Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 27.01.2021 01:05:36
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра «Вычислительная техника»

УТВЕРЖДАЮ
Проректор по учебной работе
по учебной работе
«Юго-Западный государственный университет»
_____ О.Г. Локтионова
«27» 01 2021 г.

КОДИРОВАНИЕ ИНФОРМАЦИИ ВО ВНЕШНИХ ЗАПОМИНАЮЩИХ УСТРОЙСТВАХ ЭВМ

Методические указания по выполнению лабораторной работы по дисциплине « ЭВМ и периферийных устройств»

для студентов специальности 090301

Курск 2021

УДК 621.(076.1)

Составитель: Д.Б. Борзов

Рецензент

Кандидат технических наук, доцент *В.С. Панишев*

Кодирование информации во внешних запоминающих устройствах ЭВМ: методические указания по выполнению лабораторной работы / Юго-Зап. гос. ун-т; сост.: Д.Б. Борзов. - Курск, 2017. - 21 с.: ил. 11, табл. 2. - Библиогр.: с. 21.

Содержаться теоретическое описание принципов кодирования информации во внешних запоминающих устройствах ЭВМ, приводятся описание способов кодирования различными методами.

Методические рекомендации соответствуют требованиям программы, утвержденной учебно-методическим объединением по специальностям «Информатика и вычислительная техника» (ИВТ) дневной и заочной форм обучения.

Текст печатается в авторской редакции

Подписано в печать . Формат 60x84 1/16.
Усл.печ. л. 1,0. Уч.-изд. л. 0,9. Тираж 100 экз. Заказ. 45
Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

1. Цель лабораторной работы

Целью лабораторной работы является изучение современных методов кодирования и декодирования информации во внешних запоминающих устройствах (ВЗУ), закрепление полученных знаний по принципам и способам кодировании информации.

2. Физические основы магнитной записи

2.1. Принципы записи информации

Наиболее распространенными способами регистрации информации в ВЗУ является магнитная и оптическая запись на поверхности подвижного носителя. Рассмотрим физические основы магнитной и оптической записи [1]. Материал запоминающего слоя магнитного носителя можно представить множеством хаотически расположенных магнитных доменов (рис. 1).

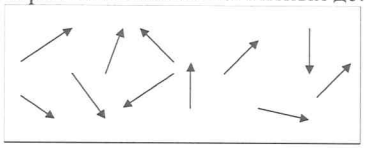


Рис. 1. Расположение доменов в ненамагниченном магнитом поле

Магнитим слой носителя выполняется из магнитотвердого материала со значениями коэрцитивной силы Hc=(12–180) кА/м. Это делается для того, чтобы запоминающий слой случайно не перемагничивался магнитными полями двигателей и других устройств, имеющихся в компьютере. Под действием внешнего магнитного поля с напряженностью H≥Hc домены ориентируются вдоль силовых линий этого поля. В результате возникает свое "внутреннее" поле ферромагнетика с напряженностью J, равной сумме элементарных магнитных моментов намагниченного материалала. В ферромагнетиках напряженность внутреннего поля J может значительно превышать напряженность внешнего поля H. Суммарное магнитное поле характеризуется магнитной индукцией B.

$$B = \mu_0 \left(J + H \right) \qquad (1)$$

Домены сохраняют ориентацию в течение длительного времени и после окончания действия внешнего магнитного поля. Величина магнитного поля в этом случае характеризуется значением остаточной индукции Br (Рис. 2).

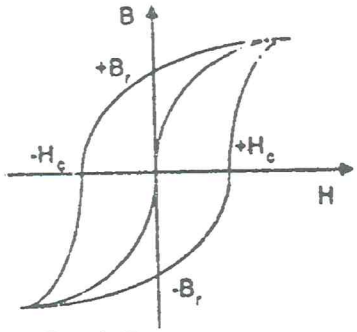


Рис. 2. Петля гистерезиса

Внешнее магнитное поле, ориентирующее домены в определенном направлении, создастся при записи информации магнитной головкой. Если поле приводит к ориентации доменов в плоскости носителя (рис. 3), то магнитную запись называют горизонтальной, а если перпендикулярно плоскости носителя - вертикальной. Как видно из рис. 3 ориентация доменов может быть во взаимопротивоположных направлениях, что соответствует значениям +Br и –Br на петле гистерезиса. Эта способность сохранять намагниченность после снятия внешнего магнитного поля и используется для запоминания информации.

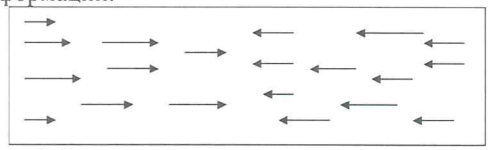


Рис. 3. Ориентация доменов при горизонтальной магнитной записи

2.2. Горизонтальная магнитная запись

В магнитных накопителях для записи и чтения информации используются универсальные магнитные головки [1,2]. Головка состоит из магнитопровода 2 (рис. 4) с воздушным зазором 3. На магнитопроводе располагается обмотка 1. При протекании тока Iw по обмотке создается магнитный поток, который в основном замыкается по магнитопроводу через воздушный зазор. Если в области воздушного зазора находится запоминающий слой магнитного носителя 5, то большая часть магнитного потока замыкается через него, поскольку магнитное сопротивление воздушного зазора велико. Под действием магнитного потока домены ориентируются вдоль силовых линий магнитного поля, т. е. в горизонтальном направлении, отсюда и название записи. Если изменить направление тока записи, то ориентация доменов также изменится на противоположную. Поскольку носитель движется, его участки оказываются намагниченными до значения +Br или -Br (рис. 2). Участки носителя, где происходит смена направления намагниченности, называют переходами. Количество переходов,

5.5. Краткое описание теории кодирования и декодирования кода Хэмминга.
5.6. Оценки информационных параметров полученных кодов (см. п. 3.1).

6. Контрольные вопросы
6.1. Обосновать необходимость многоуровневой системы кодирования в ВЗУ.
6.2. Что представляет собой энтропия источника?
6.3. Объяснить порядок построения кода Хаффмана с основанием равным двум.
6.4. Построить (15,11) – код Хэмминга.
6.5. Декодировать слово 110010000101111 (15,11) – кода Хэмминга.
6.6. Привести примеры кодирования и декодирования основных типов канальных кодов.
6.7. Как оценить суммарную эффективность системы многоуровневого кодирования в ВЗУ?
6.8. На каком свойстве основан принцип записи информации на магнитных носителях?
6.9. Пояснить принцип горизонтальной магнитной записи.
6.10. Пояснить способ вертикальной магнитной записи.

Библиографический список

1. Периферийные устройства [текст]: Учеб. Пособие/ К.Л. Пестерев, И.С. Захаров; Курск. гос. тех. ун–т. Курск, 1999. 205 с.
2. Танэнбаум, Э. Архитектура компьютера [текст]/ Э. Таненбаум. 4-е изд. СПб.: Питер, 2003. 704 с.
3. Гук, М. Аппаратные интерфейсы ПК [текст]/ М. Гук. СПб.: Питер, 2003. 528 с.
4. Столлингс, В.. Современные компьютерные сети [текст]/ В. Столлингс. 2-е изд. Питер, 2003. 783 с.
5. Марков А.А. Введение в теорию кодирования [текст]/ А.А. Марков. М.: Наука. Гл. ред. физ. –мат. Литературы, 1982. 192 с.
6. Рыжков, В.А. Сергеев Н.П., Раков Б.М. Внешние ЗУ на магнитных носителях [текст]/ В.А. Рыжков, Н.П. Сергеев, Б.М. Раков. М.: Энергия, 1978. 224 с.
7. Финк, Л.М. Теория передачи дискретных сообщений [текст] / Л.М. Финк. М.: Сов. радио, 1970. 728 с.

C	1100	11110
D	1101	01101
E	1110	01110
F	1111	01111

Длительность тактового интервала при RLL - кодировании равна 2,6 мкс. Двоичный код полубайта заменяется RLL - кодом в соответствии с таблицей 2. Полученная последовательность нулей и единиц кодируется методом без возврата к нулю. Этот метод кодирования очень похож на метод Габбора.

Для записи на магнитные диски чаще используется разработанный фирмой IBM так называемый RLL 2,7 - метод. Цифры в обозначении кода соответствуют максимальной и минимальной длине последовательности логических нулей, содержащихся в кодируемых данных между двумя единицами. Следовательно, кодом RLL 2,7 могут быть закодированы данные, содержащие между единицами минимум два и максимум 7 нулей.

Помимо метода RLL 2,7 используются методы RLL 1,7, RLL 1,8, RLL2.8.

4. Порядок выполнения работы

4.1. Изучить данные методические указания

4.2. Построить код Хаффмана для символьной строки, заданной преподавателем.

4.2.1. Пользуясь изложенной в п. 3.1.2 методикой составить таблицу объединения символов (табл. 1).

4.2.2. Построить кодовое дерево.

4.2.3. Сформировать кодовый словарь.

4.2.4. По формулам (8), (9) и (10) рассчитать соответственно величины H, $\bar{n}$, $Rсж$.

4.3. Построить код Хэмминга для исходной символьной строки.

4.3.1. Использовать в качестве исходной последовательности строку, полученную для кода Хаффмана.

4.3.2. Разбить исходную последовательность на блоки по 4 разряда.

4.3.3. Получить проверочные разряды, на основе формулы (14).

4.4. Записать итоговую кодовую последовательность.

4.5. По формулам (17) и (18) рассчитать значения информационных параметров использованного канального кода.

4.6. Подготовить отчет.

5. Содержание отчета

5.1. Тема лабораторной работы.

5.2. Цель работы.

5.3. Исходные данные для кодирования.

5.4. Решение задачи построения кода Хаффмана с необходимыми пояснениями (см. п. 3.1.2).

размещенных на единице длины носителя, называют физической плотностью записи.

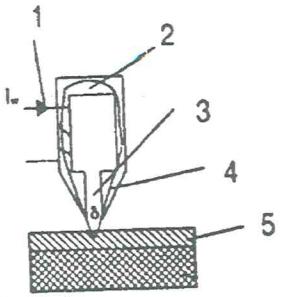


Рис. 4. Горизонтальная магнитная запись

Для считывания информации используется или явление электромагнитной индукции, или магниторезистивный эффект. В первом случае обмотка головки подключается к схеме усилителя и, следовательно, напряжение к ней не прикладывается. Головка остается неподвижной, а магнитный носитель совершает или прямолинейное (в НМЛ), или вращательное движение (в НМД). При прохождении под головкой перехода потокосцепление ее обмотки изменяется и в ней индуктируется импульс ЭДС. Величина ЭДС пропорциональна скорости изменения магнитного потока

$$e = -N d\Phi / dt, \qquad (2)$$

где N – число витков обмотки головки.

Если используется магниторезистивный эффект, то головка содержит слой из полупроводникового материала, сопротивление которого изменяется под воздействием магнитного поля. Этот слой называют магниторезистором. При прохождении отпечатков мимо магниторезистора, его сопротивление изменяется и, следовательно, изменяется величина тока. Поскольку отпечатки очень быстро проходят мимо магниторезистора, изменение тока имеет импульсную форму.

2.3. Вертикальная магнитная запись

Для получения такой записи магнитный поток должен быть направлен перпендикулярно поверхности носителя, т.е. он должен пересекать запоминающий слой [1,2]. Конструкция головки записи–считывания и магнитного носителя показаны на рис 5. Головка записи-считывания имеет два полюса - записывающий 1 и замыкающий 2. Записывающий полюс имеет площадь поперечного сечения меньше, чем замыкающий полюс. Носитель информации двухслойный, верхний слой 4 служит для хранения информации и выполняется из материала с большой коэрцитивной силой, нижний слой 5 выполняется из магнитомягкого материала и служит для замыкания магнитного потока. Если по обмотке 3 головки пропустить электрический ток, то возбуждается магнитный поток, который стремится замкнуться по пути с наименьшим магнитным сопротивлением.

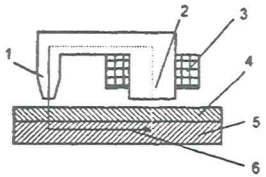


Рис. 5. Вертикальная магнитная запись

Силовые линии магнитного потока проходят через записывающий полюс 1, запоминающий слой 4, железоникелевую основу носителя 5 и вновь попадают в замыкающий полюс 2. В воздушном зазоре под записывающим полюсом создается сильное магнитное поле с напряженностью

$$H_{зап} = \frac{\Phi}{\mu_0 S_{зап}}, \qquad (3)$$

где $S_{зап}$ – площадь поперечного сечения записывающего полюса в воздушном зазоре между головкой и носителем.

Величина этого магнитного поля превышает коэрцитивную силу магнитотвердого материала запоминающего слоя, и поэтому он перемагничивается. Поскольку магнитное сопротивление материала запоминающего слоя велико, магнитный поток в основном замыкается по нижнему слою, имеющему значительно меньшее магнитное сопротивление, и поэтому он пересекает верхний слой в перпендикулярном направлении и перемагничивает его.

В воздушном зазоре под замыкающим полюсом напряженность магнитного поля определяется выражением

$$H_0 = \frac{\Phi}{\mu_0 S_0}, \qquad (4)$$

где S_0- площадь поперечного сечения замыкающего полюса.

Поскольку площадь поперечного сечения замыкающего полюса S_0 больше, чем площадь поперечного сечения записывающего полюса $S_{зап}$ то, как следует из выражений (1) и (2), напряженность магнитного поля под замыкающим полюсом получается меньше. Поперечное сечение S_0 подбирают таким, чтобы выполнялось неравенство H_0«Нс, т.е. чтобы напряженность магнитного поля под замыкающим полюсом была меньше коэрцитивной силы. Поэтому запоминающий слой под замыкающим полюсом не перемагничивает. Силовые линии магнитного поля пересекают запоминающий слой перпендикулярно поверхности носителя, в результате чего под записывающим полюсом, происходит вертикальная ориентация доменов вдоль силовых линий магнитного поля, поэтому запись называется вертикальной.

3. Проблема кодирования в ВЗУ

Алфавит входного языка ЭВМ составляет совокупность элементарных символов, с помощью которых на языке записываются программы решаемых

помощью битов, располагаемых посередине тактовых интервалов, в таком тактовом интервале бит синхронизации отсутствует. За счет этого возрастает емкость. Скорость передачи данных увеличивается до 250 Кбит/с. Число бит данных, приходящихся на единицу длины дорожки, удваивается по сравнению с FM.

Для получения сигналов, соответствующих логическим единицам, электронная схема должна выделить импульсы ЭДС, находящиеся посередине тактов. Временные диаграммы, поясняющие данный метод, приведены на рис. 11.

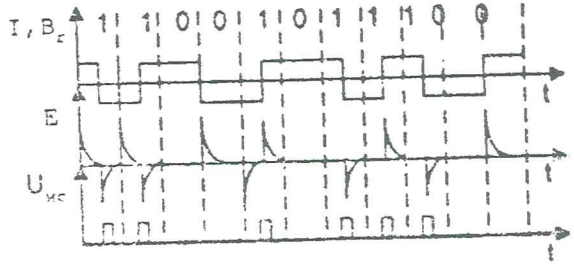


Рис. 11. Временная диаграмма для метода модифицированной частной модуляции

3.3.5. RLL-метод

В этом методе полностью исключаются биты синхронизации, за счет чего информационная емкость диска возрастает. При кодировании RLL-методом каждый байт, поступающих данных, разделяется на два полубайта, а затем полубайты кодируются специальным 5- разрядным кодом. Из 32 комбинаций 5- разрядного кода для кодирования методом RLL используются 16, в которых число рядом стоящих нулей не превышает двух. Перевод шестнадцатеричных чисел в кодовые группы RLL приведен в таблице 2.

Таблица 2

Шестнадцатеричное число	Двоичный полубайт	Кодовая группа RLL
0	00000	11001
1	0001	11011
2	0010	10010
3	0011	10011
4	0100	11101
5	0101	10101
6	0110	10110
7	0111	10111
8	1000	11010
9	1001	01001
A	1010	01010
B	1011	01011

Из временных диаграмм на рис. 9 видно, что направление тока в обмотке записывающей головки изменяется в середине каждого такта, если в следующем такте записывается такой же знак. Если знаки разные, то ток изменяет направление в момент времени, соответствующий границе такта. Аналогично изменяется и индукция носителя B(t). На графике E(t) показано, что, когда индукция изменяется от +Br к -Br , в обмотке считывающей головки индуцируется отрицательный импульс ЭДС, а когда от -Br к +Br - положительный импульс ЭДС. Отрицательные импульсы ЭДС, находящиеся на границах витков, соответствуют "0" информации, а положительные - единицам. Положительные и отрицательные импульсы ЭДС, появляющиеся в середине тактов, являются ложными и должны подавляться логическими схемами.

3.3.3. Метод частотной модуляции

В данном методе кодирования предварительно старая запись. стирается и запоминающий слой намагничивается до определенного значения Br. При этом методе в начале битового промежутка (такта) записывается бит синхронизации, а в середине битового промежутка – бит данных. В случае диска диаметром 133 мм и скорости вращения диска 300 об/мин каждый битовый элемент имеет длительность 8 мкс, следовательно, бит данных записывается через 4 мкс после бита синхронизации. Таким образом, в битовых элементах, где записываются логические единицы, частота импульсов в два раза превышает частоту импульсов в битовых элементах, где записываются логические нули, т.е. единицы модулируются частотой, отсюда и название метода. Временные диаграммы, поясняющие этот метод, приведены на рис. 10.

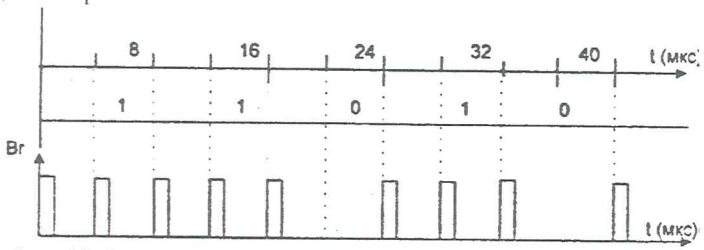


Рис. 10. Временная диаграмма для метода частной модуляции

3.3.4. Метод модифицированной частной модуляции

В данном методе изменение магнитного потока кодируется в зависимости от значения предыдущего и последующего бита. При записи единиц изменение магнитного потока происходит посередине тактового интервала. При записи нулей изменение магнитного потока происходит в начале тактового интервала, если предыдущий записываемый знак не единица. В противном случае изменение тока в магнитной головке блокируется. Метод модифицированной частотной модуляции MFM позволяет вдвое уменьшить длительность тактовых интервалов. Логические единицы записываются с

задач [3]. Стандартные алфавиты ЭВМ содержат, как правило, графические символы (A, B, C, …), арабские цифры (0, 1, 2, …), ограничители (+, X, (, ...) и управляющие символы ("возврат каретки", "забой", "стоп", …). Для того, чтобы иметь возможность представлять в ЭВМ информацию, выраженную символами алфавита, необходимо сами символы закодировать определенными наборами двоичных цифр.

В ЭВМ в качестве внутренних приняты 8-разрядные коды. При таком кодировании каждому символу полного алфавита ставится в соответствие 8-разрядная машинная единица - байт. Используются два способа 8-разрядного кодирования [2]:

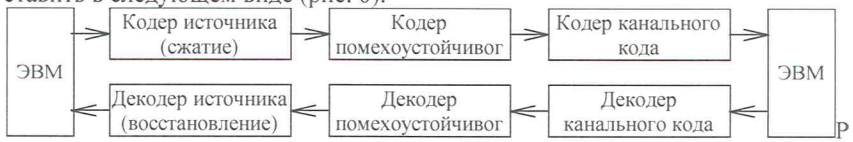
8-разрядкый код для обмена информацией (КОИ-8);

двоичный код для обмена информацией (ДКОИ).

С помощью специальной команды многоразрядные десятичные числа, вводимые в ЭВМ, могут быть упакованы в формате, в котором на каждую цифру и знак выделяется полубайт. Однако при выдаче из машины результатов производится распаковка. Таким образом, в информации, выводимой из ЭВМ в ВЗУ, каждый символ алфавита представляется 8-разрядным двоичным набором.

При записи информации на магнитный или оптический носители указанные 8-разрядные двоичные наборы требуют дополнительных преобразований по следующим причинам. Во-первых, частота встречаемости различных символов полного алфавита (а следовательно и их 8-раэрядннх представлений) неодинакова. Поэтому, используя специальные неравномерные коды, разработанные в теории информации, можно закодировать более часто встречающиеся символы более короткими кодовыми словами, что обеспечит сокращение (сжатие) объема исходной информации. Во-вторых, вследствие дефектов носителя и различных помех в тракте записи–воспроизведения ценность информации может быть утрачена из-за искажения данных ошибками. Одним из основных методов защиты от ошибок является в настоящее время применение помехоустойчивого кодирования. При этом за счет введения в информационную последовательность дополнительных избыточных разрядов появляется возможность не только обнаруживать, но и исправлять некоторые конфигурации ошибок. В-третьих, информация, записываемая на носитель, должна подвергаться канальному кодированию для обеспечения требуемой величины отношения сигнал/шум и надежного выделения синхросигналов в процессе воспроизведения.

Таким образом, иерархию уровней кодирования в ВЗУ можно представить в следующем виде (рис. 6):



ис. 6. Иерархия уровней кодирования в ВЗУ

Функции кодирования-декодирования (кодеки) реализуются как правило, средствами контроллера ВЗУ.

3.1. Алгоритмы сжатия-восстановления информации

Информация, записываемая или считываемая в ВЗУ, имеет свою количественную меру [5-7]. Эту меру в теории информации определяют как

$$J = \log N = n \cdot \log m, \qquad (5)$$

где N - общее число сообщений, которое может быть составлено из n элементов, каждый из которых имеет m возможных состояний.

Наиболее удобен выбор основания логарифма, равного двум. При этом за единицу количества информации принято считать информацию сообщения, содержащего один элемент (n=1), который может принимать два равновероятных состояния (m=2).

Тогда $J = 1 \cdot \log_2 2 = 1$ бит

Если все сообщения N равновероятны и вероятность каждого сообщения P=1/N, то количество информации можно выразить через вероятности поступления сообщений:

$$J = \log_2 N = \log_2 \frac{1}{P} = -\log_2 P \qquad (6)$$

Если сообщения неравновероятны и независимы друг от друга, используют понятие средней информации

$$\overline{J} = -\sum_{i=1}^{m} p_i \log_2 P_i, \qquad (7)$$

где p_i - вероятность i-го сообщения.

Среднее количество информации на один элемент сообщения называется энтропией:

$$H = \sum_{i=1}^{m} p_i \log_2 p_i. \qquad (8)$$

Величина H представляет собой меру беспорядочности источника сообщений и характеризует среднюю степень неопределенности состояния этого источника. Когда все m различных состояний источника равновероятны, энтропия максимальна. Если сообщения неравновероятны, среднее количество информации в одном сообщении будет меньшим. Таким образам, задача кодирования источника заключается в построении такого кода, в котором среднее количество информации, приходящееся на один элемент кода, стремится к максимальной энтропии.

Оценить эффективность кодирования можно путем сопоставления рассчитанных значений H и среднего числа разрядов $\overline{n}$ в кодовой последовательности, приходящихся на один символ:

$$\overline{n} = \sum_{i=1}^{m} p_i n_i, \qquad (9)$$

Так как при записи нулей на носителе переходы не остаются, то при чтении нельзя определить, какое их количество было записано. Поэтому для распознавания информации при считывании используют синхронизирующие импульсы, задающие границы каждого такта записи, график Uc(t). При считывании признаком 1 является совпадение информационного импульса e(t) с синхронизирующим импульсом, а признаком 0 – наличие только синхронизирующего импульса. Для получения информационного сигнала используют конъюнктор, на один вход которого подаются синхроимпульсы, а на другой – сформированный и усиленный сигнал воспроизведения информации. На выходе конъюнктора получают информационный сигнал, соответствующий записанной последовательности нулей и единиц. Синхроимпульсы могут быть записаны на специальной дорожке того же носителя.

3.3.2. Метол фазовой модуляции

Процесс записи разбивается на такты постоянной длительности (рис. 9).

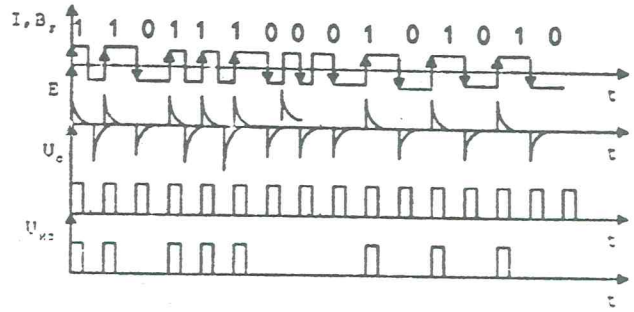


Рис. 9. Временная диаграмма для метода фазовой модуляции

Здесь на рис. 9 – I и Br – графики изменения тока и индукции; E - график ЭДС; Uc- напряжение синхроимпульсов, вырабатываемых электронной схемой; Uис - напряжение информационного сигнала.

В каждом такте производится запись одной "1" или одного "0". На границе каждого такта происходит смена направления тока в записывающей головке и, следовательно, смена магнитных состояний носителя. Полярность тока изменяется в одном направлении при записи "0" (например, от положительной к отрицательной) и в противоположном направлении при записи "1". Происходит как бы изменение фазы тока. Если должен быть записан тот же знак, что и в предыдущем такте, то в середине такта изменяется направление тока в записывающей головке. Если должен быть записан другой знак, то изменение направления тока в середине такта не происходит. При считывании 1 и 0 распознаются по полярности импульса ЭДС в головке считывания в первом полутакте, т.е. такой способ обладает свойством самосинхронизации.

где d - минимально допустимое число, нулей между символами $K_j^i = 1$.

Рассмотрим некоторые канальные коды с использованием формализованного представления процедур кодирования и декодирования.

3.3.1. Метод записи без возврата к нулю с переключением по единицам (БВН-1)

Временные диаграммы для метода БВН-1 приведена на Временные диаграммы для метода БВН-1 приведена на Рис. 8.

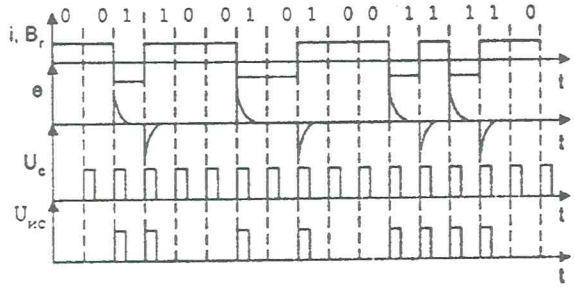


Рис. 8. Временная диаграмма для кодирования методом без возврата к нулю с переключением по единицам

Здесь на рис. 8, i – ток в головке записи, Br – намагниченность носителя, e – ЭДС в обмотке головки считывания, Uc – синхроимпульсы, Uис – напряжение информационного сигнала.

При записи единицы ток в обмотке записи меняет направление, магнитный поток в головке записи также изменяет направление и магнитный слой носителя перемагничивается из состояния насыщения одного знака в состояние насыщения другого знака. При записи нуля направление тока в обмотке магнитной головки записи не изменяется. Диаграмма изменения тока в зависимости от записываемой последовательности "1" и "0" показана на графике i(t). Если пренебречь гистерезисом, то временная диаграмма изменения индукции в носителе B(t) будет совпадать с временной диаграммой тока. Из этой диаграммы видно, что при записи каждой единицы в магнитном слое носителя остаются отпечатки - переходы от одного направления намагниченности к другому, а при записи нулей направление намагниченности не меняется. При считывании информации в обмотке считывающей головки индуктируются импульсы ЭДС, когда под головкой $e = -N\dfrac{d\Phi}{dt}$ движется участок носителя с переходом. Когда проходят участки носителя с постоянным значением индукции, то величина магнитного потока также остается постоянной, и в соответствии с приведенным уравнением ЭДС не возникает. Импульсы ЭДС показаны на графике e(t).

где n_i – количество кодовых разрядов в закодированном i-м сообщении.

Более простой оценкой эффективности является расчет коэффициента сжатия информации:

$$R_{сж} = \frac{Q_u}{Q_k}, \qquad (10)$$

где Q_u и Q_k - соответственно суммарное количество бит в исходных данных и суммарное количество бит в их кодовом представлении.

3.1.1. Код Хаффмана

Код, предложенный Хаффманом [4,7], является одним из наиболее эффективных неравномерных кодов с основанием r, позволяющий для заданного источника в высокой степени сократить избыточность сообщения.

Заданность источника предполагает, что до начала кодирования известны (или рассчитаны) вероятности (частоты) всех символов алфавита сообщений. Алгоритм кодирования заключается в следующем.

Все символы алфавита в количестве m выписываются в столбец в порядке убывания вероятностей. Для однозначности декодирования в случае равной вероятности нескольких символов можно принять соглашение о том, что выше в столбце будет располагаться символ, встретившийся ранее при подсчете частот.

Если число m-1 не кратно r-1, то в алфавит добавляются дополнительные символы, которым приписываются вероятности, равные нулю, так чтобы для объема полученного алфавита m' соблюдалось условие кратности m'-1 числу r-1.

Затем r последних символов полученного алфавита объединяются в один вспомогательный символ, которому приписывается суммарная вероятность. Алфавит с дополнительным символом (но без объединенных) вновь выписывается в столбец в порядке убывания вероятностей и формируется вспомогательный символ путем объединения r последних символов нового алфавита. Указанная процедура продолжается до тех пор, пока не останется алфавит, состоящий из r символов. Этим символам приписываются одноразрядные кодовые комбинации 0, 1, …, r-1.

Если в числе оставшихся r символов есть такие, которые принадлежат исходному алфавиту (т.е. не полученные в результате объединения других символов), то они оказываются закодированными одноразрядными кодовыми комбинациями. Для тех символов, которые получены в результате объединения r символов исходного алфавита, в кодовые комбинации дописываются вторые разряды 0, 1, …, r-1 так, что эти символы оказываются закодированными двухразрядными комбинациями. Если среди этих двухразрядных комбинаций имеются такие, которые соответствуют символам, также полученным в результате объединения, то к этим комбинациям приписываются третьи разряды и т.д., пока все символы исходного алфавита не окажутся закодированными.

Полученный в результате кодирования, словарь, где символам, положенным в порядке убывания вероятностей, поставлены в соответствие

кодовые комбинации, используется при кодировании и декодировании. В процессе кодирования вместо очередного символа передается его словарь Хаффмана, а в процессе декодирования путем просмотра словаря, начиная с кодовой комбинации наиболее вероятностного символа, однозначно восстанавливаются символы исходного алфавита.

3.1.2. Пример построения кода Хаффмана

Предположим, что а ВЗУ должна быть записана текстовая строка "ЛАБОРАТОРНАЯ РАБОТА" состоящая из 19 символов. Считая, что на каждый символ выделяется один байт, получим требуемый объем памяти $Q_u = 152$ *бит*.

Если учесть, что среди 19 указанных символов есть повторяющиеся, и закодировать строку кодом Хаффмена, затраты памяти можно уменьшить.

Рассчитаем частоты P_i встречаемости символов в рассматриваемом сообщении путем деления числа появлений t_i символа на общую длину сообщения, равную 19. Выпишем символы в порядке убывания частот и поставим им в соответствие кодовые комбинации кода Хаффмана согласно п.3.1.1. Принцип объединения символов иллюстрируется в табл.1.

Чтобы составить кодовую комбинацию, соответствующую данному символу, необходимо проследить путь перехода символа по строкам и столбцам табл.1. Для наглядности строится кодовое дерево. Из корня дерева направляются две ветви, причем ветвям с большей вероятностью (в нашем случае: 0,579) присваивается разряд 0, с меньшей (0,422) - 1. Такое последовательное ветвление продолжаем до тех пор, пока не дойдем до вероятности каждого символа исходного алфавита. Кодовое дерево для рассматриваемого примера представлено на рис.7.

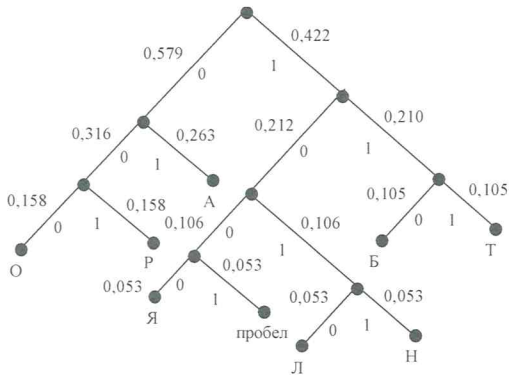


Рис. 7. Пример кодового дерева

применяют специальные коды (канальные кода), обеспечивающие необходимые значения параметров сигнала записи и воспроизведения.

В процессе кодирования группа из *m* двоичных разрядов преобразуется в кодовую группу из *n* двоичных разрядов ($m<n$). Будем обозначать разряды исходной группы символами $A_1^i, A_2^i, ..., A_m^i$, где *i* - номер группы (*i= ... -2, -1, 0, +1,+2,...)*.

Каждому символу A_j^i на входе кодера канального кода ставится в соответствие импульс (при $A_j^i = 1$) или пауза (при $A_j^i = 0$) в такте T_U. За один такт T_U передается один бит информации, а длительность такта связана со скоростью передачи информации.

Разряды кодовой группы обозначим символами $K_1^i, K_2^i, ..., K_m^i$, где $K_j^i = 0$ или $K_j^i = 1$, причем символы K_j^i в отличие от A_j^i характеризуются длительностью такта T_k.

Справедливо равенство:

$$T_K = \frac{mT_U}{n} \qquad (16)$$

При преобразовании значений символов K_j^i в импульсы тока записи будем полагать, что $K_j^i = 1$ определяет формирование перепада, (фронта или среза импульса) тока записи в начале такта T_k, а $K_j^i = 0$ сохраняет в такте T_k установленный в предыдущем такте уровень тока записи. Длительности импульсов тока записи будут, таким образом, кратны длительности такта Тк, а число их возможных значений равно числу частот (периодов записи) в канальном коде.

К канальным кодам, широко используемым в ВЗУ, откосятся двухчастотные (Манчестер, Манчестер I, ФМ, ЧМ, M2FM, 2/3 БВНМ), трехчастотные (*MFM ,4/5БВНМ 3/4БВНМ*) и многочастотные (*БВН, БВНМ*).

Среди информационных параметров канальных кодов обычно выделяют:

1. Удельную информационную плотность кода (информативность кодового символа) J_k, определяющую количество информации, передаваемой одним кодовым символом K_j^i

$$J_k = \frac{m}{n}, \text{ бит} \qquad (17)$$

2) информативность перепада тока записи (в магнитных ВЗУ - перепада намагниченности):

$$J_П = \frac{(d+1)m}{n}, \text{ бит/перепад} \qquad (18)$$

проверочных разрядов от информационных, матрицу $\overline{H}$ целесообразно преобразовать к виду

$$,\overline{H} = \overline{A} \mid \overline{I}_{n-k} \qquad (13)$$

где $\overline{I}_{n-k}$ – квадратная единичная подматрица размера $r \times r$.

Преобразование матрицы осуществляется путем сложения по модулю два различных комбинаций r её строк. Подставив (13) в (11) и учитывая, что все операции сложения в (11) выполняются по модулю два, получим выражения для определения проверочных символов:

$$x_5 = x_2 + x_3 + x_4;\ x_6 = x_1 + x_3 + x_4;\ x_7 = x_1 + x_2 + x_4 \qquad (14)$$

Рассмотрим пример кодирования. Пусть из кодера Хаффмана в кодер кода Хэмминга поступает последовательность 110010101... Разбиваем данную последовательность на информационные блоки по 4 разряда: 1100 1010 I... К каждому информационному блоку добавляем 3 проверочных разряда, вычисленные согласно (14): 1100110 1010101 I...., и передаем их на вход канального кодера.

Исправление возникшей ошибки не вызывает особых затруднений. Если вместо кодового слова $\overline{X}$ в декодер кода Хэмминга поступило сообщение $\overline{U}$, необходима вычислить

$$\overline{H} \cdot \overline{u}^T = \overline{S}, \qquad (15)$$

где вектор $\overline{S}$ называется синдромом и его нулевое значение указывает либо на то, что ошибки отсутствуют, либо на то, что они не обнаружены, а ненулевое значение $\overline{S}$ совпадает с значением одного из столбцов матрицы $\overline{H}$, указывая, что в разряде, соответствующем номеру этого столбца, произошла ошибка.

Например, пусть передавалось кодовое слово $\overline{X} = (1100110)$, а в декодер кода Хэмминга поступило слово $\overline{u} = (1101110)$ с ошибкой в разряде x4. Тогда, вычисляя (15) с матрицей $\overline{H}$ вида (12) получим $\overline{S} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, что соответствует четвертому столбцу матрицы, и четвертый разряд слова $\overline{u}$ должен быть инвертирован. Отбрасывая после декодирования проверочные разряди, передаем в декодер кода Хаффмана восстановленную последовательность.

3.3. Алгоритмы канального кодирования.

Код, получаемый на выходе кодера помехоустойчивого кода, обычно неоптимален для записи, так как выбирается из соображений, не учитывающих особенности записи в накопитель [5,6]. Поэтому для записи в ВЗУ

Таблица 1.

i	Сим-волы	ti	Pi	Объединение символов						
				1	2	3	4	5	6	7
1	А	5	0,263	0,263	0,263	0,263	0,263	0,316	0,316	0,579
2	О	3	0,158	0,158	0,158	0,158	0,158	0,263	0,263	0,422
3	Р	3	0,158	0,158	0,158	0,158	0,158	0,212	0,212	0,422
4	Б	2	0,105	0,106	0,106	0,210	0,210	0,210		
5	Т	2	0,105	0,105	0,106	0,106	0,158			
6	Л	1	0,053	0,053	0,105	0,106				
7	Н	1	0,053	0,053	0,105					
8	Я	1	0,053							
9	пробел	1	0,053							

Теперь, двигаяся по кодовому дереву сверху вниз, можно записать для каждого исходного алфавита соответствующую ему кодовую комбинацию: А-01, О-000, Р – 001, Б-110, Т-111, Л-1010, Н-1011, Я-1000, пробел-1001.

Заданная текстовая строка, закодированная кодом Хаффмана, будет иметь вид:

10100111000000101111000001101101100010010010111000011101 и занимать объем памяти Qк=56 бит.

Таким образом, рассчитанный по (10) коэффициент сжатия Rсж≈2,71. Энтропия исходного сообщения согласно (8):

$$H = -\sum_{i=1}^{9} p_i \log_2 p_i = -0.263\log_2 0.263 - 2\cdot 0.158\log_2 0.158 -$$

$$-2\cdot 0.105\log_2 0.105 - 4\cdot 0.053\log_2 0.053 \approx 2.93$$

а среднее число разрядов $\overline{n}$, приходящихся на один символ сообщения, по формуле (9):

$$\overline{n} = \sum_{i=1}^{9} p_i n_i = 0.263\cdot 2 + 2(0.158\cdot 3) + 2(0.105\cdot 3) + 4(0.053\cdot 4) \approx 2.95$$

Так как значение $\overline{n}$ почти достигает величины энтропии, кодирование текстовой строки в нашем примере кодом Хаффмана устранило практически всю естественную избыточность сообщения.

3.2. Помехоустойчивое кодирование информации кодом Хэмминга

Защита информации от ошибок в ВЗУ базируется на достижениях теории помехоустойчивого кодирования, отличительными чертами которой являются намеренное введение избыточности и дискретный характер сигналов [7]. Дискретность позволяет описывать наборы сигналов символами некоторого абстрактного алфавита. Намеренное введение избыточности обеспечивает возможность исправления ошибок при использовании сложных комбинаторных конструкций. Это позволяет сформулировать задачи помехоустойчивого кодирования как чисто математические.

Различают два различных класса стратегий кодирования - блочное и неблочное кодирование. Блочное кодирование используется в ВЗУ наиболее часто. Сущность его состоит в том, что последовательность символов, поступающее на вход кодера, разбивается на блоки (векторы) по k символов в каждом: $\overline{a_1} = (a_1, a_2, ..., a_k)$, $\overline{a_2} = (a_{k+1}, a_{k+2}, ..., a_{2k})$,.... Кодер преобразует каждый входной k–блок $\overline{a_i}$ в выходной n–блок $\overline{x_i} = \overline{x}(\overline{a_i}) = (x_1(\overline{a_i}), x_2(\overline{a_i}), ..., x_n(\overline{a_i}))$ таким образом, чтобы различным входным блокам соответствовали различные выходные. Заметим, что во входной последовательности символов все символы a_i имеют одинаковую длину.

Таким образом, кодирование ставит в соответствие каждому входному "укрупненному" символу $\overline{a}$ некоторый выходной укрупненный символ $\overline{x}(\overline{a})$. Совокупность всех таких различных символов $\overline{x}(\overline{a})$ называют блочным кодом длины n мощности $M = q^k$, где q - число различных

символов a_i. Скорость кода в q-ричных единицах измерения определяется величиной $A = \frac{k}{n}$.

Помехоустойчивые коды могут иметь систематический вид, когда *n-k* избыточных (проверочных) сигналов располагаются подряд в младших или старших позициях кодового слова. В противном случае коды называются несистематическими.

Из теории помехоустойчивого кодирования известно, что двоичные (q=2) блочные линейные систематические *(n, k)*–коды с конструктивным кодовым расстоянием *d* способны исправить в кодовом слове длины *n* бит *t* ошибочных бит, если $d \geq 2t+1$, где *k* - количество информационных разрядов в кодовом слове.

Конструктивное кодовое расстояние *d* равно минимальному числу разрядов, на которое отличаются друг от друга два любых кодовых слова из всего множества кодовых слов. Величина *r=n-k* определяет количество проверочных разрядов кодового слова.

Значения проверочных разрядов (n,*k*)–кода выбираются так, чтобы

$$\overline{H}\cdot\overline{x}^T = 0, \qquad (11)$$

где $\overline{X} = (x_1, x_2, ..., x_n)$ – вектор кодового слова, $\overline{H}$ – матрица размера $r \times n$ с линейно-независимыми строками, называемая проверочной матрицей, *T* - символ транспонирования.

Линейный код определяется как множество решений (11). Процедура кодирования линейных кедов наиболее проста, так как для её выполнения достаточно хранить в памяти только *r* векторов матрицы $\overline{H}$. Для нелинейного кода мощностью $M = q^k$ необходимо помнить все q^k кодовых последовательностей.

В классе двоичных линейных кодов одни из самых простых процедур кодирования и декодирования имеют систематические коде Хэмминга [2].

Рассмотрим их особенности на примере (7,4) - кода Хемминга с *d=3* и $n = 2^r - 1$. Проверочная матрица $\overline{H}$ этого кода имеет размерность 3×7. Столбцами $\overline{H}$ являются всевозможные ненулевые двоичные наборы длины *r* , которые располагают так, чтобы *i*-й слева столбец h_i представлял собой двоичное разложение числа *i* (старшие разряда сверху):

$$\overline{H} = \begin{pmatrix} 0\,0\,0\,1\,1\,1\,1 \\ 0\,1\,1\,0\,0\,1\,1 \\ 1\,0\,1\,0\,1\,0\,1 \end{pmatrix} \qquad (12)$$

Кодовое олово $\overline{x}$ имеет вид $\overline{X} = (x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7)$, где разряды x5, *x6*, *x7* – проверочные. Чтобы получить зависимость значений