

## Глава 4. Физическое и логическое кодирование данных

Каждый вид компьютеров имеет свой внутренний вид кодирования для представления данных – символьной и текстовой информации. Наиболее часто используются коды ASCII (American Standard Code for Information Interchange, американский стандартный код для обмена информацией) и EBCDIC (Extended Binary Coded Decimal Interchange Code, расширенный двоично-десятичный код обмена информацией).

ASCII представляет собой 8-битную кодировку для представления десятичных цифр, латинского и национального алфавитов, знаков препинания и управляющих символов. Первую половину кодовой таблицы (0 — 127) занимают символы US-ASCII, которые включают 95 печатаемых и 33 управляющих символа (разработана ANSI – Американским институтом национальных стандартов). Вторая половина таблицы (128 — 255) содержит национальные шрифты (кириллица) и символы псевдографики. Этот код используется в персональных компьютерах и в несовместимых с IBM больших машинах. На больших компьютерах (мейнфреймах) используется 8-битовый код EBCDIC, разработанный компанией IBM. При передаче данных от одного компьютера к другому может потребоваться перекодировка символов, которая осуществляется системным МО передающего или принимающего компьютеров. Эти действия являются функциями уровня представления модели OSI. Далее рассмотрим наиболее часто применяемые методы кодирования на физическом уровне.

Большинство компьютеров для представления «0» и «1» оперирует стандартными уровнями сигналов (логическими уровнями), которые определяются видом микросхем. TTL-логика представляет 0,5В как «0», и 5В как «1». ECL и CMOS-логики представляют -1,75В как «0», и -0,9В как «1». Для передачи данных, например, в оптоволоконных системах, в трансивере (приемопередатчике) устанавливается специальный чип, обрабатывающий любую логику и выдающий управляющий сигнал источнику света с конвертацией 0,5В и 5В TTL в 0 мА и 50 мА соответственно (включи свет, выключи свет).

В большинстве компьютерных сетей цифровые данные передаются при помощи цифрового сигнала, т.е. последовательностью импульсов. Для передачи данных может использоваться более двух уровней сигнала, при этом единичный импульс сигнала может представлять не один бит, а группу бит. Возможна обратная ситуация, когда для передачи одного бита может использоваться два импульса сигнала.

При цифровой передаче используют потенциальные и импульсные коды. В потенциальных кодах для представления логических единиц и нулей используется только значение сигнала в течение битового интервала, а фронты сигнала, формирующие законченные импульсы, во внимание не принимаются. Импульсные коды представляют логический ноль и логическую единицу перепадом потенциала определенного направления. В значение импульсного кода включается весь импульс вместе с его фронтами.

Сигнал в виде импульсной последовательности имеет бесконечный спектр. Основная энергия сигнала сосредоточена в диапазоне частот от нуля до частоты  $f=1/t_0$  (первый лепесток энергетического спектра сигнала), где  $t_0$  – бодовый интервал, то есть длительность единичного импульса линейного сигнала.

Теоретически в соответствии с пределом Найквиста максимально допустимая скорость изменения значений дискретного сигнала ( $B=1/t_0$ , скорость передачи в Бодах) при передаче последовательности прямоугольных импульсов по каналу связи, эквивалентом которого является идеальный ФНЧ с прямоугольной АЧХ и линейной ФЧХ и с частотой среза  $f_{гр}$ , равна  $B_{max}=2f_{гр}$ . Указанное ограничение связано с наличием переходных процессов на выходе ФНЧ, при этом время нарастания/спада фронта сигнала определяется как

$$t_n = \frac{1}{2f_p}$$

При максимально допустимой скорости передачи сигнала  $t_0=t_n$ . Если интервал  $t_0 < t_n$ , происходит недопустимое искажение формы сигнала на выходе канала и, как следствие, ошибки при приеме. Таким образом, скорость передачи информации  $N$  (Бит/с) зависит от скорости передачи сигнала  $B$  (Бод) и выбранного метода кодирования сигнала на физическом уровне.

Используемый метод физического цифрового кодирования должен достигать несколько целей:

1. Обеспечивать наименьшую ширину спектра сигнала при заданной скорости передачи информации  $N$  (Бит/с). Минимизировать величину постоянной составляющей в спектре линейного сигнала.
2. Обеспечивать приемнику возможность тактовой синхронизации (clocking). Так называемые самосинхронизирующиеся коды позволяют приемнику выделять из принимаемого цифрового потока колебание тактовой частоты и затем формировать из него тактовые импульсы при любой статистике битового потока на входе передатчика.
3. Обладать способностью распознавать ошибки
4. Обладать низкой стоимостью реализации.

Рассмотрим методы физического цифрового кодирования сигналов.

### Потенциальный код без возвращения к нулю NRZ (Non Return to Zero)

Сигнал в линии имеет два уровня: нулю соответствует нижний уровень, единице - верхний. Переходы происходят на границе битового интервала. При передаче последовательности единиц сигнал не возвращается к нулю в течение битового интервала.

Рассмотрим частные случаи передачи данных кодом NRZ (рис.4.1): чередующаяся последовательность нулей и единиц, последовательность нулей и последовательность единиц. Определим частоту основной гармоники спектра сигнала в каждом из этих случаев.

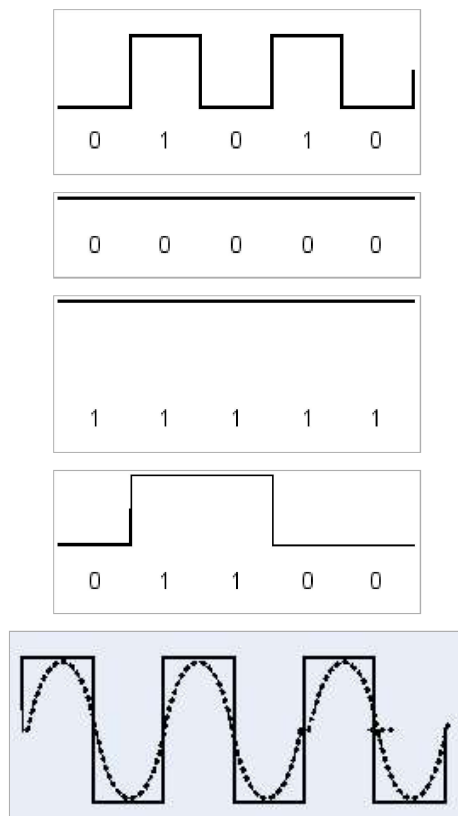


Рис.4.1. Кодирование по методу NRZ

При чередовании единиц и нулей и скорости передачи  $N$  (Бит/с) период основной гармоники в спектре сигнала равен  $T=2t_0=2/N$  (с). Частота основной гармоники  $f_0$  равна  $f_0=N/2$  (Гц).

При передаче только единиц, или только нулей сигнал в линии представляет собой постоянный ток.

Спектр реального сигнала постоянно меняется в зависимости от того, какова структура данных, передаваемых по линии связи. При передаче длинных последовательностей нулей или единиц, спектр сигнала сдвигается в сторону низких частот. Линейный сигнал NRZ обычно содержит постоянную составляющую и не всегда обеспечивает приемнику возможность синхронизироваться с поступающим сигналом. С другой стороны код NRZ прост в реализации, обладает хорошей помехоустойчивостью (из-за двух резко отличающихся

уровней сигнала). Основная энергия сигнала в коде NRZ сосредоточена на частотах от 0 до  $N/2$  (Гц).

В чистом виде код NRZ в сетях не используется. Тем не менее, используются его различные модификации, в которых с успехом устраняют как плохую самосинхронизацию, так и наличие постоянной составляющей.

Потенциальный код с инверсией при единице NRZI (Non Return to Zero with ones Inverted, NRZI)

При этом методе кодирования передаче нуля соответствует уровень сигнала, который был установлен в предыдущем битовом интервале (уровень сигнала не меняется), а при передаче единицы – уровень изменяется на противоположный.

Код используется при передаче по оптоволоконным кабелям, где приемник устойчиво распознает два состояния сигнала - свет и темнота.

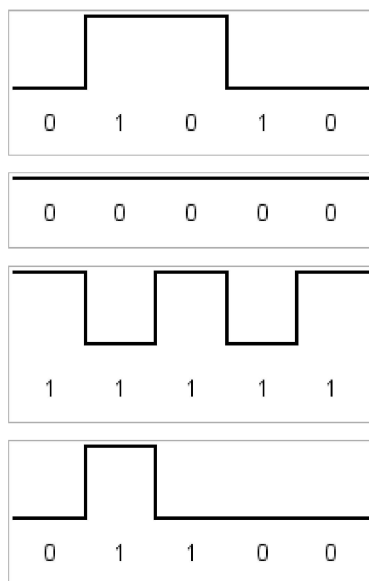


Рис.4.2. Кодирование по методу NRZI

Определим частоты основных гармоник линейного сигнала для частных случаев битовых последовательностей.

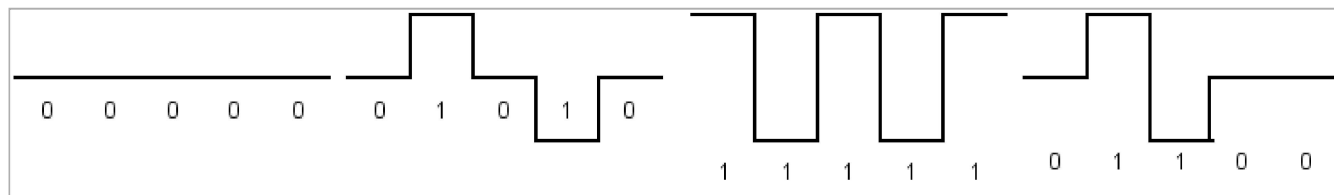
Для последовательности чередующихся единиц и нулей период сигнала равен  $T=4t_0$  (с), основная частота сигнала  $f_0=N/4$  (Гц), при последовательности единиц –  $f_0=N/2$  (Гц), при передаче последовательности нулей  $f_0=0$  - постоянный ток в линии (или отсутствие света).

Код NRZI использует только два уровня сигнала и поэтому обладает хорошей помехоустойчивостью. Максимальную энергию имеют спектральные составляющие сигнала около частоты  $N/4$  (Гц). Следует отметить, что код NRZI стал основным при разработке улучшенных методов кодирования для систем передачи данных.

### Метод биполярного кодирования с альтернативной инверсией (Bipolar Alternate Mark Inversion, AMI, квазитроичный код)

В этом методе используются три значения сигнала – «-1», «0» и «+1». Для различения трех уровней необходимо лучшее соотношение сигнал/шум на входе приемника. Дополнительный уровень требует увеличение мощности передатчика примерно на 3 дБ для обеспечения той же достоверности приема бит на линии. Это общий недостаток кодов с несколькими состояниями сигнала по сравнению с двухуровневыми кодами.

Для кодирования логического нуля используется нулевой потенциал, логическая единица кодируется попеременно либо положительным потенциалом, либо отрицательным.



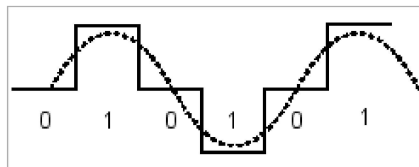


Рис.4.3. Квазитричное кодирование (AMI)

При передаче любой битовой последовательности сигнал в линии не содержит постоянную составляющую. При передаче единиц основная гармоника сигнала находится на частоте  $f_0 = N/2$  (Гц). В случае чередующегося набора единиц и нулей основная гармоника находится на частоте  $f_0 = N/4$  (Гц), что в два раза меньше чем у кода NRZ. Но остается проблема синхронизации при передаче последовательности нулей.

В целом, для различных комбинаций бит использование кода AMI приводит к более узкому спектру сигнала, чем для кода NRZ. Код AMI предоставляет также некоторые возможности по распознаванию ошибочных сигналов. Так, нарушение строго чередования полярности сигналов говорит о ложном импульсе или исчезновении корректного импульса.

### Манчестерский код (Manchester)

Манчестерский код относится к самосинхронизирующимся кодам и имеет два уровня, что обеспечивает хорошую помехозащищенность. Каждый битовый интервал делится на две части. Информация кодируется перепадом уровня, происходящим в середине каждого интервала.

Единица кодируется перепадом от высокого уровня сигнала к низкому, а ноль - обратным перепадом. В начале каждого битового интервала может происходить служебный перепад сигнала (при передаче несколько единиц или нулей подряд).

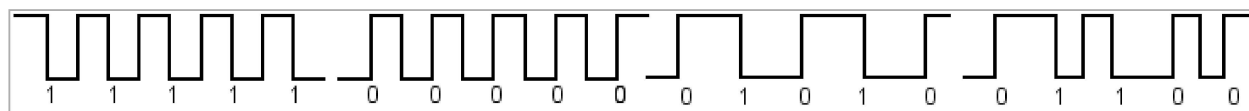


Рис.4.4. Манчестерское кодирование

При передаче любой битовой последовательности сигнал не содержит постоянную составляющую. Длительность единичного импульса линейного сигнала  $t_0$  равна половине битового интервала, то есть  $B = 2N$ . Частота основной гармоники сигнала зависит от характера битовой последовательности и находится в диапазоне  $f_0 = N/2 - N$  (Гц).

Манчестерский код используется в сетях Ethernet со скоростью передачи 10 Мбит/с (спецификация 10Base-T).

В настоящее время разработчики пришли к выводу, что во многих случаях рациональнее применять потенциальное кодирование, ликвидируя его недостатки с помощью так называемого *логического кодирования* (см. ниже в этом разделе).

### Потенциальный код 2B1Q

Это потенциальный код с четырьмя уровнями сигнала для кодирования данных. Название отражает суть кодирования – каждые два бита (2B) передаются за один такт сигналом определенного уровня (1Q). Линейный сигнал имеет четыре состояния.

Дибиту «00» соответствует потенциал -2,5 В (-3), «01» - потенциал -0,833 В (-1), «11» - потенциал +0,833 В (+1), «10» - потенциал +2,5 В (+3). Скорость передачи сигнала В при таком кодировании в 2 раза меньше скорости передачи информации N. На рис.4.6 изображен сигнал, соответствующий последовательности бит: 01 01 10 00.

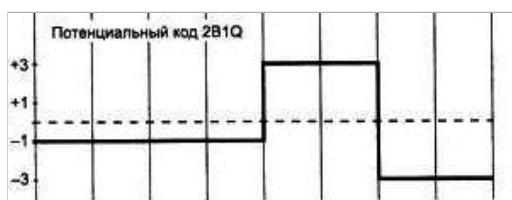


Рис.4.6. Сигнал в коде 2B1Q

Основная частота сигнала не превышает  $f_0 = N/4$  Гц. Однако для реализации этого метода кодирования мощность передатчика должна быть выше, чтобы четыре значения потенциала

четко различались приемником на фоне помех.

### Код MLT3 (Multi Level Transmission - 3).

Используются три уровня линейного сигнала: «-1», «0», «+1». Логической единице соответствует обязательный переход с одного уровня сигнала на другой. При передаче логического нуля изменение уровня линейного сигнала не происходит.

При передаче последовательности единиц период изменения уровня сигнала включает четыре бита. В этом случае  $f_0 = N/4$  (Гц). Это максимальная основная частота сигнала в коде MLT-3. В случае чередующейся последовательности нулей и единиц основная гармоника сигнала находится на частоте  $f_0 = N/8$  (Гц).

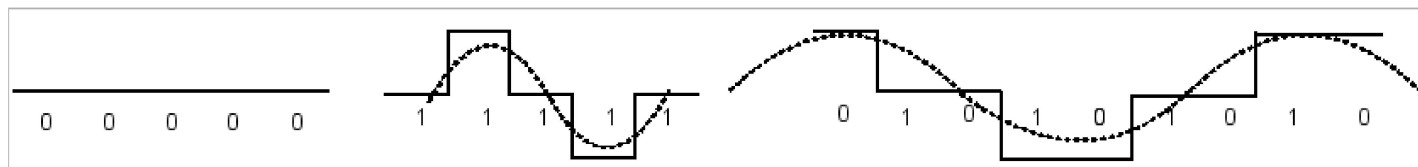


Рис.4.7. Сигнал в коде MLT-3

Логическое кодирование выполняется передатчиком до физического кодирования, рассмотренного выше, обычно средствами физического уровня. На этапе логического кодирования борются с недостатками методов физического цифрового кодирования - отсутствие синхронизации, наличие постоянной составляющей. Таким образом, сначала с помощью средств логического кодирования формируются исправленные последовательности данных, которые потом с помощью методов физического кодирования передаются по линиям связи.

Логическое кодирование подразумевает замену бит исходной информационной последовательности новой последовательностью бит, несущей ту же информацию, но обладающей, кроме этого, дополнительными свойствами.

Различают два метода логического кодирования:

- избыточные коды;
- скремблирование.

*Избыточные коды* основаны на разбиении исходной последовательности бит на группы и замене каждой исходной группы в соответствии с заданной таблицей кодовым словом, которое содержит большее количество бит.

*Логический код 4B/5B* заменяет исходные группы (слова) длиной 4 бита словами длиной 5 бит. В результате, общее количество возможных битовых комбинаций  $2^5=32$  больше, чем для исходных групп  $2^4=16$ . В кодовую таблицу включают 16 кодовых слов, которые не содержат более двух нулей подряд, и используют их для передачи данных. Код гарантирует, что при любом сочетании кодовых слов на линии не могут встретиться более трех нулей подряд.

Остальные комбинации кода используются для передачи служебных сигналов (синхронизация передачи, начало блока данных, конец блока данных, управление передачей). Неиспользуемые кодовые слова могут быть задействованы приемником для обнаружения ошибок в потоке данных. Цена за полученные достоинства при таком способе кодирования данных - снижение скорости передачи полезной информации на 25%.

Имеются также коды и с тремя состояниями сигнала, например, в коде 8B/6T для кодирования 8 бит исходной информации используются кодовые слова из 6 элементов, каждый из которых может принимать одно из трех значений. Избыточность кода 8B/6T выше, чем кода 4B/5B, так как на  $2^8=256$  исходных комбинаций приходится  $3^6=729$  результирующих комбинаций.

В *коде 8B/10B* каждые 8 бит исходной последовательности заменяются десятью битами кодового слова. При этом на 256 исходных комбинаций приходится 1024 результирующих комбинаций. При замене в соответствии с кодовой таблицей соблюдаются следующие правила:

- ни одна результирующая комбинация не должна иметь более 4-х одинаковых бит подряд;
- ни одна результирующая комбинация не должна содержать более 6 нулей или 6 единиц;

Все рассмотренные избыточные коды применяются в сетях Ethernet, которые нашли самое широкое распространение. Так, код 4B/5B используется в стандартах 100Base-TX/FX, а код 8B/6T — в стандарте 100Base-T4, который в настоящее время практически уже не используется. Код 8B/10B используется в стандарте 1000Base-X, код 64/66 в стандарте 10 GbE (когда в качестве среды передачи данных используется оптоволокно).

Осуществляют логическое кодирование сетевые адаптеры. Поскольку, использование таблицы перекодировки является очень простой операцией, метод логического кодирования избыточными кодами не усложняет функциональные требования к этому оборудованию.

Для обеспечения заданной пропускной способности линии передатчик, использующий избыточный код, должен работать с повышенной скоростью (тактовой частотой). Так, для обеспечения скорости передачи информации 100 Мбит/с с использованием кодирования 4B/5B + NRZI передатчик должен работать на скорости 125 МБод. При этом спектр линейного сигнала расширяется. Тем не менее, спектр сигнала избыточного потенциального кода уже спектра сигнала в манчестерском коде, что оправдывает дополнительный этап логического кодирования, а также работу приемника и передатчика на повышенной скорости.

**Скремблирование** представляет собой "перемешивание" исходной последовательности данных таким образом, чтобы вероятность появления единиц и нулей на линии становилась близкой 0,5. Устройства (или программные модули), выполняющие такую операцию, называются скремблерами (scramble - свалка, беспорядочная сборка).

Скремблер в передатчике выполняет преобразование структуры исходного цифрового потока. Дескремблер в приемнике восстанавливает исходную последовательность бит. Практически единственной операцией, используемой в скремблерах и дескремблерах, является XOR - "побитное исключающее ИЛИ" (сложение по модулю 2).

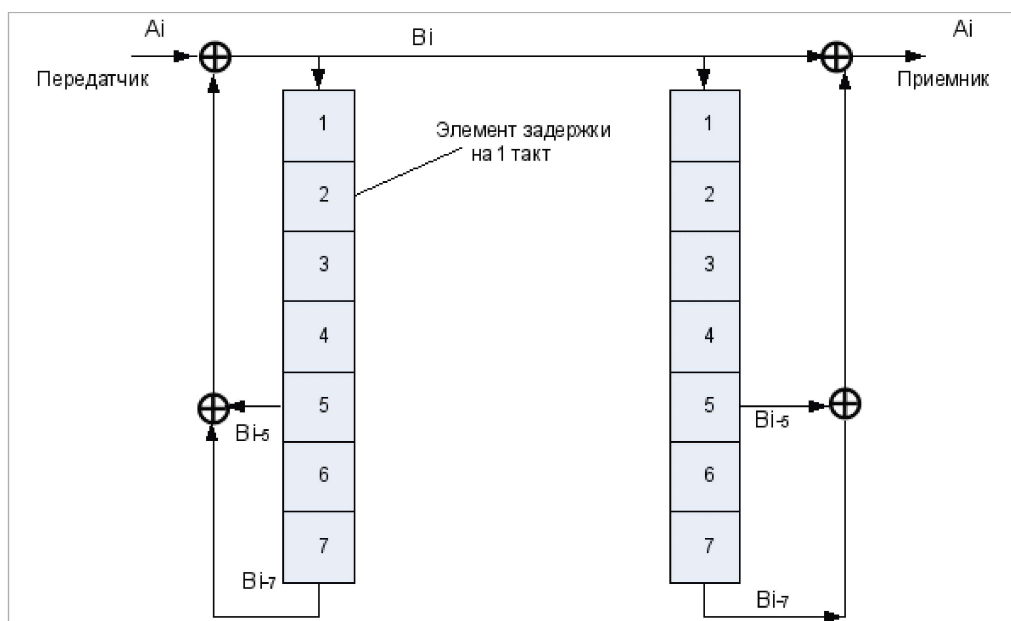


Рис.4.8. Вариант реализации скремблирования

Пусть, например, скремблер реализует соотношение  $B_i = A_i + B_{i-5} + B_{i-7}$ .

Здесь  $B_i$  – бит результирующего кода, полученный на  $i$ -м такте работы скремблера;  $A_i$  – бит исходного кода, поступающий в передатчике на вход скремблера на  $i$ -м такте;  $B_{i-5}$  и  $B_{i-7}$  – биты результирующего кода, полученные на предыдущих тактах работы скремблера, соответственно на « $i-5$ » и « $i-7$ » тактах.

Дескремблер в приемнике восстанавливает исходную последовательность, используя соотношение  $C_i = B_i + B_{i-5} + B_{i-7} = (A_i + B_{i-5} + B_{i-7}) + B_{i-5} + B_{i-7} = A_i$

## Выводы

В сетях высокоскоростной передачи данных используются различные виды логического и физического кодирования данных. Логическое кодирование выполняется передатчиком до физического кодирования средствами физического уровня. На этапе логического кодирования борются с недостатками методов физического цифрового кодирования - отсутствие синхронизации, наличие постоянной составляющей. Сформированные

исправленные последовательности данных затем с помощью методов физического кодирования передаются по линиям связи.

### Дополнительная информация.

Руководство по технологиям объединенных сетей. Cisco systems. Inc. Вильямс. 2004

### Контрольные вопросы.

1. Какой вид логического кодирования сигнала применяется в высокоскоростных оптоволоконных системах?
  - A. NRZ
  - B. Manchester
  - C. 8B6T
  - D. 8B10B
2. Декодируйте линейный сигнал 2B1Q: один символ – один элемент сигнала +1 -1 +1 -1 +3 +3 -3 +1
  - A. 1 1 0 1 1 1 0 1 1 0 1 0 0 0 1 1
  - B. 1 0 0 1 1 1 0 0 0 1 0 1 1 0 1 0
  - C. 0 1 0 1 0 1 0 0 1 1 0 1 1 0 0 0
  - D. 0 1 0 1 1 1 0 0 0 1 0 1 0 0 1 0
3. Определите скорость передачи линейного сигнала (В, Бод) при манчестерском кодировании, если скорость передачи информации составляет N=64Кбит/с:
  - A. 131072 Бод
  - B. 65536 Бод
  - C. 64 Кбод
  - D. 128 Кбод