j равномерными двоичными кодами так, чтобы двоичный код символа был бы двоичным представлением его индекса i.

3.7. ИДИ генерирует символы, образующие алфавит $X=\{x_j : p(x_j)=1/n; n=8; j=\overline{0,n-1};\}$. Закодируйте символы x_j двоичными равномерными *соседними* кодами.

3.8. Массив Q передаваемых команд характеризуется информационным объемом $V_Q=120$. Каждая из команд кодируется для передачи по двоичному каналу связи с помехами равномерным помехозащищенным (n,k) – кодом, в котором $k-$число информационных разрядов кода, $(n=k+m)$ – полное число разрядов помехозащищенного кода (ПЗК), $m-$число проверочных разрядов кода. Чему равно число $k-$информационных разрядов кода?

3.9. На основе информационной части с числом разрядов k примера 3.8 строятся два ПЗК (n_1,k) и (n_2,k) для передачи по одному и тому же двоичному КС, первый для реализации помехозащиты в режиме *обнаружения*, второй – в режиме *исправления.* Какое число разрядов больше n_1 или n_2?

3.10. На основе информационной части с числом разрядов k примера 3.8 строятся два ПЗК (n_1,k) и (n_2,k) для передачи по двум различным двоичным КС, характеризующимися вероятностями искажения бита кода $p=p_{01}=p_{10}=10^{-3}$ и $p=p_{01}=p_{10}=10^{-4}$ соответственно. Помехозащита обоих кодов реализуется в режиме *исправления*. Какое число разрядов больше n_1 или n_2?

3.11. Кодовая помехозащищенная последовательность ПЗК $(15,11)$ $y=000111100000110$ на передающей стороне скремблируются псевдослучайной последовательностью периода $T=15$ $\varsigma=100110101111000$. Как будет выглядеть скремблированная последовательность $y_c=(y+\varsigma)\bmod 2$?

3.12. Скремблированная двоичная последовательность примера 3.11 дескремблируется на приемной стороне той же скремблирущей псевдослучайной последовательностью периода $T=15$

$\varsigma = 100110101111000$. Как будет выглядеть дескремблированная последовательность $y_{дс} = (y_c + \varsigma) \bmod 2$?

Решение вариантов задач

Задача 3.7. ИДИ генерирует символы, образующие алфавит $X = \{x_j : p(x_j) = 1/n; n = 8; j = \overline{0, n-1};\}$. Закодируйте символы x_j двоичными равномерными *соседними* кодами.

Решение

Решение задачи опирается на определение *соседнего* кода (3.4). Применительно к условиям решаемой задачи (3.4) принимает вид $d\{K(x_{j-1}), K(x_j)\} = d\{K(x_j), K(x_{j+1})\} = 1$. Наиболее распространенными схемами формирования соседних кодов являются схема Грея и схема Джонсона. Результаты кодирования символов x_j по этим схемам приведены в таблице 3.2. Там же приведены обыкновенные двоичные коды на все сочетания, построенные в соответствие с условием задачи 3.6.

Таблица 3.2

№ п/п	Кодируемые переменные $x_j (j = \overline{0,7})$	Кодовые комбинации $K\{x_j\}$		
		Обыкновенные двоичные, представляющие двоичный код числа j	Соседние двоичные по схеме Грея	Соседние двоичные по схеме Джонсона
1.	x_0	000	000	0000
2.	x_1	001	001	0001
3.	x_2	010	011	0011
4.	x_3	011	010	0111
5.	x_4	100	110	1111
6.	x_5	101	111	1110
7.	x_6	110	101	1100
8.	x_7	111	100	1000

Нетрудно видеть, что соседние коды, построенные по схеме Джонсона, обладают избыточностью, чего не наблюдается в кодах Грея. ■