МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)

КАФЕДРА РАДИОТЕХНИЧЕСКИХ СИСТЕМ (РТС)

Ю.П. Акулиничев, А.С. Бернгардт

ТЕОРИЯ ЭЛЕКТРИЧЕСКОЙ СВЯЗИ

Учебное пособие

Для студентов технических направлений и специальностей

Томск 2015

Ю.П. Акулиничев, А.С. Бернгардт

ТЕОРИЯ ЭЛЕКТРИЧЕСКОЙ СВЯЗИ: Учебное пособие. –Томск: 2015. - 196 с.

Рассматриваются общие положения статистической теории передачи сигналов. Вводятся математические модели сигналов и помех, описаны основные преобразования, которым подвергаются сигналы в процессе их передачи и приема. Рассмотрены базовые элементы устройств, осуществляющих указанные преобразования. Рассмотрены также базовые методы аналоговой и цифровой модуляции, используемые в современных системах передачи информации. Большое внимание уделено вопросам кодирования сигналов: для сокращения избыточности или для повышения помехозащищенности. Освещены методы приема сигналов в когерентных и некогерентных системах связи при наличии шума, а также в условиях многолучевости. Дан анализ методов многостанционного доступа с частотным, временным и кодовым разделением каналов. Определены принципы обмена информацией в телекоммуникационных сетях.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ…….…….. **5**
1 Математические модели сигналов и помех ……..…. 7
1.1 Информация и сигналы…….……..... 7
1.2 Цифровые сигналы..……..……..... 9
1.3 Дискретные сигналы..……..……..…….. 12
1.4 Непрерывные сигналы..……..……..…... 14
1.5 Аддитивные и мультипликативные помехи..…..... 15
1.6 Методы аналитического и геометрического представления сигналов и помех..……..…….. 18
2 ПРЕОБРАЗОВАНИЯ СИГНАЛОВ ……..……. 24
2.1 Модель системы передачи информации..……..……. 24
2.2 Элементы преобразователей..……..……. 30
2.3 Преобразование неэлектрических сигналов в электрические..……. 31
2.4 Квантование по времени непрерывного сигнала..……. 32
2.5 Модуляция импульсной несущей дискретным сигналом..…...…… 36
2.6 Аналого-цифровое и цифроаналоговое преобразования..……. 37
2.7 Линейная цифровая фильтрация и генерирование последовательностей символов……..……. 40
2.8 Модуляция гармонической несущей цифровым сигналом..…….. 44
2.9 Корреляционный прием и согласованная фильтрация..……. 51
2.10 Модуляция гармонической несущей непрерывным сигналом..…... 52
2.11 Каналы передачи сигналов..……... 54
2.12 Последовательный и параллельный способы передачи..……. 57
2.13 Статистический синтез цифровой системы передачи информации **……** 59
3 КОДИРОВАНИЕ КАНАЛА ……... 64
3.1 Корректирующие коды ……..……..…… 64
3.2 Линейные блочные коды ……..……..... 67
3.3 Коды Хэмминга ……..……... 74
3.4 Циклические коды ……..... 76
3.5 Декодирование в СПИ с каналом переспроса ……..… 84
3.6 Свёрточные коды ……..…… 86
3.7 Перемежение символов …… 89
3.8 Комбинирование кодов ……..……... 91
4 КОДИРОВАНИЕ ИСТОЧНИКА ……..... 94
4.1 Собственная информация и избыточность (цифровые сигналы) …. 94
4.2 Кодирование источника ……..…… 98
4.3 Взаимная информация ……..... 104
4.4 Пропускная способность канала и теоремы о кодировании в цифровом канале с помехами …..... 106

4.5 Пропускная способность непрерывного канала с шумом ….…….. 110
5 ДЕМОДУЛЯЦИЯ ЦИФРОВЫХ СИГНАЛОВ 114
..... 114
5.1 Роль априорной информации……..... 116
5.2 Когерентные системы …….....…….. 122
5.3 Некогерентные системы..…… 124
5.4 Частично-когерентные системы…….…. 127
5.5 Прием сигнала в условиях многолучевости……... 131
5.6 Регенерация цифрового сигнала в ретрансляторах……
5.7 Особенности СПИ, в которых применяется помехоустойчивое 133
6 кодирование ……..…….....…
МНОГОКАНАЛЬНАЯ ПЕРЕДАЧА И МНОГОСТАНЦИОННЫЙ 138
ДОСТУП …….. 138
6.1 Методы многостанционного доступа ……... 141
6.2 Многостанционный доступ с частотным разделением каналов…... 144
6.3 Многостанционный доступ с временным разделением каналов … 147
6.4 Многостанционный доступ с кодовым разделением каналов…..… 149
6.5 Синхронизация в СПИ с многостанционным доступом...……..... 154
7 6.6 Расширение спектра сигнала……. **159**
ПРИНЦИПЫ ПОСТРОЕНИЯ СЕТЕЙ ЭЛЕКТРОСВЯЗИ……... 159
7.1 Определения, классификация, структуры сетей……. 160
7.2 Коммутация каналов и коммутация пакетов…….. 164
7.3 Центры коммутации……..…..……..….
7.4 Дейтаграммный метод передачи и передача с предварительным 169
установлением соединения……..… 173
7.5 Элементы теории телетрафика……..……
ЗАКЛЮЧЕНИЕ…..……...…. 180
ЛИТЕРАТУРА…….…..... 181
ПРИЛОЖЕНИЯ……..……..... 183
УСЛОВНЫЕ ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ 189
ГЛОССАРИЙ ……. 190

ВВЕДЕНИЕ

Считается, и не без оснований, что XXI век – это век информатизации. Поэтому знание фундаментальных закономерностей и методов передачи, преобразования и хранения информации для современного человека необходимо не менее чем знание основных законов физики, химии и других естественных наук. В подтверждение достаточно лишь отметить, что вся электронная аппаратура создается для передачи и обработки сигналов, несущих информацию. Исключения настолько редки, что придется хорошо подумать, чтобы привести хотя бы 2-3 соответствующих примера.

Системы передачи информации (СПИ) являются одной из самых динамичных отраслей мировой экономики. В указанных системах, особенно цифровых, воплощены практически все новейшие достижения науки и технологии, применяемые и в других радиотехнических системах. Нет ничего удивительного, что книги, содержащие более или менее связное описание лишь основных методов, имеют весьма солидный объем [1–3 и др.].

Главная цель, которая преследовалась при выборе способа подачи материала, – это максимальная доступность иногда даже в ущерб строгости и полноте освещения отдельных вопросов. Если у читателя сложится цельное представление о предмете, о назначении и взаимосвязи применяемых методов, это будет хорошей базой для более де- тального изучения избранных тем по другой учебной и научной литературе.

Чтобы материал, представленный в книге, не воспринимался как простой набор фактов, необходимо подчеркнуть следующее.

Во-первых, начальный этап любого исследования, зачастую требующий наибольшей квалификации, – это систематизация и количественное описание того, что известно об объекте исследования. Для СПИ это означает разработку математических моделей сигналов и помех. Поэтому нет смысла приступать к изучению курса, не вспомнив хотя бы основные понятия теории вероятностей и теории электрических цепей и сигналов.

Во-вторых, прием сигналов всегда требует вынесения серии решений в условиях частичной неопределенности. Поэтому нужно осознать, что не каждое из этих принятых решений оказывается правильным, поэтому при выборе методов передачи и приема стремятся к тому, чтобы доля удачных решений была максимальной.

В-третьих, улучшение одних характеристик СПИ неизбежно сопровождается ухудшением других, поэтому никогда нельзя рассматривать лишь одну сторону медали. Таким образом, проектирование СПИ и ее элементов сводится к количественному обоснованию цепи компромиссных вариантов.

В теории СПИ очень широко используются различные математические методы. Хотя следует признать, что количество основных идей не так уж ве-

лико. Поэтому за всеми математическими выражениями не мешало бы прежде всего разглядеть соответствующую идею. Недаром же большинство лучших современных методов обработки сигналов первоначально было предложено практиками на основе инженерной интуиции, и лишь потом математики строго обосновали их оптимальность.

Нет сомнения, что знание основ теории и техники передачи информации существенно облегчит освоение последующих специальных курсов. Формы разные, но суть одна – передача информации!

1 МАТЕМАТИЧЕСКИЕ МОДЕЛИ СИГНАЛОВ И ПОМЕХ

1.1 Информация и сигналы

Невозможно дать строгое и полное описание любого из природных объектов или явлений. Поэтому любое грамотное исследование всегда начинается с построения его *математической модели*. Это – упрощенное количественное описание объекта или явления, пригодное для решения конкретной задачи. Ясно, что различным задачам могут удовлетворять совершенно разные модели одного и того же явления.

В связи с этим сомнительно, что кто-то может дать строгое и удовлетворяющее всех определение понятия “информация”. Принято считать, что *информация – это сведения, знания, новости, идеи и т.п.* независимо от формы их представления.

Считают также, что информация – не материя, т.е. наличие её невозможно обнаружить с помощью органов чувств или приборов.

Ещё полагают, что информация – не энергия, поскольку она, в отличие от энергии, не может проявлять себя, превратившись в работу.

Чтобы не создалось впечатление, что весь курс будет посвящен изучению способов передачи чего-либо бестелесного, дадим ещё одно определение. *Сигнал – это материальный переносчик информации*, и информация, в нем содержащаяся, трансформируется в изменение некоторых свойств этого материального объекта. Поэтому вполне естественно, что изучение закономерностей передачи информации мы, из соображений наглядности, зачастую будем заменять изучением свойств сигналов в процессе их передачи от источника к получателю.

Итак, ясно, что передача информации невозможна без передачи сигналов. Обратное утверждение справедливо далеко не всегда. Если сведения, содержащиеся в сигнале, уже достоверно известны получателю, такой сигнал не несёт информации. С другой стороны, для менее информированного получателя тот же сигнал может содержать информацию.

Следовательно, информация – адресное понятие. Всегда можно указать два конечных объекта в цепи преобразований сигнала. Это – источник информации и получатель. *Источник информации* – это объект, текущее состояние которого X интересует получателя. Передаваемое сообщение X, трансформируясь в процессе передачи из пункта, где расположен источник, в пункт, где расположен получатель, превращается в итоге в *принимаемый сигнал* Y. Получатель, наблюдая конкретную реализацию y, пытается определить, каково конкретное значение сообщения x.

Отсюда видно, что у получателя всегда должна быть хотя бы некоторая неопределенность относительно текущего состояния источника X. Если такая неопределенность отсутствует (источник гарантированно должен быть в

единственном состоянии x_o, и получателю это уже известно), тратить ресурсы на прием сигнала *y* нет смысла (следовательно, и изучать в этой ситуации нечего).

Удобный математический аппарат для описания объектов, ожидаемое состояние которых является неопределенным, дает *теория вероятностей*. Поэтому при рассмотрении СПИ все сигналы в цепи, связывающей источник *X* с получателем *Y*, рассматривают как взаимозависимые случайные объекты.

Напомним **важнейшие понятия теории вероятностей**. *Опыт* — это совокупность действий, выполняемых по заранее оговоренным либо общепринятым правилам, например, подбрасывание монеты; извлечение карты из колоды; наблюдение погоды в конкретный день и час; измерение напряжения в конкретной точке устройства и т.д. *Событие A называется случайным событием, если при проведении опыта оно может произойти, а может не произойти, причем исход заранее неизвестен*. Примеры: выпадение герба при подбрасывании монеты; извлечение из колоды бубновой десятки; снегопад в полдень 20 января 2010 года; величина напряжения меньше 2 В.

Вероятность $p(A)$ события A — это число, характеризующее степень возможности появления этого события. Событие называется неслучайным (детерминированным), если $p(A)=0$ (это невозможное событие) либо $p(A)=1$ (это достоверное событие). Для случайного события $0 \le p(A) \le 1$.

Для некоторых событий вероятность их появления можно точно указать сразу на основе теоретических предпосылок. Например, вероятность выпадения герба равна 0,5. Для многих других событий остается лишь экспериментальный способ оценки вероятности, т.е. $p(A_J) \approx m/n$, где n — общее количество проведенных опытов; m — количество опытов, в которых появилось событие A. Очевидно, что точность оценки возрастает при увеличении количества опытов n.

Как правило, бывают известны последствия, которые повлечет за собой появление (и непоявление) события A, поэтому естественно желание заранее, до проведения опыта принять *решение* относительно того, как наилучшим образом подготовиться к преодолению этих последствий. Проблема в том, что способы подготовки к обоим исходам могут быть диаметрально противоположными, исключающими друг друга. Например, если $p(A)=0{,}98$, то мы, скорее всего, примем решение готовиться к тому, что событие A появится.

Конечно, мы можем ошибиться в своих прогнозах. Тем не менее, правильность избранной стратегии подтверждает следующее соображение. Если будем следовать ей при проведении большого количества опытов, то начинает работать закон больших чисел и проявляются некоторые закономерности. В нашем примере при проведении n = 10000 опытов событие A появится примерно m = 9800 раз и не появится всего в 200 опытах.

Решение будет еще более обоснованным, если есть возможность количественно определить величины рисков (потерь), связанных с каждым из возможных исходов опыта (кстати, величина потерь со знаком минус означает соответствующую прибыль).

Рассмотрим пример. Некто предлагает вам поиграть по следующим правилам: бросаете игральную кость и, если выпадает шестерка (вероятность этого события равна 1/6), вы платите 6 рублей; при другом исходе вы получаете 1 рубль. Итак, величина потерь X — это *дискретная случайная величина*, она имеет два возможных значения: $x_1 = 6$ и $x_2 = -1$. Стоит ли играть на таких условиях? На первый взгляд, игра кажется справедливой, тем не менее, подсчитаем среднюю величину потерь. Положим, будет проведено 600 опытов. Шестерка выпадет примерно 100 раз, при этом вы потеряете 600 рублей. Иной исход будет приблизительно в 500 опытах, и вы получите 500 рублей. Баланс нежелательный, средняя величина потерь в расчете на один опыт составит 1/6 рубля. В теории вероятностей эта числовая характеристика называется *математическим ожиданием случайной величины X*. Как мы только что убедились, она вычисляется по формуле $m_x = x_1 p(x_1) + x_2 p(x_2) = 6 \cdot 1/6 - 1 \cdot 5/6 = 1/6$.

Можно определить *полную группу несовместных случайных событий* $A_1, A_2, \ldots, A_m$. Это значит, что в результате опыта обязательно произойдет одно из перечисленных событий, но только одно из них. Для каждого события A_j существует вероятность его появления $p(A_j)$, причем $p(A_1) + p(A_2) + \ldots + p(A_m) = 1$. Например, принято сообщение, состоящее из 200 букв. Тогда полной будет, скажем, такая группа случайных событий: A_1 — в тексте нет ошибок; A_2 — в тексте одна ошибка; A_3 — в тексте от двух до пяти ошибок; A_4 — в тексте более пяти ошибок.

Определив исходный пункт для построения модели сигналов, во всех последующих разделах курса мы будем заниматься вопросами вероятностного описания сигналов в различных звеньях тракта передачи, при этом основное внимание будем уделять тем характеристикам сигнала, которые в наибольшей степени зависят от передаваемого сообщения *X*. Способы определения количества информации, содержащейся в сигнале, будут рассмотрены в разделе 4.

1.2 Цифровые сигналы

Сигнал X, который может находиться лишь в одном из m возможных состояний, называется m-ичным символом. Перечень всех возможных значений символа $x_1, x_2, \ldots, x_m$, образующих полную группу несовместных событий, называется алфавитом, а число m – основанием кода (системы счисления).

Полное вероятностное описание символа дает его ряд распределения

x_j	x_1 x_2 … x_m
p	p_1 p_2 … p_m

(1.1)

При этом сумма чисел-вероятностей во второй строке равна единице.

Примеры: 1) X – русская буква, m=33, её возможные значения (алфавит): а,б,…,я. Есть экспериментальные данные о вероятностях появления каждой из букв [10].

2) X – десятичная цифра, m=10, алфавит: 0,1,2,…,9. Если не учитывать нашу любовь к круглым числам, можно считать, что все цифры алфавита имеют одинаковую вероятность по 0,1.

3) X – двоичная цифра (бит = bit = binary digit), алфавит: 0,1.

Конечно, формально можно представить символ, для которого m=1, но принимать такой сигнал нет смысла (см. разд. 1.1). Поэтому *бит – это сигнал простейшего вида*, и благодаря этому свойству двоичные сигналы нашли широчайшее применение в различных устройствах передачи, хранения и преобразования сигналов.

Для букв алфавита можно использовать любые (разумеется, различные) названия и обозначения – от этого сущность сигнала не изменится. *Если буквам алфавита приписать числовые значения, то символ X превращается в дискретную случайную величину*. Для СПИ это – типичная ситуация, поскольку различные состояния символа передаются путем задания конкретных значений частоты, фазы, амплитуды и иных количественных параметров сигнала.

Практически любую числовую характеристику дискретной случайной величины X можно определить как математическое ожидание (М) некоторой известной функции $y = \varphi(x)$

$$\mathrm{M}\left[\varphi(X)\right] = \sum_{j=1}^{m} \varphi\left(x_j\right) p_j. \qquad (1.2)$$

В частности, вычисление по формуле (1.2) дает: математическое ожидание m_x, если $\varphi(x) = x$; дисперсию D_x, если $\varphi(x) = (x - m_x)^2$; пятый начальный момент, если $\varphi(x) = x^5$; число m, если $\varphi(x) = 1/p(x)$, и т.д.

Любое из чисел, определяемых по формуле (1.2), не дает полного описания свойств символа X в отличие от ряда распределения (1.1), но их использование порой имеет смысл благодаря компактности представления (число, а не таблица).

При передаче каждому из возможных значений символа приписывается определенный смысл в соответствии с заранее оговоренной таблицей кодирования, т.е. эта таблица известна получателю. Например, при m=2 сообщение о возможном приезде конкретного лица может иметь следующий вид: 1 – приедет, 0 – не приедет. Если использовать восьмеричный символ, то можно передать более детальное (информативное) сообщение, составив таблицу,

например, так: 0 – не приедет; 1 – приедет в понедельник; …; 7 – приедет в воскресенье. Отсюда вывод: *чем больше m, тем больше информации (сведений) можно вложить в данный символ*. Но использование символов с большим *m* порождает ряд проблем (см., например, разд. 2.), поэтому часто для передачи сообщения используют не один, а несколько *m*-ичных символов.

Кодовая комбинация $X^{[1]}, X^{[2]}, ..., X^{[n]}$, состоящая из *m*-ичных символов, нередко называется *кодовым словом* длины *n*. Другое возможное название - это *n*-разрядное *m*-ичное число. Пусть все *n* символов, входящих в это слово, имеют один и тот же алфавит $x_1, x_2, ..., x_m$.

Пока рассматриваем эту комбинацию как единое целое и перечислим все её возможные состояния:

$$x_1^{[1]}, x_1^{[2]}, ..., x_1^{[n]}; \quad x_2^{[1]}, x_1^{[2]}, ..., x_1^{[n]}; \quad ...; \quad x_m^{[1]}, x_m^{[2]}, ..., x_m^{[n]}. \tag{1.3}$$

Число этих состояний равно

$$M = m^n. \tag{1.4}$$

Отсюда следует, что комбинация символов, в принципе, ничем не отличается от одного *M*-ичного символа, если не считать использования громоздких обозначений для каждой из *M* букв алфавита (1.3). И в этом случае для полного описания сигнала нужно знать ряд распределения (1.1), который в первой строке содержит возможные значения сигнала (1.3), а во второй строке – числа, т.е. вероятности появления каждого из этих значений

$$p\left(x_i^{[1]}, x_j^{[2]}, ..., x_\kappa^{[n]}\right), \tag{1.5}$$

причем сумма этих *M* вероятностей, естественно, равна единице.

Далее, формула (1.2) определяет способ вычисления любой числовой характеристики и для кодовой комбинации, и в этом случае она принимает вид

$$\begin{aligned} &\mathrm{M}\left[\varphi\left(X^{[1]}, X^{[2]}, ..., X^{[n]}\right)\right] = \\ &= \sum_{i=1}^{m}\sum_{\tau=1}^{m}\cdots\sum_{\kappa=1}^{m}\varphi\left(x_i^{[1]}, x_\tau^{[2]}, ..., x_\kappa^{[n]}\right)p\left(x_i^{[1]}, x_\tau^{[2]}, ..., x_\kappa^{[n]}\right). \end{aligned} \tag{1.6}$$

Имея ряд распределения для комбинации в целом, можно найти вероятностные и числовые характеристики для любого символа или группы символов, входящих в комбинацию.

Для примера рассмотрим комбинацию X,Y, состоящую из двух (n=2) *m*-ичных символов. Имея m^2 вероятностей совместного появления значений x_j и y_k (1.5), по формуле сложения вероятностей можно найти вероятности для каждого из символов

$$p\left(x_j\right) = \sum_{\kappa=1}^{m} p\left(x_j, y_\kappa\right), \quad p\left(y_\kappa\right) = \sum_{j=1}^{m} p\left(x_j, y_\kappa\right). \tag{1.7}$$

Далее, воспользовавшись формулой умножения вероятностей

$$p\left(x_j, y_\kappa\right) = p\left(x_j\right)p\left(y_\kappa / x_j\right) = p\left(y_\kappa\right)p\left(x_j / y_\kappa\right), \tag{1.8}$$

можно построить *m* условных рядов распределения символа *X*, по одному ря-

ду для каждого из значений $k = 1,2,..., m$,

$$p(x_j / y_к) = \frac{p(x_j, y_к)}{p(y_к)} \tag{1.9}$$

и аналогично m условных рядов распределения для Y

$$p(y_к / x_j) = \frac{p(x_j, y_к)}{p(x_j)}. \tag{1.10}$$

Для определения какой-либо условной числовой характеристики символа X воспользуемся той же формулой (1.2), но подставлять нужно условные вероятности, соответствующие конкретному значению y символа Y.

Два символа X и Y являются *независимыми*, если $p(x_j / y_к) = p(x_j)$ и, следовательно, $p(y_к / x_j) = p(y_k)$ для всех возможных значений j и k. Независимость символов упрощает их описание, поскольку любая условная характеристика символа равна соответствующей безусловной.

Описание последовательности зависимых символов для большинства ситуаций, представляющих практический интерес, является хотя и очень простым по смыслу, но порой чрезвычайно громоздким. Одной из самых простых моделей, учитывающих зависимость символов в их бесконечной последовательности $..., X^{[k-1]}, X^{[k]}, X^{[k+1]}, ...$, является простая марковская цепь

$$p\left(x_i^{[k+1]} / x_j^{[k]}, x_s^{[k-1]}, ...\right) = p\left(x_i^{[k+1]} / x_j^{[k]}\right), \tag{1.11}$$

то есть влияние всех предшествующих символов передается “по цепочке”, в итоге значение текущего символа формально зависит от одного предыдущего.

Многие сообщения по своей природе являются *цифровыми, т.е. состоящими из последовательности символов*. Таковы тексты, различные числовые массивы. Более того, в современной технике связи очень часто сообщения, которые по своей природе не относятся к цифровым (например, звуки и изображения), преобразуются в последовательность двоичных символов.

1.3. Дискретные сигналы

Рассмотрим такие случайные функции времени $X(t)$, которые являются дискретными во времени t, но непрерывны по величине. Поведение такого сигнала можно полностью описать последовательностью $X^{[1]}, X^{[2]}, ..., X^{[n]}$, состоящей из n *непрерывных* случайных величин. Напомним, что в цифровом сигнале каждая из этих величин является дискретной.

Начнем с простейшего случая $n = 1$, т.е. сигнал – это просто **непрерывная случайная величина** X. Большинство физических величин (температура, давление, скорость, напряжение и т.п.) являются непрерывными случайными величинами, т.е. каждая из них может принять любое значение внутри некоторого непрерывного интервала.

Полное описание непрерывной случайной величины содержит её плотность вероятностей $W(x)$. Это неотрицательная функция, удовлетворяющая условию нормировки

$$\int_{-\infty}^{\infty} W(x)dx = 1. \tag{1.12}$$

Любую из числовых характеристик находим с помощью формулы, подобной (1.2),

$$M[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x)W(x)dx, \tag{1.13}$$

задав функцию $\varphi(x)$ соответствующего вида. Одной из самых полезных моделей является *нормальная (гауссовская) случайная величина*, для которой кривая плотности вероятностей – колокольного вида

$$W(x) = \frac{1}{\sqrt{2\pi}\sigma_x} exp\left[-\frac{(x-m_x)^2}{2\sigma_x^2}\right], \tag{1.14}$$

где m_x и σ_x – математическое ожидание и среднеквадратическое отклонение соответственно.

Полной характеристикой **последовательности непрерывных случайных величин** $X^{[1]}, X^{[2]}, ..., X^{[n]}$ является совместная плотность вероятностей $W\left(x^{[1]}, x^{[2]}, ..., x^{[n]}\right)$.

Для определения числовых характеристик используется аналог формулы (1.6)

$$\begin{aligned} &\mathrm{M}\left[\varphi\left(X^{[1]}, X^{[2]}, ..., X^{[n]}\right)\right] = \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi\left(x^{[1]}, x^{[2]}, ..., x^{[n]}\right) W\left(x^{[1]}, x^{[2]}, ..., x^{[n]}\right) dx^{[1]} \cdots dx^{[n]}. \end{aligned} \tag{1.15}$$

Чтобы использовать более компактные обозначения, систему случайных величин удобно представить в виде вектора-строки

$$\mathbf{X} = \left(X^{[1]}, X^{[2]}, ..., X^{[n]}\right), \tag{1.16}$$

а вектор её конкретных значений точно так же обозначить строчной буквой **x**.

Наиболее часто используются следующие числовые характеристики: вектор-строка математических ожиданий

$$\mathbf{m_x} = \left(m^{[1]}, m^{[2]}, ..., m^{[n]}\right) = \mathrm{M}\left[\left(X^{[1]}, X^{[2]}, ..., X^{[n]}\right)\right] \tag{1.17}$$

и ковариационная матрица **R**, элементами которой служат числа

$$R_{j,k} = \mathrm{M}\left[\left(X^{[j]} - m_j\right) \cdot \left(X^{[k]} - m_k\right)\right], \quad j,k = 1,2,...,n. \tag{1.18}$$

Для независимых случайных величин нет необходимости использовать понятие условной плотности вероятностей, поэтому

$$W\left(x^{[1]}, x^{[2]}, ..., x^{[n]}\right) = W_1\left(x^{[1]}\right) W_2\left(x^{[2]}\right) \cdots W_n\left(x^{[n]}\right) \tag{1.19}$$

и, следовательно, ковариационная матрица **R** является диагональной.

Система n гауссовских случайных величин описывается плотностью вероятностей

$$W(\mathbf{x}) = (2\pi)^{-n/2} (\det \mathbf{R})^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{m}_\mathbf{x}) \mathbf{R}^{-1} (\mathbf{x}^\mathbf{T} - \mathbf{m}_\mathbf{x}^\mathbf{T}) \right], \qquad (1.20)$$

где $\det\mathbf{R}$ – определитель матрицы $\mathbf{R}$,

$\mathbf{x}^\mathrm{T}$ – вектор-столбец $\mathbf{x}$.

1.4 Непрерывные сигналы

Математической моделью непрерывного (аналогового) сигнала является непрерывный случайный процесс $X(t)$. Такой процесс характеризуется ансамблем (множеством) реализаций. Ничтожная часть этого множества для примера показана на рис. 1.1 (кстати, если бы эти 5 реализаций, изображенных на рисунке, составляли всё множество, то их с полным правом можно было бы считать сигналами, отображающими 5 возможных значений пятиричного символа).

Вид каждого элемента этого множества, в принципе, известен заранее,

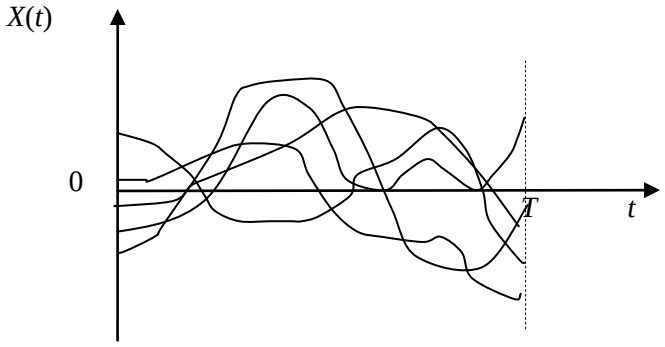

Рис. 1.1. Часть ансамбля реализаций непрерывного случайного процесса

а вся случайность заключается в том, что заранее неизвестно, которая из этих реализаций появится.

Вероятностное описание непрерывного процесса традиционно начинается с того, что мысленно проводят его квантование по времени с шагом $\Delta t = T/n$, где n – количество отсчетов, получающихся в результате квантования. В итоге непрерывный случайный процесс $X(t)$ заменяется последовательностью непрерывных случайных величин-отсчетов $X^{[1]}, X^{[2]}, ..., X^{[n]}$. Далее для этой системы случайных величин определяют желаемые вероятностные характеристики по методике, описанной в разд. 1.3. И, наконец, предполагают, что предельные значения этих характеристик при $\Delta t \to 0$ являются характеристиками самого случайного процесса $X(t)$.

Наиболее важные числовые характеристики непрерывного случайного

процесса: математическое ожидание $m_x(t)$ –детерминированная функция (аналог вектора (1.17)) и ковариационная функция $R(t_1,t_2)$ (аналог (1.18)).

Для *стационарного случайного процесса* математическое ожидание – это константа m_x, а ковариационная функция является детерминированной функцией одной переменной $\tau = t_1 - t_2$.

Для стационарного случайного процесса наглядной характеристикой является спектр плотности мощности $G(\omega)$, связанный с ковариационной функцией $R(\tau)$ парой преобразований Фурье

$$G(\omega) = \int_{-\infty}^{\infty} R(\tau)\exp(-i\omega\tau)d\tau,$$
$$R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega)\exp(i\omega\tau)d\omega. \qquad (1.21)$$

Понятие непрерывной случайной функции широко применяется для описания многих реальных сигналов. В частности, звуковой сигнал – это непрерывный случайный процесс $X(t)$. Изображение – это случайная функция $X(y_1, y_2)$ двух пространственных координат y_1 и y_2 (случайную функцию нескольких переменных называют случайным полем).

1.5 Аддитивные и мультипликативные помехи

Сигнал на выходе передатчика $x(t)$, распространяясь вдоль линии, связывающей пункты передачи и приема, изменяет свою форму. В итоге на выходе линии (на входе приемника) имеем сигнал $y(t)$, отличающийся от переданного. Проблема заключается в том, что эти изменения случайны (детерминированные изменения сигнала, возникающие в линии, в принципе, всегда могут быть точно учтены и полностью скомпенсированы, поэтому пока мы их не будем рассматривать). Такая *случайная трансформация сигнала в линии называется воздействием случайной помехи.*

Помеха $Z(t)$ называется **аддитивной** (от английского add-прибавлять), если

$$Y(t) = kX(t) + Z(t), \qquad (1.22)$$

где *k* – постоянный коэффициент передачи линии.

Причиной возникновения аддитивной помехи являются существующие всегда тепловой шум, электромагнитные поля, создаваемые посторонними источниками излучения естественного и искусственного происхождения, и т.п.

В качестве *модели аддитивной помехи обычно принимают белый шум*, т.е. стационарный гауссовский случайный процесс, имеющий нулевое математическое ожидание и равномерный и бесконечно широкий спектр плотно-

сти мощности. Если реализацию белого шума $Z(t)$ подать на вход идеального полосового фильтра, обладающего единичным коэффициентом передачи в полосе частот

$$\Delta f = f_{\text{â}} - f_{\text{í}}, \tag{1.23}$$

то на выходе фильтра будем иметь реализацию также стационарного гауссовского случайного процесса $V(t)$ с нулевым математическим ожиданием, называемого ограниченным по полосе белым шумом.

При любых $f_{\text{â}} \geq f_{\text{í}} \geq 0$ его дисперсия (средняя мощность шума) равна

$$\sigma_{\text{ø}}^2 = P_{\text{ø}} = N_0 \Delta f, \tag{1.24}$$

а n отсчетов $v_1, ..., v_n$, взятых с шагом $\Delta t = 1/(2\Delta f)$, определяемым теоремой отсчетов Найквиста-Котельникова, являются независимыми случайными величинами и описываются n-мерной совместной плотностью вероятностей

$$W(v_1, ..., v_n) = W(v_1) \cdots W(v_n) = \\ = \frac{1}{(2\pi)^{n/2} \sigma_{\text{ø}}^n} \exp\left(-\frac{1}{2\sigma_{\text{ø}}^2} \sum_{j=1}^{n} v_j^2 \right). \tag{1.25}$$

Величина N_0 – это спектральная плотность мощности белого шума и в соответствии с (1.24) она имеет размерность энергии Вт·с.

Мультипликативная помеха (от английского multiply-умножать) обычно является следствием многолучевости при радиосвязи в условиях, когда электрические свойства атмосферы или рельеф подстилающей поверхности случайным образом изменяются в пространстве и во времени (распространение радиоволн в городе или в неоднородной атмосфере).

Общий подход к описанию мультипликативной помехи заключается в том, что линия с многолучевостью представляется как линейный фильтр со случайно изменяющимися параметрами

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(t, \omega) S_x(\omega) \exp(i\omega t) d\omega, \tag{1.26}$$

где $S_x(\omega)$ – спектр сигнала $x(t)$,

$H(t, \omega)$ – передаточная функция этого фильтра, которая является комплексной случайной функцией двух переменных (для обычного фильтра с постоянными параметрами она зависит лишь от ω).

Наиболее важные числовые характеристики статистически однородной передаточной функции – это ширина $\Delta t_{н}$ её временной ковариационной функции

$$R(\Delta t) = \mathrm{M}\left[H(t, \omega) H^*(t + \Delta t, \omega) \right] \tag{1.27}$$

и ширина $\Delta \omega_{н}$ её частотной ковариационной функции

$$R(\Delta \omega) = \mathrm{M}\left[H(t, \omega) H^*(t, \omega + \Delta \omega) \right]. \tag{1.28}$$

Практически для всех реальных каналов $H(t, \omega)$ медленно изменяется

по t и ω ($\Delta t_н$ имеет порядок долей секунды, $\Delta f_н = \Delta\omega_н / 2\pi$ имеет порядок десятков килогерц и даже более), и в итоге

$$\Delta t_f \cdot \Delta f_f >> 1. \qquad (1.29)$$

Благодаря последнему свойству удается наблюдать почти в чистом виде мультипликативные помехи двух типов.

Если входной сигнал узкополосный в том смысле, что его спектр $S_x(\omega)$ сосредоточен в узкой полосе частот $\Delta\omega << \Delta\omega_н$, можно приближенно считать, что передаточная функция канала практически постоянна в полосе частот $\Delta\omega$, то есть $H(t,\omega) \approx H(t,\omega_0)$, где ω_0 – центральная частота спектра. Тогда из (1.26) имеем

$$y(t) \approx H(t,\omega_0)\, x(t), \qquad (1.30)$$

поскольку $x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)\exp(i\omega t)\, d\omega$.

Отсюда видно, что помеха проявляется как случайные изменения во времени коэффициента передачи, и это приводит к случайным медленным флуктуациям амплитуды и фазы сигнала. Если при этом длительность элементов (импульсов, например), из которых состоит сигнал, меньше чем $\Delta t_н$, то форма импульсов практически не меняется. Тем не менее, форма спектра выходного сигнала $S_y(\omega)$ случайным образом отличается от $S_x(\omega)$, при этом спектр в среднем несколько расширяется (величина расширения спектра имеет порядок $1/\Delta t_f$).

Такой тип помехи классифицируется как *временные селективные замирания* (фединг). Флуктуации уровня сигнала приводят к тому, что качество связи случайно изменяется во времени, а в отдельные периоды, при наиболее глубоких замираниях, связь может нарушаться полностью.

Наличие помехи второго типа можно обнаружить, если одновременно передавать два узкополосных сигнала, имеющие центральные частоты ω_1 и ω_2. Их замирания будут дружными (почти синхронными во времени), если $|\omega_1 - \omega_2| << \Delta\omega_н$. В противном случае, при большом разнесении по частоте $|\omega_1 - \omega_2| > \Delta\omega_н$, замирания сигналов практически независимы.

Помеха второго типа преобладает в случае, когда длительность сигнала $\Delta t << \Delta t_н$, и поэтому можно приближенно считать $H(t,\omega) \approx H(t_0,\omega)$, где t_0 – момент времени, соответствующий центру этого сигнала (импульса). Тогда из (1.26) получаем

$$S_y(\omega) \approx H(t_0,\omega) S_x(\omega), \qquad (1.31)$$

т.е. многолучевой канал воздействует на передаваемый сигнал как фильтр с постоянной во времени, но случайной частотной характеристикой.

Сравнив выражения (1.30) и (1.31), мы можем считать помеху второго типа как *частотные селективные замирания*. Более того, приведенную выше

характеристику помехи первого типа можно перенести на помеху второго типа, поменяв местами переменные t и ω, т.е. сам сигнал $y(t)$ и его спектр $S_y(\omega)$. Случайные искажения спектра сигнала приводят к тому, что импульсы в процессе передачи случайным образом искажаются, растягиваются (рассеиваются) во времени тем больше, чем больше отношение $\Delta\omega/\Delta\omega_н$, при этом средняя величина рассеяния во времени имеет порядок $1/\Delta f_i$.

Понятно, что помехоустойчивость приема сигнала, форма которого случайна, ниже по сравнению с ситуацией, когда форма ожидаемого сигнала известна заранее. Более того, рассеяние во времени приводит к наложению соседних импульсов друг на друга и к их дополнительным искажениям. Такое явление известно как *межымпульсная или межсимвольная интерференция*.

Если $\Delta t << \Delta t_н$ и $\Delta\omega << \Delta\omega_н$, то форма сигнала не искажается

$$y(t) = H(t_0, \omega_0)\, x(t), \qquad (1.32)$$

хотя при этом модуль и аргумент постоянного коэффициента передачи $H(t_0, \omega_0)$ являются случайными величинами. В итоге это определяет случайность амплитуды и фазы принимаемого сигнала $y(t)$. Если же $\Delta t > \Delta t_н$ и $\Delta\omega > \Delta\omega_н$, то существенно совместное влияние мультипликативной помехи обоих типов.

Более того, многолучевость приводит ещё и к появлению *пространственных селективных замираний*, т.е. к случайным искажениям в пространстве амплитуд и фаз волн в пункте приема. В итоге флуктуации сигналов, принимаемых двумя разнесенными в пространстве антеннами, становятся всё менее зависимыми при увеличении разнесения. Для большинства каналов при разнесении антенн на расстояние в $(10\ldots100)\,\lambda$ эти флуктуации становятся практически независимыми.

Наличие любой помехи (аддитивной или мультипликативной) приводит к тому, что принимаемый сигнал $y(t)$ практически всегда отличается по форме от переданного сигнала $x(t)$.

1.6 Методы аналитического и геометрического представления сигналов и помех

Вероятностные модели сигналов и помех, описанные в разд. 1.1–1.5, вполне пригодны в качестве исходного материала для решения большинства задач теории электрической связи. Поэтому разработка новых моделей не является целью данного раздела. Здесь мы напомним те известные образы функций, задающих форму сигналов и помех, которые позволят в дальнейшем упростить и сделать более наглядным описание операций (не меняя их сути), которые мы будем проводить над сигналами.

Как было отмечено, любой сигнал (с помехой или без нее) – это функция времени $U(t)$, характеризующая изменение напряжения или тока в некоторой точке тракта передачи.

Узкополосный сигнал, у которого ширина спектра $\Delta\omega$ намного меньше несущей частоты ω_0, более удобно представить в виде

$$U(t) = A(t)\cos\theta(t) = A(t)\cos[\omega_0 t + \varphi(t)], \tag{1.33}$$

где функции $A(t) \geq 0$ и $\varphi(t)$ в явном виде задают законы амплитудной и фазовой модуляции соответственно. Зачастую ещё более удобна запись такого сигнала в виде суммы двух колебаний

$$U(t) = C(t)\cos\omega_0 t + S(t)\sin\omega_0 t, \tag{1.34}$$

где $C(t) = A(t)\cos\varphi(t)$ и $S(t) = -A(t)\sin\varphi(t)$ – *квадратурные составляющие* сигнала.

И, наконец, для действительного узкополосного сигнала широко используется представление в виде "*аналитического сигнала*"

$$\dot{u}(t) = A(t)\exp\left[i\omega_0 t + i\varphi(t)\right] = \dot{U}(t)\exp(i\omega_0 t), \tag{1.35}$$

т.е. комплексной функции, у которой реальная часть совпадает с (1.33)

$$U(t) = \mathrm{Re}[\dot{u}(t)]. \tag{1.36}$$

Функция $\dot{U}(t)$ может быть названа комплексной огибающей высокочастотного сигнала $U(t)$ и определяется следующим образом

$$\begin{aligned}\dot{U}(t) &= A(t)\exp\left[i\varphi(t)\right] = \\ &A(t)\cos\varphi(t) + iA(t)\sin\varphi(t) = C(t) - iS(t).\end{aligned} \tag{1.37}$$

Основное достоинство приведенных способов описания узкополосного сигнала $U(t)$ проистекает из того, что функции $A(t), \varphi(t), C(t), S(t), \dot{U}(t)$ - это медленноменяющиеся функции, т.е. каждая из них относительно мало изменяется на интервале времени $T = 2\pi/\omega_0$, равном периоду несущей частоты.

Легко убедиться в этом преимуществе, если воспользоваться **векторным представлением сигнала** $U(t)$, т.е. отображать его в виде вектора на плоскости в декартовой системе координат xOy так, что в любой момент времени длина вектора равна $A(t)$, а угол с осью Ox равен $\theta(t)$. Поскольку полная фаза сигнала $\theta(t)$ растет довольно быстро, вектор будет совершать в среднем f_0 оборотов в секунду. Использование понятия аналитического сигнала (или квадратурных составляющих) позволяет применить то же представление к комплексной величине $\dot{U}(t)$. В итоге скорость вращения вектора окажется значительно меньше, позволяя более детально проследить динамику развития колебаний. В этой ситуации направления двух осей координат символизируют колебания одной и той же частоты f_0, но сдвинутые по фазе на 90^o.

Другая возможность для создания довольно наглядного образа сигнала – это использование **геометрического представления сигнала**. Такая возможность вытекает из естественной или искусственно созданной дискретности сигнала $U(t)$ во времени и может применяться к сигналам с любой шириной спектра. Например, цифровой сигнал (последовательность символов-цифр) $S_1, S_2, ..., S_n$ можно формально считать элементами вектора

$$\mathbf{S} = (S_1, S_2, ..., S_n) \tag{1.38}$$

или координатами точки в *n*-мерном пространстве.

После квантования по времени непрерывного сигнала $U(t)$ с шагом Δt на интервале $0 \le t \le T$ также получаем последовательность чисел-отсчётов $u_1, u_2, ..., u_n$

$$U(t) \to \mathbf{U} = (U_1, U_2, ..., U_n) = (u_1, u_2, ..., u_n)\sqrt{\Delta t}, \tag{1.39}$$

которую аналогично можно считать элементами вектора **U**. Переход к пределу при $\Delta t \to 0$, $n \to \infty$ позволяет не учитывать ошибки квантования по времени (ввиду возможности использования такого перехода каждый элемент вектора **u** сразу умножили на постоянную величину $\sqrt{\Delta t}$).

Итак, каждому сигналу соответствует своя точка (вектор) в *n*-мерном пространстве (не путать с описанным выше векторным представлением колебаний). Поэтому посмотрим, какие свойства сигналов отображают основные геометрические понятия.

1) Норма (длина) вектора **U** в евклидовом пространстве равна

$$|\mathbf{U}| = \sqrt{\sum_{j=1}^{n} U_j^2}. \tag{1.40}$$

Если исходный сигнал $U(t)$ был непрерывным, то перейдем к пределу и получим

$$|\mathbf{U}| = \sqrt{\lim_{\Delta t \to 0} \sum_{j=1}^{n} u_j^2 \cdot \Delta t} = \sqrt{\int_0^T U(t)\,dt} = \sqrt{E_u}, \tag{1.41}$$

где E_u – энергия сигнала (по умолчанию принимается, что входное сопротивление цепи $R_{вх} = 1$Ом). Возможные значения $0 \le |\mathbf{U}| < \infty$. Обратим внимание на этот факт: длина вектора, соответствующего сигналу, зависит не от напряжения, тока или мощности, а от *энергии* этого сигнала. Это обстоятельство определяет для цифровых сигналов способ введения такого важного понятия, как отношение сигнал/шум.

Если сигнал (1.38) – это последовательность двоичных символов (нулей и единиц), то удобнее использовать *n*-мерное пространство Хэмминга. В нем длина вектора численно равна его весу $|\mathbf{S}| = \mathrm{W}(\mathbf{S})$, т.е. определяется как количество единиц в этом векторе. Возможные значения $|\mathbf{S}| = 0; 1; 2; ...; n$ – это целые числа.

2) Расстояние между двумя непрерывными сигналами $U(t)$ и $V(t)$, т.е. расстояние между точками, отображающими эти сигналы в евклидовом пространстве,

$$d(\mathbf{U},\mathbf{V}) = \sqrt{\lim_{\Delta t \to 0} \sum_{j=1}^{n} \left(u_j - v_j\right)^2 \Delta t} = \sqrt{\int_0^T \left[U(t) - V(t)\right]^2 dt} \qquad (1.42)$$

также может принимать любые неотрицательные значения $0 \le d < \infty$.

Расстояние между двумя двоичными последовательностями **S** и **C** – это расстояние Хэмминга

$$d(\mathbf{S},\mathbf{C}) = \mathrm{W}(\mathbf{S} \oplus \mathbf{C}) = \sum_{j=1}^{n} \left(S_j \oplus C_j\right), \qquad (1.43)$$

где $\oplus$ – знак суммирования по модулю 2 (напомним, что $0 \oplus 0 = 0,\ \ 0 \oplus 1 = 1 \oplus 0 = 1,\ \ 1 \oplus 1 = 0$). Другими словами, расстояние между двумя комбинациями численно равно количеству символов, в которых они отличаются одна от другой. Возможные значения расстояния Хэмминга $d = 0; 1; 2; \ldots; n.$.

3) Скалярное произведение двух векторов **U** и **V**, отображающих непрерывные сигналы, равно

$$(\mathbf{U},\mathbf{V}) = \lim_{\Delta t \to 0} \left(\sum_{j=1}^{n} u_j v_j \Delta t \right) = \int_0^T U(t) V(t)\, dt. \qquad (1.44)$$

Эти два сигнала считаются ортогональными, если $(\mathbf{U},\mathbf{V}) = 0.$

4) Линейная независимость системы N функций $U_1(t), U_2(t), \ldots, U_N(t)$, как и системы, состоящей из N векторов, заключается в том, что никакая из этих функций не может быть представлена в виде взвешенной суммы остальных функций, например

$$U_1(t) \ne \sum_{j=2}^{N} K_j U_j(t) \qquad (1.45)$$

при любых значениях коэффициентов K_j.

Для системы, состоящей из N двоичных n-разрядных комбинаций, определение линейной независимости аналогично, но при этом предполагается поэлементное суммирование комбинаций по mod 2, а весовые коэффициенты K_j – двоичные цифры.

Перечисленные свойства системы N функций $U_1(t), U_2(t), \ldots, U_N(t)$ в евклидовом пространстве удобно задать в виде симметрической матрицы Грама

$$\mathbf{\Gamma} = \begin{Vmatrix} \Gamma_{11} & \Gamma_{12} & \ldots & \Gamma_{1n} \\ \Gamma_{21} & \Gamma_{22} & \ldots & \Gamma_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \Gamma_{n1} & \Gamma_{n2} & \ldots & \Gamma_{nn} \end{Vmatrix}, \qquad (1.46)$$

где $\Gamma_{jk} = \left(\mathbf{U_j}, \mathbf{U_k}\right) = \int_{-\infty}^{\infty} U_j(t) U_k(t) dt$.

Тогда

$$\left|\mathbf{U_j}\right| = \sqrt{\Gamma_{jj}}, \quad d(\mathbf{U_j}, \mathbf{U_k}) = \Gamma_{jj} + \Gamma_{kk} - 2\Gamma_{jk}. \tag{1.47}$$

Система функций линейно независима, если определитель матрицы Грама не равен нулю

$$\det \mathbf{\Gamma} \neq 0, \tag{1.48}$$

а ортогональной системе функций соответствует диагональная матрица Грама.

Система функций называется нормированной, если все сигналы обладают единичной энергией ($\Gamma_{jj} = 1$ при всех j). Для такой системы величина $\rho_{jk} = \Gamma_{jk}$ называется коэффициентом корреляции сигналов $U_j(t)$ и $U_k(t)$, а расстояние между этими сигналами равно

$$d(\mathbf{U_j}, \mathbf{U_k}) = \sqrt{2(1-\rho_{jk})}. \tag{1.49}$$

В заключение следует сделать одно замечание. Используя последовательность чисел-отсчётов $u_1, u_2, ..., u_n$ для описания непрерывного сигнала, мы произвольно задавали их количество и даже использовали условие $\Delta t \to 0$, $n \to \infty$. Отсюда может сложиться впечатление, что всегда можно сформировать сколь угодно независимых или ортогональных сигналов. Теоретически это верно, но на практике всегда существуют ограничения на длительность сигнала $t_c \leq T$ и ширину его спектра $f_c \leq F$. Тогда максимальное число отсчетов, значения которых можно задавать произвольно, не нарушая заданных ограничений, называется *числом степеней свободы сигнала* и равно

$$n_c = 2TF. \tag{1.50}$$

Вопросы для самоконтроля по главе 1

1. В чем разница понятий “информация” и “сигнал”?
2. Приведите примеры радиоэлектронных устройств, предназначенных не для передачи информации.
3. Назовите два основных признака того, что сигнал не несет информации.
4. Почему для математического описания сигналов используют вероятностные модели?
5. Может ли детерминированный сигнал переносить информацию?
6. Какие случайные события (величины) называются независимыми?
7. Что нужно задать для полного вероятностного описания: символа? последовательности символов?
8. Сформулируйте, в чем состоит отличие цифрового сигнала от дискретного, от непрерывного.

9. Сформулируйте, в чем состоит отличие цифрового сигнала от дискретного, от непрерывного.
10. Что нужно задать для полного вероятностного описания: последовательности отсчетов сигнала? непрерывной случайной функции?
11. В чем отличие аддитивной помехи от мультипликативной? Приведите примеры каналов связи с такими помехами.
12. Какие преимущества дает представление сигналов как элементов векторного пространства?
13. В чем отличие Евклидова пространства от пространства Хэмминга?
14. Будут ли линейно независимые сигналы ортогональными?

2 ПРЕОБРАЗОВАНИЯ СИГНАЛОВ

2.1 Модель системы передачи информации

На рисунке 2.1 представлена укрупненная схема системы передачи информации (СПИ).

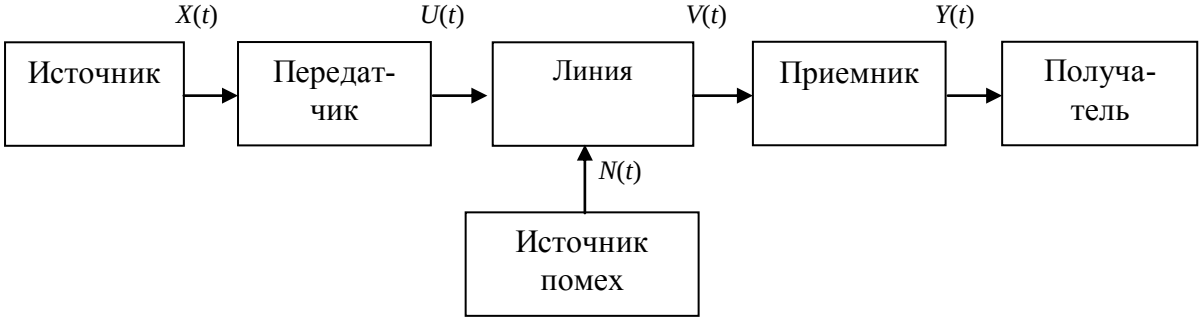

Рис. 2.1. Укрупненная схема СПИ

Источник информации – объект, состояние которого интересует получателя. Обозначим передаваемое сообщение как $X(t)$.

Получатель – объект, например человек, который, наблюдая конкретную реализацию сигнала $y(t)$, пытается как можно точнее определить значение переданного сообщения $x(t)$, т.е. сформировать оценку $\hat{x}(t) = f[y(t)]$. Зачастую такая оценка формируется уже в приемнике.

Линия – это среда, в которой распространяются сигналы из пункта передачи в пункт приема (электрический или оптический кабель; атмосфера, в которой распространяются радиоволны).

Источник помех – условное обозначение совокупности объектов, воздействие которых приводит к тому, что сигнал на выходе линии $v(t)$ случайным образом отличается от сигнала $u(t)$ на ее входе. В разделе 1.5 были рассмотрены основные виды помех. Следует отметить, что, в принципе, подобные помехи существуют во всех точках тракта передачи от источника информации до получателя, но степень искажения сигнала в наибольшей мере зависит от отношения энергий сигнала и помехи (отношение сигнал/помеха). При передаче по линии большой протяженности сигнал существенно ослабляется, поэтому решающим оказывается влияние помех, возникающих в линии передачи.

В широком смысле функции *передатчика* можно определить как согласование источника информации с линией. Более конкретно, передатчик выполняет такие функции, как преобразование передаваемого сообщения в форму, пригодную для передачи сигнала по конкретной линии (перевод сообщений в форму электрических или оптических сигналов; квантование непрерывного сигнала, преобразование его в последовательность символов;

модуляция несущей передаваемым сообщением и т.п.).

Задача приемника заключается в том, чтобы по принятой реализации $v(t)$ *угадать, какая именно из возможных реализаций сообщения* $x(t)$ *была передана*.

Можно сформулировать две крупные **статистические задачи, относящиеся к СПИ в целом**.

Задача *статистического анализа* конкретной СПИ заключается в определении количественных значений *основных показателей (скорости передачи информации и помехоустойчивости)* в ситуации, когда заданы способы преобразований сигналов во всех звеньях тракта и их количественные характеристики.

Задача *статистического синтеза* СПИ заключается в выборе наилучших способов преобразований сигналов, позволяющих построить оптимальную систему. При этом бессмысленно требование одновременно обеспечить максимальные значения обоих показателей, поскольку повышение скорости передачи при прочих постоянных условиях, как правило, ведет к снижению помехоустойчивости и наоборот. Поэтому *оптимальной можно назвать такую СПИ, которая при заданной скорости передачи обеспечивает максимально возможную помехоустойчивость, либо при заданной помехоустойчивости – максимальную скорость*. При корректном решении задачи синтеза оба критерия приводят к одному и тому же результату.

Наиболее сложной в математическом отношении является задача совместной оптимизации способа передачи и способа приема. Впервые такую задачу удалось решить (разумеется, частично, по отношению к кодеру и декодеру цифровой СПИ) в 1948 году Клоду Шеннону – творцу теории информации. Насколько сложной (с точки зрения теории и при практической реализации) является задача полного синтеза СПИ даже при самых благоприятных условиях ее функционирования, мы попытаемся показать в разд. 2.13. Поэтому наибольшее количество работ посвящено задаче частичного синтеза СПИ, а именно, оптимизации способа приема при заданных характеристиках передатчика и линии с помехами. Более того, по той же причине зачастую рассматриваются “наилучшие” способы реализации тех или иных операций над сигналами в передатчике или в приемнике (кавычки здесь использованы потому, что при этом обычно нет уверенности в том, являются ли наилучшими и другие операции над сигналами в остальных звеньях передатчика и приемника).

Поэтому и мы при изучении курса обычно будем говорить об оптимизации отдельных звеньев передатчика и приемника. Для этого кратко охарактеризуем назначение и суть **основных преобразований**, которым подвергается электрический сигнал в передатчике и приемнике СПИ. При этом из соображений удобства будем оперировать либо самим сигналом $u(t)$, принимающим действительные значения, либо его спектром $S(\omega)$. Известно, что

обе функции связаны парой преобразований Фурье

$$S(\omega) = \int_{-\infty}^{\infty} u(t)\exp(-i\omega t)dt,$$
$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)\exp(i\omega t)d\omega \quad , \qquad (2.1)$$

где i – мнимая единица.

Из преобразований **аналоговых сигналов** назовем хорошо известные и широко применяемые и в других радиоэлектронных устройствах линейные обратимые преобразования, такие как усиление сигнала, его сдвиг по оси времени (задержка) без изменения формы, сдвиг его спектра по оси частот (преобразование частоты) без изменения формы.

Модуляция – это нелинейная операция, в результате которой спектр сигнала переносится в область более высоких частот и зачастую изменяется его форма. Необходима для приведения сигнала к виду, пригодному для передачи по линии конкретного типа. Очевидно, что демодуляция это обратное преобразование сигнала в приемнике.

Следующие две операции направлены на упрощение структуры сигнала и поэтому не являются вполне обратимыми, то есть при обратном преобразовании восстановить исходный сигнал удается лишь с некоторой погрешностью.

Во-первых, это квантование по времени, в результате которого непрерывная случайная функция $u(t)$ на интервале времени $0 \le t \le T$ заменяется набором чисел – непрерывных случайных величин $u_1,...,u_n$. В общем виде линейная процедура квантования может быть определена набором соотношений

$$u_j = \int_{-\infty}^{\infty} u(t)\, f_j(t)dt, \quad j = 1,...,n, . \qquad (2.2)$$

где $f_j(t)$ – базисные (весовые) функции. С геометрической точки зрения (разд. 1.6) это – операции проектирования вектора- сигнала на базисные векторы.

Существует набор ортогональных базисных функций, при использовании которого по ряду чисел $u_1,...,u_n$ исходный сигнал восстанавливается по формуле

$$\hat{u}(t) = \sum_{j=1}^{n} u_j f_j(t) \qquad (2.3)$$

и при этом обеспечивается минимальная среднеквадратическая ошибка

$$\sigma = \sqrt{\int_{-\infty}^{\infty} \left|u(t) - \hat{u}(t)\right|^2 dt} \, . \qquad (2.4)$$

Тогда (2.3) называется разложением Карунена-Лоэва [2,9].

На практике обычно используется более простой способ: измеряют

значения функции $u(t)$ с шагом Δt (разд. 2.4).

Во-вторых, это квантование по уровню, то есть преобразование непрерывной случайной величины u, заданной на интервале ($u_н$, $u_в$), в дискретную случайную величину (m-ичный символ). Обычно m=2k, где k – целое число, и для реализации этого преобразования достаточно последовательно провести k операций сравнения величины u с заданными опорными значениями (разд. 2.6).

Дискретный сигнал $u_0,...,u_{n-1}$ (разд. 1.3) обычно получается в результате квантования по времени на интервале $0 < t < n\Delta t$ с шагом Δt непрерывного сигнала $u(t)$, поэтому он может подвергаться тем же преобразованиям, но в несколько иной форме.

Преобразования (2.1) проводятся в форме прямого и обратного дискретных преобразований Фурье (ДПФ)

$$S_k = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} u_j f^{-jk}, \quad u_j = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} S_k f^{jk}, \qquad (2.5)$$

где $f = \exp(2\pi i/n)$, $j,k = 0,\ldots,n-1$.

Числа $S_0,\ldots,S_{n-1}$ можно с некоторым приближением считать отсчетами его спектра $S(f)$, взятыми с шагом $\Delta f = \frac{1}{2n\Delta t}$ при условии, что весь спектр заключен в интервале $0...\frac{1}{2\Delta t}$.

Такие преобразования выполняются в числовом процессоре с использованием операций над числами с плавающей точкой, при этом применяется более экономный алгоритм быстрого преобразования Фурье (БПФ).

Фильтрация дискретного сигнала проводится при помощи цифрового фильтра (разд. 2.7).

Пожалуй, наиболее разнообразным и интересным оказывается набор преобразований **цифровых сигналов**.

Сигналом, применяемым для передачи символа, обычно является импульс тока или напряжения, поэтому естественно, что имеется широкий набор элементарных преобразований, необходимых для преобразования формы импульсов, изменения порядка их следования, хранения, выполнения ряда логических операций и т.п. (разд.2.2).

Более сложные преобразования, в которых участвуют целые группы символов, объединены под общим названием “кодирование”. Типичная ситуация такова. На вход кодирующего устройства поступает последовательность m-ичных символов длиной k (всего возможны $N_{вх} = m^k$ вариантов), и в ответ кодер выдает на выход последовательность таких же m-ичных символов длиной n (всего возможны $N_{вых} = m^n$ вариантов).

Различают три вида кодирования

1. *Кодирование источника* с целью сокращения избыточности, содер-

жащейся в сигнале, то есть представление сигнала $u(t)$ в наиболее экономном виде так, чтобы в конечном итоге произведение его длительности на занимаемую полосу частот было минимальным. Например, *безызбыточное кодирование* позволяет передать сообщение, используя минимальное количество символов n в последовательности на выходе кодера, в итоге оказывается $n < k$. Добиться такого результата можно лишь в том случае, когда возможные сигналы на входе кодера имеют существенно разные вероятности. Точнее, среди возможных последовательностей имеются столь маловероятные экземпляры, что можно даже не резервировать сигналы для их передачи, будучи почти уверенными в том, что ни одна из них на входе кодера не появится. Результат такого кодирования – это повышение скорости передачи информации (разд. 4.2).

2. *Канальное (помехоустойчивое) кодирование*, т.е. формирование цифрового сигнала $u(t)$ таким образом, чтобы в процессе приема появилась возможность обнаружить и исправить большую часть ошибок, возникших в принятом сигнале $v(t)$ из-за влияния помех. Эта операция, наоборот, обычно требует введения в передаваемый сигнал дополнительных, избыточных элементов, в итоге окажется $n > k$. Обсуждению проблем помехоустойчивого кодирования почти целиком посвящен разд. 3.

3. *Шифрование*, т.е. использование такого кода, способ декодирования которого известен лишь законному адресату (при этом обычно $n = k$).

Можно заметить, что передача сигналов во многом напоминает обычную транспортировку грузов из одного пункта в другой. Различные предметы перевозят лишь потому, что они несут в себе возможность удовлетворения некоторых человеческих потребностей (сигналы передают ради информации, в них содержащейся). В чем высшая суть этих потребностей, вряд ли кто сможет сформулировать, хотя и предлагается немало способов, как количественно оценить степень их удовлетворения (в настоящее время имеется несколько различных способов определения количества информации, содержащейся в сигнале).

Вагон, движущийся по рельсам (линия передачи), имеет ограниченный объем, поэтому естественной является задача экономной укладки груза, даже путем частичной разборки некоторых громоздких конструкций (задача безизбыточного кодирования), при этом в пункте назначения проводится сборка.

Разумеется, такая экономная укладка допустима лишь в случае, когда в пути отсутствуют факторы, способные испортить груз (помехи). Для защиты от вредных воздействий приходится перевозить дополнительные предметы – упаковку (помехоустойчивое кодирование). В пункте назначения упаковка, выполнившая свою функцию, отделяется от полезного груза (удаление избыточных символов после обнаружения и исправления ошибок).

Итак, можно наметить следующий порядок преобразований сигнала в

передатчике цифровой системы.

1) Аналого-цифровое преобразование, т.е. преобразование непрерывного сигнала в последовательность символов.

2) Кодирование источника.

3) Шифрование.

4) Канальное кодирование.

5) Модуляция гармонической несущей передаваемым сигналом с целью переноса его спектра в выделенную полосу частот.

6) Расширение спектра (такое преобразование, в результате которого ширина спектра сигнала $u(t)$ на входе линии оказывается существенно больше, чем та ширина, которая была бы минимально необходима для нормальной передачи этого сообщения), например, для обеспечения скрытности передачи.

Разумеется, не все перечисленные операции необходимо проводить в каждой СПИ, и последовательность перечисленных преобразований может несколько отличаться от указанной. В приемнике, естественно, проводятся обратные преобразования.

В заключение отметим те дополнительные операции с сигналами, которые нужно проводить в том случае, когда **для одновременной передачи сигналов от нескольких источников используются общие ресурсы.**

1) На рис. 2.1. изображена схема симплексной СПИ, предназначенной для передачи информации в одном направлении. Для организации одновременной двусторонней передачи, например, в телефонии, с использованием общей линии, необходимо решить проблему *дуплексирования*, т.е. сформировать сигналы $u_{\text{п}}(t)$ и $u_{\text{о}}(t)$ таким образом, чтобы при приеме разделять сигналы прямого и обратного направлений без их взаимного искажения.

2) Для одновременной передачи сигналов от n источников, находящихся на одном конце общей линии, к n получателям, находящимся на другом ее конце, осуществляется *уплотнение (мультиплексирование)* предварительно преобразованных сигналов (обычно это суммирование). Для этого передаваемые сигналы должны иметь такой вид, который позволяет в пункте приема выделять каждый сигнал из суммы и направлять своему получателю без взаимных искажений. Такая система называется *многоканальной (n-канальной) СПИ.*

3) В системах радиосвязи обычно возникает потребность в совместном использовании общего ретранслятора-приемопередатчика, находящегося на базовой станции в системах сотовой связи или на ИСЗ. При этом все источники и все получатели находятся в различных пунктах. Такая организация совместной работы называется *многостанционным (множественным) доступом*, при этом, в основном, используются те же методы, что и при многоканальной связи.

4) При построении системы обмена информацией (сети электросвязи),

предназначенной для обслуживания большого количества пользователей, исходят из того, что большинству абонентов не нужен постоянно закрепленный канал электросвязи, поскольку у них лишь эпизодически возникает потребность в передаче сравнительно коротких сообщений тому или иному получателю. Поэтому общие ресурсы, например линии, способны обеспечить одновременную передачу сообщений лишь какой-то части от общего количества абонентов по мере возникновения необходимости. Для таких сетей характерна операция *коммутации (переключения) линий* по мере поступления требований от абонентов, чтобы обеспечить передачу сообщения (или его части-пакета) в пункт назначения.

2.2 Элементы преобразователей

В качестве напоминания перечислим те элементарные технические устройства, которые служат “кирпичиками” при создании преобразователей сигнала в тракте передачи.

Ячейка памяти (рис. 2.2а) предназначена для хранения значения одной *M*-ичной цифры.

Очень часто *M*=2, и тогда это может быть, например, обычный триггер.

Регистр памяти (*n*-разрядный) – это набор независимых ячеек памяти, предназначенный для хранения комбинации, состоящей из *n*-символов (рис. 2.2б).

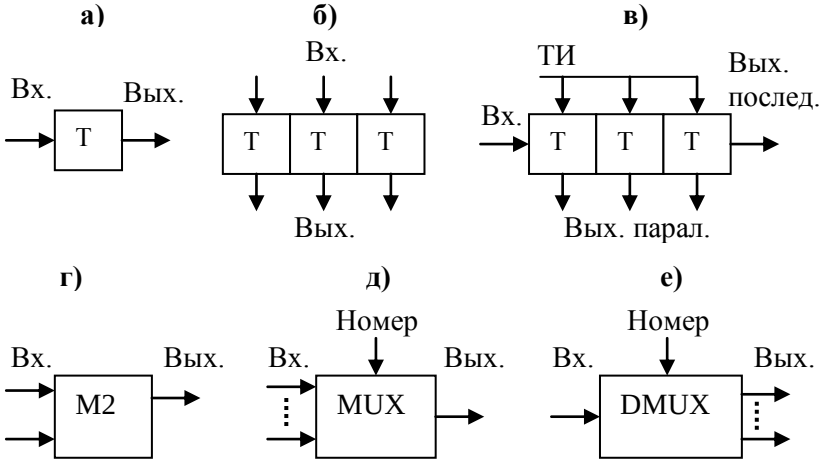

Рис. 2.2. Элементы преобразователей сигналов:
а) ячейка памяти; б) регистр памяти;
в) регистр сдвига; г) сумматор mod2;
д) мультиплексор; е) демультиплексор.

Регистр сдвига (*n*-разрядный) построен на основе *n*-разрядного реги-

стра памяти и предназначен для организации движения очереди символов. При одновременной подаче тактового импульса (ТИ) на все ячейки, содержимое каждой ячейки переносится в следующую, а в первую ячейку записывается очередной входной символ. Считывание содержимого регистра может проводиться как в параллельной форме, так и в последовательной (символ за символом).

Двоичный счетчик состоит из n триггеров, соединенных последовательно и работающих в счетном режиме. При подаче на вход двоичного числа (0 или 1), это число суммируется с n-разрядным двоичным числом, содержащимся в ячейках счетчика.

Если на вход счетчика подавать периодическую последовательность тактовых импульсов (единиц) с периодом Δt, то его состояния циклически повторяются, при этом длительность цикла равна $T=2^n\Delta t$.

Логические схемы И, ИЛИ, НЕ в комментариях не нуждаются.

Сумматор по модулю 2 (mod 2) имеет два входа и один выход (рис. 2.2г). При одновременной подаче на его входы двоичных символов u_1 и u_2 двоичный символ на выходе $v=u_1\oplus u_2$ определяется по правилам:

$$0\oplus 0=0,\quad 0\oplus 1=1,\quad 1\oplus 0=1,\quad 1\oplus 1=0. \qquad (2.6)$$

При таком подходе сложение и вычитание оказываются эквивалентными.

Мультиплексор (электронный переключатель) имеет $N=2^n$ входов и один выход (рис. 2.2д) и содержит n-разрядный регистр памяти, в котором в двоичной форме записан номер того входа, который в данный момент соединен с выходом. При обновлении содержимого регистра (при записи нового номера), происходит переключение на соответствующий вход. Вместо регистра памяти можно использовать n-разрядный двоичный счетчик и подавать на его вход периодическую последовательность тактовых импульсов. Тогда мультиплексор будет подключать входы в порядке возрастания их номеров, и этот процесс будет циклическим.

Демультиплексор (рис. 2.2е) имеет один вход и N выходов. Процесс переключения происходит точно так же, как в мультиплексоре.

2.3 Преобразование неэлектрических сигналов в электрические

Впервые такой преобразователь, пригодный для широкого использования, был предложен в 1837 году. Он содержал телеграфный ключ и таблицу азбуки Морзе и был предназначен для *преобразования текста* в последовательность импульсов электрического тока.

Более современный метод был принят в качестве стандарта в 1932 году для использования в телеграфии и известен как международный телеграфный код №2 (МТК-2). Взамен каждого знака передаваемого текста генерируется

кодовая комбинация, состоящая из пяти двоичных символов длительностью по 20 мс каждый, т.е. техническая скорость передачи составляет V_T=50 имп/с=50 Бод. Чтобы обеспечить взаимную синхронизацию передатчика и приемника, применяется старт-стопный метод передачи: в начале комбинации передается стартовый импульс длительностью 20 мс, а в конце – стоповый импульс длительностью 30 мс. В итоге для передачи одного знака требуется не менее 0,15 с. Этот стандарт давно устарел и в развитых странах не используется.

В 1968 году для телеграфии и для *передачи данных* (информации, передаваемой от компьютера к компьютеру) предложено использовать телефонные линии и 7-битовый код (КОИ-7, МТК-5, ASCII). В последующем был принят ряд стандартных скоростей передачи в диапазоне от 50 до 33600 бит/с.

Преобразователь *звукового сигнала* в электрический – это микрофон. После длительных экспериментов было найдено компромиссное решение, и *в качестве стандарта для телефонного сигнала (передача речи) установлена полоса частот 300 – 3400 Гц*. При передаче сигналов радиовещания по линии связи выделяется более широкая полоса частот (30 – 15000 Гц для систем высшего класса).

Преобразователем *неподвижного изображения* в аналоговый электрический сигнал (факсимильная связь, факс) является фотопреобразователь при использовании метода сканирования (строчной развертки). Скорость сканирования подобрана таким образом, чтобы сигнал на выходе факс-аппарата занимал полосу частот 300 – 3400 Гц и для его передачи можно было бы использовать существующие линии телефонной связи.

Преобразование *движущегося изображения* в электрический сигнал (телевидение) осуществляется тем же методом. В России принят стандарт 625 строк/кадр, 25 кадров/с, соотношение высоты к ширине кадра равно 3:4, поэтому полоса частот видеосигнала от 50 Гц до 6 МГц.

Важный вид электросвязи – это *телеметрия*, т.е. передача сигналов от датчиков, измеряющих текущие значения физических параметров некоторого объекта. Такие сигналы могут быть как аналоговыми, так и цифровыми, а значения их основных параметров могут отличаться на несколько порядков, в зависимости от назначения системы.

В системах *телеуправления* передаваемые сигналы – это команды, и диапазон возможных значений их параметров также чрезвычайно широк.

2.4 Квантование по времени непрерывного сигнала

Квантование (дискретизация) по времени – это замена непрерывной функции $u(t)$, заданной на интервале $0 \le t \le T$, последовательностью ее мгно-

венных значений-отсчетов $u_1=u(t_1)$, $u_2=u(t_2)$, … , $u_n=u(t_n)$. Они измеряются в моменты времени t_1, t_2, …, t_n, отстоящие один от другого на величину $\Delta t=t_j-t_{j-1}=T/n$, называемую *шагом квантования*.

Возможность точного восстановления исходной функции $u(t)$ по серии ее отсчетов доказана в *теореме отсчетов Найквиста–Котельникова*: любая функция, имеющая спектр, заключенный в ограниченной полосе частот 0...$F_в$, полностью определяется последовательностью отсчетов, взятых с шагом

$$\Delta t \le 1/(2F_в). \tag{2.7}$$

Другими словами, частота квантования $F_к=1/\Delta t$ должна удовлетворять условию $F_к \ge 2F_в$.

При доказательстве теоремы сформулирован и способ восстановления сигнала $u(t)$

$$u(t)=\sum_{j=1}^{n} u_j h(t-j\Delta t), \tag{2.8}$$

$$\text{где}\quad h(t)=\frac{\sin(\pi t/\Delta t)}{(\pi t/\Delta t)}. \tag{2.9}$$

Кстати, формула (2.8) является разложением Карунена-Лоэва (2.3) лишь в случае, когда сигнал $u(t)$ есть стационарный случайный процесс, спектр плотности мощности которого целиком заключен в полосе частот 0...$F_в$ и равномерен в ней.

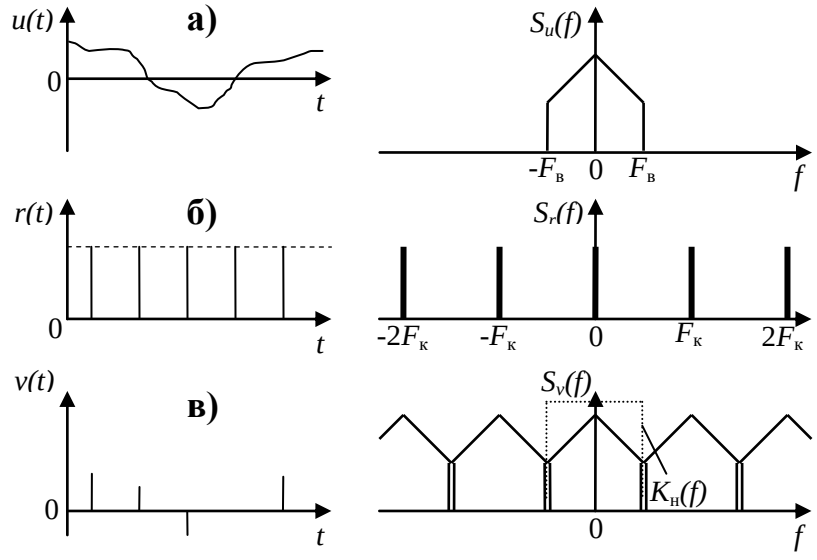

Рис. 2.3. Функции, участвующие в процедуре квантования во времени, и их спектры:
а) передаваемое непрерывное сообщение $u(t)$;
б) импульсная несущая $r(t)$;
в) последовательность импульсов-отсчетов $v(t)$

Рис. 2.3 иллюстрирует соотношение между спектрами функций, участвующих в процессе квантования во времени.

Поскольку серия коротких импульсов-отсчетов $v(t)$ получается в результате перемножения функций $u(t)$ и $r(t)$, спектр произведения $v(t)$ равен свертке функций $S_u(f)$ и $S_r(f)$, в итоге он, подобно спектру $S_r(f)$, также является периодической функцией и занимает бесконечно большую полосу.

При анализе рис.2.3в становится очевидным способ восстановления переданного сообщения $u(t)$: для этого серию отсчетов $v(t)$ нужно подать на вход идеального ФНЧ, имеющего частотную характеристику прямоугольной формы (пунктирная линия на рис. 2.3в)

$$K(f)=\begin{cases}1, & |f|\le F_{\text{с}},\\ 0, & |f|>F_{\text{с}},\end{cases} \tag{2.10}$$

где $F_{\text{с}} = F_{\text{в}} = F_{\text{к}}/2$ – частота среза ФНЧ.

Импульсная характеристика такого фильтра совпадает с функцией (2.9), тогда выражение (2.8), в сущности, описывает операцию свертки функций $v(t)$ и $h(t)$.

На практике точному восстановлению сообщения $u(t)$ мешает ряд обстоятельств. Укажем **основные причины возникновения ошибок при восстановлении сообщения $u(t)$.**

Сигнал $d(t)$ на выходе ФНЧ-демодулятора с частотной характеристикой $K(f)$ имеет спектр

$$S_d(f)=K(f)S_v(f), \tag{2.11}$$

и тогда сигнал ошибки $e(t)=u(t)-d(t)$ имеет спектр

$$S_e(f)=S_u(f)-S_d(f), \tag{2.12}$$

а нормированная величина среднего квадрата ошибки определяется выражением

$$\sigma^2=\frac{\sigma_e^2}{\sigma_u^2}=\frac{\int\limits_{-\infty}^{\infty}|e(t)|^2\,dt}{\int\limits_{-\infty}^{\infty}|u(t)|^2\,dt}=\frac{\int\limits_{-\infty}^{\infty}|S_e(f)|^2\,df}{\int\limits_{-\infty}^{\infty}|S_u(f)|^2\,df}. \tag{2.13}$$

Первая причина появления ошибок – это *вынужденное нарушение требований теоремы отсчетов, т.е. квантование с частотой* $F_{\text{к}}<2F_{\text{в}}$. Кстати, очевидно, что здесь, в отличие от идеального случая, возможен некоторый произвол при выборе частоты среза $f_{\text{с}}$ ФНЧ-демодулятора.

На рис.2.4. видно, что в этом случае происходит перекрытие отдельных частей спектра сигнала $v(t)$. Из-за этого возможно появление ошибок двух видов.

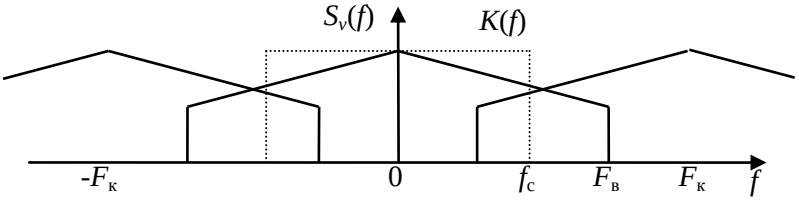

Рис. 2.4. Спектр серии отсчетов в случае $F_к<2F_В$ и частотная характеристика ФНЧ-демодулятора (пунктирная линия)

Ошибка первого рода – это пропадание высокочастотных составляющих сигнала, имеющих частоты $f_u>f_с$.

Ошибка второго рода – это трансформация составляющих спектра сигнала $u(t)$ с частотами $f_с<f_u<F_в$, и в результате на выходе ФНЧ – демодулятора появляются новые составляющие с частотами $f_d = F_к - f_u$ (вспомните, как на киноэкране ведут себя колеса экипажа, когда он набирает скорость).

Если выбрать $f_с = F_к - F_в$, то будет присутствовать лишь ошибка первого рода, а при $f_с = F_в$ – лишь ошибка второго рода. Для большинства реальных сигналов $u(t)$ амплитуды спектральных составляющих убывают с ростом частоты (как на рис. 2.3.), поэтому для минимизации среднего квадрата ошибки следует взять $f_с = F_к/2$, хотя при этом совместно существуют ошибки обоих видов. В этом последнем случае полезным окажется следующий прием: можно искусственно удовлетворить требованиям теоремы отсчетов, а именно, перед квантованием сообщение пропустить через такой же идеальный ФНЧ с частотой среза $f_с = F_к/2$. Из рис. 2.4. хорошо видно, что такая дополнительная фильтрация позволит уменьшить средний квадрат ошибки в два раза, т.к. она устраняет возможность появления ошибок второго рода.

Вторая причина появления ошибок – это *непрямоугольность частотных характеристик фильтров*, в частности ФНЧ-демодулятора (рис. 2.5). В результате, во-первых, происходят частотные искажения сигнала в области $|f|<F_в$, и, во-вторых, в спектре сигнала на выходе демодулятора появляются новые составляющие с частотами $|f|>F_в$.

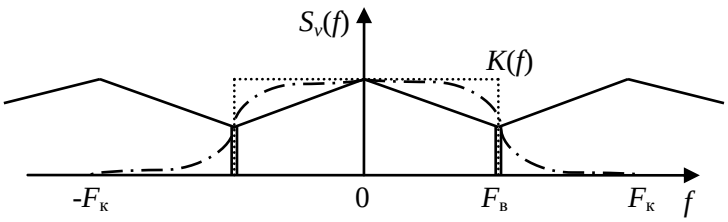

Рис. 2.5. Спектр серии отсчетов (сплошная линия), частотная характеристика ФНЧ-демодулятора в идеальном случае (пунктир) и частотная характеристика реального ФНЧ (штрих-пунктир)

Третья причина обусловлена тем, что, в соответствии с теорией преобразований Фурье, *любая функция конечной длительности во времени имеет бесконечно широкий спектр и наоборот*. Поэтому, даже если исходное сообщение $u(t)$ имело ограниченный спектр $0...F_B$, но из бесконечной серии его отсчетов, взятых с шагом $\Delta t = 1/2F_B$, по линии связи была передана конечная последовательность, состоящая из n отсчетов, при восстановлении отрезка сигнала $u(t)$, даже на интервале его передачи, возникнет ошибка.

Фактически, при передаче отрезка сигнала нарушены требования теоремы отсчетов. Рис. 2.4. иллюстрирует приемы точного анализа величины ошибки. Грубо величину ошибки можно определить по формуле

$$\sigma^2 \approx 1/n, \qquad (2.14)$$

которая для статистически стационарного сигнала при $n >> 1$ дает вполне удовлетворительную точность.

Очевидно, что в реальных условиях неизбежно совместное появление ошибок всех видов, поэтому частоту квантования $F_к$ выбирают с некоторым запасом. Например, стандартный аналоговый телефонный сигнал занимает полосу частот 300...3400 Гц, а *стандартная частота квантования для него установлена* $F_к = 8000$ *отсчетов/с.*

2.5 Модуляция импульсной несущей аналоговым сигналом

В результате квантования по времени непрерывная функция $u(t)$ заменяется последовательностью отсчетов. В итоге сообщение становится дискретным во времени, но каждый из отсчетов является реализацией непрерывной случайной величины.

Существуют три основных метода аналоговой передачи последовательности отсчетов.

Амплитудно-импульсная модуляция (АИМ): передаваемый сигнал есть периодическая последовательность прямоугольных импульсов, следующих с периодом Δt и имеющих одинаковую длительность τ, но амплитуда каждого импульса пропорциональна значению передаваемого сообщения (отсчета) в соответствующий момент времени (рис. 2.6).

При *широтно-импульсной модуляции* (ШИМ) амплитуда импульсов не меняется, но длительность каждого импульса пропорциональна значению соответствующего отсчета. На практике более удобна односторонняя ШИМ (ОШИМ), когда длительность каждого импульса изменяется за счет сдвига одного фронта, например заднего, а положение переднего фронта всегда совпадает с тактовым моментом времени (моментом взятия отсчета).

При *время-импульсной модуляции* (ВИМ) амплитуды и длительности импульсов постоянны, а задержка каждого импульса относительно тактового момента времени пропорциональна величине отсчета.

Разновидности АИМ, ШИМ и ВИМ, а также формы спектров таких сигналов детально рассматриваются, например, в [11]. Теоретически спектр каждого из сигналов бесконечно широкий, и приходится ограничивать полосу передаваемых частот интервалом $0...f_c$, где $f_c \approx (2...4)/\Delta t$. При уменьшении f_c увеличивается длительность фронтов импульсов, это затрудняет измерение информационного параметра (амплитуды, длительности или задержки импульса).

Универсальный метод демодуляции – это использование ФНЧ с часто-

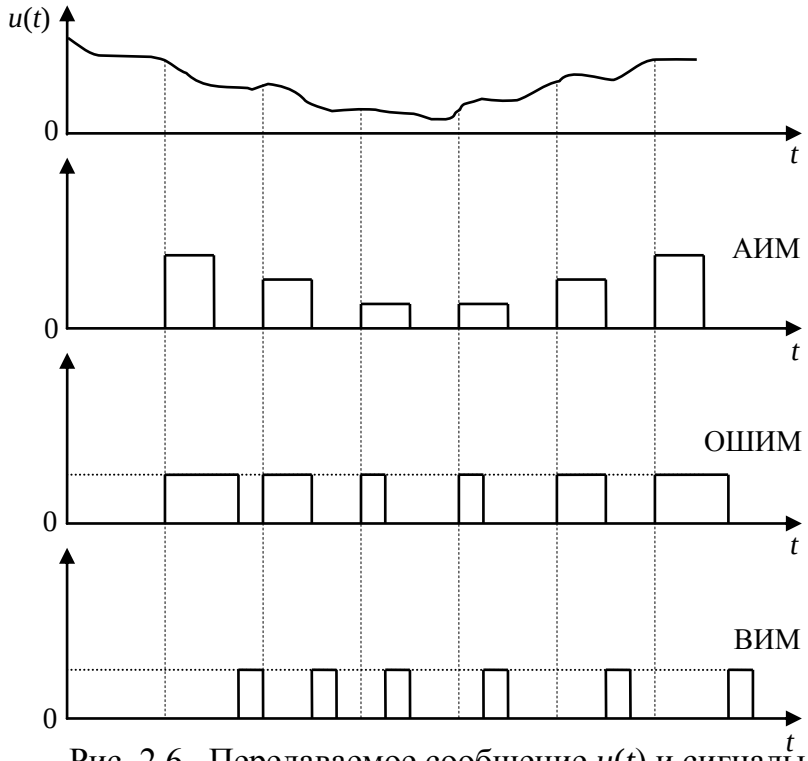

Рис. 2.6. Передаваемое сообщение $u(t)$ и сигналы с АИМ, ОШИМ, ВИМ

той среза $|f|>F_в$. Из энергетических соображений сигнал с ВИМ перед демодуляцией обычно предварительно преобразуют в сигнал с ОШИМ.

В настоящее время перечисленные методы передачи используются сравнительно редко. Предпочтение отдается цифровым методам, когда значение каждого отсчета представляется в виде комбинации из k двоичных символов.

2.6 Аналого-цифровое и цифроаналоговое преобразования

Цель аналого-цифрового преобразования (АЦП) – заменить непрерывную функцию u(t) последовательностью M-ичных символов.

АЦП осуществляется в два этапа. Первый этап – это квантование во времени непрерывного сигнала $u(t)$ (разд. 2.4). В результате получается последовательность импульсов-отсчетов, следующих с шагом Δt (рис. 2.3в).

Второй этап – это оцифровка каждого отсчета. Подготовка к этой операции заключается в следующем. Диапазон возможных значений напряжений (u_{min}, u_{max}) делится на M интервалов длиной

$$\Delta u = (u_{max} - u_{min})/M \qquad (2.15)$$

каждый. Величина Δu называется *шагом квантования по уровню* (рис. 2.7). Далее интервалы нумеруют M-ичными цифрами снизу вверх, начиная с цифры 0.

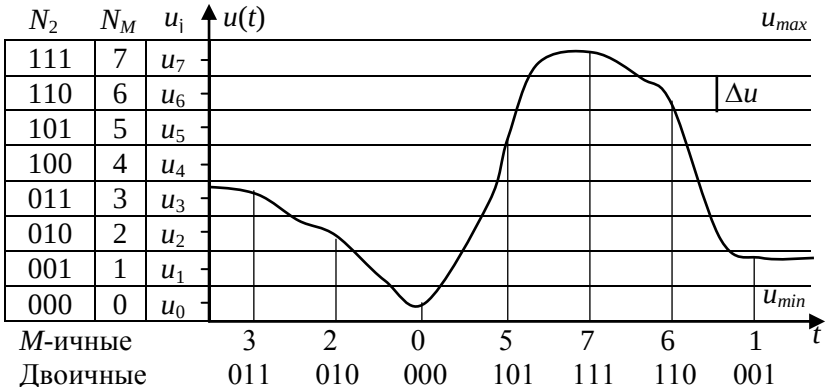

Рис. 2.7. Пример аналого-цифрового преобразования непрерывного сигнала $u(t)$ для M=8 (κ=3)

Для оцифровки очередной импульс-отсчет ”прикладываем” к шкале и заменяем его M-ичной цифрой N_M, равной номеру того интервала, в который попала вершина импульса. В итоге, вместо серии, состоящей из n отсчетов, передается последовательность M-ичных цифр длиной n.

Поскольку в технике наиболее часто для передачи используются двоичные символы, то значение M выбирают по формуле

$$M = 2^k, \qquad (2.16)$$

и номер очередного интервала передается в виде k-разрядной двоичной комбинации (N_2 на рис.2.7). В итоге количество передаваемых символов-цифр увеличивается в k раз.

Обратное, т.е. **цифроаналоговое преобразование** (ЦАП) также проводится в два этапа.

Первый этап – это формирование импульсов, соответствующих каждой M-ичной цифре (или k-разрядной комбинации при передаче двоичных символов). В качестве подготовки к этой операции внутри каждого из M интервалов выбирается точка. Обычно она соответствует центру интервала – это точки u_0, u_1, …, u_{M-1} на рис.2.7. Тогда амплитуда очередного импульса, соответствующего цифре j, устанавливается равной u_j. Таким образом, в итоге проводится округление значения каждого отсчета.

Ошибка такого округления $e = u - u_j$, где u – истинное значение отсчета. Если шаг квантования Δu достаточно мал (соответственно, число интервалов M велико), величина случайной ошибки e имеет почти равномерное распределение в интервале ширины Δu, поэтому среднеквадратическое значение ошибки квантования по уровню равно

$$\sigma_k = \Delta u/(2\sqrt{3}). \qquad (2.17)$$

Второй этап ЦАП – это преобразование серии импульсов-отсчетов в непрерывную функцию – проводится как обычно (разд. 2.4) при помощи ФНЧ с прямоугольной частотной характеристикой.

Такое квантование по уровню, когда все M интервалов имеют одинаковую длину (2.15), называется *равномерным* и является наиболее простым в техническом отношении. При фиксированном значении M эта процедура обеспечивает минимальное значение среднеквадратической ошибки квантования лишь для таких сигналов $u(t)$, у которых величина напряжения равномерно распределена в интервале (u_{min}, u_{max}). Реальные сигналы, такие как звуковой, характеризуются распределением вероятностей, имеющим узкий центральный пик и длинные “хвосты” (рис.2.8).

Для таких сигналов предпочтительней неравномерное квантование по уровню, когда длины интервалов берут меньшими в области наиболее вероятных значений напряжения и увеличивают их в области малых вероятностей.

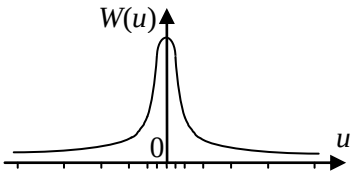

Рис. 2.8. Распределение вероятностей уровня сигнала и шкала неравномерного квантования

Тот же эффект достигается более простыми техническими средствами. Сначала сигнал $u(t)$ подвергают *компрессии по уровню*, т.е. пропускают через устройство, амплитудная характеристика которого линейна лишь в области малых значений входного напряжения, а при других значениях описывается логарифмической кривой (рис.2.9). В итоге распределение напряжения сигнала на выходе компрессора становится существенно ближе к равномерному.

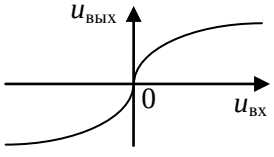

Рис. 2.9. Амплитудная характеристика компрессора

После компрессии проводят стандартную процедуру АЦП с равномерным квантованием. В пункте приема, после стандартного ЦАП следует провести обратное нелинейное преобразование, хотя, как показывает практика, нередко можно обойтись и без него.

Для проведения АЦП телефонного сигнала используют стандартные значения параметров: $F_к$=8000 отсчетов/с, k=8 бит/отсчет, т.е. M=256. В итоге получается стандартный цифровой телефонный сигнал, который характеризуется скоростью цифрового потока на выходе АЦП, равной V=8000×8=64 кбит/с [2,6].

Серия двоичных импульсов, получаемая в результате такого преобразования телефонного сигнала, называется сигналом с *импульсно-кодовой модуляцией* (ИКМ).

При записи музыкальных произведений на компакт-диск значения этих параметров равны: $F_к$=44100 отсчетов/с, k=16 бит/отсчет, т.е. M=65536, V=705,6 кбит/с.

2.7 Линейная цифровая фильтрация и генерирование последовательностей символов

На вход цифрового фильтра подается сколь угодно длинная последовательность чисел …, u_j, u_{j+1},…, а на выходе формируется последовательность чисел …, v_j, v_{j+1},…, синхронная во времени с входной последовательностью. В линейном фильтре очередной выходной символ v_j получается в результате линейной операции, проводимой с входными символами $...u_{j-1}$, u_j, поступившими к данному моменту.

Для описания последовательностей чисел удобно использовать полиномы

$$\begin{aligned} &...,u_j,u_{j+1},u_{j+2}... \to \\ &\to u(x)=...+u_j\cdot x^j+u_{j+1}\cdot x^{j+1}+u_{j+2}\cdot x^{j+2}+...; \\ &...,v_j,v_{j+1},v_{j+2},... \to \\ &\to v(x)=...+v_j\cdot x^j+v_{j+1}\cdot x^{j+1}+v_{j+2}\cdot x^{j+2}+... \end{aligned} \qquad (2.18)$$

В математике то же самое представление применяют в виде z-преобразования, которое получается из (2.18) путем формальной замены $x = z^{-1}$.

Например, пусть начало последовательности на входе имеет следующий вид: u_0, u_1, u_2…=3, 5, 7, 1, 0, 1, …, тогда ей соответствует полином $u(x) = 3 + 5x + 7x^2 + x^3 + x^5 + ...$

Трансверсальный фильтр r-го порядка (фильтр с конечной импульсной характеристикой) содержит r-разрядный регистр сдвига, управляемый последовательностью тактовых импульсов (ТИ), синхронных во времени с моментами поступления очередных входных символов. Ради наглядности на

схеме рис. 2.10 обозначения входных и выходных символов выстроены в порядке очереди, движущейся слева направо. Значения сомножителей g_0, $g_1,\ldots,g_r$, участвующих в операциях умножения, – это константы, задаваемые заранее, а сумматор ($\sum$) проводит обычное арифметическое суммирование.

Из схемы видно, что очередной символ выходной последовательности вычисляется по формуле

$$v_j = \sum_{k=0}^{r} u_{j-k} g_k, \tag{2.19}$$

которая в дискретной форме задает операцию свертки отсчетов входного сигнала и отсчетов g_0, $g_1,\ldots,g_r$ импульсной характеристики этого фильтра.

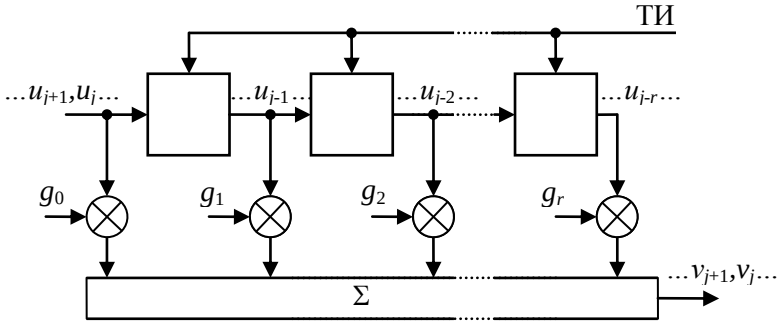

Рис. 2.10. Трансверсальный цифровой фильтр

Ради удобства все эти отсчеты обычно нормируют, т.е. умножают на постоянное число так, чтобы было g_0=1.

Легко проверить, что соотношение между полиномами определяется формулой

$$v(x) = u(x)g(x), \tag{2.20}$$

где полином $g(x)$, коэффициенты которого равны отсчетам импульсной характеристики фильтра, называется его *передаточной функцией*. Пусть последовательности чисел …, u_j, $u_{j+1},\ldots$ и …, v_j, $v_{j+1},\ldots$ – это отсчеты двух непрерывных сигналов $U(t)$ и $V(t)$ с ограниченным спектром, полученные с шагом Δt, который удовлетворяет требованиям теоремы отсчетов. Тогда частотную характеристику $K(\omega)$ непрерывного линейного фильтра, который преобразовывал бы сигнал $U(t)$ в сигнал $V(t)$, можно получить из передаточной функции эквивалентного цифрового фильтра путем замены $x = \exp(i\omega\Delta t)$.

Если число отсчетов сигнала на входе $u_0,\ldots, u_r$ конечно, то можно построить цифровой *фильтр, согласованный с сигналом* $U(t)$. Его импульсная характеристика является зеркальным отображением ряда отсчетов сигнала $g_0 = u_r, \ldots, g_r = u_0$, и числовая последовательность на выходе фильтра совпадает с отсчетами автокорреляционной функции сигнала $U(t)$. Если же на вход фильтра поступает другой сигнал $X(t)$, то на выходе имеем отсчеты взаимнокорреляционной функции этих сигналов $U(t)$ и $X(t)$.

Рекурсивный фильтр r-го порядка (фильтр с бесконечной импульсной характеристикой) отличается тем, что содержит цепи обратной связи.

Наиболее наглядно схему такого фильтра можно определить следующим образом. Потребуем, чтобы этот фильтр выполнял операцию, обратную операции (2.19), тогда каждый символ на выходе v_j можно определять, решив систему уравнений вида

$$u_j = \sum_{k=0}^{r} v_{j-k} g_k = v_j g_0 + v_{j-1} g_1 + \ldots + v_{j-r} g_r. \qquad (2.21)$$

Отсюда, полагая, что g_0=1, получаем

$$v_j = u_j - v_{j-1} g_1 - v_{j-2} g_2 - \ldots - v_{j-r} g_r. \qquad (2.22)$$

Видно, что фильтр, схема которого изображена на рис.2.11, выполняет именно такую операцию. Очевидно, что для полиномов справедливо соотношение

$$v(x) = u(x)/g(x), \qquad (2.23)$$

истинность которого при желании можно еще раз проверить, проводя деление полиномов столбиком.

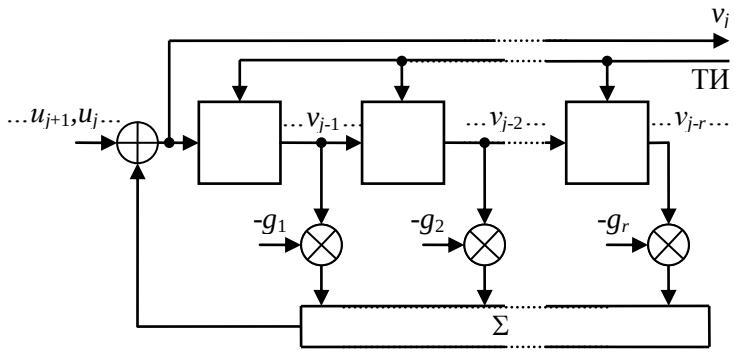

Рис. 2.11. Рекурсивный цифровой фильтр

Итак, схема рисунка 2.10 соответствует операции *умножения* входного полинома на полином $g(x)$, а схема рис. 2.11 – операции *деления* на $g(x)$. Следовательно, последовательность $g_0, g_1, \ldots, g_r$ ***не является*** отсчетами импульсной характеристики рекурсивного фильтра рис. 2.11. Чтобы найти эти отсчеты $h_0, h_1, \ldots$, нужно вычислить соответствующий полином, проведя операцию деления

$$h(x) = 1/g(x). \qquad (2.24)$$

Очевидно, что эта последовательность будет иметь бесконечную длину, что и определяет название фильтра. Тогда вместо (2.23) можно использовать соотношение, подобное формуле (2.20)

$$v(x) = u(x)h(x). \qquad (2.25)$$

Отметим особенности построения и основные свойства *двоичного цифрового фильтра*, в котором все последовательности u_j, v_j, g_j состоят из двоичных цифр (нулей или единиц). Во-первых, отпадает необходимость использования специальных элементов-перемножителей двух сигналов, по-

скольку умножение сигнала на 1 или 0 соответствует включению или выключению этого сигнала (есть проводник для подачи этого сигнала или его нет). Во-вторых, все операции суммирования двоичных чисел выполняются по модулю два.

И, наконец, имеется еще одно важное применение рекурсивного цифрового фильтра, то есть, регистра сдвига с обратными связями. Если вход фильтра отключить от источника сигнала (либо, то же самое, продолжить подавать серию нулей), то при поступлении очередного тактового импульса на выходе схемы будет появляться очередное число v_j, то есть, на выходе фильтра будем иметь бесконечную числовую последовательность. Ее вид зависит от исходного состояния регистра. Если во всех ячейках были нули, то очевидно, что выходная последовательность никакого интереса не представляет, поскольку будет состоять из одних нулей. В противном случае на выходе получим ненулевую периодическую последовательность, то есть регистр сдвига с обратными связями превратился в **генератор числовой последовательности.**

Наибольший интерес, благодаря своим замечательным свойствам, вызывают *двоичные псевдослучайные последовательности (ПСП) максимальной длины (М-последовательности).* Первое название обусловлено тем, что при заданном количестве r ячеек в регистре сдвига (рис. 2.11) такая последовательность имеет максимально-возможный период (количество тактов, через которое вид последовательности в точности повторяется), и он равен

$$M = 2^r - 1. \qquad (2.26)$$

Такие последовательности могут быть получены лишь в схемах, построенных на основе полиномов $g(x)$ определенного вида. Примеры таких полиномов для $r = n - k \le 11$ даны в Приложении 1, а о некоторых их свойствах будет дополнительно сказано в разд. 3.5.

Итак, укажем основные свойства ПСП.

1) Периодичность (2.26) .

2) В любом отрезке последовательности, длина которого равна периоду, количество нулей и единиц почти одинаково (нулей на один меньше).

3) В любом отрезке последовательности, длина которого равна M+1, среди M пар типа 00, 01, 10, 11, образованных из рядом стоящих символов, каждый тип встречается почти одинаково часто (пара 00 встречается на один раз меньше, чем любая из трех остальных).

4) Такое же равенство частот наблюдается для всевозможных троек, четверок и т.д. соседних символов вплоть до r-разрядных комбинаций.

5) Если из бесконечной ПСП взять два M-разрядных отрезка, сдвинутых на произвольное число тактов, не кратное периоду M, то эти две конечные последовательности оказываются почти ортогональными: при попарном сравнении символов, стоящих на одинаковых позициях, количество совпадающих пар на одну меньше чем количество несовпадающих.

6) Спектр плотности мощности биполярного двоичного сигнала, переносящего значения элементов ПСП (0 – прямоугольный импульс положительной полярности, 1 – импульс отрицательной полярности той же длительности τ), имеет почти такую же форму, как спектр плотности мощности одиночного импульса. Автокорреляционная функция отрезка ПСП длиной в период M, совпадающая по форме с сигналом на выходе согласованного филь-

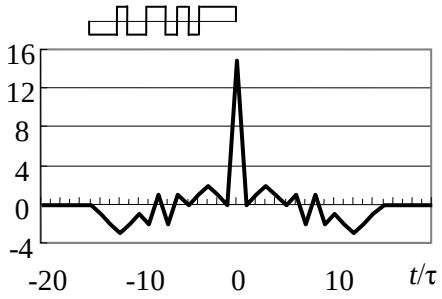

Рис. 2.12. Сигнал на выходе фильтра, согласованного с импульсным двоичным входным сигналом (сверху), заданным последовательностью вида:
–1, –1, –1, 1, –1, –1, 1, 1, –1, 1, –1, 1, 1, 1, 1

тра (рис. 2.12), имеет узкий центральный пик и малые боковые лепестки, то есть, имеет почти такой же вид, как автокорреляционная функция последовательности случайных независимых импульсов. Итак, *ПСП обладает почти такими же свойствами, как длинная истинно случайная последовательность*, полученная, например, в результате многократного подбрасывания монеты. Обратите внимание на регулярное использование слова “почти”, ограничительный смысл которого тем меньше, чем больше период M. В то же время ПСП не является случайной, так как, зная схему генератора (производящий полином $g(x)$) и исходную двоичную комбинацию в регистре, мы можем предсказать ее значения на сколь угодно тактов вперед.

Такая последовательность детерминирована, но для постороннего наблюдателя обладает свойствами случайной последовательности, поэтому она нашла многочисленные применения, например, при кодовом разделении каналов (разд. 6.4).

2.8 Модуляция гармонической несущей цифровым сигналом

Простейший способ передачи в двоичном канале – это применение однополярного двухуровневого кодирования без возврата к нулю, то есть без пауз между соседними символами, когда символу 1 соответствует прямоугольный импульс длительности τ, а символу 0 – пауза той же длительности ($u(t)$ на рис. 2.13). Этот способ применяется во внутри- и межблочных соединениях в радиоэлектронной аппаратуре, а также при передаче сигналов по

медному кабелю на небольшие расстояния.

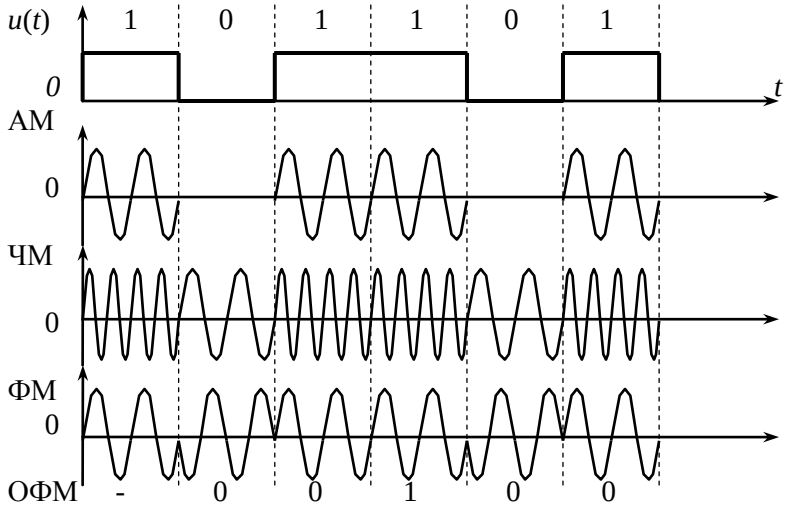

Рис.2.13. Методы модуляции гармонической несущей последовательностью двоичных символов

При использовании импульсов прямоугольной формы для передачи символов сигналы, соответствующие символам с разными порядковыми номерами, оказываются взаимно *ортогональными*. Они не перекрываются во времени, поэтому интеграл (1.44) для них равен нулю. Благодаря этому возможно разделение таких сигналов *без взаимных помех* при помощи простых линейных устройств, таких как демультиплексор или корреляционный приемник (разд. 2.9).

К сожалению, в полной мере воспользоваться этим свойством не удается. Спектр такого сигнала случаен, поскольку случайна сама передаваемая последовательность символов, но среднее значение квадрата модуля спектра по форме повторяет квадрат модуля спектра одиночного импульса (рис. 2.14). Спектр сигнала бесконечно широк, и естественно, что при передаче

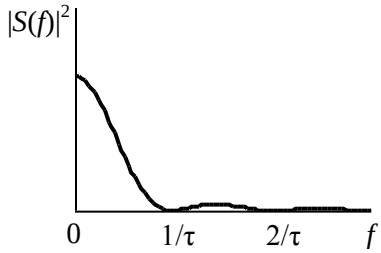

Рис. 2.14. Спектр плотности мощности последовательности двоичных импульсов

любого из сигналов приходится ограничивать занимаемую им бесконечно большую полосу частот.

Обычно сохраняют лишь полосу $(0,\ldots,F_в)$, занимаемую главным лепестком спектра, или даже несколько меньше,

$$\text{где} \quad F_в = (0{,}6\ldots1{,}0)/\tau = (0{,}6\ldots1{,}0)V_т, \tag{2.27}$$

$V_т$ – техническая скорость передачи, импульсов/с.

Напомним, что при сужении передаваемой полосы частот в области верхних частот растягиваются фронты радиоимпульсов, поэтому происходит наложение соседних импульсов друг на друга и их взаимное искажение. В итоге снижается надежность распознавания, какое именно значение двоичного символа (0 или 1) было передано с данным импульсом. Такое явление называется *межсимвольной интерференцией*.

Тем не менее, существует способ передачи в ограниченной полосе без межсимвольной интерференции. Вместо прямоугольных импульсов нужно использовать импульсы в виде функции Котельникова

$$u(t) = \frac{\sin(\pi t/\tau)}{\pi t/\tau}. \tag{2.28}$$

Пример такого сигнала при передаче последовательности двоичных символов приведен на рис. 2.15.

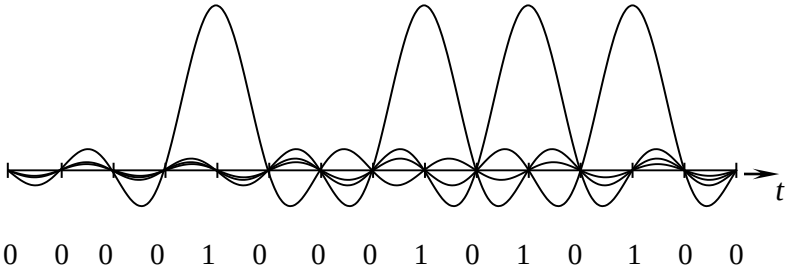

Рис. 2.15. Слагаемые, формирующие сигнал на входе линии при передаче последовательности двоичных символов импульсами в виде функции Котельникова

Импульсы перекрываются во времени, тем не менее, они взаимно ортогональны. Фильтр, согласованный с таким сигналом – это идеальный ФНЧ с частотой среза $F_в = 1/(2\tau)$, поэтому после фильтрации форма полезного сигнала не меняется, уменьшается лишь мощность аддитивного белого шума. Для вынесения решения о значении очередного символа нужно просто использовать отсчет сигнала в соответствующий тактовый момент времени. Заметьте, что значения всех остальных импульсов в этот момент равны нулю, то есть межсимвольная интерференция отсутствует. Итак, техническая скорость передачи $V_т = 2F_в$.

Доказано, что не существует других способов, которые при условии

отсутствия межсимвольной интерференции позволили бы достигнуть большего значения $V_т$ в той же ограниченной полосе частот [1,2]. Это же утверждение справедливо и в том случае, когда передача осуществляется с применением одного из методов модуляции гармонической несущей (радиоимпульсов, передаваемых в полосе частот Δf). Для достижения максимального значения

$$V_т = 1/\tau = \Delta f \qquad (2.29)$$

огибающая радиоимпульса должна иметь вид функции Котельникова. Использование импульсов такого вида все-таки вызывает технические трудности.

Традиционный метод *модуляции (манипуляции)* гармонической несущей цифровым сигналом состоит в том, что каждый из M-ичных символов, поступающих на вход модулятора, превращается в отрезок синусоидального колебания (прямоугольный радиоимпульс) длительностью τ.

Методы модуляции двоичным сигналом (M=2) – амплитудная, частотная и фазовая (рис.2.13).

Для получения сигнала с *амплитудной модуляцией* (АМ) достаточно иметь генератор несущего колебания с частотой $f_о$ и переключатель (мультиплексор с двумя входами), управляемый входной последовательностью двоичных символов (рис.2.16).

Частотный модулятор (ЧМ) содержит два генератора различных частот и переключатель. Для реализации *фазовой модуляции* (ФМ) используются сигналы, снимаемые с прямого и инверсного выходов одного генератора, т.е. начальная фаза выходного колебания задается правилом

$$\varphi = \begin{cases} 0^\circ, & S = 1, \\ 180^\circ, & S = 0, \end{cases} \qquad (2.30)$$

где S – значение передаваемого символа.

Для всех трех методов типична операция переключения (выбора) одного из двух сигналов, поэтому часто вместо термина ”модуляция” применяется термин “манипуляция”.

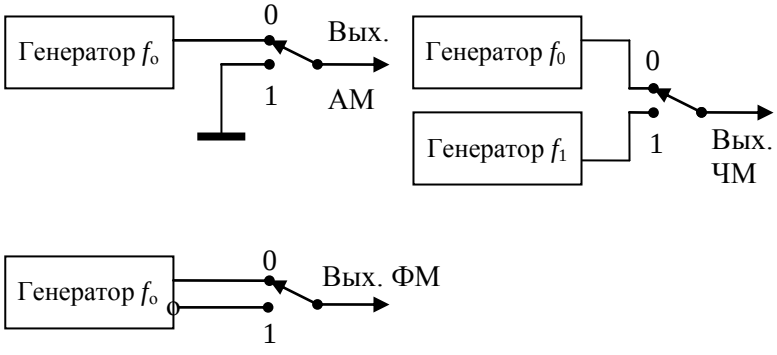

Рис. 2.16. Схемы, иллюстрирующие принципы АМ, ЧМ и ФМ при передаче двоичных символов

На практике, чтобы избежать использования генераторов, обладающих чрезвычайно высокой стабильностью частоты, вместо ФМ применяется *относительная фазовая манипуляция* (ОФМ), при которой начальная фаза текущего радиоимпульса φ_j определяется по отношению к начальной фазе предыдущего импульса φ_{j-1}, то есть

$$\varphi_j - \varphi_{j-1} = \begin{cases} 0^\circ, & S = 1, \\ 180^\circ, & S = 0. \end{cases} \qquad (2.31)$$

Внешне сигналы ФМ и ОФМ не отличаются, обладают одинаковым спектром плотности мощности, но при демодуляции, естественно, дают различные последовательности символов (рис. 2.13).

Спектр любого из модулированных сигналов – случайная функция, поскольку модулирующий сигнал случаен. Сигнал АМ – это сумма прямоугольных радиоимпульсов одинаковой длительности τ со случайными независимыми амплитудами, поэтому спектр плотности мощности АМ сигнала совпадает по форме со спектром плотности мощности одиночного радиоимпульса и содержит, кроме того, гармонику с частотой несущей f_o (рис. 2.17)

$$G(f) = \delta(f - f_0) + \left\{ \frac{\sin[\pi\tau(f - f_0)]}{\pi\tau(f - f_0)} \right\}^2. \qquad (2.32)$$

Аналогичную форму имеют спектры плотности мощности ФМ и ОФМ сигналов, но гармоники с частотой несущей они не содержат.

Сигнал ЧМ формально можно представить как сумму двух АМ сигналов, модулированных противоположными двоичными последовательностями и имеющих разные несущие частоты f_0 и f_1, поэтому его спектр плотности мощности можно приближенно представить, как сумму спектров плотности мощности этих АМ сигналов. Минимальная величина разнесения частот f_0 и f_1, при которой сигналы становятся ортогональными, равна

$$|f_o - f_1| = 1/\tau. \qquad (2.33)$$

Именно этот случай представлен на рис. 2.17в.

Спектр любого из модулированных сигналов также бесконечно широк, и приходится ограничивать занимаемую им полосу частот.

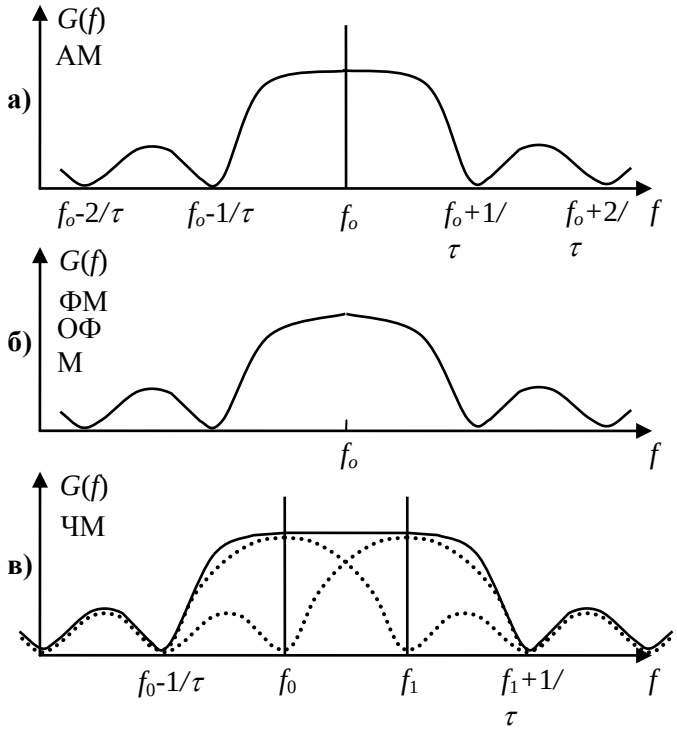

Рис. 2.17. Спектры плотности мощностиАМ, ФМ, ОФМ и ЧМ сигналов

Если воспользоваться критерием (2.34) и учесть, что спектр модулированного сигнала имеет две боковые полосы, то

$$\Delta f = (1,2...2,0)/\tau = (1,2...2,0)V_{\text{ò}}. \qquad (2.34)$$

Для сигнала ЧМ при таком подходе требуется полоса частот в 1,5…2 раза шире.

При осуществлении **многопозиционных методов манипуляции** число M возможных сигналов выбирается из знакомого соотношения $M = 2^k$, где k – целое число.

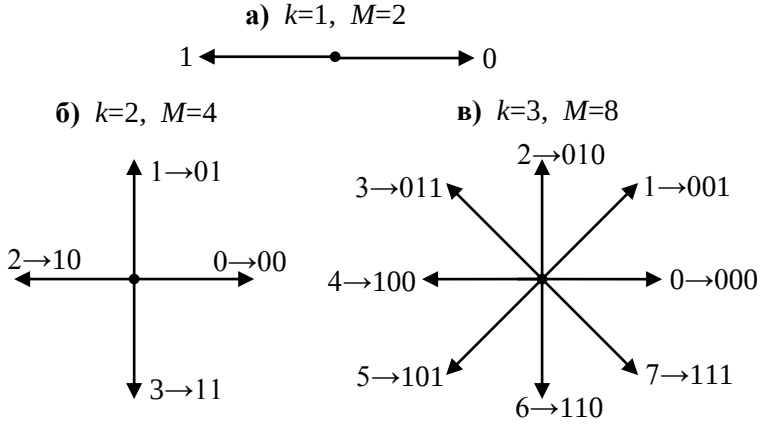

Рис. 2.18. Векторное представление колебаний при многократной ФМ

Возможные сигналы при многократной *(k-кратной) ФМ* удобно изображать в виде векторов на плоскости с декартовыми координатами *C* и *S* (рис. 2.15), отображающими квадратурные колебания (разд. 1.6).

Против каждого вектора указана *M*-ичная цифра и *k*-разрядная двоичная комбинация, которой этот вектор соответствует. Классическая ФМ (рис. 2.13) может быть названа однократной, и, как уже отмечалось, двум различным значениям *M*-ичной цифры соответствуют колебания с начальной фазой 0^o и 180^o. При двукратной ФМ возможна передача сигнала с одним из четырех значений начальной фазы (0^o, 90^o, 180^o, 270^o), т.е. при помощи одного радиоимпульса можно сразу передать значения двух битов, и т.д.

M-позиционная квадратурно-амплитудная манипуляция (КАМ-*M*) предусматривает совместное изменение амплитуд и фаз передаваемых импульсов-символов. На рис. 2.19 каждый из возможных сигналов отображает-

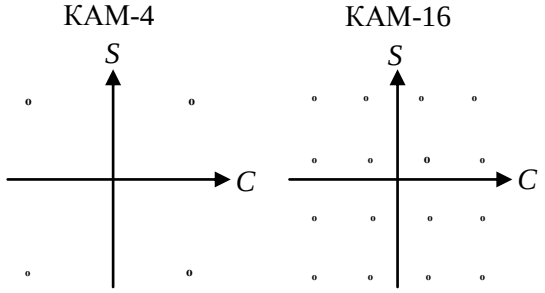

Рис. 2.19. Геометрические образы *M* позиционных сигналов с квадратурно-амплитудной модуляцией

ся не вектором, а точкой, соответствующей концу этого вектора. Кстати, сетка, в узлах которой расположены эти точки, не обязательно должна иметь квадратные ячейки; иногда, особенно если *k* – большое и нечетное число, предпочтительнее использование треугольной или радиально-круговой сетки.

Легко заметить, что сигнал с КАМ-4 модулирован только по фазе и отличается от сигнала с двукратной ФМ лишь сдвигом фазы всех возможных сигналов на 45^o.

Один импульс на выходе модулятора КАМ-16 способен перенести значение 16-ичной цифры, либо сразу значения четырех битов, соответствующих этой цифре.

Можно догадаться, что и *все многопозиционные методы манипуляции обычно применяются в относительной форме*, т.е. фаза очередного импульса отсчитывается не по отношению к фазе местного опорного колебания, а по отношению к фазе предыдущего импульса.

Наиболее привлекательное достоинство многопозиционных методов – это возможность, не изменяя длительности импульсов, и, следовательно, занимаемой полосы частот, в *k* раз увеличить битовую скорость передачи.

С другой стороны, если среднее значение энергии всех возможных сигналов неизменно, при увеличении M уменьшается расстояние между сигналами, и, следовательно, возрастает вероятность ошибки при различении этих сигналов на фоне помех. *Поэтому применение многопозиционных методов манипуляции возможно лишь при передаче сигналов по линии, вносящей малые искажения и обеспечивающей высокое отношение сигнал/шум.*

Наиболее широко применяемые методы в настоящее время – это ОФМ, двукратная и трехкратная ОФМ (радиорелейные и спутниковые линии, модемы), КАМ-16, …, КАМ-256 (модемы).

При многопозиционной АМ нерационально используется мощность передатчика, при ЧМ – полоса частот, поэтому такие методы практически не применяются.

В заключение отметим, что не всегда в качестве несущей используются гармонические колебания. Можно взять периодическую последовательность чередующихся положительных и отрицательных импульсов длительностью $\tau/2$ каждый (такая последовательность называется меандром) и модулировать ее последовательностью двоичных символов по закону ФМ. Такой способ представления двоичного символа в виде пары одинаковых разнополярных импульсов называется манчестерским кодом.

2.9 Корреляционный прием и согласованная фильтрация

Корреляционный приемник (синхронный детектор) – это принципиальный элемент демодулятора любого из цифровых сигналов, описанных в разд. 2.8. Он преобразует непрерывный сигнал $u(t)$, например радиоимпульс, поступающий на его вход, в число v в соответствии с правилом

$$v = \int_{t_н}^{t_к} u(t)w(t)dt, \qquad (2.35)$$

где $t_н$ и $t_к$ – моменты начала и окончания полезного сигнала на входе,

$w(t)$ – весовая (опорная) функция.

Сравнивая (2.35) и (1.44), видим, что величина v, вычисляемая корреляционным приемником, равна скалярному произведению векторов **u** и **w**, отображающих сигналы $u(t)$ и $w(t)$, т.е. пропорциональна длине проекции вектора **u** на направление, задаваемое вектором **w**.

Обычно функция $u(t)$ на входе приемника – это сумма полезного сигнала (форма которого $u_c(t)$, за исключением множителя a, известна) и случайного шума $u_ш(t)$

$$u(t) = a \cdot u_c(t) + u_ш(t), \qquad (2.36)$$

поэтому и выходная величина v является суммой полезной v_c и шумовой $v_ш$ составляющих

$$v = v_c + v_ш = a\int_{t_н}^{t_к} u_c(t)w(t)dt + \int_{t_н}^{t_к} u_ш(t)w(t)dt. \qquad (2.37)$$

Возвращаясь к геометрическим образам сигналов, отметим, что величина скалярного произведения ($\mathbf{u_c,w}$) максимальна, если $\mathbf{w} = \mathbf{u_c}$, т.е. весовая функция должна совпадать по форме с ожидаемым полезным сигналом

$$w(t) = \begin{cases} u_c(t), & t_н \le t \le t_к, \\ 0, & \text{вне этого интервала.} \end{cases} \qquad (2.38)$$

При технической реализации корреляционного приемника в качестве перемножителя функций $u(t)$ и $w(t)$ применяется фазовый детектор (ФД), а в качестве интегратора – интегрирующая RC-цепь (в этом случае следует лишь предусмотреть возможность обнуления выходного напряжения в момент $t_к$).

Линейный *фильтр* с постоянными параметрами, *согласованный с сигналом* $u_c(t)$, имеет импульсную характеристику

$$g(\tau) = u_c(t_к - \tau), \qquad (2.39)$$

которая является зеркальным отображением этого сигнала относительно точки $t = t_к$. Сигнал $v(t)$ на выходе такого фильтра определяется интегралом свертки

$$v(t) = \int_{-\infty}^{t} u(t_1)g(t - t_1)dt_1 = \int_{-\infty}^{t} u(t_1)u_c(t_к - t + t_1)dt_1. \qquad (2.40)$$

Если взять отсчет этого напряжения в момент $t=t_к$ и учесть (2.38), то получим

$$v(t_к) = \int_{t_н}^{t_к} u(t_1)w(t_1)dt_1. \qquad (2.41)$$

Таким образом, пара “согласованный фильтр – отсчетное устройство” выполняет те же операции, что и корреляционный приемник. Кстати, сейчас во многих случаях намного проще оказывается провести квантование сигналов по времени и выполнять операции корреляционного приема или согласованной фильтрации в цифровом виде (разд. 2.7).

2.10 Модуляция гармонической несущей непрерывным сигналом

Сами методы модуляции хорошо известны – это АМ и ЧМ. Поэтому наша задача – лишь уточнить, какова та минимальная полоса частот, которая необходима для передачи модулированного сигнала.

Спектр сигнала АМ состоит из несущей f_o и двух боковых полос, которые по форме повторяют спектр передаваемого сообщения $u(t)$ (рис. 2.17б).

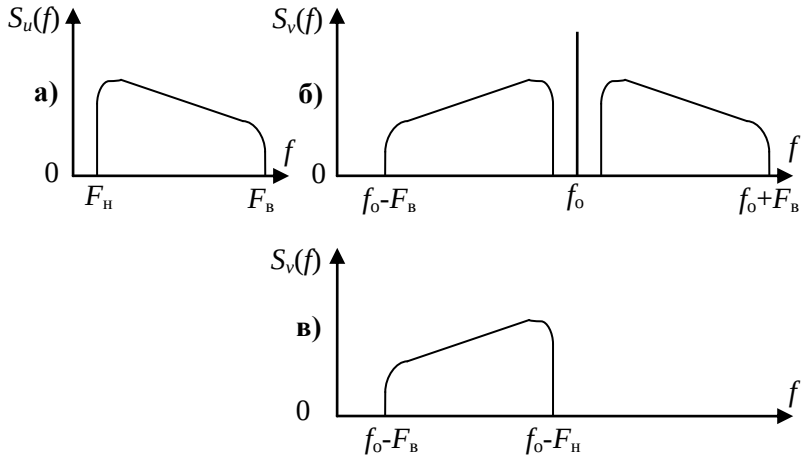

Рис. 2.20. Формы спектров: передаваемогосообщения (а),
АМ сигнала (б) и АМ ОБП сигнала (в)

Ширина спектра равна $\Delta f_{АМ} = 2F_в$. В частности, для передачи стандартного телефонного сигнала ($F_н$=300 Гц, $F_в$=3400 Гц) требуется полоса частот шириной 6,8 кГц.

Несущая не содержит информации о переданном сообщении, да и вторая боковая полоса является зеркальным отображением первой. Следовательно, более экономичным является способ *амплитудной модуляции с одной боковой полосой (АМ ОБП)* – обычно оставляют нижнюю боковую полосу (рис. 2.20в). Здесь для передачи требуется полоса частот $\Delta f_{ОБП} = F_в - F_н$, что для телефонного сигнала составляет всего 3,1 кГц.

Если для демодуляции АМ сигнала используют простой амплитудный детектор (выпрямитель плюс ФНЧ), то для демодуляции сигнала АМ ОБП необходимо применить фазовый детектор (ФД). На опорный вход ФД подают гармоническое колебание с частотой $f_г$ от местного генератора. В идеале должно быть $f_г = f_о$. Если же эти частоты различны, при демодуляции возникают искажения сообщения: сдвиг всех частот вверх, если $f_г > f_о$, и вниз, если $f_г < f_о$. Для телефонии принята допустимая величина сдвига $|f_г - f_о| < 10$ Гц.

Сигнал с частотной модуляцией (ЧМ) можно представить в виде

$$v(t) = A\cos\left\{\left[\omega_o + \omega_д \frac{u(t)}{\sqrt{P_u}}\right]t\right\}, \qquad (2.42)$$

где A – амплитуда, $f_о = \omega_о/2\pi$ – несущая частота,

$f_д = \omega_д/2\pi$ – девиация частоты,

P_u – средняя мощность сигнала $u(t)$.

Определение формы спектра ЧМ-сигнала – трудная задача. Если девиация частоты $f_д$ существенно меньше, чем верхняя граничная частота $F_в$ передаваемого сообщения $u(t)$, то полоса частот, занимаемая сигналом, при-

мерно равна $2F_в$, как при обычной АМ. Если же $f_д > F_в$, сигнал с ЧМ становится широкополосным, и для его удовлетворительной передачи требуется полоса частот [7]

$$\Delta f_{ЧМ} \approx (2...4)\, f_д. \qquad (2.43)$$

Для многоканальной передачи телефонных сигналов по медному кабелю применяется АМ ОБП, а в радиосвязи широкое применение нашла ЧМ с полосой частот передаваемого сигнала 10..30 кГц (радиорелейные, сотовые и спутниковые линии) [4].

2.11 Каналы передачи сигналов

Каналом называется совокупность объектов (линия передачи и некоторые устройства преобразования сигналов в передатчике и приемнике), предназначенных для передачи сообщений определенного вида. Каналы можно классифицировать по разным видам признаков.

В разд. 1 были охарактеризованы три класса сообщений и сигналов в зависимости от их *структуры*: непрерывные, дискретные и цифровые. Соответственно можно назвать три вида каналов.

Непрерывный (аналоговый) канал предназначен для передачи сигналов, представляющих реализации непрерывной случайной функции времени $u(t)$ (разд. 1.4), и он постоянно занят во время передачи этого сигнала. Любая линия связи (медный или оптический кабель, радиолиния) даже без каких-либо дополнительных устройств может выполнять роль аналогового канала. Одна из основных характеристик такого канала – его полоса пропускания. Поэтому непременной составляющей любого аналогового канала является совокупность частотно-селективных элементов, определяющих эту полосу. Например, простейший телефонный канал является аналоговым и имеет полосу пропускания шириной 3100 Гц, лежащую в интервале 300 – 3400 Гц (разд. 2.2).

В аналоговом канале процедура построения функции $\hat{x}(t)$, с максимальным приближением отображающей переданное сообщение $x(t)$ в условиях, когда приемнику доступна лишь реализация $v(t)$ полезного сигнала с шумом, называется задачей *фильтрации случайного процесса* $v(t)$ [9]. Показателем качества приема служит обычно функция ошибки $\Delta x(t) = x(t) - \hat{x}(t)$. Эта функция случайна. Заметим, что в аналоговом канале она практически никогда не равна нулю, то есть в таком канале переданное сообщение в любой момент воспроизводится с погрешностью. Довольно часто количественным показателем служит среднеквадратическая величина этой погрешности.

Дискретный канал предназначен для передачи дискретного сигнала, который, как правило, получается в результате квантования по времени непрерывного сигнала и поэтому представляет собой последовательность им-

пульсов-отсчетов. В состав дискретного канала обязательно входят устройства тактовой синхронизации, которые указывают те моменты времени, в которые следует ожидать поступления очередного импульса на вход приемника (дают команду приемнику начать обработку очередного импульса). Итак, одна из важнейших характеристик дискретного канала – это тактовая частота. Например, в дискретном канале, предназначенном для передачи телефонного сигнала после его квантования по времени, тактовая частота равна 8 кГц.

Приемник в дискретном канале по каждому принятому импульсу, случайным образом искаженному помехами, пытается определить, какой величины была амплитуда x этого импульса при его передаче, то есть до искажения помехами. Это задача *оценки параметра* [9], и результатом ее решения является оценка (число) $\hat{x}$. И в этом случае погрешность может быть большой или маленькой, но она присутствует при оценке любого импульса-отсчета.

Цифровой М-ичный канал предназначен для последовательной передачи *М*-ичных символов. При передаче применяется один из методов, описанных в разд. 2.8. Для передачи каждого символа, независимо от того, какой у него порядковый номер и которое из *М* возможных значений он принял, отводится одинаковый интервал времени τ, то есть каждый символ имеет вид импульса того или иного вида, размещенного на этом интервале. Очевидно, что и в цифровом канале нужна система тактовой синхронизации, и тактовая частота входит в число основных параметров. Принято называть ее *технической скоростью передачи* и измерять в бодах (1 Бод = 1 импульс в секунду). При использовании сложных методов модуляции, например КАМ, каждый передаваемый импульс несет значения нескольких двоичных символов, поэтому техническую скорость передачи можно измерять и в бит/с. Например, стандартный двоичный цифровой телефонный канал характеризуется технической скоростью передачи в 64 кбит/с (разд. 2.6).

Таким образом, в состав цифрового канала кроме линии передачи обязательно входят устройство формирования сигналов (например, модулятор) в пункте передачи и решающее устройство (например, демодулятор) в пункте приема. Решающее устройство в момент окончания очередного импульса должно указать, которое из *М* возможных значений символа было передано при помощи данного импульса. Это задача *проверки статистических гипотез* [9], и результом является *решение* $\hat{x}$, принятое приемником, разумеется, лишь одно из *М* возможных. Поскольку в линии передачи действуют помехи, случайным образом искажающие форму сигнала, вынесение такого решения является далеко не тривиальной задачей, и понятно, что демодулятор иногда ошибается.

В нормальных условиях функционирования СПИ при передаче последовательности символов, составляющих сообщение, значения большинства из них (а, может быть, и всех) приемнику удается угадать, т.е. значения этих

символов (или даже всего сообщения) передаются *абсолютно точно*. И лишь некоторая их часть восстанавливается с ошибкой.

В этом заключается принципиальное отличие цифровых методов передачи от всех остальных. Цифровая СПИ не признает “полутонов”: она либо выполняет свои функции идеально, либо совершает ошибки, и такие части сообщения могут искажаться довольно существенным образом либо утрачиваются вообще.

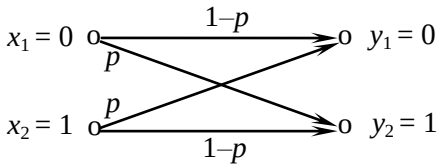

Рис. 2.21. Переходные вероятности двоичного симметричного канала

Популярная модель цифрового канала – это двоичный симметричный канал с независимыми ошибками, который можно описать при помощи графа, изображенного на рис. 2.19, где x_1 и x_2 – возможные значения символа на входе (например, 0 и 1), y_1 и y_2 – возможные значения символа на выходе. Симметрия канала проявляется в том, что вероятности ошибок обоих видов (1 – выбор y_2 при условии, что передан x_1; 2 – выбор y_1 при условии, что передан x_2) равны между собой. Эта величина p называется *битовой вероятностью ошибки* в канале (во всем мире широко используется аббревиатура BER = Bit Error Rate).

Если параметры линии передачи постоянны, а единственной помехой является аддитивный белый гауссовский шум, то не так уж сложно применить такой метод вынесения решения, при котором *ошибки в различных символах будут независимы*, то есть вероятность ошибки при приеме данного символа не изменяется в зависимости от того, появлялись ли ошибки при приеме предыдущих символов. Иногда такой канал называют каналом без памяти.

Как уже отмечалось, спектр прямоугольных импульсов бесконечно широк (см. рис.2.14 и 2.18), а полоса пропускания линии ограничена. Из-за этого возникает межсимвольная интерференция. Она не приводит к зависимости ошибок в канале или к их группированию, но величина BER при этом возрастает.

Если же в канале имеется еще и мультипликативная помеха, то в те периоды времени, когда уровень сигнала велик, ошибки происходят редко, но на интервалах, соответствующих замираниям сигнала, частота появления ошибок резко возрастает и даже происходит *группирование ошибок*. В таком случае говорят, что возникают *пачки (пакеты) ошибок,* причем длина каждого пакета, разумеется, случайна. Здесь одной величины p недостаточно для

полного определения надежности передачи в таком канале, так как ошибки становятся зависимыми, нужно характеризовать и длину пакета.

В состав цифрового канала могут входить несколько последовательно включенных устройств формирования сигнала в передатчике, например, кодирующее устройство и модулятор. Соответственно, и процесс вынесения решения в приемнике также может осуществляться в несколько этапов. На всех этапах, кроме последнего, возможны три режима. Опишем их на примере работы демодулятора двоичных импульсов. В *жестком режиме* вынесения решения демодулятор по каждому принятому импульсу выносит конкретное решение (0 или 1). В *мягком режиме* демодулятор не называет конкретное значение очередного символа, а выдает число p_0, равное вероятности того, что с данным импульсом был передан символ 0. Ясно, что при этом $p_1 = 1 - p_0$.

В *мягком режиме со стираниями* (это промежуточный вариант) демодулятор цифрового канала настроен таким образом, что он выдает конкретное значение (0 или 1) очередного символа лишь в том случае, когда принятый импульс оказался не слишком искаженным помехами, поэтому нет особых сомнений относительно того, что именно было передано. В остальных случаях, когда неопределенность велика, демодулятор отказывается назвать значение символа. Такой символ считается *стертым*, но позиция символа, разумеется, сохраняется с тем, чтобы декодирующее устройство впоследствии могло поставить на эту позицию конкретное значение. Отсюда понятно, что последнее решающее устройство в цепи может работать лишь в жестком режиме.

Может показаться странным, но это факт: при всей его неопределенности решение, выдаваемое в мягком режиме, оказывается более информативным для последующих устройств, хотя их структура при этом усложняется. Пример из жизни: слушая прогноз погоды, мы предпочли бы фразу “дождь будет с вероятностью 0,7” услышать целиком вместо урезанного и категоричного варианта “дождь будет”. Поэтому нет ничего удивительного в том, что в современных СПИ мягкий режим работы решающих устройств используется все шире (см. разд. 5.7).

2.12 Последовательный и параллельный способы передачи

Естественной является *последовательная* передача сообщения по мере появления элементов этого сообщения на выходе источника.

В качестве альтернативы можно представить *параллельный способ* передачи сообщения, состоящего из N символов. Для этого нужно ресурсы канала распределить между N субканалами, тогда в каждом из субканалов можно одновременно передать лишь по одному символу сообщения. При

этом длительность символа в каждом из субканалов можно увеличить в N раз, следовательно, во столько же раз уменьшится требуемая полоса частот, так что время передачи сообщения и суммарная полоса частот не изменятся (это мы и называем “распределить ресурсы канала”).

Ясно, что при передаче длинного сообщения организовать такое большое количество субканалов не всегда удается, поэтому на практике применяют комбинированный, *последовательно-параллельный* способ передачи. Здесь передаваемую последовательность символов делят на группы (комбинации) по n символов в каждой, и для их одновременной передачи используют n параллельных субканалов.

Назовем несколько наиболее применяемых методов организации параллельных субканалов.

Во-первых, это метод *пространственного разделения*. Здесь символы одновременно передаются по n параллельным шинам (обычно медным проводам). Такой способ применяется лишь при передаче на очень малые расстояния, например, между блоками аппаратуры.

Во-вторых, весьма популярна одновременная передача двух символов на двух *квадратурных составляющих*, то есть в каждом из двух субканалов используются несущие той же частоты, но сдвинутые по фазе на 90˚. Легко убедиться, что если каждую из квадратурных составляющих подвергнуть АМ с удалением из спектра несущей, а затем эти сигналы сложить, то есть передать по общей линии, то получим сигнал с КАМ (рис. 2.19).

В-третьих, существует метод передачи с использованием *ортогональных частотно-разделенных сигналов* (*Orthogonal Frequency-Division Multiplexing* (*OFDM*)). Для его реализации в полосе частот $\Delta f = 1/\tau$, минимально необходимой для последовательной передачи радиоимпульсов длительности τ (2.29), организуют n гармонических поднесущих с шагом по частоте, равным $\Delta f / n$. Символы комбинации передают одновременно на этих поднесущих, используя один из методов манипуляции, описанных в разд. 2.8 (кроме ЧМ, для которой нужны две поднесущие). В каждом субканале передаются радиоимпульсы прямоугольной формы, при этом форму спектра каждого из них задает функция Котельникова, сдвинутая по оси частот. Представить, как выглядят спектры сигналов во всех субканалах при передаче комбинации из $n = 15$ двоичных символов с АМ, можно, обратившись к тому же рис. 2.15 и заменив в нем время t на частоту f.

Благодаря тому, что спектр сигнала в каждом из субканалов стал в n раз уже, общее ограничение полосы частот очень мало искажает его форму (следовательно, мало искажается и форма самих импульсов), что обеспечивает малость межсимвольной интерференции как между символами, разнесенными во времени (в одном субканале), так и между символами, передаваемыми на разных поднесущих.

Здесь без потери взаимной ортогональности сигналов можно использо-

вать и КАМ, то есть параллельно передавать разные символы на обеих квадратурных составляющих.

Этот метод известен уже несколько десятилетий, но начал применяться лишь в последние годы при передаче данных по медному кабелю, но особенно широко в системах радиодоступа. Количество поднесущих n может составлять сотни и даже тысячи. Представить себе передатчик или приемник, в котором имеется n комплектов аппаратуры для одновременной обработки такого количества сигналов, довольно трудно. Поэтому операция формирования суммы сигналов из частотно-разнесенных субканалов и операция разделения этих сигналов выполняются после их АЦП в цифровом сигнальном процессоре, реализующем процедуры прямого и обратного быстрого преобразования Фурье (2.5).

Четвертый метод параллельной передачи символов комбинации можно назвать методом *кодового разделения* (*Code Division Multiplexing* (*CDM*)). Здесь все n символов комбинации передаются одновременно и полностью занимают один и тот же интервал времени длительности τ, при этом спектр каждого из них размещен в одной и той же полосе частот Δf. Возможность разделения этих символов в пункте приема обусловлена тем, что перед подачей на вход линии каждый из этих символов (радиоимпульсов) подвергается дополнительной внутриимпульсной двоичной фазовой манипуляции, при этом в разных субканалах применяются разные законы манипуляции так, чтобы все передаваемые сигналы оказались взаимно ортогональными.

Сейчас мы обсуждали, как поток символов, создаваемый одним источником, передавать одновременно по n каналам. Но актуальна и обратная задача. Допустим, мы имеем n потоков символов, одновременно создаваемых разными источниками информации. Их можно передать одновременно по n параллельным каналам, но ведь можно использовать и один канал с большей величиной $V_т$ для последовательной передачи всех поступающих символов. Поэтому обсуждение подобных проблем мы продолжим в разд. 6.

2.13 Статистический синтез цифровой системы передачи информации

В заключение раздела посмотрим, во что выливается проблема полного статистического синтеза цифровой системы передачи информации. Исследуем идеализированную ситуацию, к которой можно лишь приблизиться на практике:

1) источник информации выдает одно из M возможных сообщений x_1, … ,x_M, причем все они равновероятны;

2) имеется набор сигналов известной формы $u_1(t)$, …, $u_M(t)$, заданных на интервале $0 < t < T$, и передатчик при появлении сообщения x_k передает в линию сигнал $u_k(t)$;

3) линия передачи идеальна, то есть в ней имеется лишь аддитивная помеха (1.22) в виде белого шума, а значение коэффициента передачи линии K известно, в итоге форма любого из ожидаемых сигналов на входе приемника полностью известна;

4) задана простая функция потерь, то есть, потери считаются одинаковыми при возникновении ошибок любого вида.

Зададим полосу частот $\Delta f = f_{â} - f_{í}$, в которой заведомо полностью помещается спектр любого из сигналов, и запишем совместную плотность вероятности для $n = T/\Delta t$ отсчетов сигнала с шумом на входе приемника при условии, что передается сигнал $u_k(t)$

$$W(v_1,...,v_n / u_{k,1},...,u_{k,n}) = W(v_1 / u_{k,1}) \cdots W(v_n / u_{k,n}) = \\ = \frac{1}{(2\pi)^{n/2}\sigma_{ø}^{n}} \exp\left(-\frac{\Delta t}{2\sigma_{ø}^{2}} \sum_{j=1}^{n} \left(v_j - u_{k,j} \right)^2 \right). \qquad (2.44)$$

Если задана простая функция потерь, то средняя величина потерь минимальна, если приемник выбирает тот из M сигналов, для которого вероятность его передачи при наблюдаемой реализации $v_1,...,v_n$ максимальна [9]. Это метод максимума апостериорной вероятности.

Поскольку все M значений сигнала равновероятны, метод максимума апостериорной вероятности превращается в метод максимального правдоподобия [9], то есть приемник выбирает тот из M сигналов, для которого величина (2.44) оказалась максимальной.

Это тот сигнал $u_k(t)$, для которого сумма в (2.44) минимальна (см. (1.42)). Итак, пользуясь геометрическими представлениями, имеем M векторов $\mathbf{u_1},\ldots,\mathbf{u_M}$, отображающих возможные значения полезного сигнала на входе приемника, и вектор $\mathbf{v}$, отображающий имеющуюся реализацию сигнала с шумом. Приемник должен выбрать тот из векторов $\mathbf{u_1},\ldots,\mathbf{u_M}$, который оказался наиболее близким к вектору $\mathbf{v}$.

Определив алгоритм работы приемника, попытаемся выяснить, как нужно задать M векторов $\mathbf{u_1},\ldots,\mathbf{u_M}$, чтобы минимизировать полную вероятность ошибки. Плотность вероятности (2.44) обладает центральной симметрией, то есть ее значение зависит лишь от расстояния между $\mathbf{u_k}$ и $\mathbf{v}$, но не от направления. Отсюда следует, что векторы $\mathbf{u_1},\ldots,\mathbf{u_M}$ должны располагаться на максимально возможных расстояниях друг от друга.

Теоретически проблема решается чрезвычайно просто – нужно увеличить мощность передатчика, и будем иметь любые сколь угодно большие расстояния. Только практики будут не в восторге от такого решения. Поэтому проблему лучше решать при условии, что средняя величина энергии сигнала E_c фиксирована.

Во все условия и выражения, которые мы приводили, M возможных сигналов входят на равных основаниях. Поэтому можно ожидать, что если и сама система сигналов будет обладать подобной симметрией, будем иметь некоторый экстремум (максимум или минимум) для полной вероятности

ошибки.

Действительно, В.А. Котельниковым было показано [1,6,7], что минимум полной вероятности ошибки достигается в том случае, когда система сигналов обладает следующими свойствами:

1) все сигналы обладают одинаковой энергией (длины всех векторов одинаковы);

2) расстояния d между всеми сигналами одинаковы и максимальны.

Сигналы, удовлетворяющие этим условиям, называются *симплексными*.

Для такой системы сигналов после ее нормировки все элементы на главной диагонали матрицы Грама (1.47) равны единице, а все остальные элементы также принимают одно и то же значение ρ. Найдем это значение.

Очевидно, что справедливо неравенство

$$\sum_{j=1}^{n}\left(\sum_{k=1}^{M} u_{k,j}\right)^2 \Delta t \geq 0. \qquad (2.45)$$

Заменим квадрат суммы двойной суммой и учтем условие нормировки и (1.44)

$$\sum_{j=1}^{n}\sum_{k=1}^{M}\sum_{m=1}^{M} u_{k,j}u_{m,j}\Delta t = \sum_{k=1}^{M}\sum_{m=1}^{M}\Gamma_{k,m} = M + (M^2 - M)\rho \geq 0. \qquad (2.46)$$

Максимум расстояния между векторами обеспечивается при наименьшем значении ρ, для которого имеем

$$\rho = -\frac{1}{M-1}. \qquad (2.47)$$

Расстояние между любыми двумя сигналами из (1.50) равно

$$d(\mathbf{U_j}, \mathbf{U_k}) = \sqrt{\frac{2M}{M-1}}. \qquad (2.48)$$

На рис. 2.22 изображены симплексные системы векторов для $M = 2$ и $M = 3$. При M = 2 векторы соответствуют двум противоположным сигналам. Именно такие сигналы формируются при ФМ (рис. 2.18а). Таким образом, метод ФМ занимает особое место среди других методов модуляции в двоичных системах.

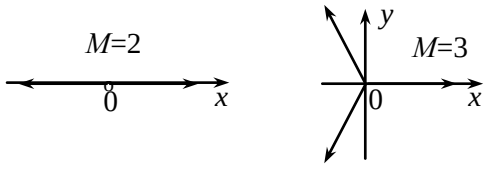

Рис. 2.22. Симплексные системы векторов для M=2 и M=3

Во-первых, если реализованы оговоренные выше условия, он должен обеспечить минимально-возможную вероятность ошибки при приеме двоичного сигнала.

Во-вторых, он широко используется при генерировании различных си-

стем сигналов при $M > 2$. Дело в том, что непосредственное генерирование симплексных сигналов при $M >> 1$ связано с определенными трудностями. Значительно проще реализовать это в два этапа: сначала генерируется последовательность, содержащая $M - 1$ двоичный символ (симплексный код), а затем эта последовательность подается на двоичный ФМ-модулятор. Сигналы (не обязательно симплексные), полученные таким образом, называются *сигнально-кодовыми конструкциями.*

Из (2.47) видно, что при $M >> 1$ симплексная система сигналов мало отличается от *ортогональной системы*, для которой $\rho = 0$. Наконец, промежуточное положение занимает *биортогональная система*, в которой для каждого сигнала $\mathbf{u_k}$ существует противоположный сигнал $-\mathbf{u_k}$, а все остальные сигналы ортогональны сигналу $\mathbf{u_k}$**.** Как будет показано в разд. 5, высокая энергетическая эффективность при использовании подобных систем сигналов может быть достигнута лишь при $M >> 1$.

Итак, для одного частного идеализированного случая мы решили задачу статистического синтеза всей системы передачи информации, и уже на этом примере увидели те трудности, которые придется преодолеть при практической реализации оптимальной СПИ.

Во-первых, это трудности при генерировании самих сигналов, но, как уже отмечалось, они частично преодолеваются путем использования сигнально-кодовых конструкций.

Во-вторых, каждый сигнал из системы, содержащей $M = 2^k$ таких сигналов, занимает довольно широкую полосу частот. Необходимая полоса была бы намного меньше, если бы то же сообщение передавалось обычным способом при помощи комбинации, состоящей из k двоичных символов.

В-третьих, провести обработку принимаемого сигнала в соответствии с выражением (2.44) можно при помощи корреляционного приемника. В итоге в пункте приема нужно иметь M параллельно работающих корреляционных приемников, каждый из которых настроен на прием своего сигнала. Эта процедура называется “прием сигнала в целом”.

В-четвертых, если условия работы СПИ начать приближать к реальным (неравновероятность сигналов на выходе источника информации; нестабильность параметров устройств, генерирующих и обрабатывающих сигналы; наличие мультипликативной помехи и т.п.), задача статистического синтеза всей СПИ становится неразрешимой.

Это еще раз подтверждает мысль, высказанную в разд. 2.1: сложную задачу по необходимости приходится разбивать на ряд более простых задач, то есть обработку сигналов и в процессе передачи, и в процессе приема проводить поэтапно, при этом нет полной уверенности в том, что на каждом этапе мы идем к конечной цели кратчайшим путем.

Вопросы для самоконтроля по главе 2

1. Опишите этапы аналого-цифрового преобразования непрерывного сигнала.
2. Опишите этапы цифро-аналогового преобразования.
3. Изобразите обобщенную модель системы передачи информации. Опишите функции кодера и декодера.
4. Приведите несколько примеров преобразователей сообщения в первичный сигнал.
5. Зачем нужна модуляция? Назовите виды аналоговой модуляции гармонической несущей.
6. Назовите способы манипуляции гармонической несущей. Чем обусловлен выбор того или иного способа?
7. Каковы недостатки многопозиционных методов манипуляции гармонической несущей?
8. Из каких соображений выбирается шаг квантования непрерывного сигнала: по напряжению?
9. Из каких соображений выбирается шаг квантования непрерывного сигнала:по времени?
10. Дайте определения терминов: сообщение, сигнал, помеха, канал связи, линия связи, многоканальная связь, многостанционный доступ, техническая скорость передачи.
11. Почему шаг квантования непрерывного сигнала по времени выбирается меньше того значения, которое следует из теоремы отсчетов?
12. Укажите стандартную частоту квантования во времени (отсчетов/с) телефонного сигнала.
13. Укажите количество разрядов в стандартном АЦП, применяемом при преобразовании телефонного сигнала.
14. Что удобнее применять на практике –коррелятор или согласованный фильтр?

3 КОДИРОВАНИЕ КАНАЛА.

3.1 Корректирующие коды

Наличие помех в линии передачи сигналов приводит к тому, что при демодуляции принимаемого сигнала с помехой демодулятор не всегда верно угадывает значение очередного *m*-ичного символа, т.е. вместо истинного значения x_i выдает другое значение x_k, взятое из алфавита $x_1, x_2, ..., x_m$. В таком случае говорят, что в данном символе произошла ошибка. Ошибки возникают случайным образом, поэтому нет гарантии того, что в принятом сигнале они отсутствуют.

В таких условиях дальнейшая обработка (декодирование) цифрового сигнала с выхода демодулятора основана на анализе взаимосвязи между символами в последовательности. Если декодер приходит к выводу, что такая последовательность символов в принципе не могла быть передана, то такое решение называется *обнаружением ошибок*. На данном этапе декодер не пытается выяснить, какие именно символы приняты ошибочно и каковы их истинные значения, важно лишь установить факт, что в принятой последовательности хотя бы один символ содержит ошибку. Таким образом, этап обнаружения ошибок завершается выдачей решения в двоичной форме: “нет ошибок”, “есть хотя бы одна ошибка”. Если принято первое решение, декодирование очередной принятой последовательности на этом фактически завершается.

Если в системе передачи информации отсутствует обратный канал, то есть, нет возможности сообщить в пункт передачи о принятом решении, декодер после обнаружения ошибок предпринимает попытку восстановить переданное сообщение. Если ему удалось угадать положение ошибочных символов и их истинные значения, считают, что произошло *исправление ошибок*.

Очевидно, что угадывание далеко не всегда завершается успехом, поэтому не все возможные сочетания ошибок могут быть исправлены и даже обнаружены. В итоге иногда на выходе декодера всё-таки будут появляться последовательности символов, содержащие ошибки. Поэтому основной задачей теории помехоустойчивого кодирования является поиск таких последовательностей символов, при использовании которых процент ошибочных комбинаций на выходе декодера был бы как можно ниже.

Итак, *корректирующим называется такой код, использование которого даёт возможность в процессе приёма цифрового сигнала обнаруживать в нём наличие ошибок и, возможно, даже исправлять некоторые из ошибочных символов.*

Среди корректирующих кодов наибольшее распространение получили *блочные двоичные коды*, то есть передача двоичного сообщения производится блоками, причём каждый блок содержит *n* двоичных символов. Кодирова-

ние и декодирование каждого блока производится независимо от других блоков.

Напомним, что *расстояние Хэмминга* (1.43) d_{ab} между двумя кодовыми комбинациями **a** и **b** численно равно количеству символов, в которых эти комбинации отличаются одна от другой. Например, пусть n=5, **a**=01101, **b**=00001, тогда $d_{ab}=2$.

Вес комбинаций равен количеству единиц, содержащихся в ней. Например, W(**a**)=3, W(**b**)=1. Очевидно, что $0 \le W \le n$.

Расстояние Хэмминга удобно вычислять, пользуясь операцией суммирования по модулю 2

$$d_{ab} = W(\mathbf{a}+\mathbf{b}). \qquad (3.1)$$

В дальнейшем суммирование двоичных кодовых комбинаций или их элементов предполагается проводить лишь по mod 2, даже если используется обычный знак суммирования.

Вектор ошибки **e** – это n-разрядная двоичная комбинация, в которой положение единиц указывает на положение ошибочных символов в принятой комбинации **y**. Пусть при этом передана комбинация **x** , тогда

$$\mathbf{y}=\mathbf{x}+\mathbf{e}. \qquad (3.2)$$

Например, если n=6, **x**=011010, **e**=010100, то **y**=001110.

Кратность ошибки q – это количество ошибочных символов в принятой комбинации

$$q=W(\mathbf{e})=W(\mathbf{x}+\mathbf{y})=d_{xy}, \qquad (3.3)$$

причём $0 \le q \le n$.

Величина q случайна, в каналах с независимыми ошибками её математическое ожидание $m_q=pn$, где p – вероятность появления ошибки в одном символе на выходе демодулятора (битовая вероятность ошибки). В реальных каналах связи отношение сигнал/помеха обычно настолько велико, что $pn \ll 1$. Это значит, что большинство комбинаций будут приняты без ошибок (q=0), изредка будут встречаться комбинации, содержащие где-то один ошибочный символ (q=1), ещё реже – комбинации с двукратными ошибками и т.д. То есть, принятая комбинация лежит недалеко от переданной.

Очевидно, что никакой код не способен обнаружить и исправить все возможные ошибки вплоть до $q=n$. Возможности любого кода ограничены, поэтому в первую очередь нужно направить усилия на борьбу с теми видами ошибок, которые встречаются наиболее часто, то есть с ошибками малых кратностей.

Способ кодирования полностью определён, если задана *кодовая таб-*

Таблица 3.1 – Пример кодовой таблицы (n=6)

Буква	Комбинация
а	000000
б	000111
в	111000
г	111111

лица, в которой перечислены все возможные сообщения и соответствующие им n-разрядные кодовые комбинации. Например, если каждая комбинация соответствует одной букве алфавита, а в используемом алфавите всего 4 буквы, кодовая таблица может иметь следующий вид (табл.3.1).

Можно вычислить расстояние d_{jk} между двумя любыми комбинациями в кодовой таблице. Минимальное его значение
называется кодовым расстоянием и служит одной из важнейших характери-

$$d_{код} = \min_{j \neq k} d_{jk} \qquad (3.4)$$

стик выбранного способа кодирования. Это показатель отличия двух наиболее близких комбинаций в таблице. В приведённом примере
$d_{аб} = 3, d_{ав} = 3, \ldots, d_{аг} = 6$, в итоге $d_{код} = 3$.

Довольно трудоёмкий, зато универсальный способ декодирования – это *декодирование по минимуму расстояния*. Принятая комбинация сравнивается поочерёдно со всеми комбинациями в таблице. Если такая комбинация есть в таблице, то декодирование заканчивается. Конечно, при этом нет полной уверенности в том, что в принятой комбинации отсутствуют ошибки, хотя это весьма вероятно. Просто здесь нет лучшего варианта, поскольку обычно равновозможна передача любой из комбинаций, входящих в таблицу.

В противном случае (принятая комбинация не совпадает ни с одной комбинацией в кодовой таблице) следует несомненный вывод: в принятой комбинации есть ошибка и, может быть, не одна (произошло обнаружение ошибок).

Логически рассуждая и рассматривая примеры, легко убедиться в том, что *можно гарантированно обнаруживать любые ошибки кратности*

$$q_{\text{о}} \leq d_{код} - 1. \qquad (3.5)$$

В приведённом примере (табл. 3.1) код обнаруживает все однократные и двукратные ошибки. Этот же код способен обнаруживать также некоторые ошибки более высоких кратностей, но не все. В частности, ошибки в первых трёх символах кодовой комбинации не будут замечены.

После обнаружения ошибок в СПИ с обратным каналом посылают запрос на повторную передачу данной комбинации. При отсутствии канала переспроса следует попытаться исправить ошибки, то есть найти в кодовой

таблице комбинацию, наиболее близкую к той, которая была принята (наиболее правдоподобную). Даже если в таблице нашлась лишь одна такая комбинация, это ещё не служит гарантией того, что ошибки исправлены верно. *Можно гарантированно исправить любые ошибки кратности*

$$q_{\text{и}} \le \frac{d_{код} - 1}{2}. \qquad (3.6)$$

Например, код (табл. 3.1) способен исправить любую однократную ошибку. Убедитесь на примерах в том, что ресурсы кода используются полнее, если $d_{код}$ является нечётным числом. Кроме того, видно, что исправлять ошибки труднее, нежели их обнаруживать. Разумеется, возможны случаи, когда исправляются некоторые ошибки более высоких кратностей, чем следует из (3.6), но не все.

3.2 Линейные блочные коды

Когда количество символов *n* в кодовой комбинации велико (на практике оно может составлять несколько тысяч), кодовая таблица блочного кода может содержать огромное количество комбинаций. Кодирование и декодирование оказываются слишком трудоёмкими. Поэтому вполне естественно желание разработать такие коды, которые позволяют проводить эти операции, не обращаясь к кодовой таблице, а используя лишь некий набор свойств, присущий всем комбинациям в кодовой таблице. Именно такую возможность предоставляет применение линейного блочного кода.

Любой *линейный блочный код* обозначается как (*n*,*k*)-код, где: *k* - количество информационных символов, то есть длина двоичной комбинации, поступающей на вход кодера; *n* - количество символов в комбинации на его выходе. На вход кодера может поступать любая *k*–разрядная комбинация. Число таких комбинаций равно

$$N = 2^k. \qquad (3.7)$$

Если бы кодирование не проводилось (*n*=*k*, то есть входная комбинация сразу поступает на выход кодера), такой код не обладал бы никакими корректирующими свойствами, поскольку для него $d_{код} = 1$, и любые ошибки в принятой комбинации оказались бы незамеченными.

Поэтому для корректирующего кода *n*>*k*, и разность *r*=*n*-*k* есть число проверочных символов. Значения этих символов вычисляется по информационным символам с использованием ряда линейных операций.

Очевидно, что для самого простого кода *r*=1. Такой код называется *кодом с проверкой на чётность* (*n*, *n*-1). На вход кодера поступает комбинация $\mathbf{a} = (a_1, a_2, ..., a_k)$. Кодер повторяет значения этих символов $s_1 = a_1, ..., s_k = a_k$ и добавляет к ним ещё один, проверочный символ (напоминаем об операции mod2)

$$s_{k+1} = a_1 + a_2 + ... + a_k . \qquad (3.8)$$

В итоге любая комбинация **s** на выходе кодера содержит чётное количество единиц. Это и есть то самое свойство, которое присуще всем переданным комбинациям. Например, при k=7 имеем: **a**=0110100→**s**=01101001; **a**=0000000→**s**=00000000; **a**=1111111→**s**=11111111. При декодировании достаточно провести общую проверку на чётность

$$c_1 = s_1 + ... + s_n . \qquad (3.9)$$

Если проверка прошла ($c_1 = 0$), то это лишь означает, что такая комбинация могла быть передана. Если же проверка не прошла ($c_1 = 1$), то в принятой комбинации, несомненно, есть ошибки.

Кодовое расстояние для кода с проверкой на чётность равно двум, поэтому в соответствии с (3.5) он обнаруживает любую однократную ошибку. Более того, он способен обнаружить любую ошибку нечётной кратности, но исправлять ошибки он не может.

Операции (3.8) и (3.9) – чрезвычайно простые. Для реализации каждой из них достаточно иметь один счётный триггер. Тогда сразу возникает идея для повышения корректирующей способности кода использовать не одну, а несколько проверок на чётность. При этом очевидно, что в каждой проверке участвуют не все n символов, а лишь их часть. А в разных проверках должны участвовать разные группы символов.

В качестве примера рассмотрим код (5,2). Зададим его в *систематической форме*, то есть первые k символов выходной комбинации должны повторять входные информационные символы, а на оставшихся r=n-k=3 позициях размещаются проверочные символы, которые предстоит вычислить при кодировании, то есть

$$\mathbf{s}=(\mathbf{a},\mathbf{b}), \qquad (3.10)$$

где **a** – вектор-строка информационных символов, **b** – вектор-строка проверочных символов. В частности, рассмотренный выше простой код с проверкой на чётность тоже является систематическим.

Итак, для кода (5,2) из $N_1 = 2^n = 32$ всевозможных пятиразрядных комбинаций в кодовую таблицу включим лишь такие комбинации, которые удовлетворяют всем трём заданным проверкам на чётность (количество проверок на чётность всегда равно r). Результаты проведения этих r проверок для конкретной комбинации представим в виде вектора-строки $\mathbf{c} = (c_1, c_2, ..., c_r)$, который имеет медицинское название “синдром”, то есть, сочетание признаков, характеризующих определённое состояние. Допустим, для нашего кода мы выбрали следующую систему проверок

$$c_1 = s_1 + s_3; \quad c_2 = s_1 + s_2 + s_4; \quad c_3 = s_2 + s_5 , \qquad (3.11)$$

то есть, в первой проверке участвуют символы, стоящие на первой и третьей позициях, и т.д. В кодовую таблицу (типа табл. 3.1) включим лишь те комбинации, для которых вектор **c**=**0**. Оказывается, что их всего 4, как и следовало

ожидать (3.7).

Таблица 3.2. Кодовая таблица кода (5,2)

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

По приведённой таблице (табл. 3.2) легко найти, что $d_{код} = 3$, то есть заданная нами система проверок на чётность (3.11) действительно однозначно определяет систематический корректирующий код, способный не только обнаружить любые однократные и двукратные ошибки, но даже исправлять все однократные ошибки.

Систему проверок (3.11) можно более наглядно представить в виде рис.3.1.

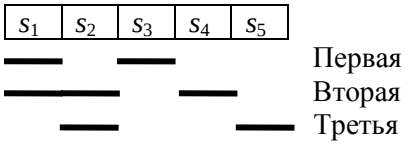

Рис. 3.1. Схема проверок на чётность для кода (5,2)

Первый этап декодирования – обнаружение ошибок – фактически сводится к проведению *r* проверок на чётность (3.11) в принятой кодовой комбинации. Считается, что ошибок нет, если **c**=**0**.

Если **c**≠**0**, второй этап – исправление ошибок – также можно провести, пользуясь схемой рис.3.1 и рассматривая различные варианты. Например, если принята комбинация **y**=11110, то находим **c**=011 и делаем вывод о наличии ошибок в принятой комбинации. Исходя из возможностей данного кода ($d_{код} = 3$), делаем предположение, что произошла однократная ошибка (гарантированно исправлять ошибки более высоких кратностей этот код все равно не может (3.6)). Рассматривая разные варианты расположения этой одиночной ошибки, видим, что наблюдаемый результат **c**=011 мог быть получен лишь в случае, когда эта ошибка расположена во втором символе, следовательно, **x**=10110 и **a**=10.

Чтобы не перебирать различные варианты каждый раз при декодировании очередной кодовой комбинации, полезно заранее заготовить таблицу, в которой перечислить все возможные векторы исправляемых ошибок (однократных в нашем примере) и соответствующие им значения вектора-

синдрома (табл. 3.3). Если бы код был способен исправлять ещё и все двукратные ошибки, в эту таблицу следовало бы включить дополнительно $C_5^2 = 10$ векторов таких ошибок, в частности 11000, 10100 и т.д.

Таблица 3.3. Таблица соответствия вектора ошибки и синдрома для (5,2)-кода

e	**c**
10000	110
01000	011
00100	100
00010	010
00001	001

Из табл. 3.3 видно, что анализируемый код (5,2) действительно способен исправить любую однократную ошибку, поскольку все значения **c** различны, а каждому **c** соответствует единственный вектор ошибки **e**. Легко убедиться, что для двукратных ошибок такого однозначного соответствия уже не будет.

Чтобы завершить обсуждение тех вычислений, которые нужно проводить при кодировании и декодировании линейных блочных кодов, полезно продолжить их формализацию, в частности, ввести для них более компактную форму записи.

Вместо схемы рис. 3.1 удобно использовать *проверочную матрицу* **H**, содержащую r строк и n столбцов и состоящую из 0 и 1. Каждая строка соответствует одной проверке на четность, причем положение единиц в этой строке указывает на позиции символов кодовой комбинации, участвующих в данной проверке. Для рассмотренного кода (5,2) такая матрица имеет вид

$$\mathbf{H} = \begin{bmatrix} 10100 \\ 11010 \\ 01001 \end{bmatrix}. \qquad (3.12)$$

Если линейный блочный код к тому же является систематическим, матрицу **H** можно условно разделить на два блока $\mathbf{H}=(\mathbf{Q},\mathbf{I_r})$, причем левый блок имеет размеры $(r \times k)$, а правый блок - это единичная матрица $(r \times r)$. Для нашего примера

$$\mathbf{Q} = \begin{pmatrix} 10 \\ 11 \\ 01 \end{pmatrix}, \quad \mathbf{I_r} = \begin{pmatrix} 100 \\ 010 \\ 001 \end{pmatrix}. \qquad (3.13)$$

Операция вычисления синдрома (3.11) имеет следующий вид

$$\mathbf{c} = \mathbf{s}\mathbf{H}^{\mathrm{T}}, \qquad (3.14)$$

где $\mathbf{H}^{\mathrm{T}}$ - транспонированная матрица **H**.

Кодер систематического кода заполняет первые *k* позиций информационными символами, поступившими на его вход. На оставшиеся *r* позиций он ставит проверочные символы, при этом вычисляет их значения так, чтобы для *n*-разрядной выходной комбинации оказалось **c**=**0**. Для рассмотренного примера $s_1 = a_1$, $s_2 = a_2$, а из соотношений (3.11) находим три проверочных символа, завершая кодирование:

$$s_3 = s_1, \;\; s_4 = s_1 + s_2, \;\; s_5 = s_2 . \qquad (3.15)$$

Первый этап декодирования принятой *n*-разрядной комбинации **y** – это обнаружение ошибок, и он также определяется соотношением (3.14)

$$\mathbf{c} = \mathbf{y}\mathbf{H}^{\mathrm{T}}. \qquad (3.16)$$

Напомним, что в соответствии с (3.2) вектор ошибок **e** позволяет выразить принятую комбинацию **y** через переданную **s**, то есть **y**=**s**+**e**. Подставим эту сумму в (3.16) и напомним, что для любой передаваемой комбинации $\mathbf{s}\mathbf{H}^{\mathrm{T}} = \mathbf{0}$. В итоге получим

$$\mathbf{c} = (\mathbf{s}+\mathbf{e})\mathbf{H}^{\mathrm{T}} = \mathbf{s}\mathbf{H}^{\mathrm{T}} + \mathbf{e}\mathbf{H}^{\mathrm{T}} = \mathbf{e}\mathbf{H}^{\mathrm{T}}, \qquad (3.17)$$

то есть, синдром, вычисленный по принятой комбинации, не зависит от того, какая комбинация из кодовой таблицы была передана, а определяется лишь вектором ошибок. Именно по этой причине при обсуждении операций декодирования у нас не возникло потребности обращаться к кодовой таблице, а ограничились лишь анализом предполагаемого характера возникающих ошибок. Нетрудно догадаться, что и табл. 3.3. была составлена с использованием формулы (3.17), то есть $\mathbf{c} = \mathbf{e}\mathbf{H}^{\mathrm{T}}$.

Не нужно строить код так, чтобы результат какой-либо проверки на чётность следовал из остальных проверок – это пустая трата ресурсов. Кроме того, любой из *n* символов комбинации должен участвовать хотя бы в одной проверке. При выполнении этих условий любой матрице **H** соответствует единственная матрица **G** размера $(k \times n)$, называемая *производящей (генераторной) матрицей* данного кода. Обе матрицы связаны уравнением

$$\mathbf{G}\mathbf{H}^{\mathrm{T}} = \mathbf{0}, \qquad (3.18)$$

где **0** – нулевая матрица размера (k×r).

Если задана матрица **H**, то, решив уравнение (3.18), можно найти матрицу **G**. Процедура решения оказывается простой, если код является систематическим. Тогда и матрицу **G** условно можно разбить на два блока $\mathbf{G} = (\mathbf{I_k}, \mathbf{Q}^{\mathrm{T}})$, где $\mathbf{I_k}$ – единичная матрица $(k \times k)$. Используя блочные представления обеих матриц, имеем

$$\mathbf{G}\mathbf{H}^{\mathrm{T}} = (\mathbf{I_k}, \mathbf{Q}^{\mathrm{T}}) \begin{pmatrix} \mathbf{Q}^{\mathrm{T}} \\ \mathbf{I_r} \end{pmatrix} = \mathbf{Q}^{\mathrm{T}} + \mathbf{Q}^{\mathrm{T}} = \mathbf{0}, \qquad (3.19)$$

то есть построенная таким образом матрица **G** удовлетворяет уравнению (3.18). В частности, для (5,2)-кода (3.11) матрица **G** имеет вид

$$\mathbf{G} = \begin{pmatrix} 10110 \\ 01011 \end{pmatrix}. \qquad (3.20)$$

Из сказанного ясно, что матрица **G**, как и матрица **H**, полностью определяет код. Поэтому другая интерпретация тех же процессов кодирования и декодирования может быть основана на матрице **G**.

Кодирование линейным блочным кодом можно проводить по формуле

$$\mathbf{s} = \mathbf{aG}. \qquad (3.21)$$

Используя определение (3.18), видим, что синдром для полученной таким образом кодовой комбинации всегда равен нулю

$$\mathbf{c} = \mathbf{sH}^{\mathrm{T}} = \mathbf{aGH}^{\mathrm{T}} = \mathbf{a0} = \mathbf{0}, \qquad (3.22)$$

то есть эта комбинация входит в кодовую таблицу.

Возможность декодирования (вычисления синдрома) принятой комбинации с использованием матрицы **G** становится вполне очевидной, если код является систематическим (3.10). Из принятой комбинации $\mathbf{s}_{\text{п}} = (\mathbf{a}_{\text{п}}, \mathbf{b}_{\text{п}})$ берут вектор-строку информационных символов $\mathbf{a}_{\text{п}}$ и по ней заново проводят кодирование, например, по формуле (3.21), получая новое значение вектора-строки проверочных символов $\mathbf{b}_{\text{н}}$

$$\mathbf{s}_{\text{н}} = \mathbf{a}_{\text{п}}\mathbf{G} = \mathbf{a}_{\text{п}}(\mathbf{I}_{\text{k}}, \mathbf{Q}^{\mathrm{T}}) = (\mathbf{a}_{\text{п}}, \mathbf{a}_{\text{п}}\mathbf{Q}^{\mathrm{T}}) = (\mathbf{a}_{\text{п}}, \mathbf{b}_{\text{н}}). \qquad (3.23)$$

По определению (3.16), вектор-синдром для принятой комбинации $\mathbf{s}_{\text{п}}$ равен

$$\mathbf{c} = \mathbf{s}_{\text{п}}\mathbf{H}^{\mathrm{T}} = (\mathbf{a}_{\text{п}}, \mathbf{b}_{\text{п}}) \begin{pmatrix} \mathbf{Q}^{\mathrm{T}} \\ \mathbf{I}_{\text{r}} \end{pmatrix} = \mathbf{a}_{\text{п}}\mathbf{Q}^{\mathrm{T}} + \mathbf{b}_{\text{п}} = \mathbf{b}_{\text{н}} + \mathbf{b}_{\text{п}}, \qquad (3.24)$$

то есть для вычисления синдрома достаточно сложить два вектора проверочных символов: принятый и вновь вычисленный.

Кроме способности обнаруживать и исправлять ошибки, блочные коды обладают еще и способностью *восстанавливать стертые символы* (см. разд. 2.11).

Допустим, проведено кодирование рассмотренным кодом (5,2), и комбинация на входе декодера имеет вид 1X10X, где буквой X обозначены позиции стертых символов. Считаем, что остальные символы ошибок не содержат. В первой проверке на четность (рис. 3.1) стертые символы не участвуют. Чтобы удовлетворить требованиям второй проверки, на вторую позицию в этой комбинации нужно поставить 1, тогда она примет вид 1110X. Из третьей проверки видно, что и последний символ должен быть единицей. В итоге принято решение, что была передана четвертая комбинация из кодовой таблицы.

Можно на примерах убедиться в том, что код гарантированно может восстановить любые $q_{\text{с}}$ стертых символов, если $q_{\text{с}}$ удовлетворяет тому же условию (3.5).

В заключение сделаем несколько **замечаний.**

1) Любой линейный блочный код обладает следующим свойством: если

из кодовой таблицы взять две (или более) комбинаций и сложить их (разумеется, по модулю 2), то получим комбинацию, принадлежащую той же таблице.

2) Операцию умножения вектора-строки на матрицу удобно провести следующим образом: элементы этого вектора записываем против строк матрицы, а затем суммируем те строки, против которых оказались единицы.

3) По определению (3.4) кодовое расстояние нужно находить по кодовой таблице. Для линейных блочных кодов существует менее трудоёмкий способ: к матрице **G** нужно дописать строку, состоящую из нулей, и тем же способом (3.4) определить минимальное расстояние для полученной таблицы.

4) Каждая строка *проверочной матрицы* **H** соответствует одной проверке на четность, причем положение единиц в этой строке указывает на позиции символов кодовой комбинации, участвующих в данной проверке.

5) Каждый столбец *проверочной матрицы* **H** соответствует синдрому для однократного ошибочного символа, номер которого в кодовой комбинации совпадает с номером столбца. Это позволяет по матрице **H** определить *корректирующую способность* кода. Так. Код способен *обнаружить только однократную ошибку*, если все столбцы *ненулевые и есть одинаковые*. Будет обнаружена *двукратная ошибка* или *исправлена однократная*, если все *столбцы ненулевые и разные*. В общем случае *кодовое расстояние равно минимальному числу линейно зависимых столбцов матрицы* ***H*.**

6) Одновременная одинаковая перестановка столбцов матриц **G** и **H** даёт новый код, эквивалентный исходному, поскольку это приводит к аналогичной перестановке соответствующих символов во всех кодовых комбинациях. Тот же результат будет, если один из двух столбцов заменить их суммой. Многократное повторение подобных операций можно применить, например, для приведения кода к систематической форме.

7) Применяются два способа реализации кодера и декодера: программный (при использовании компьютера) и аппаратный. При аппаратной реализации кодирования для любого линейного блочного кода можно применить такой универсальный (следовательно, не самый экономный) способ: входную k - разрядную последовательность информационных символов последовательно, символ за символом, ввести в k – разрядный входной регистр сдвига, чтобы иметь возможность проводить одновременно операции со всеми символами; при помощи набора сумматоров по модулю 2 вычислить раздельно все n символов кодовой комбинации по формуле (3.21) с учётом примечания 2 (для систематического кода можно ограничиться вычислением лишь проверочных символов по формулам типа (3.15)); при помощи n-входового мультиплексора вывести последовательно все n символов в нужном порядке (разумеется, и для вывода символов можно применить n-разрядный регистр сдвига, предварительно записав в него n вычисленных символов кодовой

комбинации).

Универсальный способ декодирования предполагает следующие действия: *n* принимаемых символов последовательно вводятся в *n*-разрядный регистр сдвига; при помощи набора сумматоров по модулю 2 вычисляют все *r* элементов синдрома по формулам типа (3.11) (можно подсчитать потребное количество сумматоров даже с учётом дублирования некоторых операций); при помощи логического устройства (назовём его анализатором синдрома), пользуясь таблицей (**e**–**c**), по вычисленному значению **c** находят **e**, то есть номера ошибочных символов; при помощи *k*-входового мультиплексора принятые информационные символы последовательно подают на вход сумматора по модулю 2, а на второй вход сумматора в том же порядке с анализатора синдрома подают найденные элементы вектора ошибок, в итоге в этом сумматоре ошибки последовательно исправляются в процессе вывода информационных символов.

8) Избыточность (в техническом смысле) для линейного блочного кода равна

$$R = r / n. \qquad (3.25)$$

Увеличить кодовое расстояние (и корректирующую способность) кода при заданном *n* удаётся лишь ценой увеличения избыточности, поэтому *основной девиз помехоустойчивого кодирования – минимальное значение r при заданных n и* $d_{код}$.

9) Код, укороченный по сравнению с исходным систематическим кодом, получают следующим образом: для передачи сообщения используют лишь последние *m* позиций *k*-разрядного вектора **a**, а первые *k*-*m* позиций заполняют нулями, затем кодируют обычным образом. По линии связи передаётся укороченная комбинация, а в пункте приёма перед декодированием записывают нули на недостающие позиции. При таком укорочении $d_{код}$ и *r* не изменяются, но избыточность возрастает за счёт уменьшения *n*, поэтому этот способ следует применять лишь в случае крайней необходимости.

3.3 Коды Хэмминга

Продолжим рассмотрение линейных блочных кодов, обладающих кодовым расстоянием $d_{код} = 3$ и в соответствии с (3.6) способных исправить любую одиночную ошибку в принятой *n*-разрядной комбинации.

Таблица (**e**–**c**) типа табл. 3.3 содержит *n* строк, при этом все значения *r*-разрядного вектора **c** должны быть различными. Тогда максимально-возможное значение *n* определяется формулой

$$n = n_{\max} = 2^r - 1, \qquad (3.26)$$

где вычитание единицы обусловлено тем, что нулевая комбинация, соответствующая отсутствию ошибок, в таблице не содержится.

Коды, обладающие такими параметрами, называются кодами Хэмминга. К ним относятся коды (3,1), (7,4), (15,11), (31,26) и т.д.

Кстати, код (5,2), рассмотренный в разделе 3.2, также имеет $d_{код}=3$, но не является кодом Хэмминга.

Напомним, что эта таблица (**e-c**) построена с использованием формулы $\mathbf{c}=\mathbf{eH}^{\mathrm{T}}$, и в ней значения вектора-синдрома совпадает с соответствующими столбцами матрицы **H**, поэтому даже нет нужды в построении этой таблицы. К сожалению, такое простое соответствие характерно только для кодов, исправляющих одиночные ошибки.

Итак, для кода Хэмминга в качестве столбцов матрицы **H** нужно записать различные *r*-разрядные двоичные числа, кроме нулевого (красота идеи построения кода Хэмминга станет позже более понятной, если их записать в порядке возрастания).

Для примера запишем матрицу **H** для (7,4)-кода

$$\mathbf{H}=\begin{bmatrix}0001111\\0110011\\1010101\end{bmatrix}. \qquad (3.27)$$

В качестве проверочных для кода Хэмминга можно считать любые *r* символов, но вычисление их значений при кодировании слегка упрощается, если в качестве проверочных взять те символы, которые охвачены лишь одной проверкой на чётность. В нашем примере это s_1, s_2, s_4, тогда информационными символами будут s_3, s_5, s_6, s_7. Перестановкой символов код можно привести к систематическому виду, но пользы от такой операции в данном случае пока не видно.

Кодирование для (7,4)-кода проводится следующим образом: четыре информационных символа a_1, a_2, a_3, a_4, поступившие на вход кодера, записываются на позиции s_3, s_5, s_6, s_7, а затем находят три проверочных символа по той же методике, как и (3.15)

$$s_1=s_3+s_5+s_7,\quad s_2=s_3+s_6+s_7,\quad s_4=s_5+s_6+s_7. \qquad (3.28)$$

Декодирование начинается с вычисления синдрома (3.16). Если **c**=**0**, считают, что ошибок нет. В противном случае вычисленное значение **c** совпадает с номером ошибочного символа, записанным в двоичной форме.

Для (7,4)-кода вычисление элементов синдрома проводится по формулам

$$c_1=s_4+s_5+s_6+s_7,\quad c_2=s_2+s_3+s_6+s_7,\quad c_3=s_1+s_3+s_5+s_7. \qquad (3.29)$$

Например, для принятой комбинации **y**=0111100 имеем **c**=000, поэтому считываем значения y_3, y_5, y_6, y_7 и подаём их на выход декодера (проверочные символы свою роль уже сыграли и получателю они не нужны). Для комбинации 1010001 имеем **c**=101, то есть ошибка в пятом символе – его нужно изменить на обратный (0 на 1).

Код Хэмминга определён формулой (3.26), поэтому он является опти-

мальным кодом, то есть, среди всех кодов с заданными значениями r и $d_{код}=3$ *нет другого кода, который бы обладал меньшей избыточностью (3.25)*. Более того, он относится к классу *совершенных кодов*, то есть, исправляя любую однократную ошибку, он не способен исправить ни одной ошибки более высокой кратности. Хотя обнаружить он может и однократные, и двукратные ошибки. Поэтому при обнаружении ошибок ($\mathbf{c}\neq\mathbf{0}$) дальнейшее декодирование проводится в предположении, что в кодовой комбинации произошла более вероятная, однократная ошибка. Таким образом, совершенство кода заключается в том, что он все ресурсы направляет на исправление однократных ошибок.

Это стремление просматривается в схеме проверок, построенной наиболее рациональным образом, если вспомнить, что результат каждой проверки выдаётся в двоичной форме (да - нет). Представим себе, что перед декодированием комбинация дополнена слева ещё одним знаком $s_0=0$, символизирующим отсутствие ошибок в остальных символах. Тогда проведение проверок на чётность напоминает процедуру угадывания числа, которое может принять любое из 2^r значений. Схема проверок на чётность указывает, каким должен быть смысл каждого из задаваемых вопросов при условии, что на каждый из них даётся ответ “да” или “нет”. Схема проверок кода Хэмминга такова, что позволяет уверенно определить заданное число при минимальном количестве вопросов *r*.

3.4 Циклические коды

Следует отметить, что циклический код – это одна из самых ярких иллюстраций того, как разумное приложение результатов абстрактной математической теории позволило создать весьма простые и эффективные технические устройства – кодеры и декодеры.

Циклический код – это разновидность линейных блочных кодов, поэтому для него справедливо все, что было сказано в разделах 3.1 и 3.2. В частности, сумма двух комбинаций, взятых из кодовой таблицы, дает комбинацию, принадлежащую той же таблице. Циклические коды обладают еще одним, дополнительным признаком: комбинация, полученная после циклической перестановки символов в последовательности, взятой из кодовой таблицы, принадлежит той же таблице.

Стимулом для создания таких кодов явилось желание использовать для кодирования и декодирования простые устройства, подобные линейным цифровым фильтрам (разд. 2.7).

Поэтому и здесь для описания *n*-разрядных комбинаций **s**, состоящих из *M*-ичных символов, удобно использовать полиномы (2.13)

$$\mathbf{s}=(s_0,s_1,...,s_{n-1})\rightarrow s(x)=s_0x^0+s_1x^1+...+s_{n-1}x^{n-1}, \qquad (3.30)$$

где s_j=0; 1;…; *M*-1.

Сразу следует запомнить те особые правила обращения с полиномами, которые обусловлены особыми свойствами циклических кодов по сравнению с другими цифровыми последовательностями.

Во-первых, мы ограничимся изучением лишь двоичных кодов, когда любой символ s_j в любой точке тракта передачи может принимать лишь одно из двух значений: 0 или 1 (кстати, общая теория циклических кодов охватывает и недвоичные коды, причем при больших значениях M код приобретает некоторые новые свойства, но эта теория, естественно, несколько сложнее). Чтобы обеспечить это ограничение, суммирование символов (коэффициентов полиномов) нужно проводить по модулю 2 (смотри формулы (2.1)). В дальнейшем, ради простоты, будем использовать знак обычного суммирования, например,

$$x+x=x(1+1)=0, \quad x^5+x^5+x^5=x^5, \quad (1+x^3+x^4)+(1+x^2+x^4)=x^2+x^3.$$

Во-вторых, циклический код блочный, следовательно, сигналы, получаемые в результате любых преобразований, - это комбинации, состоящие из n бит. Поэтому при умножении полиномов принято соглашение

$$x^{mn}=1, \qquad (3.31)$$

где m – любое целое неотрицательное число, следовательно, $x^{mn}+1=x^{mn}-1=0$, $x^n+1=0$, $x^{mn+j}=x^j$ и т.д.

Например, при n=5 имеем

$(1+x^4)(x^2+x^3)=x^2+x^3+x^6+x^7=x^2+x^3+x+x^2=x+x^3$. В итоге любой полином будет содержать члены с x в степени, не превышающей n-1.

Итак, при умножении полиномов используются соглашения о суммировании mod 2 и (3.31), а правила деления полиномов, разумеется, вытекают из правил их умножения.

Поскольку коэффициенты полиномов – целые (двоичные) числа, при делении полинома на полином получаем полином-частное и полином-остаток. Например, при n=5 после выполнения операции $(x+x^4)/(1+x^2)$ имеем частное $(1+x^2)$ и остаток $(1+x)$. В отличие от случая фильтрации обычных сигналов (разд. 2.7) при работе с циклическим кодом не имеет значения, каким получилось частное, важно лишь, каково значение остатка. Для обозначения полинома, соответствующего остатку от деления $a(x)$ на $b(x)$, будем использовать один из двух вариантов:

$$\mathrm{res}(x) \quad \text{или} \quad a(x) \bmod b(x), \qquad (3.32)$$

при этом видно, что максимальный показатель степени полинома res(x) не превышает p-1, где p – степень полинома $b(x)$.

В частности, и соглашение (3.31) означает, что завершающая операция над полиномом $a(x)$, полученным в результате каких-либо преобразований, - это вычисление остатка

$$\mathrm{res}(x)=a(x) \bmod (x^n+1). \qquad (3.33)$$

Другая особенность, вытекающая из формул (3.31) и (3.33), заключает-

ся в том, что циклическая перестановка символов вправо в n-разрядной комбинации отображается умножением соответствующего полинома на x. Например, при n=7 для комбинации **s**=0110011, $s(x) = x + x^2 + x^5 + x^6$ после циклической перестановки имеем $\mathbf{s}^* = 1011001$, то есть $s^*(x) = 1 + x^2 + x^3 + x^6 = xs(x)$.

Способы кодирования и декодирования конкретным циклическим кодом полностью определяются его *производящим (порождающим, генераторным) полиномом* степени r

$$g(x) = g_0 + g_1 x + ... + g_r x^r, \qquad (3.34)$$

где $g_j = 0;1$, причём всегда $g_0 = g_r = 1$. Формально этот полином тоже содержит n членов, просто у него $s_j = 0$ при $j > r$.

Фундаментальное свойство циклического кода состоит в том, что полином s(x), соответствующий любой разрешённой (передаваемой) комбинации **s***, делится без остатка на производящий полином*

$$s(x) \bmod g(x) = 0. \qquad (3.35)$$

Отсюда следует метод декодирования принимаемой комбинации $v(x)$: нужно вычислить остаток res(x) от деления этого полинома на $g(x)$. Степень полинома res(x) не превышает r-1, то есть этот остаток содержит r бит. Если оказалось, что $\text{res}(x) \neq 0$, то это несомненно указывает на наличие ошибок в принятой комбинации (обнаружение ошибок).

Напомним, что переданная **s** и принятая **v** комбинации связаны соотношение (3.2), откуда следует

$$v(x)=s(x)+e(x), \qquad (3.36)$$

где $e(x)$ – полином ошибок. Тогда с учётом (3.35) имеем соотношение, аналогичное формуле (3.17),

$$\text{res}(x)=v(x)\bmod g(x)=s(x)\bmod g(x)+e(x)\bmod g(x)=e(x)\bmod g(x), \qquad (3.37)$$

то есть величина остатка не зависит от того, какая комбинация была передана, а зависит лишь от того, какие произошли ошибки. Таким образом, остаток res(x), вычисленный при декодировании, обладает тем же свойством, что и синдром **c** любого линейного блочного кода.

Отсюда вытекает первое требование, которому должен удовлетворять “хороший” производящий полином: если нужно обнаруживать все ошибки кратности $q \le q_o$, то для всех таких полиномов $e(x)$ при вычислении по формуле (3.43) должен получаться *ненулевой остаток*.

Из описанной процедуры декодирования становится очевидным и способ исправления ошибок: вычислив остаток для принятой комбинации $v(x)$, нужно подставить его значение в уравнение (3.37) $\text{res}(x)=e(x)\bmod g(x)$ и найти из него вектор ошибок $e(x)$. Тогда очевидно второе требование к полиному $g(x)$: если нужно исправлять все ошибки кратности $1 \le q \le q_и$, то всем таким полиномам $e(x)$ должны соответствовать различные ненулевые остатки, иначе не удастся однозначно решить уравнение (3.37). Кстати, максимальное количество различных ненулевых остатков определяется очевидным образом

$$N_{res} = 2^r - 1, \qquad (3.38)$$

а число различных однократных ошибок равно n, число двукратных ошибок равно C_n^2 и т.д.

Обратимся к проблеме *кодирования*. Если полином $a(x)$ степени k-1, соответствующий входной комбинации информационных символов, умножить на производящий полином

$$s(x)=a(x)g(x), \qquad (3.39)$$

то получим полином, удовлетворяющий условию (3.35) (сравните с формулой (3.21)). Код, формируемый таким образом, не является систематическим.

Для *кодирования в систематической форме* (информационные символы a_j последовательно переносятся в кодовую комбинацию **s** и занимают k первых позиций) можно применить другой метод

$$s(x) = x^r a(x) + [x^r a(x)] \bmod g(x). \qquad (3.40)$$

Операции, определяемые этим выражением, можно детализировать следующим образом:

1) умножение $a(x)$ на x^r означает, что k информационных символов сдвигается на r позиций вправо и в итоге занимают k старших (по степеням x) разрядов; r разрядов слева при этом оказываются нулевыми;

2) вычисляется r-разрядный остаток res(x) от деления полинома $x^r a(x)$, соответствующего полученной таким образом комбинации;

3) эти r элементов остатка помещаются на r нулевых позиций комбинации, полученной в пункте 1, в качестве проверочных символов.

Легко проверить, что такая комбинация удовлетворяет условию (3.35). Коды, полученные по формулам (3.39) и (3.40), эквивалентны, но это разные коды, то есть им соответствуют разные кодовые таблицы. Но поскольку оба кода обладают одним и тем же свойством (3.35), алгоритмы обнаружения и исправления ошибок для них должны быть одинаковы.

Требования, предъявляемые к полиному $g(x)$ кода, способного обнаруживать и исправлять ошибки заданных кратностей, свидетельствуют, что далеко не всякий полином степени r может быть производящим полиномом. В результате математического анализа, основанного на общих положениях алгебры, было показано, что параметры и производящий полином хорошего циклического кода должны обладать следующими свойствами.

Во-первых,

$$n = 2^p - 1, \qquad (3.41)$$

где p – целое число $(p \geq 3)$;

$$r \leq pq_{\text{и}}, \qquad (3.42)$$

где $q_{\text{и}}$ - максимальная кратность гарантированно исправляемых ошибок, при этом кодовое расстояние $d_{код} = 2q_{и} + 1$. При $r << k$ это неравенство обычно обращается в равенство.

Во-вторых,

$$g(x) \bmod b(x) \neq 0 \qquad (3.43)$$

для любого $b(x)$, если $b(x) \neq 1$ и $b(x) \neq g(x)$. Это значит, что многочлен $g(x)$ является неприводимым (простым), то есть не делится без остатка ни на какой другой многочлен, кроме 1 и самого себя.

В-третьих,

$$g(x)h(x) = x^n + 1 = 0, \qquad (3.44)$$

то есть двучлен $x^n + 1$ делится без остатка на $g(x)$ (сравните это с формулой (3.18)).

Коды, обладающие указанным набором свойств, называются кодами БЧХ. Таблицы производящих полиномов для таких кодов даны в ряде книг [2,8]. Некоторые из полиномов приведены в Приложении 1.

Обратите внимание, что при $q_и = 1$, $d_{код} = 3$ для (3.48) имеем $p=r$ и для (3.41) получим

$$n = 2^r - 1. \qquad (3.45)$$

Таким образом, *самый простой циклический код БЧХ эквивалентен коду Хэмминга (3.26) и, следовательно, также является оптимальным и совершенным.*

Полином степени k, входящий в (3.44),

$$h(x) = h_0 + h_1 x + ... + h_k x^k \qquad (3.46)$$

называется *проверочным полиномом* данного циклического кода, причём всегда $h_0 = h_k = 1$. Таблицы проверочных полиномов не приводятся, поскольку его легко можно найти из (3.44).

Вполне естественно, что для кодирования и декодирования вместо полинома $g(x)$ можно использовать полином $h(x)$, при этом *фундаментальное свойство* (3.41) формулируется следующим образом

$$s(x)h(x) \bmod (x^n + 1) = s(x)h(x) = 0, \qquad (3.47)$$

то есть произведение $s(x)h(x)$ делится без остатка на $x^n + 1$, при этом обозначение $\bmod(x^n + 1)$ опущено, поскольку эта операция уже была нами принята в соглашениях (3.31) и (3.33). Кстати, циклический код – это линейный блочный код, поэтому неудивительно, что эта формула напоминает (3.22).

Разумеется, это соотношение также можно использовать для декодирования, то есть для обнаружения ошибок. Итак, и кодер, и декодер можно строить либо на основе производящего полинома, либо на основе проверочного. Во всех случаях кодер и декодер содержат регистр сдвига и сумматоры по модулю 2, причём количество ячеек регистра сдвига равно степени полинома. На практике обычно применяют коды с малой избыточностью, для которых $r << k$, поэтому кодер и декодер выгоднее строить на основе производящего полинома $g(x)$.

Завершая обсуждение общих положений, отметим, что для кодирования циклическим кодом и декодирования не требуется использования даже производящей и проверочной матриц, не говоря уже о кодовой таблице. Тем

не менее, эти матрицы существуют и имеют следующий вид:

$$\mathbf{G}=\begin{bmatrix} g(x) \\ xg(x) \\ x^2g(x) \\ \ldots\ldots\ldots\ldots \\ x^{r-1}g(x) \end{bmatrix}, \qquad \mathbf{H}=\begin{bmatrix} (h(x))_c \\ (xh(x))_c \\ (x^2h(x))_c \\ \ldots\ldots\ldots\ldots \\ (x^{k-1}h(x))_c \end{bmatrix}, \tag{3.48}$$

где $(h(x))_c$ - полином, согласованный с $h(x)$, то есть коэффициенты этого полинома записаны в обратном порядке $h_{cj}=h_{n-1-j}$. Например, при n=7 для $h(x)=1+x+x^2+x^4$ имеем $(h(x))_c=x^2+x^4+x^5+x^6$.

Рассмотрим конкретный **пример**. Для (n,k) – кодов с кодовым расстоянием три, исправляющих все однократные ошибки (циклических кодов Хэмминга), в Приложении 1 приведены следующие возможные значения параметров: (7,4), (15,11), (31,26), и т.д. Проанализируем наиболее короткий код (7,4). Для него даны два варианта производящего полинома: $g(x)=1+x+x^3$ либо $g(x)=1+x^2+x^3$. Коды, построенные на основе любого из них, оказываются разными, но эквивалентными. Выберем для определённости первый полином $g(x)=1+x+x^3$.

Далее найдём проверочный полином из соотношения (3.44), для этого выполним деление $(1+x^n)$ на $g(x)$ столбиком (рис.3.2).

$$\begin{array}{rl|l} & \mathbf{x^7+1} & \mathbf{x^3+x+1} \\ + & \underline{\mathbf{x^7+x^5+x^4}} & \hline \mathbf{x^4+x^2+x+1} \\ & \mathbf{x^5+x^4+1} & \\ + & \underline{\mathbf{x^5+x^3+x^2}} & \\ & \mathbf{x^4+x^3+x^2+1} & \\ + & \underline{\mathbf{x^4+x^2+x}} & \\ & \mathbf{x^3+x+1} & \\ + & \underline{\mathbf{x^3+x+1}} & \\ & \mathbf{0} & \end{array}$$

Рис.3.2. Вычисление проверочного полинома для (7,4)-кода

Таким образом, проверочный полином имеет вид $h(x)=1+x+x^2+x^4$.

Пользуясь способом (3.54), запишем производящую и проверочную матрицы

$$\mathbf{G}=\begin{bmatrix} 1101000 \\ 0110100 \\ 0011010 \\ 0001101 \end{bmatrix}, \quad \mathbf{H}=\begin{bmatrix} 0010111 \\ 0101110 \\ 1011100 \end{bmatrix}. \tag{3.49}$$

Убедитесь, что эти матрицы удовлетворяют соотношению (3.18).

Схема кодера, реализующего метод (3.45) в соответствии с рис. 2.10, приведена на рис. 3.3. Ячейки регистра обнуляются, затем при замкнутом

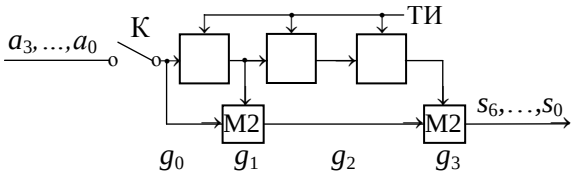

Рис. 3.3. Схема кодера *несистематического* (7,4)-кода на основе производящего полинома

ключе К в кодер вводятся 4 информационных символа, при этом в схеме вычисляются и выводятся 4 первых символа $s_0,...,s_3$ кодовой комбинации. Далее ключ К размыкается, и в течении следующих трёх тактов вычисляются и вы-

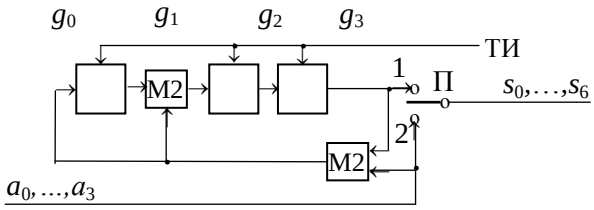

Рис. 3.4. Схема кодера *систематического* (7,4)-кода на основе производящего полинома

водятся значения s_4, s_5, s_6.

Схема кодера систематического кода, реализующая метод (3.34), дана на рис. 3.4.

Это схема деления полинома $x^r a(x)$ на полином $g(x)$, но, в отличие от схемы рис.2.11, она построена несколько иначе, поскольку ориентирована не на вычисление частного, а на вычисление остатка. Регистр обнуляется, переключатель П переводится в положение 2, и в течение 4 тактов в кодер вводят 4 информационных символа. Одновременно эти символы поступают на выход в качестве информационной части кодовой комбинации $s_6 = a_3,..., s_3 = a_0$ (обратите внимание, что ввод и вывод производятся в обратном порядке, начиная с коэффициентов при старших степенях x). В это время в регистре с обратными связями начинается вычисление остатка. Затем переключатель П переводится в положение 1, отключая вход, и в течение следующих 3 тактов

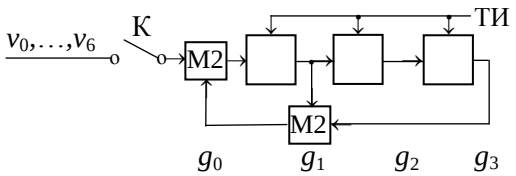

Рис. 3.5. Схема декодера (7,4)-кода на основе производящего полинома

завершается вычисление остатка и его элементы выводятся на месте проверочных символов кодовой комбинации

Схема декодера, приведённая на рис. 3.5, также ориентирована на вычисление остатка от деления полинома $v(x)$ на полином $g(x)$ (формула (3.43)), поэтому она практически не отличается от предыдущей схемы. Регистр обнуляется, ключ К замыкается, и в течение 7 тактов в регистр вводятся 7 символов принятой комбинации, начиная с v_6 (старшего разряда). После завершения ввода в ячейках регистра оказывается записанным значение остатка $\mathrm{res}(x)$. Если $\mathrm{res}(x)=0$ (в 3 ячейках регистра оказались нули), считают, что в принятой комбинации ошибок нет, и на этом декодирование заканчивается.

В противном случае $(\mathrm{res}(x)\neq 0)$, ключ К размыкается (фактически это соответствует ситуации, когда последующие входные символы равны нулю), и подают ещё 7 тактовых импульсов, формально продолжая процесс деления. Номер такта, на котором в ячейках регистра оказывается комбинация 100 (это случай $\mathrm{res}(x)=1$, учитывается обратный порядок следования символов), и есть номер ошибочного символа. После этого исправление этого символа и отделение информационных символов – это тривиальные операции.

Чтобы не было сомнений, заметим, что ожидаемая комбинация 100 в течение 7 тактов обязательно появится, причём один раз. Дело в том, что *при отключении входа схема рис. 3.5 превращается в генератор двоичной псевдослучайной М-последовательности* с периодом (2.21) равным 7. Свойства *М*-последовательности таковы, что в течение периода в ячейках генератора обязательно по одному разу оказываются записанными все *r*-разрядные двоичные комбинации, кроме нулевой, разумеется.

Обратите внимание, что наличие сумматоров на стыках между ячейками регистра сдвига определяется значениями коэффициентов полинома $g(x)$: сумматор есть, если $g_j=1$, и он отсутствует, если $g_j=0$. Таким образом, *зная g(x), можно сразу составить схемы кодера и декодера для любого циклического кода (или схему генератора М-последовательности).*

Как показывает практика, формальное запоминание приведённых соотношений и схем мало что даёт для понимания сути вопроса, если детально не разобрать хотя бы десяток конкретных численных примеров. В частности, если нужно закодировать информационную последовательность 0011, для которой $a(x)=x^2+x^3$, то, проследив такт за тактом работу схемы рис.3.4 и записывая содержимое ячеек, можно убедиться в том, что на выходе появится комбинация 1100010, которой соответствует полином $s(x)=x+x^5+x^6$ (значение этого полинома вычислите отдельно по формуле (3.46) и сравните с комбинацией).

В заключение обязательно нужно отметить следующее обстоятельство. Благодаря циклическому свойству, при кодировании и декодировании циклического кода удаётся организовать конвейерную обработку длинных по-

следовательностей символов в устройствах, содержащих удивительно малое количество простейших логических элементов. Например, для кода (255,247) кодер и декодер содержат по 8 ячеек регистра сдвига и по 4 сумматора по модулю 2. Благодаря этому ценному качеству *среди всевозможных линейных блочных кодов именно циклические коды нашли наиболее широкое применение.*

Более того, в последнее время нашли широкое применение *недвоичные циклические коды – это коды Рида-Соломона*.

Каждые символ кодовой комбинации такого кода является M-ичной цифрой, где, как обычно, $M = 2^m$, причем m – целое число. Как и у любого другого линейного блочного (n,k)-кода, кодовое слово кода Рида-Соломона (обозначается RS(n,k)) содержит n символов, из них k информационных и r проверочных, причем $n = M–1$.

Гарантировано, что при декодировании в кодовом слове будут обнаружены и исправлены $t = r/2$ символов независимо от их расположения внутри кодового слова. Параметр t носит название корректирующей способности кода. Либо могут быть восстановлены r стертых символов.

RS(n,k)-код также полностью определяется производящим полиномом степени r, но коэффициенты полинома – это M-ичные цифры. Кодер и декодер строятся по схеме регистра сдвига с обратными связями, но ячейки регистра рассчитаны на хранение M-ичных цифр, такими же являются и весовые коэффициенты, а суммирование проводится по модулю M.

На практике при использовании RS(n,k)-кода последовательность двоичных символов, поступающих на вход кодера, разбивают на группы, допустим, по $k = 8$ бит в каждой, и долее с каждым байтом оперируют как с 256-ичным символом. В итоге кодовое слово будет фактически содержать $8 \cdot (2^8 - 1) = 2040$ бит. Если код содержит, например, 40 проверочных символов, то он может исправить до 160 ошибочных бит при условии их кучного расположения в байтах. Отсюда видно, что RS(n,k)-код неплохо работает при наличии пакетов ошибок.

3.5 Декодирование в СПИ с каналом переспроса

Рассмотрим ситуацию, когда кроме прямого канала, предназначенного для передачи информации из пункта А в пункт В, существует ещё вспомогательный обратный канал из пункта В в пункт А, называемый каналом переспроса. Единственное назначение этого канала – повысить достоверность передачи информации в прямом канале.

Пакеты, передаваемые в прямом канале, кодируются блочным кодом, и в *пункте В осуществляется лишь обнаружение ошибок в принятой n-разрядной комбинации*. После завершения этой процедуры в канал переспроса передаётся всего один бит, два возможных значения которого имеют, до-

пустим, следующий смысл: 0 – ошибок не обнаружено, нужно передавать следующую комбинацию; 1 – обнаружение ошибки, нужно ещё раз передать данную комбинацию. Возможны такие неблагоприятные ситуации, когда повторная передача будет проводиться более чем один раз.

В такой системе получателю будет выдана ошибочная комбинация, если произойдёт хотя бы одно из событий:

1) в принятой комбинации произошли ошибки слишком высокой кратности $q \geq d_{код}$, которые не способен обнаружить применяемый код (см. формулу (3.5)). Вероятность такого события

$$P_{\text{п}} = \sum_{q=d_{код}}^{n} P(q), \tag{3.50}$$

где $P(q)$ – вероятность того, что ровно q бит в принятой n-разрядной комбинации являются ошибочными. В канале с независимыми ошибками эту вероятность можно найти по биномиальной формуле Бернулли

$$P(q) = C_n^q p^q (1-p)^{n-q}, \tag{3.51}$$

где $C_n^q = \dfrac{n!}{q!(n-q)!}$ - число сочетаний из n по q,

p – вероятность ошибки в одном символе, то есть в длинной последовательности символов, передаваемых в прямом канале, одна ошибка появляется в среднем на $1/p$ символов;

2) произошла ошибка в обратном канале. Обозначим вероятность этого события $P_{\text{о}}$.

В итоге вероятность выдачи ошибочной комбинации найдём по формуле сложения вероятностей

$$P_{\text{ош}} = P_{\text{п}} + P_{\text{о}} - P_{\text{п}}P_{\text{о}}. \tag{3.52}$$

На практике обычно $P_{\text{п}} << 1$, $P_{\text{о}} << 1$, поэтому $P_{\text{ош}} \approx P_{\text{п}} + P_{\text{о}}$.

В СПИ без канала переспроса в пункте В приходится не только обнаруживать, но и исправлять ошибки, при этом вероятность выдачи ошибочной комбинации равна

$$P_{\text{ош1}} = \sum_{q=q_{\text{и}}+1}^{n} P(q), \tag{3.53}$$

где $q_{\text{и}}$ - максимальная кратность исправляемых ошибок (3.6).

Численные расчёты показывают, что при той же величине избыточности $R=r/n$ СПИ с обратным каналом обеспечивает большую помехоустойчивость, если ошибки в прямом канале происходят редко $(pn << 1)$, а обратный канал ещё более надёжен $(P_{\text{о}} < P_{\text{п}})$.

Рассмотрим *пример*. Пусть задано $p = 10^{-4}$, $P_{\text{о}} = 10^{-8}$, в прямом канале применяется код Хэмминга (127,120). Тогда имеем

$$P_{\text{п}} = P(3) + \ldots + P(127) \approx P(3) =$$
$$\frac{125 \cdot 126 \cdot 127}{6} \cdot 10^{-12} \cdot 0,9999^{124} \approx 3,3 \cdot 10^{-7},$$

причём опущены слагаемые $P(4)$, $P(5)$,…, вносящие несущественный вклад. Для итоговой вероятности ошибок имеем

$$P_{\text{ош}} = 3,3 \cdot 10^{-7} + 10^{-8} - 3,3 \cdot 10^{-15} \approx 3,4 \cdot 10^{-7}.$$

В отсутствие канала переспроса тот же код используем для исправления ошибок (всего лишь однократных) и имеем

$$P_{\text{ош1}} = P(2) + \ldots + P(127) \approx P(2) \approx$$
$$\frac{126 \cdot 127}{2} \cdot 10^{-8} \cdot 0,9999^{125} \approx 7,9 \cdot 10^{-5},$$

то есть вероятность ошибки намного выше.

Вероятность повторной передачи комбинации в СПИ с переспросом в нашем примере

$$P_{\text{повт}} = P(1) + P(2) \approx P(1) \approx 1,26 \cdot 10^{-2},$$

то есть в среднем примерно одна из 80 комбинаций будет передаваться повторно, что ненамного увеличивает избыточность, хотя ресурс времени для повторных передач, разумеется, нужно выделять.

Если декодер выдал ошибочную комбинацию, то это будет другая комбинация из кодовой таблицы, скорее всего, наиболее близкая к той комбинации, которая в действительности была передана (вспомним, что в реальных СПИ наиболее вероятно появление ошибок малых кратностей). Следовательно, в такой комбинации из общего количества n символов лишь $d_{код}$ символов являются ошибочными. Тогда *битовая вероятность ошибки на выходе декодера* приближенно равна

$$p_b \approx P_{\text{ош}} \frac{d_{код}}{n}, \qquad (3.54)$$

то есть на выходе декодера одна ошибка появляется в среднем на $1/p_b$ бит.

Естественно, что метод переспроса достаточно широко применяется в современных цифровых СПИ.

3.6 Свёрточные коды

Среди неблочных кодов наибольшее применение нашли сверточные коды. На выходе кодера формируется непрерывная последовательность связанных между собой двоичных символов так, что в этой последовательности нельзя выделить блоки, независимые от других, для того, чтобы декодировать их раздельно.

Тем не менее ради удобства свёрточный код также характеризуют двумя положительными числами n и k. Отношение этих чисел k/n называется *степенью кодирования* $(k < n)$. Это значит, что в бесконечно длинной после-

довательности на выходе кодера на каждые n передаваемых символов приходится k информационных и $r = n - k$ проверочных символов. Избыточность кода при этом, как обычно, определяется отношением $R = r/n$.

Кодер свёрточного кода содержит K-разрядный регистр сдвига, n многовходовых сумматоров mod2 и n-входовый мультиплексор (рис. 3.6). Целое число K называется длиной кодового ограничения по входу.

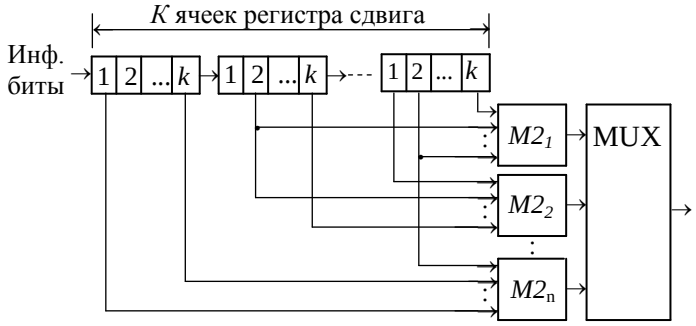

Рис. 3.6. Общая схема кодера сверточного кода со степенью кодирования k/n

В течение k тактов в регистр сдвига быстро вводятся очередные k информационных символов, а затем мультиплексор совершает цикл опроса сумматоров, выдавая очередные n символов на выход кодера. Далее указанный цикл многократно повторяется в той же последовательности.

Способ подключения каждого сумматора к ячейкам регистра сдвига определяется либо вектором связи $g = (g_0, g_1, ..., g_{K-1})$, либо соответствующим ему генераторным полиномом $g(x) = g_0 + g_1 x + ... + g_{K-1} x^{K-1}$, причём обычно на вход одного сумматора подаётся не более чем по одному символу из каждой k-разрядной информационной подпоследовательности.

Таким образом, *для полного описания способа кодирования нужно задать n векторов связи (генераторных полиномов)*. Если k первых генераторных полиномов равны $g_0(x) = 1$, $g_1(x) = x$, ..., $g_{k-1}(x) = x^{k-1}$, получается систематический свёрточный код, у которого из каждых n позиций первые k позиций занимают информационные символы.

В качестве **примера** рассмотрим очень простой систематический свёрточный код, у которого довольно низкая степень кодирования k/n = 1/3, высокая избыточность R = 2/3 и небольшая длина кодового ограничения K=3. Зададим для этого кода три генераторных полинома

$$g_1(x) = 1, \quad g_2(x) = 1 + x, \quad g_3(x) = 1 + x^2, \tag{3.55}$$

из которых $g_1(x)$ генерирует (точнее, тождественно передаёт) информацион-

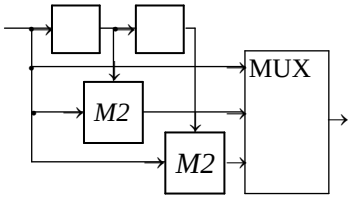

Рис. 3.7. Схема кодера систематического сверточного кода для k/n=1/3

ную последовательность (И), а $g_2(x)$ и $g_3(x)$ генерируют первую (П1) и вторую (П2) проверочные последовательности соответственно. Схема кодера дана на рис. 3.7.

В таблице 3.4 для заданной информационной последовательности показано, каковыми оказываются обе проверочные последовательности и последовательность на выходе кодера.

Существуют разные способы декодирования свёрточных кодов. Один из самых простых – это *метод порогового декодирования*.

Таблица 3.4. Последовательности символов в кодере свёрточного кода k/n=1/3

И	1	0	1	1	0	0	1	…
П1	.	1	1	0	1	0	1	…
П2	.	…	0	1	1	1	1	…
Выход кодера	.	…	110	101	011	001	111	…

Опишем его суть на приведённом примере. На очередном шаге, переходя к рассмотрению очередного информационного символа, пытаются ответить на вопрос, есть ли в этом символе ошибка. Для этого по принятым информационным символам по тем же правилам (рис. 3.7) вычисляют две новые проверочные последовательности. В итоге для анализа имеется пять последовательностей: информационная, две принятые проверочные и две новые проверочные (табл. 3.5).

Таблица 3.5. Последовательности симвлов, участвующие в пороговом декодировании

Принятые	и	и	**и**	и	и	и
	п1	п1	**п1**	**п1**	п1	п1
	п2	п2	**п2**	п2	**п2**	п2
Новые	п1	п1	**п1**	**п1**	п1	п1
	п2	п2	**п2**	п2	**п2**	п2

Далее попарно сравнивают принятые и вычисленные проверочные символы, причём в этой процедуре участвуют лишь те проверочные символы, которые зависят от данного информационного символа (они помечены в табл. 3.5). Выносимое решение (исправлять данный информационный символ или оставить без изменения) определяется количеством совпадающих пар. Если число совпадающих пар меньше порога, равного двум, информационный символ изменяют на противоположный. Рассмотрев различные варианты расположения ошибок в полученной пятёрке принятых символов, следует убедиться, что таким способом выбирается наиболее правдоподобное ло-

кальное решение.

Кстати, этот же метод декодирования можно представить в другой интерпретации: рассматриваются два возможных варианта значений текущего информационного символа (0 или 1) и для каждого варианта вычисляются ожидаемые значения всех пяти символов. Выбирается тот вариант, для которого ожидаемые значения символов в этой пятёрке ближе к тому, что наблюдается в действительности (по количеству совпадений).

Именно эта идея реализуется в более сложных алгоритмах декодирования, причём рассматриваются различные варианты не для одного, а для целой серии предыдущих информационных символов. Ширина окна декодирования – это число последних принятых символов, которое нужно хранить в декодере. С увеличением этой ширины увеличивается объём вычислений, но зато все лучше используются потенциальные возможности кода. Предел улучшения обычно наступает, когда ширина окна существенно превышает полную длину кодового ограничения.

Разумный компромисс обеспечивают хорошие современные алгоритмы декодирования, основанные на том, что анализ выбранного варианта заканчивается досрочно и он отбрасывается, если в процессе анализа уже наблюдаются существенные расхождения ожидаемого с действительным. Один из самых эффективных – алгоритм декодирования Витерби, позволяющий приблизиться к потенциальным характеристикам кода, но у него объем вычислений при декодировании растет экспоненциально с увеличением длины кодового ограничения, поэтому она обычно не более десяти.

Одно из главных достоинств свёрточного кода – простота реализации, позволяющая проводить кодирование с высокой скоростью. Другое достоинство – это способность обнаруживать и исправлять пакеты (вспышки) ошибок, если длина пакета меньше чем K.

Корректирующая способность несистематических кодов несколько выше, но они могут быть катастрофическими, то есть после серии неудачных исправлений может возникнуть бесконечное число ошибок на выходе декодера. Нужно, чтобы n производящих многочленов не имели общих делителей.

Для ослабления подобных эффектов иногда используют следующий приём: непрерывный свёрточный код искусственно превращают в “блочный”, то есть периодически прекращают подачу информационных символов на вход кодера на время, достаточное для того, чтобы вывести содержимое кодера и очистить регистры декодера.

3.7 Перемежение символов

Анализ свойств помехоустойчивых кодов показывает, что корректирующая способность любого блочного кода ограничена. В частности, код мо-

жет правильно воспроизвести кодовую комбинацию лишь в том случае, когда количество ошибочных символов в ней мало, а именно, удовлетворяет условию (3.6).

В канале с независимыми ошибками кодовое расстояние $d_{код}$ всегда можно выбрать так, что подавляющее большинство ошибочных комбинаций будет удовлетворять этому условию, следовательно, ошибки в них будут исправлены. Главное, что избыточность при этом не слишком велика, то есть применение кодирования оказывается эффективным.

В канале с группированием ошибок при том же их общем количестве оказывается, что большинство комбинаций вообще не содержит ошибок. В остальных комбинациях их так много, что код не может их исправить. И в том, и в другом случае применение корректирующего кода оказывается бесполезным и даже вредным: энергию, которая затрачивается на передачу проверочных символов, лучше было бы целиком вложить в информационные символы, то есть вообще отказаться от помехоустойчивого кодирования. Подобный эффект производит и воздействие импульсных помех.

То же самое, правда, в более завуалированной форме, происходит и при использовании сверточных кодов.

Из сказанного ясно, что если уж использовать помехоустойчивое кодирование, то очень полезным был бы способ, позволяющий ошибки из пакетов более или менее равномерно распределять по кодовым комбинациям. Именно для этого предложен метод, который носит название перемежение (interliving).

Наиболее широко применяется перемежение по времени (последовательный способ передачи), хотя в последнее время в системах радиосвязи начинает использоваться перемежение по частоте (параллельный способ передачи).

Весьма популярен табличный способ перемежения по времени. Для этого n-разрядные комбинации с выхода кодера записываются в виде строк прямоугольной таблицы. После заполнения m-строк считывание и передача в линию производится по столбцам. В приемнике осуществляется обратное преобразование: запись по столбцам, считывание и подача на вход декодера – по строкам. Это эквивалентно перемещению символов во времени, показанному на рис. 3.8. В показанном примере, если в линии возникает пакет,

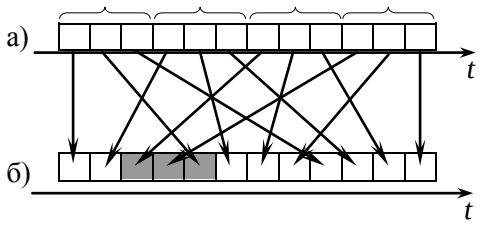

Рис. 3.8. Схема, иллюстрирующая табличный способ перемежения по времени (n = 3, m = 4):
а) последовательность символов с выхода кодера;
б) последовательность символов, подаваемых в линию

состоящий из $q = 3$ ошибок (эти позиции затемнены), то после деперемежения окажется лишь по одной ошибке в 1, 3 и 4 комбинациях.

Если число строк m в таблице больше, чем максимальная возможная длина пакета ошибок, то в каждой кодовой комбинации окажется не более одного ошибочного символа. Если используется хотя бы код Хэмминга, то все они будут исправлены.

Перемежение по частоте имеет смысл проводить лишь в том случае, когда на разных ортогональных несущих (см. разд. 2.12) одновременно передаются символы не одной, а нескольких кодовых комбинаций.

Строго говоря, само по себе перемежение не является способом помехоустойчивого кодирования, так как избыточность при этом не увеличивается. Следовательно, польза от перемежения может быть лишь тогда, когда оно проводится после кодирования корректирующим кодом.

3.8 Комбинирование кодов

Рассмотрим ситуацию, когда передаваемое сообщение участвует не в одном, как это предполагалось ранее, а в нескольких актах помехоустойчивого кодирования. Существуют последовательные и параллельные методы комбинирования кодов.

Код, получающийся в результате *последовательной* реализации нескольких процедур кодирования (кодами одного типа или разными), называется *каскадным*. Допустим, на каждом этапе применяются линейные блочные коды. Тогда первый кодер, получив k информационных символов, добавляет к ним r_1 проверочных символов и все вместе подает на вход второго кодера в качестве информационных символов. Тот, в свою очередь, добавляет к ним еще r_2 проверочных символов и т.д. Первый код в этой цепи называется внутренним кодом.

В итоге после прохождения цепочки из p кодеров на выходе имеем комбинацию линейного блочного кода, содержащую те же k информационных символов и $r = r_1 + \ldots + r_p$ проверочных символов. Возрастают избыточность кода и его корректирующая способность.

Пример 1. Двумерный композиционный код получается следующим образом. Прямоугольная таблица, содержащая n_1 столбцов и n_2 строк, делится на четыре блока. Левый верхний блок (k_1 столбцов и k_2 строк) заполняется информационными символами от источника. Затем каждая строка отдельно кодируется (n_1,k_1) линейным блочным кодом, при этом $r_1 = n_1 - k_1$ проверочных символов каждой строки помещают в правый верхний блок. Далее каждый столбец кодируется (n_2,k_2)-кодом. В итоге получим новый (n,k)-код, где $n = n_1n_2$, $k = k_1k_2$. На рис. 3.9 приведен пример кодирования, когда на обоих этапах используется код с проверкой на четность.

1 0 0 1 1 0 1	0
0 0 0 1 1 1 0	1
1 1 0 0 1 1 0	0
0 1 0 0 1 0 1	1

Рис. 3.9. Пример кодирования двумерным композиционным кодом (32,21)

Одиночная ошибка нарушает условие четности в соответствующих строке и столбце, поэтому ее положение может быть определено, но две ошибки код уже исправить не может. Однократное кодирование кодом с проверкой на четность позволяет лишь обнаруживать одиночные ошибки, но не исправлять их. Итак, применение двукратного кодирования позволило повысить корректирующую способность кода. При этом интересно посмотреть, велика ли цена. Из Приложения 1 мы видим, что примерно те же значения параметров имеет (31,21)-код БЧХ, но он способен исправлять и любые двукратные ошибки.

Пример 2. Поступающие на вход кодера $k = 2^r - 1 - r$ информационных символов кодируются кодом Хэмминга, а затем к полученным $n = k + r$ символам добавляется еще один проверочный символ, чтобы организовать общую проверку на четность. В итоге кодовое расстояние кода повышается с трех до четырех, и он приобретает способность обнаруживать ошибки до трехкратной включительно.

Декодирование кода также проводится в два этапа. На первом этапе принятая комбинация подвергается общей проверке на четность, в результате которой получают ответ на вопрос, четной или нечетной является кратность ошибок. Затем последний проверочный символ отбрасывается, и по оставшимся символам вычисляется синдром по правилам декодирования кода Хэмминга. Сопоставление результатов на обоих этапах позволяет даже определить кратность ошибки (сам по себе код Хэмминга на это не способен) и в зависимости от этого предпринять нужные действия (если ошибка однократная, код Хэмминга ее может исправить; если двукратная или трехкратная, нужно сделать запрос на повторную передачу комбинации). Снова отметим, что такими же характеристиками обладает код Рида-Малера с кодовым расстоянием, равным четырем.

Множество примеров показывает, что всегда можно найти такой код, который при тех же значениях *n* и *k* при использовании одноэтапной процедуры кодирования обеспечивает лучшую (или не худшую) помехоустойчивость, чем каскадный код, формируемый в несколько этапов. Зачем же тогда применяются каскадные коды? Лишь потому, что на каждом этапе можно использовать относительно простые и короткие коды и менее трудоемкие методы кодирования и декодирования. В частности, при записи на компакт-

диск требуется применять код, способный исправлять пакеты, содержащие тысячи ошибок. И он реализуется путем многократного последовательного выполнения операций перемежения и кодирования кодом Рида-Соломона.

Возможно и *параллельное использование нескольких процедур кодирования*. Здесь последовательность информационных символов одновременно кодируется, допустим, двумя систематическими кодами. Лучше, если на вход второго кодера она будет подаваться после перемежения. Далее по линии передаются сами информационные символы и проверочные символы с выходов обоих кодеров. Очевидно, что в отличие от кодирования, имеет смысл лишь *совместная работа декодеров путем взаимного обмена информацией в процессе декодирования*. Именно на этих принципах построено использование *турбо-кодов*. Дальнейшее развитие этих идей дано в разд. 5.7.

Вопросы для самоконтроля по главе 3

1. Чем определяется корректирующая способность кода? Поясните на примере.
2. Какие коды называются корректирующими?
3. Что значит “обнаружить ошибки” при декодировании кодовой комбинации?
4. Что значит “исправить ошибки” при декодировании кодовой комбинации?
5. Каков характерный признак, позволяющий отличить кодовую таблицу линейного блочного кода от кодовых таблиц других кодов?
6. Что такое проверочная матрица линейного блочного кода? Как она используется при обнаружении ошибок в принятой комбинации?
7. Каков характерный признак, позволяющий отличить кодовую таблицу циклического кода от кодовых таблиц других кодов?
8. Чему равно количество комбинации в кодовой таблице линейного блочного кода?
9. Почему в проверочной матрице не может быть нулевых: столбцов? строк?
10. Какой смысл имеют строки проверочной матрицы?
11. По каким признакам можно определить, что проверочная матрица принадлежит коду, способному исправить любую одиночную ошибку?
12. Чем обусловлена популярность циклических кодов? Из каких логических элементов состоят кодер и декодер?
13. В чем заключается фундаментальное свойство комбинаций циклического кода?
14. Может ли помехоустойчивый код быть безизбыточным?
15. Почему декодирование по минимуму расстояния применяется

редко?

16. Являются ли сверточные коды блочными и чем обусловлена их популярность?

17. Какова цель перемежения символов?

18. Какие способы комбинирования кодов используют в системах связи?

4 КОДИРОВАНИЕ ИСТОЧНИКА

4.1 Собственная информация и избыточность (цифровые сигналы)

Сначала определим, сколько информации содержится в сообщении X на выходе источника информации. Сообщение имеет цифровую форму, и ради упрощения обозначений полагаем, что X – это m-ичный символ, а $x_1, x_2, \ldots, x_m$ – его возможные значения.

Чтобы исключить влияние остальных элементов тракта передачи на рис. 2.1 (передатчик, линия, приемник), полагаем, что получатель имеет возможность непосредственно наблюдать сообщение, выдаваемое источником, то есть $Y=X$. Конечная цель получателя заключается в том, чтобы определить, какое именно значение x_j появилось на выходе источника. Эту задачу он может решить столь же успешно, если находится на другом конце цифрового *канала без помех*, в котором каждому из возможных сигналов y_k соответствует единственное значение x_j, поэтому по принятому сигналу всегда можно безошибочно восстановить переданное сообщение. Например, канал, в котором производится тождественное преобразование передаваемого сообщения $x_j \rightarrow y_j$, является самой простой и наглядной формой канала без помех.

Пусть задано полное вероятностное описание переданного сообщения X (его ряд распределения)

x_j	x_1	x_2	…	x_m
$p(x_j)$	$p(x_1)$	$p(x_2)$	…	$p(x_m)$

(4.1)

Даже при этом *до передачи сообщения при $m \geq 2$ всегда существует неопределенность относительно того, какое именно значение появится на выходе источника*. После появления конкретного значения, например x_3, неопределенность исчезает, и объем знаний наблюдателя возрастает (не знал – узнал).

Количество информации, получаемой наблюдателем при появлении значения x_j, Клод Шеннон предложил определять по формуле

$$I(x_j) = -\log_a p(x_j) \qquad (4.2)$$

и назвал эту величину *собственной информацией* символа x_j.

Выбор основания логарифма a определяет единицу количества информации. Если a=2, то единица называется двоичной (бит), при $a = e \approx 2{,}72$ –

натуральной, а при a = 10 – десятичной. Переход от одной системы логарифмов к другой равносилен изменению единицы количества информации

$$\log_b p = \log_b a \cdot \log_a p, \qquad (4.3)$$

то есть 1 нат. ед.=$\log_2 e$ бит≈1,443 бит,

1 дес. ед.=$\log_2 10$ бит≈3,32 бит.

В техники связи наиболее широко применяются двоичные сигналы, обладающие равновозможными состояниями x_1 и x_2. Тогда количество информации, содержащейся в любом из этих сигналов, одинаково и равно $I(x_1) = I(x_2) = -\log(1/2) = 1$ бит.

Поэтому двоичные единицы информации используются наиболее часто. В дальнейшем по умолчанию мы также будем использовать двоичные единицы (отметим, что не следует путать *два различных значения слова "бит"*: 1) двоичный сигнал; 2) двоичная единица количества информации).

Определение (4.2) не является единственно возможным. Предложены и другие способы определения количества информации (см. разд. 4.6), но к настоящему времени именно теория информации по Шеннону позволила получить наибольшее количество полезных для практики результатов.

Свойства собственной информации:

1) $I(x_j)\geq 0$, причем знак равенства имеет место лишь в ситуации, когда X – детерминированный сигнал, т.е. $p(x_1)=1$, следовательно, любые другие состояния сигнала невозможны (m=1);

2) величина собственной информации тем больше, чем менее вероятно данное состояние;

3) обычно заранее не известно, сколько информации будет содержать сигнал, поскольку заранее не известно, какое именно значение x_j появится на выходе источника.

Последнее обстоятельство позволяет саму величину $I_j=I(x_j)$ считать одним из возможных значений некоторой дискретной случайной величины, имеющей ряд распределения

I_j	I_1	I_2	…	I_m
$p(x_j)$	$p(x_1)$	$p(x_2)$	…	$p(x_m)$

. (4.4)

Например, пусть четверичный сигнал с двукратной ФМ характеризуется значениями начальной фазы

$x_j=\varphi_j$	0°	90°	180°	270°
$p(x_j)$	0,10	0,60	0,18	0,12

(4.5)

тогда ряд распределения для величины собственной информации имеет вид

I_j=-log$p(x_j)$	3,32	0,74	2,47	3,06
$p(x_j)$	0,10	0,60	0,18	0,12

.

Математическое ожидание случайной величины “собственная информация” равно

$$H(X) = \mathrm{M}\left[-\log p(X)\right] = -\sum_{j=1}^{m} p(x_j)\log p(x_j) \tag{4.6}$$

и называется *энтропией сигнала X*.

Для примера (4.5) имеем $H(X)$=–0,1×log0,1–0,6×log0,6––0,18×log0,18–0,12×log0,12= 1,587 бит.

Чтобы упростить подобные однотипные вычисления, в Приложении 2 дана таблица значений функции $h(p)= -p\log(p)$, а на рис. 4.1 – ее график.

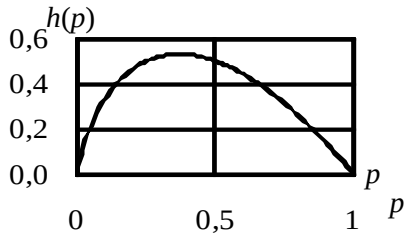

Рис. 4.1. Вспомогательная функция $h(p)= -p\log_2 p$

Свойства энтропии:

1) $H(X)\geq 0$, причем знак равенства имеет место лишь при детерминированном сигнале X (это легко показать, убедившись в том, что $h(p)=0$ лишь при $p=0$ и при $p=1$);

2) *сигнал X, имеющий m возможных состояний, обладает максимальной энтропией*

$$H(x)\leq H_{max}=\log m \tag{4.7}$$

в случае, когда все эти состояния равновероятны, т.е. $p(x_j)=1/m$.

В справедливости этого утверждения можно убедиться, проверив, что “выравнивание” вероятностей любых двух состояний ведет к увеличению энтропии. Действительно, имеем

$$2h\left(\frac{p_1+p_2}{2}\right) \geq h(p_1)+h(p_2), \tag{4.8}$$

поскольку функция $h(p)$ – выпуклая.

Эти два свойства показывают, что энтропия, в дополнение к своему основному назначению (среднее количество собственной информации в символе X) может использоваться как *мера неопределенности исхода опыта над случайным объектом X*.

Энтропия $H(X)$ – детерминированная величина, поэтому ее удобнее использовать для описания информационного содержания сигналов, нежели

набор значений (4.2).

Чтобы увеличить количество возможных вариантов N, передаваемое сообщение обычно формируют в виде последовательности длины n, состоящей из m-ичных символов $\mathbf{X} = X^{[1]}, X^{[2]}, ..., X^{[n]}$, где каждый символ может принять одно из значений $x_1, x_2, ..., x_m$, при этом $N = m^n$.

Собственная информация последовательности $x_i^{[1]}, x_j^{[2]}, ..., x_k^{[n]}$ определяется тем же способом (4.2)

$$I(x_i^{[1]}, x_j^{[2]}, ..., x_k^{[n]}) = -\log p(x_i^{[1]}, x_j^{[2]}, ..., x_k^{[n]}) \qquad (4.2a)$$

и, естественно, обладает теми же свойствами. К ним можно добавить лишь *свойство аддитивности*, вытекающее из формулы умножения вероятностей,

$$\begin{aligned} I(x_i^{[1]}, x_j^{[2]}, x_q^{[3]}, ..., x_r^{[n-1]}, x_k^{[n]}) &= -\log p(x_i^{[1]}) - \\ &- \log p(x_j^{[2]} / x_i^{[1]}) - \log p(x_q^{[3]} / x_i^{[1]}, x_j^{[2]}) - ... - \\ &- \log p(x_k^{[n]} / x_i^{[1]}, x_j^{[2]}, ..., x_r^{[n-1]}) = \\ &= I(x_i^{[1]}) + I(x_j^{[2]} / x_i^{[1]}) + I(x_q^{[3]} / x_i^{[1]}, x_j^{[2]}) + ... + \\ &+ I(x_k^{[n]} / x_i^{[1]}, x_j^{[2]}, ..., x_r^{[n-1]}), \end{aligned} \qquad (4.9)$$

где обозначение $I(a/\mathbf{b})$ использовано для определения *условной собственной информации*, т.е. информации, содержащейся в символе a при условии, что к этому времени наблюдателю уже известны значения всех символов, входящих в группу $\mathbf{b}$.

Энтропия последовательности символов также определяется способом (4.6)

$$\begin{aligned} H(\mathbf{X}) &= H(X^{[1]}, X^{[2]}, ..., X^{[n]}) = \\ &= -\sum_{i=1}^{m}\sum_{j=1}^{m}...\sum_{k=1}^{m} p(x_i^{[1]}, x_j^{[2]}, ..., x_k^{[n]}) \log p(x_i^{[1]}, x_j^{[2]}, ..., x_k^{[n]}). \end{aligned} \qquad (4.6a)$$

Вычисляя формально математические ожидания обеих частей равенства (4.9), видим, что и энтропия последовательности обладает свойством аддитивности

$$\begin{aligned} H(\mathbf{X}) &= H(X^{[1]}) + H(X^{[2]} / X^{[1]}) + ... + \\ &+ H(X^{[n]} / X^{[1]}, X^{[2]}, ..., X^{[n-1]}). \end{aligned} \qquad (4.10)$$

Уточняя свойство экстремальности (4.7), отметим, что и энтропия последовательности максимальна

$$H(\mathbf{X}) \le H_{n\max} = \log N = n \log m = nH_{\max}, \qquad (4.11)$$

когда все N состояний равновероятны. Это возможно лишь в том случае, когда и *для каждого символа все его m состояний равновероятны, а символы в последовательности независимы.*

С этим свойством связано весьма важное понятие **избыточности сигнала**. Сигнал $\mathbf{X}$ обладает избыточностью, если количество информации (эн-

тропия) $H(\mathbf{X})$, содержащейся в нем, меньше того максимального количества $H_{n\max}$,которое он в принципе мог бы содержать при тех же m и n. Коэффициент избыточности определяется по формуле

$$R = 1 - H(\mathbf{X}) / H_{n\max} \qquad (4.12)$$

и может принимать значения $0 \le R \le 1$. Иногда удобнее использовать другую формулу, приводящую к тому же результату,

$$R = 1 - L_{\min} / L = L_r / L, \qquad (4.13)$$

где L – среднее количество символов в сообщении (среднее значение может быть даже дробным);

$L_{\min}$ – минимальное среднее количество символов, необходимое для того, чтобы вместить то же количество информации, следовательно, величина L_r есть среднее количество избыточных (“лишних”) символов.

Несмотря на внешнее сходство формул (3.25) и (4.13), они определяют избыточность по-разному: в первом случае – в техническом смысле, а во втором – в более общем, информационном смысле. Результаты вычислений по обеим формулам для сигнала на выходе декодера совпадают в единственном случае, когда двоичный сигнал на его входе не содержит избыточности, то есть все его $N=2^k$ комбинаций равновероятны.

Избыточность письменных текстов в любом из европейских языков довольно велика: она составляет 70–80%. Использование различных сокращений есть способ уменьшения избыточности. В иных случаях избыточность сознательно увеличивают, например, указывая сумму прописью в финансовых документах. Использование корректирующих кодов также основано на введении в передаваемый сигнал избыточных, проверочных символов.

Из экономических соображений всегда хотелось бы сформировать передаваемый сигнал, не содержащий избыточности, однако ставить это главной целью при кодировании допустимо лишь при передаче сигнала в канале без помех. В канале с помехами передаваемый сигнал должен содержать избыточность тем большую, чем выше уровень помех. Именно благодаря этой избыточности удается обеспечить требуемое качество передачи за счет того, что в процессе приема появляется возможность обнаруживать и исправлять большинство ошибок, возникающих в канале.

4.2 Кодирование в канале без помех

Сообщение на выходе любого реального источника информации (текст, речь, изображение и т.п.) обладает избыточностью. Нельзя сказать, что она бесполезна (см. вышеприведенные заметки о роли избыточности), но в подавляющем большинстве случаев введена далеко не самым рациональным образом. Поэтому перед тем, как проводить какие-то преобразования сообщения, лучше от этой избыточности избавиться.

Преобразование цифрового сигнала, направленное на сокращение из-

быточности, называется кодированием источника. Раньше тот же процесс называли кодированием в дискретном канале без помех.

Например, при двоичном кодировании алфавита, содержащего $s = 5$ букв, кодовая таблица может иметь следующий вид (табл. 4.1). Для каждой буквы x_j в таблице указана вероятность ее появления $p(x_j)$ и соответствующая кодовая комбинация, а также длина l_j этой комбинации (в общем случае комбинации могут содержать разное количество символов).

Обычно кодовая таблица используется многократно (для передачи многобуквенного текста), и наиболее объективным показателем экономичности кода является средняя длина кодовой комбинации

$$L = \sum_{j=1}^{s} l_j p(x_j) = 3 \cdot 0,2 + ... + 6 \cdot 0,1 = 3,2 \text{ бит / букву.} \qquad (4.14)$$

При этом избыточность сигнала на выходе кодера можно определить по формуле (4.13), полагая, что $L_{\min}$ есть минимально возможное среднее количество m-ичных символов, приходящихся на одну букву текста.

Попытаемся найти ответы на два вопроса:

1) каково численное значение теоретического предела $L_{\min}$?

2) как нужно кодировать, чтобы достигнуть этого предела или хотя бы к нему приблизиться?

Кстати, очевидно, что для алфавита табл. 4.1 существуют лучшие коды, нежели код, указанный в примере.

Таблица 4.1. Пример кодовой таблицы

x_j	$p(x_j)$	Кодовые слова	l_j
а	0,2	101	3
б	0,5	010	3
в	0,1	0	1
г	0,1	1001	4
д	0,1	110000	6

Проведем простые, хотя и не очень строгие, рассуждения, позволяющие ответить на первый вопрос.

Собственная информация буквы x_j равна $I(x_j) = -\log p(x_j)$. Столько же информации должно содержать j-е кодовое слово.

С другой стороны, максимальное среднее количество информации, которое может перенести один m-ичный символ на выходе кодера, в соответствии с (4.7) равно $H_{\max} = \log m$. Тогда количество таких символов, требуемое для размещения имеющейся информации, должно удовлетворять очевидному неравенству

$$l_j \geq \frac{I(x_j)}{H_{\max}} = \frac{-\log p(x_j)}{\log m}. \qquad (4.15)$$

Неравенство справедливо и для математических ожиданий обеих частей

$$\sum_{j=1}^{s} l_j p(x_j) \geq \frac{1}{\log m}\left[-\sum_{j=1}^{s} p(x_j)\log p(x_j)\right]. \qquad (4.16)$$

Тогда на основании определений (4.6) и (4.14) имеем

$$L \geq \frac{H(X)}{\log m} = L_{\min}. \qquad (4.17)$$

Таким образом, ответ на первый вопрос получен.

Выясним, при каких условиях неравенство (4.17) может обратиться в равенство, т.е. когда существует безызбыточный код. Для этого необходимо, чтобы каждое слагаемое в (4.16) обратилось в равенство, т.е. обратились в равенства все *s* выражений (4.15). Левая часть l_j – это количество символов в *j*-й комбинации, т.е. положительное целое число. Поэтому равенство в (4.15) возможно лишь в том случае, когда и правая часть – положительное целое число. В итоге равенство в (4.17) возможно лишь в случае, когда распределение вероятностей букв таково, что каждая из вероятностей может быть представлена в виде отрицательной целочисленной степени основания кода *m*

$$p(x_j) = m^{-l_j}. \qquad (4.18)$$

Приведенные соотношения отображают содержание доказанной строго *теоремы Шеннона о кодировании в дискретном канале без помех.*

При кодировании алфавита X с энтропией H(X) средняя длина кодовой комбинации не может быть меньше H(X)/logm. Если вероятности букв не являются отрицательными целочисленными степенями числа m, то точное достижение указанной нижней грани невозможно, но при кодировании достаточно длинными блоками к ней можно сколь угодно приблизиться.

Последнее утверждение теоремы, о кодировании блоками, будет проиллюстрировано позже.

Формула (4.15) дает лишь весьма общее указание о способе кодирования: более вероятным буквам должны соответствовать более короткие кодовые комбинации. Это – *принцип статистического кодирования.*

Предложено несколько способов кодирования с малой избыточностью. *Наилучший из них –* ***код Хафмана***. Величина *L*, получаемая при кодировании этим кодом, не больше, чем при кодировании любым другим кодом.

Процедура кодирования двоичным ($m = 2$) кодом Хафмана предусматривает ($s - 1$)-кратное повторение одной и той же операции, включающей следующие действия:

1) буквы алфавита, полученного на предыдущем шаге, располагаются в порядке убывания их вероятностей;

2) проводится сокращение алфавита на одну единицу, т.е. две наименее вероятные буквы объединяются и заменяются одной новой буквой, для кото-

рой вероятность ее появления, естественно, равна сумме вероятностей объединяемых букв.

Процесс кодирования удобно графически иллюстрировать при помощи горизонтально расположенного дерева (слева – ветви, справа – корень), у которого всегда две ветви объединяются в одну, более крупную. Объединяемые ветви обозначаются двоичными цифрами: верхняя – 1, нижняя – 0. Чтобы записать кодовое слово, соответствующее данной букве, нужно двигаться к ней от корня дерева и считывать эти двоичные цифры.

Пример. Закодировать буквы алфавита, приведенного в табл. 4.2, и оценить избыточность полученного кода.

Таблица 4.2 – Пример ряда распределения вероятностей букв алфавита

x_j	x_1	x_2	x_3	x_4	x_5	x_6
$p(x_j)$	0,1	0,2	0,25	0,05	0,25	0,15

Энтропия источника сообщений равна

H(X) = –0,1 × log0,1 – … – 0,15 × log0,15 = 2,423 бит/букву.

Очевидно, что из-за неравновероятности букв алфавита сообщение на выходе источника имеет избыточность

$$R_s = 1 - H(X)/H(X)_{\max} = 0,0627,$$

где $H(X)_{\max} = \log s = 2,585$ бит/букву.

Кодовое дерево показано на рис. 4.2, где в разрывах ветвей указаны суммарные вероятности объединяемых букв.

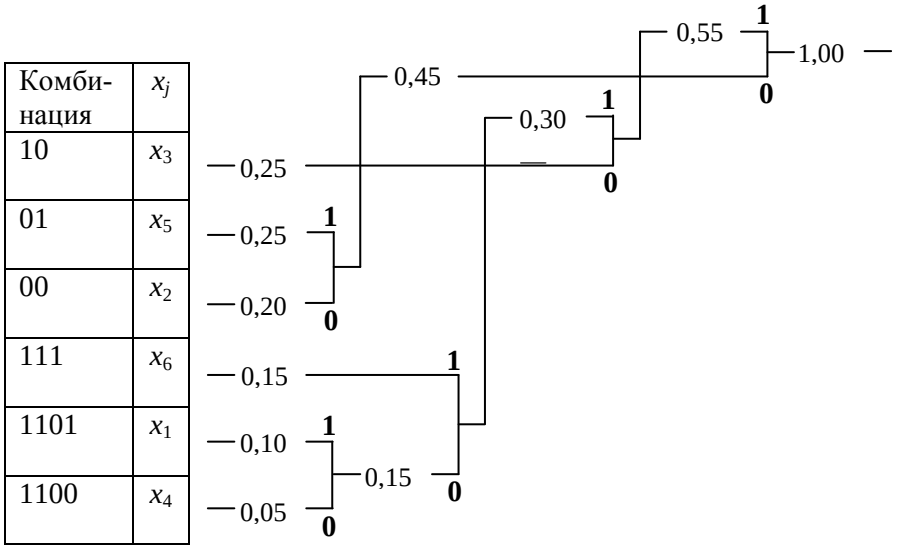

Рис. 4.2. Кодовое дерево кода Хафмана

Средняя длина кодового слова (4.14) на выходе кодера равна $L = 2 \times 0,25 + \ldots + 4 \times 0,05 = 2,45$ (бит/букву), и избыточность (4.13) этого сигнала

составляет $R = 1 – 2{,}423/2{,}45 = 0{,}011$.

Итак, алфавит источника удалось закодировать с малой избыточностью, даже меньшей, чем избыточность входного сигнала.

Следует отметить, что использование кода Хафмана позволяет получить сигнал с минимально возможной избыточностью, но из этого не следует, что избыточность сигнала на выходе кодера всегда будет меньше, чем избыточность исходного алфавита. Возможна такая, на первый взгляд парадоксальная ситуация, когда все s букв алфавита равновероятны и, следовательно, входной сигнал не содержит избыточности, но выходной сигнал избыточен. Поэтому еще раз подчеркнем, что кодирование с нулевой избыточностью возможно лишь для распределений вероятностей специального вида, а именно, когда вероятности всех букв могут быть представлены в форме (4.18).

Другой весьма хороший код – это **код Шеннона-Фано**. Здесь также буквы алфавита предварительно располагаются в порядке убывания вероятностей. Кодирование двоичным кодом проводится в несколько этапов:

1) весь алфавит делится на две группы (верхнюю и нижнюю) так, чтобы суммарные вероятности в каждой из групп были по возможности равны друг другу (к сожалению, если вероятности всех букв не могут быть представлены в виде (4.18), при выполнении этой операции возможен субъективный подход); затем для каждой буквы записывается первый символ кодового слова: 1– для букв, попавших в верхнюю группу, и 0 – для букв из нижней группы;

2) каждая группа делится, в свою очередь, на две равновероятные подгруппы, и в каждом кодовом слове записывается второй символ по тому же принципу.

Такое деление и дописывание символов проводится до тех пор, пока в каждой подгруппе не останется по одной букве.

Если $m > 2$, на каждом шаге проводится деление группы букв на m равновероятных подгрупп и для их нумерации используются символы $0,1,...,(m–1)$.

Обратите внимание, что при кодировании и кодом Хафмана, и кодом Шеннона-Фано в самой структуре кодовой таблицы уже заложена возможность разделения при приеме последовательно передаваемых кодовых комбинаций. Например, при кодировании кодом рис. 4.2 принятая двоичная последовательность 0011000100111однозначно декодируется как последовательность букв $x_2x_4x_5x_2x_6$. Это обусловлено тем, что в кодовой таблице никакая комбинация не является началом другой, более длинной комбинации.

При использовании обоих кодов любая, даже одиночная ошибка в принятой последовательности лишает приемник возможности правильно декодировать ее оставшуюся часть, но эту проблему обсуждать не будем, поскольку, напомним, эти коды предназначены исключительно для кодирования при отсутствии помех (реально, в ситуации, когда возможность появле-

ния ошибки пренебрежимо мала).

Кодирование блоками, упоминаемое в теореме Шеннона, предполагает использование тех же кодов, но кодовая таблица содержит не отдельные буквы передаваемого текста, а их комбинации (двухбуквенные, трехбуквенные и т.д.), при этом, естественно, нужно указать вероятности появления этих комбинаций. При кодировании n-буквенных комбинаций объем кодовой таблицы равен $N = s^n$, т.е. экспоненциально растет с увеличением длины блоков. Положительный эффект при этом заключается в том, что вероятности всех n-буквенных блоков все более точно могут быть представлены в форме (4.18), т.е. избыточность может быть сделана сколь угодно малой.

Главный недостаток обоих кодов заключаются в том, что для их применения нужно знать вероятности появления букв (или их комбинаций при кодировании блоками), а на практике такая благоприятная ситуация возникает чрезвычайно редко.

Код Лемпела-Зива [1–3] свободен от этого недостатка. Здесь кодовая таблица, изначально почти пустая, заполняется одновременно в пунктах передачи и приема в процессе кодирования (декодирования), причем в эту таблицу вносятся лишь такие все более длинные отрезки передаваемого сообщения, которые еще не встречались ранее. Каждому отрезку в таблице присваивается n-разрядный номер. При внесении очередной записи (строки) в таблицу передается блок, содержащий:

1) номер отрезка, уже имеющегося в таблице;

2) символ, следующий в передаваемом сообщении за этим отрезком.

Рассмотрим пример кодирования двоичным кодом Лемпела-Зива двоичного сообщения 0010101000101. Сообщение короткое, поэтому можно взять небольшое значение $n = 3$.

На рис. 4.3 показаны двоичные последовательности на входе и выходе кодера, причем ради наглядности они разбиты на части, соответствующие отдельным шагам. Сплошные линии указывают, какие номера извлекаются из кодовой таблицы и подаются на выход кодера. Штриховые линии показывают путь внесения очередных записей в кодовую таблицу (первая запись “пробел” и его номер 000 внесены в таблицу заранее). Убедитесь, что по передаваемой последовательности в пункте приема можно в том же порядке заполнять кодовую таблицу и декодировать сигнал.

Для дальнейшего сокращения избыточности полезно те отрезки в кодовой таблице, для которых уже существуют всевозможные более длинные по-

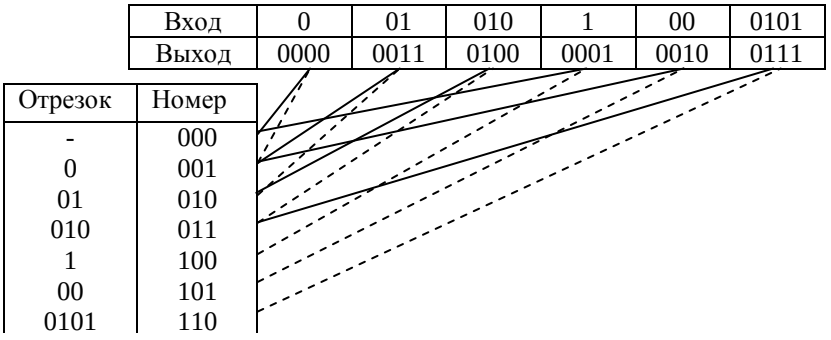

Рис. 4.3. Пример кодирования кодом Лемпела-Зива

следовательности (длина больше на единицу), исключать из таблицы и повторно использовать освободившиеся номера. Например, поле записи отрезка 00 можно было бы исключить отрезок 0, т.к. для него уже существуют более длинные 01 и 00, но мы эту операцию не выполняли.

Из приведенного примера видно, что код Лемпела-Зива даже увеличивает избыточность при кодировании коротких сообщений (на входе было 13 бит, а на выходе стало 24), однако при кодировании более длинных сообщений это соотношение постепенно улучшается. Доказано, что асимптотически (при увеличении длины сообщения) избыточность этого кода стремится к нулю.

Код Лемпела-Зива является основой многих процедур архивирования файлов.

Все рассмотренные выше методы – это *универсальные методы, позволяющие кодировать любое цифровое сообщение без потери информации*.

Разработаны также *специализированные методы*, каждый из которых предназначен для кодирования сообщений определенного вида (голосовых, видеоизображений и т.п.) и основан на учете особенностей данного вида [2–4]. Кодирование такими методами производится с частичной потерей информации, что выливается в погрешности при восстановлении сообщения в приемнике. Можно увеличить степень сжатия за счет увеличения погрешностей.

4.3 Взаимная информация

В канале с помехами возможны ситуации, когда принятое значение m-ичного символа y_k отличается от переданного значения x_j, хотя в нормальных условиях все-таки более вероятно, что $k=j$. В принципе, возможно появление любой пары x_j, y_k.

Количество информации, извлекаемой получателем из принятого сигнала y_k, когда в действительности передавалось сообщение x_j, К. Шеннон предложил определять по формуле

$$\boldsymbol{I}(x_j; y_k) = \log \frac{\boldsymbol{p}(x_j / y_k)}{\boldsymbol{p}(x_j)}. \qquad (4.19)$$

Отметим некоторые очевидные свойства величины (4.19).

1) Свойство симметрии

$$\boldsymbol{I}(x_j; y_k) = \boldsymbol{I}(y_k; x_j) = \log \frac{\boldsymbol{p}(x_j / y_k)}{\boldsymbol{p}(x_j)} =$$

$$= \log \frac{\boldsymbol{p}(y_k / x_j)}{\boldsymbol{p}(y_k)} = \log \frac{\boldsymbol{p}(x_j, y_k)}{\boldsymbol{p}(x_j)\boldsymbol{p}(y_k)} \qquad (4.20)$$

вытекает из формулы умножения вероятностей (1.8). Вследствие того, что значение переданного сообщения x_j и значение принятого сигнала y_k входят в (4.20) одинаковым образом, эта величина получила более короткое название

взаимной информации между x_j и y_k.

2) В канале без помех возможны лишь пары x_j, y_j, поэтому для остальных пар находить взаимную информацию просто нет смысла. Для таких пар имеем

$$\boldsymbol{I}(\boldsymbol{x}_j;\boldsymbol{y}_j)=\boldsymbol{I}(\boldsymbol{x}_j)=\boldsymbol{I}(\boldsymbol{y}_j), \tag{4.21}$$

т.к. в таком канале всегда $p(x_j/y_j)= p(y_j/x_j)=1$. В канале без помех, как уже отмечалось, вся информация, создаваемая источником, достигает получателя.

3) Взаимная информация есть разность двух собственных информаций

$$\boldsymbol{I}(\boldsymbol{x}_j;\boldsymbol{y}_k)=\boldsymbol{I}(\boldsymbol{x}_j)-\boldsymbol{I}\left(\boldsymbol{x}_j/\boldsymbol{y}_k\right)=\boldsymbol{I}(\boldsymbol{y}_k)-\boldsymbol{I}\left(\boldsymbol{y}_k/\boldsymbol{x}_j\right). \tag{4.22}$$

И в канале с помехами при большом отношении сигнал/помеха наиболее вероятным является $k=j$, поэтому обычно $p(x_j/y_j)>p(x_j)$ и, следовательно, $I(x_j/y_j)<I(x_j)$, тогда $I(x_j;y_j)>0$. Но иногда все-таки случается $k\neq j$, и в этой ситуации вполне возможно, что $I(x_j/y_k)>I(x_j)$ и, следовательно, количество получаемой информации отрицательно $I(x_j;y_k)<0$.

4) Если X и Y независимы (помеха полностью подавляет полезный сигнал), всегда $I(x_j;y_k)=0$, т.к. $P(x_j/y_k)=P(x_j)$.

Математическое ожидание случайной величины “взаимная информация” равно

$$\boldsymbol{I}(\boldsymbol{X};\boldsymbol{Y})=\sum_{\boldsymbol{j}=1}^{\boldsymbol{m}}\sum_{\boldsymbol{k}=1}^{\boldsymbol{m}}\boldsymbol{p}(\boldsymbol{x}_j,\boldsymbol{y}_k)\boldsymbol{I}(\boldsymbol{x}_j;\boldsymbol{y}_k) \tag{4.23}$$

и называется *средней взаимной информацией между X и Y*, причем вместо $I(x_j;y_k)$ можно подставить любое из трех соотношений (4.20).

Средняя взаимная информация также обладает свойствами: симметрии $I(X;Y)=I(Y;X)$; в канале без помех $I(X;Y)=H(X)=H(Y)$; для независимых X и Y всегда $I(X;Y)=0$, а в общем случае имеем

$$\boldsymbol{I}(\boldsymbol{X};\boldsymbol{Y})=\boldsymbol{H}(\boldsymbol{X})-\boldsymbol{H}\left(\boldsymbol{X}/\boldsymbol{Y}\right). \tag{4.24}$$

Напомним, что $H(X)\geq 0$ и $H(X/Y)\geq 0$.

Воспользовавшись неравенством Иенсена для логарифмической функции при $Z>0$

$$\mathrm{M}\left[\log \boldsymbol{Z}\right]\leq \log \boldsymbol{M}\left[\boldsymbol{Z}\right] \tag{4.25}$$

и, положив $Z_{j,k}=P(x_j)/P(x_j/y_k)$, можно показать, что

$$\boldsymbol{I}(\boldsymbol{X};\boldsymbol{Y})\geq 0. \tag{4.26}$$

Отсюда, кстати, следует, что всегда $H(X/Y)\leq H(X)$.

В свете сказанного величины, входящие в формулу (4.24), приобретают вполне определенный смысл. Информация $H(X)$, созданная источником, делится на две части:

$I(X;Y)$ – то, что доходит до получателя;

$H(X/Y)$ – то, что потеряно в канале из-за воздействия помех на сигнал.

Соотношение между этими двумя частями в первую очередь зависит от отношения сигнал/помеха.

Приведенные выше соотношения, конечно, по своей сути справедливы и в более общем случае, когда переданное сообщение $\mathbf{X}=X^{[1]},\ldots, X^{[n]}$ и принятый сигнал $\mathbf{Y}=Y^{[1]},\ldots, Y^{[n]}$ – это последовательности длины n, состоящие из m-ичных символов.

Взаимная информация между конкретными реализациями этих последовательностей равна

$$\boldsymbol{I}(\boldsymbol{x}_i^{[1]},...,\boldsymbol{x}_j^{[n]};\boldsymbol{y}_q^{[1]},...,\boldsymbol{y}_s^{[n]}) = \log\frac{\boldsymbol{p}\left(\boldsymbol{x}_i^{[1]},...,\boldsymbol{x}_j^{[n]}/\boldsymbol{y}_q^{[1]},...,\boldsymbol{y}_s^{[n]}\right)}{\boldsymbol{p}(\boldsymbol{x}_i^{[1]},...,\boldsymbol{x}_j^{[n]})} \qquad (4.27)$$

и в дополнение к перечисленным выше свойствам обладает еще и свойством аддитивности (способ формального выражения этого свойства зависит от того, какой из многочисленных вариантов формулы умножения вероятностей мы изберем для представления числителя и знаменателя в выражении (4.27)).

Средняя взаимная информация между последовательностями X и Y

$$\begin{aligned}\boldsymbol{I}_n &= \boldsymbol{I}(\boldsymbol{X}^{[1]},...,\boldsymbol{X}^{[n]};\boldsymbol{Y}^{[1]},...,\boldsymbol{Y}^{[n]}) = \\ &= \sum_{i=1}^{m}...\sum_{j=1}^{m}\sum_{q=1}^{m}...\sum_{s=1}^{m}\boldsymbol{p}(\boldsymbol{x}_i^{[1]},...,\boldsymbol{x}_j^{[n]},\boldsymbol{y}_q^{[1]},...,\boldsymbol{y}_s^{[n]})\times \\ &\times\boldsymbol{I}(\boldsymbol{x}_i^{[1]},...,\boldsymbol{x}_j^{[n]};\boldsymbol{y}_q^{[1]},...,\boldsymbol{y}_s^{[n]})\end{aligned} \qquad (4.28)$$

является основной характеристикой, используемой для описания возможностей цифровых информационных систем.

Для нее также справедливо соотношение (4.24)

$$\boldsymbol{I}_n = \boldsymbol{H}_n - \boldsymbol{H}_{n\,\text{усл}}, \qquad (4.29)$$

где $\boldsymbol{H}_n = \boldsymbol{H}(\boldsymbol{X}^{[1]},...,\boldsymbol{X}^{[n]})$,

$\boldsymbol{H}_{n\,\text{усл}} = \boldsymbol{H}(\boldsymbol{X}^{[1]},...,\boldsymbol{X}^{[n]}/\boldsymbol{Y}^{[1]},...,\boldsymbol{Y}^{[n]})$ – энтропия и условная энтропия передаваемого сообщения $\mathbf{X}$ соответственно. Поскольку все члены в (4.29) возрастают при увеличении длины последовательностей n, удобнее провести нормировку и представить это соотношение в виде

$$\boldsymbol{I} = \boldsymbol{H} - \boldsymbol{H}_{\text{усл}}, \qquad (4.30)$$

где $H=H_n/n$ – *скорость создания информации*;

$I=I_n/n$ – *скорость передачи информации* (логичнее было бы назвать эту величину скоростью приема информации, но первый термин уже прочно вошел в употребление);

$H_{\text{усл}}=H_{n\,\text{усл}}/n$ – *скорость потерь информации*.

Все три величины измеряются в битах на символ. Поделив эти величины на длительность сигнала, соответствующего одному символу, будем измерять их в битах в секунду.

4.4 Пропускная способность канала и теоремы о кодировании в цифровом канале с помехами

Скорость передачи информации (4.30) – это характеристика СПИ в целом (рис.2.1), поэтому она зависит от многих факторов: производительности источника H, способа передачи (кодирования), способа приема и характеристик линии передачи (полосы пропускания, отношения сигнал/помеха и т.п.).

Пропускная способность канала

$$\boldsymbol{C} = \max \boldsymbol{I} \quad - \tag{4.31}$$

это наибольшая возможная в этом канале скорость передачи информации (бит/символ или бит/с) при заданных ограничениях на значения ряда его физических параметров.

Получается, что всегда $I \le C$. Чтобы достигнуть скорости передачи, равной пропускной способности канала, нужно иметь достаточно производительный источник информации и использовать оптимальные способы передачи и приема.

Например, *пропускная способность m-ичного канала без помех* на основании (4.11) и (4.21)

$$\boldsymbol{C} = \log \boldsymbol{m}, \ \frac{\text{бит}}{\text{символ}}. \tag{4.32}$$

Чтобы добиться максимальной скорости передачи информации, необходимо и достаточно, чтобы сигнал на входе канала не содержал избыточности.

В качестве второго примера найдем **пропускную способность двоичного симметричного канала с ошибками** (рис. 4.4), обозначив **X** и **Y**=**X**⊕**E** – последовательности символов на его входе и выходе соответственно; **E** – случайный вектор ошибок (смотри также (3.2)); $\begin{matrix} \boldsymbol{p} = \boldsymbol{p}\left(\boldsymbol{y}^j = 1 / \boldsymbol{x}^j = 0\right) = \\ \boldsymbol{p}\left(\boldsymbol{y}^j = 0 / \boldsymbol{x}^j = 1\right) - \end{matrix}$ вероятность появления ошибки в очередном, j-м двоичном символе (битовая вероятность ошибки), при этом предполагается, что ошибки в отдельных символах независимы (см. также (3.57)). Число возможных значений каждой из n-разрядных последовательностей равно $N=2^n$.

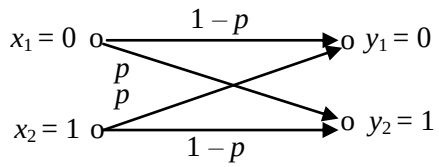

Рис. 4.4 – Переходные вероятности двоичного симметричного канала

Пользуясь свойством симметрии (4.20), запишем

$$\boldsymbol{I}(\mathbf{X};\mathbf{Y}) = \boldsymbol{H}(\mathbf{Y}) - \boldsymbol{H}\left(\mathbf{Y} / \mathbf{X}\right). \tag{4.33}$$

Ошибки в симметричном канале не зависят от значений переданных символов, поэтому величина

$$\boldsymbol{H}\left(\mathbf{Y}/\mathbf{X}\right)=\boldsymbol{H}\left(\mathbf{E}\right)=\boldsymbol{n}\left[-\boldsymbol{p}\cdot\log\boldsymbol{p}-(1-\boldsymbol{p})\cdot\log(1-\boldsymbol{p})\right] \qquad (4.34)$$

также не зависит от вероятностных характеристик переданного сигнала **X** и является энтропией случайного вектора ошибок **E**. Следовательно, чтобы максимизировать величину (4.33), нужно формировать передаваемый сигнал **X** таким образом, чтобы $H(\mathbf{Y})=H(\mathbf{Y})_{\max}=n\log m=n$ при $m=2$. Это возможно лишь в том случае, когда все символы в последовательности **Y** независимы, а каждый символ с одинаковой вероятностью принимает значение 0 или 1 (сигнал **Y** не содержит избыточности). Это, в свою очередь, при $p \neq 0{,}5$ возможно лишь тогда, когда передаваемый сигнал **X** также не содержит избыточности. В итоге получим формулу для пропускной способности

$$\boldsymbol{C}=1+\boldsymbol{p}\cdot\log\boldsymbol{p}+(1-\boldsymbol{p})\cdot\log(1-\boldsymbol{p}),\quad \frac{\text{бит}}{\text{дв.симв}}. \qquad (4.35)$$

Эта зависимость представлена на рис. 4.5 (анализируя график, учтите, что при $p=1$ после инвертирования выходного сигнала мы фактически полу-

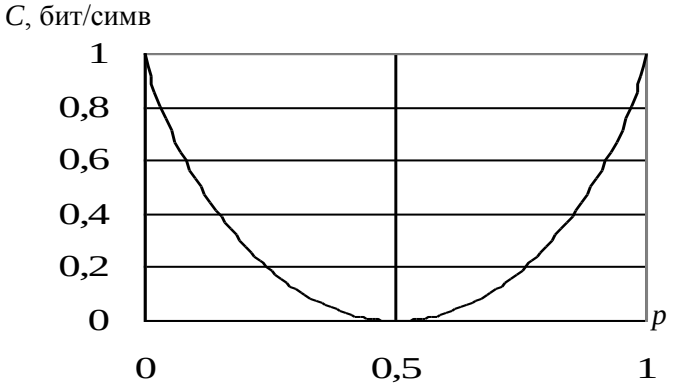

Рис. 4.5. Пропускная способность двоичного симметричного канала без памяти

чаем канал без ошибок).

Как было показано в разделе 3.1, сигнал, не обладающий избыточностью, не предоставляет возможностей обнаруживать и исправлять ошибки. Таким образом, *применение корректирующего кода и передача информации с максимально возможной скоростью I=C – понятия несовместимые*. Выясним, возможно ли это при $I<C$.

“Удлиним” канал, т.е. включим в его состав кодер и декодер, и обозначим **A** и **B** – цифровые сигналы на его входе и выходе соответственно. Напомним, что способ кодирования определен, если из общего количества N возможных последовательностей **X** выбраны $N_A<N$ последовательностей, признанных в качестве разрешенных к передаче (включенных в кодовую таблицу), при этом каждому из N_A возможных сообщений **a** соответствует своя последовательность $\mathbf{x}=f(\mathbf{a})$. В свою очередь, функция $\mathbf{b}=g(\mathbf{y})$ определяет способ декодирования. Тогда имеем

$$\boldsymbol{I}(\mathbf{A};\mathbf{B})=\boldsymbol{H}(\mathbf{A})-\boldsymbol{H}\left(\mathbf{A}/\mathbf{B}\right). \qquad (4.36)$$

Принципиальный вопрос теории помехоустойчивого кодирования в упрощенной постановке можно сформулировать следующим образом: существуют ли такие преобразования f и g, которые позволяют канал с ошибками после "удлинения" превратить в канал без ошибок, т.е. обеспечить $H(\mathbf{A}/\mathbf{B})=0$.

Ответ на этот вопрос дают **теоремы Шеннона о кодировании в дискретном канале с шумом** (1948 год).

Прямая теорема. *Если скорость создания информации H источником на входе канала с ошибками с пропускной способностью C меньше пропускной способности, то существует такой код, который способен обеспечить сколь угодно малую вероятность ошибки при декодировании, при этом скорость передачи информации I может быть сколь угодно близка к скорости ее создания H.*

При доказательстве теоремы предполагалось, что код случаен, т.е. отбор N_A разрешенных кодовых комбинаций **X** из общего их количества N произведен случайным образом. Далее записывалось соотношение для вероятности правильного декодирования $P_{\text{пр}}$, т.е. вероятности того, что принятая комбинация **y** окажется ближе к передаваемой комбинации **x**, нежели к любой другой разрешенной комбинации. Поскольку для случайного кода $P_{\text{пр}}$ – тоже случайная величина, находилось ее математическое ожидание $\mathrm{M}[P_{\text{пр}}]$ (среднее значение). Затем осуществлялся предельный переход при $n\to\infty$ и оказалось, что $\mathrm{M}[P_{\text{пр}}]\to 1$, следовательно, вероятность ошибки при декодировании $\mathrm{M}[P_{\text{ош}}]\to 0$, $\mathrm{M}[H(A/B)]\to 0$, $M[I(A;B)]\to H(A)$. И, наконец, очевидно, что в этом ансамбле случайных кодов существует хотя бы один код, для которого характеристики помехоустойчивости не хуже средних, т.е. $P_{\text{ош}}\le \mathrm{M}[P_{\text{ош}}]$.

В частности, если все последовательности **A, X, У, B** – двоичные, и все N_A значений сообщения **A** равновероятны, условие $H<C$ означает

$$\boldsymbol{H}=\frac{\log \boldsymbol{N_A}}{\boldsymbol{n}}=\frac{\boldsymbol{k}}{\boldsymbol{n}}<\boldsymbol{C} \qquad (4.37)$$

где k – длина последовательности **A**, т.е. количество информационных символов при использовании линейного блочного кода (n,k).

Прямая теорема не утверждает, что возможна безошибочная передача ($P_{\text{ош}}=0$), можно лишь обеспечить любое сколь угодно малое ненулевое значение $P_{\text{ош}}$. Как показывает численный анализ, в том числе и для известных линейных блочных кодов, чем меньше требуемое значение $P_{\text{ош}}$, тем большими значениям n и k должен характеризоваться корректирующий код (при этом кодер и декодер становятся более сложными, и возрастают задержки при передаче и приеме).

Обратная теорема. *Если $H>C$, никакой код не позволит получить сколь угодно малую вероятность ошибки $P_{\text{ош}}$ и обеспечить скорость потерь информации $H_{\text{усл}}$ меньшую, чем $H–C$.*

Из (4.30) имеем $H_{\text{усл}}=H–I$. По определению (4.31) $I\le C$, следовательно, $H_{\text{усл}}\ge H–C>0$. Но $P_{\text{ош}}\to 0$ возможно лишь при $H_{\text{усл}}\to 0$, откуда и следует утвер-

ждение теоремы.

Кстати, если $H>C$, увеличение длины кодовых комбинаций n, наоборот, ведет к увеличению вероятности ошибки $P_{\text{ош}}$.

Доказательство теорем Шеннона дало мощный толчок развитию теории помехоустойчивого кодирования. Основное практическое их применение очевидно: прежде чем конструировать корректирующий код с заданными характеристиками для применения с конкретным источником информации и конкретным каналом, полезно вычислить и сравнить H и C.

4.5 Пропускная способность непрерывного канала с шумом

Большая часть из того, что было сказано об информационных характеристиках цифровых сигналов, по сути, применимо и к непрерывным сигналам. В частности, если X и Y – непрерывные случайные величины, то величина взаимной информации между их значениями x и y, подобно (4.19), равна

$$I(x;y)=\log\frac{W(x/y)}{W(x)}, \tag{4.38}$$

а величина средней взаимной информации

$$I(X;Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}W(x,y)I(x;y)dxdy. \tag{4.39}$$

Принципиальное различие между цифровыми и непрерывными сигналами заключается в следующем. Для непрерывного сигнала бессмысленным является понятие “канал без помех”. Действительно, если мы говорим о передаче в канале без помех, это значит, что получатель имеет возможность точно определить значение переданного сообщения. При угадывании переданного значения непрерывного сигнала всегда будем допускать ошибку, пусть очень маленькую (вероятность попадания в точку на непрерывном отрезке есть бесконечно малая величина). Таким образом, *неопределенность ожидаемого значения непрерывной случайной величины теоретически бесконечно велика*.

По этой причине следует внимательно относиться к соотношениям, в которых фигурирует понятие энтропии непрерывной случайной величины

$$H_{\text{д}}(X)=-\int_{-\infty}^{+\infty}W(x)\log W(x)dx. \tag{4.40}$$

Как видно, под знаком логарифма здесь стоит размерная величина (напомним, что размерность $W(x)$ есть размерность величины x^{-1}). Поэтому величина $H(X)$ зависит от выбора единиц, в которых измеряется значение x. Величины (4.38) и (4.39) определены корректно и такой зависимости не имеют.

Величине (4.40) можно приписать следующий смысл. Она показывает, насколько неопределенность ожидаемого значения случайной величины X

больше, чем неопределенность ожидаемого значения другой, эталонной случайной величины $X_э$, имеющей равномерное распределение на интервале единичной длины. Чтобы отметить эту особенность, $H_д(X)$ называют дифференциальной энтропией.

Один из наиболее важных результатов теории информации – это **формула для пропускной способности непрерывного канала с шумом**.

Предполагается, что:

1) параметры источника информации постоянны во времени, то есть передаваемый сигнал $X(t)$ – это стационарный случайный процесс;

2) линия передачи обладает заданной полосой пропускания Δf и ее параметры также постоянны, то есть она не вносит никаких искажений в передаваемый сигнал;

3) единственной помехой на входе приемника является аддитивный белый гауссовский шум со спектральной плотностью $N_о$.

Пропускная способность канала равна

$$C = \Delta f \log(1 + P_c / P_ш), \quad \text{бит/с}, \tag{4.41}$$

где $P_с$ и $P_ш$ – средние мощности полезного сигнала и шума на выходе канала (на входе приемника) соответственно, причем для белого шума из (1.24) имеем $P_ш = N_о \Delta f$.

Чтобы не изображать семейство графиков, введем формально нормирующую величину

$$\Delta f_н = P_с / N_о, \tag{4.42}$$

тогда зависимость пропускной способности от обоих параметров Δf и $P_с$ можно представить одной кривой (рис. 4.6).

Из приведенной зависимости видно, что:

1) предельное значение скорости передачи информации на единицу мощности полезного сигнала на входе приемника равно $C_{max} = (\log_2 e)/N_о \approx 1{,}4427/N_о$ (бит/Вт·с) и может быть достигнуто лишь асимптотически при $\Delta f \to \infty$, то есть предельная энергетическая эффективность СПИ, равная $1{,}4427/N_о$ (бит/Дж) достигается при ее весьма низкой частотной эффективности; но тогда и $P_с/P_ш \to 0$, поэтому потребуется применять чрезвычайно мощный (и сложный) корректирующий код;

2) имеет смысл расширять полосу пропускания канала до тех пор, пока $P_ш = N_о \Delta f < P_с$, так как при дальнейшем увеличении полосы прирост пропуск-

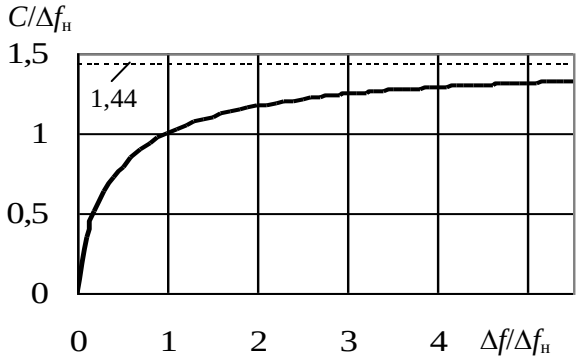

Рис. 4.6. Пропускная способность непрерывного канала с белым шумом

ной способности существенно замедляется.

Чтобы обеспечить скорость передачи, равную пропускной способности канала, нужно, чтобы и полезный сигнал на входе линии являлся гауссовским случайным процессом с равномерным спектром плотности мощности в полосе Δf, то есть был подобен белому шуму. Разработка и техническая реализация аналоговых алгоритмов такого преобразования для большинства реальных сообщений сопряжены с большими трудностями. Значительно проще во многих случаях оказывается использовать цифровые методы передачи аналоговых сигналов.

Заключительные замечания о значении теории информации.

1) Любая теория информации оперирует лишь вероятностными характеристиками сигналов и является математической теорией. Поэтому она универсальна, то есть может быть применена к любым сигналам независимо от их физической формы.

2) Основные результаты теории информации имеют вид неравенств, то есть, она определяет предельные, потенциальные характеристики СПИ, не всегда указывая пути для их достижения. Наглядное сравнение: если нас интересует, каков объем сосуда сложной формы, то, определив его максимальный линейный размер d, мы уже знаем, что этот объем никак не может быть больше $\pi d^3/6$ (для тех, кто не знает, каким замечательным свойством обладает объем шара диаметра d, пути достижения этого предельного значения остаются весьма туманными).

3) Статистические теории информации не учитывают ценности конкретного сообщения для его получателя. Попытки исправить это положение не привели к осязаемым результатам, поскольку ценность сообщения – понятие субъективное. Здесь лучше придерживаться девиза цивилизованной обслуживающей организации: если потребитель платит за передачу сообщения, значит, для него это важно.

Итак, теория информации – не средство на все случаи жизни, но примененная “к месту”, она может существенно ускорить и упростить процесс проектирования и анализа СПИ.

Вопросы для самоконтроля по главе 4

1. Что такое собственная информация и энтропия дискретной случайной величины?
2. При каких условиях максимальна энтропия совокупности двух символов и чему она равна?
3. Что такое избыточность дискретного источника?
4. Может ли равномерный код быть оптимальным (безизбыточным)?
5. Дайте определение взаимной информации переданного и принятого символов. Как влияет на ее величину интенсивность помех в канале

связи?

6. Что такое избыточность сигнала? В каких случаях она полезна, а когда нет?
7. Когда полезно применять кодирование с малой избыточностью?
8. Какой смысл вкладывают в понятия: “кодирование источника”? “канальное кодирование”?
9. Каково значение минимально-возможной средней длины кодовой комбинации?
10. Всегда ли удается закодировать сигнал так, чтобы избыточность на выходе кодера была нулевой?
11. Когда полезно кодировать блоки букв, а не отдельные буквы?
12. Какой способ разделения кодовых комбинаций применяется в кодах, обладающих малой избыточностью?
13. В чем заключается главный недостаток кодов Хафмана и Шеннона-Фано?
14. Откуда берется кодовая таблица, используемая при кодировании кодом Лемпела-Зива?
15. От чего зависит пропускная способность непрерывного канала связи с аддитивным белым шумом?

5 ДЕМОДУЛЯЦИЯ ЦИФРОВЫХ СИГНАЛОВ

5.1 Роль априорной информации

В цифровой системе передачи информации (СПИ) передаваемый сигнал является последовательностью M-ичных символов, при этом, как правило,

$$M = 2^k. \qquad (5.1)$$

Значение каждого символа передается при помощи радиоимпульса прямоугольной формы, используя один из методов модуляции, описанных в разд. 2.8. В процессе передачи сигнала по линии эти импульсы случайным образом искажаются (обычно это происходит из-за наличия мультипликативной помехи) и появляется аддитивная помеха (разд. 1.5).

Демодулятор к моменту окончания очередного принимаемого импульса должен указать (вернее, угадать), которое из M возможных значений символа было передано с данным импульсом. Очевидно, что иногда демодулятор будет выдавать ошибочные решения, поэтому желательно применять такой способ обработки импульса, который при заданных характеристиках сигналов и помех обеспечивает *минимум полной вероятности ошибки* P. Это и есть главный критерий качества приема в цифровой СПИ.

Чтобы следовать этому критерию, для обработки очередного импульса нужно использовать такой алгоритм, который учитывает все сведения об ожидаемом импульсе, которые известны к данному моменту (момент прихода и длительность, несущая частота и начальная фаза, характер искажений огибающей и закон паразитной внутриимпульсной фазовой модуляции и т.п.). Это и есть априорные сведения, и чем их больше, тем меньше будет вероятность ошибок, допускаемых демодулятором, который учитывает эти сведения.

Некоторые параметры ожидаемого импульса известны с высокой точностью. Например, в канале с постоянными параметрами принимаемый сигнал повторяет по форме передаваемый, поэтому известны длительность импульса и его амплитуда. Считается, что известны также текущие значения несущей частоты и частоты повторения импульсов, но здесь многое зависит от стабильности генераторов, задающих эти частоты в передатчике, и генераторов, воспроизводящих эти же колебания в приемнике. Фактически эти генераторы играют роль часов, по которым планируется график формирования (обработки) сигнала в передатчике (приемнике). Чем выше синхронность хода часов приемника по отношению к часам передатчика, тем выше качество приема.

Те устройства приемника, которые обеспечивают условия, при которых частота и даже текущая фаза генератора несущей в приемнике с достаточной точностью совпадают с соответствующими параметрами генератора в пере-

датчике, называются системой обеспечения *когерентности*. Аналогично, система тактовой *синхронизации* формирует в приемнике тактовые импульсы, которые с достаточной точностью указывают момент времени, когда начинается очередной принимаемый импульс (следовательно, заканчивается предыдущий).

Обеспечить качественную синхронизацию обычно бывает проще, нежели когерентность, поскольку несущая частота, как правило, существенно выше тактовой частоты. В связи с этим принято рассматривать три типа систем по *степени их когерентности*.

Когерентной называется СПИ, в которой ожидаемые значения начальных фаз всех принимаемых импульсов (вплоть до окончания сеанса связи) известны заранее, и эти сведения используются при демодуляции импульсов. Другими словами, генераторы несущей в передатчике и приемнике должны обладать столь высокой стабильностью, чтобы фазы выдаваемых колебаний не расходились заметно в течение сеанса связи. Когерентная СПИ – это идеал, который используется лишь для сравнения с другими СПИ, реализуемыми практически.

Частично-когерентной называется СПИ, в которой ожидаемые значения начальных фаз всех принимаемых импульсов заранее неизвестны, но в процессе приема они оцениваются, и эти сведения используются при демодуляции импульсов. Другими словами, генератор несущей в приемнике при помощи устройства фазовой автоподстройки частоты (ФАПЧ) постоянно синхронизируется с генератором передатчика. Подстройка осуществляется по самому принимаемому сигналу. И в этом случае генераторы несущей должны обладать некоторой стабильностью частоты, достаточной, например, для того, чтобы в случае разрыва цепи ФАПЧ фазы выдаваемых колебаний не расходились заметно хотя бы в течение нескольких сотен импульсов, что вполне реализуемо. Поэтому именно *частично-когерентную СПИ на практике обычно называют когерентной*.

Некогерентной называется СПИ, в которой ожидаемые значения начальных фаз всех принимаемых импульсов неизвестны и не оцениваются в процессе приема (ФАПЧ не применяется). Прием очередного импульса рассматривается как прием сигнала со случайной начальной фазой, равномерно распределенной в интервале $0–2\pi$. Для этого генераторы несущей в передатчике и приемнике должны обладать лишь такой стабильностью, чтобы фазы выдаваемых колебаний не расходились заметно хотя бы в течение одного импульса. Поэтому некогерентная СПИ оказывается проще и дешевле.

Очевидно, что когерентная СПИ, в принципе, может обеспечить меньшую вероятность ошибки, чем частично-когерентная и, в еще большей степени, некогерентная СПИ.

Подобным образом можно классифицировать цифровые СПИ как синхронные и асинхронные. В *синхронной* СПИ передача каждого символа (им-

пульса) начинается в тактовый момент времени, при этом периодическая последовательность тактовых импульсов генерируется постоянно. Благодаря этому в приемнике имеется возможность осуществлять ФАПЧ местного генератора тактовых импульсов по принимаемому сигналу и предсказывать ожидаемые моменты прихода для большого количества ближайших символов.

В *асинхронной* СПИ передача первого символа кодовой комбинации начинается в произвольный момент времени, правда, остальные символы следуют за ним через равные интервалы известной величины. В связи с этим в начале каждой кодовой комбинации обязательно нужно передавать дополнительный, *стартовый* импульс, которые запускает в приемнике ждущий генератор тактовых импульсов с той же частотой повторения. Таким способом предсказывается ожидаемое время прихода всех импульсов, но только для данной кодовой комбинации.

В современных СПИ передача символов, как правило, проводится в синхронном режиме, а асинхронный режим в случае необходимости применяется лишь для передачи достаточно длинных кодовых комбинаций (см. разд. 6.5).

Другой фактор, определяющий уровень априорных сведений о сигнале, это мультипликативная помеха. При воздействии мультипликативной помехи в виде временных селективных замираний амплитуды и начальные фазы принимаемых импульсов достаточно медленно, но случайным образом изменяются во времени. Демодуляция превращается в прием импульсов известной формы, но со случайными значениями амплитуд и начальных фаз, при этом, как правило, удается обеспечить частичную когерентность СПИ.

При воздействии мультипликативной помехи в виде частотных селективных замираний форма принимаемых импульсов становится случайной, но неизменной во времени, и демодуляция превращается в прием импульсов неизвестной (случайной) формы.

При воздействии мультипликативной помехи общего вида (1.26) форма принимаемых импульсов становится случайной и при этом медленно изменяется (флуктуирует) во времени, увеличивается межсимвольная интерференция. Оптимальный прием таких сигналов существенно усложняется, при этом заметно увеличивается вероятность ошибки.

Таким образом, *при выборе способа демодуляции импульсов* и при вычислении достигаемой при этом вероятности ошибки *необходимо четко определить характер аддитивных и мультипликативных помех*, воздействующих на сигнал.

Рассмотрим сначала задачу демодуляции **двоичного импульса**, переданного в канале с постоянными параметрами при воздействии аддитивного

белого гауссовского шума. В таких условиях, хотя бы в принципе, можно обеспечить когерентность СПИ и использовать любой из способов манипуляции гармонической несущей: АМ, ЧМ или ФМ (разд. 2.8).

Демодулятор к моменту окончания очередного импульса должен определить, которое из двух значений символа (0 или 1) было передано либо, что то же самое, который из двух сигналов ($u_0(t)$ или $u_1(t)$) был передан. В терминах статистической радиотехники это – классическая задача различения двух сигналов известной формы на фоне аддитивного белого шума.

Доказано, что *оптимальный способ приема* включает два этапа:

1) преобразование сигнала на входе приемника

$$u(t) = u_c(t) + u_{ø}(t), \tag{5.2}$$

состоящего из реализации полезного сигнала $u_c(t)$ и реализации шума $u_{ø}(t)$, в число v при помощи корреляционного приемника (2.35)

$$v = \int_{t_i}^{t_ê} u(t)w(t)dt; \tag{5.3}$$

2) вынесение решения относительно предполагаемого значения $\hat{s}$ переданного символа s путем сравнения полученного значения v с пороговым значением (порогом) $v_п$ по правилу

$$\hat{s} = \begin{cases} 0, & \text{если } v > v_n, \\ 1, & \text{если } v \le v_n. \end{cases} \tag{5.4}$$

Весовая функция должна совпадать по форме с разностным сигналом

$$w(t) = u_0(t) - u_1(t). \tag{5.5}$$

Обычно символы-импульсы следуют один за другим без пустых защитных интервалов, поэтому при использовании АМ и ФМ генератор опорного сигнала $w(t)$ превращается в генератор гармонического колебания несущей частоты.

Поскольку в (2.37) величина $v = v_c + v_ш$ образуется как сумма детерминированной v_c и случайной гауссовской $v_ш$ составляющих, она также является случайной величиной, имеющей нормальное распределение вероятности

$$W(v) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left(v - m_j\right)^2}{2\sigma^2} \right], \qquad j = 0;1. \tag{5.6}$$

Как видно из (2.37), значение математического ожидания

$$m_j = \mathrm{M}[v_c] + \mathrm{M}[v_ш] = \int_{t_н}^{t_к} u_j(t)w(t)dt + 0 =$$

$$= \begin{cases} \int_{t_н}^{t_к} u_0(t)\left[u_0(t) - u_1(t)\right]dt, & \text{если } s = 0, \\ \int_{t_н}^{t_к} u_1(t)\left[u_0(t) - u_1(t)\right]dt, & \text{если } s = 1 \end{cases} \tag{5.7}$$

зависит от того, который из двух сигналов передавался. Если энергии обоих сигналов одинаковы (ЧМ, ФМ), то $m_1 = -m_0$.

Значение среднеквадратического отклонения получим с учетом того, что корреляционная функция белого шума – это дельта-функция [1,2], а она обладает фильтрующим свойством,

$$\sigma^2 = \mathrm{M}\left[v_ø\, v_ø\right] = \int_{t_i}^{t_ê}\int_{t_i}^{t_ê} \mathrm{M}\left[u_ø(t_1)u_ø(t_2)\right] w(t_1)w(t_2)dt_1dt_2 =$$

$$= N_o \int_{t_i}^{t_ê}\int_{t_i}^{t_ê} \delta\left(t_1 - t_2\right) w(t_1)w(t_2)dt_1dt_2 = N_o \int_{t_i}^{t_ê} w(t)^2 dt \qquad (5.8)$$

и видим, что оно не зависит от того, который из сигналов передавался.

Отношение сигнал/шум по энергии

$$q^2 = \frac{\left(m_0 - m_1\right)^2}{\sigma^2} = \frac{\int_{t_i}^{t_ê}\left[u_0(t) - u_1(t)\right]^2 dt}{N_o} \qquad (5.9)$$

является отношением энергии разностного сигнала к спектральной плотности шума.

На рис. 5.1 изображены условные плотности вероятности величины v для случая, когда используется ФМ. Пороговое значение v_n принято равным нулю.

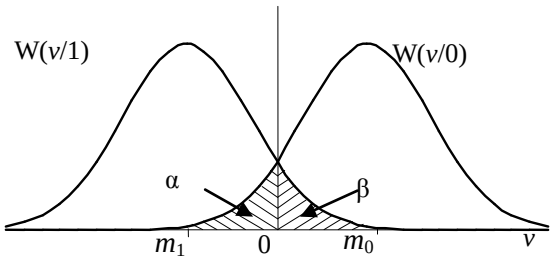

Рис. 5.1. Условные плотности вероятности сигнала на выходе интегратора корреляционного приемника

Из рисунка видно, что возможна ситуация, когда передается символ 0, но величина v оказывается ниже порога, поэтому демодулятор в соответствии с инструкцией (5.4) выносит решение о том, что передан символ 1. Такой исход может быть назван *ошибкой первого рода*, и вероятность такой ошибки равна

$$\alpha = \int_{-\infty}^{v_п} W(v/0)dv\,. \qquad (5.10)$$

Аналогично, *ошибка второго рода* возникает тогда, когда демодулятор выдает символ 0, но фактически передавался символ 1. Вероятность появления такой ошибки равна

$$\beta = \int_{v_п}^{-\infty} W(v/1)dv. \qquad (5.11)$$

При увеличении порога возрастает величина α, но зато уменьшается β, при сдвиге порога влево эта зависимость меняется на обратную. Обычно в СПИ символы 0 и 1 передаются с одинаковой вероятностью, а неприятности, обусловленные появлением ошибок обоих видов, одинаковы. Тогда объективной характеристикой качества приема становится полная вероятность ошибки (битовая вероятность ошибки или, в англоязычной литературе, Bit Error Rate (BER))

$$p = P(0)\alpha + P(1)\beta = 0{,}5(\alpha + \beta). \qquad (5.12)$$

Легко убедиться, что для сигналов ФМ эта величина минимальна, если $v_п = 0$, как показано на рис. 5.1, и равна

$$p = \alpha = \beta = 1 - \Phi(q/2), \qquad (5.13)$$

где $\Phi(q)$ – интеграл вероятности (таблица значений интеграла вероятности приведена в Приложении 3).

Обратите внимание, что BER зависит лишь от отношения (5.9) энергии *разностного сигнала* к спектральной плотности шума и, разумеется, уменьшается при увеличении этого отношения).

Посмотрим, как величина q^2 зависит от отношения сигнал/шум в обычном смысле, то есть от отношения *средней энергии импульса* на входе демодулятора к спектральной плотности шума, полагая символы 0 и 1 равновероятными:

$$q_c^{\,2} = P(0)\frac{m_0^2}{\sigma^2} + P(1)\frac{m_1^2}{\sigma^2} = \frac{1}{2N_o}\int_{t_i}^{t_ê}\left[u_0^2(t) + u_1^2(t)\right]dt = \frac{E_0 + E_1}{2N_o}. \qquad (5.14)$$

Для сигнала с ФМ имеем E_0=E_1, поэтому из (5.9) получим (см. также рис. 2.18а) $q = 2q_c$.

Для сигнала с АМ это соотношение является наилучшим, когда один из двух сигналов, скажем, $u_1(t) = 0$, и имеем $q = \sqrt{2}q_c$.

Для сигнала с ЧМ это соотношение зависит от модуля разности несущих частот $\Delta f = |f_0 - f_1|$ и равно

$$q = \sqrt{2}\left[1 - \frac{\sin(\pi\Delta f\tau)}{\pi\Delta f\tau}\right]q_c. \qquad (5.15)$$

Если $\Delta f = n/\tau$, где n – целое положительное число, сигналы $u_0(t)$ и $u_1(t)$ становятся ортогональными, и $q = \sqrt{2}q_c$, то есть помехоустойчивость АМ и ЧМ сигналов одинакова. Именно для этого случая для когерентной СПИ на рис. 5.2 показана зависимость вероятности ошибки, вычисленной по формуле (5.13), от отношения сигнал/шум q_c.

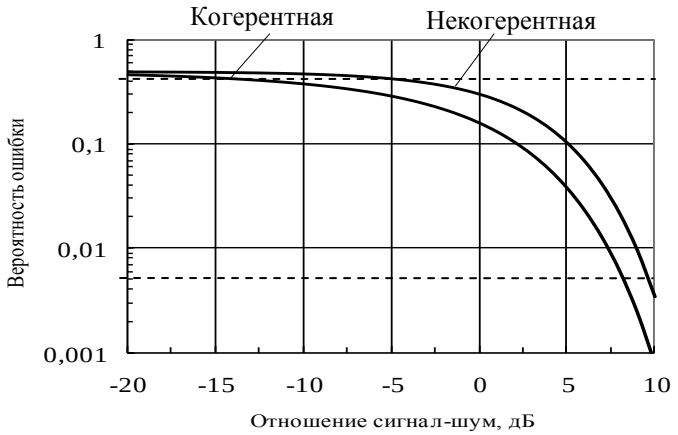

Рис. 5.2. Зависимость вероятности ошибки от отношения сигнал/шум q_c в СПИ с двумя ортогональными сигналами

Итак, отношение сигнал/шум максимально и, следовательно, BER при демодуляции минимальна для сигналов с ФМ. Более того, доказано, что *ФМ – это идеал, то есть при передаче двоичных импульсов при заданной величине q_c не существует другого способа модуляции, который при приеме на фоне белого шума обеспечил бы меньшее значение BER, чем ФМ.*

Рассмотрим методы **демодуляции сигналов при использовании многопозиционных методов модуляции**, то есть в случае, когда передаваемый радиоимпульс прямоугольной формы может принять (как правило, с одинаковой вероятностью) одно из $M=2^k$ значений: $u_0(t)$, $u_1(t)$,…, $u_{M-1}(t)$, отличающихся друг от друга начальными фазами и даже амплитудами (при использовании КАМ-M). Напомним, что M-ичный номер сигнала при передаче обычно заменяет k-разрядную комбинацию двоичных символов (см., например, рис. 2.18), поэтому полная вероятность ошибки при демодуляции принимаемого импульса P не совпадает с BER.

Здесь демодуляция также осуществляется в два этапа, но на первом этапе при помощи двух параллельно включенных корреляционных приемников вычисляются две величины v_1 и v_2, соответствующие двум квадратурным составляющим принимаемого сигнала с шумом (1.34), который одновременно подается на входы обоих приемников. Очевидно, что для этого в качестве опорного колебания $w(t)$ на вход первого приемника подаются колебания несущей с “нулевой” начальной фазой (ось C на рис. 2.19, хотя для сигнала с двукратной ФМ значительно удобнее задать начальную фазу, равную $\pi/4$). Опорное колебание на входе второго приемника отличается по фазе на $\pi/2$.

Для выбора способа обработки на втором этапе (вынесение решения) учтем, что областью возможных значений величин v_1 и v_2 является вся плоскость CS на рис. 2.18 и 2.19. Правило вынесения решения очевидно: в качестве переданного считают то значение полезного сигнала из M возможных, которое оказывается наиболее близким к точке с координатами (v_1,v_2), отображающей принятую реализацию сигнала с шумом. Значение передаваемого

символа, выбранное таким способом, называется *максимально правдоподобным*. Чтобы при приеме серии импульсов каждый раз не определять M расстояний, следует заранее плоскость CS разбить на M частей, соответствующих каждому из сигналов (пример такого разбиения штриховыми линиями для сигнала с КАМ-16 дан на рис. 5.3).

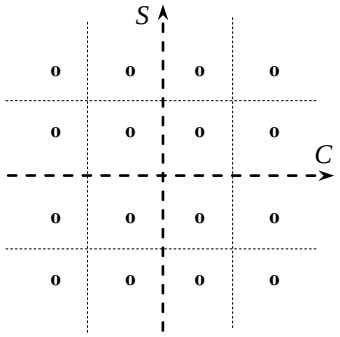

Рис. 5.3. Подмножества, соответствующие каждому из сигналов КАМ-16

Существуют ошибки $M(M–1)$ видов, причем, в отличие от двоичного случая, ошибки разных видов неравновероятны. Поскольку нормальная двумерная совместная плотность вероятности шумовых составляющих координат точки (v_1,v_2) имеет круговую симметрию, из всех возможных видов ошибки наиболее вероятно попадание точки, соответствующей принятому сигналу с шумом, в соседнюю область (по горизонтали либо по вертикали). При этом чаще возникают ошибки при демодуляции сигналов, которым соответствуют внутренние точки на диаграмме типа изображенной на рис. 5.3 (у них больше соседей).

Тогда точное и приближенное (с некоторым превышением) выражения для полной вероятности ошибки при демодуляции принимаемого импульса при $M = m^2$ имеют вид

$$P = 4p\left(1-1/m\right) - 4p^2\left(1-2/m+1/m^2\right) \approx 4p\left(1-1/m\right), \qquad (5.16)$$

где $p = 1-\hat{O}(\Delta q/2)$ – вероятность ошибки (5.13) при демодуляции двоичного ФМ сигнала,

Δq^2 – отношение энергии *разности двух ближайших сигналов* к спектральной плотности шума.

Для двукратной и трехкратной ФМ (рис. 2.18) при Δq >4…6 формула для полной вероятности ошибки имеет следующий вид:

$$P \approx 2\left[1-\Phi(\Delta q/2)\right]. \qquad (5.17)$$

Формулы (5.16) и (5.17) тем точнее, чем больше M и Δq.

Очевидно, что увеличение битовой скорости передачи, которое достигается при использовании многопозиционных методов модуляции, потребует существенного увеличения средней энергии полезного сигнала на входе приемника, пропорциональной величине q_c^2. *Здесь частотная эффективность*

системы улучшается за счет снижения ее энергетической эффективности. В таблице 5.1 приведены значения отношения $(q_c/\Delta q)^2$ для разных видов модуляции.

В заключение отметим, что при использовании многопозиционных методов модуляции BER зависит еще и от способа кодировки возможных зна-

Таблица 5.1. Значения отношения $(q_c/\Delta q)^2$

Вид модуляции	$(q_c/\Delta q)^2$
ФМ	0,25
КАМ-4	0,5
двукратная ФМ	0,5
трехкратная ФМ	1,71
КАМ-16	2,5
КАМ-64	10,5
КАМ-256	42,5
КАМ-1024	170,5

чений *M*-ичного символа при помощи *k*-разрядных двоичных комбинаций. Наилучшим является такой код, когда соседним значениям импульса соответствуют комбинации, отличающиеся лишь в одном символе. Коды, показанные на рис. 2.18б и 2.18в, не удовлетворяют этому требованию. В частности, при использовании трехкратной ФМ для преобразования восьмиричной цифры в двоичную комбинацию существует код Грея (0↔000; 1↔001; 2↔011; 3↔010; 4↔110; 5↔111; 6↔101; 7↔100). Здесь при возникновении наиболее вероятных ошибок из трех битов комбинации ошибочным оказывается лишь один, в итоге для BER имеем $p = P/3$.

5.3 Некогерентные системы

В некогерентной СПИ могут использоваться лишь АМ или ЧМ, поэтому в такой системе невыгодно применять многопозиционные методы передачи. В связи с неопределенностью начальной фазы ожидаемого импульса во время приема не имеет смысла понятие разностного сигнала. Оптимальным способом предварительной обработки сигнала с белым шумом и в этом случае является корреляционный прием, но необходимо использовать два (для АМ) или четыре (для ЧМ) параллельно включенных приемника. Каждая пара приемников настроена на прием двух квадратурных составляющих импульса определенного вида, поэтому в ней используются опорные колебания соответствующей несущей частоты, но сдвинутые по фазе на 90° (напомним, что при АМ лишь один из двух символов является импульсом, а второй – это пауза, поэтому нужна лишь одна пара приемников).

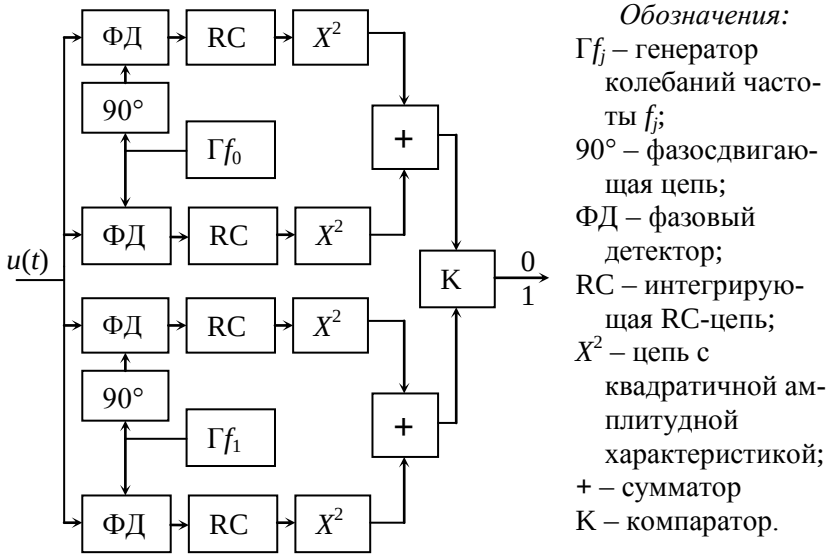

Рис. 5.4. Функциональная схема демодулятора некогерентного сигнала с ЧМ

На рис. 5.4 приведена функциональная схема демодулятора некогерентного сигнала с ЧМ, причем фазовые детекторы играют роль перемножителей высокочастотного входного сигнала и гармонического опорного колебания, а RC-цепь с постоянной времени, намного превышающей длительность импульса входного сигнала, выполняет функции интегратора (см. формулу (2.27)). Компаратор выносит решение “передан символ 0”, если в момент окончания импульса напряжение на выходе верхнего сумматора оказалось больше, чем на выходе нижнего.

Обратите внимание, что на схеме нужно было бы показать цепи, по которым подаются тактовые импульсы, задающие моменты срабатывания компаратора и осуществляющие обнуление интегрирующей цепочки непосредственно перед началом каждого импульса.

Вероятность ошибки при демодуляции двоичного импульса (BER) минимальна при заданной величине q_c, когда передается один из двух ортогональных сигналов равных энергий, например, с ЧМ, и определяется формулой

$$p = P = 0{,}5\exp\left(-q_c^2/2\right). \qquad (5.18)$$

Видно, что и в некогерентной СПИ вероятность ошибки определяется величиной энергии принимаемого импульса по отношению к спектральной плотности шума. Эта зависимость также представлена на рис. 5.2, чтобы можно было сравнить качество передачи сигнала в когерентной и некогерентной системах при прочих равных условиях.

Как и следовало ожидать (см. разд. 5.1), BER в когерентной системе всегда меньше чем в некогерентной. При приеме сильного сигнала (величина BER мала, нижняя штриховая прямая на рис. 5.2), чтобы в некогерентной си-

стеме обеспечить то же значение BER, которое достигается в когерентной СПИ, нужно лишь немного увеличить мощность передатчика – на единицы и даже доли децибела (напомним, что в децибелах по формуле $10\cdot\lg(E_1/E_2)$ выражается отношение энергий или мощностей двух сигналов). При приеме очень слабого сигнала это отношение возрастает до единиц и десятков децибел, то есть именно в этих условиях преимущество когерентной СПИ особенно ощутимо. Отметим, кстати, что в нормальных условиях функционирования СПИ величина q_c обычно превышает 10–15 дБ.

5.4 Частично-когерентные системы

В частично-когерентной системе оптимальный демодулятор должен выполнять те же операции (5.3) и (5.4), но текущая фаза генератора опорного колебания $w(t)$ подстраивается под изменения этого параметра в принимаемом сигнале при помощи системы ФАПЧ. Схема такого генератора для **СПИ с АМ или ЧМ** представлена на рис. 5.5, где обозначено: ФД – фазовый детектор; ФНЧ – фильтр нижних частот; ГУН – генератор, управляемый напряжением. Это простейшая система автоматического регулирования. В

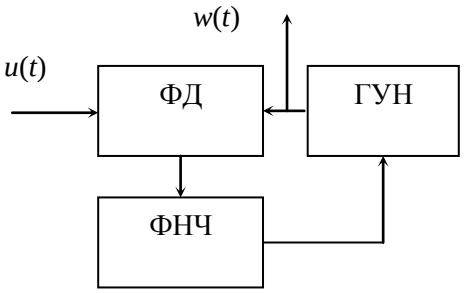

Рис. 5.5. Генератор опорного колебания с петлей ФАПЧ

ней при рассогласовании фазы генератора относительно входного сигнала в ту или другую сторону на выходе ФД появляется напряжение соответствующего знака. Оно воздействует на частоту и, следовательно, фазу генератора. Фильтр НЧ, у которого частота среза во много раз меньше частоты повторения импульсов, служит для сглаживания флуктуаций управляющего напряжения, обусловленных шумом.

Если в СПИ для каждого из возможных сигналов имеется противоположный сигнал (однократная или многократная ФМ либо ОФМ, КАМ), такой генератор работать не будет, так как в процессе фильтрации в ФНЧ эти сигналы, появляясь в среднем одинаково часто, практически уничтожают друг друга (в спектре каждого из таких сигналов отсутствует составляющая несущей частоты).

По этой причине в частично-когерентной системе перечисленные методы модуляции применяются лишь в относительной форме. Это позволяет

проводить ФАПЧ не на самой несущей частоте, а на одной из ее гармоник.

В частности, в **двоичной СПИ с ОФМ** подстройку генератора следует проводить на удвоенной несущей частоте (рис. 5.6). При удвоении частоты начальная фаза каждого из импульсов также удваивается, в итоге на выходе получается непрерывное колебание, что обеспечивает устойчивую работу цепи ФАПЧ с высокой точностью.

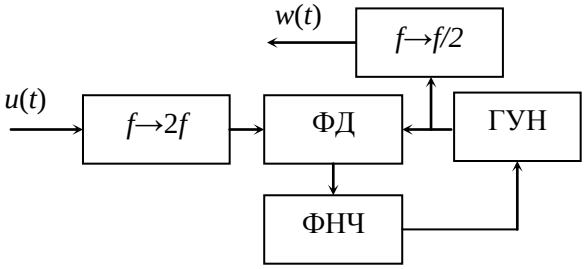

Рис. 5.6. Генератор опорного колебания с петлей ФАПЧ на удвоенной несущей частоте

Тем не менее, применять ФМ в такой системе не следует, поскольку формируемое опорное колебание все-таки обладает некоторой неопределенностью. Классический тип делителя частоты – это счетный триггер. В зависимости от случайного начального состояния этого триггера на его выходе формируется одна из двух противоположных последовательностей. В итоге, хотя и обеспечивается хорошая взаимная когерентность колебаний $w(t)$ и колебаний принимаемого сигнала, но $w(t)$ может быть либо синфазна, либо противофазна по отношению к сигналу. В первом случае сигнал с ФМ принимается нормально, во втором случае все символы на выходе демодулятора изменяются на обратные. Это явление называется обратной работой детектора. Конечно, можно было бы периодически передавать какую-нибудь настроечную последовательность символов, чтобы определять, в котором из двух режимов находится демодулятор, но, как показала практика, намного проще и надежнее применять ОФМ.

Тогда устройство, выносящее решение по каждому принимаемому импульсу в соответствии с (5.4) так, как будто бы принимается ФМ-сигнал, должно быть дополнено схемой, сравнивающей решения, вынесенные по каждым двум соседним импульсам. Такая схема состоит из ячейки сдвига и сумматора по модулю 2 (рис. 5.7) и одинаково хорошо работает как при прямой, так и при обратной работе демодулятора (5.4). *Описанный способ демодуляции сигнала с ОФМ называется когерентным*.

Вероятность ошибки на первом этапе $p_{\text{ФМ}}$, описываемом выражением

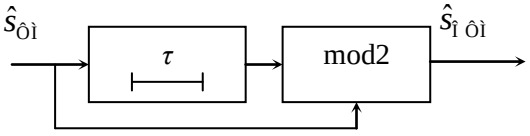

Рис. 5.7. Схема сравнения решений, вынесенных по соседним импульсам

(5.4), определяется формулой (5.13). Ошибка на втором этапе (рис. 5.7) произойдет в одном из двух случаев:

1) на первом этапе при демодуляции первого из двух импульсов произошла ошибка, второй импульс принят правильно;

2) первый импульс принят правильно, второй – с ошибкой.

В итоге битовая вероятность ошибки при демодуляции сигнала с ОФМ равна

$$p_{\text{î ôì}} = 2p_{\text{ôì}}\left(1-p_{\text{ôì}}\right) = 2\hat{\text{O}}(q/2)\left(1-\hat{\text{O}}(q/2)\right). \qquad (5.19)$$

На практике обычно q>>1, откуда $1-p_{\text{ôì}} \approx 1$, поэтому

$$p_{\text{î ôì}} \approx 2p_{\text{ôì}}\ , \qquad (5.20)$$

то есть *BER в частично- когерентной СПИ с ОФМ всего лишь в два раза больше чем BER в идеальной, когерентной СПИ с ФМ.* Этим объясняется столь широкое использование ОФМ и когерентного метода приема таких сигналов.

Существует **также некогерентный метод демодуляции сигнала с ОФМ.** Для реализации этого метода на фазовый детектор приемника, рассчитанного на демодуляцию ФМ-сигналов и, следовательно, работающего в соответствии с (5.3) и (5.4), подается в качестве опорного колебания $w(t)$ предыдущий принятый радиоимпульс. Его можно получить с выхода линии задержки.

Битовую вероятность ошибки можно вычислить по формуле (5.18), но величину q_c^2 следует удвоить, так как решение выносится после приема составного сигнала, содержащего два импульса. Здесь величина BER больше, чем при использовании когерентного метода приема, так как опорное колебание кроме составляющей, когерентной с полезным принимаемым сигналом, содержит также шумовую составляющую. Эта разница тем больше, чем меньше отношение сигнал/шум.

При использовании **многопозиционных методов модуляции**, кроме АМ и ЧМ, необходимо проводить ФАПЧ опорного генератора на гармониках несущей частоты более высокого порядка. Так, *в СПИ с двукратной ФМ или с КАМ генератор подстраивается по четвертой гармонике, а для трехкратной ФМ нужно перейти на восьмую гармонику*.

В заключение отметим интересную особенность СПИ, использующих относительные методы манипуляции в любой форме. Это – *парное группирование ошибок*, то есть при увеличении отношения сигнал/шум возрастает относительная доля ошибок, поражающих два соседних символа, по сравнению с одиночными ошибками. Это свойство нужно учитывать при выборе корректирующего кода. Например, код с проверкой на четность не будет замечать такие ошибки, если внутри кодовой комбинации окажутся оба символа пары, что весьма вероятно при использовании длинных комбинаций.

5.5 Прием сигнала в условиях многолучевости

Как видно из обсуждения, проведенного в разд. 5.1, реализация когерентной СПИ была бы возможна лишь при использовании канала передачи сигналов, идеального в том смысле, что его амплитудно-частотная и фазочастотная характеристики неизменны во времени и, следовательно, могут быть точно известны в пункте приема. Частично-когерентная и некогерентная СПИ также работают в условиях, почти идеальных, когда фазовый сдвиг в канале медленно меняется во времени одинаково на всех частотах. Обеспечить такие условия можно при передаче сигнала по кабельной линии либо при радиосвязи, когда радиотрасса проходит в однородной среде вдали от объектов, могущих вызвать отражение или рассеяние радиоволн, например, в линии спутниковой связи.

При распространении радиоволн в неоднородной тропосфере на расстояние всего лишь в десятки километров, над неровной земной или водной поверхностью и, тем более, в условиях города поле, падающее на приемную антенну, является суммой большого количества волн, часто называемых лучами. Эти волны обусловлены отражениями первичного поля от различных объектов (зданий, деревьев, автомобилей, холмов и т.п.), находящихся вблизи трассы. Разделить эти лучи и определить параметры каждого из них обычно не представляется возможным, поэтому полагают, что амплитуды и задержки (следовательно, и фазовые сдвиги) всех лучей случайны и даже взаимно независимы. Для полноты картины добавим, что обычно все эти параметры еще и случайным образом изменяются во времени. Такой канал называется *каналом со случайно изменяющимися параметрами*, а искажения сигнала, передаваемого в таком канале, классифицируются как воздействие *мультипликативной помехи* (см. разд. 1.5).

На рис. 5.8 в качестве примера приведена одна из реализаций случайно изменяющейся амплитуды принимаемого сигнала.

Даже если среднее значение амплитуды $A_с$ достаточно велико, существуют такие случайно расположенные интервалы времени случайной длительности (помечены жирными линиями), когда амплитуда падает ниже уровня $A_п$, при котором еще возможен нормальный прием сигнала, то есть его демодуляция с заданной величиной BER. Можно считать, что при этом связь нарушается. Очевидно, что величина $A_п$ определяется уровнем аддитивного шума на входе приемника.

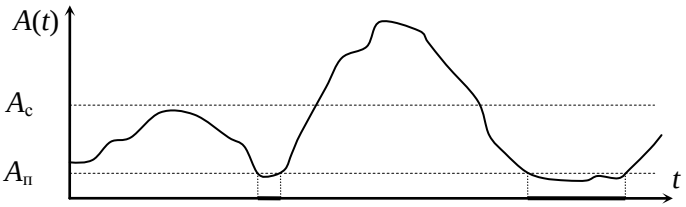

Рис. 5.8. Пример флуктуаций амплитуды принимаемого сигнала в канале с многолучевостью

Обычно полагают, что случайная амплитуда сигнала имеет релеевское распределение вероятности

$$W(A)=\frac{A}{\sigma^2}\exp\left(-\frac{A^2}{2\sigma^2}\right), \qquad A\geq 0, \tag{5.21}$$

где σ – среднеквадратическое значение флуктуаций каждой из квадратурных составляющих сигнала, при этом $A_с$=1,2533σ
(рис. 5.9). Вероятность того, что в заданный момент времени амплитуда сигнала окажется ниже порога $A_п$, определяется обычным образом

$$P_з=\int_0^{A_п} W(A)dA=1-\exp\left(-\frac{A_п^2}{2\sigma^2}\right) \tag{5.22}$$

и численно равна площади заштрихованного участка на рис. 5.9. Часто $A_п$ <<σ, поэтому (5.22) упрощается

$$P_з\approx\frac{A_п^2}{2\sigma^2} \quad \text{при } A_п<<\sigma. \tag{5.23}$$

Тогда величину среднего суммарного времени нарушения связи на интервале T можно определить как $T_ç=TP_ç$.

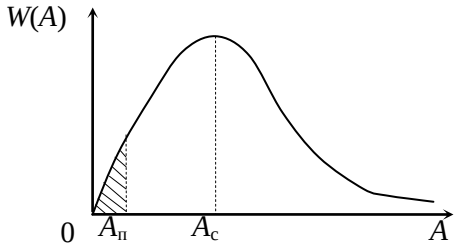

Рис. 5.9. Релеевское распределение вероятности амплитуды сигнала

Конечно, уменьшить это время можно путем увеличения мощности передатчика, но существует более изящный и, можно сказать, более эффективный способ. Он называется *разнесенным приемом* и применяется как в аналоговых, так и в цифровых СПИ.

Для реализации ***N*-кратного разнесенного приема**, во-первых, организуют N параллельных каналов (ветвей разнесения) для одновременной передачи одного и того же сигнала и, во-вторых, применяют специальное устройство для того, чтобы из этих N принятых сигналов сформировать один сигнал, обладающий лучшими свойствами, нежели сигналы в ветвях разнесения.

Если флуктуации сигналов в ветвях разнесения полностью идентичны

(коэффициент корреляции для любой пары r=1), сигнал на выходе устройства комбинирования не приобретет новых свойств, так как прерывание связи будет происходить одновременно во всех ветвях разнесения. *Флуктуации сигналов в ветвях разнесения должны быть независимы.* При этом условии вероятность одновременного прерывания связи во всех ветвях разнесения минимальна (строго говоря, наилучший вариант – это прием двух сигналов при $r = -1$, но такого в реальных каналах, к сожалению, не наблюдается).

В настоящее время чаще других применяются следующие **способы организации *N* ветвей разнесения**.

Пространственный разнесенный прием, когда прием одного и того же сигнала осуществляется на N приемных антенн, разнесенных в пространстве на расстояние порядка 10…100 длин волн (лучше поперек трассы), чтобы обеспечить независимость флуктуаций, причем меньшие значения характерны для условий распространения в городе.

Частотный разнесенный прием, когда передача (соответственно, и прием) одного и того же сообщения производится с использованием общей передающей (приемной) антенны на N несущих, разнесенных на 100…10000 кГц.

Пространственно-частотный разнесенный прием, когда сочетаются передача на $N_ч$ несущих и одновременный прием на $N_п$ антенн, при этом получается $N = N_ч N_п$ ветвей разнесения. Весьма популярен метод четырехкратного разнесения (две несущих и две антенны).

В результате комбинирования N разнесенных сигналов амплитуда сигнала на выходе устройства комбинирования становится более устойчивой (уменьшается размах флуктуаций относительно среднего значения), вследствие чего одновременно улучшаются две важные характеристики цифровой СПИ: $P_з$ и BER. Обычно применяется один из трех **методов комбинирования.**

Автовыбор максимального сигнала, то есть на выход устройства комбинирования поступает тот из сигналов, который в данный момент имеет максимальную амплитуду. Пусть отношение сигнал/шум достаточно велико для того, чтобы вероятность нарушения связи без применения разнесения $P_{з1}$ можно было определить по формуле (5.23). Тогда при использовании разнесенного приема связь будет нарушена, если одновременно во всех независимых ветвях амплитуда окажется ниже порога, в итоге $P_{зN} = (P_{з1})^N$. Например, $P_{з1} = 0{,}1$, тогда при использовании даже двукратного разнесения $P_{з2} = 0{,}01$.

Признаем, что приведенный пример не вполне удачен потому, что при переходе к N-кратному разнесению предполагается соответствующее увеличение суммарной энергии принятых сигналов, а это, очевидно, само по себе должно привести к улучшению характеристик СПИ. Поэтому более корректным будет сравнение в одинаковых условиях, когда *при переходе к разнесенному приему суммарная энергия принимаемых сигналов остается неизмен-*

ной, при этом, естественно, отношение сигнал/шум q_c^2 в каждой ветви разнесения уменьшается в N раз. Тогда для приведенного примера величина σ^2 в каждой ветви уменьшается вдвое, поэтому имеем $P_{з2} = 0{,}04$.

Тем не менее, из приведенного примера не следует делать поспешный вывод о том, что применение разнесенного приема даже при неизменной суммарной энергии принимаемых сигналов всегда полезно, поскольку должным образом не учтено влияние негативного фактора, а именно, уменьшения отношения сигнал/шум. Более объективным будет сравнение методов при дополнительном требовании обеспечения одного и того же значения BER, что мы сделаем позже, а пока определим остальные методы комбинирования сигналов.

Линейное сложение сигналов, то есть на выход устройства комбинирования поступает сумма сигналов из ветвей разнесения

$$u(t) = \sum_{j=1}^{N} u_j(t) \ . \tag{5.24}$$

Оптимальное сложение сигналов, то есть на выход устройства комбинирования поступает взвешенная сумма сигналов из ветвей разнесения

$$u(t) = \sum_{j=1}^{N} k_j(t) u_j(t), \tag{5.25}$$

где $k_j(t)$ – весовая функция, модуль которой пропорционален текущему значению амплитуды сигнала $u_j(t)$ в j-й ветви разнесения, в итоге более сильные сигналы суммируются с большим весом.

Существуют два варианта для реализации любого из методов сложения: до детектора (суммирование радиоимпульсов на высокой или промежуточной частоте) и после детектора (суммирование низкочастотных импульсов с шумом с выхода обычного линейного детектора, применяемого в каждой ветви). *При суммировании до детектора обязательной является операция предварительного взаимного фазирования всех радиосигналов*, чтобы текущие фазы всех слагаемых были одинаковы.

Эти три метода комбинирования расположены в порядке возрастания сложности их реализации и, одновременно, в порядке убывания величины BER, которую они способны обеспечить.

На рис. 5.10 дана зависимость **битовой вероятности ошибки** при различении двух равновероятных ортогональных сигналов от кратности разнесения при различных значениях суммарного отношения сигнал/шум в предположении, что флуктуации от импульса к импульсу независимы. Расчет проведен для способа оптимального сложения. Хотя кратность разнесения N – это целое число, ради наглядности на рисунке проведены сплошные линии.

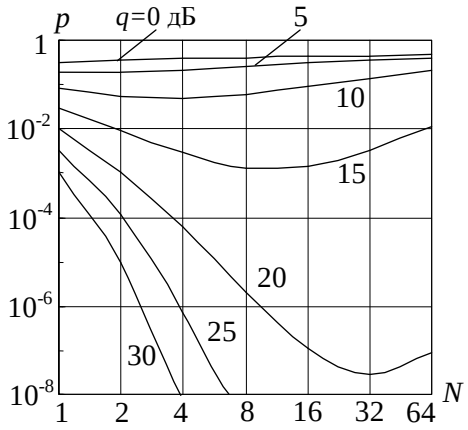

Рис.5.10. Зависимость битовой вероятности ошибки при различении двух равновероятных ортогональных сигналов от кратности разнесения при различных значениях отношения сигнал/шум

Видно, что применение разнесенного приема даже при фиксированной суммарной энергии сигналов, если суммарное отношение сигнал/шум больше 5-10 дБ, позволяет обеспечить меньшую величину BER по сравнению с вариантом неразнесенного приема. *При любом отношении сигнал/шум существует оптимальное значение $N_{опт}$, при котором вероятность ошибки минимальна.* При q > 15…20 дБ для оптимальной кратности имеем $N_{îïò} \approx q^2/3,2$. Прием слабых сигналов лучше проводить без разнесения. Таким образом, использование разнесенного приема эффективно лишь в том случае, когда принимаемые сигналы достаточно сильные.

5.6 Регенерация цифрового сигнала в ретрансляторах

При передаче на большие расстояния сигнал может быть ослаблен настолько, что прием становится невозможным. Чтобы избежать этого, на определенных расстояниях вдоль линии связи (3…5 км при передаче по медному кабелю, 40…60 км в радиорелейной линии) устанавливают промежуточные пункты, в которых производится восстановление ослабленного сигнала.

Используются два режима обработки принимаемого цифрового сигнала в таком ретрансляторе:

1) обычное усиление до того уровня, который этот сигнал имел на выходе предыдущего ретранслятора;

2) усиление, демодуляция каждого принятого импульса, то есть вынесение решения о том, какое именно значение символа передается, и формирование импульса правильной формы в соответствии с принятым решением (только в этом режиме, строго говоря, производится *регенерация сигнала*).

В первом режиме при передаче сигнала вдоль линии в нем *накапливаются шумы*, вносимые каждым из ретрансляторов. В итоге на выходе *N*-пролетной линии отношение сигнал/шум по энергии оказывается в *N* раз меньше того, что было бы при передаче по более короткой, однопролетной линии

$$q_N^2 = q_1^2 / N . \qquad (5.26)$$

Естественно, это влечет увеличение битовой вероятности ошибки в пункте приема.

Во втором режиме шумы не накапливаются ($q_N^2 = q_1^2$), но зато *накапливаются ошибки*, иногда возникающие в ретрансляторах при демодуляции (вынесении решений).

Итак, при увеличении длины линии помехоустойчивость передачи в обоих случаях ухудшается, вопрос лишь в том, насколько быстро. В качестве примера вычислим BER для некогерентной двухпролетной СПИ, использующей ортогональные сигналы.

По формуле (5.18) находим BER для одного пролета $p_1 = 0,5\exp\left(-q_1^2/2\right)$, тогда на выходе линии без регенерации имеем $p_2 = 0,5\exp\left(-q_c^2/4\right) = \sqrt{p_1/2}$.

На выходе линии с регенерацией сигнала ошибка происходит в одном из двух случаев:

1) ошибка при передаче на первом пролете, на втором ошибки нет;

2) ошибка при передаче на втором пролете, на первом ошибки нет.

В итоге для BER получим $p_2 = 2p_1(1-p_1)$. Обычно СПИ работает в условиях, когда $p_1 << 1$. Тогда величина BER при использовании регенерации примерно в $1/\sqrt{8p_1}$ раз меньше, чем в системе без регенерации, причем разница тем существенней, чем больше отношение сигнал/шум.

Мы рассмотрели пример *регенерации сигнала на низшем, первом уровне* – уровне “модуляции-демодуляции”. Если в СПИ применяется помехоустойчивое кодирование, то возможна *регенерация сигнала еще и на втором уровне* “кодирования-декодирования”, что дополнительно повышает помехоустойчивость СПИ.

В аналоговых СПИ регенерация сигнала невозможна. Если в цифровой СПИ в большинстве случаев удается *точно* воспроизвести значения переданного сигнала, то в аналоговой СПИ из-за воздействия помех при демодуляции принятого сигнала ошибки, пусть малые, происходят всегда. Вследствие этого регенерировать, то есть воспроизвести неискаженный сигнал не удается.

Возможность регенерации сигнала – это огромное преимущество цифровых методов передачи перед аналоговыми.

Отметим, что платой за это преимущество является повышение стоимости линии передачи. Поэтому при передаче по медному кабелю на один обслуживаемый регенерационный пункт приходится 20…40 обычных усили-

тельных пункта (их называют необслуживаемыми регенерационными пунктами).

5.7 Особенности СПИ, в которых применяется помехоустойчивое кодирование

Начнем с вопроса: **всегда ли применение помехоустойчивого кодирования повышает помехоустойчивость СПИ?** На первый взгляд, следует дать положительный ответ, поскольку в процессе декодирования удается обнаружить и даже исправить значительную часть ошибок, возникающих в процессе демодуляции принимаемых символов. Если же учесть, что бесплатных преимуществ не бывает, то ответ не будет столь однозначным.

Например, если в СПИ применить линейный блочный код, то вместо k информационных символов за то же время придется передать $n = k + r$ символов. Каждый передаваемый импульс при этом станет короче, и при той же мощности передатчика энергия каждого импульса и, следовательно, отношение сигнал/шум q_c^2 на входе демодулятора станут в n/k раз меньше. Результатом является увеличение BER. Другими словами, мы увеличиваем количество ошибочных символов на входе декодера для того, чтобы он мог исправить некоторые из них.

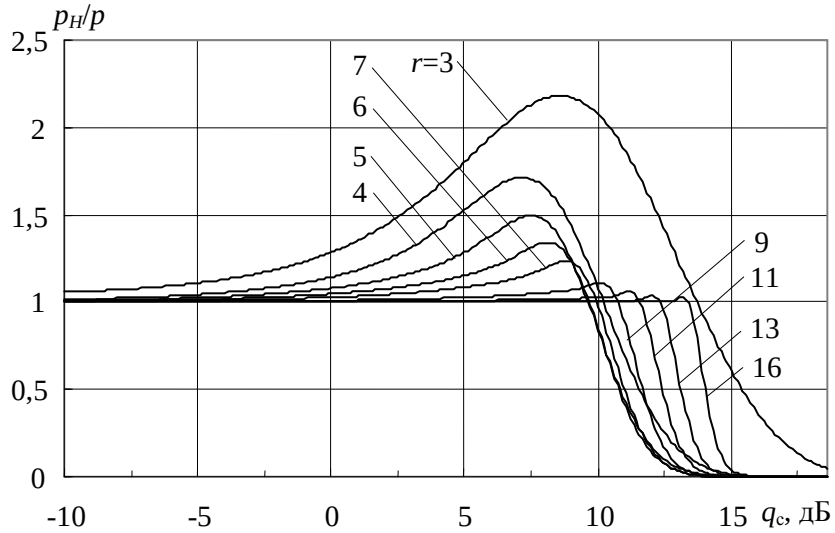

Рис.5.11. Отношение BER на выходе декодера кода Хэмминга с r проверочными символами к величине BER в некогерентной СПИ без кодирования

На рис. 5.11 приведены результаты расчета отношения битовой вероятности ошибки на выходе декодера кода Хэмминга с r проверочными символами к величине битовой вероятности ошибки в СПИ без кодирования. Сравнение проведено в одинаковых условиях, то есть при той же мощности передатчика некогерентной СПИ, причем q_c – это отношение сигнал/шум в СПИ без кодирования. Демодулятор работает в жестком режиме. Канал пере-

спроса отсутствует, то есть код Хэмминга используется для *исправления ошибок*.

Видно, что при приеме слабых сигналов (отношение сигнал/шум меньше 9,3 дБ) применение любого из кодов Хэмминга лишь ухудшает верность передачи. В таблице 5.2 даны минимальные значения отношения сигнал/шум (в децибелах) в СПИ без кодирования, при которых использование кода Хэмминга уменьшает BER.

Из данных, приведенных в таблице, видна закономерность, характерная для кодов с большим значением n: чем длиннее комбинации, тем больше минимальное значение q_c, при котором имеет смысл применять данный код.

Таблица 5.2. Минимальное значение q_c, при котором использование кода Хэмминга уменьшает BER

r	3	4	5	6	7	8
(n,k)	(7,4)	(15,11)	(31,26)	(63,57)	(127,120)	(255,247)
q_c, дБ	13,3	9,8	9,3	9,3	9,6	9,9

r	9	10	11	12
(n,k)	(511,502)	(1023,1013)	(2047,2036)	(4095,4083)
q_c, дБ	10,3	10,8	11,2	11,6

Эта же мысль подтверждается данными таблицы 5.3, в которой приведены интервалы значений отношения сигнал/шум, в которых данный код является наилучшим среди всех кодов Хэмминга.

Таблица 5.3. Интервал значений q_c, в котором код Хэмминга с указанным числом проверочных символов обеспечивает минимум BER

r	0 (без кодирования)	5	6	7	8
q_c, дБ	<9,3	9,3-10,0	10,0-12,5	12,5-14,8	14,8-17,1

Вторая проблема, которую мы рассмотрим, это целесообразность использования **жесткой или мягкой процедур принятия решения.**

Рассмотрим сначала процедуру принятия решения в демодуляторе двоичной СПИ, которая определяется формулой (5.4). Эту операцию можно назвать двоичным квантованием непрерывной величины v, поступающей с выхода корреляционного приемника, поскольку область возможных значений этой величины ($-\infty<v<\infty$) разбита на два интервала точкой $v_п$. Решающее устройство указывает лишь номер интервала $\hat{s}$ (0 или 1), в который попало значение v для данного принятого импульса. Такая процедура принятия решения называется *жесткой*.

Возможен и другой подход, когда область возможных значений величины v разбита на три интервала точками $v_{п0}$ и $v_{п1}$ (рис. 5.12). Тогда решение, которое должен вынести демодулятор, определяется следующей инструкцией

$$\hat{s} = \begin{cases} 0, & \text{если } v \geq v_{n0}, \\ \text{символ не определен}, & \text{если } v_{n1} > v > v_{n0}, \\ 1, & \text{если } v \leq v_{n1}. \end{cases} \qquad (5.27)$$

Такая процедура принятия решения называется *мягкой*, при этом демодулятор производит троичное квантование непрерывной величины v.

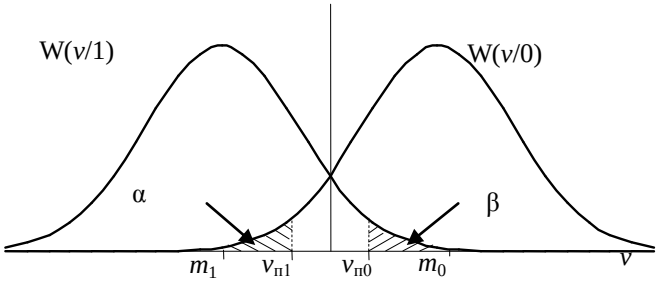

Рис. 5.12 – Условные плотности вероятности сигнала на выходе интегратора корреляционного приемника

Из рис. 5.12 хорошо видно, что демодулятор указывает конкретное значение символа лишь в тех случаях, когда наблюдаемое значение v является достаточно большим или достаточно маленьким, то есть имеется высокая степень уверенности в том, какое именно значение символа передается в данный момент. Считают, что в остальных случаях *символ стирается* (позиция данного символа, разумеется, сохраняется, и на эту позицию следующие устройства должны все-таки поставить 0 или 1).

При использовании мягкой процедуры уменьшаются вероятности ошибок α, β и p, но появляется некоторое количество стертых символов. Поэтому *мягкую процедуру демодуляции можно использовать лишь в том случае, когда применяется помехоустойчивое кодирование, чтобы декодер корректирующего кода мог восстановить значения стертых символов.*

Наличие стертых символов усложняет работу декодера (при использовании сверточного кода и алгоритма декодирования Витерби это усложнение незначительно), но зато существенно повышает помехоустойчивость СПИ.

Можно развить эту идею дальше, и область возможных значений величины v разбить не на два, а, скажем, на восемь интервалов. Тогда демодулятор в качестве решения будет выдавать одну из следующих комбинаций: 111, 110, 101, 100, 000, 001, 010, 011, где первый бит указывает на предполагаемое значение принимаемого двоичного символа, а два остальных бита характеризуют степень надежности этого решения. Использование такой еще более мягкой процедуры позволяет после декодирования обеспечить то же значе-

ние BER, что и в жестком режиме, но требуемая мощность сигнала примерно на 2 дБ меньше [2]. Декодирование при этом становится еще сложнее.

Дальнейшее увеличение числа интервалов N (переход к абсолютно мягкой процедуре работы демодулятора) также повышает помехоустойчивость СПИ, но выигрыш уже незначителен. Эта предельная ситуация носит название *приема в целом*.

Мягкий режим работы демодулятора и декодеров нередко применяют в условиях комбинирования кодов. Весьма характерным является использование **турбо-кодов.**

Они основаны на нескольких принципах.

1) Использование нескольких кодирующих (соответственно, декодирующих устройств). В первом из предложенных турбо-кодов (1993 год) применялись два систематических сверточных кода со степенью кодирования 1/2. В линию передачи подавались информационные символы и обе проверочные последовательности.

2) На вход первого кодера последовательность информационных символов подается непосредственно, а на вход второго – после перемежения.

3) Демодулятор приемника работает в мягком режиме.

4) В процессе декодирования производится несколько итераций, на каждой из которых уточняются оценки вероятностей символов. Оценки, полученные одним декодером, передаются другому декодеру в качестве априорных данных.

5) На каждой итерации декодеры принимают мягкие решения, и лишь последняя завершается жестким решением.

По своим свойствам турбо-коды близки к случайным кодам и их энергетическая эффективность высока. Уже у первого кода она оказалась лишь на 0,5 дБ ниже теоретического предела [2].

Вопросы для самоконтроля по главе 5

1. Дайте определения когерентной и некогерентной СПИ.
2. Сформулируйте задачу, решаемую демодулятором сигнала в цифровой СПИ.
3. Опишите преобразования принимаемых импульсов при демодуляции двоичного сигнала, не искаженного в канале передачи.
4. Приведите схему оптимального демодулятора двоичных сигналов в когерентной СПИ.
5. Приведите схему оптимального демодулятора двоичных сигналов в частично-когерентной СПИ при использовании ОФМ.
6. Запишите формулу для определения полной вероятности ошибки на выходе оптимального демодулятора двоичных сигналов в когерентной СПИ.

7. Запишите формулу для определения полной вероятности ошибки на выходе оптимального демодулятора двоичных сигналов в частично-когерентной СПИ при использовании ОФМ.
8. Почему в цифровых СПИ не применяются методы ОФМ с кратностью большей трех?
9. В которой из радиолиний – ‘Земля – ИСЗ’ или ‘ИСЗ – Земля’ – можно обеспечить более высокое качество передачи и почему?
10. Покажите, что при большом отношении сигнал/шум некогерентная СПИ мало уступает когерентной СПИ.
11. Укажите физические явления, приводящие к тому, что передаточные характеристики канала связи становятся случайными.
12. Дайте определение многолучевой линии.
13. При каких условиях становятся существенными искажения сигнала, обусловленные частотно-селективными замираниями?
14. Какой метод является основным для повышения устойчивости связи в каналах с замираниями?
15. Опишите методы комбинирования разнесенных сигналов.
16. Почему применение автоматической регулировки усиления при одиночном приеме сигнала в канале с замираниями не уменьшает вероятности ошибки при его демодуляции?
17. Всегда ли целесообразно применять помехоустойчивое кодирование для уменьшения итоговой битовой вероятности ошибки?
18. При каких условиях можно использовать мягкую процедуру вынесения решения в процессе приема цифровых сигналов?
19. Какой обработке подвергается цифровой сигнал в регенераторе?
20. Перечислите преимущества цифровых СПИ перед аналоговыми.
21. Какова суть порогового эффекта, характерного для цифровых СПИ и проявляющегося при изменении уровня полезного сигнала по отношению к уровню помех.

6 МНОГОКАНАЛЬНАЯ ПЕРЕДАЧА И МНОГОСТАНЦИОННЫЙ ДОСТУП

6.1 Методы многостанционного доступа

Система передачи информации из одного пункта в другой называется *многоканальной* (*n*-канальной), если в ней по одному кабелю осуществляется *одновременная передача n сообщений*.

Многостанционным доступом (МД) называется метод, позволяющий *n* станциям радиосвязи (возможно, разбросанным на большой территории) *совместно использовать ресурсы одного ретранслятора*.

В сущности, оба определения совпадают; различие, определяющее область использования каждого из них, заключается лишь в том, какая среда передачи сигналов используется: кабель или эфир. Поэтому нет ничего удивительного в том, что возможные *методы реализации обеих технологий одинаковы*.

На рис. 6.1 изображена упрощенная (содержащая лишь принципиально необходимые элементы) схема передающей и приемной частей многоканальной СПИ, предназначенной для передачи в одном направлении, где обозначено: $И_1$, …, $И_n$ – источники информации (сообщений); ГП – генераторы поднесущих; $М_1$, …, $М_n$ – канальные модуляторы; УУК – устройство уплотнения каналов; УРК – устройство разделения каналов; $Д_1$, …, $Д_n$ – канальные демодуляторы; $П_1$, …, $П_n$ – получатели сообщений.

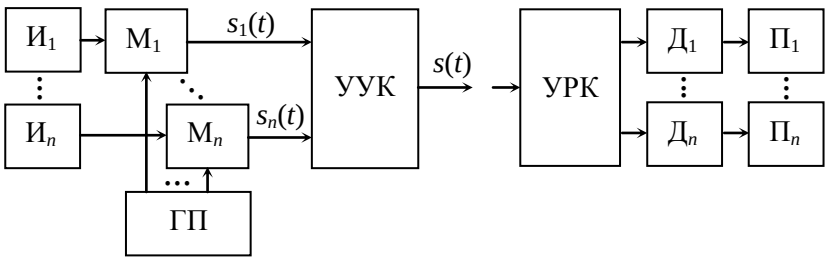

Рис. 6.1. Упрощенная схема многоканальной СПИ

Генератор поднесущих выдает *n* периодических сигналов (гармонических, последовательностей импульсов и т.п.), имеющих разную форму в разных каналах. Вид поднесущих и, следовательно, возможные способы их модуляции определяются методом уплотнения каналов. Модулированные поднесущие $s_1(t)$, …, $s_n(t)$ называются *канальными сигналами*. Операция уплотнения каналов – это обычное суммирование канальных сигналов

$$s(t) = \sum_{j=1}^{n} s_j(t). \qquad (6.1)$$

В результате этого получается *групповой сигнал* $s(t)$.

Для разделения каналов, как правило, применяются также линейные устройства. Считают, что в приемнике происходит разделение канальных сигналов без возникновения *междуканальных помех*, если выполняется следующее условие: при любых значениях сообщений, передаваемых в каналах, напряжение на *j*-м выходе линейного устройства разделения каналов равно нулю тогда и только тогда, когда $s_j(t)=0$.

Известно, что для выполнения этого условия необходимо и достаточно, чтобы *система n функций, описывающих канальные сигналы, была линейно независимой при любых значениях передаваемых сообщений*. Формально это условие проверяется так: система линейно независима, если не равен нулю определитель матрицы Грама (1.47)

$$\det \mathbf{\Gamma} \neq 0, \qquad (6.2)$$

Если выполняется лишь это условие, устройство разделения каналов может оказаться достаточно сложным. Поэтому обычно выдвигают еще одно, дополнительное требование: канальные сигналы должны быть к тому же попарно ортогональны

$$\Gamma_{jk} = \int_{-\infty}^{\infty} s_j(t)s_k(t)dt = \begin{cases} E_j, & j=k, \\ 0, & j\neq k, \end{cases} \qquad (6.3)$$

где E_j – энергия *j*-го канального сигнала. При этом матрица Грама становится диагональной.

Приведем **примеры типичных СПИ, использующих ортогональные канальные сигналы.**

1) Канальные сигналы не перекрываются во времени (рис. 6.2). Условие ортогональности (6.3) здесь, несомненно, выполняется, так как при любом *t* подынтегральное выражение равно нулю при $j\neq k$. Это метод *временного разделения каналов* (ВРК). Используются также термины: временное мультиплексирование, многостанционный доступ с временным разделением каналов (МДВР), Time Division Multiple Access (TDMA).

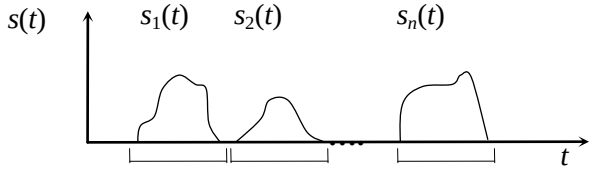

Рис. 6.2 Канальные сигналы, не перекрывающиеся во времени

Фактически линия связи поочередно подключается на короткое время к каждому из источников. Поэтому принципиальными элементами такой СПИ являются мультиплексор (устройство уплотнения каналов) и демультиплексор (устройство разделения каналов), всегда работающие синхронно (рис.

6.3). Поэтому *неотъемлемой составной частью любой СПИ с ВРК является система синхронизации приемного устройства.*

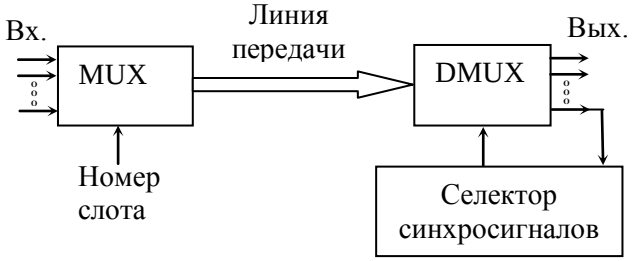

Рис. 6.3. Принципиальные элементы СПИ с ВРК

Важно отметить, что длительность канального интервала (слота) мала (в противном случае у пользователей исчезнет эффект “одновременности” передачи), поэтому за время одного подключения к линии удается передать лишь малую часть сообщения: один бит, один байт, один отсчет либо пакет, то есть небольшую группу подобных элементов. Чтобы успеть опросить все источники и неоднократно повторять этот процесс, все перечисленные элементы сигналов источников должны быть сжаты во времени по крайней мере в n раз.

2) Спектры канальных сигналов не перекрываются на оси частот. Спектр группового сигнала (сумма спектров канальных сигналов) показан на рис. 6.4. Условие ортогональности (6.3) здесь выполняется для спектров, следовательно, и для функций времени. Это метод *частотного разделения каналов* (ЧРК). Используются также термины: частотное мультиплексирование, многостанционный доступ с частотным разделением каналов (МДЧР), Frequency Division Multiple Access (FDMA).

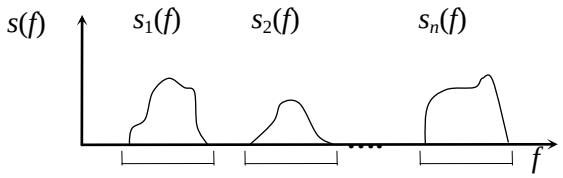

Рис. 6.4. Спектры канальных сигналов,
не перекрывающиеся на оси частот

Каждой паре “источник-получатель” выделена своя полоса частот. Поэтому принципиальными элементами такой СПИ являются блок n генераторов гармонических колебаний разных частот-поднесущих в передатчике и *набор полосовых фильтров для разделения канальных сигналов в приемнике*. Поскольку невозможно построить канальные полосовые фильтры с идеальной прямоугольной частотной характеристикой, между полосами, занятыми канальными сигналами, всегда предусматривают пустые *защитные частотные интервалы*.

3) Все канальные сигналы передаются одновременно (перекрываются

во времени) в одной общей полосе частот (их спектры также перекрываются), но взаимная ортогональность сигналов обеспечена тем, что все они имеют разную форму. Тогда *устройство разделения каналов – это набор n параллельно включенных корреляционных приемников или согласованных фильтров* (разд. 2.9), на входы которых подается принимаемый групповой сигнал. В каждом приемнике весовая функция $w(t)$ совпадает по форме с соответствующим канальным сигналом, поэтому он реагирует лишь на сигнал данного канала. Последнее требование удается выполнить лишь в цифровой СПИ.

Такой метод называется *многоканальной передачей с разделением по форме сигналов*. В качестве набора канальных сигналов можно использовать любую известную из курса математики систему ортогональных функций (функции Эрмита, Лагерра, Чебышева и т.п.), но все эти функции имеют довольно сложную форму. Поэтому для практического использования предложен более простой вариант метода, когда в качестве канальных сигналов применяются импульсы с внутриимпульсной фазовой модуляцией, причем закон модуляции каждого канального сигнала задается индивидуальной двоичной кодовой комбинацией.

В таком виде метод получил название *кодового разделения каналов* (КРК). Используются также термины: кодовое мультиплексирование, многостанционный доступ с кодовым разделением каналов (МДКР), Code Division Multiple Access (CDMA).

На практике при использовании любого метода уплотнения канальных сигналов после их разделения в приемнике возникают *междуканальные помехи*, то есть сигнал на выходе канального демодулятора зависит не только от того, что было передано в данном канале (это норма), но и от того, что передавалось в других каналах (иначе говорят, что отсутствует полная развязка между каналами). Если канальные сигналы во всех звеньях сквозного тракта передачи остаются ортогональными, это, по определению, гарантирует отсутствие междуканальных помех. Поэтому *общей причиной возникновения междуканальных помех являются такие искажения канальных сигналов в процессе их передачи, которые приводят к нарушению их ортогональности.*

В разд. 6.2–6.4 более подробно рассмотрены особенности каждого из трех методов передачи.

6.2 Многостанционный доступ с частотным разделением каналов

Метод частотного разделения каналов для передачи сигналов с использованием общей среды основан на традиционных методах радиотехники и начал широко применяться в начале 20 века после изобретения удобного прибора для усиления сигналов – электронной лампы.

Как уже было сказано, поднесущие – это гармонические колебания разных частот. Поэтому при передаче непрерывных сигналов можно использовать АМ, АМ ОБП, ЧМ (разд. 2.10), а для передачи цифровых сигналов – методы, описанные в разд. 2.8. После каждого канального модулятора, как правило, включают полосовой канальный фильтр, идентичный канальному фильтру в приемнике, служащий для устранения нежелательных продуктов модуляции, частоты которых оказались вне полосы, отведенной для данного канала.

Отметим две особенности, характерные для систем, использующих АМ ОБП. Во-первых, сам процесс модуляции обычно осуществляют в два приема: сначала формируют сигнал с обычной двухполосной АМ, а затем при помощи канального полосового фильтра в спектре этого сигнала оставляют лишь одну (чаще нижнюю) боковую полосу, подавляя несущую и вторую боковую полосу. Во-вторых, в приемнике в качестве демодулятора следует использовать фазовый детектор, на второй вход которого нужно подавать гармоническое колебание с частотой несущей. Это колебание вырабатывается местным генератором, поэтому его частота может несколько отличаться от частоты несущей данного канала в передатчике. В итоге при демодуляции возникает характерное искажение сообщения: все составляющие в его спектре вместе сдвигаются вверх или вниз, причем величина и направление сдвига равны соответственно величине и знаку разности несущих частот. Поэтому при многоканальной передаче телефонных сигналов существует норма: отличие несущих частот не должно превышать 10 Гц.

Отметим две главные причины возникновения междуканальных помех.

1) Неидеальность (непрямоугольность) частотной характеристики канального полосового фильтра, например, в приемнике либо смещение его центральной частоты, превышающее ширину защитного частотного интервала. Тогда на вход демодулятора наряду с сигналом данного канала будут поступать некоторые частотные составляющие, принадлежащие одному или двум соседним канальным сигналам. Влияние более далеких каналов, конечно, будет намного слабее, поэтому такая помеха называется *помехой по соседнему каналу*. Для уменьшения уровня такой помехи можно, во-первых, использовать более совершенные фильтры, что заметно повышает их стоимость и массогабаритные характеристики, и, во-вторых, увеличивать защитные частотные интервалы, что, очевидно, приведет к увеличению полосы частот, необходимой для передачи группового сигнала.

2) Нелинейность амплитудной характеристики группового тракта. Как известно, нелинейные искажения даже при передаче колебания одной частоты приводят к появлению дополнительных колебаний кратных частот (высших гармоник). Групповой сигнал, как видно на рис. 6.4, это сумма колебаний разных частот, поэтому нелинейные искажения приводят к появлению множества новых колебаний с различными суммарными и разностными ча-

стотами (комбинационных составляющих), которых не было в спектре передаваемого сигнала. Ясно, что при большом числе каналов в полосу пропускания канального полосового фильтра в приемнике будет попадать большое количество таких комбинационных составляющих, образованных из сигналов других каналов. После демодуляции они будут восприниматься как дополнительный шум. Такая помеха называется *комбинационной помехой.*

Назовем два способа уменьшения комбинационных помех.

1) Уменьшение амплитуд канальных сигналов и, в итоге, амплитуды группового сигнала, чтобы работать в пределах сравнительно линейного участка амплитудной характеристики группового тракта. Платой за это является уменьшение отношения сигнал/шум в каждом канале (полезный сигнал на входе приемника уменьшили, а шум остался прежним). Кроме того, работа усилителя мощности в линейном режиме обычно приводит к снижению его КПД, что крайне нежелательно, например, в спутниковых ретрансляторах, где мощность источника питания (солнечной батареи) ограничена.

2) Выбор частот поднесущих таким образом, чтобы наиболее сильные комбинационные составляющие оказались за пределами полосы, используемой для передачи группового сигнала. Это сделать тем проще, чем меньше число каналов. Поэтому при большом числе каналов выгодно использовать многоступенчатую, древовидную, иерархическую структуру устройства уплотнения каналов, чтобы на каждом уровне блок уплотнения имел малое число входов.

Эти принципы закреплены в действующем стандарте, определяющем *структуру и параметры устройства частотного уплотнения аналоговых телефонных сигналов* для их передачи по медному кабелю (Приложение 4). Предусмотрены четыре ступени уплотнения, при этом во всех ступенях применяется амплитудная модуляция с нижней боковой полосой.

На 12 входов первой ступени уплотнения поступают 12 стандартных аналоговых телефонных сигналов с полосой 300-3400 Гц каждый. Блок содержит 12 генераторов поднесущих с частотами 64, 68, 72, … ,104, 108 кГц (отсюда видно, что на каждый канал отводится полоса 4 кГц), 12 канальных модуляторов для формирования АМ канальных сигналов, разнесенных по оси частот, и 12 канальных полосовых фильтров для преобразования АМ в АМ ОБП. В итоге (см. рис. 2.17) на выходе блока первый канальный сигнал занимает полосу 60,4-63,7 кГц, второй канальный сигнал – полосу 64,4-67,7 кГц и т.д. (рис. 6.5). Между канальными сигналами предусмотрены защитные интервалы шириной по 0,9 кГц каждый. Вместе 12 канальных сигналов образуют *первичный групповой сигнал*, занимающий полосу 60,4-107,7 кГц. Обычно указывают округленные значения 60 – 108 кГц, то есть ширина спектра первичного группового сигнала равна 4·12 = 48 кГц.

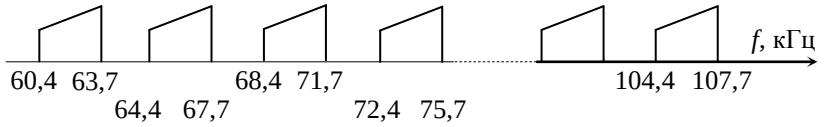

Рис. 6.5. Спектр первичного группового сигнала

Три последующие ступени организованы по тому же принципу и имеют по 5 входов каждая. В частности, на входы второй ступени уплотнения поступают 5 первичных групповых сигналов, занимающих полосу 60–108 кГц каждый, а на выходе получается *вторичный групповой сигнал*, спектр которого лежит в полосе 312–352 кГц (60 телефонных каналов на 240 кГц).

6.3 Многостанционный доступ с временным разделением каналов

Метод временного разделения каналов основан на использовании большого количества разнообразных импульсных устройств, поэтому он начал широко применяться в 60-х годах 20 века, когда появилась удобная элементная база – транзисторы и транзисторные микросхемы.

Существуют два основных режима временного мультиплексирования сигналов.

Режим синхронной (циклической) передачи основан на поочередном, строго периодическом подключении источников сообщений к линии связи всегда в одной и той же последовательности. Благодаря этому принадлежность каждого элемента группового сигнала конкретному источнику определяется его временным положением (номером канального интервала) внутри кадра (цикла), поэтому достаточно лишь изредка, например, в конце каждого цикла передавать специальные синхросигналы, позволяющие приемнику работать синхронно с передатчиком (рис. 6.6). Можно использовать аналоговые либо цифровые методы передачи.

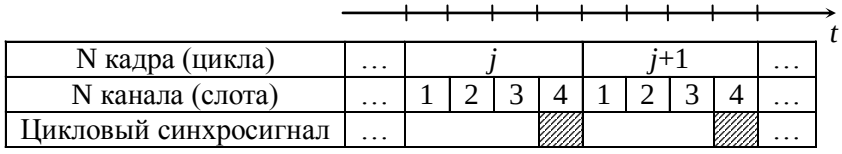

Рис. 6.6. Временная диаграмма передачи в синхронной трехканальной СПИ с временным разделением каналов (циклическая передача)

Режим асинхронной передачи применяется только в цифровых СПИ. При подключении источника к линии передается пакет, то есть небольшая часть сообщения, например файла, содержащая обычно от сотен до нескольких тысяч бит. Порядок подключения источников к линии – произвольный, по мере возникновения необходимости, поэтому циклы, как таковые, здесь отсутствуют (рис. 6.7). В начальной части каждого пакета имеется заголовок,

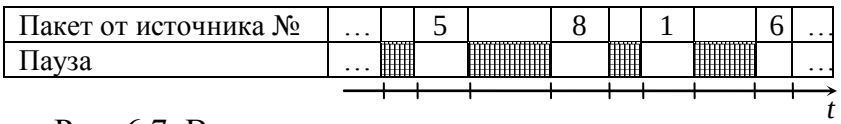

Рис. 6.7. Временная диаграмма передачи пакетов в асинхронной СПИ с временным разделением каналов

где, кроме прочего, обязательно указан номер получателя. Особенности и преимущества асинхронного метода передачи мы обсудим позже, в разд. 7.2.

В **синхронной СПИ** в течение цикла источник подключается к линии один раз, следовательно, количество подключений в секунду равно частоте передачи циклов. В аналоговой СПИ, ориентированной на передачу телефонных сигналов, в каждом слоте передается один отсчет этого сигнала, поэтому частота следования циклов должна составлять 8000 циклов в секунду, и длительность цикла равна 125 мкс. Отсчеты можно передавать с использованием любого из методов модуляции импульсной несущей – АИМ, ШИМ или ВИМ (разд. 2.5).

В цифровой синхронной СПИ длительность кадра задается также 125 мкс из тех же соображений, тогда в слоте тоже передается по одному отсчету в цифровой форме от каждого источника, то есть по 8 бит (см. разд. 2.6). При этом скорость передачи в групповом сигнале $V = 8000 \cdot 8 \cdot n$ бит/с, где n – фактическое число каналов (слотов в цикле), включая и служебный канал, предназначенный для передачи циклового синхросигнала.

Передача каждого бита осуществляется при помощи обычных импульсов прямоугольной формы (1 – импульс, 0 – пауза), но чаще применяется один из методов модуляции гармонической несущей цифровым сигналом (разд. 2.8). Определенными достоинствами обладает также метод фазовой манипуляции при использовании несущей в виде периодической последовательности прямоугольных импульсов при их скважности, равной двум (так называемый меандр). Если период несущей равен длительности бита, то после модулятора ему будет соответствовать пара импульсов противоположной полярности – манчестерский код.

Теоретически спектр прямоугольных импульсов бесконечно широк, поэтому при передаче приходится его ограничивать (разд. 2.8). Это вызывает искажения формы принимаемых импульсов, в частности растягивание их фронтов, что, во-первых, приводит к *межсимвольной интерференции* (взаимное искажение соседних импульсов), во-вторых, к возникновению *помех по соседнему каналу* (взаимное искажение соседних импульсов, принадлежащих разным каналам, то есть стоящих на краях слотов). Те же неприятности возникают и при передаче группового сигнала в канале с мультипликативной помехой в виде частотных селективных замираний.

Параметры цифровых СПИ с временным разделением каналов также стандартизованы. Первый из наиболее важных стандартов носит название **плезиохронной цифровой иерархии** (ПЦИ) – Plesyochronous Digital Hierarchy (PDH) – и определяет способ мультиплексирования стандартных цифровых телефонных сигналов для их передачи по медному кабелю (Приложение 5). Напомним (разд. 2.6), что стандартный цифровой телефонный сигнал (DSO) получается в результате АЦП стандартного аналогового телефонного сигнала и характеризуется скоростью 64 кбит/с.

Первая ступень уплотнения PDH (аппаратура ИКМ-30) рассчитана на организацию 32 каналов методом временного побайтного мультиплексирования, в итоге на выходе формируется цифровой поток E1 со скоростью 64·32 = 2048 кбит/с. Из них 2 канала являются служебными (передача сигналов цикловой синхронизации, передача команд для управления коммутирующими приборами на АТС и др.), а остальные 30 – пользовательскими.

Каждая из ступеней более высокого уровня (аппаратура ИКМ-120, ИКМ-480, ИКМ-1920) рассчитана на уплотнение четырех цифровых потоков низшего уровня методом временного побитного мультиплексирования, а пятый уровень пока не имеет аппаратурной реализации (и, по-видимому, не будет иметь из-за конкуренции с методами передачи по оптическому кабелю).

Дополнительные функции, выполняемые аппаратурой этих четырех уровней, обусловлены следующим обстоятельством. Плезиохронная цифровая иерархия создавалась для работы в почти синхронных (плезиохронных) сетях связи, когда скорости однотипных цифровых потоков E1, E2 или E3 на входах каждой ступени несколько отличаются друг от друга. Поэтому при мультиплексировании четырех входных потоков возникает проблема выравнивания их скоростей. Она решается путем вставки “пустых” битов, если скорость данного входного потока оказалась ниже скорости мультиплексирования в этой ступени, либо путем передачи “лишних” битов по служебным каналам в противном случае. Такой сложный способ передачи потока по нескольким путям существенно затрудняет его извлечение из группового сигнала в пункте приема.

Другой важный стандарт носит название **синхронной цифровой иерархии** (СЦИ) – Synchronous Digital Hierarchy (SDH) – и определяет способ мультиплексирования цифровых потоков для их передачи по оптоволоконному кабелю (Приложение 6). Аппаратура SDH не предназначена для мультиплексирования мелких цифровых потоков, поступающих от пользователей (для этого используется аппаратура PDH, например ИКМ-30), она применяется для транспортировки более крупных потоков между узлами сети. Синхронная цифровая иерархия создавалась для работы в синхронных сетях связи, где существует синхронизация всех узлов сети из единого центра. Благодаря этому скорости передачи всех однотипных потоков строго одинаковы.

Передача цифровых сигналов на каждом уровне мультиплексирования осуществляется блоками длительностью 125 мкс. Такой блок называется *синхронным транспортным модулем уровня N* (Synchronous Transport Module of level N) или, коротко, STM-N. Например, STM-1 содержит 2430 байт (то есть 19440 бит), из которых 2340 байт – это полезная нагрузка, а остальной объем (90 байт) предназначен для передачи служебной информации.

Полезная нагрузка STM-1 не заполняется путем циклического опроса входов, здесь принят иной, более гибкий и универсальный принцип, пригод-

ный как для синхронного, так и для асинхронного режимов передачи. Она заполняется одним или несколькими “виртуальными контейнерами”, то есть порциями входных потоков. Каждый контейнер в своей служебной части содержит числовой адрес того узла сети, в котором его следует “выгрузить”. Стандартизованы контейнеры пяти типов разных объемов, при этом мелкие контейнеры могут помещаться в более крупные. Такая организация транспортировки позволяет в любом узле сети быстро и просто находить нужные контейнеры без распаковки всего содержимого STM-1, а также вставлять другие контейнеры на свободные места.

Синхронный транспортный модуль уровня 4 заполняется один за другим четырьмя STM-1. Такая же операция проделывается и с модулями более высоких уровней.

6.4 Многостанционный доступ с кодовым разделением каналов

Этот метод применяется только в двоичных цифровых СПИ с ФМ. Начнем с обычной одноканальной СПИ с ФМ, где для передачи каждого бита от источника используется радиоимпульс прямоугольной формы, начальная фаза которого определяется значением этого бита, например, 0° при передаче единицы и 180° при передаче нуля (рис. 2.12). Длительность этого информационного импульса $\tau_{\text{è}} = 1/V$, где V – скорость передачи, бит/с. Ширина спектра такого сигнала примерно равна $\Delta f \approx 2/\tau_{\text{è}} = 2V$. При малой скорости передачи сигнал занимает узкую полосу частот.

Для организации кодового разделения каналов в многоканальной СПИ каждый канальный информационный импульс (рис. 6.8а) подвергается дополнительной внутриимпульсной фазовой модуляции (рис. 6.8б). Для этого импульс разбивается на *B* частей равной длины (на *B* более коротких импульсов-чипов), при этом *число B называется базой сигнала*. Для данного

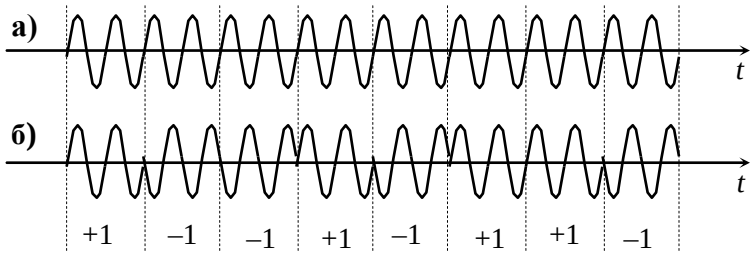

Рис. 6.8. Преобразование информационного импульса в сигнал с внутриимпульсной ФМ:
а) до преобразования; б) после преобразования.

канала раз и навсегда задана *B*-разрядная двоичная последовательность, определяющая закон внутриимпульсной фазовой модуляции, причем элемен-

ты этой последовательности удобно обозначить как +1 и –1 вместо традиционных 1 и 0. Далее каждый элемент сигнала умножается на соответствующий элемент последовательности, то есть остается без изменений, если множитель равен +1 и инвертируется в случае –1.

В итоге ширина спектра канального сигнала увеличивается в B раз по сравнению с традиционным методом передачи без внутриимпульсной модуляции, так как теперь сигнал фактически состоит из более коротких импульсов. Кстати, метод внутриимпульсной ФМ является самым распространенным методом искусственного расширения спектра сигнала и часто применяется в других областях радиоэлектроники, например в радиолокации.

Все канальные сигналы передаются одновременно в одной и той же широкой полосе частот. Каждому канальному сигналу присваивается своя двоичная последовательность, определяющая закон внутриимпульсной ФМ, так, чтобы сигналы всех каналов были взаимно ортогональны. Очевидно, что для проверки ортогональности даже не нужно вычислять значение интеграла (6.3). Два канальных ФМ сигнала ортогональны, если в двоичных последовательностях, определяющих законы их внутриимпульсной ФМ, число совпадающих элементов равно числу несовпадений. Например, если одному из каналов присвоена последовательность, показанная на рис. 6.8, то другим каналам можно присвоить, допустим, такие последовательности: +1, –1, +1, –1, –1, +1, –1, +1; +1, –1, –1, +1, +1, –1, –1, +1, …

В системах МДКР операции выделения нужного канального сигнала из суммарного, группового сигнала и демодуляции можно совместить. Для этого можно использовать корреляционный приемник (или заменяющий его согласованный фильтр) и в качестве опорного сигнала (2.35) использовать целиком импульс с внутриимпульсной ФМ. Сравнение с порогом величины v, выданной корреляционным приемником, позволит вынести решение относительно значения переданного двоичного символа. Таким образом, в системах МДКР *устройство разделения каналов и канальные демодуляторы – это набор, состоящий из* n *корреляционных приемников.*

Максимальное число ортогональных последовательностей равно B, следовательно, именно таково количество каналов, работающих без взаимных помех.

Для $B = 2^m$, где m – целое положительное число, существует простой способ определения системы ортогональных двоичных последовательностей в виде строк матрицы Адамара. Матрица Адамара второго порядка имеет вид

$$\mathbf{H}_2 = \begin{Vmatrix} +1 & +1 \\ +1 & -1 \end{Vmatrix}. \qquad (6.4)$$

Матрица порядка $2p$ состоит из блоков – матриц порядка p

$$\mathbf{H}_{2p} = \begin{Vmatrix} \mathbf{H}_p & \mathbf{H}_p \\ \mathbf{H}_p & -\mathbf{H}_p \end{Vmatrix}. \qquad (6.5)$$

Например, $\mathbf{H}_4 = \begin{Vmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{Vmatrix}$.

Главный недостаток такой системы канальных сигналов состоит в том, что не все они имеют широкий спектр. Например, сигнал, задаваемый первой строкой матрицы Адамара, фактически не содержит внутриимпульсной модуляции, поэтому его спектр не расширяется. Из-за этого для практического применения подбирают последовательности, обладающие лучшими спектральными свойствами. Хорошие результаты дает использование отрезков *псевдослучайной двоичной последовательности* (ПСП) (см. разд. 2.7).

Причиной **междуканальных помех** в системе МДКР могут служить любые искажения (линейные и нелинейные) группового сигнала, приводящие к нарушению взаимной ортогональности канальных сигналов. Во-первых, это ограничение полосы пропускания группового тракта и мультипликативная помеха в канале передачи, во-вторых, нелинейные искажения в оконечном каскаде передатчика.

Иногда возможности для возникновения междуканальных помех создаются уже на этапе проектирования СПИ. Например, использование отрезков ПСП для внутриимпульсной ФМ. Как уже отмечалось, они не вполне ортогональны. Другая причина заключается в том, что ради упрощения СПИ не предусматривается взаимная синхронизация каналов, то есть модулирующие последовательности в разных каналах произвольно сдвинуты во времени, что нарушает их ортогональность.

Метод МДКР используется пока лишь в гражданских и военных системах радиосвязи с подвижными объектами. Следует упомянуть стандарт IS-95, в соответствии с которым построены сотовые системы CDMA и спутниковая система связи с подвижными объектами Globalstar (B=64, ширина спектра группового и каждого из канальных сигналов равна 1,25 МГц).

6.5 Синхронизация в СПИ с многостанционным доступом

Проблема синхронизации не возникает лишь в аналоговых СПИ с ЧРК. *Все элементы, предназначенные для преобразования импульсных сигналов, в остальных СПИ работают под управлением тактовых импульсов, задающих ритм их работы*. В качестве исключения можно назвать ФНЧ-демодулятор в приемнике, преобразующий последовательность импульсов-отсчетов в непрерывный сигнал.

Работой передающей части **многоканальной СПИ в синхронном режиме передачи** управляет достаточно стабильный первичный генератор тактовых импульсов. Период следования этих импульсов равен длительности самого короткого элемента группового сигнала. В цифровой СПИ с ВРК это

один бит, в СПИ с КРК это один из *B* элементов информационного бита. Например, в аппаратуре ИКМ-30 тактовые импульсы имеют частоту 2048 кГц.

Тактовые импульсы более высоких порядков формируются путем деления частоты первичного генератора. Например, в той же аппаратуре ИКМ-30 каждый канал подключается на время передачи 8 бит, поэтому канальные импульсы имеют частоту 256 кГц; цикл состоит из 32 канальных интервалов, поэтому начало каждого цикла задают цикловые импульсы частоты 8 кГц, формируемые из канальных импульсов делением их частоты на 32; различаются четные и нечетные циклы. Далее в ИКМ-30 частота цикловых импульсов делится на 16 для формирования сверхциклов. Для управления работой АЦП на каждом из 30 входов нужны тактовые импульсы двух видов: с частотой 64 кГц для формирования очередного бита и с частотой 8 кГц для запуска цикла преобразования следующего отсчета. Если в СПИ применяется блочный корректирующий код, то нужен еще один делитель частоты для формирования импульсов, задающих момент начала очередной кодовой комбинации, и т.д.

Точно такой же комплект генераторного оборудования используется для управления работой всех устройств в приемной части СПИ. *Система синхронизации приемного устройства* обеспечивает условия, при которых тактовые импульсы всех видов в приемнике формируются одновременно с соответствующими тактовыми импульсами в передатчике (точнее, с одной и той же задержкой, равной текущему времени распространения сигнала от передатчика к приемнику).

Чтобы система синхронизации могла выполнять свои функции, кроме сигналов, несущих информацию о передаваемых сообщениях, необходимо еще передавать специальные *синхросигналы* разных видов. Поэтому в передающей части СПИ имеются *формирователи синхросигналов* разных видов, запускаемые соответствующими тактовыми импульсами от генераторного оборудования. В приемной части СПИ имеются *селекторы синхросигналов*, которые постоянно сравнивают временное положение принимаемых информационных и синхронизирующих сигналов, с одной стороны, и тактовых импульсов от местного генераторного оборудования, с другой стороны, и по мере необходимости подают команды на коррекцию последних.

Специальные синхросигналы в момент начала очередного мельчайшего элемента группового сигнала, например бита, не передают. Для работы селектора тактовых импульсов в приемнике достаточно тех изменений в сигнале, которые происходят при переходе от 1 к 0 или от 0 к 1 (при ФМ это скачки фазы на 180°). Правда, иногда в сигнале встречаются длинные серии нулей или единиц, тогда эти скачки отсутствуют, и работа селектора затрудняется. Чтобы облегчить работу селектора тактовых импульсов, каждый бит передают не одним, а двумя импульсами разных полярностей (манчестер-

ским кодом), вводят специально паузы между символами и т.п. К сожалению, все эти меры несколько расширяют спектр передаваемого сигнала, поэтому применяются далеко не всегда.

Видимых границ между кодовыми комбинациями, принадлежащими соседним каналам, в групповом сигнале нет. Тем не менее, канальные синхросигналы, как правило, не передают. Достаточно того, что передается кадровый синхросигнал в начале каждого кадра (сигналы точного времени ведь не передают ежесекундно или ежеминутно, достаточно передавать раз в час). Для передачи кадрового синхросигнала выделяется отдельный канал, и в нем в цифровой СПИ передается специальная кодовая комбинация той же длины, что и в любом другом канале. Например, в аппаратуре ИКМ-30 в начале каждого четного цикла на 2-8 битовых позициях передается синхросигнал вида 0011011. Единственное отличие циклового синхросигнала от комбинаций в информационных каналах состоит в том, что он всегда один и тот же, а комбинации в информационных каналах изменяются.

Исходя из принципа формирования тактовых импульсов разных уровней (деление частоты тактовых импульсов первого уровня), легко догадаться, что *селектор тактовых импульсов первого уровня должен быть аналоговым, а селекторы более высоких уровней – это цифровые устройства*. Правда, в последнее время и для выделения тактовых импульсов нередко применяется цифровой селектор, но, в сущности, в нем выполняются те же операции в цифровой форме.

В качестве первичного генератора в приемнике применяется генератор тактовых импульсов, управляемый напряжением (ГУН), например генератор импульсов частоты 2048 кГц в ИКМ-30, в котором осуществляется *фазовая автоподстройка частоты* по схеме рис. 5.5. На вход фазового детектора-дискриминатора подается принимаемый групповой сигнал после детектора (амплитудного, фазового либо частотного в зависимости от вида модуляции). Нужно отметить лишь одно обстоятельство: если символы передаются импульсами длительности τ без пауз между ними, то в спектре такого сигнала отсутствует составляющая с частотой повторения символов $F_c = 1/\tau$ (эта составляющая совпадает с частотой, на которой спектральная плотность равна нулю), то есть отсутствует информация для работы дискриминатора. Поэтому продетектированный сигнал, перед тем как подать его на вход дискриминатора, подвергают дополнительной обработке. Например, этот сигнал можно пропустить через дифференцирующую RC-цепь и затем инвертировать импульсы отрицательной полярности. В итоге на вход дискриминатора будут подаваться короткие импульсы в те моменты, когда символ изменяется на противоположный.

В любом аналоговом устройстве автоматического регулирования имеются погрешности трех видов: систематическая, то есть постоянное смещение временного положения тактовых импульсов, его медленные случайные

вариации (вандер) и быстрые случайные флуктуации их временного положения (джиттер).

Главная причина появления постоянной *систематической ошибки* – неточная первоначальная настройка элементов, например, отличие собственной частоты ГУН (при нулевом управляющем напряжении) от частоты первичного генератора в передатчике, либо изменение настройки в процессе старения элементов. Наличие ФАПЧ уменьшает, но не сводит эту погрешность к нулю.

Вандер обусловлен динамическими ошибками следящей системы при медленных изменениях частот генераторов либо задержки в канале передачи.

Джиттер вызывается наличием быстрых флуктуаций частот обоих генераторов, влиянием шума и замираний принимаемого сигнала.

Поскольку тактовые импульсы управляют работой демодулятора, наличие погрешностей всех видов приводит к увеличению вероятности ошибки при демодуляции. На селекторы синхросигналов верхних уровней поступает сигнал с выхода демодулятора, поэтому эти же погрешности, в конечном итоге, увеличивают и вероятность сбоя синхронизации на этих уровнях.

Любой *цифровой селектор синхросигналов* в конце каждого периода (этот период равен периоду повторении синхросигнала данного уровня) выносит одно из двух решений: 1) синхронизация имеется; 2) произошел сбой синхронизации.

Чтобы вынести решение, селектор анализирует принимаемый сигнал в течение n предыдущих периодов и подсчитывает количество m случаев, когда на позиции, указанной соответствующим тактовым импульсом от местного генераторного оборудования, форма принимаемого сигнала соответствует ожидаемому синхросигналу. Эта операция называется накоплением импульсов. Если $m \geq m_{\text{п}}$, где $m_{\text{п}}$ – заранее установленное пороговое значение, то принимается первое из двух возможных решений, в противном случае – второе.

Например, в ИКМ-30 можно каждый раз анализировать n=5 подряд идущих четных циклов и установить $m_{\text{п}} = 4$. Генератор четных цикловых импульсов в каждом из этих пяти циклов выделяет один слот из тридцати двух. Если в четырех или пяти выделенных слотах оказалась комбинация 0011011, то следует принять решение о наличии синхронизации.

В работе цифрового селектора синхросигналов возможны *ошибки двух видов*. Допустим, генераторное оборудование сфазировано правильно, но селектор принял решение о сбое синхронизации. Назовем эту ситуацию ошибкой первого рода. Если же при неверной фазировке генераторного оборудования селектор принял решение о наличии синхронизации, то это – ошибка второго рода.

Предположим, что в приведенном выше примере для ИКМ-30 ошибки

на выходе демодулятора отсутствуют. Тогда при $n = 5$ следует принять $m_п = 5$, и возможно лишь возникновение ошибки второго рода в ситуации, когда во всех пяти помеченных пользовательских каналах появится комбинация 0011011. Так как в пользовательских каналах имеется последовательность практически независимых равновероятных символов, то вероятность такого события $p_2 = 0,5^{35} \approx 2,9 \cdot 10^{-11}$ довольно мала.

В реальной ситуации ошибки на выходе демодулятора встречаются, поэтому в работе цифрового селектора синхросигналов возможны ошибки обоих видов, и следует подобрать такое значение порога $m_п$, при котором минимальна полная вероятность ошибки.

Если селектор принял решение о сбое синхронизации, то он переходит в режим поиска, то есть проверяет подряд все слоты до тех пор, пока не встретится комбинация 0011011. После этого переходит в нормальный режим, то есть режим слежения, при этом не исключено, что через n периодов он снова повторит поиск.

Естественно, что *требования к помехоустойчивости селектора синхросигналов существенно выше, чем требования к помехоустойчивости других решающих устройств*. Например, ошибка в одном бите на выходе демодулятора исказит лишь одну кодовую комбинацию, а сбой синхронизации на некоторое время нарушит работу всей многоканальной СПИ. Уменьшить вероятность нарушения синхронизации можно, увеличив значение параметра n, но при этом возрастает время восстановления синхронизации.

Решение проблемы синхронизации в **системах радиосвязи с многостанционным доступом** имеет ряд особенностей. Если в многоканальной СПИ формирование цифрового сигнала, кодирование, модуляция и уплотнение каналов осуществляются *в пункте передачи* под управлением общего генераторного оборудования, то в системах радиосвязи *передатчик, устройство уплотнения* канальных сигналов (базовая станция в системах сотовой связи, ретранслятор на ИСЗ в системах спутниковой связи и т. п.) и *приемник находятся в разных пунктах*. Здесь приходится налаживать совместную синхронную работу не двух, а *трех* объектов в условиях, когда расстояния между ними меняются во времени (следовательно, меняются и задержки сигнала).

В таких системах, как правило, осуществляется двусторонняя передача информации, то есть *на обоих участках имеются два равноценных канала*: в *прямом и обратном направлениях*. Здесь *ведущая роль принадлежит ретранслятору*, поэтому в нем находится генераторное оборудование, задающее временной график работы всей системы, и формирователи синхросигналов. Приемники всех абонентов синхронизируют свое генераторное оборудование по сигналам, передаваемым ретранслятором, и это же оборудование управляет работой своего передатчика.

Остается лишь один вопрос: с каким опережением должен работать пе-

редатчик, чтобы его сигнал достиг ретранслятора к тому моменту, когда начнется нужный временной слот? Для оценки текущего времени запаздывания на пути от станции до ретранслятора каждая станция периодически принимает свой сигнал, возвратившийся от ретранслятора, и измеряет его запаздывание относительно переданного сигнала.

Работа СПИ **в асинхронном режиме передачи** (передача пакетов по мере возникновения требований) имеет ряд особенностей. *Все современные протоколы требуют, чтобы пакет содержал целое число байтов*. Поэтому на первом уровне (биты) или даже на первых двух уровнях (биты и байты) способ синхронизации не отличается от того, что характерно для СПИ в синхронном режиме передачи.

Опишем два типичных способа обеспечения синхронизации на третьем уровне, где система синхронизации должна указать приемнику байты, соответствующие началу и концу пакета.

1) В начале и в конце каждого пакета передается *флаг*, то есть специальная восьмибитовая комбинация, допустим, 01111110. Такая комбинация не может встретиться нигде больше внутри пакета. Естественно, что последовательность символов от источника информации может не удовлетворять этому требованию, поэтому она подвергается перекодировке: после серии из пяти единиц вставляется дополнительный нуль. Например, последовательность 0101111110011111110 преобразуется в 01011111000111110110. Селектор синхросигнала проверяет все восьмибитовые последовательности и первую обнаруженную комбинацию 01111110 идентифицирует как начало пакета, а вторую – как его конец.

2) Номер байта, соответствующий началу пакета, и количество байтов в пакете передаются по отдельным, служебным каналам.

Как видно на рис. 6.7, существуют свободные интервалы, не занятые для передачи пакетов. В действительности в это время передается “пустая” последовательность символов для того, чтобы обеспечить устойчивую работу селекторов синхросигналов в приемной части СПИ.

6.6 Расширение спектра сигнала

Завершая обсуждение свойств сигналов, используемых в СПИ с кодовым разделением каналов, рассмотрим основные способы **расширения спектра сигнала.**

Расширение спектра сигнала – это процедура, в результате которой спектр цифрового сигнала становится шире, чем это минимально необходимо для его передачи, а процесс расширения происходит при участии числовой последовательности, не зависящей от передаваемого сообщения [2]. Применение внутриимпульсной кодовой ФМ – это один из двух основных способов расширения спектра сигнала. Как мы уже увидели (разд. 6.4), здесь оба при-

знака присутствуют.

Другой способ расширения спектра – это скачкообразное *переключение несущей частоты сигнала*.

Для этого заранее устанавливают (точнее, закладывают в программу работы синтезатора несущей) ряд фиксированных частот $f_1, \ldots, f_M$, и переключение (выбор номера очередной частоты) производится через равные промежутки времени τ в соответствии с очередным элементом псевдослучайной управляющей М-ичной числовой последовательности.

Возможен режим *быстрого переключения* ($\tau \ll \tau_{\text{è}}$), то есть несущая частота переключается B раз за время передачи одного информационного символа. Свойства таких сигналов весьма близки к свойствам сигналов с внутриимпульсной кодовой ФМ, но создание столь быстродействующего синтезатора частот – сложная задача.

В более простом режиме *медленного переключения* частоты ($\tau \gg \tau_{\text{è}}$) в интервале τ передается серия информационных импульсов на неизменной несущей частоте, и в итоге в пределах этого интервала получившийся сигнал ничем не отличается от обычного узкополосного сигнала.

Каковы преимущества сигналов с расширенным спектром?

Если база сигнала велика, при передаче в радиоканале спектр сигнала распределяется по широкой полосе частот, так что на любом участке отношение средних мощностей сигнала и шума может быть даже намного меньше единицы. В таких условиях обеспечивается *скрытность* передачи, поскольку уверенно обнаружить сам факт наличия полезного сигнала уже не удается. Очевидно, что скрыть одновременную работу большого количества источников таким образом уже труднее. Сигнал с широкой полосой меньше подвержен воздействию узкополосной помехи, поражающей лишь часть его спектра.

Сигнал с расширенным спектром обладает и некоторой криптоустойчивостью, поскольку для его нормального приема (сжатия) недостаточно знать несущую частоту и обеспечить синхронизацию приемника, нужно еще знать числовую последовательность, управляющую процессом расширения спектра. С этим же связана и повышенная устойчивость таких сигналов к воздействию широкополосной помехи. Такая помеха хорошо воздействует на корреляционный приемник лишь в случае, когда она похожа на полезный сигнал, но для создания такой помехи также нужно знать расширяющую числовую последовательность.

Иногда можно встретить утверждение, что сигналы с расширенным спектром способны обеспечить меньшую величину вероятности ошибки при демодуляции, чем узкополосные сигналы. Это утверждение ошибочно, по крайней мере, по отношению к СПИ, использующей канал с постоянными параметрами, поскольку в нем BER зависит лишь от отношения энергии информационного импульса к спектральной плотности шума, а при расширении

спектра эта энергия не увеличивается.

Зато *в канале с мультипликативной помехой* преимущества сигналов с расширенным спектром (точнее, возможности, которыми еще нужно уметь воспользоваться) не подвергаются сомнению.

Действительно, спектр узкополосного сигнала может целиком оказаться внутри частотного интервала, на котором коэффициент передачи линии очень мал, как показано на рис. 6.9, что приведет к появлению пакета ошибок, а то и к прерыванию связи. Конечно, через некоторое время форма этой

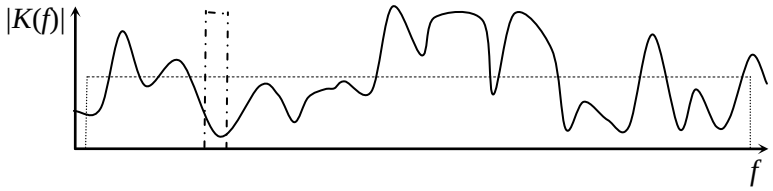

Рис. 6.9. Примеры “мгновенной” амплитудно-частотной характеристики канала с частотно-селективными замираниями (сплошная линия), спектров узкополосного (штрих-пунктир) и широкополосного (штрих) сигналов

случайной АЧХ изменится и, возможно, уровень сигнала возрастет, но часть сообщения уже повреждена.

Суммарная мощность широкополосного сигнала на выходе линии при любых изменениях формы ее АЧХ тоже меняется, но значительно меньше. Из этого пока не следует, что, обрабатывая этот сигнал при помощи корреляционного приемника, обеспечим почти столь же надежную связь, как в канале с постоянными параметрами. Общий принцип уже был сформулирован в разд. 5.5: если есть мультипликативная помеха, то нужно организовать какое-то количество независимых ветвей разнесенных передачи и (или) приема и сформировать сигнал на входе решающего устройства путем комбинирования этих разнесенных сигналов.

На первый взгляд кажется, что путем расширения спектра организуется частотно-разнесенная передача, но при этом, в отличие от традиционного метода частотного разнесения (разд. 5.5), во-первых, на каждой частоте передается не целый сигнал (информационный бит), а только его часть, и, во-вторых, здесь невозможно спектр принимаемого сигнала разделить на конечное число независимо флуктуирующих составляющих. Поэтому, не вдаваясь в тонкости классификации, рассмотрим применяемые методы комбинирования частей такого сигнала.

Выделение составляющих сигнала, их раздельная обработка и комбинирование возможны в частотной и временной областях.

Обработка в частотной области сводится к использованию адаптивного эквалайзера, то есть линейного фильтра с изменяемой частотной характеристикой. Такая возможность обусловлена тем, что частотная характеристика канала изменяется достаточно медленно, то есть, она почти не изменя-

ется за время, в течение которого можно принять десятки и сотни информационных импульсов. По этим импульсам приемник оценивает форму их спектра $S_{\text{вых}}(f)$. Форма спектра сигнала на входе канала $S_{\text{âõ}}(f)$ известна, поэтому оценивается форма частотной характеристики канала $K(f) \approx S_{\text{вых}}(f)/S_{\text{вх}}(f)$. Частотная характеристика эквалайзера вычисляется очевидным образом $K_{\text{ý}}(f) = K(f)^{-1}$, в итоге частотная характеристика сквозного тракта становится почти равномерной и, следовательно, уменьшаются искажения формы принимаемых импульсов.

Для некоторых видов модуляции сигналов (например, КАМ либо разновидности ФМ) утверждение о том, что в пункте приема известна форма спектра сигнала, не соответствует действительности, так как из-за случайности информационных импульсов известна лишь форма спектра плотности мощности сигнала. Поэтому периодически приходится передавать обучающую последовательность импульсов, когда вместо случайной серии информационных символов на вход модулятора подается известная серия битов. Такой цикл адаптации эквалайзера повторяется десятки и даже сотни раз в секунду.

Как видно из приведенного описания, адаптивный эквалайзер можно применять не только для приема сигналов с искусственно расширенным спектром, но и в любой СПИ, использующей для передачи достаточно короткие импульсы, как это делается, например, в сотовой системе стандарта GSM.

Обработка во временной области предполагает выполнение следующих операций.

1) Принимаемый сигнал, состоящий из серии информационных импульсов длительностью $\tau_{\text{è}}$ с внутриимпульсной ФМ (рис. 6.8б), подают на вход фильтра, согласованного с таким импульсом. Согласованный фильтр сжимает каждый импульс в *B* раз, так что каждая квадратурная составляющая сигнала имеет вид короткого импульса с множеством малых боковых лепестков (рис. 2.12).

Поскольку при наличии многолучевости в пункте приема имеется большое количество сигналов, отраженных от разных объектов и поэтому имеющих случайные амплитуды и случайные задержки, на выходе согласованного фильтра присутствует сумма таких независимых откликов, и в итоге имеем сигнал, рассеянный, расширенный во времени.

2) С шагом, равным длительности сжатого импульса, измеряют значения каждой квадратурной составляющей. Таким образом из суммарного сигнала выделяют *n* независимых слагаемых, каждое из которых содержит передаваемое сообщение. Далее эти разнесенные сигналы можно комбинировать одним из известных методов (разд. 5.5).

Rake (грабли) – это название, которое дали первому приемнику, в котором был реализован этот способ разнесенных передачи-приема. В настоящее

время он все шире применяется в широкополосных системах передачи, например, в сотовых системах связи с кодовым разделением каналов.

Если расширение спектра производится путем медленного переключения частоты, возможности борьбы с мультипликативной помехой существенно ниже. Единственное, что дает этот метод - это ограничение времени, в течение которого система будет непрерывно работать в неблагоприятных условиях. Ясно, что в большинстве случаев этот интервал не превышает периода переключения частоты τ.

В заключение следует подчеркнуть, что положительные качества сигналов с расширенным спектром достигаются за счет существенного снижения частотной эффективности СПИ. Поэтому такие сигналы, как правило, используются лишь в условиях многостанционного доступа, когда эта широкая полоса совместно используется многими абонентами.

Вопросы для самоконтроля по главе 6

1. При каком необходимом условии возможно линейное разделение канальных сигналов без взаимных помех в многоканальной СПИ?
2. Поясните принципы частотного, временного и кодового разделения каналов.
3. Почему в многоканальной СПИ канальные сигналы взаимно ортогональны, хотя это не является обязательным условием для их разделения без междуканальных помех.
4. Из каких соображений выбирается длительность цикла передачи в многоканальной СПИ с ВРК?
5. Что используется в качестве устройств уплотнения и разделения канальных сигналов в многоканальной СПИ: с ЧРК? с ВРК? с КРК?
6. Почему возникает необходимость осуществления тактовой и цикловой синхронизации приемника в многоканальной СПИ.
7. Укажите причину возникновения междуканальных помех в многоканальных СПИ: с ЧРК? с ВРК? с КРК?
8. Какие требования предъявляются к расширяющим последовательностям при кодовом разделении канала?
9. Дайте определения синхронного и асинхронного режимов передачи цифровых сигналов.
10. Укажите области применения корреляционного приемника.
11. Укажите основные различия двух стандартов многоканальной передачи с временным мультиплексированием: PDH и SDH.
12. Чем определяется максимальное число канальных сигналов в многоканальной СПИ с КРК?
13. Почему в многоканальных СПИ: с ЧРК требуется высокая линейность амплитудной характеристики группового тракта?

7 ПРИНЦИПЫ ПОСТРОЕНИЯ СЕТЕЙ ЭЛЕКТРОСВЯЗИ

7.1 Определения, классификация, структуры сетей

Канал – это совокупность технических средств, обеспечивающих передачу сообщений определенного вида от оконечного оборудования одного пользователя к оконечному оборудованию другого. Различают каналы: симплексные, предназначенные для передачи только в одном направлении; дуплексные для одновременной передачи в обоих направлениях; полудуплексные, позволяющие передавать в обоих направлениях, но поочередно.

Сеть электросвязи – это совокупность технических средств, обеспечивающая обмен информацией между множеством пользователей, разнесенных в пространстве. Основные составляющие сети представлены на рис. 7.1.

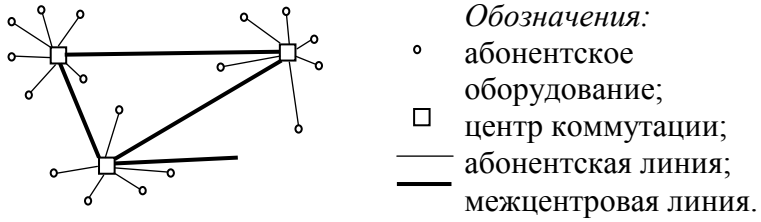

Рис. 7.1. Состав сети электросвязи

Абонентские линии в большинстве случаев являются одноканальными, межцентровые линии, как правило, многоканальные. В сети в любой момент времени существует какое-то количество каналов. Канал, связывающий двух абонентов сети, называется *некоммутируемым* (арендованным, постоянным), если он постоянно готов к передаче, так как не используется другими абонентами. *Коммутируемый канал* создается лишь на время передачи сообщения либо его части (пакета), в остальное время эти технические средства используются для создания подобных каналов для других абонентов. То есть, передачу сообщения можно начать лишь после того, как сеть произведет нужное соединение.

Первичная сеть – это линии и коммутационное оборудование, служащее для образования каналов со стандартными характеристиками, пригодных для передачи сообщений разных видов. *Вторичная сеть* – это структура, организующая обмен сообщениями определенного вида с использованием арендованных ресурсов первичной сети и соответствующего оконечного оборудования. Примеры вторичных сетей – это телефонная сеть, сеть факсимильной связи, телеграфная сеть, сеть передачи газетных полос, сети радио- и телевизионного вещания, компьютерные сети и т.д.

Сеть сигнализации – это сеть, предназначенная для обслуживания первичной сети и используемая для передачи команд, управляющих работой центров коммутации каналов.

По площади охвата различают локальные, местные, городские, зональ-

ные, общегосударственные и глобальные сети.

В зависимости от принадлежности сети бывают государственными, ведомственными и частными.

Структура сети в значительной степени определяется принятым в ней способом коммутации.

Пусть центры коммутации в сети отсутствуют, коммутация производится самим отправителем сообщения. Единственной возможной структурой является *полносвязная* (каждый с каждым) (рис. 7.2а). Эта структура самая живучая (в случае аварии найдется обходный путь) и самая дорогостоящая. Таким образом соединены, например, центры зон в телефонной сети страны.

Пусть центры коммутации в сети отсутствуют, но коммутация производится получателем сообщения. Единственной возможной структурой является *общая шина* (общая среда передачи) (рис. 7.2б). Это наиболее популяр-

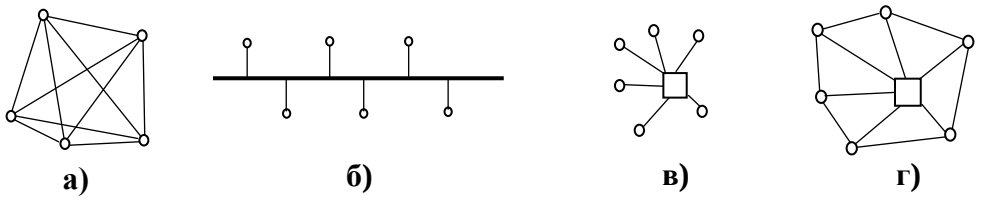

Рис. 7.2. Варианты структуры сети: а) полносвязная;
б) общая шина; в) звезда; г) радиально-кольцевая

ная структура для локальных компьютерных сетей, сетей вещания.

При наличии центров коммутации наиболее распространенными являются структуры типа “звезда” и “радиально-кольцевая” (рис. 7.2в и 7.2г) для небольших сетей (сеть предприятия, например) и *радиально-узловая* (рис. 7.1) для более крупных сетей, включая и глобальные. Сети такого типа построены, как правило, по иерархическому принципу, то есть имеют центры коммутации различных уровней, например, АТС предприятия, района, городская АТС для междугородной связи и.т.д.

И, наконец, сети, содержащие центры коммутации, подразделяются на сети с *коммутацией каналов* (аналоговые либо цифровые) и сети с *коммутацией пакетов* (только цифровые).

7.2 Коммутация каналов и коммутация пакетов

В сети с *коммутацией каналов* (КК) передачу сообщения (телеграмма, файл, телефонный разговор) можно начать лишь после того как будет организован сквозной канал от источника к получателю. Если хотя бы один из участков занят, инициатор соединения получает от сети отказ (например, всем знакомые частые гудки) и через некоторое время может повторить попытку. Пример – телефонная сеть.

В сети с *коммутацией пакетов* (КП) очередной пакет (малая часть сообщения) передается до ближайшего центра коммутации, там он вносится в память, ставится в очередь и передается дольше, как только найдется свободный канал в межцентровой линии в нужном направлении. В следующем центре коммутации повторяется то же самое. Примеры – телеграфная сеть, компьютерные сети. В такой сети вызывающий абонент не получает отказов, но возможны непредсказуемые задержки в доставке пакетов.

Условия применения метода КК: у всех абонентов имеется оборудование, полностью совместимое по всем характеристикам: по методам модуляции, кодирования, шифрования, скорости передачи и т.п.

Условия применения метода КП: у абонентов имеется оборудование, входящее в перечень разрешенного оборудования (этот перечень может быть довольно большим); абоненты согласны с тем, что пакеты будут доставляться с задержкой, величина которой случайна.

Выбор того или другого метода коммутации определяется, в первую очередь, характером нагрузки, создаваемой источниками информации в этой сети. В этой классификации с одной стороны можно расположить источники, создающие *постоянную нагрузку* во время передачи сообщения. Типичным представителем является источник телевизионного сигнала: независимо от характера передаваемого сюжета через каждые 1/25 секунды нужно передавать новый кадр. Характеру такого источника лучше соответствует метод КК.

С другой стороны расположим источники, создающие *пульсирующую нагрузку*. Типичный представитель – это компьютер в сети. Например, во время интерактивного обмена (диалога) лишь изредка возникает необходимость в передаче сравнительно коротких сообщений. В паузах канал связи не нужен. Метод КП и был предложен для согласования метода передачи с характером таких источников.

В той или иной мере *пульсирующий характер имеет нагрузка, создаваемая большинством источников*. Например, во время сеанса телефонной связи мы 50% времени молчим, слушая собеседника (если разговаривают женщина и мужчина, то для него этот процент значительно выше). Много времени занимают паузы между словами, фразами, когда можно канал не занимать.

Сравнительная характеристика методов КК и КП.

1) Использование метода КП благодаря возможности создать очередь позволяет *обеспечить почти стопроцентное использование пропускной способности межцентровых линий*, а в сетях с КК к этому пределу приблизиться нельзя, поскольку вероятность создания сквозного канала падает с увеличением нагрузки в сети. Например, пусть для создания сквозного канала нужно соединить 10 независимых участков, и для каждого вероятность наличия свободного канала равна 0,7. Тогда вероятность установить соединение с

первой попытки равна $P = 0{,}7^{10} = 0{,}028$, то есть недопустимо мала. Это является самым главным достоинством метода КП, поскольку стоимость линий составляет большую часть стоимости любой сети, особенно кабельной.

2) После установления соединения в сети с КК абоненту гарантируется определенное *качество обслуживания*, то есть фиксированная полоса пропускания канала и, как следствие, скорость передачи не ниже заранее определенного значения, фиксированные значения битовой вероятности ошибки и задержки сигнала в канале, причем все эти параметры не зависят от возникающих потребностей других абонентов. В сети с КП все параметры, определяющие качество обслуживания, зависят от активности других абонентов, поэтому случайным образом изменяются во времени и в час наибольшей нагрузки (ЧНН) могут принять недопустимые значения.

3) В сети с КК задержка случайной величины возникает перед началом передачи сообщения, во время попыток установить соединение, а в сети с КП перед началом передачи задержки нет, но зато случайная задержка возникает в процессе передачи.

4) В сети с КК характеристики сквозного тракта в наибольшей степени определяются характеристиками наихудшего звена в цепи, а при КП передачу пакетов в этом звене можно вести с пониженной скоростью, с использованием более мощного кодирования, направляя часть пакетов в обход и т.п.

5) В сети с КП передача между соседними центрами коммутации ведется в некоммутируемых каналах, обладающих скорректированными частотной и фазовой характеристиками; в центрах коммутации осуществляется двухуровневая регенерация сигнала (модулятор-демодулятор, кодер-декодер); обычно используется канал переспроса. Благодаря этому сеть с КП способна обеспечить на несколько порядков более низкую вероятность ошибки по сравнению с сетью с КК.

6) В сети с КП каждый пакет имеет заголовок, в котором содержится вся служебная информация, необходимая для его доставки по назначению (адреса источника и получателя, порядковый номер пакета в серии, значения других параметров данного пакета, требования к качеству обслуживания и т.п.). При КК такая информация передается только один раз, во время установления соединения, поэтому формально в сети с КП избыточность (отношение длины заголовка к длине пакета) выше, чем в сети с КК. Фактически ситуация изменяется на обратную, если учесть, что в паузах между соседними пакетами канал используется для передачи пакетов от других абонентов (см. п.1).

7) В сети с КК работу по организации соединения, например, неоднократный набор номера, включая даже выбор маршрута (в телефонной сети нужно знать код страны, код города, номер абонента), выполняет вызывающий абонент; в сети с КП большую часть этой работы выполняет сеть.

8) В сети с КК при аварии на любом из участков сквозной канал разру-

шается, а в сети с КП центры коммутации могут самостоятельно направить остальные пакеты по обходному пути, а в оконечном центре коммутации произвести сбор и расстановку поступивших пакетов в соответствии с их порядковыми номерами.

9) В сети с КК радикальная модернизация, включающая изменение способов модуляции, кодирования и т.п., может быть проведена лишь во всей сети сразу; в сети с КП такую модернизацию можно проводить по частям.

Проведенное сравнение показывает, что *сеть с коммутацией пакетов обладает большей гибкостью и способна обеспечить более высокую надежность и достоверность передачи сообщений при более эффективном использовании ресурсов сети, а сеть с коммутацией каналов обеспечивает более высокую гарантию стабильного качества обслуживания, хотя в целом его уровень может быть ниже.*

Тогда возникает вопрос: насколько универсален метод КП, в какой мере такая сеть пригодна для передачи всевозможных сообщений, особенно сообщений от источников, создающих постоянную нагрузку? В частности, пригодна ли такая сеть для организации телефонной связи? На этот вопрос можно ответить утвердительно хотя бы потому, что уже существует и успешно развивается междугородная и международная IP-телефония, то есть ведение телефонных переговоров через сеть Интернет, в которой применяется метод коммутации пакетов.

Оказывается, вопрос не в том, существуют ли задержки в процессе передачи пакетов, а в том, насколько велика средняя величина задержки и, главное, насколько велики случайные вариации этой задержки. Например, все уже привыкли к тому, что даже в сети с КК при передаче сигнала через спутниковый ретранслятор на геостационарной орбите (высота около 36 тыс. км) существует задержка примерно в 0,3 секунды. Это, конечно, создает кому-то небольшие неудобства при ведении телефонного диалога, но для телевизионного вещания и, особенно, при обмене файлами можно было бы смириться и с большей задержкой. Главное, что ее величина постоянна.

Следует отметить одно замечательное свойство, вытекающее из закона больших чисел теории вероятностей: чем больше абонентов обслуживает сеть с КП и ее отдельные участки, тем относительно стабильнее становится суммарный поток требований абонентов и, следовательно, тем меньше становятся средняя величина задержки при передаче пакетов и величина ее случайных вариаций. Разумеется, все это только при условии, что пропускная способность участков сети выше, чем суммарная нагрузка, создаваемая абонентами. Относительная величина необходимого превышения (запаса) также уменьшается с увеличением количества абонентов (подробнее об этом см. разд. 7.5).

Таким образом, *при передаче больших потоков информации в сети с КП имеется реальная возможность обеспечить высокое качество обслужи-*

вания при высокой эффективности использования ресурсов сети.

Справедливости ради отметим, что и в сети с КК при увеличении потоков повышается эффективность, но это никак не уменьшает преимуществ сети с КП, отмеченных выше.

Современная тенденция при создании новых сетевых технологий заключается в разработке "гибридных" методов с тем, чтобы использовать достоинства обоих методов и, по возможности, ослабить их недостатки. За основу берется метод коммутации пакетов и дополняется некоторыми процедурами, присущими методу коммутации каналов, например, процедурой предварительного создания сквозного "виртуального" канала (разд. 7.4).

7.3 Центры коммутации

Принципы построения и функции **центра коммутации каналов** рассмотрим на примере автоматической телефонной станции (АТС).

Основная задача центра КК – это организация по требованию физического соединения любой из *n* входных линий с любой другой из *n* выходных линий и удержание этого соединения на время передачи сообщения. Центр КК должен одновременно поддерживать до *m* таких соединений. Непосредственно эту задачу решает устройство коммутации (коммутационное поле).

Чтобы решить эту задачу, центр коммутации должен принимать и анализировать сообщения, поступающие из сети сигнализации, формировать подобные сообщения и направлять их в другие центры КК, контролировать состояние каналов передачи и вырабатывать соответствующие команды при изменении ситуации, вести учет и тарификацию произведенных соединений, проводить тестирование состояния оборудования, организовывать предоставление абонентам ряда дополнительных услуг и тому подобное. Все эти функции выполняет устройство управления центра коммутации.

По-видимому, можно выделить три основных этапа развития АТС.

Первое поколение – это декадно-шаговые (электромеханические) АТС. Элементом устройства коммутации является декадно-шаговый искатель. Он содержит 10 неподвижных контактов, расположенных по окружности, и один подвижный контакт, поэтому способен производить гальваническое соединение любой из десяти входных линий с одной выходной линией либо, при обратном включении, одной входной с любой из десяти выходных (рис. 7.3).

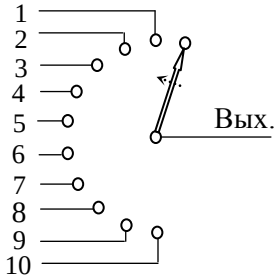

Рис. 7.3. Схема декадно-шагового искателя

Перемещение подвижного контакта из исходного положения производится электромагнитом, на который подается серия, содержащая от 1 до 10 импульсов. При подаче очередного импульса контакт перемещается на один шаг. Генератором этой серии импульсов при наборе очередной цифры телефонного номера является дисковый номеронабиратель в телефонном аппарате вызывающего абонента. В итоге для организации соединения, задаваемого, например, шестизначным номером, и удержания его на все время телефонного разговора будут задействованы шесть декадно-шаговых искателей, находящихся даже в разных АТС, расположенных вдоль линии связи.

Оборудование декадно-шаговой АТС имеет большие массу и размеры, высокую стоимость, для него характерны значительные колебания сопротивления контактов, открытых для внешних воздействий. Последнее обстоятельство является причиной кратковременных (порядка единиц миллисекунд) замираний сигнала, что особенно пагубно влияет на передачу данных. Декадно-шаговые АТС морально устарели и сняты с производства, хотя еще кое-где находятся в эксплуатации.

Второе поколение – это координатные АТС (*АТС с пространственной коммутацией*). Элементом устройства коммутации является многократный координатный соединитель (МКС). Он содержит набор из *n* “горизонтальных” и *n* “вертикальных” шин, при этом в каждом узле имеется управляемый ключ (рис. 7.4). Ключ – это малоразмерный герметизированный контакт (гер-

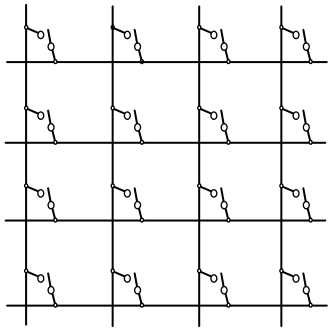

Рис. 7.4. Схема многократного координатного соединителя

кон), в котором замыкание контактов происходит под воздействием внешнего магнитного поля, создаваемого миниатюрным электромагнитом. Управляющие напряжения на эти электромагниты подаются от устройства управления АТС. В итоге любой вход можно соединить с любым выходом, причем одновременно поддерживать до *n* соединений.

Координатные АТС обладают теми же недостатками, но в значительно меньшей степени. Благодаря тому, что все МКС работают по сигналам из устройства управления АТС, то есть из специализированного компьютера, есть возможность предоставлять абонентам ряд дополнительных услуг.

Третье поколение – это электронные (*цифровые*) *АТС*. Элементом устройства коммутации является пара мультиплексор-демультиплексор (рис.

7.5), откуда следует, что в АТС этого типа производятся такие же преобразо-

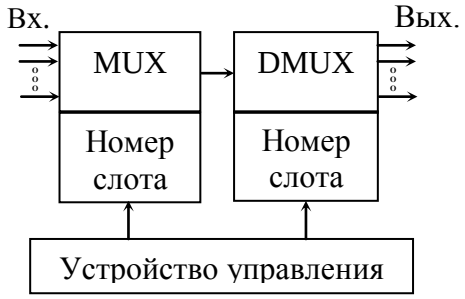

Рис. 7.5. Схема коммутатора временных интервалов

вания сигналов, как в многоканальной цифровой СПИ с ВРК (рис. 6.3). Имеются лишь два отличия. Во-первых, не нужен селектор синхросигналов, так как мультиплексор и демультиплексор находятся рядом и работают по командам от общего устройства управления. Второе отличие связано с порядком подключения *n* входов и *n* выходов. В СПИ с ВРК в любой момент времени номер А входной линии, подключаемой к выходу мультиплексора, совпадает с номером В той выходной линии, которая подключается к входу демультиплексора. В коммутаторе АТС эти два номера всегда различны, хотя в целом все *n* пар номеров повторяются в том же порядке от цикла к циклу (через 125 мкс) до тех пор, пока не закончится какой либо разговор либо не поступит требование на новое соединение. Тогда происходит переформирование пар, но оно, разумеется, не затрагивает номера слотов тех абонентов, которые еще не закончили разговор.

Например, при *n*=4 номера чередуются в следующем порядке

Номер А	…	1	2	3	4	1	2	3	4	1	2	3	4	…
Номер В	…	3	4	2	1	3	4	2	1	3	4	2	1	…

Это значит, что в первом слоте первый вход подключен к третьему выходу, во втором слоте второй вход подключен к четвертому выходу и так далее в каждом цикле. Такой способ носит название *коммутация временных интервалов*. Нет никаких препятствий для того, чтобы несколько входных линий соединить с одной выходной линией для организации совместного разговора.

За время одного подключения (в одном слоте) передается один отсчет телефонного сигнала в цифровой форме, то есть 8 бит. Поэтому все аналоговые телефонные сигналы, приходящие по абонентским линиям, на входе в АТС подвергаются аналого-цифровому преобразованию, а все сигналы, направляемые в абонентские линии – цифро-аналоговому преобразованию.

Нередко (когда оба абонента обслуживаются одной АТС) оба преобразования делаются лишь для того, чтобы осуществить коммутацию в АТС.

Основная задача **центра коммутации пакетов** та же – это организация по требованию физического соединения любой из *n* входных линий с любой другой из *n* выходных линий, но *сохраняется это соединение лишь на короткое время – на время передачи одного пакета*. Непосредственно эту задачу решает устройство коммутации цифровых сигналов.

Чтобы решить эту задачу, центр коммутации должен производить демодуляцию принимаемых импульсов и декодирование комбинаций корректирующего кода (проводить обнаружение ошибок), выделять и анализировать заголовки пакетов (хотя бы для того, чтобы определить, в каком направлении его отправить дальше либо куда направить запрос на повторную передачу пакета в случае обнаружения в нем ошибок), с учетом текущей нагрузки и состояния линий выбрать направление дальнейшей передачи пакета и поставить его в очередь, провести кодирование и модуляцию перед отправкой пакета, контролировать состояние каналов передачи, вести учет величины трафика и тарификацию, проводить тестирование состояния оборудования и тому подобное.

Приведенный далеко не полный перечень показывает, что функции центра КП намного сложнее и разнообразнее, чем функции центра КК, поэтому нередко центр КП – это многопроцессорная вычислительная система.

Очевидно, что в больших сетях с коммутацией пакетов нет необходимости в том, чтобы все центры коммутации были способны выполнять все перечисленные функции в полном объеме, а в малых сетях многие из функций вообще не нужны. Поэтому существуют центры коммутации пакетов разных классов. Укажем некоторые из них, учитывая, что одна из важнейших разновидностей сетей с коммутацией пакетов – это компьютерные сети.

Мост (bridge) делит общую среду, в которой передаются пакеты, например, общую шину (рис. 7.2б), на части (подсети) и выпускает пакет за пределы данной подсети лишь в случае необходимости. Имеет один процессорный блок.

Коммутатор (switch) имеет процессор на каждом порте, то есть это – совокупность мостов. Способен обрабатывать пакеты независимо в параллельном режиме.

Маршрутизатор (router) способен выбирать наиболее рациональный маршрут, используя явную числовую адресацию подсетей с учетом их иерархии. Способен преобразовывать формат пакета для связи подсетей с разной сетевой технологией.

Шлюз (gateway) обладает полным набором функций и может объединять сети с разными типами системного и прикладного программного обеспечения.

Кроме того, для каждого вида сообщений существуют оптимальные

значения параметров пакетов. Например, при передаче телевизионного или телефонного сигнала желательно использовать короткие пакеты, чтобы не увеличивать задержек, связанных с формированием, а также обработкой принимаемых пакетов. При передаче файлов допустимо использование более длинных пакетов для уменьшения избыточности. Отметим еще, что создано огромное количество вторичных сетей с коммутацией пакетов самого разнообразного назначения.

Все это выдвигает на первый план **проблему унификации, стандартизации** структуры и параметров пакетов, а также способов их обработки на разных этапах передачи. Особое внимание должно быть уделено стандартизации *интерфейсов* (стыков), то есть границам раздела двух устройств или систем.

Широко применяемый метод стандартизации – это разработка протоколов обмена информацией. *Протокол* – это соглашение, касающееся управления процедурами информационного обмена между взаимодействующими объектами, то есть управления форматом сообщений, формированием контрольной информации, потоком команд и действиями, которые предпринимаются в случае обнаружения ошибок. Предполагается, что эти объекты не имеют средств прямого обмена информацией, но могут пересылать данные через соответствующий локальный интерфейс протоколам нижнего уровня, пока не будет достигнут самый нижний – физический. Информация передается в удаленный пункт с использованием протокола самого низкого уровня и далее продвигается вверх через систему интерфейсов до соответствующего уровня в пункте назначения. Для каждой сети нужен набор протоколов, определяющих способ взаимодействия удаленных объектов на данном уровне, и набор интерфейсов, определяющих способ взаимодействия между соседними уровнями. Такой набор называется пакетом (стеком) протоколов. Именно он определяет способ обработки пакетов в центрах коммутации.

В настоящее время разработано большое количество протоколов разного уровня, поэтому создана концепция, согласно которой связное оборудование с различными протоколами может объединяться в систему с помощью сети передачи данных. Такой стандарт стандартов называется *семиуровневой эталонной моделью взаимодействия открытых систем*. Он определяет семь уровней взаимодействия, начиная с нижнего: физический, канальный, сетевой, транспортный, сеансовый, представления данных и прикладной. Первые три уровня относятся к передаче и маршрутизации данных. Последние три уровня обслуживают пользовательские приложения. Четвертый уровень обеспечивает стыковку трех нижних и трех верхних уровней. Эта модель пока не реализована в виде законченного множества протоколов для всех уровней, но уже дает возможность упорядочить имеющиеся протоколы.

7.4 Дейтаграммный метод передачи и передача с предварительным установлением соединения

Дейтаграмма (datagram) – это самостоятельный пакет, в заголовке которого содержится полная информация, необходимая для его передачи от источника к получателю независимо от всех предыдущих и последующих пакетов, даже если они являются частями одного сообщения. Поэтому в каждом центре коммутации каждую приходящую дейтаграмму обрабатывают как новое сообщение, вследствие чего во время сеанса связи двух абонентов пакеты могут приходить в пункт назначения по разным путям, с различным запаздыванием и даже не в том порядке, в каком они были переданы. Некоторые пакеты могут быть потеряны (искажены из-за воздействия помех). Дейтаграммный способ передачи – это наиболее полное воплощение идей метода коммутации пакетов.

Примером сети, в которой используется дейтаграммный метод передачи, является Интернет (стек протоколов TCP/IP). Заголовок пакета содержит следующую информацию: номер версии протокола IP; длина заголовка; требуемый тип обслуживания (малая задержка, высокая надежность и т.п.); общая длина пакета; общая длина, флаг и величина смещения фрагмента (информация, необходимая для сборки пакета из фрагментов, на которые в случае необходимости он был разбит в одном из центров коммутации); время жизни (число маршрутизаторов на пути пакета, после прохождения которых пакет уничтожается как заблудившийся); тип протокола верхнего уровня, нужного для обработки данных в пакете; контрольная сумма (проверочные биты) заголовка; полный числовой адрес отправителя; полный числовой адрес получателя и прочее.

Этот список приведен лишь для того, чтобы показать, что длина заголовка велика. Вслед за заголовком идет информационная часть пакета, которая также кодируется циклическим кодом с 16 проверочными символами. Чтобы избыточность была не слишком большой, информационную часть следует увеличить, в итоге общая длина пакета может доходить до 64 кбайт.

Сеть, в которой применяется дейтаграммный метод передачи, функционирует по принципу самоорганизации, она не имеет общего управления, хотя есть общие базы данных. Используются довольно простые протоколы, в узлах сети имеется минимум информации о проходящих пакетах и качестве обслуживания. Все это обеспечивает высокую живучесть сети и открывает широкие возможности для ее наращивания и изменения структуры.

С другой стороны, в каждом узле (центре коммутации) в случае обнаружения ошибок в предыдущий узел посылается запрос на повторную передачу пакета. Послать такой запрос можно только после завершения приема всего пакета. Из-за этого при передаче длинных пакетов по низкоскоростным линиям, характеризующимся высокой вероятностью появления ошибок, воз-

никают большие случайные задержки. В итоге чрезвычайно трудно обеспечить высокое качество обслуживания.

Таким образом, дейтаграммный метод малопригоден для скоростной высококачественной передачи больших объемов информации. Это положение может измениться, если на всех участках тракта от источника к получателю будут использоваться каналы с высокой пропускной способностью и столь малой вероятностью появления ошибок, при которой повторные передачи пакетов станут весьма редкими.

Указанный недостаток дейтаграммного метода привел к разработке **методов передачи с коммутацией пакетов с предварительным установлением соединения**, основанных на идее создания виртуального канала.

Первый пакет из серии пакетов, представляющей сообщение, которое предстоит передать конкретному получателю, направляется до ближайшего центра коммутации. Этот пакет имеет заголовок большого объема и подобен заголовку дейтаграммы, то есть содержит всю необходимую информацию для его доставки в пункт назначения. Поэтому он передается таким же образом, как передавалась бы дейтаграмма. Разница заключается в следующем. При использовании дейтаграммного метода центру коммутации, через который проходит очередной пакет, нет необходимости запоминать сведения, содержащиеся в заголовке этого пакета, так как в заголовке следующего пакета будет то же самое. *При передаче пакетов с предварительным установлением соединения в каждом центре коммутации, через который проходит первый пакет серии, запоминаются сведения, содержащиеся в его заголовке, для того, чтобы знать, что нужно делать с остальными пакетами этой же серии*.

Все остальные пакеты серии пойдут по тому же маршруту, поэтому их заголовки могут быть очень короткими. В заголовке каждого пакета достаточно указать его порядковый номер в серии (во избежание путаницы при повторных передачах пакетов с ошибками) и краткое указание на принадлежность данного пакета к конкретной серии, так как каждый центр коммутации одновременно обслуживает несколько серий.

Таким образом первый пакет серии прокладывает путь остальным пакетам, то есть создает сквозной *виртуальный канал*. Он так называется потому, что существует постоянно в течение всего времени передачи серии пакетов не как реальное соединение линий передачи, а лишь в виде записей в центрах коммутации вдоль маршрута. Виртуальный канал превращается в физический канал лишь на короткое время передачи очередного пакета от одного центра коммутации к соседнему. Последний пакет серии несет команду удалить из памяти центров коммутации те сведения, которые были переданы с первым пакетом, после чего данный виртуальный канал просто перестает существовать (произошло разъединение абонентов). Если во время передачи серии на одном из участков произошла авария, то, в отличие от дейтаграммного метода, передача пакетов прерывается, и виртуальный канал раз-

рушается. Чтобы продолжить передачу, нужно заново осуществить процедуру установления соединения (все описанное напоминает то, что происходит в сети с коммутацией реальных каналов).

Центр коммутации, направляя очередной пакет к соседнему центру, вносит в заголовок этого пакета короткий порядковый номер, который при установлении соединения был присвоен данной серии. На каждом исходящем направлении ведется собственная нумерация, при этом номера, освободившиеся после окончания какой-либо серии, используются для нумерации вновь созданных соединений. Такой короткий номер, называемый *номером исходящего виртуального канала*, вполне заменяет длинный полный адрес пункта назначения. Для следующего центра коммутации этот номер играет роль *номера входящего виртуального канала*, и его вполне достаточно, чтобы определить, к какой серии принадлежит пакет. Направляя пакет дальше, центр коммутации вносит в заголовок этого пакета свой номер исходящего виртуального канала.

Главное преимущество методов пакетной передачи с предварительным установлением соединения по сравнению с дейтаграммным методом – это существенно меньшая длина заголовков передаваемых пакетов, что заметно уменьшает избыточность. Можно использовать более короткие пакеты, что также повышает скорость и достоверность передачи. С другой стороны, обработка пакетов в центрах коммутации усложняется.

Одним из первых протоколов такого типа, определяющих принципы функционирования вторичных сетей на основе низкоскоростных цифровых или снабженных модемами аналоговых каналов, является *протокол X.25* (1974 год). Очередной пакет (кадр) в начале и в конце имеет синхросигнал – флаг 01111110. Номер виртуального канала содержит от 6 до 10 бит. Для обнаружения ошибок применяется кодирование циклическим кодом с 16 проверочными символами. Если обнаружены ошибки, в предыдущий узел посылается запрос на повторную передачу пакета, что приводит к дополнительным задержкам. Протокол X.25 еще слабо использует преимущества предварительного установления соединения, поэтому гарантии обеспечения требуемого качества обслуживания не намного выше по сравнению с дейтаграммным методом.

Следующий шаг по использованию этих преимуществ был сделан при разработке *технологии Frame Relay*, ориентированной на использование оптоволоконных линий. Они отличаются высокой верностью передачи, то есть количество ошибочных пакетов оказывается на 1-3 порядка меньше, чем при передаче с использованием других линий. Учитывая это, во всех промежуточных узлах проводится обнаружение ошибок лишь в заголовках пакетов, а информационная часть не проверяется. Если в заголовке ошибок не обнаружено, то сразу, не дожидаясь окончания ввода всего пакета, начинается его передача к следующему узлу. Если в заголовке обнаружены ошибки, пакет

уничтожается.

Обнаружение ошибок в информационной части пакета проводится лишь в оконечном узле. Если в пакете обнаружены ошибки, запрос на его повторную передачу (и передачу потерянных пакетов) направляется в первый узел. Это, конечно, немного увеличивает нагрузку в сети, но зато существенно ускоряет обработку пакетов в узлах сети.

Использование оптоволоконных линий предполагается и в еще более современной технологии, называемой асинхронным режимом переноса, но обычно применяется сокращенное название *ATM (Asynchronous Transfer Mode)*. В дополнение к тому, что было сказано о технологии Frame Relay, все пакеты, называемые ячейками АТМ, имеют фиксированную и довольно малую длину – 53 байта. Из них 5 байт – это заголовок, остальные 48 байт – информационная часть. Из 40 бит заголовка 8 позиций отведено для проверочных символов циклического кода с порождающим полиномом $g(x) = x^8 + x^2 + x + 1$.

Длина ячейки выбрана на основе компромисса между потребностями телефонии (желательно использовать короткие пакеты для уменьшения задержек) и передачи данных (желательно использовать более длинные пакеты для уменьшения избыточности). Единый размер ячеек независимо от их содержимого заметно упрощает и ускоряет их обработку в узлах сети.

Второе отличие сети АТМ от остальных сетей с коммутацией пакетов – это *гарантия обеспечения заявленного качества обслуживания,* разумеется, лишь после установления соединения (напомним, что именно это мы отметили как главное достоинство метода коммутации каналов).

Посылая запрос на установление соединения, абонент заявляет свои требования к средней скорости передачи, допустимым величинам задержек и вероятности ошибок. Сеть анализирует требования тех абонентов, для которых она уже установила соединения (разумеется, виртуальные), и оценивает свои возможности удовлетворить требования новой заявки. Если возможность есть, заявка принимается, если нет, то следует отказ в установлении соединения. Вызывающий абонент может сделать новую попытку (как в сети с коммутацией каналов).

Поскольку передача пакетов – процесс случайный, сеть может оценить свои возможности лишь в вероятностном смысле. Может возникнуть ситуация, когда на каком-либо участке нагрузка превышает его пропускную способность, поэтому часть пакетов придется выбросить. В первую очередь уничтожаются пакеты от тех абонентов, которые превышают свои требования, но уровень приоритета при этом учитывается, как и в любой другой системе.

Технология АТМ универсальна (в ячейках АТМ можно размещать биты цифрового сигнала любого происхождения) и пригодна для передачи информации со скоростями до 10 Гбит/с и даже выше. Ячейки АТМ переносятся в битовом потоке, сформированном в некотором другом стандарте, приня-

том в первичной сети, например PDH или SDH. Очевидно, что второй стандарт подходит несравненно больше. Ячейки ATM просто размещают одну за другой в виртуальном контейнере четвертого порядка VC-4 до его заполнения, и этот контейнер помещают в STM-1.

Таким образом, новые сетевые технологии демонстрируют явную тенденцию к соединению положительных особенностей двух методов коммутации.

7.5 Элементы теории телетрафика

Кроме индивидуальных абонентского оборудования и абонентских линий в сети имеется и оборудование общего пользования (центры коммутации, межцентровые абонентские линии и т.п.). Если бы число обслуживающих приборов N в таком общем оборудовании (например, число одновременно поддерживаемых соединений в центре коммутации АТС) было равно числу абонентов M, обслуживаемых данной АТС, то каждый абонент всегда получал бы обслуживание по первому требованию, и обсуждать в данном разделе было бы нечего.

Очевидно, что такая идеальная система экономически неэффективна, поскольку у каждого абонента лишь иногда, случайным образом возникает необходимость провести сеанс связи. Поэтому всегда $N < M$, но с уменьшением N ухудшается качество обслуживания, поскольку часть заявок не будет обслужена либо будет обслужена после ожидания в очереди. Те методы, которые мы обсуждали в разд. 7.2–7.4, были разработаны для разрешения этого противоречия, но никаких количественных оценок, характеризующих качество обслуживания, мы пока не делали.

Именно это составляет предмет теории массового обслуживания и теории очередей, причем первые исследования в области теории телетрафика были проведены датским ученым А.К. Эрлангом методами теории вероятностей [4,5].

Ключевым является понятие нагрузки. Начнем с *идеальной системы обслуживания* ($N = M$). Пусть поведение абонента номер 1 характеризуется безразмерной стационарной случайной функцией его активности

$$a_1(t) = \begin{cases} 1, & \text{когда он передает информацию,} \\ 0, & \text{в остальное время.} \end{cases} \qquad (7.1)$$

Математическое ожидание этой функции, то есть среднее значение активности абонента

$$m_1 = P\left[a_1(t) = 1\right] = A_1 \quad \text{Эрл} \qquad (7.2)$$

численно равно вероятности того, что в произвольно выбранный момент времени абонент активен. Эта также безразмерная величина носит название *нагрузки, создаваемой данным абонентом,* и измеряется в эрлангах. Экспериментально ее можно оценить как

$$A_1 \approx T_è / T, \qquad (7.3)$$

где T_u – суммарное время передачи информации за длительный промежуток времени T.

Поскольку активность абонента имеет импульсный характер, полезно ввести еще два параметра:

C_1 – среднее количество требований на обслуживание (вызовов или пакетов), поступающих от абонента за время T (обычно это час наибольшей нагрузки (ЧНН)),

τ – средняя продолжительность обслуживания одного вызова.

Тогда в соответствии с (7.3)

$$A_1 = \frac{C_1 \tau}{T} = \lambda_1 \tau, \qquad (7.4)$$

где $\lambda_1 = C_1/T$ – средняя интенсивность поступления вызовов от абонента.

Например, каждый квартирный телефон в ЧНН создает нагрузку A_1 = 0,05–0,1 Эрл, при этом средняя продолжительность разговора τ = 3–4 мин и λ_1 = 1–2 вызова в час.

Далее полагаем, что функции активности всех абонентов обладают одинаковыми свойствами, тогда показатели суммарной активности абонентов равны

$$a(t) = \sum_{k=1}^{M} a_k(t), \quad C = C_1 M, \quad \lambda = \lambda_1 M, \quad A = A_1 M = \lambda \tau, \qquad (7.5)$$

где *A – общая поступающая нагрузка.*

Считаем, что все абоненты действуют независимо один от другого. Вероятность того, что в произвольно выбранный момент времени ровно q абонентов находятся на обслуживании, находим по биномиальной формуле

$$P(q) = \frac{M!}{q! \cdot (M-q)!} A_1^q (1 - A_1)^{M-q}. \qquad (7.6)$$

На рис. 7.6 для примера приведен график биномиального распределения суммарной активности 30 абонентов учрежденческой АТС для A_1=0,1, то есть общая поступающая нагрузка A = 3 и, в принципе, могла бы быть обслужена тремя приборами. Здесь лишь в 0,0089% времени для обслуживания требуется больше 10 приборов, а вероятность того, что для обслуживания требуется больше 15 приборов $P(q > 15) = 3{,}7 \cdot 10^{-9}$ чрезвычайно мала.

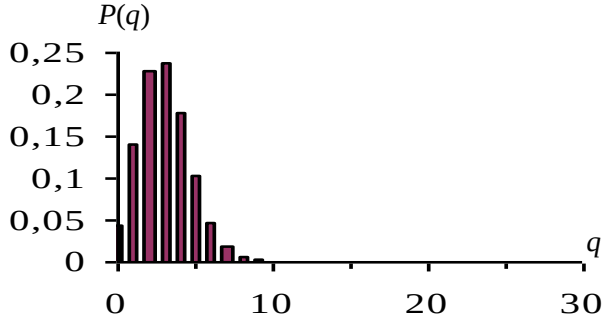

Рис. 7.6. Биномиальное распределение при A_1 = 0,1 и M = 30

Данные примера показывают, что количество приборов обслуживания N может быть во много раз меньше числа абонентов M. Таков принцип построения всех систем массового обслуживания.

Различают две группы методов анализа: для систем с потерями и для систем с ожиданием. В *системе с потерями* в случае, когда общая нагрузка превышает пропускную способность системы, равную N, избыточный трафик отбрасывается, в итоге *обслуженная нагрузка* A_o меньше чем поступающая A. Таков характер обслуживания в сетях с коммутацией каналов.

В *системе с ожиданием* избыточный трафик сохраняется, вызовы находятся в очереди, пока не освободятся устройства, которые могут их обслужить. Если, в идеале, длина очереди, играющей роль буфера, сглаживающего колебания поступающей нагрузки, не ограничена, и $A < N$, то $A_o = A$, то есть вся поступающая нагрузка когда-то будет обслужена. В реальных условиях длина очереди ограничена, и система с ожиданием частично приобретает черты системы с потерями. Таков характер обслуживания в сетях с коммутацией пакетов.

При численном анализе системы массового обслуживания в дополнение к показателям активности абонентов необходимо задать вероятностные характеристики потока требований, то есть потока заявок на установление соединения при КК либо потока пакетов при КП. Простейшая и довольно реалистическая модель, для которой получено большинство результатов, основана на следующих предположениях:

1) в любой достаточно малый интервал времени может поступить только одно требование;

2) вероятность $p = \lambda \cdot \Delta t$ поступления требования в любой достаточно малый интервал времени Δt пропорциональна длине этого интервала;

3) вероятность поступления требования в любой фиксированный интервал времени не зависит от количества требований, поступающих в других интервалах (очевидно, что это предположение основано на условии $N << M$);

4) длительность обслуживания одного вызова t_c без учета времени ожидания в очереди является непрерывной случайной величиной и подчиняется экспоненциальному закону распределения

$$P(t_c > t) = \exp(-t/\tau). \qquad (7.7)$$

Случайно выбранный интервал t разобьем на n малых интервалов $\Delta t = t/n$ и воспользуемся перечисленными предположениями и формулой умножения вероятностей независимых событий, чтобы вычислить вероятность того, что в течение интервала t не поступит ни одного требования,

$$P_0(\lambda t) = (1-p)^n = (1-\lambda \Delta t)^{t/\Delta t} = \exp(-\lambda t). \qquad (7.8)$$

Пользуясь теми же предположениями, получена более общая формула, по которой вычисляют вероятность того, что в течение интервала t поступят ровно j требований

$$P_j(\lambda t) = \left[(\lambda \Delta t)^j / j! \right] \exp(-\lambda t) . \qquad (7.9)$$

Это – закон Пуассона, поэтому такой *поток требований называется пуассоновским*. Формула (7.8) является частным случаем (7.9) при $j = 0$.

В **системе с потерями** ситуация, когда общая активность абонентов превышает пропускную способность системы (число приборов обслуживания N), называется *блокировкой*. В приведенном примере (рис. 7.6) при $N = 10$ вероятность блокировки равна $8{,}9 \cdot 10^{-5}$.

В *системе с явными потерями* заблокированные вызовы (получившие отказ при первой попытке) теряются, то есть абоненты не делают повторных попыток установить соединение, а направляют это сообщение другими, например обходными путями.

Вероятность блокировки в такой системе определяется по известной формуле Эрланга

$$B = \frac{A^N}{N! \left[1 + A + A^2/2! + A^3/3! + A^4/4! + \ldots + A^N/N! \right]} . \qquad (7.10)$$

На рис. 7.7 дана зависимость вероятности блокировки от нагрузки, поступающей на канал обслуживания, в системе с явными потерями. В связи с потерей некоторых требований нагрузка, обслуженная каждым прибором, несколько меньше поступающей нагрузки и равна

$$A_o / N = (1 - B) A / N . \qquad (7.11)$$

Хорошо видно, что *при увеличении размеров системы и соответствующем возрастании нагрузки A и числа приборов обслуживания N эффективность использования приборов обслуживания A_o/N возрастает и приближается к единице при том же качестве обслуживания* (при той же вероятности блокировки). Например, при $B = 0{,}001$ и $N = 4$ увеличение общей нагрузки в

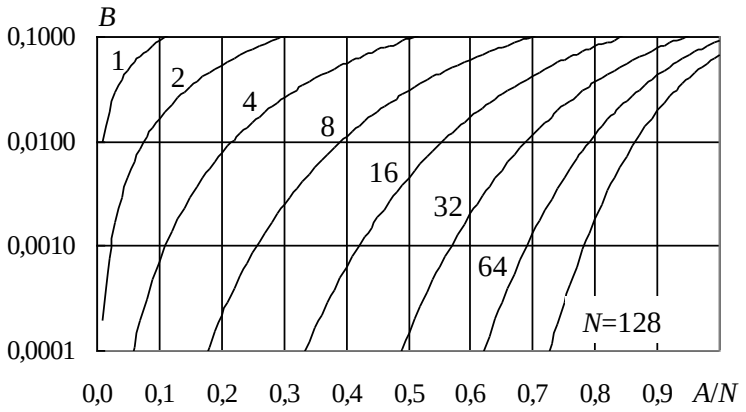

Рис. 7.7. Вероятность блокировки в зависимости от нагрузки на канал обслуживания в системе с явными потерями

сети в 5 раз потребует лишь двукратного увеличения числа приборов обслуживания (с 4 до 8) за счет того, что нагрузка, которую обрабатывает каждый прибор, увеличивается примерно в 2,5 раза.

Очевидно, что объединение малых обслуживающих систем в большую систему позволяет повысить ее эффективность, то есть при использовании меньшего числа приборов обеспечить обслуживание той же нагрузки с той же вероятностью блокировки.

Плата за высокую эффективность большой системы – это ее повышенная чувствительность к перегрузкам, поскольку она имеет меньший запас по пропускной способности. Например, при той же вероятности блокировки B = 0,001 в малой системе при $N = 4$ увеличение нагрузки на 10% приведет к несущественному возрастанию величины B, в то время как в более крупной системе при $N = 128$ то же возрастание нагрузки увеличивает вероятность блокировки на порядок.

Более реалистический вариант системы с потерями – это *система с повторными вызовами*. Она анализируется при следующих предположениях.

1) Все заблокированные вызовы в очередь не ставятся, но абоненты делают многократные повторные попытки до тех пор, пока не получат обслуживания.

2) Каждая из повторных попыток проводится через случайный интервал времени после предыдущей, причем длительности всех интервалов взаимно независимы, а средняя длина каждого интервала превышает среднее время обслуживания абонента τ.

Для интенсивности поступления вызовов первой попытки мы уже использовали обозначение λ. Для вероятности блокировки в этой системе с повторными вызовами используем обозначение B_r.

Поскольку доля заблокированных вызовов равна B_r, то вскоре произойдет $B_r\lambda$ повторных попыток, из них часть снова будет заблокирована и так далее. В итоге в состоянии статистического равновесия системы общая интенсивность поступления вызовов определяется как сумма убывающей геометрической прогрессии

$$\lambda_r = \lambda + B_r\lambda + B_r^2\lambda + B_r^3\lambda + ... = \lambda/(1-B_r), \qquad (7.12)$$

где B_r – вероятность блокировки, рассчитанная для такой же системы с явными потерями, но при интенсивности нагрузки λ_r.

В этом уравнении две неизвестные величины B_r и λ_r, поэтому его можно решить методом последовательных приближений по методике расчета системы с явными потерями (7.10): сначала для фактического значения λ по (7.10) находим первое приближение для B_r и используем его для вычисления первого приближения для λ_r по формуле (7.12). Далее это значение λ_r подставляем в (7.10), находим второе приближение для B_r и так далее.

Обычно вероятность блокировки невелика ($B \ll 1$), поэтому хорошая точность получается уже после первого приближения

$$\lambda_r \approx \lambda/(1-B), \qquad (7.13)$$

откуда видно, что повторные вызовы не сильно увеличивают нагрузку, так как $(1-B) \approx 1$. Итак, B_r определяем из (7.10) с использованием (7.13). Кстати,

распространять это утверждение на очень большие системы следует с осторожностью, так как они весьма чувствительны даже к малым перегрузкам.

В **системе с ожиданием** избыточный трафик помещается в очередь и сохраняется до начала обслуживания. При анализе примем те же простые предположения: $A < N$; $M >> N$; длина очереди не ограничена; длительность обслуживания вызова подчиняется экспоненциальному закону распределения (7.7).

Вероятность того, что вызов сталкивается с перегрузкой системы и, следовательно, помещается в очередь, также была определена Эрлангом

$$P_{î÷} = \frac{NB}{N - A(1-B)}, \qquad (7.14)$$

где B – вероятность блокировки в системе с явными потерями (7.10).

Вероятность того, что вызов, оказавшийся в очереди, ждет начала обслуживания не более времени t, равна

$$P_{î÷}(t_{î÷} \le t) = \exp\left[-\frac{(N-A)t}{\tau}\right], \qquad (7.15)$$

где τ – средняя длительность обслуживания (не ожидания), определенная в (7.7). Отсюда видно, что среднее время ожидания в очереди равно

$$\tau_{î÷} = \frac{\tau}{N-A}. \qquad (7.16)$$

Обратите внимание, что в формулах (7.14) – (7.16) фигурирует не отношение A/N, а разность этих величин $N–A$. Чем больше эта разность, тем меньше вероятность того, что вызов не будет обслужен сразу и попадет в очередь, и тем меньше время ожидания в очереди, если он туда попал. Очевидно, что в крупной системе, при больших значениях A и N, легче обеспечить большое значение этой разности даже при очень высокой эффективности использования обслуживающих приборов ($A/N \approx 1$), нежели в малой системе.

Итак, приведенные соотношения количественно доказывают справедливость тех утверждений, которые были высказаны в конце разд. 7.2.

Вопросы для самоконтроля по главе 7

1. Дайте определения симплексного, дуплексного и полудуплексного каналов.
2. Дайте определения методов передачи с коммутацией каналов и коммутацией пакетов.
3. Перечислите условия, при которых возможно применение метода: коммутации каналов; коммутации пакетов.
4. Дайте сравнительную характеристику методов коммутации каналов и коммутации пакетов.
5. Охарактеризуйте дейтаграммный метод передачи в сети.

6. В чем различие физического и виртуального каналов?
7. Почему декадно-шаговые АТС морально устарели?
8. Что такое коммутация временных интервалов?
9. Как осуществляется коммутация в координатной АТС?
10. Каково назначение сети сигнализации?
11. Перечислите основные типы топологических структур сетей электросвязи и укажите, при каких условиях их целесообразно применять.
12. Какие числовые характеристики используются для описания активности абонента в системе массового обслуживания?
13. Дайте определение единиц, в которых измеряется величина трафика.
14. Почему в больших сетях электросвязи можно более эффективно использовать оборудование при том же качестве обслуживания?
15. Дайте определение средней интенсивности поступления вызовов от абонента.
16. При каких условиях поток требований на обслуживание можно назвать пуассоновским?
17. При каких условиях число обслуживающих приборов может быть много меньше числа абонентов в сети?
18. Дайте определения систем обслуживания с явными потерями и с повторными вызовами.
19. Что зависит от предельной длины очереди в системе массового обслуживания с ожиданием?
20. Приведите формулу для среднего времени ожидания в системе массового обслуживания.
21. При каких условиях в системе массового обслуживания с ожиданием можно обеспечить любую малую вероятность блокировки?

ЗАКЛЮЧЕНИЕ

В книге представлена лишь малая часть того огромного арсенала методов, который используется в современных системах обмена информацией. Да и сами методы очерчены схематично, на уровне основной идеи, почти без всяких тонкостей, хотя на практике они могут иметь большое значение. Главное, чтобы система обмена информацией воспринималась не просто как набор модуляторов, кодеров и прочих устройств, а как единое целое, функционирующее ради решения общей задачи. Если к этому добавить уверенность в том, что любое улучшение одних характеристик системы передачи информации обязательно нужно сопоставлять с необходимыми для этого затратами либо ухудшением других характеристик, то читатель вполне созрел для более глубокого изучения интересующих его вопросов.

Когда идет речь о каком-либо преобразовании сигналов в присутствии шума, помех, то можно смело считать, что вы не добрались до его сути, пока не получили ответы на следующие вопросы:

1) Что из себя представляет полезная часть сигнала на входе (символ, последовательность символов, непрерывную функцию и т.п.) и какие вероятностные характеристики известны?

2) Каков тип помех (аддитивные, мультипликативные или иные), воздействующих на полезный сигнал до и в процессе данного преобразования? Каковы для них законы распределения вероятностей и числовые характеристики?

3) Какова цель данного преобразования? Если это решающее устройство (демодулятор, декодер и т.п.), то перечислить и дать ясное определение каждому из решений, которые могут быть приняты этим устройством.

4) Какова цена каждой из возможных ошибок, которые могут быть допущены в процессе вынесения решений, либо искажений сигнала при его преобразовании?

5) Есть ли в известной литературе описания преобразований сигналов, совпадающих с данным преобразованием по пп. 1-4? Имеются ли сведения о том, что какое-то из этих преобразований является оптимальным?

Успехов в дальнейшем освоении экспоненциально расширяющегося набора средств, применяемых для передачи информации!

ЛИТЕРАТУРА

ЦИТИРОВАННЫЕ ИСТОЧНИКИ

1. Радиосистемы передачи информации : Учебное пособие для вузов / В. А. Васин [и др.]; ред.: И. Б. Федоров, В. В. Калмыков. - М. : Горячая линия-Телеком, 2005.
2. Зюко А.Г. Теория электрической связи / А.Г. Зюко, Д.Д. Кловский, В.Н. Коржик, М.В Назаров / Под ред. Д.Д. Кловского. – М.: Радио и связь, 1997. – 432 с.
3. Скляр Б. Цифровая связь. Теоретические основы и практическое применение, 2-е изд.: Пер. с англ. – М: Изд. дом. “Вильямс”, 2003. – 1104 с.
4. Прокис Дж. Цифровая связь. Пер. с англ. / Под ред. Д.Д. Кловского. – М: Радио и связь, 2000. – 800 с.
5. Акулиничев Ю.П. Теория электрической связи: Учебное пособие. – СПб.: Издательство «Лань», 2010. – 240 с.: ил.
6. Уайндер С. Справочник по технологиям и средствам связи. Пер. с англ. – М.: Мир, 2000. – 429 с.
7. Гаранин М.В., Журавлев В.И., Кунегин С.В. Системы и сети передачи информации: Учеб. пособие для вузов. – М.: Радио и связь, 2001. – 336 с.6.
8. Беллами Дж. Цифровая телефония. – М.: Эко Трендз, 2004. – 640 с.
9. Хетагуров Я.А. Повышение надежности цифровых устройств методами избыточного кодирования / Я.А. Хетагуров, Ю.Н. Руднев – М.: Энергия, 1974. – 272 с.
10. Левин Б.Р. Теоретические основы статистической радиотехники. Книга вторая. Изд. второе, перераб. и доп. – М.: Сов. радио, 1975. – 392 с.
11. Цимбал В.П. Задачник по теории информации и кодированию. – Киев: Вища школа, 1976. – 276 с.
12. Борисов Ю.П., Пенин П.И. Основы многоканальной передачи информации. – М.: Связь, 1967.

Дополнительная литература

1. Акулиничев Ю.П., Дроздова В.И. Сборник задач по теории информации. – Томск: ТГУ, 1976. – 146 с.
2. 6. Кловский Д.Д., Шилкин В.А. Теория электрической связи. Сб. задач и упражнений. Учебное пособие для вузов. – М.: Радио и связь, 1990. – 280 с.
3. Лосев Ю. И. , Плотников Н.Д. Основы теории передачи данных. Сборник задач. – Киев: Вища школа, 1977. – 160 с.
4. 8. Харкевич А.А. Борьба с помехами. – М.: Наука, 1965.

5. Клюев Л.Л. Теория электрической связи / Л.Л. Клюев.– Минск : Дизайн ПРО, 1998. – 336 с.
6. Ипатов В. Широкополосные системы и кодовое разделение сигналов. Принципы и приложения. Москва: Техносфера, 2007. - 488с.
7. Радиотехнические системы передачи информации. Учебное пособие для вузов/ В.А.Борисов и др. Под ред. В.В. Калмыкова. – М.: Радио и связь, 1990. – 304 с.
8. 19. Крук Б.И., Попантонопуло В.Н., Шувалов В.П. Телекоммуникационные системы и сети: Учебное пособие. В 3 томах. Том 1 – Современные технологии/ Под ред. проф. В.П. Шувалова. – Изд. 3-е, испр. и доп. – М.: Горячая линия –Телеком 2004. – 646 с.: ил.
9. Пенин П.И., Филиппов М.И. Радиотехнические системы передачи информации. Учебное пособие для вузов. – М.: Радио и связь, 1984. – 256 с.
10. Левин Л.С., Плоткин В.А. Цифровые системы передачи информации. – М.: Радио и связь, 1982. – 216 с.
11. Передача дискретных сообщений. Под. ред. В.П. Шувалова.– М.: Радио и связь, 1990.
12. Баева Н.Н., Бобровская И.К., Брескин В.А., Федорова Е.Л. Многоканальная электросвязь и РРЛ. – М.: Радио и связь, 1984.
13. Системы связи и радиорелейные линии. Под ред. Н.И. Калашникова. – М.: Связь, 1977.
14. Горелов Г.В., Фомин А.Ф., Волков А.А., Котов В.К. Теория передачи сигналов на железнодорожном транспорте: Учебник для вузов ж.-д. транспорта. – М.: Транспорт, 2001. – 415 с.
15. Норенков И.П., Трудоношин В.А. Телекоммуникационные технологии и сети. 2-е изд., испр. и доп. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2000. – 248 с.
16. Ратынский М.В. Основы сотовой связи/ Под ред. Д.Б. Зимина, 2-е изд. - М.: Радио и связь, 2000. - 248 с.
17. Ирвин Дж., Харль Д. Передача данных в сетях: инженерный подход. Пер. с англ. - Спб.: БХВ-Петербург, 2003. - 448 с.
18. Иванов В.И., Гордиенко В.Н., Попов Г.Н. и др. Цифровые и аналоговые системы передачи. Под. ред. В.И. Иванова. – 2-е изд. – М.: Горячая линия-Телеком, 2003. – 232 с.
19. Многоканальная связь. Под. ред. И.А. Аболица. – М.: Связь, 1971.

ПРИЛОЖЕНИЯ

Приложение 1

Коэффициенты (в восьмеричной форме) для некоторых производящих полиномов циклических кодов БЧХ [2,8]

n	k	$q_и$	$g(x)$	n	k	$q_и$	$g(x)$
7	4	1	13; 15	127	120	1	211; 217;
15	11	1	23; 31		113	2	235
	7	2	721		106	3	41567
	5	3	2467		99	4	11554743
31	26	1	45; 75; 67		92	5	3447023271
	21	2	3551; 3557				624730022
	16	3	107657		85	6	327
	11	5	5423325				130704476
	6	7	313365047		78	7	322273
63	57	1	103; 147;				2623000216
	51	2	155				6130115
	45	3	12471;	255	247	1	435; 551
	39	4	11625		239	2	267543
	36	5	1701317		231	3	156720665
	30	6	166623567		223	4	756266
			1033500423				41375
	24	7	1574641		215	5	23157564
			65547				726421
	18	10	17323260	511	502	1	1021; 1131
			404441	1023	1013	1	2011; 2415
	16	11	136302651	2047	2036	1	4005; 4445
			2351725				
	10	13	633114136				
			7235453				
	7	15	472622305				
			527250155				
			5231045543				
			503271737				

Примечание: длинные последовательности записаны в две строки.

Пример: 721→111010001→ $g(x)=x^8+x^7+x^6+x^4+1$.

Приложение 2

Таблица значений вспомогательной функции $h(p)= -p\log_2 p$

p	$h(p)$	p	$h(p)$	p	$h(p)$	p	$h(p)$
0,00	0,0000	0,26	0,5053	0,51	0,4954	0,76	0,3009
0,01	0,0664	0,27	0,5100	0,52	0,4906	0,77	0,2903
0,02	0,1129	0,28	0,5142	0,53	0,4854	0,78	0,2796
0,03	0,1518	0,29	0,5179	0,54	0,4800	0,79	0,2687
0,04	0,1858	0,30	0,5211	0,55	0,4744	0,80	0,2575
0,05	0,2161	0,31	0,5238	0,56	0,4684	0,81	0,2462
0,06	0,2435	0,32	0,5260	0,57	0,4623	0,82	0,2348
0,07	0,2686	0,33	0,5278	0,58	0,4558	0,83	0,2231
0,08	0,2915	0,34	0,5292	0,59	0,4491	0,84	0,2113
0,09	0,3127	0,35	0,5301	0,60	0,4422	0,85	0,1993
0,10	0,3322	0,36	0,5306	0,61	0,4350	0,86	0,1871
0,11	0,3503	0,37	0,5307	0,62	0,4276	0,87	0,1748
0,12	0,3671	0,38	0,5305	0,63	0,4199	0,88	0,1623
0,13	0,3826	0,39	0,5298	0,64	0,4121	0,89	0,1496
0,14	0,3971	0,40	0,5288	0,65	0,4040	0,90	0,1368
0,15	0,4105	0,41	0,5274	0,66	0,3956	0,91	0,1238
0,16	0,4230	0,42	0,5256	0,67	0,3871	0,92	0,1107
0,1	0,434						

7	6	0,43	0,5236	0,68	0,3783	0,93	0,0974
0,18	0,4453	0,44	0,5211	0,69	0,3694	0,94	0,0839
0,19	0,4552	0,45	0,5184	0,70	0,3602	0,95	0,0703
0,20	0,4644						
		0,46	0,5153	0,71	0,3508	0,96	0,0565
0,21	0,4728	0,47	0,5120	0,72	0,3412	0,97	0,0426
0,22	0,4806	0,48	0,5083	0,73	0,3314	0,98	0,0286
0,23	0,4877	0,49	0,5043	0,74	0,3215	0,99	0,0144
0,24	0,4941	0,50	0,5000	0,75	0,3113	1,00	0,0000
0,25	0,5000						

Приложение 3

Таблица значений интеграла вероятности $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{z}\exp\left(-\frac{x^2}{2}\right)dx$

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

$$1-\Phi(z)=\Phi(-z)\approx\frac{1}{\sqrt{2\pi}z}\exp\left(-\frac{z^2}{2}\right) \quad \text{при } z>3$$

$$\Phi(z)\approx 0,5+\frac{z}{\sqrt{2\pi}} \quad \text{при } |z|<<1.$$

Приложение 4

Параметры стандартного устройства частотного уплотнения аналоговых телефонных сигналов с полосой 300-3400 Гц
по ГОСТ 21655-87

Номер ступени	Частоты поднесущих, кГц	Полосы сигналов, кГц	
		канальных	группового
1	64, 68, 72, … , 104, 108	60-64, 64-68, 68-72,…, 100-104, 104-108	60-108
2	420, 468, 516, 564, 612	312-360, 360-408, 408-456, 456-504, 504-552	312-552
3	1364, 1612, 1860, 2108, 2356	812-1052, 1060-1300, 1308-1548, 1556-1796, 1804-2044	812-2044

Приложение 5

Плезиохронная цифровая иерархия (европейский вариант)
по ГОСТ 26886-86 в соответствии с рекомендацией G.703.

Уровень цифровой иерархии	*V*, кбит/с	*N*	*K*	Тип аппаратуры
0 (основной) – DSO	64	1	–	
1 (первичный) – E1	2 048	30	30	ИКМ-30
2 (вторичный) – E2	8 448	120	4	ИКМ-120
3 (третичный) – E3	34 368	480	4	ИКМ-480
4 (четверичный) – E4	139 264	1 920	4	ИКМ-1920
5 (пятиричный) – E5	564 992	7 680	4	

(*V* – скорость передачи на выходе;
N – количество основных цифровых каналов;
K – коэффициент мультиплексирования входов)

Приложение 6

Синхронная цифровая иерархия

Уровень цифровой иерархии	Скорость передачи, Мбит/с	Коэффициент муль-типлексирования входов
STM-0	51,84	–
STM-1	155,52	3
STM-4	622,08	4
STM-16	2 488,32	4
STM-64	9 953,28	4
STM-256	39 813,12	4

УСЛОВНЫЕ ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ.

АИМ — амплитудно-импульсная модуляция.
АМ — амплитудная модуляция.
АМ ОБП — амплитудная модуляция с одной боковой полосой.
АТС — автоматическая телефонная станция.
АЦП — аналого-цифровое преобразование.
ВИМ — время-импульсная модуляция.
ОФМ — относительная фазовая модуляция.
СПИ — системы передачи информации.
ФМ — фазовая модуляция.
ЦАП — цифроаналоговое преобразование.
ЧМ — частотная модуляция.
ШИМ — широтно-импульсная модуляция.

ГЛОССАРИЙ.

Алфавит – перечень всех возможных значений символа $x_1,x_2,\ldots,x_m,$. образующих полную группу несовместных событий.

АИМ – амплитудно-импульсная модуляция.

Аналоговый сигнал – непрерывный сигнал непрерывного времени.

АМ – амплитудная модуляция.

АМ ОБП – амплитудная модуляция с одной боковой полосой.

Асинхронная СПИ – передача первого символа кодовой комбинации начинается в произвольный момент времени.

ATM (Asynchronous Transfer Mode) – способ асинхронной передачи разнородной информации по асинхронным сетям связи.

АТС – автоматическая телефонная станция.

АЦП – аналого-цифровое преобразование.

Бит – двоичная цифра (bit = binary digit).

Блокировка вызова – потеря вызова, отказ в предоставлении ресурсов сети.

Групповой сигнал – сумма канальных сигналов, передаваемых по линии связи.

Дейтаграмма (datagram) – это самостоятельный пакет, в заголовке которого содержится полная информация, необходимая для его передачи от источника к получателю и доставляемый адресату независимо от других пакетов.

ВИМ – время-импульсная модуляция.

ВРК (Time Division Multiple Access – TDMA). –временное разделение каналов(сигналов). Принцип разделения передаваемых сигналов, базирующийся на периодическом выделении каждому сигналу определенного промежутка времени.

Вторичная сеть – это структура, организующая обмен сообщениями определенного вида с использованием арендованных ресурсов первичной сети и соответствующего оконечного оборудования.

Демодуляция – выделение из меняющихся во времени параметров высокочастотного модулированного сигнала исходного низкочастотного сигнала.

Дискретизация – замена непрерывной функции ее дискретными значениями, отсчитываемыми через определенные интервалы времени.

Дискретный сигнал – дискретный сигнал дискретного времени.

Информация –это сведения, знания, новости, идеи и т.п. независимо от формы их представления

Источник информации – это объект, текущее состояние которого *X* интересует получателя.

КАМ-М – М-позиционная квадратурно-амплитудная манипуляция, предусматривает совместное изменение амплитуд и фаз передаваемых им-

пульсов-символов.

Канал – это совокупность технических средств, обеспечивающих передачу сообщений определенного вида от оконечного оборудования одного пользователя к оконечному оборудованию другого

Канальное (помехоустойчивое) кодирование – формирование цифрового сигнала таким образом, чтобы в процессе приема появилась возможность обнаружить или исправить большую часть ошибок, возникших в принятом сигнале из-за влияния помех.

Квантование – замена отсчетных значений сигнала разрешенными значениями.

Когерентная СПИ – ожидаемые значения начальных фаз всех принимаемых импульсов (вплоть до окончания сеанса связи) известны заранее.

Кодирование источника – сокращение избыточности, содержащейся в сигнале источника.

Код Хэмминга –систематический код с кодовым расстоянием $d_0=3$.

Кодовое расстояние – минимальное для данного кода расстояние Хэмминга.

КРК (Code Division Multiple Access – CDMA).– *кодовое разделение каналов*. Используют также термины: кодовое мультиплексирование, многостанционный доступ с кодовым разделением каналов (МДКР).Сигналы передают в одно время и в одной полосе частот. Разделение сигналов происходит по коду (форме).

КК (коммутация каналов) – это организация по требованию физического соединения любой из *n* входных линий с любой другой из *m* выходных линий и удержание этого соединения на время передачи сообщения.

Коммутатор – коммутационная схема (матрица) с n входами и m выходами, в которой соединение входа с выходом устанавливается с помощью одного коммутационного элемента.

Коммутируемый канал – создается лишь на время передачи сообщения либо его части (пакета)

Коммутация пакетов — это организация по требованию физического соединения любой из *n* входных линий с любой другой из *m* выходных линий, но *сохраняется это соединение лишь на короткое время – на время передачи одного пакета*.

Коммутация сообщений – разновидность коммутации, при которой сообщения передаются целиком.

Контейнер – используется для размещения цифровых потоков от различных источников перед передачей их в сетях с синхронной цифровой иерархией.

Линия – это среда, в которой распространяются сигналы из пункта передачи в пункт приема

Математическая модель – это упрощенное количественное описание объекта или явления, пригодное для решения конкретной задачи.

М*ежсимвольная интерференция* – наложение соседних импульсов друг на друга и их взаимное искажение.

М*ногоканальная* система – по одной линии связи осуществляется *одновременная передача* n *сообщений*

Многостанционный доступ (МД) – метод, позволяющий *n* станциям радиосвязи (возможно, разбросанным на большой территории) *совместно использовать ресурсы одного ретранслятора*.

Модуляция – процесс, заключающийся в изменении во времени параметров высокочастотного гармонического колебания (амплитуды, частоты, фазы) по закону изменения во времени низкочастотного сигнала, подлежащего передаче.

Некоммутируемый (арендованный, постоянный) канал – постоянно готов к передаче, так как не используется другими абонентами

Основание кода (объем алфавита) – число возможных значений символов источника. Основой цифровой канал –канал со скоростью передачи 64 Кбит/с.

ОФМ – относительная фазовая модуляция.

Пакет – часть сообщения, представленная в виде блока с заголовком, имеющего установленный формат (структуру данных) и ограниченную длину, и передаваемая по сети как единое целое.

Перемежение (interliving) – перемешивание символов, способ, позволяющий ошибки из пакетов более или менее равномерно распределять по кодовым комбинациям.

Первичная сеть – это линии и коммутационное оборудование, служащее для образования каналов со стандартными характеристиками, пригодных для передачи сообщений разных видов.

Плезиохронная цифровая иерархия (Plesyochronous Digital Hierarchy - PDH) – иерархия скоростей при объединении цифровых потоков на сетях без общей синхронизации.

Пропускная способность канала – это наибольшая возможная в этом канале скорость передачи информации (бит/символ или бит/с) при заданных ограничениях на значения ряда его физических параметров.

Разделимый код – код, в кодовых комбинациях которого элементы можно разделить на информационные и проверочные.

Разнесенный прием – прием одного сигнала по нескольким каналам, разнесенным по частоте, времени, в пространстве и т. д..

Расстояние Хэмминга – степень различия комбинаций кода. Определяется для любых двух комбинаций числом несовпадающих символов.

Регенерация – восстановление формы искаженного двоичного цифрового сигнала после передачи его на определенное расстояние.

Сеть сигнализации – это сеть, предназначенная для обслуживания первичной сети и используемая для передачи команд, управляющих работой центров коммутации каналов

Сеть электросвязи – это совокупность технических средств, обеспечивающая обмен информацией между множеством пользователей, разнесенных в пространстве.

Сверточные коды – непрерывные коды.

Сигнал – материальный переносчик информации с помощью физической величины, изменение одного или нескольких параметров которой отображает сообщение.

Синхронизация – процесс установления и поддержания определенных временных соотношений между двумя или несколькими процессами.

Синхронной СПИ – передача каждого символа (импульса) начинается в тактовый момент времени

Синхронная цифровая иерархия(SDH) – иерархия скоростей при объединении цифровых потоков, разрешенная к применению на сетях общей синхронизации.

Скорость передачи информации –с реднее число бит, передаваемых одним символом или в секунду.

Скремблирование – сложение цифрового сигнала по модулю два с псевдослучайной последовательностью.

СПИ – системы передачи информации.

ФМ – фазовая модуляция.

Циклический код – разновидность линейного блочного кода, основное свойство которого заключается в следующем: комбинация, получаемая путем циклической перестановки элементов разрешенной кодовой комбинации, также является разрешенной.

Цифровой сигнал – последовательность, состоящая изчередующихся случайным образом импульсов двух типов, один из которых обозначает 1, а второй 0.

ЦАП – цифроаналоговое преобразование

ЧМ – частотная модуляция.

ЧРК – частотное разделение каналов(сигналов) –принцип разделения передаваемых сигналов, базирующийся на отведении каждому сигналу своей полосы частот в заданном частотном диапазоне.

ШИМ – широтно-импульсная модуляция.

Энтропия –мера неопределенности источника сообщений, производительность источника(бит/символ или бит/с).

OFDM (Orthogonal Frequency-Division Multiplexing) – модуляция ортогональных несущих, базовая технология беспроводных широкополосных систем связи.