

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЦИФРОВОЙ РАДИОСВЯЗИ

*Допущено Министерством образования Республики Беларусь
в качестве учебного пособия для студентов
учреждений высшего образования
по радиотехническим специальностям*

Минск БГУИР 2012

УДК 621.396.2(075.8)
ББК 32.884.1я73
Т33

А в т о р ы:

Н. И. Листопад, В. М. Козел, В. В. Дубровский, К. Л. Горбачев, К. А. Ковалев

Р е ц е н з е н т ы:

кафедра программного обеспечения вычислительной техники
и автоматизированных систем
Белорусского национального технического университета
(протокол №2 от 22.09.2011);

заведующий кафедрой информационных систем и технологий
учреждения образования «Белорусский государственный
технологический университет»,
доктор технических наук, профессор П. П. Урбанович

Т33 **Теоретические основы цифровой радиосвязи : учеб. пособие /**
Н. И. Листопад [и др.]. — Минск : БГУИР, 2012. — 330 с.
ISBN 978-985-488-849-1.

В учебном пособии рассмотрены общетеоретические вопросы построения систем цифровой радиосвязи, включая вопросы теории информации; дискретизации и квантования аналоговых сигналов; кодирования источника; помехоустойчивого кодирования и перемежения данных; теории криптографии; полосовой и низкочастотной модуляции; синхронизации и множественного доступа к каналам связи. Приведены основные принципы построения различных элементов систем цифровой радиосвязи.

Пособие предназначено для студентов радиотехнических и телекоммуникационных специальностей, а также может быть полезно для магистрантов, аспирантов и научных сотрудников, занимающихся вопросами цифровой связи.

УДК 621.396.2(075.8)
ББК 32.884.1я73

ISBN 978-985-488-849-1

© УО «Белорусский государственный
университет информатики
и радиоэлектроники», 2012

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	7
1. ОСНОВНЫЕ ПОКАЗАТЕЛИ КАЧЕСТВА КАНАЛОВ ЦИФРОВОЙ РАДИОСВЯЗИ.....	8
2. МОДЕЛИ КАНАЛОВ ЦИФРОВОЙ РАДИОСВЯЗИ.....	12
2.1. Классификация каналов связи по характеру среды распространения электромагнитных волн. Характеристики каналов связи.....	12
2.2. Физические процессы, обуславливающие возникновение искажений сигнала в радиоканале.....	13
2.3. Паразитные эффекты при приёме и обработке сигналов в системах цифровой радиосвязи.....	16
2.3.1. Межсимвольная интерференция.....	16
2.3.2. Джиттер.....	16
2.3.3. Глазковая диаграмма.....	18
2.4. Математическая модель канала с замираниями.....	19
2.5. Характеристики канала с замираниями.....	21
2.6. Примеры математических моделей каналов.....	23
2.6.1. Неизбирательные каналы (НИК).....	23
2.6.2. Частотно-избирательный канал (ЧИК).....	23
2.6.3. Каналы с временной избирательностью (ВИК).....	25
2.6.4. Каналы с частотной и временной избирательностью (ЧВИ).....	26
2.7. Характеристики многолучёвого канала в условиях прямой видимости.....	27
2.8. Модели ошибок в каналах цифровой радиосвязи.....	29
3. ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ.....	31
3.1. Мера количества информации в дискретном сообщении.....	31
3.2. Энтропия дискретного источника с независимым выбором сообщений.....	32
3.3. Энтропия дискретного источника с зависимыми сообщениями.....	33
3.4. Избыточность источника.....	34
3.5. Производительность источника.....	35
3.6. Совместная энтропия двух источников.....	36
3.7. Скорость передачи и пропускная способность канала связи.....	37
3.8. Энтропия непрерывного источника случайной величины и её свойства.....	41
3.9. Пропускная способность непрерывного канала.....	43
3.10. Предельные информационные характеристики канала связи.....	46
3.11. Расширение границы Шеннона (BLAST – Bell Labs Layered Space Time).....	48
4. ДИСКРЕТИЗАЦИЯ И КВАНТОВАНИЕ СИГНАЛОВ.....	52
4.1. Дискретизация во времени.....	52
4.1.1. Вводные замечания.....	52
4.1.2. Представление аналоговых сигналов в дискретной форме.....	52
4.1.3. Теорема В. А. Котельникова.....	53
4.1.4. Дискретизация полосовых сигналов.....	56
4.1.5. Дискретизация сигналов в реальных системах.....	58
4.1.6. Теорема отсчётов в частотной области.....	60
4.1.7. Выводы и практические рекомендации.....	60
4.2. Квантование по уровню.....	61
4.2.1. Место квантования сигнала в процедуре преобразования аналоговых сигналов в цифровые.....	61
4.2.2. Величина шума квантования.....	66
4.2.3. Техническая реализация схем квантования по уровню.....	66
4.2.4. Шумовые эффекты в цифровых системах.....	68
4.2.5. Неравномерное квантование сигналов.....	69
4.2.6. Дифференциальное квантование.....	72

5. КОДИРОВАНИЕ ИСТОЧНИКА	74
5.1. Понятие префиксного кода.....	74
5.2. Неравенство Крафта.....	75
5.3. Алгоритм Хаффмана	76
5.4. Алгоритм Лемпеля–Зива–Велча (LZW).....	77
5.5. Алгоритм голосового кодирования HVXC	80
6. ОСНОВЫ КРИПТОГРАФИИ	83
6.1. Симметричные криптосистемы.....	83
6.2. Криптосистема с открытым ключом	86
6.3. Криптографический протокол.....	88
6.4. Управление ключами	90
7. ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ В СИСТЕМАХ ЦИФРОВОЙ РАДИОСВЯЗИ	94
7.1. Необходимость канального кодирования	94
7.2. Общие принципы помехоустойчивого кодирования и декодирования	95
7.3. Классификация помехоустойчивых кодов.....	97
7.4. Блочное линейное кодирование	98
7.5. Древовидные коды	99
7.6. Кодирование методом укрупнения алфавита	101
7.7. Определение алгебраических структур.....	101
7.8. Циклические коды.....	103
7.9. Наиболее распространенные циклические коды.....	105
7.10. Мажоритарно декодируемые коды	107
7.11. Декодирование циклических кодов.....	107
7.12. Сверточные коды.....	109
7.13. Основные принципы каскадного кодирования	112
7.14. Сверточные турбокоды	113
7.15. Блочные турбокоды	116
8. ПЕРЕМЕЖЕНИЕ ДАННЫХ	118
8.1. Общие сведения о перемежении данных	118
8.2. Кодирование информации в каналах связи на базе стандарта TETRA.....	119
8.3. Блочное перемежение	121
8.4. Перемежение по N блокам	123
9. ФОРМИРОВАНИЕ ЦИФРОВЫХ ПОТОКОВ	124
9.1. Вводные замечания	124
9.2. Цифровые иерархии	125
9.2.1. Плезиохронная цифровая иерархия.....	125
9.2.2. Структура цифрового потока E1	125
9.2.3. Мультиплексирование цифровых потоков	132
9.2.4. Синхронная цифровая иерархия	139
9.2.5. Методы асинхронной передачи	143
10. МОДУЛЯЦИЯ В ОСНОВНОЙ (ИНФОРМАЦИОННОЙ) ПОЛОСЕ ЧАСТОТ (BASEBAND MODULATION)	147
10.1. Основные принципы модуляции	147
10.2. Коды линии	149
10.2.1. Код без возвращения к нулю (NRZ)	149
10.2.2. Манчестерский код (Manchester code).....	150
10.2.3. Код с возвращением к нулю (RZ)	151
10.2.4. Код с чередующейся инверсией (AMI)	152
10.2.5. Биполярный код высокой плотности порядка n (HDB-n)	153
11. ПОЛОСОВАЯ МОДУЛЯЦИЯ/ДЕМОДУЛЯЦИЯ (модуляция/демодуляция радиосигналов).....	155

11.1. Основные принципы полосовой модуляции	155
11.2. Краткий обзор методов цифровой (дискретной) модуляции	156
11.3. Представление радиосигнала в комплексном виде. Аналитический сигнал	160
11.3.1. Комплексные числа	160
11.3.2. Комплексное представление радиосигнала	162
11.3.3. Замечания по поводу отрицательных частот	164
11.3.4. Заключительные замечания	166
11.4. Прямой и обратный переход между радиосигналом и его комплексным представлением	167
11.5. Квазиоптимальное принятие решений при приеме сигналов с дискретной модуляцией	170
11.5.1. Общая методология синтеза решающих правил при квазиоптимальном приеме сигналов с дискретной модуляцией	170
11.5.2. Синтез функции правдоподобия для канала связи с аддитивным белым гауссовским шумом	173
11.5.2.1. Белый гауссовский шум	173
11.5.2.2. Комплексное представление полосовых случайных процессов	177
11.5.2.3. Функция правдоподобия для канала связи с аддитивным белым гауссовским шумом	180
11.5.3. Обобщенные решающие правила при квазиоптимальном приеме сигналов с дискретной модуляцией	182
11.5.3.1. Демодуляция информационного символа в целом	182
11.5.3.2. Поэлементная демодуляция информационного символа	184
11.6. Помехоустойчивость процесса демодуляции	188
11.6.1. Общее соотношение для вероятности ошибочного приема	188
11.6.2. Случай ортогональных сигналов	193
11.6.3. Переход от вероятности ошибки на символ модуляции к вероятности ошибки на элемент (бит) модуляции	194
11.7. Основные методы дискретной модуляции	199
11.7.1. Фазовая манипуляция (PSK)	199
11.7.1.1. Бинарная фазовая манипуляция (BPSK)	199
11.7.1.2. М-ичная фазовая манипуляция (M-PSK)	208
11.7.1.3. Квадратурная фазовая манипуляция и квадратурная фазовая манипуляция со сдвигом (QPSK и OQPSK)	244
11.7.1.4. Явление обратной работы.	
Относительная (дифференциальная) фазовая манипуляция (DPSK, DQPSK)	225
11.7.2. Частотная манипуляция (FSK)	233
11.7.2.1. Бинарная частотная манипуляция (BFSK)	233
11.7.2.2. М-позиционная (многопозиционная) частотная манипуляция (M-FSK)	240
11.7.3. Непрерывная фазовая модуляция (CPM)	
11.7.3.1. Общее представление	244
11.7.3.2. Наиболее часто используемые виды формирующего импульса частотной манипуляции	246
11.7.3.3. Общие принципы демодуляции сигналов с CPM	248
11.7.3.4. Минимальная манипуляция (MSK)	252
11.7.3.5. Гауссовская минимальная манипуляция (GMSK)	256
11.7.4. Квадратурная амплитудная модуляция (Quadrature Amplitude Modulation QAM)	257
11.8. Спектральные характеристики излучения при дискретной полосовой модуляции	262

11.8.1. Основные принципы определения спектральных характеристик сигналов с дискретной полосовой модуляцией	262
11.8.2. Спектральные характеристики сигналов с линейной модуляцией	265
11.8.3. Основные виды формирующих импульсов	268
11.9. Сигнально-кодовые конструкции (Trellis Coded Modulation)	272
11.9.1. Место и роль сигнально-кодовых конструкций в цифровой радиосвязи	272
11.9.2. Согласование канала кодом Грея	277
11.9.3. Решетчатое кодирование (Trellis Coded Modulation)	279
11.9.3.1. Общие понятия о решетчато-кодовой модуляции	279
11.9.3.2. Разбиение исходного ансамбля сигналов	281
11.9.3.3. Конструирование сигнально-кодовой конструкции	282
11.9.3.4. Сравнение помехоустойчивости кодированной и некодированной схем	291
11.10. Заключение	292
12. СИНХРОНИЗАЦИЯ	290
12.1. Общая задача синхронизации при приеме сигналов с дискретной модуляцией	290
12.2. Методы восстановления несущего колебания при полосовой демодуляции	292
12.2.1. Максимально правдоподобная оценка фазы	292
12.2.2. Методы синхронизации по несущему сигналу с обратной связью по решению	308
12.2.3. Методы синхронизации по несущему колебанию, не управляемые решением	296
12.2.3.1. Синхронизация при приеме сигнала с $\pm\pi/2$ -BPSK	297
12.2.3.2. Синхронизация при приеме сигнала с M-PSK	299
12.2.3.3. Синхронизация при приеме сигнала с QPSK и OQPSK	302
12.2.3.4. Синхронизация при приеме сигнала с QAM	304
12.2.3.5. Заключительные замечания	306
12.3. Методы восстановления тактовой синхронизации	307
12.3.1. Тактовая синхронизация управляемая решением	308
12.3.2. Тактовая синхронизация, не управляемая решением	309
12.4. Кадровая синхронизация	312
12.5. Заключительные замечания	313
13. УПЛОТНЕНИЕ И МНОЖЕСТВЕННЫЙ ДОСТУП	314
13.1. Общие принципы осуществления множественного доступа к каналам связи	314
13.2. Множественный доступ с частотным разделением (FDMA – Frequency Division Multiple Access)	317
13.3. Множественный доступ с временным разделением (TDMA – Time Division Multiple Access)	318
13.4. Метод множественного доступа с кодовым разделением (CDMA – Code Division Multiple Access)	319
13.5. Метод доступа с пространственным разнесением пользователей (SDMA – Space Division Multiple Access)	321
13.6. Метод доступа, основанный на использовании поляризационных свойств электромагнитного излучения (PDMA – Polarization Division Multiple Access)	321
13.7. Мультиплексирование с ортогональным частотным разделением (OFDM – Orthogonal Frequency Division Multiplexing)	322
ЗАКЛЮЧЕНИЕ	235
ЛИТЕРАТУРА	326

ВВЕДЕНИЕ

Цифровая радиосвязь в сравнении с аналоговой находит все большее распространение в силу очевидных преимуществ цифровых систем перед аналоговыми. Речь идет в первую очередь о таких преимуществах, как более высокая достоверность и качество приема/передачи информации.

Простейшая структурная схема системы цифровой связи состоит из передатчика, канала связи и приемника. К передатчику подключается источник информации, а к приемнику – получатель. При этом передатчик выполняет функции преобразования источника сигнала к виду, необходимому для его передачи по каналу связи: дискретизацию и квантование информационного сообщения, кодирование источника, шифрование, помехоустойчивое кодирование, модуляцию в информационной полосе частот, полосовую модуляцию, множественный доступ к каналу связи. С помощью приемника осуществляется извлечение информации, переданной по каналу связи, путем преобразований, обратных тем, что были произведены в передатчике.

В главе 1 представлены основные показатели качества каналов цифровой радиосвязи. В основном это параметры, используемые для анализа характеристик бинарного канала связи согласно рекомендациям ИТУ-Т G.821, G.826 и M.2100.

В главе 2 описаны модели каналов, используемые для передачи цифровой информации.

Основы теории информации изложены в главе 3. Описываются как непосредственно основы теории информации, так и информационные характеристики и параметры каналов связи.

Глава 4 посвящена вопросам преобразования аналогового сообщения в цифровую последовательность путем дискретизации и квантования.

В главе 5 представлены основные алгоритмы, используемые при кодировании источника сигнала.

Основы криптографии изложены в главе 6. Рассмотрены основные крипто-системы и протоколы.

В главе 7 излагаются теоретические основы помехоустойчивого кодирования, производится классификация и описание наиболее распространенных помехоустойчивых кодов.

Глава 8 посвящена такому фундаментальному компоненту помехоустойчивого кодирования, как перемежение. Рассмотрена суть операции перемежения, новые подходы построения перемежителей.

Основы формирования цифровых потоков изложены в главе 9. Представлена цифровая иерархия цифровых потоков данных, синхронные и асинхронные системы передачи.

Теоретические основы модуляции в информационной полосе частот изложены в главе 10. Рассмотрены основные коды линии связи.

Модуляция и демодуляция радиосигналов подробно рассмотрены в главе 11. Здесь же представлены основные сигнально-кодовые конструкции, используемые для оптимального приема сигналов.

Вопросы синхронизации при приеме сигналов с дискретной модуляцией изложены в главе 12.

Общие принципы и основы практической реализации систем уплотнения и множественного доступа представлены в главе 13.

1. ОСНОВНЫЕ ПОКАЗАТЕЛИ КАЧЕСТВА КАНАЛОВ ЦИФРОВОЙ РАДИОСВЯЗИ

Качество передачи информации для цифровых линий связи характеризуется рядом параметров. В основном это параметры, используемые для анализа характеристик бинарного канала согласно рекомендациям ITU-T G.821, G.826 и M.2100.

AS (availability seconds, измеряется в секундах, с) – время готовности канала; вторичный параметр, равный разности между общей длительностью теста и временем неготовности канала. AS (измеряется в %) – относительное время готовности канала; параметр, характеризующий готовность канала, выраженный в процентах. В отличие от AS (с), AS (%) является первичным параметром и входит в число основных параметров рекомендации G.821. Его можно интерпретировать как вероятностную меру качества канала, предоставляемого пользователю.

BBE (background block error – блок с фоновой ошибкой) – блок с ошибками, не являющийся частью SES (определение SES см. ниже), применяется при анализе ошибок по блокам. Является важным параметром, вошедшим в рекомендацию ITU-T G.826.

BIT или BIT ERR (bit errors – число ошибочных бит) – параметр, используемый при анализе канала на наличие битовых ошибок, является числителем в выражении для расчета BER (определение BER см. ниже). Битовые ошибки подсчитываются только во время пребывания канала в состоянии готовности.

EB (error block – число ошибочных блоков) – параметр, используемый при анализе канала на наличие блоковых ошибок, является числителем в выражении для расчета BLER. Блоковые ошибки подсчитываются только во время пребывания канала в состоянии готовности.

BLER – (block error rate – частота блоковых ошибок, коэффициент ошибок по блокам) – редко применяемый на практике параметр, равный отношению числа ошибочных блоков данных к общему числу переданных блоков. Ошибочным блоком считается блок, содержащий хотя бы один ошибочный бит. Обычно значение параметра BLER больше (хуже), чем параметра BER. Его целесообразно измерять только в тех сетях передачи данных, где информация передается блоками фиксированного размера, а параметр BLER является важной характеристикой канала с учетом кадровой структуры передачи.

Например, для сетей ATM принята кадровая структура передачи в виде кадров длиной 53 бита. Ошибочный кадр уничтожается (дискартируется). В этом случае можно рассматривать кадр ATM как блок длиной в 53 бита, а эквивалентом BLER будет параметр ошибки по кадрам CER (Cell Error Rate). В другом примере в качестве эквивалента блока может выступать сверхцикл ИКМ, а эквивалентом BLER будет ошибка по CRC.

BBER (background block error – rate – коэффициент ошибок по блокам с фоновыми ошибками) – отношение числа блоков с фоновыми ошибками к

общему количеству блоков в течение времени готовности канала, за исключением всех блоков в течение SES. Является важным параметром, вошедшим в рекомендацию ITU-T G.826.

BER или RATE (bit error rate – частота битовых ошибок, коэффициент ошибок по битам) – основной параметр в системах цифровой передачи, равный отношению числа битовых ошибок к общему числу бит, переданных за время проведения теста по каналу, находящемуся в состоянии готовности. При обнаружении десяти последовательных секундных интервалов, сильно пораженных ошибками (SES), анализатор переключается на подсчет времени неготовности канала. При этом измерение BER прерывается до восстановления работоспособности канала. Таким образом, управляемые проскальзывания, связанные с потерей одного или нескольких циклов информации, практически не влияют на значение BER.

Измерения параметра BER универсальны в том смысле, что не требуют наличия цикловой и сверхцикловой структуры в измеряемом потоке, однако требуют передачи специальной тестовой последовательности и могут быть проведены только в случае полного или частичного отключения цифрового канала от полезной нагрузки.

CLKSLIP или SLIP (clock slips – число тактовых проскальзываний) – параметр, характеризующийся числом синхронных управляемых проскальзываний, появившихся с момента начала теста. Проскальзыванием называется повторение или исключение группы символов в синхронной или плезиохронной последовательности двоичных символов в результате различия между скоростями считывания и записи в буферной памяти. Поскольку проскальзывание ведет к потере части информации, что в свою очередь ведет к потере цикловой синхронизации, на практике используются эластичные управляемые буферы с возможностью управления проскальзываниями. В этом случае проскальзывания называются управляемыми. В наибольшей степени параметр CLKSLIP связан с параметром неготовности канала (UAS). Сопоставление CLKSLIP и UAS позволяет выявить причину неготовности канала, в частности, связана ли она с нарушением синхронизации. Значение параметра CLKSLIP зависит от размера имитируемого прибором буфера, который может составлять от 1 бита до нескольких килобайт.

CRC ERR (CRC errors – число ошибок) – параметр ошибки, измеренный с использованием циклового избыточного кода (CRC), распространенный параметр определения ошибок реально работающего канала без его отключения и без передачи тестовой последовательности. Необходимым условием измерения параметра CRC является наличие механизма формирования кода в аппаратуре передачи. Встроенные средства самодиагностики большей части современных цифровых систем передачи используют именно этот механизм. Таким образом, при измерении параметра CRC можно не только оценить частоту ошибок, но и проверить работу системы самодиагностики.

При использовании CRC часто возникает вопрос о необходимости измерения одновременно с ним и параметра BER. Здесь необходимо учитывать две особенности применения CRC. Во-первых, каждая ошибка CRC не

обязательно связана с ошибкой одного бита информации. Несколько битовых ошибок в одном сверхцикле могут дать только одну ошибку CRC для блока. Таким образом, при использовании CRC можно говорить не об истинном уровне ошибок в канале, а только об оценке их величины. Тем не менее CRC является удобным методом контроля ошибок при проведении сервисного наблюдения за работающим каналом, когда практически невозможно измерить реальные параметры битовых ошибок.

DGRM (degraded minutes – количество минут деградации качества) – несколько временных интервалов, продолжительностью 60 с каждый, когда канал находится в состоянии готовности, но BER менее заданной величины.

EFS (error free seconds – время, свободное от ошибок, с) – отражает время, в течение которого сигнал был правильно синхронизирован, а ошибки отсутствовали, т. е. общее время пребывания канала в состоянии безошибочной работы.

EFS (error free seconds – процент времени, свободного от ошибок) – то же, что и предыдущий параметр, только выраженный в процентах по отношению к общему времени с момента начала тестирования.

ES (errors seconds – длительность поражения сигнала ошибками, с) – количество секунд с ошибками; параметр показывает интервал времени поражения всеми видами ошибок в канале, находящемся в состоянии готовности. ES связан с другими параметрами соотношением: $AS = ES + EFS$.

ES (errors seconds – процент поражения сигнала ошибками) – параметр связан с EFS (%) соотношением: $ES (\%) + EFS (\%) = AS (\%)$.

LOSS (loss of signal seconds – длительность потери сигнала, с) – параметр характеризует интервал времени, в течение которого сигнал был потерян.

PATL (pattern loss – количество потерь тестовой последовательности) – параметр, характеризующийся числом потерь тестовой последовательности, появившихся с момента начала теста.

PATLS (pattern loss seconds – продолжительность времени потери тестовой последовательности, с) – общее время потери тестовой последовательности с момента начала теста.

SES (severally errors seconds – продолжительность многократного поражения ошибками, количество секунд, пораженных ошибками, с) – интервал времени, пораженный ошибками несколько раз. Подсчет SES производится только во время готовности канала. Из определения видно, что SES – составная часть параметра ES.

Вторая интерпретация параметра SES связана с измерениями по блоковым ошибкам, тогда SES определяется как односекундный интервал времени, содержащий более 30 % блоков с ошибками.

SES (severally errors seconds – относительная продолжительность многократного поражения ошибками) – тот же параметр, но выраженный в процентах.

SESR (severally error seconds rate – коэффициент ошибок по секундам, пораженным ошибками, в %) – параметр, практически равный SES.

SLIP или CLKSLIP (clock slips – число тактовых проскальзываний) – параметр, характеризующийся числом синхронных управляемых проскальзываний, появившихся с момента начала теста. В наибольшей степени этот параметр связан с параметром неготовности канала – UAS.

Для линий аналоговой связи основным критерием качества является отношение средней мощности сигнала к средней мощности шума (S/N или SNR). В цифровой связи в качестве критерия качества чаще используется отношение E_b/N_0 , где E_b – это энергия бита. Имеется ввиду энергия бита передаваемой информации, которую можно описать как мощность сигнала S , умноженную на время передачи бита T_b . N_0 – спектральная плотность мощности шума, ее можно выразить как мощность шума N , деленную на эквивалентную шумовую полосу W . Поскольку время передачи бита и скорость передачи битов R_b взаимно обратны, T_b можно заменить на $1/R_b$:

$$\frac{E_b}{N_0} = \frac{ST_b}{N/W} = \frac{S/R_b}{N/W} = \frac{S}{N} \frac{W}{R_b}. \quad (1.1)$$

Одной из важнейших метрик качества в системах цифровой связи является зависимость вероятности появления ошибочного бита P_b от E_b/N_0 . На рис. 1.1 показан «водопадоподобный» вид большинства подобных зависимостей.

Безразмерное отношение E_b/N_0 – это стандартная качественная мера производительности систем цифровой связи. Следовательно, необходимое отношение E_b/N_0 можно рассматривать как метрику, позволяющую сравнивать качество различных систем: чем меньше требуемое отношение E_b/N_0 , тем эффективнее процесс детектирования при данной вероятности ошибки. Проведем обоснование выбора данной метрики по сравнению с аналоговыми сигналами.

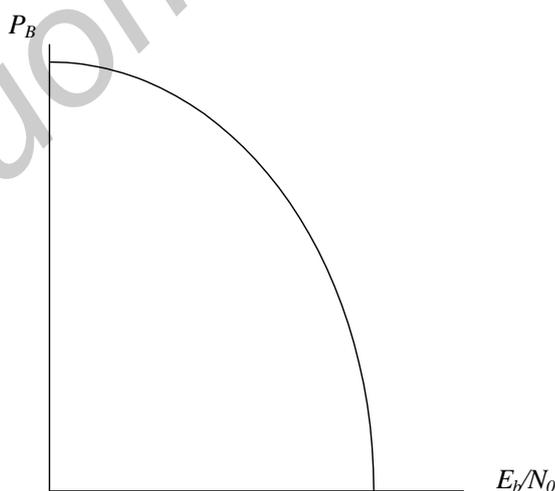


Рис. 1.1. Типовая зависимость вероятности появления ошибочного бита P_b от E_b/N_0

2. МОДЕЛИ КАНАЛОВ ЦИФРОВОЙ РАДИОСВЯЗИ

2.1. Классификация каналов связи по характеру среды распространения электромагнитных волн.

Характеристики каналов связи

По характеру среды распространения радиоволн можно выделить несколько типов каналов:

- с искусственными направляющими линиями;
- без искусственных направляющих линий.

Каждый из этих каналов имеет свои особенности и описывается математическими моделями, которые во многом зависят от рабочей частоты СЦРС.

Первая математическая модель каналов связи, используемая при проектировании систем связи и для их анализа, – двоично-симметричная модель. Для двоично-симметричного канала (ДСК) вероятность ошибки на двоичный символ принимается постоянной. Также считается, что ошибки статистически не зависят друг от друга. ДСК – канал без памяти, т. е. присутствие или отсутствие ошибки в символе не зависит от предыдущих ошибок и не влияет на последующие. Данная модель канала связи достаточно адекватна для малозашумленных каналов – например, оптических, кабельных каналов, – но неприемлема для беспроводных каналов.

Беспроводные радиолинии более точно описываются математическими моделями, учитывающими затухание сигнала и рассеяние (дисперсию) его энергии. Эти модели учитывают корреляцию между ошибками в потоке информационных символов. Поэтому дисперсионный канал с замираниями (ДКЗ) сигнала позволяет адекватно определить большинство характеристик СЦРС, работающей в реальных условиях.

Моделирование ДКЗ основано на использовании сведений о расчётных физических или наблюдаемых параметрах канала. В середине 60-х гг. прошлого века использовались детерминированные параметры модели канала (время задержки сигнала, коэффициент передачи). Впоследствии оказалось, что эффективнее вводить в модель изменяемые во времени параметры, что хорошо согласовалось с нестационарной природой канала с замираниями.

ДКЗ описывается как линейный фильтр с изменяющимися во времени параметрами. На математическую модель канала передачи информации влияют не только физические характеристики среды распространения радиоволн, но и подвижность приёмного (передающего) устройства. В системах подвижной радиосвязи сигнал в точке приёма состоит из суммы сигналов, приходящих несколькими путями за счёт отражений от различных природных и искусственных объектов. Такой механизм формирования сигнала в антенне приёмника есть результат прохождения сигнала через многолучевой канал (МЛК) связи. Сигнал, прошедший через МЛК, представляет собой гауссовский случайный процесс, что следует из центральной предельной

теоремы А. М. Ляпунова. В этом случае полное описание канала связи сводится к определению комплексного коэффициента передачи канала, из которого находится корреляционная функция или спектральная плотность мощности сигнала в точке приёма.

При наличии прямой видимости результирующий сигнал в точке приёма состоит из аддитивной смеси прямого (главного) луча и, возможно, из нескольких дополнительных лучей.

2.2. Физические процессы, обуславливающие возникновение искажений сигнала в радиоканале

Понимание условий распространения и особенностей канала передачи информации важно для эффективного проектирования, оптимизации и эксплуатации СЦРС. Наибольшее внимание уделяют пониманию физических условий распространения радиоволн в зоне работы СЦРС. В настоящее время проведено множество экспериментов и созданы математические модели реальных радиоканалов. Отметим следующие ключевые особенности распространения радиоволн, вызывающие искажения сигнала.

Замирания. Сигнал в точке приёма является аддитивной суммой многих составляющих, приходящих с различных направлений, что является результатом многолучевого распространения. Характеристики сигнала зависят от особенностей ландшафта, наличия зданий и сооружений, что порождает случайные колебания уровня сигнала в зависимости от географического положения приёмника. При изменении положения приёмного устройства в пределах рабочей длины волны колебания уровня могут достигать величины 20 дБ и более. Это явление называют замираниями сигнала.

Сигнал с замираниями можно представить случайным процессом, состоящим из двух компонент – узкополосного случайного процесса с быстроменяющейся огибающей, значения которой подчиняются закону Релея:

$$p(x) = \begin{cases} \frac{x}{\sigma_x^2} \exp\left[-\frac{x^2}{2\sigma_x^2}\right], & x \geq 0; \\ 0, & x < 0, \end{cases} \quad (2.1)$$

где σ_x – параметр, значение которого определяет наиболее вероятное значение огибающей сигнала, и медленноменяющегося процесса, у которого огибающая описывается логнормальным законом распределения:

$$p(x) = \begin{cases} \frac{1}{x\sigma_x\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu_x)^2}{2\sigma_x^2}\right], & x \geq 0; \\ 0, & x < 0, \end{cases} \quad (2.2)$$

где σ_x и μ_x – параметры распределения, которые позволяют вычислить математическое ожидание m_x и дисперсию D_x случайной величины x :

$$\begin{cases} m_x = \exp\left[\mu_x + \frac{\sigma_x^2}{2}\right]; \\ D_x = \exp\left[2\mu_x + \sigma_x^2\right]\left(\exp\left[\sigma_x^2\right] - 1\right). \end{cases} \quad (2.3)$$

На рис. 2.1 показан график закона Релея при $\sigma_x = 1$, на рис. 2.2 – график логнормального закона при $\sigma_x = 1$ и $\mu_x = 1$.

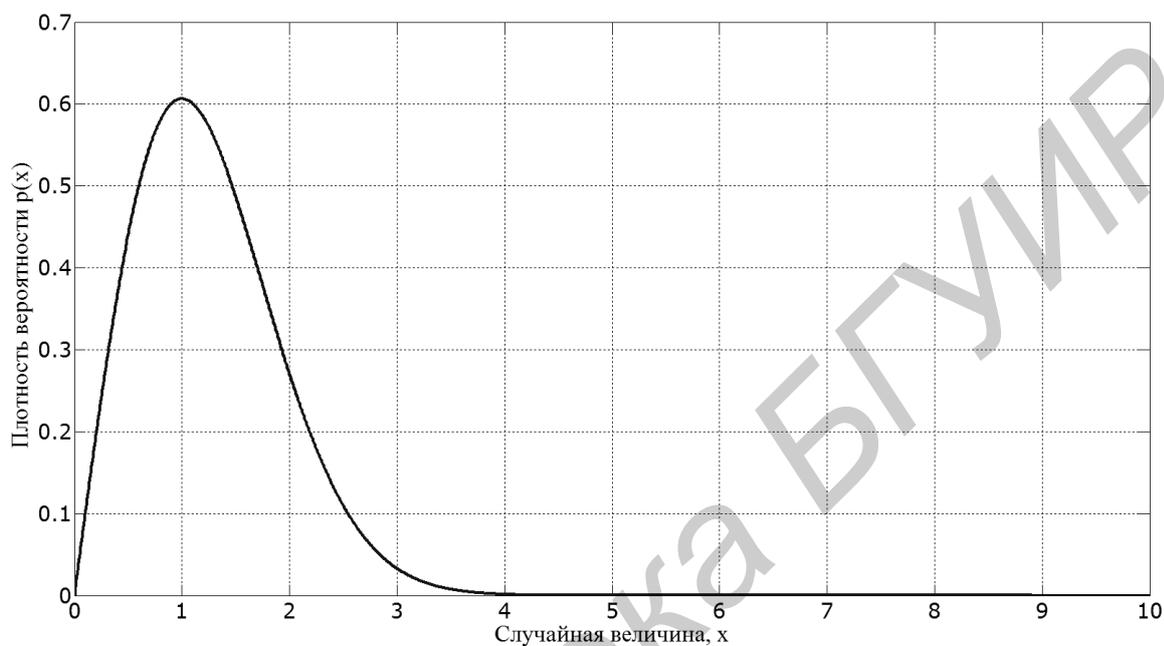


Рис. 2.1. График релейского закона распределения вероятности

Иными словами, медленные изменения математического ожидания огибающей подчиняются логнормальному закону, а быстрые флуктуации – релейскому закону. Параметры законов распределения огибающей зависят от рабочей частоты и относительной к передатчику скорости движения приёмника.

Эффект Доплера. Сущность эффекта Доплера заключается в изменении частоты принимаемого сигнала (доплеровский сдвиг частоты) при относительном движении передающего и приёмного устройств. Явление открыто австрийским физиком К. Доплером в 1842 г. Воздействие эффекта Доплера на сигнал в точке приёма усложняется тем, что в условиях многолучёвого распространения скорость приёмника относительно различных компонентов сигнала различается, что вызывает различный доплеровский сдвиг для различных компонент.

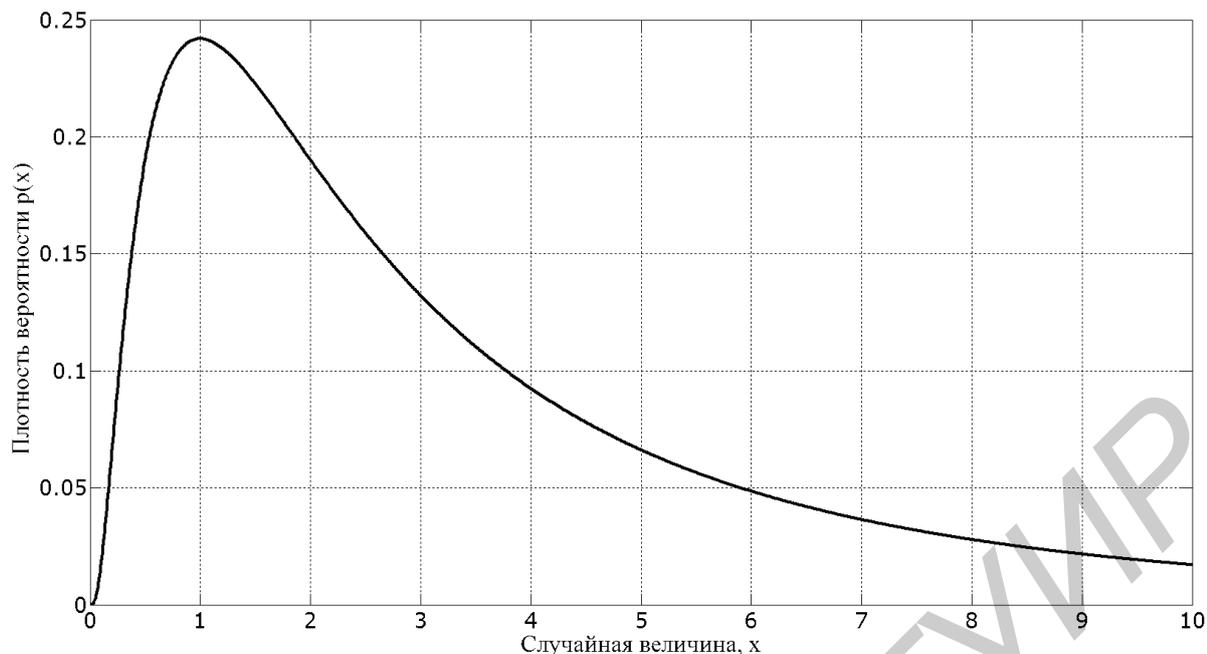


Рис. 2.2. График логнормального закона распределения вероятности

Задержки распространения сигнала. Из-за многолучевой природы распространения сигнала в зоне функционирования СЦРС в точку приёма приходят сигналы с различными задержками. Среднеквадратичное отклонение задержки может колебаться от долей микросекунды в городских условиях вплоть до ста микросекунд в холмистом пригороде. Это ограничивает максимальную полосу частот, используемых сигналом, величиной между 40 и 250 кГц. Это значение полосы частот известно как ширина полосы когерентности. Ширина полосы когерентности обратно пропорциональна среднеквадратической задержке распространения. В полосе когерентности канал имеет постоянный коэффициент передачи и линейную фазочастотную характеристику.

Для сигнала с полосой занимаемых частот большей полосы когерентности, амплитудно-частотная и фазочастотная характеристики канала становятся частотно-зависимыми. В первом приближении можно считать, что канал становится частотно-избирательным тогда, когда среднеквадратичное значение задержки сигнала больше, чем продолжительность информационного символа, что вызывает межсимвольную интерференцию (МСИ) в СЦРС. Частотно-избирательные каналы называются дисперсионными каналами, в которых наблюдаются значительные замирания, тогда как недисперсионные каналы есть каналы со сравнительно неглубокими замираниями.

Энергетический потенциал линии связи и потери распространения на трассе. Энергетический потенциал линии связи — термин, определяющий оценку мощности сигнала на приёмной стороне с учётом ослабления, вызванного потерями распространения сигнала между передатчиком и приёмником (потери на трассе).

Обычно энергетический потенциал линии связи оценивается для заданного отношения сигнал–шум (С/Ш) на входе приёмного устройства.

Суммарный шум состоит из совокупности посторонних помех и собственного шума.

2.3. Паразитные эффекты при приёме и обработке сигналов в системах цифровой радиосвязи

2.3.1. Межсимвольная интерференция

Межсимвольная интерференция (МСИ) – паразитный эффект размывания во времени радиочастотного импульса на соседние, что создаёт дополнительную помеху при обработке сигнала. МСИ является следствием частотной избирательности канала.

Наложение нескольких предыдущих символов на текущий равносильно воздействию полосового шума, что может существенно снизить энергетическую эффективность СЦРС и приводит к значительным ошибкам при декодировании сигнала.

МСИ наблюдается:

- в дисперсионных каналах с замираниями;
- в каналах с ограниченной полосой частот.

Системы связи, которые передают данные по каналам с ограниченной полосой, обычно осуществляют формирование импульса, чтобы избежать МСИ, вызванной ограничением ширины полосы. Если АЧХ канала является плоской и у формирующего фильтра конечная полоса пропускания, МСИ не возникают. В случае, если АЧХ канала неизвестна, используется адаптивный корректирующий фильтр, компенсирующий неравномерность частотной характеристики канала.

Методы борьбы с МСИ подразумевают:

- 1) адаптивное выравнивание АЧХ канала связи;
- 2) использование кодов, исправляющих ошибки;
- 3) формирование информационного сигнала с защитными интервалами между элементами.

2.3.2. Джиттер

Джиттер, или фазовое дрожание – это явление фазовой модуляции принимаемого сигнала, возникающее по причине нестабильности задающих генераторов опорных колебаний или быстрых изменений характеристик канала связи.

В СЦРС джиттер проявляется в виде случайных быстрых изменений местоположения фронтов элементарных сигналов цифрового сигнала во времени, что приводит к рассинхронизации, в результате чего возникают ошибки при декодировании сигнала. Средняя частота джиттера составляет 10 Гц и более.

Джиттер является одной из основных проблем при проектировании устройств цифровой электроники, в частности, цифровых интерфейсов.

В системах передачи информации джиттер определяется производной задержки прохождения сигнала по времени $\varphi(t) = d\tau_{\text{зад}}(t)/dt$.

Два подхода к определению джиттера:

- фазовый джиттер;
- частотный джиттер.

Быстрые колебания частоты, характеризующиеся девиацией, большей 10 Гц, относятся к джиттеру. Медленные колебания частоты с девиацией, меньшей 10 Гц, называются дрейфом фазы («vander» в англоязычной литературе).

Влияние джиттера на характеристики сигнала проявляется двояко. В случае передачи оцифрованного аналогового сигнала наличие джиттера приводит к тому, что восстановленный сигнал оказывается неравномерно дискретизированным, что может привести к значительным нарушениям в структуре аналогового сигнала. В основном такое влияние джиттер оказывает на параметры СЦРС с частотным разделением каналов. Учёт влияния джиттера в этом случае зависит от его структуры: регулярный или нерегулярный джиттер.

В случае постоянного (регулярного) джиттера, обусловленного обычно процессами мультиплексирования и регенерации, его влияние предсказуемо, носит аддитивный характер и компенсируется регенераторами и мультиплексорами. Нерегулярный (случайный) джиттер не может быть скомпенсирован и рассматривается как результат влияния аддитивного шума. Исследования показали его связь с параметрами собственного шума и возможностью накопления по мощности в составном цифровом канале. Наиболее существенным считается процесс накопления регулярного джиттера, поскольку регулярный джиттер аккумулируется быстрее нерегулярного.

При передаче данных в цифровой форме джиттер, непосредственно связанный с нарушениями в синхронизации цифрового потока, приводит к отклонению точек считывания дискретов от истинного центра.

Причины возникновения джиттера. Типы джиттера

Существует несколько основных причин возникновения джиттера, которые влияют на его структуру и воздействуют на параметры качества систем связи.

Регулярный джиттер обычно коррелирован с передаваемыми последовательностями битов. Поскольку задержки при передаче сигналов в скремблерах и кодерах обычно зависят от типа принимаемой или передаваемой последовательности, процессы в мультиплексорах и регенераторах могут вызывать регулярный джиттер.

Второй причиной возникновения регулярного джиттера могут быть нарушения в канале передачи, в частности, связанные с наличием перекрёстных помех, которые также вносят регулярный, коррелированный с последовательностями бит джиттер.

Нерегулярный джиттер обычно обусловлен электромагнитным воздействием внешних источников сигнала.

2.3.3. Глазковая диаграмма

Для оценки влияния МСИ и джиттера на сигнал и возможность его эффективной обработки используется глазковая диаграмма.

Методика построения глазковой диаграммы состоит в подаче исследуемого сигнала к вертикальным отклоняющим пластинам осциллографа, в то время как на горизонтальных отклоняющих пластинах действует пилообразный сигнал. Период пилообразного сигнала должен быть в N раз больше длительности элементарного символа сигнальной конструкции, где $N \geq 2$. Формирующееся на экране осциллографа изображение названо глазковой диаграммой из-за его схожести с разрезом человеческих глаз. Внутреннюю область глазковой диаграммы называют глазковым раскрывом.

Глазковая диаграмма при своей простоте позволяет получить ценную информацию о характере искажений сигнала. Измерения характеристик сигнала делятся на два класса.

Амплитудные измерения:

- измерение амплитуды сигнала;
- измерение амплитуды точки пересечения траекторий;
- размах области пересечения траекторий;
- амплитуда глазка;
- отношение сигнал/шум;
- качество сигнальной конструкции (равномерность распределения узлов глазковой диаграммы);
- раскрыв глазка.

Временные измерения:

- определение детерминированного джиттера;
- момент пересечения траекторий;
- задержка сигнала;
- длительность заднего фронта глазка;
- длительность переднего фронта глазка;
- интервал времени между глазками (временной разнос соседних глазков);
- временной раскрыв глазка (ширина глазка);
- девиация джиттера;
- определение случайного джиттера;
- СКО джиттера;
- величина полного джиттера.

Ширина раскрыва глазка определяет временной интервал, в течение которого обрабатываемый сигнал может быть принят и обработан без МСИ. Очевидно, что оптимальный момент времени для оценки информативного параметра сигнала находится посередине раскрыва глазка.

Чувствительность системы к погрешности синхронизации определена интервалом времени между глазками.

Амплитуда глазка для заданного временного окна позволяет оценить запас помехоустойчивости сигнала, определяемый минимальным СКО шума, требуемым для смещения амплитудного значения сигнала к центру глазка.

Увеличение интервала наблюдения при построении глазковой диаграммы позволяет оценить стационарность работы формирователя сигнала.

Многолучёвое распространение сигнала может исключить возможность однозначного определения переходов из одной сигнальной точки в другую. В результате снижается помехоустойчивость системы, а на глазковой диаграмме уменьшается размах глазка.

2.4. Математическая модель канала с замираниями

Математическая модель канала с замираниями представляет собой линейный фильтр с изменяющимися во времени параметрами. Каналы систем подвижной цифровой радиосвязи и каналы с рассеянием энергии радиоволн (дисперсионные каналы) предполагают наличие в точке приёма сигнала, являющегося суммой некоторого множества сигналов. Для описания таких каналов передачи информации используются статистические методы, сводящиеся к нахождению корреляционной функции и спектральной плотности мощности сигнала в точке приёма.

Возникновение замираний сигнала в точке приёма происходит при многолучёвом распространении радиоволн, поэтому замирания такого характера часто называют многолучёвыми замираниями. В свою очередь, на распространение радиоволн оказывает влияние ряд факторов, основные из которых перечислены ниже:

- рассеяние радиоволн (преобразование распространяющихся в одном направлении радиоволн в радиоволны, распространяющиеся в различных направлениях);

- рефракция радиоволн (изменение направления распространения радиоволн вследствие изменения скорости их распространения при прохождении их через неоднородную среду);

- дифракция радиоволн (изменение структуры поля радиоволны под влиянием препятствий, представляющих собой пространственные неоднородности среды распространения, в частности, приводящие к огибанию радиоволной этих препятствий).

Амплитуда и фаза принимаемого сигнала изменяются во времени по случайному закону по причине случайных изменений характеристик среды распространения. Результирующий сигнал имеет значительный уровень колебаний мощности как в сторону уменьшения, так и в сторону увеличения.

Дисперсионный канал с замираниями может быть представлен с использованием комплексных форм записи сигнала $s(t)$ на входе канала и на его выходе – $y(t)$, т. е. в точке излучения сигнала и в точке его приёма. Так

как большинство используемых в СЦРС сигналов есть узкополосные случайные процессы, то

$$s(t) = \operatorname{Re} \left\{ \mathbf{z}(t) e^{j2\pi f_0 t} \right\}; \quad (2.4)$$

$$y(t) = \operatorname{Re} \left\{ \mathbf{w}(t) e^{j2\pi f_0 t} \right\}. \quad (2.5)$$

Здесь f_0 – средняя частота спектра передаваемого сигнала; $\mathbf{z}(t)$, $\mathbf{w}(t)$ – комплексные огибающие сигналов $s(t)$ и $y(t)$ соответственно, определяемые следующим образом:

$$\mathbf{z}(t) = z(t) e^{j\varphi_z(t)} = \sqrt{s^2(t) + \hat{s}^2(t)} \exp \left[j \operatorname{arctg} \frac{\hat{s}(t)}{s(t)} \right]; \quad (2.6)$$

$$\mathbf{w}(t) = w(t) e^{j\varphi_w(t)} = \sqrt{y^2(t) + \hat{y}^2(t)} \exp \left[j \operatorname{arctg} \frac{\hat{y}(t)}{y(t)} \right]. \quad (2.7)$$

Здесь $\hat{s}(t)$ и $\hat{y}(t)$ – сопряжённые по Гильберту сигналы для сигналов $s(t)$ и $y(t)$ соответственно.

Для произвольной функции $x(t)$ прямое и обратное преобразование Гильберта определяется следующими выражениями:

$$\begin{cases} \hat{x}(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau; \\ x(t) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\hat{x}(\tau)}{t - \tau} d\tau. \end{cases} \quad (2.8)$$

Если предположить, что в точке приёма складывается множество сигналов, приходящих по разным путям, то для нахождения характеристик канала можно применить центральную предельную теорему А. М. Ляпунова. Импульсная характеристика канала $h(t, \xi)$ в этом случае будет представлять собой комплексный гауссовский процесс, у которого огибающая описывается распределением Релея, а начальная фаза равномерно распределена на интервале $(-\pi; \pi)$:

$$\mathbf{w}(t) = \int_{-\infty}^{+\infty} \mathbf{z}(t - \xi) h(t, \xi) d\xi, \quad (2.9)$$

где ξ – задержка сигнала.

Коэффициент передачи дисперсионного канала с замираниями определяется согласно выражению

$$K(f, t) = \int_{-\infty}^{+\infty} h(t, \xi) e^{-j2\pi f_0 \xi} d\xi. \quad (2.10)$$

Радиосигнал в точке приёма представляет собой гауссовский случайный процесс с нулевым математическим ожиданием и огибающей, распределенной по закону Релея. Также предполагается, что импульсная характеристика канала представляет собой стационарный случайный процесс. Несмотря на то что случайные замирания сигнала порождены нестационарной природой канала, на достаточно коротком интервале времени и в относительно узкой полосе частот канал передачи можно рассматривать приблизительно стационарным. Такого рода стационарность канала называется *квазистационарностью*. Каналы, стационарные в течение всего интервала наблюдения, называются стационарными в широком смысле.

Если импульсная характеристика канала $h(t, \xi)$ не зависит от задержки ξ , то говорят, что канал некоррелированно рассеивает сигнал. В случае, если изменяющаяся во времени импульсная характеристика канала имеет стационарный во времени и по частоте флуктуационный характер, то такой канал называется стационарным в широком смысле каналом с некоррелированным рассеиванием сигнала.

Стационарный в широком смысле канал передачи информации полностью определяется автокорреляционной функцией (АКФ) или спектральной плотностью мощности (СПМ) сигнала на выходе канала. В силу частотно-временной дуальности характеристик канала существует несколько определений корреляционных характеристик и СПМ при его описании.

2.5. Характеристики канала с замираниями

Частотно-временная корреляционная функция (ЧВКФ) (двухчастотная корреляционная функция)

$$R(\Omega, \tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} K^*(\omega, t) K(\omega + \Omega, t + \tau) d\omega dt. \quad (2.11)$$

Физический смысл ЧВКФ заключается в определении степени подобия сигнала $y(t)$ с его сдвинутой на время τ временной копией, спектр которой смещён на частоту Ω при условии подачи на вход канала сигнала с равномерной спектральной плотностью мощности (СПМ), равной 1.

Комплексная корреляционная функция импульсной характеристики канала

$$Q(\tau, \xi)\sigma(\Delta\xi) = \left\langle h^*(t, \xi)h(t + \tau, \xi + \Delta\xi) \right\rangle_t = \int_{-\infty}^{+\infty} h^*(t, \xi)h(t + \tau, \xi + \Delta\xi)dt. \quad (2.12)$$

Здесь угловые скобки означают операцию статистического усреднения по множеству реализаций; $\sigma(\cdot)$ – функция Хевисайда.

Функция $Q(t, \xi)$ есть результат преобразования Фурье от ЧВКФ:

$$Q(\tau, \xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} R(\Omega, \tau) e^{j\Omega(t+\xi)} d\Omega, \quad (2.13)$$

и представляет автокорреляционную функцию (АКФ) флуктуаций коэффициента передачи для задержек на интервале $(\xi; \xi + \Delta\xi)$. Предполагается, что колебания коэффициента передачи для различных задержек сигнала некоррелированы. Подобная ситуация соответствует дисперсионному каналу с некоррелированным рассеянием.

Ниже приведена функция рассеяния, определяемая как СПМ комплексного коэффициента передачи для задержки сигнала ξ :

$$S(\xi, \nu) = \int_{-\infty}^{+\infty} Q(\tau, \xi) e^{-j2\pi\nu\tau} d\tau. \quad (2.14)$$

Функция рассеяния определяет задержку сигнала и его доплеровский сдвиг. Средняя СПМ на интервале задержки $\xi + \Delta\xi$ и доплеровского сдвига $\nu + \Delta\nu$ равна $S(\xi, \nu)\Delta\xi\Delta\nu$.

Функция рассеяния является результатом операции двойного преобразования Фурье от ЧВКФ по соответствующим параметрам.

Приведённые характеристики позволяют находить свойства дисперсионных каналов с замираниями интересных с практической точки зрения:

- неизбирательные каналы;
- частотно-избирательные каналы;
- каналы с временной избирательностью;
- каналы с частотной и временной избирательностью.

2.6. Примеры математических моделей каналов

2.6.1. Неизбирательные каналы (НИК)

Неизбирательный канал – это канал, в котором не происходит ни частотной, ни временной избирательности сигнала. Такой канал иногда называют *плоским каналом*. В неизбирательном канале коэффициент передачи – случайная величина, распределённая по релеевскому закону, а фазовый сдвиг также случаен и равномерно распределен на интервале $(-\pi; \pi)$:

$$y(t) = K_0 \operatorname{Re} \left\{ \mathbf{z}(t) e^{j(2\pi f_0 t - \theta)} \right\}, \quad (2.15)$$

где K_0 и θ – статистически независимые случайные величины.

Приведенное выше выражение моделирует передачу сигнала с полосой, много меньшей интервала частотной корреляции канала и длительностью импульса, много меньшей интервала корреляции замираний. Эти два условия для большинства СЦРС не выполняются.

2.6.2. Частотно-избирательный канал (ЧИК)

Частотно-избирательный канал характеризуется неравномерной по частоте передаточной функцией. Приблизительно постоянная амплитуда и начальная фаза сигнала на выходе канала может наблюдаться лишь в течение сравнительно короткого интервала наблюдения. Для близких частот сигнала значения коэффициента передачи $K(f)$ канала имеют значительную корреляцию, в то же время для удалённых друг от друга частот значения коэффициента передачи можно считать статистически независимыми. При отсутствии корреляции между замираниями различных частотных составляющих сигнала канал называется частотно-избирательным.

В ЧИК коэффициент передачи случаен лишь для различных частот и не изменяется с течением времени. Такие каналы иногда называют плоскими во времени. Импульсная характеристика плоского во времени канала не зависит от времени: $h(t, \xi) = h(\xi)$. Канал также можно считать плоским во времени, если временные замирания сигнала малы по сравнению с длительностью элементарного символа.

Комплексная огибающая $\mathbf{w}(t)$ сигнала $y(t)$ на выходе ЧИК определяется свёрткой комплексной огибающей $\mathbf{z}(t)$ входного сигнала $s(t)$ и импульсной характеристики $h(\xi)$:

$$\mathbf{w}(t) = \int_{-\infty}^{+\infty} \mathbf{z}(t - \xi) h(\xi) d\xi. \quad (2.16)$$

Корреляционная функция ЧИК

$$\langle h^*(\xi)h(\xi + \Delta\xi) \rangle_t = Q(\xi)\sigma(\Delta\xi). \quad (2.17)$$

Функция $Q(\xi)$ в последнем выражении называется спектральной плотностью мощности задержки сигнала.

Частотная корреляционная функция, являющаяся результатом прямого преобразования Фурье от спектральной плотности задержки сигнала, определяется выражением

$$R(\Omega) = \int_{-\infty}^{+\infty} Q(\xi)e^{-j\Omega\xi}d\xi = \langle K^*(\omega)K(\omega + \Omega) \rangle_f = \int_{-\infty}^{+\infty} K^*(\omega)K(\omega + \Omega)d\omega. \quad (2.18)$$

Таким образом, ЧИК полностью описывается либо СПМ задержки сигнала, либо частотной корреляционной функцией.

В практической деятельности функция корреляции частоты $R(\Omega)$ определяется экспериментально передачей двух гармонических сигналов с различной частотой и последующей оценкой корреляции комплексных огибающих сигналов в зависимости от разноса их частот. Частотный разнос, при котором корреляция двух принимаемых сигналов максимальна, называют полосой когерентности плоского во времени дисперсионного канала. Полоса когерентности $F_{\text{ког}}$ чаще всего определяется по уровню $1/e$ или $1/2$ от максимального значения $R(\Omega)$.

СПМ задержки $Q(\xi)$ показывает распределение задержки сигнала в многолучевом канале. По данной характеристике определяется средняя задержка распространения L сигнала, а также диапазон возможных задержек ΔL , равный удвоенному СКО задержки сигнала:

$$L = \int_{-\infty}^{+\infty} \xi Q(\xi)d\xi, \quad (2.19)$$

$$\Delta L = 2\sqrt{\int_{-\infty}^{+\infty} (\xi - L)^2 Q(\xi)d\xi}. \quad (2.20)$$

Если длительность элементарного символа меньше ΔL , то в точке приёма наблюдается межсимвольная интерференция, что затрудняет или делает невозможной безошибочную демодуляцию сигнала.

Отметим, что средняя задержка и полоса когерентности – взаимно обратные величины.

2.6.3. Каналы с временной избирательностью (ВИК)

Каналы с временной избирательностью отличаются тем, что импульсная характеристика канала зависит от времени. Приблизительно постоянная амплитуда и начальная фаза сигнала $y(t)$ наблюдается лишь в течение сравнительно небольшого интервала времени. Для одинаковых или близких частот степень замираний будет коррелирована лишь в течение небольшого времени. Отсутствие корреляции коэффициента передачи на заданной частоте для различных временных сечений есть признак того, что канал является избирательным во времени.

ВИК часто называют плоским по частоте каналом. Это означает, что амплитудные и фазовые колебания одинаковы для каждой спектральной составляющей сигнала. Плоские по частоте каналы наблюдаются в случае, когда полоса частот, занимаемых сигналом, много меньше интервала корреляции канала по частоте. Низкоскоростные цифровые каналы диапазона ВЧ (3...30 МГц), например радиовещание в стандарте DRM, есть яркий пример каналов с временной избирательностью. В подобных каналах длительность информационного символа много больше средней длительности замираний.

Для плоского по частоте канала импульсный отклик определяется следующим образом:

$$\mathbf{w}(t) = \mathbf{z}(t)h(t), \quad (2.21)$$

где $h(t)$ – зависящий от времени и не зависящий от частоты импульсный отклик канала.

Комплексная корреляционная функция импульсной характеристики канала определяется следующим выражением:

$$R(0, \tau) \equiv R(\tau) = \left\langle h^*(t)h(t + \tau) \right\rangle_t = \int_{-\infty}^{+\infty} h^*(t)h(t + \tau)dt. \quad (2.22)$$

Канал с временной избирательностью полностью описывается либо функцией $R(\tau)$, либо её Фурье-образом $P(\nu)$:

$$P(\nu) = \int_{-\infty}^{+\infty} R(\tau)e^{-j2\pi\nu\tau}d\tau. \quad (2.23)$$

Функция $P(\nu)$ есть СПМ доплеровского сдвига частоты.

Экспериментальное определение корреляционной функции $R(\tau)$ предполагает излучение достаточно длительного гармонического сигнала с последующим вычислением корреляционной функции принятого процесса. Таким образом вычисляется время когерентности канала $T_{\text{ког}}$, также называемого постоянной времени замираний. Постоянная времени замираний определяется по уровню $1/e$ или $1/2$ от максимального значения функции

$R(\tau)$ – аналогично определению полосы когерентности. Постоянная времени замираний $T_{\text{ког}}$ есть мера средней длительности замираний канала.

Средний доплеровский сдвиг частоты и среднеквадратичный диапазон доплеровского сдвига частоты определяются аналогично среднему времени задержки и среднему диапазону колебаний задержки сигнала. Время когерентности канала и когерентная полоса частот являются взаимнообратными величинами, т. к. функции $P(\nu)$ и $R(\tau)$ связаны преобразованиями Фурье.

Если частота Доплера $F_{\text{доп}}$ больше полосы занимаемых сигналом частот ΔF , то говорят, что сигнал рассеян по частоте, а канал обладает свойством частотной дисперсии. Эффект Доплера есть свидетельство того, что канал передачи ведёт себя как линейный фильтр с изменяющимися параметрами. Этот факт является следствием либо взаимного движения приёмного устройства относительно передающего, либо тропосферного рассеяния сигнала при движении среды распространения радиоволн и в условиях значительных гидрометеоров.

2.6.4. Каналы с частотной и временной избирательностью (ЧВИ)

Каналы передачи сигнала, обладающие как частотной, так и временной избирательностью, называют дисперсионными в широком смысле каналами. Каналы передачи СЦРС не являются избирательными одновременно по времени и частоте. Если бы могла наблюдаться одновременная частотно-временная избирательность канала, то это означало бы, что

– полоса когерентности $F_{\text{ког}}$ одновременно и больше, и меньше полосы частот, занимаемых полезным сигналом;

– задержка распространения L одновременно и больше, и меньше длительности элементарного символа, что является противоречием.

Говоря о частотно-временной избирательности канала, полагают, что канал является избирательным по времени или частоте последовательно в течение некоторых интервалов времени.

Функция рассеяния для канала с ЧВИ определяется произведением СПМ доплеровского сдвига частоты и АКФ импульсной характеристики канала:

$$S(\xi, \nu) = P(\nu)Q(\xi). \quad (2.24)$$

Последнее выражение указывает на тот факт, что частотная и временная избирательность канала независимы друг от друга.

2.7. Характеристики многолучёвого канала в условиях прямой видимости

Импульсная характеристика МЛК в условиях прямой видимости представляется комплексным гауссовским процессом с ненулевым математическим ожиданием, огибающая которого описывается случайным процессом, распределённым по закону Райса (см. рис. 2.3):

$$p(h_0) = \begin{cases} \frac{h_0}{\sigma^2} I_0 \left[\frac{mh_0}{\sigma^2} \right] \exp \left[-\frac{h_0^2 + m^2}{2\sigma^2} \right], & h_0 \geq 0 \\ 0, & h_0 < 0, \end{cases} \quad (2.25)$$

Здесь σ – СКО порождающего гауссовского процесса; $m = \sqrt{m_I^2 + m_Q^2}$ – математическое ожидание комплексного гауссовского процесса, определяемое через МО синфазной m_I и квадратурной m_Q ;

$I_0(x) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{x \cos t} dt$ – модифицированная функция Бесселя первого рода нулевого порядка.

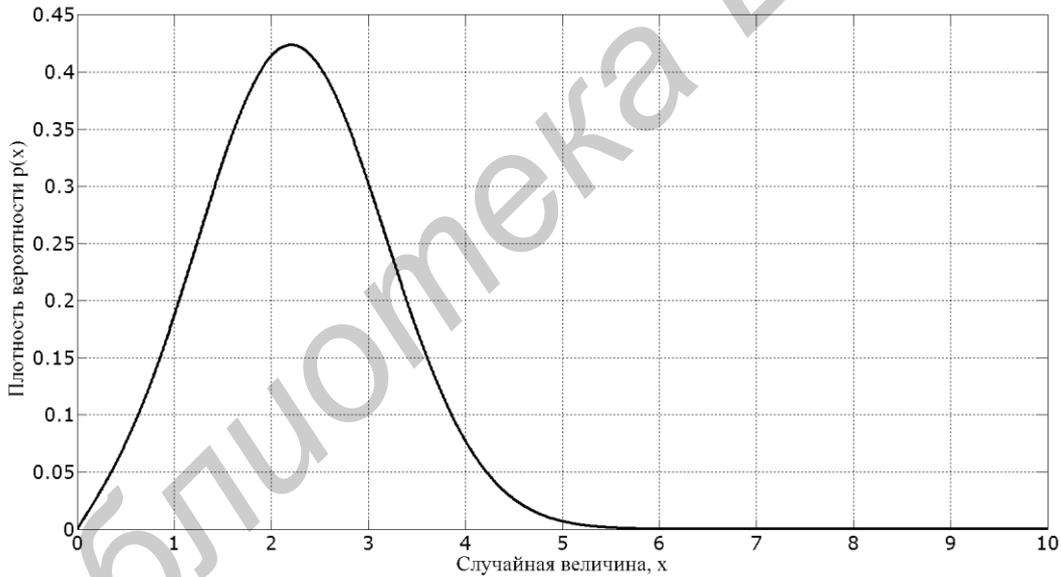


Рис. 2.3. Распределение Райса для $m=2$ и $\sigma=1$

Следует отметить, что параметры σ и m не являются СКО и МО процесса с райсовским распределением. Канал, импульсная характеристика которого описывается распределением, приведенным выше, называется каналом с райсовскими замираниями.

Коэффициент передачи канала определяется выражением

$$K(j\omega) = 1 + \sum_{i=1}^n \beta_i e^{j\omega\tau_i}, \quad (2.26)$$

в котором коэффициент передачи прямого луча нормализован, т. е. модуль ККП равен единице, а задержка сигнала равна нулю. β_i и τ_i – коэффициенты передачи и задержки сигнала, приходящего по i -му пути соответственно.

Модуль ККП для двухлучёвого канала

$$K(\omega) = \sqrt{1 + \beta^2 + 2\beta \cos \omega\tau}, \quad (2.27)$$

а аргумент

$$\varphi(\omega) = \arctg \left[\frac{\beta \sin \omega\tau}{1 + \beta \cos \omega\tau} \right]. \quad (2.28)$$

Графики модуля и аргумента ККП двухлучёвого канала приведены на рисунках ниже.

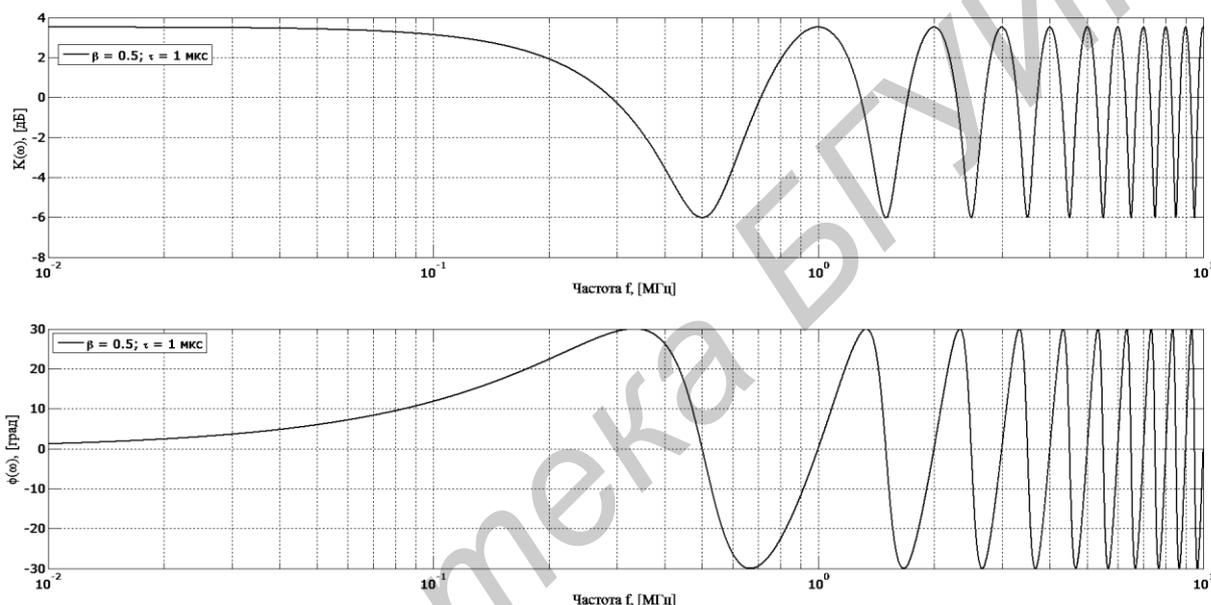


Рис. 2.4. Комплексный коэффициент передачи двухлучёвого канала при $\beta = 0,5$ и $\tau = 1$ мкс

Как видно из графиков, наибольшее подавление сигнала наблюдается на частотах

$$f_{\text{под}} = \frac{2n-1}{2} \cdot \frac{1}{\tau}, \quad n \in N,$$

и равно величине $K_{\min} = 1 - \beta$.

Множество частот $f_{\text{под}}$ иногда называют частотами режекции сигнала.

Групповое время задержки (ГВЗ) двухлучёвого канала определяется выражением

$$T(\omega) = \left| \frac{d\varphi}{d\omega} \right| = \left| \beta\tau \frac{\beta + \cos \omega\tau}{1 + \beta \cos \omega\tau + \beta^2} \right|. \quad (2.29)$$

Максимальное по модулю значение ГВЗ зависит как от задержки, так и от коэффициента передачи канала для второго луча:

$$T_{\min} = \frac{\beta\tau}{1-\beta}.$$

Частоты, при которых ГВЗ равно T_{\min} , соответствуют частотам максимального подавления сигнала.

Характеристики двухлучёвого канала имеют ясный физический смысл и достаточно просты в применении, однако большинство каналов передачи в СЦРС предполагают распространение сигнала по трём и более путям и анализ таких каналов представляет собой более сложную задачу.

Натурные испытания по определению характеристик многолучёвых каналов узкополосных СЦРС показали плодотворность задания ККП или ГВЗ канала полиномами степени M :

$$p(\omega) = C_0 + C_1\omega + C_2\omega^2 + \dots + C_M\omega^M = \sum_{i=0}^M C_i\omega^i, \quad (2.30)$$

где $p(\omega)$ – аппроксимирующая ту или иную характеристику канала функция.

Множество коэффициентов полинома определяется исходя из минимума среднеквадратичного отклонения их оценок при достаточно большом количестве опытов. Как правило, используются полиномы степени M , равной 2, 4 и 6.

Для частотно-избирательных каналов используются полиномы $M = 4$. На интервалах времени, в течение которых замирания отсутствуют или имеют плоский во времени характер, хорошие результаты при определении свойств канала дают полиномы степени $M = 0$ или $M = 2$.

Передаточная функция МЛК определяется следующим образом:

$$K(j\omega) = A_0 + \sum_{k=1}^n (A_k + jB_k)(j\omega)^k. \quad (2.31)$$

Таким образом, ККП двухлучёвого канала описывается полиномом первой степени с тремя коэффициентами $\{A_0; A_1; B_1\}$. Для каналов передачи, описание которых требует применения полиномов второй степени, потребуется уже 5 коэффициентов.

2.3. Модели ошибок в каналах цифровой радиосвязи

Для СЦРС иногда необходимо связать вероятность ошибки на символ (бит) с поведением реального канала передачи, который в большинстве случаев создаёт пакетные ошибки. В этом направлении могут быть реализованы два подхода.

Классический метод связывает скорость возникновения ошибок со степенью изменения уровня сигнала в точке приёма. В этом случае используется зависимость вероятности ошибки на символ от отношения сигнал/шум (С/Ш).

Второй подход предполагает моделирование ошибок в каналах с памятью. Основанная на сведениях о характеристиках канала передачи, модель

отражает особенности поведения канала, а статистика ошибок представляется функцией параметров канала.

Основной проблемой моделирования статистики ошибок, возникающих при приеме сигнала, является выбор адекватной модели, описывающей передаваемые данные. Модель Гильберта использует марковские цепи, состоящие из последовательностей «хороших» G , т. е. безошибочно принятых, бит, и «плохих» B , т. е. ошибочных бит. В более сложных задачах используется большее количество состояний марковской цепи.

Библиотека БГУИР

3. ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ

3.1. Мера количества информации в дискретном сообщении

Система связи служит для передачи сообщений от отправителя к получателю. Однако не всякое сообщение содержит информацию. Информация – это совокупность сведений об объекте или явлении, которые увеличивают знания потребителя об этом объекте или явлении.

В математической теории информации исходят из того, что в некотором событии (сообщении) x_i количество информации $I(x_i)$ зависит не от его конкретного содержания, степени важности и т. д., а от того, каким образом выбирается (совершается) данное событие из общей совокупности возможных событий.

В реальных условиях выбор конкретного события производится с некоторой априорной вероятностью $P(x_i)$ совершения данного события. Чем меньше эта вероятность, тем больше информации содержится в данном сообщении.

При определении количества информации исходят из следующих требований:

1. Количественная мера информации должна обладать свойством аддитивности: количество информации в нескольких независимых сообщениях должно равняться сумме количества информации в каждом сообщении.

2. Количество информации о достоверном событии ($P(x_i) = 1$) должно равняться нулю, так как такое сообщение не увеличивает наших знаний о данном объекте или явлении.

Указанным требованиям удовлетворяет логарифмическая мера, определяемая формулой

$$I(x_i) = \log_m \left(\frac{1}{P(x_i)} \right). \quad (3.1)$$

Чаще всего логарифм берется с основанием 2, реже – с основанием e :
 $I(x_i) = -\log_2(P(x_i))$ двоичных единиц информации (бит) или
 $I(x_i) = -\ln(P(x_i))$ натуральных единиц информации (нит).

Одну двоичную единицу информации содержит сообщение, вероятность выбора которого равняется $1/2$. В этом случае

$$I(x_i) = -\log_2(1/2) = \log_2(2) = 1 \text{ бит.}$$

Таким образом, 1 бит информации соответствует количеству информации, содержащейся в событии, которое имеет два равновероятных исхода.

Учитывая, что в практике передачи и преобразования информации широко применяются двоичные символы, двоичная логика, двоичные источники

сообщений и двоичные каналы передачи, наиболее часто используется двоичная единица информации (бит).

3.2. Энтропия дискретного источника с независимым выбором сообщений

В теории информации чаще всего необходимо знать не количество информации $I(x_i)$, содержащееся в отдельном сообщении (символе), а среднее количество информации в одном символе, создаваемом источником сообщений.

Если имеется ансамбль (полная группа) из k сообщений (символов) x_1, x_2, \dots, x_k , называемая алфавитом источника с вероятностями $P(x_1), P(x_2), \dots, P(x_k)$, то среднее количество информации, приходящееся на один символ, называется энтропией источника сообщений $H(x)$ и определяется формулой

$$H(x) = \sum_{i=1}^k P(x_i) I(x_i) = - \sum_{i=1}^k P(x_i) \log_2(P(x_i)). \quad (3.2)$$

Размерность энтропии – среднее количество единиц информации, приходящееся на один символ, генерируемый источником информации. Энтропия характеризует источник сообщений с точки зрения неопределённости выбора того или другого сообщения или его производительности.

Рассмотрим свойства энтропии.

1. Чем больше неопределённость выбора сообщений, тем больше энтропия. Неопределённость максимальна при равенстве вероятностей выбора каждого сообщения: $P(x_1) = P(x_2) = \dots = P(x_k) = 1/k$. В этом случае

$H_{max}(x) = \sum_{i=1}^k \frac{1}{k} \log_2(k) = \log_2(k)$ (т. е. максимальная энтропия равна логарифму от объёма алфавита). Например, при $k = 2$ (двоичный источник)

$H_{max}(x) = \sum_{i=1}^2 \frac{1}{2} \log_2(2) = \log_2(2) = 1$ бит.

2. Энтропия минимальна и равна нулю, если для одного из символов вероятность его появления равна 1, а для остальных равна 0 (выбирается всегда только один заранее известный символ, например, одна буква). Зависимость $H(x)$ от вероятностей появления символов для двоичного источника приведена на рис. 3.1.

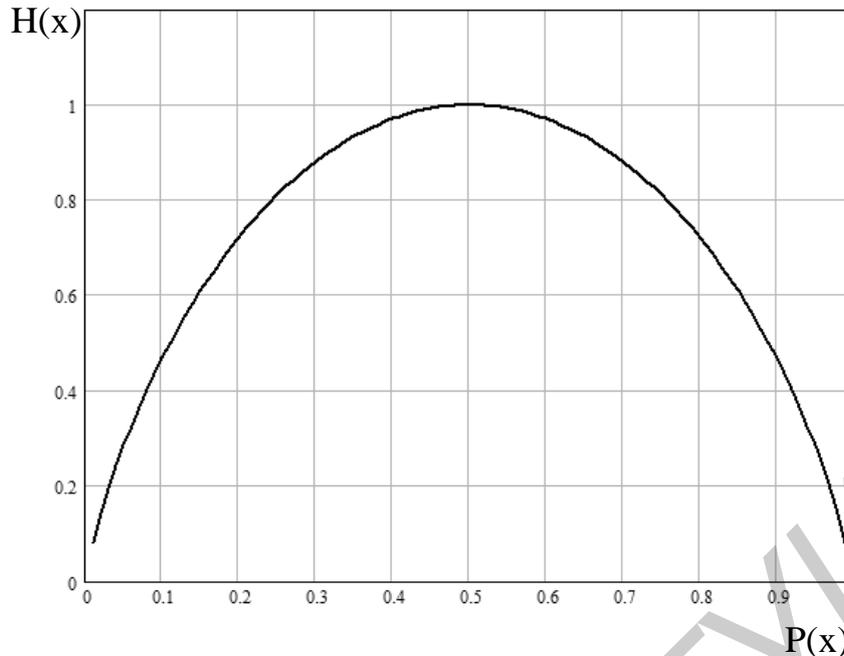


Рис. 3.1. Зависимость энтропии двоичного источника от вероятности появления символов

3. Свойство аддитивности. Пусть элементарные символы группируются в «слова» размерностью в n символов. В этом случае общее количество возможных слов (размерность нового алфавита) увеличивается с k до k^n . Таким образом, энтропия источника, генерирующего «слова» равна

$$H(x_{n \times k}) = \log_2(k^n) = n \log_2(k). \quad (3.3)$$

Таким образом, укрупнение элементарных символов в «слова» привело к увеличению энтропии в n раз, так как теперь уже слово включает в себя информацию n элементарных символов исходного алфавита.

4. Энтропия источника информации не может быть отрицательной.

3.3. Энтропия дискретного источника с зависимыми сообщениями

Ранее при определении энтропии предполагалось, что каждое сообщение (символ или слово) выбирается независимым образом. Рассмотрим более сложный случай, когда в источнике сообщений имеются корреляционные связи. Пусть выбор очередного символа сообщения зависит от конечного числа предшествующих символов n . Математической моделью такого источника является марковская цепь n -го порядка, у которой вероятность выбора очередного символа зависит от n предшествующих символов и не зависит от более ранних

Если объем алфавита источника равен k , а число связанных символов, которые необходимо учитывать при определении вероятности очередного символа, равно порядку источника n , то каждому символу может предшествовать $M = k^n$ различных сочетаний символов (состояний источника), влияющих на вероятность появления очередного символа x_i . А вероятность появления в сообщении любого из k возможных символов определяется условной вероятностью $P(x_i / x_{i-1}, x_{i-2}, \dots, x_{i-n})$ с учётом предшествующих символов $x_{i-1}, x_{i-2}, \dots, x_{i-n}$, т. е. с учётом M возможных состояний. Эти состояния обозначим как q_1, q_2, \dots, q_M .

Энтропия такого дискретного источника определяется в два этапа.

1. Вычисляются условные энтропии источника в каждом из M состояний, считаем эти состояния известными:

$$\begin{aligned} \text{для состояния } q_1: H(x/q_1) &= -\sum_{i=1}^k P(x_i/q_1) \log_2(P(x_i/q_1)); \\ \text{для состояния } q_2: H(x/q_2) &= -\sum_{i=1}^k P(x_i/q_2) \log_2(P(x_i/q_2)); \\ \text{для состояния } q_M: H(x/q_M) &= -\sum_{i=1}^k P(x_i/q_M) \log_2(P(x_i/q_M)). \end{aligned} \quad (3.4)$$

2. Далее находим энтропию источника $H(x)$ путём усреднения по всем состояниям q :

$$H(x) = \sum_{j=1}^M P(q_j) H(x/q_j). \quad (3.5)$$

Окончательно получаем

$$H(x) = -\sum_{j=1}^M P(q_j) \sum_{i=1}^k P(x_i/q_j) \log_2(P(x_i/q_j)). \quad (3.6)$$

При наличии корреляционных связей между символами в эргодическом источнике энтропия уменьшается, так как при этом уменьшается неопределённость выбора символов и в ряде случаев часть символов можно угадать по предыдущим или ближайшим символам.

3.4. Избыточность источника

Как было показано выше, энтропия максимальна при равновероятном появлении элементов сообщения и отсутствии между ними корреляционных связей. При неравномерном распределении вероятности появления символов

и при наличии корреляционных связей между символами энтропия уменьшается.

Чем ближе энтропия источника информации к максимальной, тем рациональнее работает источник. Чтобы судить о том, насколько хорошо использует источник свой алфавит, вводят понятие избыточности источника сообщений

$$g = \frac{H_{max}(x) - H(x)}{H_{max}(x)} = \frac{\log_2(k) - H(x)}{\log_2(k)}. \quad (3.7)$$

Наличие избыточности приводит к загрузке канала связи передачей лишних символов, которые не несут информации (их можно угадать и не передавая).

Однако преднамеренная избыточность в сообщениях иногда используется для повышения достоверности передачи информации – например, при помехоустойчивом кодировании в системах передачи информации с исправлением ошибок. Большую избыточность имеет любая устная речь. Например, избыточность русского языка (как и других) – около 50 %. Благодаря избыточности облегчается понимание речи при наличии дефектов в произношении или при искажениях речевых сигналов в каналах связи.

3.5. Производительность источника

Производительность источника определяется количеством информации, передаваемой в единицу времени. Измеряется производительность количеством двоичных единиц информации (бит) в секунду. Если все элементы сообщения имеют одинаковую длительность τ , то производительность источника равна

$$R = \frac{H(x)}{\tau}. \quad (3.8)$$

Если же различные элементы сообщения имеют разную длительность, то в приведённой формуле надо учитывать среднюю длительность $\langle \tau \rangle$, равную математическому ожиданию символа:

$$\langle \tau \rangle = \sum_{i=1}^k \tau_i P(\tau_i). \quad (3.9)$$

Однако в последней формуле $P(\tau_i)$ можно заменить на $P(x_i)$ (вероятность i -го элемента сообщения), так как эти вероятности равны. В результате получаем

$$\langle \tau \rangle = \sum_{i=1}^k \tau_i P(x_i), \quad (3.10)$$

а производительность источника будет равна

$$R = \frac{H(x)}{\langle \tau \rangle}. \quad (3.11)$$

Максимально возможная производительность дискретного источника равна $R_{max} = \frac{H_{max}(x)}{\tau}$.

Для двоичного источника, имеющего одинаковую длительность элементов сообщения, имеем

$$R_{max} = \frac{1}{\tau} \text{ бит/с.}$$

При укрупнении алфавита в слова по n символов, имеем

$$R_{max} = \frac{n \log_2(2)}{n\tau} = \frac{1}{\tau} \text{ бит/с.}$$

Таким образом, путём увеличения размерности алфавита увеличить производительность источника нельзя, так как в этом случае и энтропия, и длительность сообщения одновременно возрастают в одинаковое число раз.

Увеличить производительность можно путём уменьшения длительности элементов сообщения, однако возможность эта ограничивается полосой пропускания канала связи. Поэтому производительность источника можно увеличить за счет более экономного использования полосы пропускания, например, путем применения сложных многоуровневых сигналов.

3.6. Совместная энтропия двух источников

Пусть имеется два дискретных источника с энтропиями $H(x)$ и $H(y)$ и объёмами алфавитов k и m .

Объединим оба эти источника в один суммарный источник и определим совместную энтропию. Элементарное сообщение на выходе системы содержит элементарное сообщение x_i и сообщение y_j . Алфавит сложной системы будет иметь объём $k \times m$, а энтропия будет равна

$$H(x, y) = -\sum_{i=1}^k \sum_{j=1}^m P(x_i, y_j) \log(P(x_i, y_j)). \quad (3.12)$$

По теореме умножения вероятностей

$$P(x, y) = P(x)P(y/x) = P(y)P(x/y) \quad (3.13)$$

с учетом этого имеем

$$\begin{aligned} H(x, y) &= -\sum_{i=1}^k \sum_{j=1}^m P(x_i, y_j) \log(P(x_i, y_j)) = H(x) + H(y/x) = \\ &= H(y) + H(x/y). \end{aligned} \quad (3.14)$$

Здесь $H(y/x) = -\sum_i \sum_j P(x_i, y_j) \log_2(P(y_j/x_i))$ – условная энтропия источника y относительно источника x . Она показывает, какую энтропию имеют сообщения y , когда уже известно сообщение x .

Если источники независимы, то $P(y/x) = P(y)$ и $H(y/x) = H(y)$. В этом случае $H(x, y) = H(x) + H(y)$.

Если источники частично зависимы, то $H(x, y) < H(x) + H(y)$.

Если источники полностью зависимы (x и y содержат одну и ту же информацию), то $H(y/x) = 0$ и $H(x, y) = H(x) = H(y)$.

3.7. Скорость передачи и пропускная способность канала связи

В дискретной системе связи при отсутствии помех информация на выходе канала связи полностью совпадает с информацией на его входе, поэтому скорость передачи информации численно равна производительности источника сообщений:

$$C = R(x) = \frac{H(x)}{\langle \tau \rangle}. \quad (3.15)$$

При наличии помех часть информации источника теряется и скорость передачи информации оказывается меньшей, чем производительность источника. Одновременно в сообщение на выходе канала добавляется информация о помехах (рис. 3.2). Поэтому при наличии помех необходимо учитывать на выходе канала не всю информацию, сообщаемую источником, а только взаимную информацию.

В процессе передачи символы входного алфавита претерпевают искажения и преобразуются в символы выходного алфавита канала связи, которые в общем случае могут не совпадать со входными.

Под *взаимной информацией* будем понимать приращение знания о входных символах при приеме (получении) выходных символов. Данное приращение характеризуется приращением апостериорной вероятности (вероятности после наблюдения конкретного выходного символа y_j) конкретного символа входного алфавита x_i относительно априорной вероятности (вероятности, которая была до приема символа выходного алфавита) данного символа.

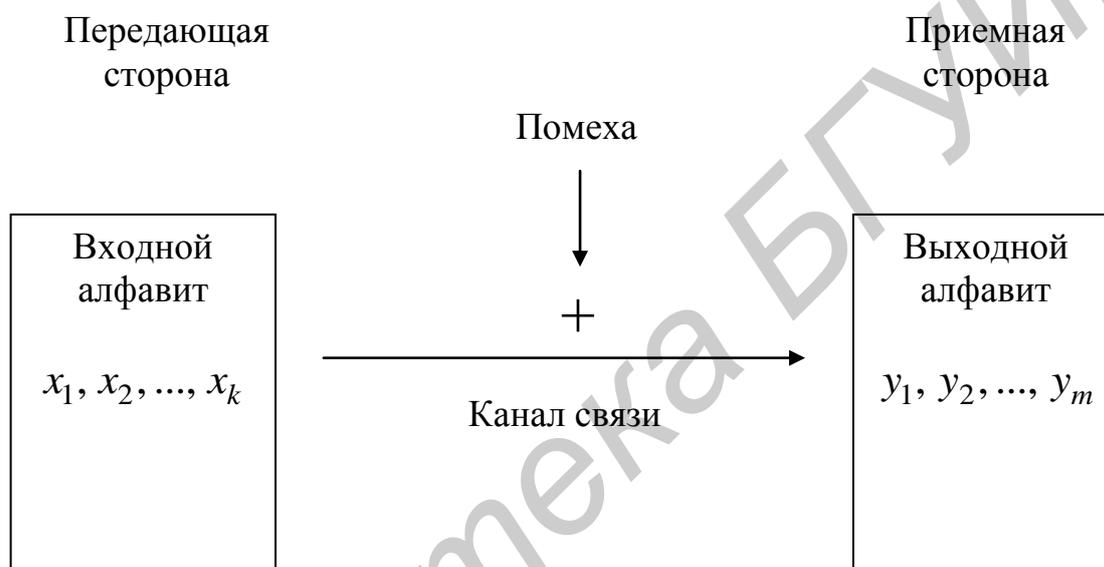


Рис. 3.2. Процесс передачи информации

В двоичных единицах информации данная взаимная информация запишется следующим образом:

$$I_{y_j \rightarrow x_i} = \log_2 \left(\frac{P(x_i / y_j)}{P(x_i)} \right) = \log_2 \left(\frac{P(x_i / y_j)P(y_j)}{P(x_i)P(y_j)} \right) = \quad (3.16)$$

$$= \log_2 \left(\frac{P(x_i, y_j)}{P(x_i)P(y_j)} \right) = I(x_i) + I(y_j) - I(x_i, y_j),$$

$$I(x_i) = -\log_2(P(x_i)), \quad (3.17)$$

$$I(y_j) = -\log_2(P(y_j)), \quad (3.18)$$

$$I(x_i, y_j) = -\log_2(P(x_i, y_j)), \quad (3.19)$$

Взаимная информация обладает свойством симметрии $I_{y_j \rightarrow x_i} = I_{x_i \rightarrow y_j}$ и характеризует количество информации, которое приносит конкретный символ выходного алфавита y_j (принятый символ) относительно конкретного символа входного алфавита x_i (переданного символа) и наоборот.

Для того чтобы охарактеризовать процесс передачи информации в целом, осуществим усреднение взаимной информации с учетом вероятностей появления входных и выходных символов.

$$\begin{aligned} H_{y \rightarrow x} = H_{x \rightarrow y} &= \sum_i \sum_j P(x_i, y_j) I_{y \rightarrow x} = \sum_i \sum_j P(x_i, y_j) I_{x \rightarrow y} = \\ &= \sum_i \sum_j P(x_i, y_j) \log_2 \left(\frac{P(x_i, y_j)}{P(x_i)P(y_j)} \right) = H(x) + H(y) - H(x, y) = \\ &= H(x) - H(x/y) = H(y) - H(y/x). \end{aligned} \quad (3.20)$$

Таким образом, информационная производительность канала связи с шумами определяется следующим образом:

$$\begin{aligned} R(x, y) &= \frac{H_{y \rightarrow x}}{\langle \tau \rangle} = \frac{1}{\langle \tau \rangle} (H(x) - H(x/y)) = R(x) - R(x/y) = \\ &= R(y) - R(y/x), \end{aligned} \quad (3.21)$$

где $R(x) = \frac{1}{\langle \tau \rangle} (H_{max}(x) - g)$ – производительность источника; $R(x/y)$ – «ненадёжность» канала, (потери) в единицу времени; $R(y)$ – энтропия выходного сообщения в единицу времени; $R(y/x) = R(n)$ – энтропия помех (шума) в единицу времени; g – избыточность источника.

Пропускной способностью канала связи (канала передачи информации) C называется максимально возможная скорость передачи информации по каналу

$$C = \max(R(x, y)). \quad (3.22)$$

Таким образом, пропускная способность канала связи равна максимальной производительности источника на входе канала, полностью согла-

сованного с характеристиками этого канала, за вычетом потерь информации в канале из-за помех.

Для бинарного канала с шумами пропускная способность равна

$$C = \max(R(x, y)) = \max(R(y) - R(y/x)). \quad (3.23)$$

В случае бинарного канала связи размерность выходного алфавита равна 2, следовательно, максимальная производительность канала связи на выходе равна

$$\max(R(y)) = \frac{1}{\langle \tau \rangle} \log_2 2 = \frac{1}{\langle \tau \rangle}, \quad (3.24)$$

и, следовательно, пропускная способность канала запишется следующим образом:

$$C = \frac{1}{\langle \tau \rangle} \max(1 - H(y/x)). \quad (3.25)$$

Определим условную энтропию

$$\begin{aligned} H(y/x) &= -\sum_{i=1}^2 \sum_{j=1}^2 P(x_i, y_j) \log_2(P(y_j/x_i)) = \\ &= -P(x_1)P(y_1/x_1) \log_2(P(y_1/x_1)) - P(x_1)P(y_2/x_1) \log_2(P(y_2/x_1)) - \\ &- P(x_2)P(y_1/x_2) \log_2(P(y_1/x_2)) - P(x_2)P(y_2/x_2) \log_2(P(y_2/x_2)) = \\ &= -P(x_1)(1 - P_e) \log_2(1 - P_e) - P(x_1)(P_e) \log_2(P_e) - \\ &- P(x_2)(P_e) \log_2(P_e) - P(x_2)(1 - P_e) \log_2(1 - P_e) = \\ &= -(P_e) \log_2(P_e) - (1 - P_e) \log_2(1 - P_e), \end{aligned} \quad (3.26)$$

где P_e – вероятность ошибочного приема.

Таким образом, пропускная способность бинарного канала с шумами равна

$$C = \frac{1}{\langle \tau \rangle} (1 - H_e) = \frac{1}{\langle \tau \rangle} (1 + P_e \log_2(P_e) + (1 - P_e) \log_2(1 - P_e)), \quad (3.27)$$

где H_e – энтропия ошибок при приеме.

В канале без помех пропускная способность равна производительности источника.

Для эффективного использования пропускной способности канала необходимо его согласование с источником информации на входе. Такое согласование возможно как для каналов связи без помех, так и для каналов с помехами на основании двух теорем, доказанных К. Шенноном.

1-я теорема (для канала связи без помех):

Если источник сообщений имеет энтропию H (бит на символ), а канал связи – пропускную способность C (бит в секунду), то можно закодировать сообщения таким образом, чтобы передавать информацию по каналу со средней скоростью, сколь угодно близкой к величине C , но не превзойти её.

К. Шеннон предложил также и метод такого кодирования, который получил название статистического, или оптимального кодирования. В дальнейшем идея такого кодирования была развита в работах Р. Фано и Д. Хаффмена и в настоящее время широко используется на практике для «сжатия сообщений».

2-я теорема (для каналов связи с помехами):

Если пропускная способность канала равна C , а производительность источника $R(x) < C$, то путём соответствующего кодирования можно передавать информацию по каналу связи со скоростью, сколь угодно близкой к C и с вероятностью ошибки, сколь угодно близкой к нулю. Если же $R(x) > C$, то можно закодировать источник таким образом, что ненадёжность будет меньше, чем $R(x) - C + \epsilon$, где ϵ – сколь угодно малая величина.

Не существует способа кодирования, обеспечивающего ненадёжность, меньшую, чем $R(x) - C$.

К сожалению, теорема К. Шеннона для каналов с шумами (помехами) указывает только на возможность такого кодирования, но не указывает способа построения соответствующего кода.

3.8. Энтропия непрерывного источника случайной величины и её свойства

Для нахождения энтропии непрерывного источника воспользуемся выражением для энтропии дискретного источника, заменив вероятности $P(x)$ на элементарные вероятности $W(x)dx$, где $W(x)$ – плотность распределения вероятности информационного сигнала источника. В результате получим

$$\begin{aligned}
H(x) &= - \int_{-\infty}^{\infty} W(x) dx \log_2(W(x) dx) = \\
&= - \int_{-\infty}^{\infty} W(x) (\log_2(W(x)) + \log_2(dx)) dx.
\end{aligned}
\tag{3.28}$$

Логарифм бесконечно малой величины (dx) равен минус бесконечности, в результате чего получаем

$$H(x) = \infty - \int_{-\infty}^{\infty} W(x) \log_2(W(x)) dx.
\tag{3.29}$$

Таким образом, энтропия непрерывной случайной величины бесконечно велика. Но так как в последнем выражении первое слагаемое (∞) от величины x или от $W(x)$ не зависит, при определении энтропии непрерывной величины это слагаемое отбрасывают, учитывая только второе слагаемое (некоторую «добавку» к бесконечности). Эта добавочная энтропия, определяемая формулой

$$H(x) = - \int_{-\infty}^{\infty} W(x) \log_2(W(x)) dx,
\tag{3.30}$$

и называется дифференциальной энтропией непрерывной случайной величины (источника).

Как и для дискретных сообщений, существуют следующие разновидности дифференциальной энтропии непрерывной величины.

Условная энтропия случайной величины y относительно случайной величины x .

$$H(y/x) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(x, y) \log_2(W(y/x)) dx dy.
\tag{3.31}$$

Совместная энтропия двух непрерывных случайных величин равна

$$H(x, y) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(x, y) \log_2(W(x, y)) dx dy.
\tag{3.32}$$

Для дифференциальной энтропии справедливы все свойства простой энтропии.

Если случайная величина ограничена в объёме $V = b - a$ (может принимать значения только в интервале от a до b), то её дифференциальная энтропия максимальна при равномерном законе распределения этой величины:

$$H_{max}(x) = \log_2(b - a). \quad (3.33)$$

Если случайная величина не ограничена в численных значениях (т. е. может изменяться в пределах от $-\infty$ до $+\infty$), а ограничена только по мощности, то дифференциальная энтропия максимальна в случае гауссовского закона распределения этой величины. Этот максимум соответствует величине

$$H_{max}(x) = \log_2 \sqrt{2\pi e \sigma^2}, \quad (3.34)$$

где σ^2 – мощность случайной величины.

Следовательно, энтропия зависит только от мощности σ^2 .

Эта очень важная формула будет использоваться позднее для определения пропускной способности непрерывного канала связи.

3.9. Пропускная способность непрерывного канала

Сигнал, отображающий непрерывное сообщение, можно рассматривать как некоторый случайный процесс, спектр которого ограничен определённой полосой частот. В соответствии с теоремой Котельникова для описания этого процесса длительностью T требуется $N = \frac{T}{\Delta t}$ отсчётов, где $\Delta t = \frac{1}{2F_{max}}$ – интервал дискретизации. Так как сигнал с ограниченным спектром полностью характеризуется своими дискретными значениями, то знание значений сигнала между этими значениями не увеличивает наших знаний о сигнале. Следовательно, при определении энтропии непрерывного сигнала достаточно учитывать только его дискретные значения, взятые в соответствии с теоремой Котельникова.

Как было показано выше, энтропия обладает свойством аддитивности. Так, если у какого-то дискретного сигнала длительностью t энтропия равна $H(x)$, то энтропия сигнала, составленного из N элементов, будет равна $N \times H(x)$. Аналогичным образом можно вычислить энтропию непрерывного сигнала длительностью T , которая будет равна

$$N \times H_1(x) = 2 \times F_{max} \times T \times H_1(x), \quad (3.35)$$

где $H_1(x)$ – энтропия одного сечения случайного сигнала, определяемая через одномерную плотность вероятности. Размерность энтропии $H_1(x)$ – бит на один отсчёт случайного сигнала (одно сечение случайного процесса).

Производительность непрерывного случайного процесса (источника) будет равна

$$R(x) = \frac{H(x)}{T} = 2 \times F_{max} \times H_1(x). \quad (3.36)$$

Таким образом, производительность эргодического источника непрерывного сигнала полностью определяется энтропией одного отсчета и удвоенной полосой частот генерируемого источником сигнала.

Если $x(t)$ – сигнал на входе канала связи, а $y(t) = x(t) + n(t)$ – сигнал на его выходе ($n(t)$ – аддитивный шум канала), то скорость передачи информации по непрерывному каналу связи будет определяться так:

$$C = \max(R(x, y)), \quad (3.37)$$

$$R(x, y) = \frac{H_{y \rightarrow x}}{\langle \tau \rangle} = \frac{1}{\langle \tau \rangle} (H(y) - H(y/x)), \quad (3.38)$$

где величину $1/\langle \tau \rangle$ надо заменить на $2F_{max}$. Предполагая, что источник сигнала согласован с каналом и его полоса пропускания $\Delta F = F_{max}$, имеем

$$R(x, y) = 2\Delta F (H(y) - H(y/x)), \quad (3.39)$$

где, как и ранее, $H(y)$ – это энтропия выходного сигнала канала связи; $H(y/x)$ – энтропия шума (это название будет разъяснено далее по тексту).

Пропускная способность равна максимально возможной скорости передачи по каналу связи, когда источник сигнала полностью согласован с характеристиками канала связи:

$$C = 2\Delta F_{max} (H(y) - H(y/x)). \quad (3.40)$$

Максимум $H(y)$ достигается в случае гауссовского закона распределения случайной величины y . При этом

$$\max H(y) = \log_2 \sqrt{2\pi e \sigma_y^2}, \quad (3.41)$$

где σ_y^2 – мощность выходного сигнала канала связи.

При учёте влияния помехи необходимо рассматривать наихудший случай, когда помеха распределена также по гауссовскому закону:

$$W_n(n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{n^2}{2\sigma_n^2}\right), \quad (3.42)$$

где σ_n^2 – мощность шума, (тем самым гарантируется пропускная способность канала не хуже определенной, при любых статистических характеристиках шума).

Условная вероятность $W(y/x)$ – это вероятность распределения случайной величины y при якобы известном заранее значении x , хотя величина x является случайной. Но так как $y(t) = x(t) + n(t)$, можно записать

$$W(y/x) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(y-x)^2}{2\sigma_n^2}\right). \quad (3.43)$$

Определим условную энтропию $H(y/x)$:

$$\begin{aligned} H(y/x) &= - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(x,y) \log_2(W(y/x)) dx dy = \\ &= - \int_{-\infty}^{\infty} W(x) \int_{-\infty}^{\infty} W(y/x) \log_2(W(y/x)) dy dx = \\ &= - \int_{-\infty}^{\infty} W(x) \log_2 \sqrt{2\pi e \sigma_n^2} dx = \log_2 \sqrt{2\pi e \sigma_n^2} = H(n). \end{aligned} \quad (3.44)$$

Отсюда видно, почему условная энтропия $H(y/x)$ называется энтропией шума.

Осуществив необходимые подстановки, получаем следующее выражение для пропускной способности канала связи:

$$C = 2\Delta F \log_2 \left(\sqrt{\frac{\sigma_y^2}{\sigma_n^2}} \right) = \Delta F \log_2 \left(\frac{\sigma_y^2}{\sigma_n^2} \right). \quad (3.45)$$

Учитывая, что рассматриваемые случайные процессы $x(t)$, $n(t)$ и $y(t) = x(t) + n(t)$ являются гауссовскими, мощность выходного процесса $y(t)$ определяется как сумма мощностей его аддитивных составляющих $\sigma_y^2 = \sigma_x^2 + \sigma_n^2$. Тогда выражение для пропускной способности будет иметь вид (формула Шеннона)

$$C = \Delta F \log_2 \left(1 + \frac{\sigma_x^2}{\sigma_n^2} \right) = \Delta F \log_2 (1 + q). \quad (3.46)$$

где $q = \frac{\sigma_x^2}{\sigma_n^2} = \frac{P_s}{P_n}$ – отношение мощности сигнала к мощности шума в канале связи.

В заключение можно отметить следующее. Для достижения скорости передачи информации по непрерывному каналу связи, близкой к пропускной способности канала связи, сигнал $x(t)$ по статистической структуре должен быть близок к флюктуационной помехе (шуму) с гауссовским законом распределения.

3.10. Предельные информационные характеристики канала связи

Преобразуем формулу Шеннона следующим образом:

$$C \leq \Delta F \log_2 \left(1 + \frac{P_s}{P_n} \right) = \Delta F \log_2 \left(1 + \frac{E_b C}{N_0 \Delta F} \right), \quad (3.47)$$

где $E_b = P_s / C$ – энергия, приходящаяся на один бит передаваемой информации; $N_0 = P_n / \Delta F$ – спектральная плотность мощности аддитивного шума, действующего в канале связи.

Выполним ряд преобразований.

$$\beta = \frac{C}{\Delta F} \leq \log_2 \left(1 + \frac{E_b C}{N_0 \Delta F} \right), \quad (3.48)$$

$$\beta \leq \log_2 \left(1 + \frac{E_b}{N_0} \beta \right), \quad (3.49)$$

$$2^\beta \leq 1 + \frac{E_b}{N_0} \beta. \quad (3.50)$$

Окончательно получим нижнюю границу отношения $\frac{E_b}{N_0}$, при которой возможна реализация заданной спектральной эффективности $\beta = \frac{C}{\Delta F}$, (бит/с)/Гц, канала связи

$$\frac{E_b}{N_0} \geq \frac{2^\beta - 1}{\beta}.$$

Рассмотрим канал связи при условии $\Delta F \rightarrow \infty$. Перепишем формулу Шеннона следующим образом:

$$1 \leq \frac{E_b}{N_0} \log_2 \left(1 + \frac{E_b C}{N_0 \Delta F} \right)^{\frac{N_0 \Delta F}{E_b C}}. \quad (3.51)$$

Учитывая, что $\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x} \right)^x = e$, получим минимально возможное отношение энергии, приходящейся на один бит передаваемой информации, к спектральной мощности шума, при котором еще возможна передача информации.

$$\frac{E_b}{N_0} \geq \frac{1}{\log_2(e)} = \ln(2) \approx 0,69. \quad (3.52)$$

Если рассмотреть полученную выше нижнюю границу для $\frac{E_b}{N_0}$, то в случае если $\beta \rightarrow 0$, т. е. $\Delta F \rightarrow \infty$ или $C \rightarrow 0$, получим $\frac{E_b}{N_0} \geq \lim_{\beta \rightarrow 0} \frac{2^\beta - 1}{\beta} = \ln(2)$. Это означает, что если $E_b / N_0 < 0,69$, передача информации невозможна. На рис. 3.3 приведена зависимость минимального требуемого отношения энергии, приходящейся на один бит передаваемой информации, к спектральной плотности шума для обеспечения заданной спектральной эффективности передачи информации.

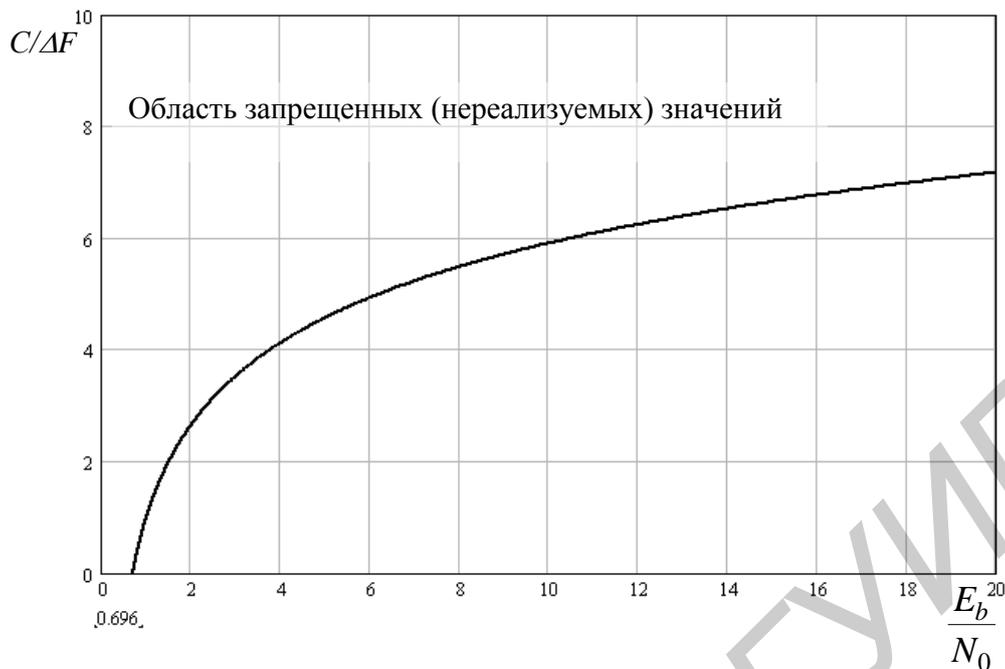


Рис. 3.3. Соотношение между спектральной эффективностью и требуемым минимальным отношением сигнал/шум

3.11. Расширение границы Шеннона (BLAST – Bell Labs Layered Space Time)

Как известно, пропускная способность канала связи с ограниченной полосой частот удовлетворяет следующему соотношению:

$$C \leq \Delta F \log_2(1 + q), \text{ бит/с,}$$

где ΔF – ширина полосы частот канала; q – мощности сигнала к мощности шума в канале связи. Является ли данная граница непреодолимой?

Рассмотрим структурную схему системы передачи информации, приведенную на рис. 3.4.

Передаваемый поток данных разбивается на несколько параллельных потоков (число потоков равно n). Далее каждый из полученных параллельных потоков подвергает модуляции свой радиосигнал. Радиосигналы всех параллельных потоков отличаются только модулирующими символами (совпадают схемы модуляции, интервалы излучения, используемые коды канала, центральные частоты, уровни и пр.) Каждый из полученных радиосигналов излучается в пространство своей антенной (антенны имеют пространственное разнесение).

На приемной стороне организуется m приемных каналов (по числу используемых разнесенных в пространстве антенн). Всегда $m \geq n$. Сигналы принятых каналов подвергаются специальной обработке, заключающейся в операции, обратной той, которой они подверглись в среде распространения

($[K(j\omega)]$) матрица $n \times m$, элементы которой представляют собой комплексные коэффициенты передачи между i -й передающей и j -й приемной антеннами).

Таким образом, вектор принимаемых сигналов $\mathbf{S}_{rx} = [S_{rx1}, \dots, S_{rxm}]$ можно записать следующим образом:

$$\mathbf{S}_{rx} = \mathbf{S}_{tx} \mathbf{K}(j\omega), \quad (3.53)$$

где $\mathbf{S}_{tx} = [S_{tx1}, \dots, S_{txn}]$ – вектор сигналов, сформированных модулятором передающей части.

С учетом операции «Компенсация канала» имеем вектор сигналов на входе демодулятора

$$\mathbf{S}_{dem} = \mathbf{S}_{rx} \mathbf{K}(j\omega)^{-1} = \mathbf{S}_{tx} \mathbf{K}(j\omega) \mathbf{K}(j\omega)^{-1} = \mathbf{S}_{tx}. \quad (3.54)$$

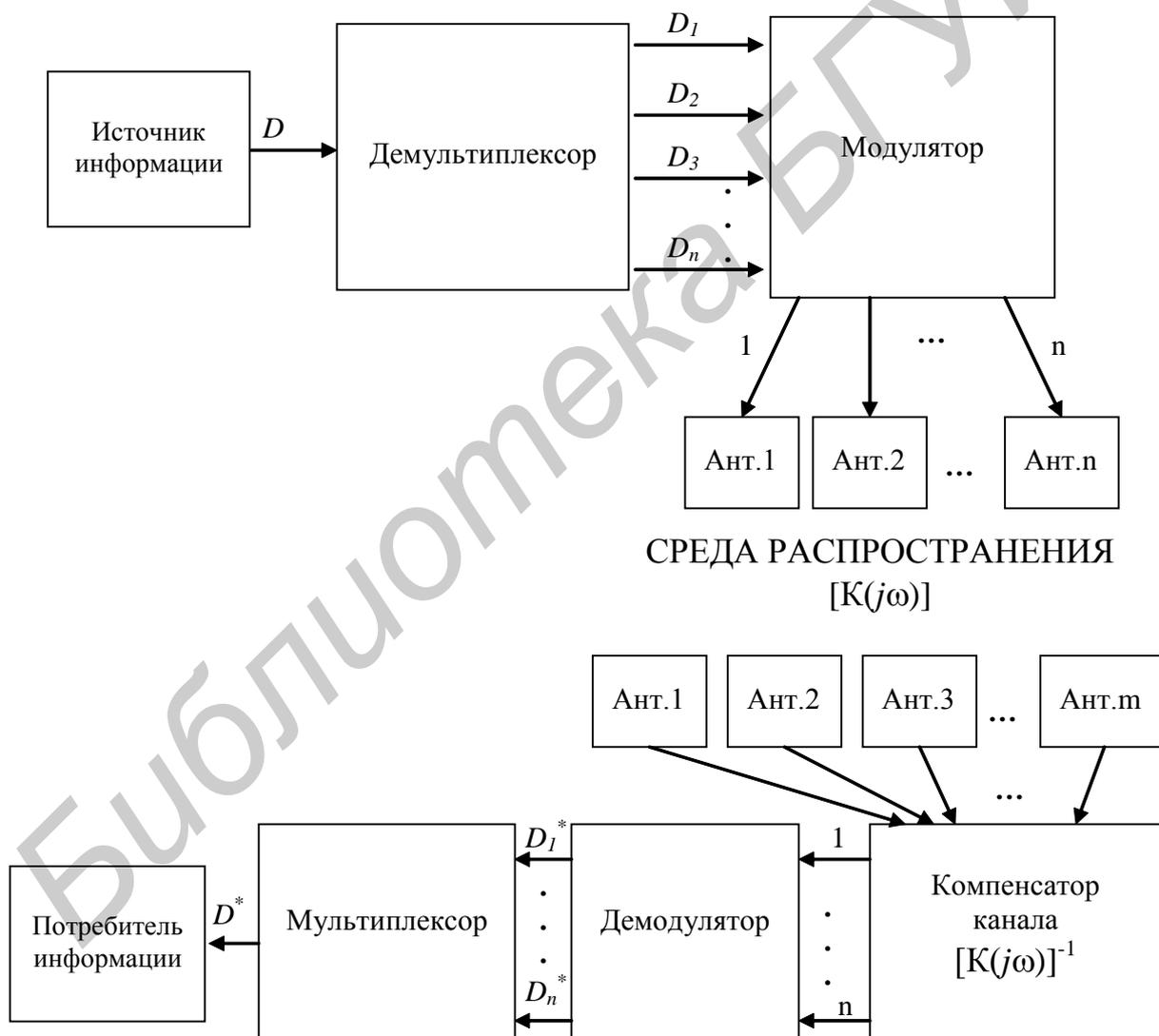


Рис. 3.4. Структурная схема, поясняющая сущность BLAST технологии

Таким образом, на входе демодулятора (соответственно на его выходе) имеем копии передаваемых сигналов или данных.

Сравним приведенную выше систему с обычной системой связи, использующей только один канал на передачу и прием. Для сравнения будем считать, что отношение q в обеих системах одинаково. Это означает, что в каждом канале «разнесенной» системы отношение сигнал/шум будет в n раз меньше q/n . Таким образом, общую пропускную способность можно определить (на основании свойства аддитивности информации) как

$$C_{\text{разнес.}} \leq n\Delta F \log_2 \left(1 + \frac{q}{n} \right), \quad (3.55)$$

а для обычной системы имеем

$$C_{\text{неразнес.}} \leq \Delta F \log_2 (1 + q) . \quad (3.56)$$

На рис. 3.5 приведены графические зависимости потенциального выигрыша в пропускной способности системы, построенной по BLAST технологии, по сравнению с обычной системой связи ($n=1$) от отношения q при различном числе каналов n разнесенной передачи информации.

Из рис. 3.5 видно, что применение технологии BLAST может увеличивать пропускную способность системы связи в разы при сохранении спектральных и энергетических характеристик канала связи. Есть ли противоречия в полученном результате? Рассмотрим процедуру «компенсации канала»:

$$\mathbf{K}(j\omega) \mathbf{K}(j\omega)^{-1} = \mathbf{I} , \quad (3.57)$$

где \mathbf{I} – единичная матрица размером $n \times n$. Другими словами, правильно осуществленная процедура «компенсации канала» эквивалентна реализации n параллельных независимых каналов. Так что полученный результат вполне адекватен и не противоречит основным положениям теории информации. Так за счет чего возникает выигрыш без дополнительных энергетических и спектральных затрат?

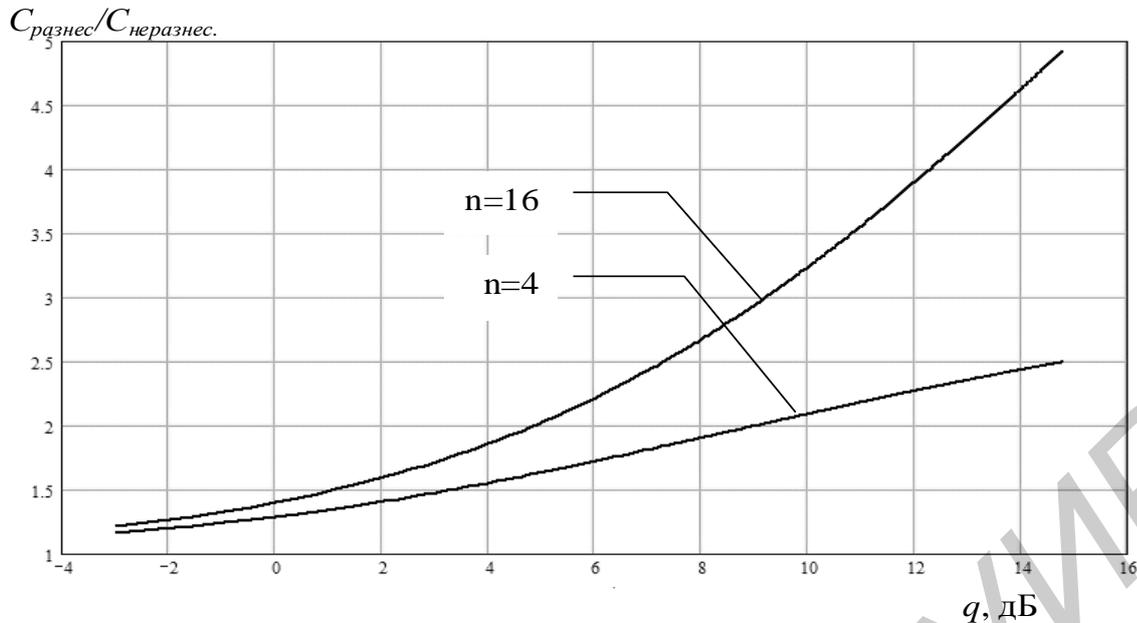


Рис. 3.5. Потенциальный выигрыш в пропускной способности

Если внимательно присмотреться к приемной части, то можно интерпретировать совокупность из m приемных антенн и компенсатора канала как антенную решетку, реализующую n своеобразных пространственных лучей, которые как бы собирают энергию, излучаемую каждой передающей антенной в различные направления (а не только в направлении на приемную систему). Это можно еще интерпретировать как собирание всевозможных пространственных лучей, вызываемых переотражениями, преломлениями и прочими эффектами распространения радиоволн. Именно за счет такого сбора «рассеянной» энергии и получается выигрыш в пропускной способности. Данное утверждение подтверждается тем, что в условиях свободного распространения радиоволн матрица $\mathbf{K}(j\omega)$ становится сингулярной (определитель равен нулю) и нахождение обратной матрицы становится невозможным.

Сложность применения технологии BLAST заключается в возможности и точности определения матрицы комплексных коэффициентов передачи каналов $\mathbf{K}(j\omega)$. Существенная нестационарность среды распространения радиоволн и необходимость точного оценивания фазовых соотношений в данной матрице являются сдерживающим фактором повышения пропускной способности. Объем обучения алгоритма «компенсации канала» накладывает ограничения на количество параллельных каналов на передающей стороне и может полностью нивелировать реализуемый выигрыш в пропускной способности. Необходимо отметить, что к настоящему времени для реализованных опытных образцов каналов передачи информации с BLAST-технологией, полученные выигрыши едва превышают значения, равные двум. При этом сложность и стоимость имеющихся BLAST решений сдерживает их широкое практическое внедрение.

4. ДИСКРЕТИЗАЦИЯ И КВАНТОВАНИЕ СИГНАЛОВ

4.1. Дискретизация во времени

4.1.1. Вводные замечания

Для передачи по каналу связи аналоговых сообщений цифровыми методами сигнал, переносящий сообщение, должен быть представлен в цифровой форме. Процедуру, реализующую данное преобразование, называют аналого-цифровым преобразованием. Первым этапом аналого-цифрового преобразования является дискретизация сигнала, в результате которого формируется последовательность выборок (отсчётов). Значения выборок соответствуют мгновенным значениям сигнала в некоторые равноотстоящие друг от друга моменты времени. Таким образом, на выходе устройства дискретизации формируется дискретный во времени сигнал с континуумом значений.

Дискретный сигнал позволяет по выборкам точно восстановить исходный аналоговый сигнал. Процесс дискретизация сигнала и его восстановления базируется на теореме В. А. Котельникова.

4.1.2. Представление аналоговых сигналов в дискретной форме

Пусть $s(t)$ – некоторый произвольный аналоговый сигнал с ограниченным спектром, определяемым в общем виде комплексной функцией угловой частоты $\mathbf{S}(\omega) \equiv S(j\omega)$. Модуль спектральной плотности сигнала обозначим как $S(\omega)$. Для сигнала с ограниченным спектром двухсторонняя спектральная плотность $S(\omega)$ при $|\omega| > \omega_m$ равна нулю. Угловая частота $\omega_m = 2\pi f_m$ называется граничной частотой спектра сигнала $s(t)$, а сам сигнал – низкочастотным. Понятие «низкая частота» в данном случае является относительным, но не абсолютным. Для сигнала, формируемого датчиком температуры воздуха на метеостанции, граничная частота будет порядка 10^{-3} Гц, для аудиосигналов – ~ 20 кГц, для телевизионного сигнала в классических вещательных системах – ~ 6 МГц. Во всех указанных примерах сигналы называются «низкочастотными», в англоязычной литературе – «baseband signals». В русскоязычной литературе иногда можно встретить понятие «видеосигнал».

Каждые T секунд у сигнала берутся выборки, т. е. в данные моменты времени определяются мгновенные значения сигнала, в результате чего формируется последовательность импульсов (отсчётов) $s_d(t)$. Значение T называется интервалом (периодом) дискретизации сигнала, величина, обратная ему, – частотой дискретизации $f_d = 1/T$. На практике чаще всего оперируют понятием частоты дискретизации, нежели интервалом дискретизации. Величина f_d измеряется в единицах циклической частоты – герцах.

Аналитически дискретный сигнал задаётся следующим образом:

$$s_d(t) = s(kT) = s(t)\delta_T(t) = s(t) \sum_{k=-\infty}^{+\infty} \delta(t - kT). \quad (4.1)$$

Здесь $k \in Z$ – номер отсчёта; $\delta_T(t)$ – опорный сигнал, представляющий собой последовательность δ -функций с периодом следования T .

Функция $s_d(t)$ представляет идеальный дискретный сигнал. При техническом осуществлении процесса дискретизации аналоговых сигналов вместо δ -функций, в силу невозможности их физической реализации, используется периодическая последовательность прямоугольных видеоимпульсов, длительность которых много меньше периода дискретизации T .

4.1.3. Теорема В. А. Котельникова

Теория дискретизации сигналов базируется на теореме отчётов, сформулированной советским академиком В. А. Котельниковым. Согласно теореме сигнал, спектр которого не превышает некоторую граничную частоту ω_m , может быть точно представлен последовательностью своих отсчётов, взятых через равные интервалы времени $T = \pi/\omega_m = 1/(2f_m)$:

$$s(t) = \sum_{k=-\infty}^{\infty} s(kT) \frac{\sin[\omega_m(t - kT)]}{\omega_m(t - kT)} = \sum_{k=-\infty}^{\infty} s(kT) \text{sinc}[\omega_m(t - kT)]. \quad (4.2)$$

Данное интерполяционное выражение является базовым в теории кодирования аналоговых источников. Следует отметить, что интервал времени T может быть меньше значения $1/(2f_m)$, но не превышает его. Минимально допустимая частота дискретизации сигнала равна удвоенной граничной частоте спектра дискретизируемого сигнала $s(t)$.

Спектр последовательности $\delta_T(t)$ определяется выражением

$$S_\delta(\omega) = \omega_d \sum_{n=-\infty}^{+\infty} \delta(\omega - n\omega_d), \quad (4.3)$$

где $\omega_d = 2\omega_m$.

Воспользовавшись свойством свёртки двух функций, получим

$$S_d(\omega) = \frac{1}{2\pi} \left[S(\omega) \otimes \omega_d \sum_{n=-\infty}^{+\infty} \delta(\omega - n\omega_d) \right] = \frac{1}{T} \sum_{n=-\infty}^{+\infty} S(\omega - n\omega_d). \quad (4.4)$$

Здесь знак \otimes означает операцию свёртки двух функций.

В частотной области операция дискретизации сигнала с ограниченным спектром означает повторение спектра исходного сигнала вдоль частотной оси. Период повторения спектра аналогового сигнала равен $\omega_d = 2\omega_m$ или $f_d = 2f_m$ (см. рис. 4.1).

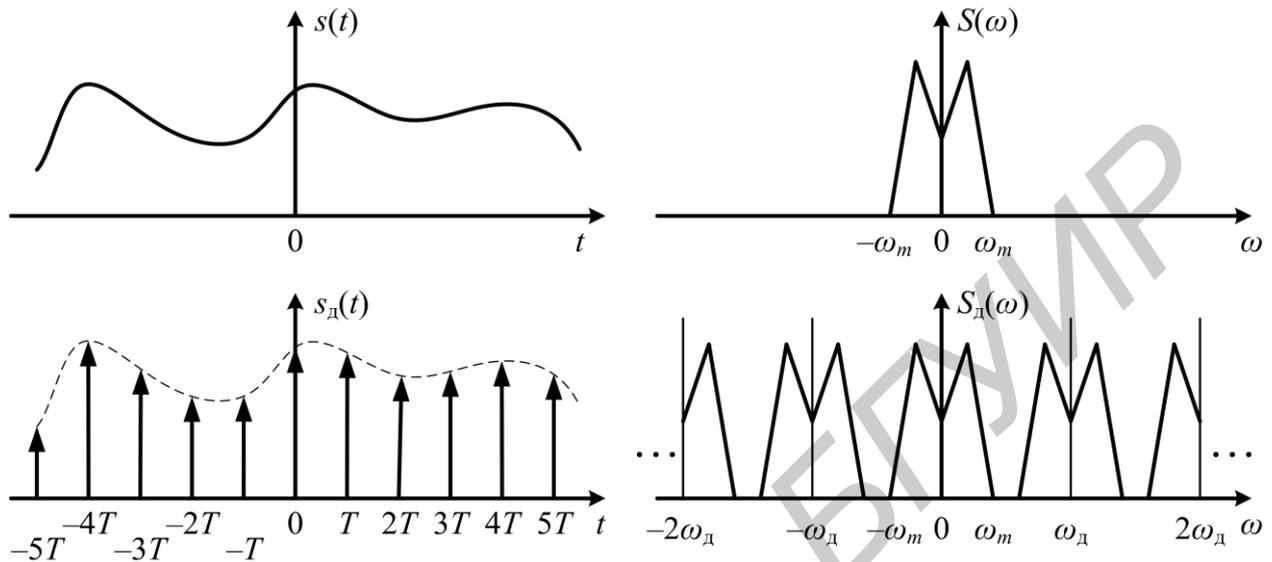


Рис. 4.1. Преобразование спектра при дискретизации сигнала

Спектр $S(\omega)$ исходного сигнала повторяется в спектре $S_d(\omega)$ дискретного сигнала, не перекрываясь только в том случае, если $\omega_d \geq 2\omega_m$ или $f_d \geq 2f_m$. Из рисунка видно, что восстановление аналогового сигнала из дискретного возможно путём пропускания последнего через идеальный фильтр нижних частот (ФНЧ) с частотой среза, заключённой в диапазоне от f_m до $f_d - f_m$. В оптимальном случае частота среза должна находиться посередине указанного интервала и быть равной $f_d/2$.

Передаточная функция идеального ФНЧ определяется выражением

$$K_{\text{ФНЧ}}(\omega) = \begin{cases} T, & |\omega| \leq \omega_m; \\ 0, & |\omega| > \omega_m, \end{cases} \quad (4.5)$$

где $\omega_m = \pi/T$. Следовательно,

$$S(\omega) = S_d(\omega)K_{\text{ФНЧ}}(\omega). \quad (4.6)$$

Найдя обратное преобразование Фурье, получим импульсную характеристику ФНЧ:

$$h_{\text{ФНЧ}}(t) = \frac{\sin \omega_m t}{\omega_m t}. \quad (4.7)$$

Вычислим обратное преобразование Фурье и получим выражение для непрерывного сигнала:

$$\begin{aligned} s(t) = s_d(t) \otimes h_{\text{ФНЧ}}(t) &= \sum_{k=-\infty}^{+\infty} s(kT) \delta(t - kT) \otimes \frac{\sin \omega_m t}{\omega_m t} = \\ &= \sum_{k=-\infty}^{+\infty} s(kT) \frac{\sin \omega_m (t - kT)}{\omega_m (t - kT)}. \end{aligned} \quad (4.8)$$

Полученное выражение совпадает с общей записью теоремы Котельникова. Теорема В. А. Котельникова налагает условие ограниченности на спектр сигнала на выходе аналогового источника и определяет минимально возможную частоту дискретизации. В случае, если частота дискретизации $f_d < 2f_m$, происходит наложение, или перекрытие, фрагментов спектра дискретного сигнала. В этом случае неискажённое восстановление аналогового сигнала из дискретного невозможно. В реальных технических устройствах кодирования аналогового источника информации частоту дискретизации выбирают несколько большей, чем того требует теорема Котельникова, создавая в спектре дискретного сигнала некоторый защитный интервал. Например, в телефонии верхняя граничная частота спектра сигнала определяется значением 3,4 кГц, однако частота дискретизации выбирается равной не 6,8 кГц, а 8 кГц. Таким образом, в спектре дискретного сигнала создаётся защитный интервал $f_d - 2f_m = 1,2$ кГц. Введение такого защитного интервала существенно облегчает проектирование и расчёт ФНЧ, восстанавливающего исходный сигнал из дискретного.

Особым случаем является дискретизация гармонического колебания с частотой f_0 . Здесь принципиально необходимо, чтобы частота дискретизации была $f_d > 2f_0$, а не $f_d = 2f_0$, как того требует теорема В. А. Котельникова. Если частота дискретизации будет точно равна удвоенной частоте колебания, то возможна ситуация, когда все отсчёты гармонического колебания будут иметь нулевые значения.

4.1.4. Дискретизация полосовых сигналов

Действительный сигнал $s(t)$ называется полосовым, если его двухсторонний спектр везде равен нулю, за исключением интервалов $\omega \in (\omega_1; \omega_2)$ и $\omega \in (-\omega_2; -\omega_1)$, где $\omega_1 = 2\pi f_1$, $\omega_2 = 2\pi f_2$. Такой сигнал можно рассматривать как узкополосный процесс, для которого выполняется условие

$$\frac{\Delta\omega}{\omega_0} = \frac{\Delta f}{f_0} \gg 1, \quad (4.9)$$

где $\Delta\omega = 2\pi\Delta f = \omega_2 - \omega_1$ – ширина полосы частот аналогового сигнала; $\omega_0 = 2\pi f_0 = (\omega_1 + \omega_2)/2$ – центральная частота спектра.

Согласно теореме В. А. Котельникова частота дискретизации должна быть не менее $2f_2$. Однако в рассматриваемом случае это значение частоты слишком велико для корректной дискретизации и последующего восстановления полосового сигнала. В действительности минимальная частота дискретизации полосового сигнала определяется нижней граничной частотой f_1 , верхней граничной частотой f_2 , а также их разницей $\Delta f = f_2 - f_1$ – шириной полосы частот (см. рис. 4.2).

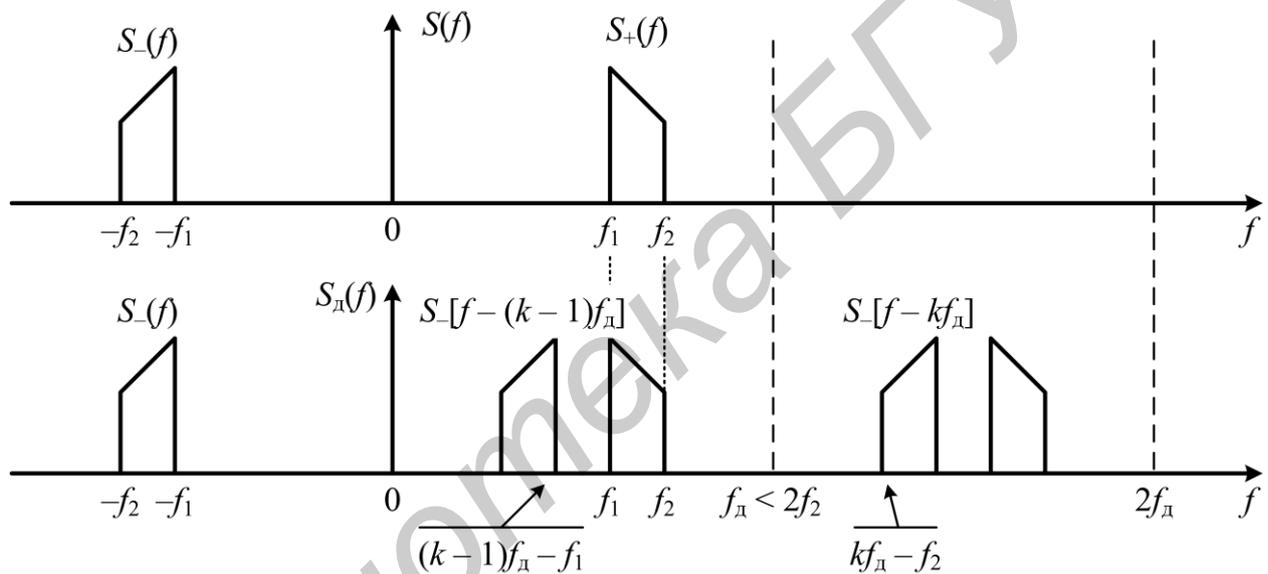


Рис. 4.2. Спектральное преобразование при дискретизации полосового сигнала

Согласно рис. 4.2, спектральная плотность исходного аналогового сигнала может быть представлена суммой двух функций, первая из которых определена для отрицательных частот, вторая – для положительных:

$$S(f) = S_-(f) + S_+(f). \quad (4.10)$$

В случае, когда $f_d < 2f_2$, при дискретизации полосового сигнала вокруг его спектра положительных частот слева и справа формируются два фрагмента его спектра отрицательных частот. Отсюда следует, что частота дискретизации f_d должна быть выбрана таким образом, чтобы верхняя частота спектрального фрагмента $S_-[f - (k-1)f_d]$ не превышала значения f_1 ,

а нижняя частота спектрального фрагмента $S_- [f - kf_d]$ была не меньше, чем значения f_2 . В этом случае спектр исходного полосового сигнала не перекрывается соседними фрагментами спектра дискретного сигнала, что допускает неискажённое восстановление аналогового сигнала из дискретного путём пропускания последнего через идеальный полосовой фильтр.

Для определения частоты дискретизации полосового сигнала запишем систему неравенств:

$$\begin{cases} (k-1)f_d - f_1 \leq f_1; \\ kf_d - f_2 \geq f_2, \end{cases} \quad (4.11)$$

где k – натуральное число.

Так как $f_1 = f_2 - \Delta f$, то выражение (4.11) можно привести к виду

$$1 \leq k \leq \frac{f_2}{\Delta f} \leq \frac{kf_d}{2\Delta f}. \quad (4.12)$$

Или

$$\frac{k-1}{2} \frac{f_d}{\Delta f} \leq \frac{f_2}{\Delta f} - 1. \quad (4.13)$$

Минимально возможная допустимая частота дискретизации полосового сигнала определяется выражением

$$f_{d \min} = \frac{2f_2}{\lfloor f_2/\Delta f \rfloor}, \quad (4.14)$$

где операция $\lfloor X \rfloor$ означает операцию округления числа X до ближайшего целого, не превышающего X .

На рис. 4.3 приведено пространство допустимых частот дискретизации полосового сигнала в зависимости от отношения верхней граничной частоты спектра сигнала к полосе занимаемых им частот. Заштрихованные области соответствуют недопустимым значениям частоты дискретизации полосового сигнала.

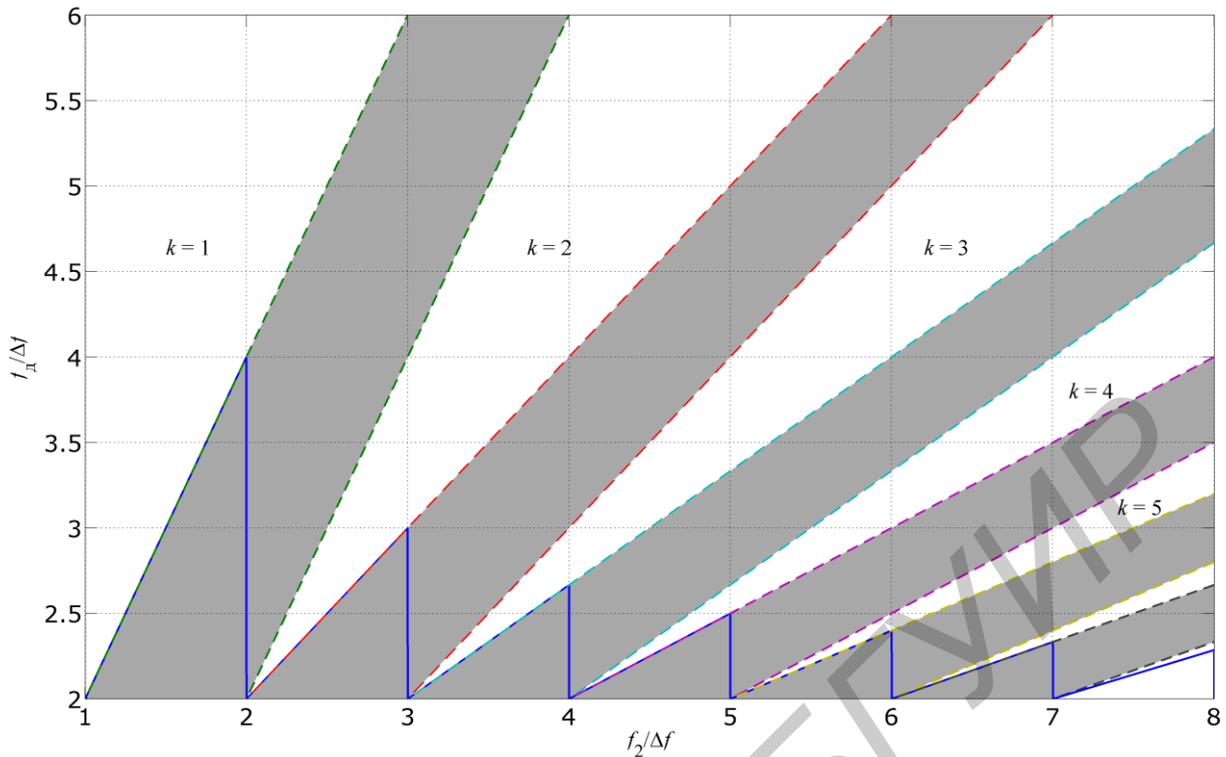


Рис. 4.3. Область допустимых значений частоты дискретизации полосового сигнала (незатонированные участки)

Если отношение $f_2/\Delta f$ является целым числом, то минимально возможная частота дискретизации полосового сигнала равна удвоенной полосе занимаемых частот, т. е. $f_d = 2\Delta f$. В этом случае фрагменты спектров $S_+(f)$ и $S_-(f)$ будут соприкасаться друг с другом и результирующий спектр будет сплошным на всём интервале частот. Если отношение $f_2/\Delta f$ не является целым, то частота дискретизации f_d может быть равна значению, взятому из некоторого интервала согласно рис. 4.3. Для большинства практических приложений значение f_d задаётся равным среднему арифметическому крайних допустимых значений. В результате формируется спектр дискретного сигнала с равными защитными интервалами между чередующимися участками $S_+(f)$ и $S_-(f)$.

4.1.5. Дискретизация сигналов в реальных системах

На практике реализовать последовательность дельта-импульсов невозможно, поэтому процесс дискретизации осуществляется в быстродействующих переключательных схемах, в которых в качестве опорных используются последовательности видеоимпульсов $x_p(t)$ определённой формы, чаще всего – прямоугольной. Рассматривают два типа дискретизации сигналов с ограниченным спектром в реальных системах.

Естественная дискретизация сигнала

Процедура естественной дискретизации сигнала $s(t)$ аналитически записывается следующим образом:

$$s_{дн}(t) = s(t)x_p(t), \quad (4.15)$$

где $x_p(t)$ – последовательность прямоугольных импульсов с единичной амплитудой, длительностью $\tau_{и}$ и периодом следования T .

В результате формируется дискретный сигнал, состоящий из последовательности импульсов с изменяющейся в течение интервалов времени $\tau_{и}$ амплитудой. Спектр опорной последовательности определяется функцией

$$X_p(\omega) = \sum_{k=-\infty}^{+\infty} C_k \delta(\omega - k\omega_d), \quad (4.16)$$

где $\omega_d = 2\pi/T$, а

$$C_k = \frac{\tau_{и}}{T} \frac{\sin[k\omega_d\tau_{и}/2]}{k\omega_d\tau_{и}/2} \exp\left[-\frac{jk\omega_d\tau_{и}}{2}\right].$$

Спектр дискретного сигнала

$$S_{дн}(\omega) = S(\omega) \otimes X_p(\omega) = \sum_{k=-\infty}^{+\infty} C_k S(\omega - k\omega_d). \quad (4.17)$$

Из этого следует, что спектр дискретного сигнала представляет собой чередующиеся через ω_d спектры исходного сигнала, умноженные на константу C_k . Следовательно, аналоговый сигнал может быть без искажений восстановлен из дискретного путём пропускания последнего через идеальный фильтр нижних частот.

Плосковершинная дискретизация сигналов осуществляется в устройствах выборки–хранения (УВХ). Дискретный сигнал для рассматриваемого случая определяется следующим выражением:

$$s_{дф}(t) = \sum_{k=-\infty}^{+\infty} s(kT)x_p(t - kT). \quad (4.18)$$

Здесь в отличие от естественной дискретизации значение амплитуды прямоугольного импульса не изменяется в течение интервала времени $\tau_{и}$ и равно значению сигнала $s(t)$ в момент взятия выборки. Используя выражение для идеального дискретного сигнала $s_d(t)$, можно записать:

$$s_{дф}(t) = x_p(t) \otimes \left[\sum_{k=-\infty}^{+\infty} s(kT)\delta(t - kT) \right] = x_p(t) \otimes s_d(t). \quad (4.19)$$

Используя свойство свёртки двух функций, найдём спектр дискретного сигнала:

$$S_{df}(\omega) = X_p(\omega)S_d(\omega) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} X_p(\omega)S(\omega - k\omega_d), \quad (4.20)$$

где

$$P(\omega) = \tau_{\text{и}} \frac{\sin[\omega\tau_{\text{и}}/2]}{\omega\tau_{\text{и}}/2} \exp\left[-\frac{j\omega\tau_{\text{и}}}{2}\right]. \quad (4.21)$$

Из полученного выражения следует, что плосковершинная дискретизация вносит амплитудные и фазовые искажения. Более высокие частотные составляющие дискретного сигнала ослабляются относительно низкочастотных. Указанный эффект называется эффектом апертурной погрешности. Апертурная погрешность при плосковершинной дискретизации сигнала может быть скорректирована эквалайзером с передаточной функцией $K_{\text{eq}}(\omega) = 1/X_p(\omega)$. Если $\tau_{\text{и}} \ll T$, то в пределах диапазона частот основного спектра $S(\omega)$ апертурные искажения незначительны и ими можно пренебречь.

4.1.6. Теорема отсчётов в частотной области

В силу частотно-временной дуальности преобразований Фурье теорему отсчётов можно сформулировать и для спектральной плотности сигнала, введя понятие сигнала, ограниченного во времени. Сигнал $s(t)$ называется ограниченным во времени, если $s(t) = 0$ всюду, за исключением интервала времени $|t| < T$.

Спектр $S(\omega)$ ограниченного во времени сигнала может быть точно представлен множеством своих значений $S(k\omega_d)$, взятых через равные интервалы частоты, не превышающие значения $\omega_d = \pi/T$:

$$S(\omega) = \sum_{k=-\infty}^{+\infty} S(k\omega_d) \frac{\sin[T(\omega - k\omega_d)]}{T(\omega - k\omega_d)}. \quad (4.22)$$

Иначе говоря, спектр, отображающий в частотной области сигнал конечной длительности, точно описывается множеством дискретных значений спектральной плотности, взятых через определённые равноотстоящие друг от друга интервалы частоты.

4.1.7. Выводы и практические рекомендации

Теорема В. А. Котельникова является «фундаментом», на котором базируется преобразование непрерывных сигналов в цифровую форму. Фактически аналоговый сигнал в виде напряжения между каким-либо узлами электронной цепи или тока на каком-либо её участке обладает информационной избыточностью, устранимой при представлении данного сигнала в виде по-

следовательности отсчётов. Сущность и практическую ценность теоремы В. А. Котельникова можно сформулировать двумя утверждениями:

– если сигнал имеет спектр, не содержащий значений выше некоторой частоты f_m Гц, то такой сигнал в полной мере описывается своими значениями – отсчётами, – взятыми через равные интервалы времени, не превышающие $1/2f_m$ с;

– сигнал может быть точно восстановлен по своим отсчётам, следующим друг за другом с частотой не менее $2f_m$ Гц.

Теорема Котельникова предполагает, что спектр сигнала строго ограничен. Известно, что сигнал с конечным спектром имеет бесконечную длительность и, соответственно, сигнал с конечной длительностью имеет спектр в бесконечно широкой полосе частот. Таким образом, конечная длительность большинства реально используемых сигналов вступает в противоречие с условием теоремы Котельникова. В связи с этим перед процессом дискретизации сигнала обязательно должна быть осуществлена процедура фильтрации сигнала.

4.2. Квантование по уровню

4.2.1. Место квантования сигнала в процедуре преобразования аналоговых сигналов в цифровые

Кодирование дискретного сигнала – это специальный тип аналого-цифрового преобразования, при котором значения дискретного сигнала в отсчётных точках кодируются последовательным двоичным кодом.

Полагается, что двоичный код состоит из последовательности слов объёмом n бит. Таким образом, количество уникальных кодовых слов $M = 2^n$. Каждое значение кодового слова соответствует некоторому известному значению выборки сигнала. Так как область допустимых значений выборок представляет собой континуум, то кодовые слова в действительности кодируют не конкретные значения, а некоторые интервалы значений выборок. Такие интервалы называют квантами, а сам дискретный сигнал предварительно подвергается квантованию. В результате квантования формируется дискретный сигнал, в котором значения выборок приравнены ближайшим разрешённым значениям – уровням квантования. Количество таких разрешённых значений не должно превышать объёма M кодовых слов.

Преобразование аналогового сигнала в цифровой предполагает осуществление трёх основных операций: дискретизации, квантования и собственно кодирования (см. рис. 4.4).

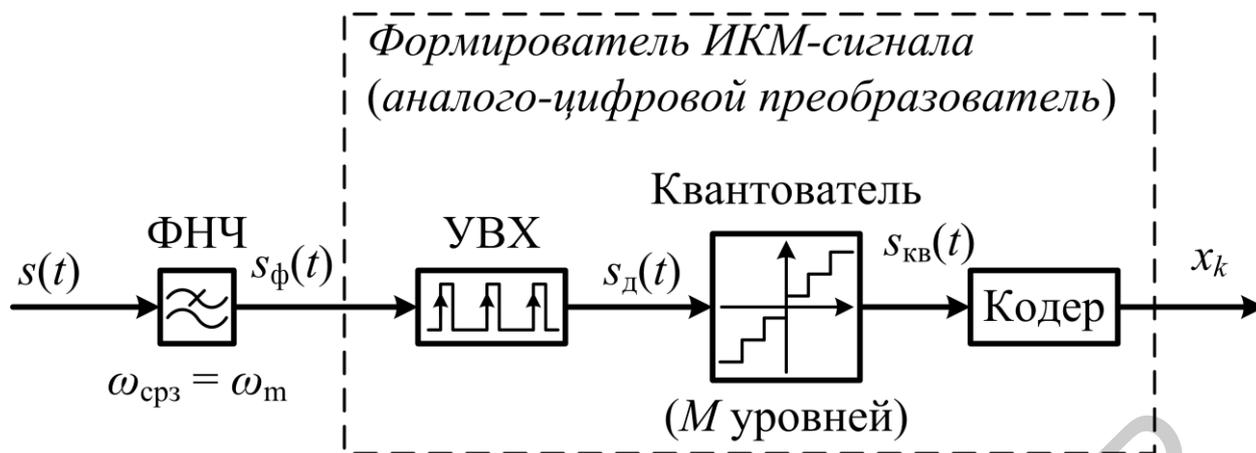


Рис. 4.4. Структурная схема формирователя цифрового сигнала

Операция дискретизации формирует последовательность равноотстоящих друг от друга плосковершинных прямоугольных импульсов, амплитуда которых равна мгновенному значению преобразовываемого сигнала в отсчётные моменты времени. Отсчётные моменты времени задаются либо одним из фронтов импульсов опорной последовательности, либо по некоторому уровню срабатывания УВХ.

Работа 8-уровневого квантователя входного аналогового сигнала показана на рис. 4.5. Сверху вниз показаны фрагмент реализации аналогового сигнала, квантованный сигнал и ошибка квантования. Приведённый квантователь называется равномерным, т. к. все кванты равны между собой (см. рис. 4.6, рис. 4.7). Чем больше уровней квантования, тем больше квантованный сигнал соответствует исходному аналоговому сигналу.

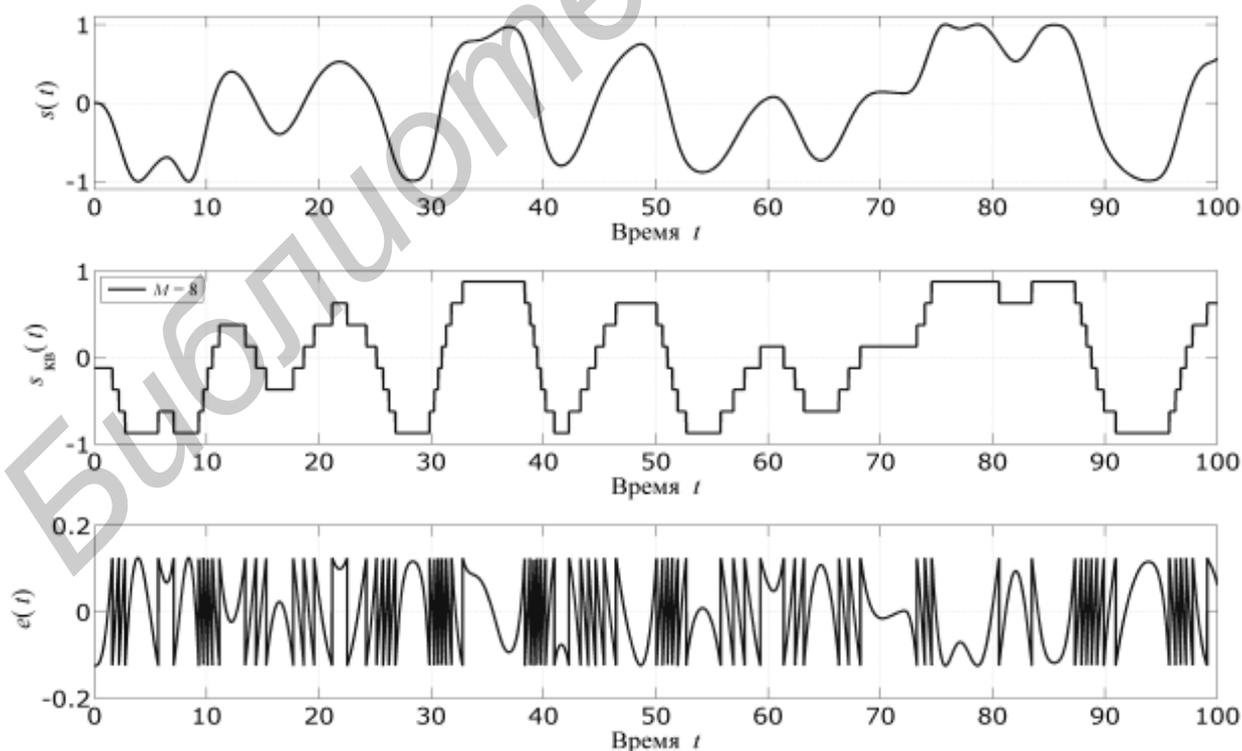


Рис. 4.5. Осциллограммы процесса квантования аналогового сигнала

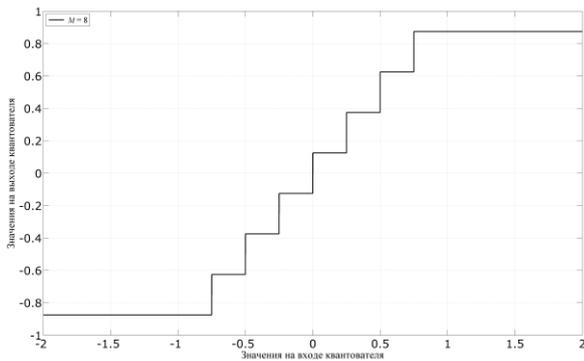


Рис. 4.6. Амплитудная характеристика квантователя с количеством уровней квантования $M = 8$

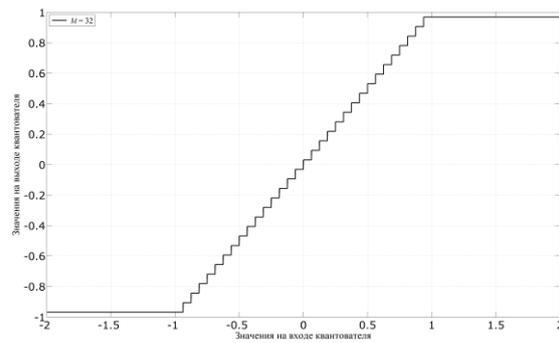


Рис. 4.7. Амплитудная характеристика квантователя с количеством уровней квантования $M = 32$

Ошибка $e(t) = s_{\text{кв}}(t) - s(t)$, неизбежно возникающая при квантовании, называется ошибкой квантования. Предельная ошибка при равномерном квантовании равна половине шага квантования:

$$e_{\text{max}} = \frac{s_{\text{max}} - s_{\text{min}}}{2M}, \quad (4.23)$$

где s_{min} и s_{max} – минимальное и максимально допустимое мгновенное значение квантуемого сигнала соответственно. Для примера на рис. 4.5 $e_{\text{max}} = 0,125$. Таким образом, даже при грамотном выборе частоты дискретизации сигнала квантование всегда вносит в цифровой сигнал аддитивные искажения, называемые цифровым шумом, или шумом квантования. Математически шум квантования является результатом округления точного значения отсчёта сигнала до ближайшего разрешённого уровня.

На выходе квантователя формируется сигнал с амплитудно-импульсной модуляцией. В практических приложениях важно, чтобы динамический диапазон сигнала не выходил за некоторый фиксированный предел, что требует нормировки аналогового сигнала. Чаще всего для удобства математических выкладок нормировка осуществляется по уровням ± 1 . Если сигнал выходит за допустимые пределы, то в квантованный сигнал вносится дополнительная помеха, обусловленная перегрузкой квантователя (см. рис. 4.8).

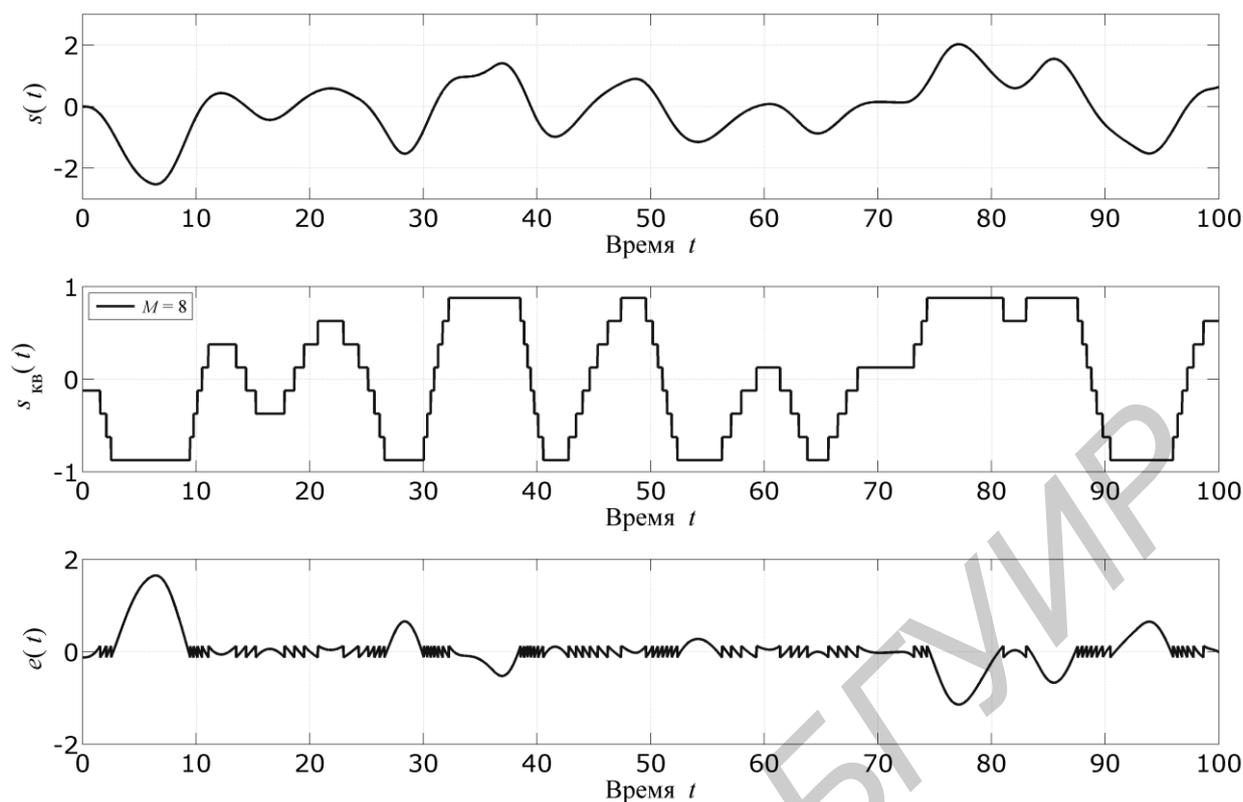


Рис. 4.8. Сущность эффекта перегрузки квантователя

Процедура дискретизации и квантования аналогового сигнала в полном объёме представлена на рис. 4.9. Сверху вниз показаны осциллограммы аналогового сигнала $s(t)$, дискретного во времени сигнала $s_{д}(t)$, дискретного во времени и квантованного по уровню сигнала $s_{д.кв}(t)$, а также шум квантования $e(t)$.

Каждое из квантованных значений дискретного во времени сигнала можно представить параллельным двоичным кодом, как показано на рис. 4.10. В целях возможности передачи параллельного двоичного кода по одному каналу код преобразовывается в последовательный.

Последовательный код формируется в течение интервала времени, не превышающего интервал дискретизации, путём последовательной передачи всех $\log(M)$ бит, кодирующих квантованное значение каждого отсчёта исходного аналогового сигнала.

Допустимо кодировать значения отсчётов в системе счисления, отличной от двоичной. Для основания системы счисления $q > 2$ количество уровней квантования $M = q^n$, где n – количество разрядов, кодирующих одно состояние. Разрешённые значения в этом случае будут кодироваться меньшим числом разрядов, однако такой вид цифрового сигнала практически нигде не используется, и в дальнейшем под цифровым сигналом будет подразумеваться только двоичный цифровой сигнал.

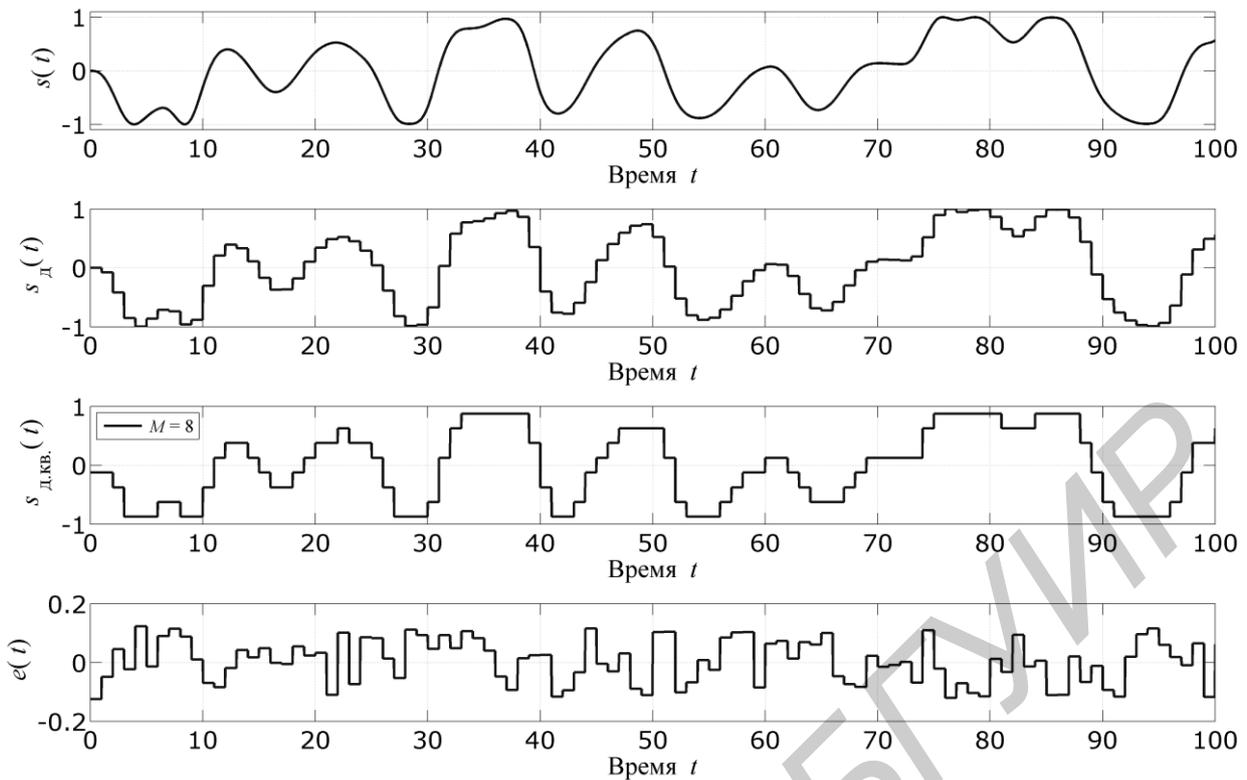


Рис. 4.9. Процесс формирования дискретного во времени квантованного сигнала

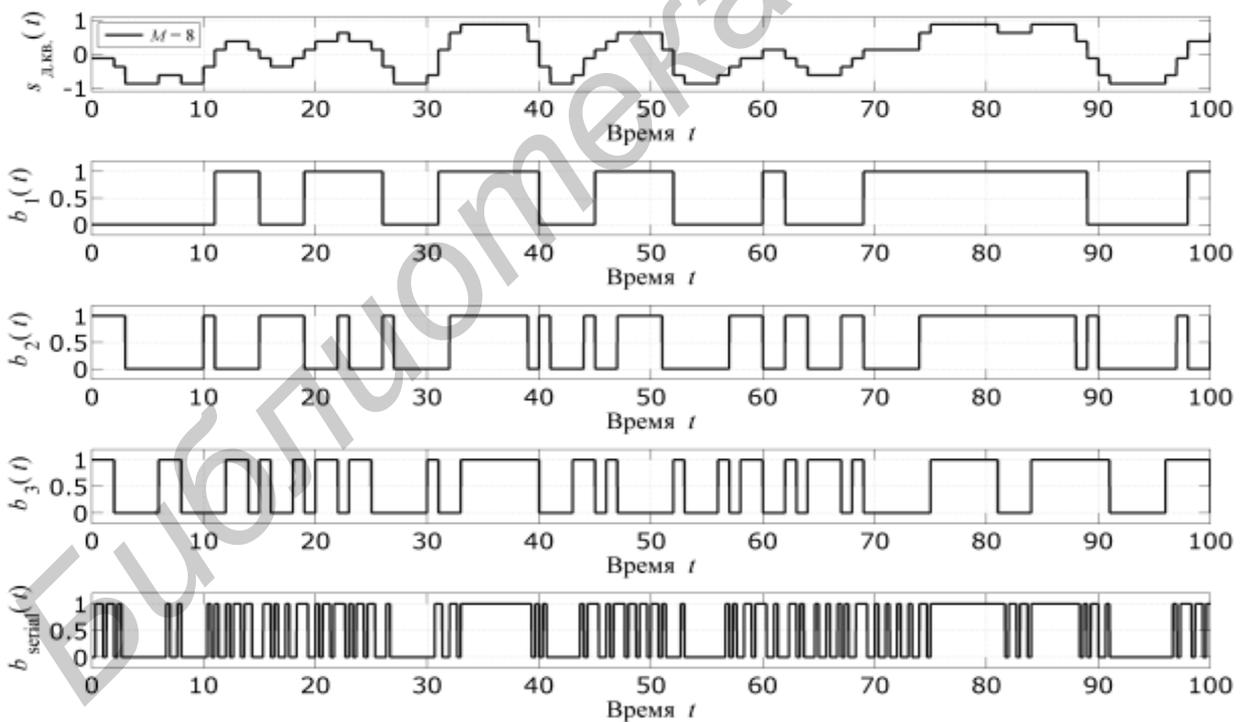


Рис. 4.10. Формирование сигнала с ИКМ:

$b_1(t)$ – первый, или старший, бит параллельного кода; $b_2(t)$ – второй бит параллельного кода; $b_3(t)$ – третий, или младший, бит параллельного кода; $b_{\text{serial}}(t)$ – последовательный код

4.2.2. Величина шума квантования

Квантователь определяет некоторый сигнал ошибки (см. сигнал $e(t)$ на рис. 4.5). Пиковое значение ошибки декодирования

$$P = \frac{1}{2M} = \frac{1}{2^{n+1}}, \quad (4.24)$$

где n – количество разрядов кодового слова.

Следовательно, для того чтобы пиковое значение не превышало P , необходимо обеспечить разрядность кода n из условия:

$$n \geq \log \frac{1}{P} - 1 = -\log P - 1 \approx -3,3221 \lg P - 1. \quad (4.25)$$

График функции для целых n представлен на рис. 4.11.

По приведённому графику можно определить минимально допустимую разрядность последовательного кода, обеспечивающую заданное пиковое значение ошибки.

4.2.3. Техническая реализация схем квантования по уровню

Схемы формирования цифрового сигнала состоят из аналого-цифрового преобразователя (АЦП) и при необходимости преобразователя параллельного кода в последовательный. Современные АЦП строятся по трём базовым схемам:

- схемы с последовательным счётом;
- схемы на основе последовательного приближения;
- схемы параллельного кодирования.

Все иные типы АЦП являются либо модификациями, либо комбинациями указанных трёх.

В момент взятия выборки в кодере с последовательным счётом одновременно с двоичным счётчиком запускается генератор линейно нарастающего напряжения (ГЛИН). Когда значение линейно нарастающего напряжения становится равным значению выборки кодируемого сигнала, с двоичного счётчика осуществляется съём данных в виде двоичного числа заданной разрядности. После этого происходит сброс счётчика и ГЛИН в состояние «нуль». Процедура повторяется в течение каждого интервала дискретизации T . Подход отличается простотой, требует минимума элементов, однако скорость преобразования сравнительно невысока, т. к. она определяется скоростью переключения состояний счётчика.

Кодер с последовательным приближением, или поиском, работает по принципу сравнения значений выборок с некоторыми пробными уровнями напряжения. Следующее пробное напряжение зависит от результата предыдущего, а интервал поиска уменьшается в два раза. В связи с тем, что поиск

начинается в середине области допустимых значений кодируемого сигнала, т. е. является грубым, а потом всё уточняется, результат достигается сравнительно быстро. Опорные уровни формируются последовательностью делителей напряжения, коммутируемых друг с другом электронными ключами. Коммутация осуществляется логическими схемами аппаратно. После того, как значение выборки не будет отличаться от сформированного опорного напряжения более, чем на половину кванта, осуществляется съём кода, или числа, кодирующего значение выборки. Подход требует большего числа прецизионных компонентов, таких как высокостабильные делители напряжения, однако скорость кодирования на порядок выше, чем у кодера с последовательным счётом, и определяется она скоростью работы коммутаторов.

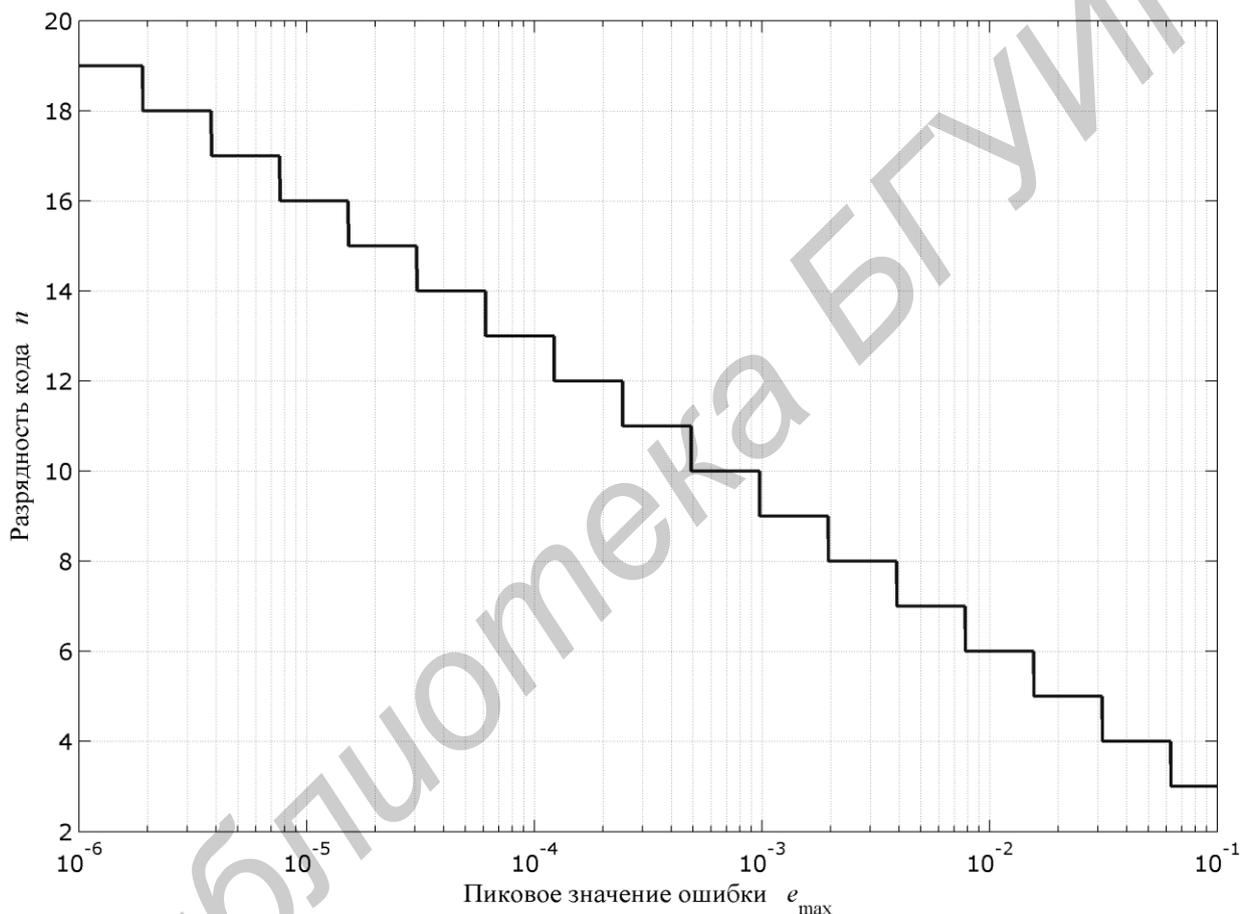


Рис. 4.11. Зависимость минимальной длины кодового слова для заданной величины ошибки декодирования цифрового сигнала

Параллельные кодеры включают в себя множество параллельно включённых компараторов, каждый из которых осуществляет сравнение выборки с одним конкретным значением опорного напряжения. Последовательность выборок подаётся одновременно на все компараторы. Множество высоких и низких уровней на выходе компараторов после процедуры двоичного преобразования (шифрации) определяет значение кодового слова ИКМ-сигнала. Подход отличается наивысшей скоростью преобразования, однако схемно наиболее громоздок.

4.2.4. Шумовые эффекты в цифровых системах

Результат декодирования цифрового сигнала представляет собой аддитивную смесь аналогового сигнала и шума. Возникновение шумовых искажений на выходе декодера обусловлено двумя факторами:

- 1) шумами квантования;
- 2) ошибками в символах цифрового сигнала (битовые ошибки).

Битовые ошибки возникают из-за помех в канале передачи информации и неравномерности амплитудно-частотной характеристики (АЧХ) канала в пределах полосы частот, занимаемой цифровым сигналом. При недостаточной фильтрации аналогового сигнала на входе кодера могут возникать дополнительные искажения за счёт перекрытия фрагментов спектра дискретного сигнала.

При определённых условиях отношение средней мощности аналогового сигнала к средней мощности шума на выходе декодера определяется выражением

$$q = \frac{M^2}{1 + 4(M^2 - 1)P_e}, \quad (4.26)$$

где P_e – вероятность появления ошибочного символа цифрового кода.

Как видно из (4.26), при отсутствии ошибок в цифровом сигнале качество декодирования определяется лишь количеством уровней квантования:

$$q_{max} = M^2. \quad (4.27)$$

С практической точки зрения шумы квантования в зависимости от условий работы декодера можно разделить на четыре группы:

- 1) шумы перегрузок;
- 2) случайные шумы;
- 3) гранулярные шумы;
- 4) шумы покоя.

Случайные шумы определяются собственно эффектом квантования выборок сигнала и наблюдаются при нормальном режиме работы квантователя, когда динамический диапазон аналогового сигнала и амплитудная характеристика квантователя согласованы.

В случае если динамический диапазон сигнала сравним с величиной кванта, наблюдается так называемый гранулярный шум. На выходе декодера наблюдается значительное ухудшение сигнал-шум, а аналоговый сигнал существенно огрубляется.

Шум покоя наблюдается при постоянных или медленно изменяющихся кодируемых сигналах. В этом случае на выходе декодера возникает некоторая осциллирующая с частотой $f_d/2$ составляющая. Шум покоя можно исключить путём режекции аналоговым фильтром спектральной компоненты

на частоте $f_d/2$ или исключением возможности переходов по уровню для неизменяющихся величин.

Если сигнал на входе кодера равен нулю, такой режим работы называется режимом простоя и в этом случае шум покоя можно исключить введением в амплитудную характеристику квантователя так называемой «мёртвой зоны» – горизонтальной ступеньки при малых входных напряжениях, в отличие от вертикальной ступеньки, как это показано на рис. 4.11.

Учитывая, что $M = 2^n$, выражение для отношения С/Ш можно записать в следующем виде:

$$q_{[\text{дБ}]} = 6,02n + \alpha. \quad (4.28)$$

где α – число, зависящее от структуры кода, формы элементарного символа, динамического диапазона и закона распределения мгновенных значений аналогового сигнала. Для цифровых символов прямоугольной формы и равномерного закона распределения аналогового сигнала $\alpha = 0$.

Из полученного выражения видно, что увеличение длины кода на один разряд увеличивает предельное отношение С/Ш приблизительно на 6 дБ. Так, для обеспечения отношения С/Ш на выходе декодера не хуже 96 дБ длину кода необходимо выбрать как минимум равной 16. Тогда полоса частот ИКМ-сигнала, определяемая первым нулём в спектре, будет в 32 раза шире полосы частот, занимаемой аналоговым сигналом.

4.2.5. Неравномерное квантование сигналов

Некоторые сигналы, например речевые, имеют плотность вероятности распределения мгновенных значений с максимумом вблизи нуля. Это значит, что наиболее вероятными являются не пиковые значения напряжений, а напряжения, имеющие малый уровень. Для таких сигналов с неравномерным распределением мгновенных значений гранулярный шум квантования является преобладающим с точки зрения качества систем передачи информации. В целях адаптации к такого рода сигналам широко используется неравномерное квантование выборок сигнала, подразумевающее уменьшение шага квантования для значений вблизи нуля.

Неравномерное квантование может быть осуществлено путём предварительного компандирования сигнала с последующей его подачей на обычный цифровой кодер. В результате компандирования происходит перераспределение мгновенных значений сигнала согласно некоторому закону. В Северной Америке и Японии в качестве компандирующих используются безынерционные нелинейные преобразователи, в которых амплитудная характеристика подчиняется μ -закону. В Европе применяется аналогичная ха-

рактеристика, подчиняющаяся несколько иному A -закону. Прямое компрессирование называется компрессией, обратное – экспандированием.

Характеристика μ -типа определена следующим образом:

$$y(t) = \text{sign}[x(t)] \frac{\ln(1 + \mu |x(t)|)}{\ln(1 + \mu)}, \quad (4.29)$$

где $x(t) \in (-1; 1)$ и $y(t) \in (-1; 1)$ – входной компрессируемый и выходной процесс соответственно; μ – константа, определяющая степень сжатия динамического диапазона сигнала.

Семейство графиков зависимости (4.29) при различных значениях параметра μ представлено на рис. 4.12.

Характеристика A -типа описывается следующей функциональной зависимостью:

$$y(t) = \begin{cases} \text{sign}[x(t)] \frac{A |x(t)|}{1 + \ln A}, & |x(t)| \in \left[0; \frac{1}{A}\right]; \\ \text{sign}[x(t)] \frac{1 + \ln[A |x(t)|]}{1 + \ln A}, & |x(t)| \in \left(\frac{1}{A}; 1\right], \end{cases} \quad (4.30)$$

где $A \geq 1$ – параметр нелинейной функции преобразования.

Семейство графиков зависимости (4.30) при различных значениях параметра A представлено на рис. 4.13.

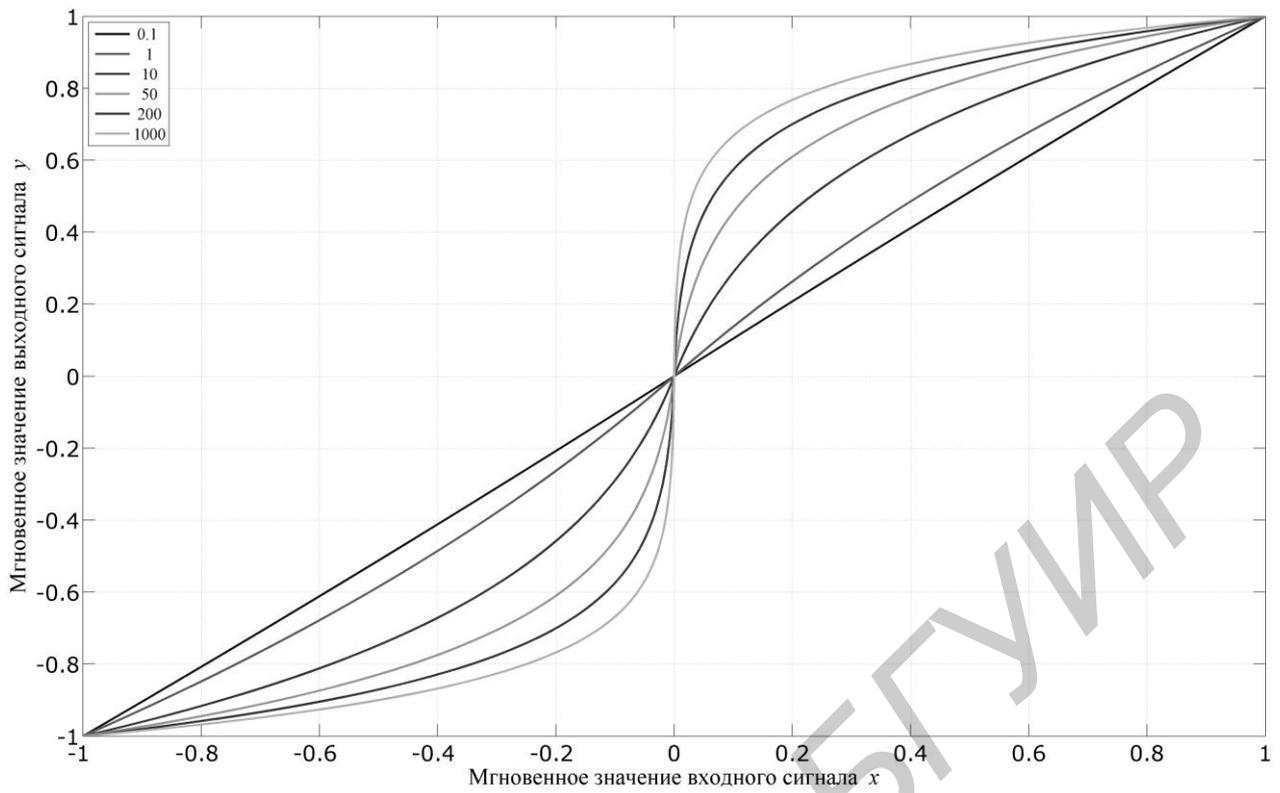


Рис. 4.12. Нелинейная характеристика μ -типа

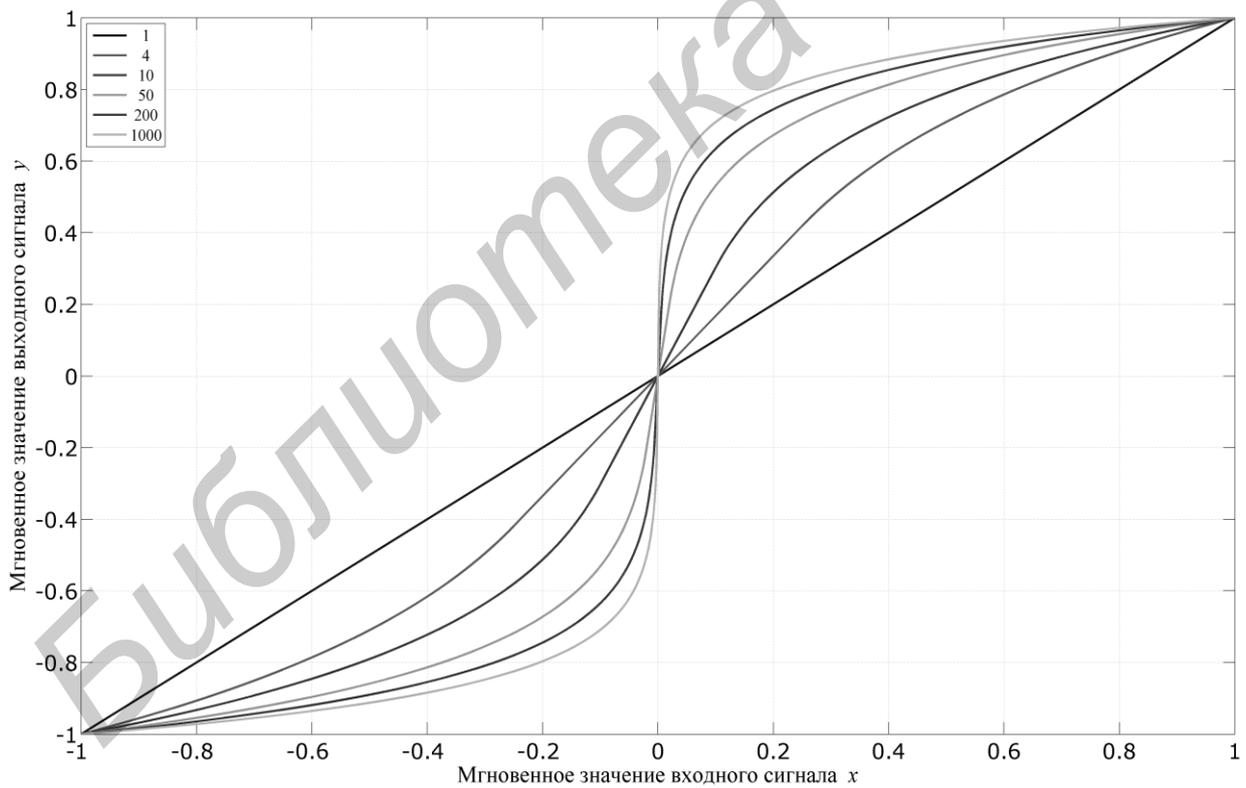


Рис. 4.13. Нелинейная характеристика A-типа

4.2.6. Дифференциальное квантование

Дифференциальное квантование имеет место, когда квантованию подвергаются не сами импульсы, а разница их амплитуд. Схема кодирования может изменять диапазон возможных значений, если разница между значениями амплитуд выходит за границы текущего диапазона возможных значений. В этом случае говорят об адаптивном дифференциальном квантовании. Такой подход позволяет учесть все возможные разницы амплитуд, но может привести к увеличению погрешности квантования.

Для успешного восстановления формы аналогового сигнала принимающей стороной достаточно информации об относительном изменении амплитуды двух соседних импульсов. Поэтому для реализации метода, называемого дельта-модуляцией, требуется только один бит, указывающий на направление изменения амплитуды.

При дельта-модуляции каждому биту соответствует целое значение амплитуды. В результате при передаче каждого значения амплитуды передается относительно небольшое количество информации об этом значении, что является основной причиной, по которой зернистость может изменить результат. Единственный способ, позволяющий удерживать оба типа искажений в допустимых пределах, – применение достаточно малого шага дискретизации. Соответственно, такой способ предполагает использование частоты дискретизации, превышающей частоту Найквиста. Для каждого значения амплитуды отводится 1 бит, поэтому битовая скорость соответствует скорости дискретизации. Этот метод разработан, чтобы сократить расход полосы пропускания, но не следует ради него отказываться от преимуществ сжатия.

Дельта-модуляция не всегда может эффективно решить поставленные перед ней задачи. Однако это под силу другой методике, основанной на дельта-модуляции. Она называется дельта-модуляцией с переменной крутизной наклона (CVSDM Continuously Variable Slope Delta Modulation). Метод CVSDM позволяет наблюдать за управлением битовым потоком. Для этого можно использовать либо последовательность единиц, увеличивающую размер шага для понижения наклона аналогового сигнала, либо последовательность нулей, уменьшающую размер шага для увеличения крутизны наклона и ослабления эффекта зернистости. В других адаптивных формах дельта-модуляции просто выполняется наблюдение за тенденциями посредством запоминания предыдущих направлений изменения амплитуды. Если для многих импульсов было замечено одно и то же направление изменения амплитуды, то можно предположить, что наклон сигнала является достаточно высоким. Чтобы уменьшить зернистость, нужно уменьшить шаг дискретизации. И наоборот, если чередуются отрицательные и положительные направления изменения амплитуды, это говорит о том, что наклон получился выше нормы. В этом случае шаг дискретизации увеличивается во избежание завышения наклона кривой.

В дополнение к временным методам, существуют и частотные методы кодирования источника информации. Они имеют ряд преимуществ. При ис-

пользовании, например SB-ADPCM, входной сигнал разбивается на несколько частотных полос, которые кодируются независимо друг от друга при помощи метода ADPCM. На принимающей стороне биты декодируются, в результате чего восстанавливается исходный полосовой сигнал.

Преимущество этого метода в том, что погрешность в каждом диапазоне частот зависит только от кодирования. Поэтому можно использовать более точные схемы кодирования и шаги квантования для тех диапазонов частот, которые наиболее важны для воспроизведения сигнала. Таким образом, погрешность в этих диапазонах частот невелика, в то время, как в других допустим более высокий уровень погрешности.

Между соседними отсчётами речевого сигнала имеется значительная корреляция, которая слабо убывает по мере увеличения интервала между отсчетами. Это означает, что речевой сигнал изменяется медленно и разность между соседними отсчетами будет иметь меньшую дисперсию, чем исходный сигнал, что позволяет применять методы разностного квантования речевого сигнала.

Линейная дельта-модуляция использует одnorазрядный (двухуровневый) квантователь и предсказатель. Восстановление аналогового сигнала из сигнала линейной ДМ осуществляется суммированием шага квантования.

Линейная дельта-модуляция технически реализуется относительно просто, но обладает рядом недостатков: перегрузка на крутизне; шум дробления (шум незанятого канала). Кроме того, для обеспечения приемлемого качества восстановления речевого сигнала требуется высокая скорость преобразования (передачи) порядка 200 кбит/с.

В адаптивной дельта-модуляции шаг квантования меняется в зависимости от крутизны исходного сигнала от минимального до максимального значения.

5. КОДИРОВАНИЕ ИСТОЧНИКА

5.1. Понятие префиксного кода

Префиксный код – это код со словом переменной длины, удовлетворяющий условию Фано (Роберт Фано 1917 г. р. – итало-американский учёный, известный специалист по теории информации).

Условие Фано. Если кодовое слово a принадлежит некоторому непустому кодовому множеству, то кодового слова ab в данном множестве не существует. Формулировка означает, что никакое кодовое слово не может быть началом другого кодового слова. Таким образом, несмотря на то что префиксный код состоит из кодовых слов разной длины, эти слова можно записывать без разделительного символа.

Примером кода, удовлетворяющего условию Р. Фано, являются телефонные номера. Если в сети существует номер 101, то номер 1012345 не может быть выдан: при наборе трёх цифр АТС прекращает «понимать» дальнейший набор и соединяет с адресатом по номеру 101. Однако для набора с сотового телефона это правило уже не действует, потому что требуется явное завершение последовательности знаков соответствующей кнопкой. Таким образом, 101, 1010 и 1012345 идентифицируются системой как различные.

Код, состоящий из слов «0», «10» и «11», является префиксным, и сообщение «01001101110» можно разбить на слова единственным образом: «0 10 0 11 0 11 10». Код, состоящий из слов 0, 10, 11 и 100, префиксным не является, и то же сообщение можно трактовать несколькими способами: «0 10 0 11 0 11 10» или «0 100 11 0 11 10». Иными словами, каждый элемент кодового множества является и кодовым словом, и префиксом, отделяющим кодовое слово от предыдущего и последующего кодового слова в информационном потоке. При формировании кодового множества недопустимые префиксы (кодовые слова) могут быть получены путём последовательного отбрасывания последнего знака определённой кодовой комбинации. Например, для кодовой комбинации «11101101» префиксами будут «1110110», «111011», «11101», «1110», «111», «11», «1».

Часть кодовой комбинации, дополняющая префикс до самой комбинации, называется суффиксом. Таким образом, множество кодовых слов может состоять либо из одних префиксов, либо из префиксов, объединённых с суффиксами.

Префиксные коды наглядно представляются с помощью кодовых деревьев. Если ни один узел кодового дерева не является вершиной данного кода, то он обладает свойствами префикса. Узлы дерева, которые не соединяются с другими, называются конечными. Комбинации, которые им соответствуют, являются кодовыми комбинациями префиксного кода. Любой код со словом фиксированной длины, очевидно, является префиксным.

5.2. Неравенство Крафта

Пусть заданы два произвольных конечных множества, которые называются соответственно кодируемым алфавитом и кодирующим алфавитом. Их элементы называются символами, а строки, или последовательности конечной длины, символов – словами. Длина слова – это число символов, из которого оно состоит.

В качестве кодирующего алфавита часто рассматривается множество $\{0, 1\}$ – двоичный алфавит.

Схемой алфавитного кодирования, или просто алфавитным кодом, называется любое отображение символов кодируемого алфавита в слова кодирующего алфавита, которые называют кодовыми словами. Пользуясь схемой кодирования, каждому слову кодируемого алфавита можно поставить его код – конкатенацию кодовых слов, соответствующих каждому символу этого слова.

Код называется разделимым, или однозначно декодируемым, если никаким двум словам кодируемого алфавита не может быть поставлен в соответствие один и тот же код.

Префиксным кодом называется алфавитный код, в котором ни одно из кодовых слов не является префиксом никакого другого кодового слова. Любой префиксный код является разделимым.

Теорема 5.1

Пусть заданы кодируемый и кодирующий алфавиты, состоящие из m и n символов соответственно и заданы желаемые длины кодовых слов: l_1, l_2, \dots, l_m . Тогда необходимым и достаточным условием существования разделимого и префиксного кодов, обладающих заданным набором длин кодовых слов, является выполнение неравенства

$$\sum_{i=1}^m n^{-l_i} \leq 1. \quad (5.1)$$

Это неравенство известно под названием неравенства Крафта–Макмиллана. Впервые оно было выведено Леоном Крафтом в его магистерской дипломной работе в 1949 г., однако он рассматривал только префиксные коды, поэтому при обсуждении префиксных кодов это неравенство часто называют просто неравенством Крафта. В 1956 г. Броквей Макмиллан доказал необходимость и достаточность этого неравенства для более общего класса кодов – разделимых кодов.

5.3. Алгоритм Хаффмана

В 1952 г. американский математик Дэвид Хаффман (1925–1999) предложил алгоритм формирования префиксного кода для компрессии без потерь (архивации).

Предположим, что нужно устранить информационную избыточность следующей символьной последовательности «ААВССD». При условии кодирования одного символа восемью битами эта последовательность занимает 7 байт. Компрессия по методу кодирования повторов даст следующий результат: «3A1B2C1D», т. е. объём сообщения стал равен 8 байтам. Алгоритм Хаффмана позволяет сократить объём исходного сообщения до 2 байт за счёт использования априорной информации о частоте появления той или иной буквы в сообщении.

Составим таблицу частот появления букв для нашего примера. Символы «А», «В», «С», «D» появляются в сообщении 3, 1, 2 и 1 раз соответственно. Затем эти сведения используются для построения так называемого двоичного дерева (рис. 5.1), используемого для генерации кодированной последовательности.

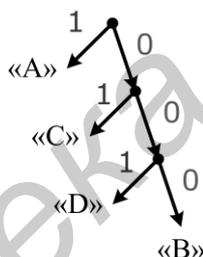


Рис. 5.1. Кодовое дерево Хаффмана

Левые ветви дерева отмечены кодом «0», а правые – «1». Слово, кодирующее определённый символ, образуется при прохождении от вершины дерева к символу. Таким образом, буква «А» кодируется единицей, буква «С» – последовательностью «01», D – последовательностью «001», а «В» – последовательностью «000», и различные символы кодируются кодовыми словами различной длины. Так как код Хаффмана является префиксным, то специального разделителя между кодовыми словами не требуется. Отсюда следует, что последовательность «1110000101001» кодирует текст «ААВВССD» тринадцатью битами.

Недостатки кода Хаффмана отмечены ниже.

1. Для последовательностей, представленных различными алфавитами и имеющими различную статистическую структуру, согласно методу необходимо генерировать различные двоичные деревья, что требует переформатирования дерева Хаффмана.

Например, в русскоязычных текстах чаще всего встречается буква «А», она будет стоять у вершины дерева и ей будет присвоен код, равный двоичной единице. В англоязычных текстах самая частая буква – «Е», и единич-

ный код будет присвоен ей. Следовательно, для различных типов данных применяется некоторое согласованное кодовое множество и в месте обработки кодированной последовательности требуется знание структуры кода.

Решение этой проблемы предполагает передачу кодовой таблицы до передачи основного информационного потока. Поэтому кодирование Хаффмана неэффективно для малых объёмов передаваемой информации.

В случае, если имеется несколько предопределённых кодовых структур, то имеется возможность передавать лишь идентификатор структуры. Разумеется, на приёмной стороне предопределённые структуры кода должны быть известны.

2. Существенно неравномерный вес символов в различных кодовых словах.

3. В случае, если кодовое слово имеет больший размер, чем может хранить стандартная ячейка памяти, возникает переполнение кодового слова.

5.4. Алгоритм Лемпеля–Зива–Велча (LZW)

Алгоритм Лемпеля – Зива – Велча (Lempel-Ziv-Welch, LZW) – это универсальный алгоритм сжатия данных без потерь, опубликованный Велчем в 1984 г. Алгоритм разработан так, чтобы его можно было быстро реализовать, но он не обязательно оптимален, поскольку при его реализации не проводится предварительный анализ входных данных, т. е. априорные сведения о кодируемой информации отсутствуют.

Данный алгоритм при сжатии (кодировании) динамически создаёт таблицу преобразования строк. Определённым последовательностям символов (словам) ставятся в соответствие группы бит фиксированной длины (обычно 12-битные). Таблица инициализируется всеми 1-символьными строками (в случае 8-битных символов – это 256 записей). По мере кодирования алгоритм просматривает сообщение символ за символом, и сохраняет каждую новую, уникальную 2-символьную строку в таблицу в виде пары код/символ, где код ссылается на соответствующий первый символ. После того как новая 2-символьная строка сохранена в таблице, на выход передаётся код первого символа. Когда на входе читается очередной символ, для него по таблице находится уже встречавшаяся строка максимальной длины, после чего в таблице сохраняется код этой строки со следующим символом на входе; на выход выдаётся код этой строки, а следующий символ используется в качестве начала следующей строки.

Алгоритму декодирования на входе требуется только закодированный текст, поскольку он может воссоздать соответствующую таблицу преобразования непосредственно по закодированному тексту.

Процедура LZW-кодирования подразумевает следующие операции.

1. Инициализация словаря всеми возможными односимвольными фразами. Инициализация входной фразы w первым символом сообщения.
2. Считать очередной символ k из кодируемого сообщения.
3. Если «конец сообщения», то выдать код для w .

4. Если фраза wk уже есть в словаре, присвоить входной фразе значение wk и перейти к шагу 2; иначе выдать код w , добавить wk в словарь, присвоить входной фразе значение k и перейти к шагу 2.

На момент своего появления алгоритм LZW давал лучший коэффициент сжатия для большинства приложений, чем любой другой хорошо известный метод того времени. Он стал первым широко используемым на компьютерах методом сжатия данных.

Пример

Необходимо закодировать следующую фразу, представленную латиницей, заглавными буквами и без пробелов: «KARLUKLARYUKRALKORALLYAKLARAUKARLAUKRALAKLARNET#». Символ решётки «#» означает конец сообщения.

Определим, что разрядность начального кода составляет 5 бит.

Этап 1. Задание основного словаря.

Таблица 5.1

Начальный словарь LZW-кодера

Символ	Порядковый номер символа	Код символа	Символ	Порядковый номер символа	Код символа
#	0	00000	N	14	01110
A	1	00001	O	15	01111
B	2	00010	P	16	10000
C	3	00011	Q	17	10001
D	4	00100	R	18	10010
E	5	00101	S	19	10011
F	6	00110	T	20	10100
G	7	00111	U	21	10101
H	8	01000	V	22	10110
I	9	01001	W	23	10111
J	10	01010	X	24	11000
K	11	01011	Y	25	11001
L	12	01100	Z	26	11010
M	13	01101			

При кодировании согласно табл. 5.5.1 сообщение длиной 48 символов займёт объём, равный 240 бит или 30 байт.

Этап 2. Кодирование сообщения и динамический синтез дополнительного словаря.

Таблица 5.2

Дополнительный словарь LZW-кодера

Группа символов	Порядковый номер группы	Код группы символов
KA	27	11011
AR	28	11100
RL	29	11101
LU	30	11110
UK	31	11111
KL	32	100000
LA	33	100001
ARY	34	100010
YU	35	100011
UKR	36	100100
RA	37	100101
AL	38	100110
LK	39	100111
KO	40	101000
OR	41	101001
RAL	42	101010
LL	43	101011
LY	44	101100
YA	45	101101
AK	46	101110
KLA	47	101111
ARA	48	110000
AU	49	110001
UKA	50	110010
ARL	51	110011
LAU	52	110100
UKRA	53	110101
ALA	54	110110
AKL	55	110111
LAR	56	111000
RN	57	111001
NE	58	111010
ET	59	111011
T#	60	111100

Сформированное согласно алгоритму сжатия Лемпеля–Зива–Велча сообщение будет представлено следующей символьной последовательностью: «K-A-R-L-U-K-0L-0AR-0Y-0UK-0R-0A-0L-0K-0O-RA-0L-0L-0Y-0A-KL-0AR-0A-0UK-0AR-LA-UKR-AL-AK-LA-0R-0N-0E-0T-0#». Нуль перед символами означает, что перед соответствующим 5-битным представлением символа необходимо передавать «0».

В двоичном коде сообщение будет выглядеть следующим образом: «01011-00001-10010-01100-10101-01011-001100-011100-011001-011111-010010-000001-001100-001011-001111-100101-001100-001100-011001-000001-100000-011100-000001-011111-011100-100001-100100-100110-101110-100001-010010-001110-000101-010100-000000» и занимать объём 204 бита. Осуществлена компрессия сообщения на 36 бит, или на 15 %. Если сообщение будет длиннее, то элементы словаря будут представлять всё более и более длинные части текста, благодаря чему повторяющиеся слова будут представлены компактно.

Для декодирования сообщения необходимо знание начального словаря. Последующие записи словаря реконструируются в процессе декодирования, т. к. они являются просто конкатенацией предыдущих записей.

Алгоритм LZW отличают высокая скорость работы как при компрессии, так и при декомпрессии, низкие требования к оперативной памяти и сравнительно простая аппаратная реализация. Кроме того, в отличие от алгоритма Хаффмана не требуется передача вместе с информацией сведений о кодовой структуре. Недостаток алгоритма – невысокая степень сжатия по сравнению с алгоритмами, в которых проводится предварительный анализ статистической структуры данных.

5.5. Алгоритм голосового кодирования HVXC

Вокодер (voice coder) HVXC обеспечивает удовлетворительное качество передачи при сверхнизких скоростях информационного потока – 2,0 и 4,0 кбит/с – и хорошее качество при скорости потока 9,6 кбит/с. Вокодер HVXC широко используется в цифровых системах радиовещания в диапазонах со сложной помеховой обстановкой. Аббревиатура HVXC расшифровывается как Harmonic Vector eXcitation Coding, что означает кодирование с гармоническим векторным возбуждением. Структурная схема кодера приведена на рис. 5.2.

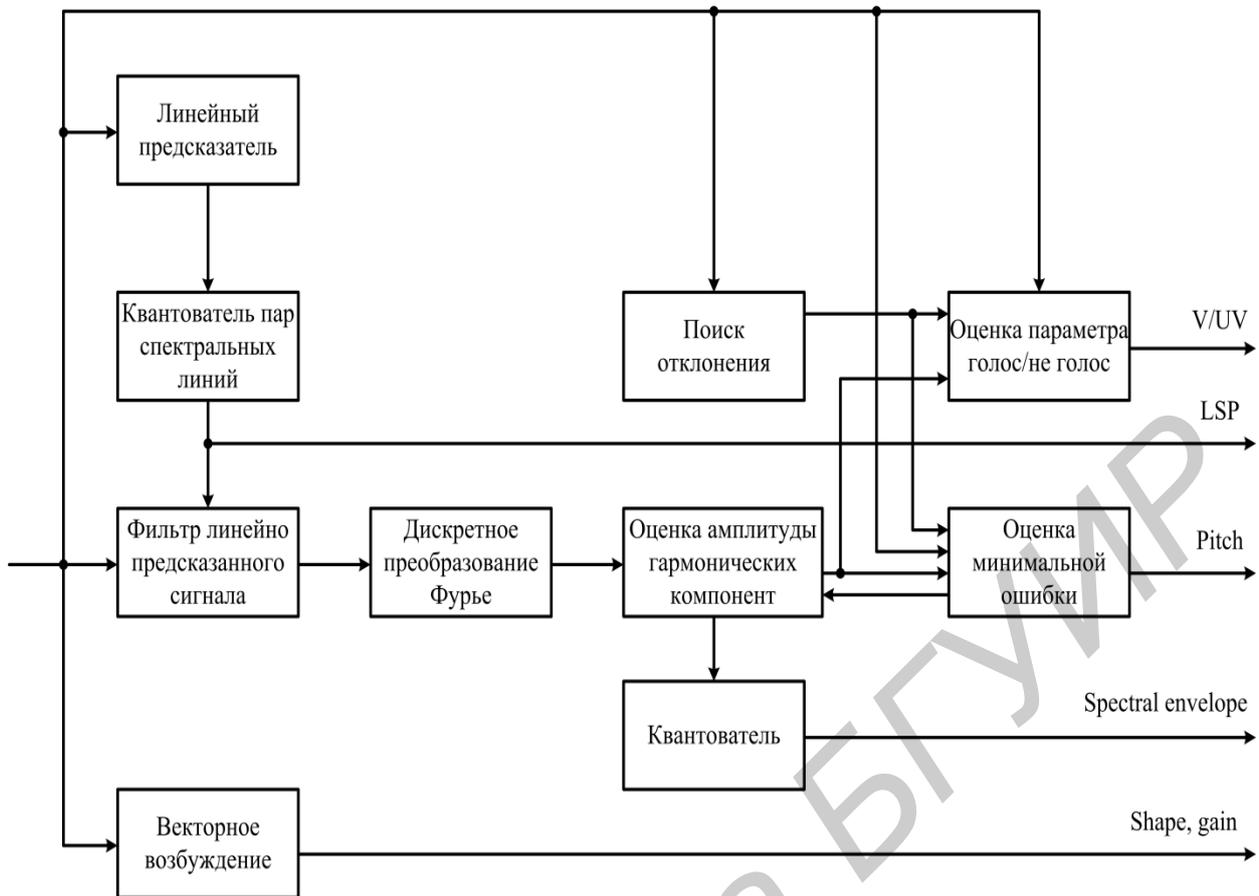


Рис. 5.2. Структурная схема голосового кодера HVXC

Поясним назначение функциональных блоков на рисунке.

1. Линейный предсказатель. Несжатый речевой сигнал с частотой дискретизации 8 кГц разбивается на фреймы длительностью 256 отсчётов и паузой между ними, равной 160 отсчётам. В блоке каждый фрейм подвергается линейному предсказанию, или экстраполяции десятого порядка. В результате формируется «синтетический» экстраполированный сигнал, который подаётся на квантователь пар спектральных линий.

2. Квантователь пар спектральных линий обеспечивает квантование параметров сигнала с частичным предсказанием и многоуровневой векторной схемой оптимального квантования.

3. Фильтр линейно предсказанного сигнала обеспечивает ограничение спектра входного сигнала, необходимое для корректного преобразования Фурье.

4. Блок векторного возбуждения кодирует голосовые (тональные) компоненты фрагменты входного сигнала.

5. Блок дискретного преобразования Фурье переносит сигнал в частотную область.

6. Блок поиска отклонения определяет предельные отклонения мгновенных значений сигнала от среднего.

7. Блок оценки амплитуды гармонических компонент определяет в спектре сигнала заданное количество наиболее значимых по мощности компонент.

8. Блок оценки «голос/не голос» на основе априорных данных о входном сигнале производит оценку его типа на заданном временном интервале.

9. Блок оценки минимальной ошибки осуществляет важную операцию по поиску оптимальной с точки зрения минимума среднеквадратичной ошибки отклонения синтетического сигнала.

На выходе вокодера формируются 5 низкоскоростных информационных потоков. Среди них: V/UV – voice/unvoice; LSP – line spectral pair; «Pitch» – отклонение; «Spectral envelope» – огибающая спектра; «Shape, gain» – форма и уровень сигнала. Сформированные вокодером потоки подвергаются последующему мультиплексированию.

Библиотека БГУИР

6. ОСНОВЫ КРИПТОГРАФИИ

Криптография – наука о методах обеспечения конфиденциальности (невозможности прочтения информации посторонним) и аутентичности (целостности и подлинности авторства, а также невозможности отказа от авторства) информации.

Изначально криптография изучала методы шифрования информации – обратимого преобразования открытого (исходного) текста на основе секретного алгоритма и/или ключа в зашифрованный текст (шифротекст). Традиционная криптография рассматривает симметричные криптосистемы, в которых зашифровывание и расшифровывание проводится с использованием одного и того же секретного ключа. Помимо этого современная криптография включает в себя асимметричные криптосистемы, системы электронной цифровой подписи (ЭЦП), хеш-функции, управление ключами, получение скрытой информации, квантовую криптографию.

6.1. Симметричные криптосистемы

Симметричные криптосистемы – способ шифрования, в котором для шифрования и расшифровывания применяется один и тот же криптографический ключ. До изобретения схемы асимметричного шифрования единственным существовавшим способом являлось симметричное шифрование. Ключ алгоритма должен сохраняться в секрете обеими сторонами. Алгоритм шифрования выбирается сторонами до начала обмена сообщениями.

Алгоритмы шифрования и дешифрования данных широко применяются в компьютерной технике в системах сокрытия конфиденциальной и коммерческой информации от злонамеренного использования сторонними лицами. Главным принципом в них является условие, что передатчик и приемник заранее знают алгоритм шифрования, а также ключ к сообщению, без которых информация представляет собой всего лишь набор символов, не имеющих смысла.

Классическим примером таких алгоритмов являются симметричные криптографические алгоритмы, перечисленные ниже:

- простая перестановка;
- одиночная перестановка по ключу;
- двойная перестановка;
- перестановка «Магический квадрат».

Простая перестановка без ключа – один из самых простых методов шифрования. Сообщение записывается в таблицу по столбцам. После того как открытый текст записан колонками, для образования шифровки он считывается по строкам. Для использования этого шифра отправителю и получателю нужно договориться об общем ключе в виде размера таблицы. Объединение букв в группы не входит в ключ шифра и используется лишь для удобства записи несмыслового текста.

Более практический метод шифрования, называемый одиночной перестановкой по ключу, очень похож на предыдущий. Он отличается лишь тем, что колонки таблицы переставляются по ключевому слову, фразе или набору чисел длиной в строку таблицы.

Для дополнительной скрытности можно повторно шифровать сообщение, которое уже было зашифровано. Этот способ известен под названием «двойная перестановка». Для этого размер второй таблицы подбирают так, чтобы длины ее строк и столбцов были другие, чем в первой таблице. Лучше всего, если они будут взаимно простыми. Кроме того, в первой таблице можно переставлять столбцы, а во второй строки. Наконец, можно заполнять таблицу зигзагом, змейкой, по спирали или каким-то другим способом. Такие способы заполнения таблицы если и не усиливают стойкость шифра, то делают процесс шифрования гораздо более интересным.

Магическими квадратами называются квадратные таблицы со вписанными в их клетки последовательными натуральными числами от 1, которые дают в сумме по каждому столбцу, каждой строке и каждой диагонали одно и то же число. Подобные квадраты широко применялись для вписывания шифруемого текста по приведенной в них нумерации. Если потом выписать содержимое таблицы по строкам, то получалась шифровка перестановкой букв. На первый взгляд кажется, будто магических квадратов очень мало. Тем не менее, их число очень быстро возрастает с увеличением размера квадрата. Так, существует лишь один магический квадрат размером 3×3 , если не принимать во внимание его повороты. Магических квадратов 4×4 насчитывается уже 880, а число магических квадратов размером 5×5 около 250 000. Поэтому магические квадраты больших размеров были хорошей основой для надежной системы шифрования, потому что ручной перебор всех вариантов ключа для этого шифра невыносим.

Рассмотрим пример шифрования короткого сообщения с использованием квадрата 4×4 .

16	3	2	13
5	10	11	8
9	6	7	12
4	15	14	1

Шифрование по магическому квадрату производится следующим образом. Например, требуется зашифровать фразу: «ПриезжаюСегодня.». Буквы этой фразы вписываются последовательно в квадрат согласно записанным в них числам: номер позиции буквы в предложении соответствует числу в магическом квадрате. В пустые клетки ставится точка.

16 .	3 и	2 р	13 д
5 з	10 е	11 г	8 ю
9 с	6 ж	7 а	12 о
4 е	15 я	14 н	1 п

После этого зашифрованный текст записывается в строку (считывание производится слева направо, построчно): «ирдзегюСжаоеянП».

При расшифровывании текст вписывается в квадрат, и открытый текст читается в последовательности чисел «магического квадрата».

В настоящее время известны следующие две разновидности симметричных шифров.

Блочные шифры. Обрабатывают информацию блоками определённой длины (обычно 64, 128 символов). К блоку применяется ключ в установленном порядке, как правило, несколькими циклами перемешивания и подстановки, называемыми раундами. Результатом повторения раундов является лавинный эффект – нарастающая потеря соответствия символов между блоками открытых и зашифрованных данных.

Поточные шифры. Шифрование проводится над каждым битом либо байтом исходного (открытого) текста с использованием гаммирования. Гаммирование (gamma хoring) – метод шифрования, основанный на «наложении» гамма-последовательности на открытый текст. Обычно это суммирование в каком-либо конечном поле (например, в поле GF(2) такое суммирование принимает вид обычного «исключающего ИЛИ»). При расшифровке операция проводится повторно, в результате получается открытый текст.

Большинство симметричных шифров используют сложную комбинацию большого количества подстановок и перестановок. Многие такие шифры выполняются в несколько (иногда до 80) проходов, используя на каждом проходе «ключ прохода». Множество «ключей прохода» для всех проходов называется «расписанием ключей» (key schedule). Как правило, оно создается из ключа выполнением над ним неких операций, в том числе перестановок и подстановок.

Операция перестановки перемешивает биты сообщения по некоему закону. В аппаратных реализациях она тривиально реализуется как перепутывание проводников. Именно операции перестановки дают возможность достижения «эффекта лавины». Операция перестановки линейна – $f(a) \text{ xor } f(b) = f(a \text{ xor } b)$.

Наиболее распространенные алгоритмы симметричных криптосистем:

- AES (Advanced Encryption Standard) – американский стандарт шифрования;
- ГОСТ 28147-89 – отечественный стандарт шифрования данных;
- DES (Data Encryption Standard) – стандарт шифрования данных в США до AES;
- 3DES (Triple-DES, тройной DES);
- RC6 (шифр Ривеста);
- Twofish;

- IDEA (International Data Encryption Algorithm);
- SEED – корейский стандарт шифрования данных;
- Camellia – сертифицированный для использования в Японии шифр;
- CAST (по инициалам разработчиков Carlisle Adams и Stafford Tavares);
- XTEA – наиболее простой в реализации алгоритм.

К основным недостаткам симметричных криптосистем следует отнести:

- сложность управления ключами в большой сети. Означает квадратичное возрастание числа пар ключей, которые надо генерировать, передавать, хранить и уничтожать в сети. Для сети в 10 абонентов требуется 45 ключей, для 100 – уже 4950, для 1000 – 499500 и т. д.;

- сложность обмена ключами. Для применения необходимо решить проблему надёжной передачи ключей каждому абоненту, так как нужен секретный канал для передачи каждого ключа обеим сторонам.

Для компенсации недостатков симметричного шифрования в настоящее время широко применяется комбинированная (гибридная) криптографическая схема, где с помощью асимметричного шифрования передаётся сеансовый ключ, используемый сторонами для обмена данными с помощью симметричного шифрования.

Важным свойством симметричных шифров является невозможность их использования для подтверждения авторства, так как ключ известен каждой стороне.

6.2. Криптосистема с открытым ключом

Криптографическая система с открытым ключом (или асимметричное шифрование, асимметричный шифр) – система шифрования, при которой открытый ключ передаётся по открытому (то есть незащищённому, доступному для наблюдения) каналу и используется для шифрования сообщения. Для расшифровки сообщения используется секретный ключ.

Идея криптографии с открытым ключом очень тесно связана с идеей односторонних функций, т. е. таких функций $f(x)$, что по известному x довольно просто найти значение $f(x)$, тогда как определение x из $f(x)$ невозможно за разумный срок.

Но сама односторонняя функция бесполезна в применении: ею можно зашифровать сообщение, но расшифровать нельзя. Поэтому криптография с открытым ключом использует односторонние функции с лазейкой. Лазейка – это некий секрет, который помогает при расшифровке. То есть существует такой y , что зная $f(x)$ и y , можно вычислить x . К примеру, если разобрать часы на множество составных частей, то очень сложно собрать вновь работающие часы. Но если есть инструкция по сборке (лазейка), то можно легко решить эту проблему.

Понять идеи и методы криптографии с открытым ключом помогает следующий пример – хранение паролей в компьютере. Каждый пользователь

в сети имеет свой пароль. При входе он указывает имя и вводит секретный пароль. Но если хранить пароль на диске компьютера, то кто-нибудь его может считать (особенно легко это сделать администратору этого компьютера) и получить доступ к секретной информации. Для решения задачи используется односторонняя функция. При создании секретного пароля в компьютере сохраняется не сам пароль, а результат вычисления функции от этого пароля и имени пользователя. Например, пользователь Алиса придумала пароль «Гладиолус». При сохранении этих данных вычисляется результат функции $f(\text{ГЛАДИОЛУС})$, пусть результатом будет строка РОМАШКА, которая и будет сохранена в системе. В результате файл паролей примет следующий вид:

Имя	$f(\text{имя_пароль})$
АЛИСА	РОМАШКА
БОБ	НАРЦИСС

Вход в систему теперь выглядит так:

Имя:	АЛИСА
Пароль:	ГЛАДИОЛУС

Когда Алиса вводит «секретный» пароль, компьютер проверяет, даёт ли функция, применяемая к ГЛАДИОЛУС, правильный результат РОМАШКА, хранящийся на диске компьютера. Стоит изменить хотя бы одну букву в имени или в пароле, и результат функции будет совершенно другим. «Секретный» пароль не хранится в компьютере ни в каком виде. Файл паролей может быть теперь просмотрен другими пользователями без потери секретности, так как функция практически необратима.

В предыдущем примере используется односторонняя функция без лазейки, поскольку не требуется по зашифрованному сообщению получить исходное. В следующем примере рассматривается схема с возможностью восстановить исходное сообщение с помощью «лазейки», т. е. труднодоступной информации. Для шифрования текста можно взять большой абонентский справочник, состоящий из нескольких толстых томов (по нему очень легко найти номер любого жителя города, но почти невозможно по известному номеру найти абонента). Для каждой буквы из шифруемого сообщения выбирается имя, начинающееся на ту же букву. Таким образом букве ставится в соответствие номер телефона абонента. Отправляемое сообщение, например «КОРОБКА», будет зашифровано следующим образом:

Сообщение	Выбранное имя	Криптотекст
К	Королёв	5643452
О	Орехов	3572651
Р	Рузаева	4673956
О	Осипов	3517289
Б	Батурин	7755628
К	Кирсанова	1235267
А	Арсеньева	8492746

Криптотекстом будет являться цепочка номеров, записанных в порядке их выбора в справочнике. Чтобы затруднить расшифровку, следует выбирать случайные имена, начинающиеся на нужную букву. Таким образом, исходное сообщение может быть зашифровано множеством различных списков номеров (криптотекстов).

Чтобы расшифровать текст, надо иметь справочник, составленный согласно возрастанию номеров. Этот справочник является лазейкой (секрет, который помогает получить начальный текст), известной только легальным пользователям. Не имея на руках копии справочника, криптоаналитик затратит очень много времени на расшифровку.

6.3. Криптографический протокол

Криптографический протокол (Cryptographic protocol) – это абстрактный или конкретный протокол, включающий набор криптографических алгоритмов. В основе протокола лежит набор правил, регламентирующих использование криптографических преобразований и алгоритмов в информационных процессах.

Существуют следующие основные виды криптографических протоколов:

- протоколы шифрования / расшифрования;
- протоколы электронной цифровой подписи (ЭЦП);
- протоколы идентификации / аутентификации;
- протоколы аутентифицированного распределения ключей.

Протоколы шифрования / расшифрования. В основе протокола этого класса содержится некоторый симметричный или асимметричный алгоритм шифрования/расшифрования. Алгоритм шифрования выполняется на передаче отправителем сообщения, в результате чего сообщение преобразуется из открытой формы в зашифрованную. Алгоритм расшифрования выполняется на приёме получателем, в результате чего сообщение преобразуется из зашифрованной формы в открытую. Так обеспечивается свойство конфиденциальности.

Для обеспечения свойства целостности передаваемых сообщений симметричные алгоритмы шифрования / расшифрования обычно совмещаются с

алгоритмами вычисления имитозащитной вставки (ИЗВ) на передаче и проверки ИЗВ на приёме, для чего используется ключ шифрования. При использовании асимметричных алгоритмов шифрования / расшифрования свойство целостности обеспечивается отдельно путем вычисления электронной цифровой подписи (ЭЦП) на передаче и проверки ЭЦП на приёме, чем обеспечиваются также свойства безотказности и аутентичности принятого сообщения.

Протоколы электронной цифровой подписи (ЭЦП). В основе протокола этого класса содержится некоторый алгоритм вычисления ЭЦП на передаче с помощью секретного ключа отправителя и проверки ЭЦП на приёме с помощью соответствующего открытого ключа, извлекаемого из открытого справочника, но защищенного от модификаций. В случае положительного результата проверки протокол обычно завершается операцией архивирования принятого сообщения, его ЭЦП и соответствующего открытого ключа. Операция архивирования может не выполняться, если ЭЦП используется только для обеспечения свойств целостности и аутентичности принятого сообщения, но не безотказности. В этом случае после проверки ЭЦП может быть уничтожена сразу или по прошествии ограниченного промежутка времени ожидания.

Протоколы идентификации/аутентификации. В основе протокола идентификации содержится некоторый алгоритм проверки того факта, что идентифицируемый объект (пользователь, устройство, процесс, ...), предъявивший некоторое имя (идентификатор), знает секретную информацию, известную только заявленному объекту, причем метод проверки является, конечно, косвенным, т. е. без предъявления этой секретной информации.

Обычно с каждым именем (идентификатором) объекта связывается перечень его прав и полномочий в системе, записанный в защищенной базе данных. В этом случае протокол идентификации может быть расширен до протокола аутентификации, в котором идентифицированный объект проверяется на правомочность заказываемой услуги.

Если в протоколе идентификации используется ЭЦП, то роль секретной информации играет секретный ключ ЭЦП, а проверка ЭЦП осуществляется с помощью открытого ключа ЭЦП, знание которого не позволяет определить соответствующий секретный ключ, но позволяет убедиться в том, что он известен автору ЭЦП.

Протоколы аутентифицированного распределения ключей. Протоколы этого класса совмещают аутентификацию пользователей с протоколом генерации и распределения ключей по каналу связи. Протокол имеет двух или трёх участников; третьим участником является центр генерации и распределения ключей (ЦГРК), называемый для краткости сервером S. Протокол состоит из трёх этапов: генерации, регистрации и коммуникации. На этапе генерации сервер S генерирует числовые значения параметров системы, в том числе свой секретный и открытый ключ. На этапе регистрации сервер S идентифицирует пользователей по документам (при личной явке или через уполномоченных лиц), генерирует ключевую и/или идентификационную информацию для каждого объекта и формирует маркер безопасности, содержа-

щий необходимые системные константы и открытый ключ сервера S (при необходимости). На этапе коммуникации реализуется собственно протокол аутентифицированного ключевого обмена, который завершается формированием общего сеансового ключа.

6.4. Управление ключами

Управление ключами состоит из процедур, обеспечивающих:

- включение пользователей в систему;
- выработку, распределение и введение в аппаратуру ключей;
- контроль использования ключей;
- смену и уничтожение ключей;
- архивирование, хранение и восстановление ключей.

Управление ключами как основа для обеспечения конфиденциальности обмена информацией, идентификации и целостности данных играет важнейшую роль в криптографии. Важным свойством хорошо спроектированной системы управления ключами является сведение сложных проблем обеспечения безопасности многочисленных ключей к проблеме обеспечения безопасности нескольких ключей, которая может быть относительно просто решена путем обеспечения их физической изоляции в выделенных помещениях и защищенном от проникновения оборудовании. В случае использования ключей для обеспечения безопасности хранимой информации субъектом может быть единственный пользователь, который осуществляет работу с данными в последовательные промежутки времени. Управление ключами в сетях связи включает, по крайней мере, двух субъектов – отправителя и получателя сообщения.

Целью управления ключами является нейтрализация таких угроз, как:

- компрометация конфиденциальности закрытых ключей;
- компрометация аутентичности закрытых или открытых ключей. При этом под аутентичностью понимается знание или возможность проверки идентичности корреспондента, для обеспечения конфиденциальной связи с которым используется данный ключ;
- несанкционированное использование закрытых или открытых ключей, например, использование ключа, срок действия которого истек.

Сроки действия ключей. Одной из важных характеристик системы управления ключами являются сроки действия ключей. Срок действия ключа означает промежуток времени, в течение которого он может быть использован доверенными сторонами.

Сокращение сроков действия ключей необходимо для достижения следующих целей:

- ограничения объема информации, зашифрованной на данном ключе, которая может быть использована для криптоанализа;
- ограничения размера ущерба при компрометации ключей;
- ограничения объема машинного времени, которое может быть использовано для криптоанализа.

В дополнение к указанной выше классификации ключей по уровням может быть введена также следующая классификация.

– ключи с длительным сроком действия. К ним относится главный ключ, часто – ключи для шифрования ключей;

– ключи с коротким сроком действия. К ним относятся ключи для шифрования данных.

Как правило, в телекоммуникационных приложениях используются ключи с коротким сроком действия, а для защиты хранимых данных – с длительным сроком действия. Необходимо иметь в виду, что термин «короткий срок действия» относится только к сроку действия ключа, а не к промежутку времени, в течение которого ключ должен оставаться в секрете. Например, к ключу, используемому для шифрования в течение только одного сеанса связи, часто предъявляется требование, чтобы зашифрованная на нём информация не могла быть вскрыта на протяжении нескольких десятков лет. В то же время электронная подпись проверяется немедленно после передачи сообщения, поэтому ключ подписи должен сохраняться в тайне в течение достаточно короткого срока.

Жизненный цикл ключей. Ключевая информация должна быть сменена до момента истечения срока действия ключа. Для этого могут быть использованы действующая ключевая информация, протоколы распределения ключей и ключевые уровни. Для того чтобы снизить ущерб от компрометации ключей, следует избегать зависимостей между действующей и устанавливаемой ключевой информацией. Например, не рекомендуется защищать очередной сеансовый ключ с помощью действующего сеансового ключа. При хранении закрытых ключей должны быть приняты меры по обеспечению их конфиденциальности и аутентичности. При хранении открытых ключей должны быть приняты меры, позволяющие проверить их аутентичность. Конфиденциальность и аутентичность могут быть обеспечены криптографическими, организационными и техническими мерами.

Все криптосистемы, за исключением простейших, в которых используемые ключи зафиксированы раз и навсегда, нуждаются в периодической замене ключей. Эта замена проводится с помощью определенных процедур и протоколов, в ряде которых используются и протоколы взаимодействия с третьей стороной. Приведём последовательность стадий, которые проходят ключи от момента установления до следующей замены (жизненный цикл ключей).

1. Регистрация пользователей. Эта стадия включает обмен первоначальной ключевой информацией, такой, как общие пароли или PIN-коды, путем личного общения или пересылки через доверенного курьера.

2. Инициализация. На этой стадии пользователь устанавливает аппаратное оборудование и/или программные средства в соответствии с установленными рекомендациями и правилами.

3. Генерация ключей. При генерации ключей должны быть приняты меры по обеспечению их необходимых криптографических качеств. Ключи могут генерироваться как самостоятельно пользователем, так и специальным

защищенным элементом системы, а затем передаваться пользователю по защищенному каналу.

4. Установка ключей. Ключи устанавливаются в оборудование тем или иным способом. При этом первоначальная ключевая информация, полученная на стадии регистрации пользователей, может либо непосредственно вводиться в оборудование, либо использоваться для установления защищенного канала, по которому передается ключевая информация. Эта же стадия используется в последующем для смены ключевой информации.

5. Регистрация ключей. Ключевая информация связывается регистрационным центром с именем пользователя и сообщается другим пользователям ключевой сети. При этом для открытых ключей создаются сертификационным центром ключевые сертификаты, и эта информация публикуется тем или иным способом.

6. Обычный режим работы. На этой стадии ключи используются для защиты информации в обычном режиме.

7. Хранение ключа. Эта стадия включает процедуры, необходимые для хранения ключа в надлежащих условиях, обеспечивающих его безопасность до момента его замены.

8. Замена ключа. Замена ключа осуществляется до истечения его срока действия и включает процедуры, связанные с генерацией ключей, протоколами обмена ключевой информацией между корреспондентами, а также с доверенной третьей стороной. Для открытых ключей эта стадия обычно включает обмен информацией по защищенному каналу с сертификационным центром.

9. Архивирование. В отдельных случаях ключевая информация после её использования для защиты информации может быть подвергнута архивированию для её извлечения со специальными целями (например, рассмотрения вопросов, связанных с отказами от цифровой подписи).

10. Уничтожение ключей. После окончания сроков действия ключей они выводятся из обращения, и все имеющиеся их копии уничтожаются. При этом необходимо следить, чтобы в случае уничтожения закрытых ключей тщательно уничтожалась и вся информация, по которой возможно их частичное восстановление.

11. Восстановление ключей. Если ключевая информация уничтожена, но не скомпрометирована (например, из-за неисправности оборудования или из-за того, что оператор забыл пароль), должны быть предусмотрены меры, дающие возможность восстановить ключ из хранимой в соответствующих условиях его копии.

12. Отмена ключей. В случае компрометации ключевой информации возникает необходимость прекращения использования ключей до окончания срока их действия. При этом должны быть предусмотрены необходимые меры оповещения абонентов сети. При отмене открытых ключей, снабженных сертификатами, одновременно производится прекращение действия сертификатов.

7. ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ В СИСТЕМАХ ЦИФРОВОЙ РАДИОСВЯЗИ

7.1. Необходимость канального кодирования

Общее требование, предъявляемое к любой СПИ, состоит в достоверном и своевременном получении заданного объёма информации из излучений с ограниченной энергией. Ошибочные, запоздалые или неполные сведения обесценивают полученную информацию, так как не позволяют оперативно принимать правильные решения.

Достоверному приёму информации по реальным радиоканалам препятствуют следующие факторы:

- случайные искажения радиосигнала при распространении электромагнитных волн;
- неизбежное наличие внешних и внутренних помех;
- техническое несовершенство устройств передачи, приёма и обработки сигналов – переносчиков информации.

Улучшение качества приёма из-за технического несовершенства аппаратуры в принципе может быть достигнуто путём её улучшения. Что же касается помех и искажений радиосигнала, то нивелирование их влияния носит принципиальный характер, обусловленный характеристиками сигнала, радиоканала и приёмника. Радиоканал является одним из самых сложных каналов передачи информации в смысле обеспечения заданного качества передачи информации. Использование СПИ в различных сферах жизнедеятельности человека ставит дополнительные требования к защите информации от попыток несанкционированного доступа к ней.

В настоящее время фундаментальные задачи передачи и приёма радиосигналов сводятся к следующим трём:

- 1) генерация сигналов, обеспечивающих повышенную помехозащищённость, скрытность и защиту информации от несанкционированного доступа;
- 2) создание больших ансамблей сигналов с «хорошими» взаимокорреляционными свойствами (для систем с множественным доступом);
- 3) оптимальный приём радиосигналов и выделение сообщений в каналах с помехами.

Во многих практических случаях приём сигналов должен осуществляться при очень малых отношениях сигнал/помеха в силу ограниченной мощности передатчика, а в ряде случаев отношение сигнал/шум, меньшее единицы, выбирается преднамеренно.

Основной задачей помехоустойчивого кодирования является решение проблемы обеспечения заданной достоверности передаваемых данных за счёт применения специальных устройств кодирования–декодирования в составе систем цифровой радиосвязи.

К наиболее эффективным методам обеспечения высокого качества цифровой передачи в условиях высокого уровня шума канала относятся уже весьма мощные в настоящее время алгоритмы кодирования/декодирования с использованием корректирующих кодов. За 50 лет своего интенсивного развития в технику связи были успешно внедрены пороговые декодеры, алгоритм Витерби, коды Рида–Соломона, каскадные схемы кодирования, а также разработки последнего времени: алгоритмы для турбокодов и многопороговые декодеры (МПД). Требования к алгоритмам исправления ошибок в каналах с шумами, в частности, спутниковых каналах, непрерывно растут, и главная проблема – декодирование с эффективностью, близкой к оптимальной по энергетике канала, но при максимально простой реализации, – еще до недавнего времени была весьма далека от своего успешного решения.

Технологической основой разработки и внедрения высокоэффективных методов канального кодирования является микроэлектроника. В настоящее время практически нет ограничений на вычислительную сложность канальных кодеров-декодеров (кодеков), как, например, при турбокодировании.

Другим, еще более эффективным подходом к решению проблемы синтеза высокоэффективных декодеров являются многопороговые декодеры. Указанный метод широко используется как для блочных, так и для свёрточных кодов, в том числе и недвоичных. В настоящее время алгоритмы типа МПД могут декодировать длинные коды почти так же, как и оптимальные переборные алгоритмы. Однако сами МПД сохраняют при этом линейную сложность реализации. При одинаковой эффективности МПД выполняет примерно в 100 раз меньшее число операций, чем соответствующие турбодекодеры. При аппаратной реализации МПД может быть в некоторых случаях более быстрым, чем турбодекодеры, почти на 3 десятичных порядка.

7.2. Общие принципы помехоустойчивого кодирования и декодирования

Применение кодов, исправляющих ошибки, является эффективным средством повышения достоверности передачи информации при сохранении неизменными скорости передачи и энергетических параметров сигнала или при снижении отношения сигнал/шум, требуемого для обеспечения заданной достоверности приема информации. Так, современные помехоустойчивые коды позволяют снизить отношение сигнал/шум до 6–8 дБ. Для примера – в космических или спутниковых системах снижение отношения сигнал/шум на 1 дБ эквивалентно снижению стоимости аппаратуры на 1 млн дол. США.

Теоретически эффективность помехоустойчивого кодирования была доказана Шенноном еще в 1948 г. Он показал, что если скорость передачи информации от некоторого источника меньше пропускной способности канала связи, то, применяя избыточное кодирование, можно обеспечить безошибочную передачу сообщения через канал связи. Внедрение помехоустой-

чивого кодирования сдерживалось отсутствием необходимых аппаратных средств.

Теория помехоустойчивого кодирования базируется на достаточно сложном математическом аппарате, в частности, на высшей алгебре.

Основу помехоустойчивого кодирования составляет введение в передаваемые сообщения некоторой избыточности. Например, в нашей речи вероятны ошибки, но это не мешает правильному пониманию содержания, если этих ошибок немного. Это касается как самих слов, так и связности слов в предложениях. Наша речь избыточна.

Введение избыточных символов пропорционально длине блока, т. е. относительное количество избыточных символов в блоках различной длины не зависит от длины блока. Чем длиннее блок, тем больше ошибок может быть исправлено и тем выше помехозащищенность передаваемого сообщения. Но с ростом длины блока растет и сложность декодера, и задержка при передаче сигнала. Большое влияние на выбор кода и алгоритма его декодирования оказывает канал связи.

Три модели каналов связи перечислены ниже.

1. Двоичный симметричный канал (ДСК) без памяти.
2. Дискретный канал без памяти.
3. Дискретный канал с памятью.

В первом случае на входе модулятора и выходе демодулятора имеются только двоичные символы (0,1) и демодулятор принимает жесткие решения – 0,1. В дискретном канале без памяти на входе имеем p -ичные символы, а на выходе – q -ичные. Дискретный канал с памятью – это канал с замираниями. Кодирование представляет собой преобразование последовательности информационных символов $\{a_i\}$ в последовательность закодированных символов $\{b_i\}$. Наиболее наглядно такое преобразование можно представить в табличном виде.

a_i	b_i
00	100
01	010
10	001
11	111

Рассмотрим общие принципы кодирования. Пусть на вход канала поступает некий сигнал $S(t)$, а на выходе имеем уже другой сигнал $S^*(t)$. Пусть эти сигналы являются элементами некоторого множества. На этом множестве вводится метрика $D(S^*, S_j)$, т. е. ищется минимальное расстояние между любыми парами сигналов $S_i(t)$ и $S_j(t)$ – элементами данного множества. Выбирается та пара сигналов, где эта метрика минимальна.

$$D(S, S^*) = \int_0^T [S(t) - S^*(t)]^2 dt. \quad (7.1)$$

Выражение (7.1) наиболее справедливо, если в канале связи рассматривается аддитивный белый гауссовский шум. Декодер, минимизирующий функцию (7.1), называют декодером максимального правдоподобия.

Вычислить среднеквадратичную метрику для всего ансамбля сигналов $S_i(t)$ достаточно сложно; индекс i принадлежит интервалу $0 \leq i \leq q^k$, q – основание кода, k – количество информационных символов в кодовой последовательности. Иногда для упрощения процедуры вычисления функции (7.1) используют не сами сигналы $S_i(t)$ и $S_i^*(t)$, а их цифровые представления. Дальнейшее упрощение процедуры вычисления – это $q = 2$. Такие двоичные решения на практике называют жесткими.

Принятие демодулятором жестких решений соответствует модели ДСК. Получаемая при таком методе приема метрика равна числу несовпадений символов между принимаемой последовательностью и опорной и называется метрикой Хемминга, или хемминговым расстоянием.

В процессе работы кодер помехоустойчивого кода в соответствии с некоторыми правилами осуществляет обработку входных символов. В зависимости от количества возможных значений Q каждого из символов все коды можно разделить на двоичные – при $Q = 2$ – и недвоичные – при $Q > 2$.

Линейные коды образуют векторное пространство и обладают следующим важным свойством: два кодовых слова можно сложить, используя подходящее определение суммы, и получить третье кодовое слово. В случае обычных двоичных кодов эта операция является посимвольным сложением двух кодовых слов по модулю 2. Данное свойство существенно упрощает процедуры кодирования и декодирования, а также задачу вычисления параметров кода, поскольку минимальное расстояние между двумя кодовыми словами при этом эквивалентно минимальному расстоянию между кодовым словом, состоящим целиком из нулей, и некоторым другим кодовым словом. Таким образом, при вычислении характеристик линейного кода достаточно рассмотреть, что происходит при передаче кодового слова, состоящего целиком из нулей.

7.3. Классификация помехоустойчивых кодов

В настоящее время разработаны и широко используются несколько различных классов помехоустойчивых кодов, отличающихся друг от друга структурой, функциональным назначением, энергетической эффективностью, алгоритмами кодирования и декодирования. Помехоустойчивые коды условно могут быть разделены на следующие классы:

- 1) блочные и свёрточные (древовидные) коды;

- 2) линейные и нелинейные коды;
- 3) систематические и несистематические коды;
- 4) циклические и нециклические коды;
- 5) двоичные и недвоичные коды;
- 6) исправляющие независимые ошибки и исправляющие пакетные ошибки;
- 7) каскадные и некаскадные коды.

Представленную выше классификацию можно разбить на две группы кодов: блочные и древовидные.

Блочный кодер – устройство без памяти, преобразующее одну последовательность в другую. Кодер древовидного кода – устройство с памятью.

Блочный код представляет собой множество последовательностей символов, которые называют кодовыми словами. При использовании двоичных блочных кодов непрерывная последовательность информационных бит разбивается на блоки по k бит. Все дальнейшие операции по кодированию и декодированию производятся с блоками. Каждому информационному блоку из k бит ставится в соответствие блок из n символов, где $n > k$. Этот блок из n символов называют кодовым словом B_i , которое передается по каналу связи, искажается и декодируется независимо от других блоков. Величину n называют длиной кодового слова. Число различных кодовых слов m в двоичном блочном коде с параметрами (n, k) равно 2^k .

Основными параметрами блочного кода являются: скорость кода $R = k/n$, минимальное кодовое расстояние d_{min} между любой парой кодовых слов данного кода и вес $w(B)$ кодового слова B , определяемый как число ненулевых символов в кодовом слове B . Последний параметр определяется минимальным расстоянием, или расстоянием Хэмминга, между любыми двумя кодовыми словами длиной n .

Любой двоичный блочный код с $d_{min} = d$ способен обнаружить любые $f = d - 1$ ошибок и исправить $t = \lfloor (d - 1) / 2 \rfloor$ ошибок. Так, при $d = 3$ код способен исправить одну ошибку и обнаружить две.

Совершенные коды позволяют исправить все ошибки кратности и ни одной ошибки более высокой кратности.

7.4. Блочное линейное кодирование

Значительная часть кодирующих схем, применяемых на практике, предполагает использование линейных кодов. Блочный код длиной n с 2^k словами называется линейным (n, k) -кодом, если его кодовые слова образуют k -мерное подпространство V_k векторного n -мерного пространства V_n . Подпространство V_k порождается базисом из k линейно независимых векторов, которые образуют строки порождающей матрицы (n, k) -кода

$$G = \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_{k-1} \end{bmatrix} = \begin{bmatrix} g_{0,0} & g_{0,1} & \cdots & g_{0,n-1} \\ g_{1,0} & g_{1,1} & \cdots & g_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ g_{k-1,0} & g_{k-1,1} & \cdots & g_{k-1,n-1} \end{bmatrix}. \quad (7.2)$$

Процесс кодирования блочного кода состоит в разбиении информационной последовательности на сообщения $U = (u_0, u_1, \dots, u_{k-1})$ длиной k и отображении этих сообщений в кодовые слова $C = (c_0, c_1, \dots, c_{n-1})$ длиной n :

$$C = UG. \quad (7.3)$$

Кодовые слова можно представлять в систематической форме $C = (U, V)$, образуя отдельно информационную часть U из k символов и проверочную часть V из $m = n - k$ символов. Порождающая матрица такого систематического кода будет иметь вид

$$G = [I_k P] = \begin{bmatrix} 1 & 0 & \cdots & 0 & g_{0,0} & g_{0,1} & \cdots & g_{0,n-1} \\ 0 & 1 & \cdots & 0 & g_{1,0} & g_{1,1} & \cdots & g_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & g_{k-1,0} & g_{k-1,1} & \cdots & g_{k-1,n-1} \end{bmatrix}. \quad (7.4)$$

7.5. Древовидные коды

Древовидные коды – это коды, у которых мгновенные значения состояния кодера и последовательные их изменения полностью описываются древовидной диаграммой. Кодер свёрточного кода является устройством с памятью, на вход которого поступают наборы из k_0 входных символов, а на выходе появляются наборы из n_0 выходных символов, которые зависят как от набора входных символов k_0 , так и от v предыдущих входных символов. Память кодера хранит $K = k_0 + v$ входных информационных символов. Параметр K называют длиной кодового ограничения.

Параметры древовидного кода: скорость кода $R = k_0 / n_0$, свободное кодовое расстояние $d_{св}$.

Рассмотрим пример кодера древовидного типа. Он представляет регистр сдвига с одним или несколькими сумматорами (рис. 7.1).

Древовидная диаграмма этого кода приведена на рис. 7.2. Вертикальные линии диаграммы называются ребрами дерева, горизонтальные – ветвями. Каждое ребро дерева соответствует входному символу кодера, каждая ветвь представляет собой последовательность, состоящую из n_0 кодирован-

ных (выходных) двоичных символов. Для рассматриваемого примера $k_0 = 1$ и $n_0 = 2$. Процедура кодирования может быть описана последовательно во времени. Входной бит 0 соответствует верхнему ребру, бит 1 – нижнему. На рис. 7.2 вершина каждой ветви (узел) обозначена числами от 0 до 3 (в двоичном представлении) в соответствии с содержимым двух левых ячеек регистра. Число означает состояние кодера. Ребра, выходящие из двух узлов, характеризуемых одинаковым состоянием, тождественны. Проследим, как формируется кодовая (выходная) последовательность символов на примере некоторой произвольной (входной) последовательности информационных бит.

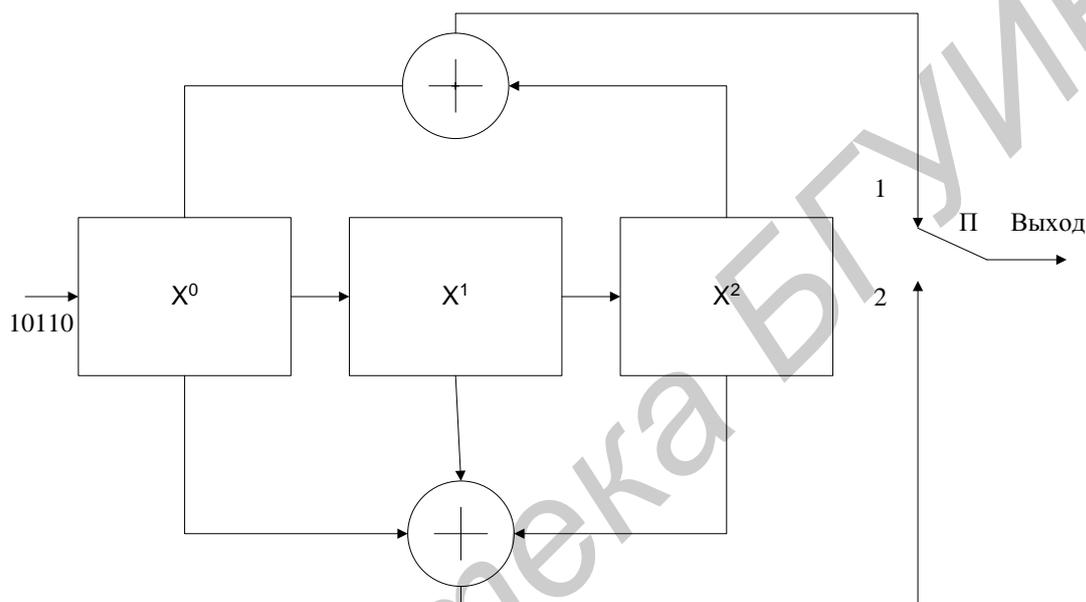


Рис. 7.1. Кодер двоичного древовидного кода с параметрами $R = 1/2$, $v = 2$, $K = 3$

Пусть последовательность информационных бит имеет вид 10110. В начальный момент времени состояние кодера 000, на вход кодера подается первый бит (1), происходит движение по нижнему ребру, на выход подается ветвь (11). На вход кодера подается второй бит (0) происходит движение по верхнему ребру, на выход подается ветвь (01). Третий бит (1) вызывает движение по нижнему ребру, на выход подается ветвь (00) и т. д. Таким образом, информационной последовательности бит (10110) соответствует кодовая последовательность символов (1101001010).

Путь формирования ходовой последовательности выделен на рис. 7.2 жирной ломаной линией.

7.6. Кодирование методом укрупнения алфавита

Данный метод передачи часто относят к методам m -ичной модуляции. Примером такой модуляции является широко используемая на практике мно-

гопозиционная ЧМ. Однако с полным основанием его относят и к методам кодирования блоковыми кодами (m -ичное кодирование). При $m \rightarrow \infty$ такое кодирование позволяет сколь угодно близко приблизиться к пределу Шеннона: $E_b / N_0 = \ln 2$.

Операция кодирования состоит в отображении блока из k последовательных независимых символов в один из $m = 2k$ подлежащих передаче сигналов, т. е. происходит переход от передачи символов двоичного алфавита к передаче символов (слов) m -ичного алфавита. Причем символы m -ичного алфавита должны приниматься в целом.

При анализе помехоустойчивости при кодировании укрупнением алфавита необходимо учитывать следующее:

– энергия символа укрупненного алфавита E_m в k раз больше энергии бита, что ведет к повышению помехоустойчивости;

– при использовании двоичных сигналов ошибка приведет к тому, что вместо передаваемого символа 1(0) можно принять только ему альтернативный 0(1). При равновероятной передаче одного из символов m -ичного алфавита ошибка может привести к приему с равной вероятностью любого из ($m-1$) оставшихся символов, что снизит помехоустойчивость системы.

Реализовано такое кодирование может быть с помощью ортогональных сигналов. Для этих целей могут быть применены различные алгоритмы построения ансамблей ортогональных сигналов. Второй путь – это использование симплексных сигналов.

Если используется некогерентный прием, т. е. параметры ЭС на приемной стороне заранее неизвестны и вычисляются статистически, то оптимальным является использование ортогональных сигналов.

Так же, как и при когерентном приеме, с увеличением m требуемое для обеспечения заданной вероятности ошибки отношение $E_c / N_0 \rightarrow \ln 2$, при этом потери полезного сигнала существенно меньше.

7.7. Определение алгебраических структур

Важными и многочисленным классом часто используемых на практике блоковых кодов являются циклические коды. Поиск лучших среди блоковых кодов, т. е. таких, которые при прочих равных условиях (одинаковых n и k) обнаруживают или/и исправляют наибольшее количество ошибок, связан с введением некоторых ограничений на их свойства или алгоритмы их построения. Кроме того, сами алгоритмы кодирования и декодирования включают алгебраические операции суммирования и умножения. Эти алгебраические операции выполняются в соответствии с правилами определения алгебраических структур, которые имеют символы, содержащиеся в кодовых словах. Далее приводится минимальная информация для определения алгебраических структур, необходимая для изучения блоковых кодов и алгоритмов их декодирования.

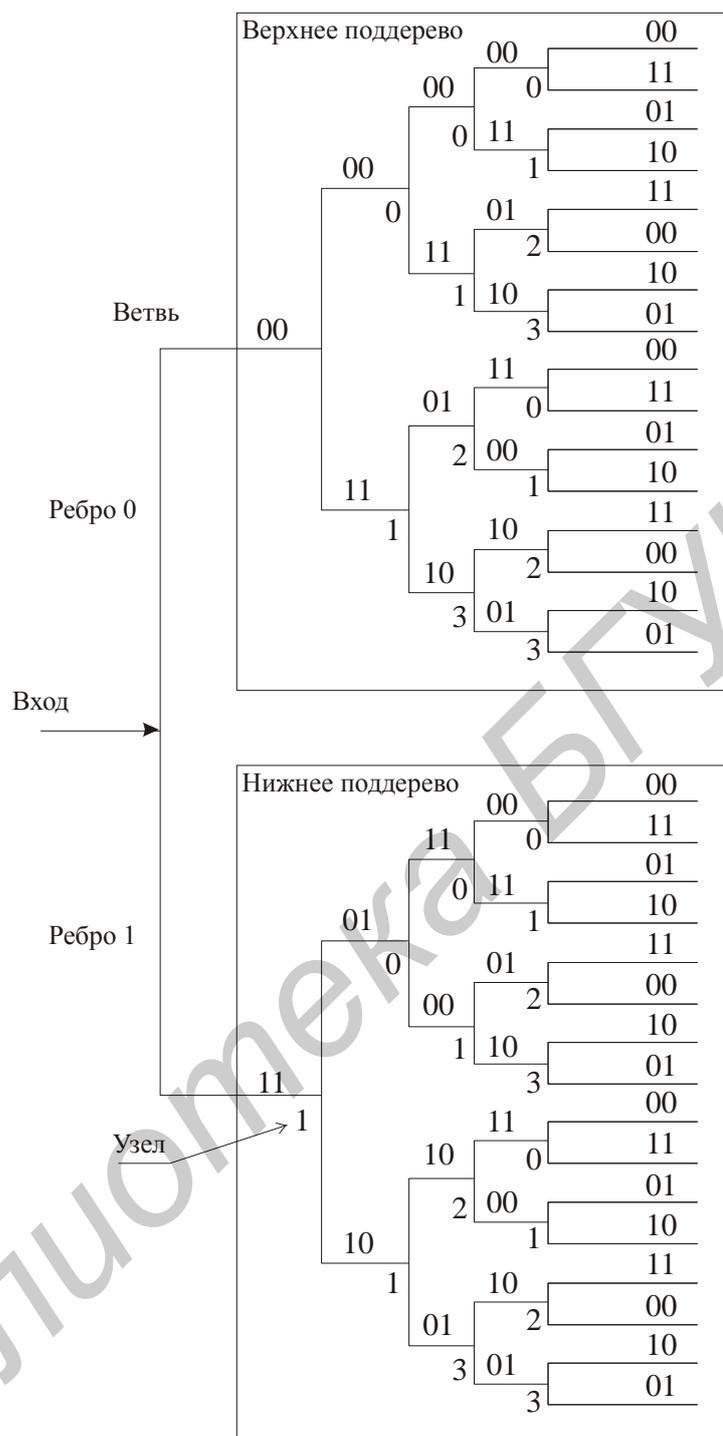


Рис. 7.2. Часть полубесконечного дерева для двоичного древовидного кода с параметрами $R = 1/2$, $v = 2$, $K=3$

Группой G называют множество элементов или символов, для которых определена одна операция (сложение или умножение) и выполняются следующие аксиомы:

– операция может быть применена к двум другим элементам группы, в результате чего получается третий элемент этой же группы;

- для любых трех элементов a, b, c справедливо следующее: $(a+b)+c = a+(b+c)$ и $(ab)c = a(bc)$;
- в группе существует единственный (нейтральный) единичный элемент. Для сложения: $a+0 = 0+a = a$. Для умножения: $1a = a1$;
- каждый элемент группы имеет единственный обратный элемент: $a+(-a) = (-a)+a = 0$. $aa^{-1} = a^{-1}a = 1$;
- если $a+b = b+a$ и $ab = ba$, то такая группа называется абелевой.

Кольцом R называют множество элементов или символов, для которых определены две операции: сложения – $(a+b)$ и умножения – ab .

Должны выполняться также следующие аксиомы:

- множество R является аддитивной абелевой группой относительно операции сложения;
- для любых a и b из множества R $ab = ba$ также является элементом R ;
- для любых a, b и c из множества R $a(bc) = (ab)c$ также является элементом R ;
- для любых a, b и c из множества R справедливы равенства: $a(b+c) = ab+ac$ и $(b+c)a = ba+ca$.

Поле F называют множество элементов, в котором определены две операции – сложение и умножение, причем это множество по операции сложения образует аддитивную абелеву группу с единичным элементом, обозначенным «0»; по операции умножения группу образуют все элементы поля, кроме 0, единичный элемент этой группы обозначается «1».

Для любого числа q , являющегося степенью простого числа, существует поле, содержащее q элементов. Поля с конечным числом элементов $q = p^i$, где p – простое число, а i – любое натуральное число, называют полями Галуа и обозначаются $GP(q)$.

При $i = 1$ элементами поля будут простые натуральные числа от 0 до $p-1$, и операции сложения и умножения над этими числами производятся по $mod\ 2$. Такие поля Галуа называются простыми. При $i > 1$ поля Галуа являются расширением простого поля.

7.8. Циклические коды

Циклическим кодом называют такой блочный код, которому вместе с некоторым кодовым словом принадлежат и другие кодовые слова, являющиеся всеми его циклическими сдвигами. Пусть $B = [b_{n-1}b_{n-2}...b_1b_0]$ – кодовое слово. Тогда последовательность $B^{(1)} = [b_{n-2}b_{n-1}...b_0b_{n-1}]$, являющаяся циклическим сдвигом B , также является кодовым словом данного циклического кода.

При операциях с циклическими кодами обычное кодовое слово B удобно представлять в виде многочлена $B(X)$ со степенью $\leq n-1$, где коэффициенты при степенях переменной величины есть символы кодового слова.

$$B(X) = b_{n-1}X^{n-1} + b_{n-2}X^{n-2} + \dots + b_1X + b_0. \quad (7.5)$$

Коэффициенты $B(X)$ есть элементы поля Галуа. Многочлен $B^{(1)}$ можно получить умножением многочлена $B(X)$ на X , т. е. это остаток от деления

$$[XB(X)] \text{ на } X^n - 1.$$

По аналогии

$$B^{(i)}(X) = [X^i B(X)] / (X^n - 1). \quad (7.6)$$

Представим также информационное сообщение в виде

$$C(X) = c_{k-1}X^{k-1} + c_{k-2}X^{k-2} + \dots + c_1X + c_0, \quad (7.7)$$

где последовательность $C = [c_{k-1}, c_{k-2}, \dots, c_1, c_0]$ – информационные символы.

Введем некоторый порождающий многочлен, который является одним из сомножителей при разложении $X^n - 1$ на множители.

$$g(X) = X^{n-k} + g_{n-k-1}X^{n-k-1} + \dots + g_1X + 1. \quad (7.8)$$

При практическом построении двоичных циклических кодов (n, k) и выборе порождающих многочленов можно воспользоваться специальной таблицей.

Можно записать без вывода:

$$B^i(X) = C^i(X)g(X), \quad (7.9)$$

где $i = 0, 1, \dots, q^k$.

Порождающий многочлен $g(X)$ является делителем двухчлена $X^n - 1$. Следовательно, справедливо равенство

$$X^n - 1 = g(X)h(X). \quad (7.10)$$

Функцию $h(X)$ называют проверочным многочленом.

В ряде случаев при реализации кодера циклического кода на основе регистра сдвига пользуются проверочным многочленом $h(X)$. Этот многочлен

также используют при построении декодеров для декодирования последовательностей, построенных по мажоритарным алгоритмам. В этом случае на основе проверочного многочлена $h(X)$ формируется $n-k$ проверочных соотношений:

$$S_{j-k} = \sum_{i=0}^k h_i b_{j-i}, k \leq j \leq n-1. \quad (7.11)$$

Соотношение (7.11) называют синдромом комбинаций B_i . Равенство нулю выражения (7.11) имеет место только в том случае, если коэффициенты b_{ij} являются коэффициентами проверочного многочлена. Это означает, если координаты синдрома S равны нулю, то для этих символов b_{ij} не искажены помехами.

7.9. Наиболее распространенные циклические коды

Коды Хемминга могут быть как двоичными, так и недвоичными. Двоичные коды Хемминга – это совершенные циклические коды с параметрами $n = 2^l - 1$, $k = 2^l - l - 1$, $d = 3$. Данные коды исправляют только одиночные ошибки. При $l = 3$ $g(X) = X^3 + X + 1$ и $h(X) = X^3 + X^2 + X + 1$ кодер кода Хемминга (7.4) на основе порождающего и проверочного многочленов представлен на рис. 7.3.

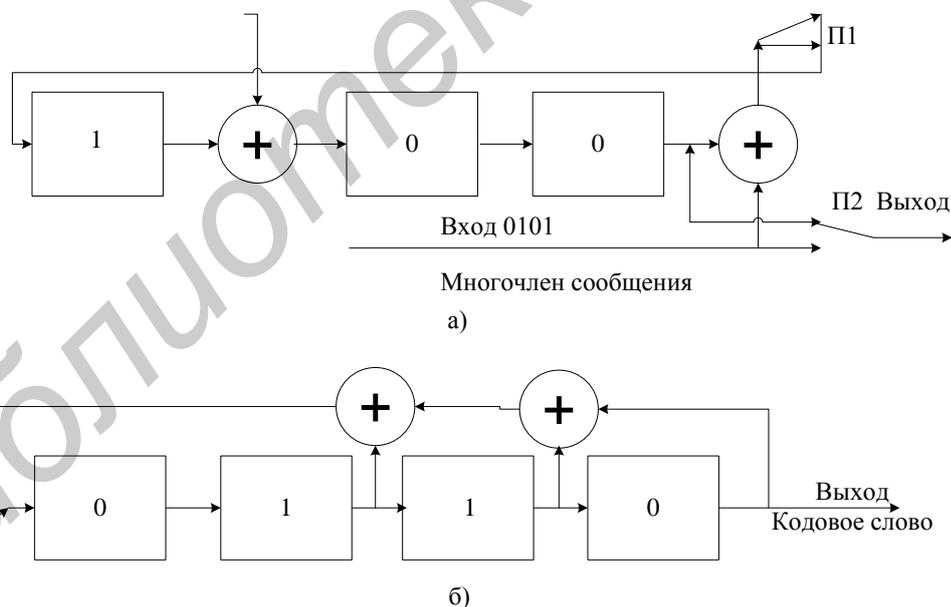


Рис. 7.3. Кодер на основе многочленов:
а – порождающего; б – проверочного

Изменения состояний регистра кодера для порождающего многочлена представлены в табличном виде ниже. В исходном состоянии в регистре – все нули.

Информация	Номер шага	Содержимое регистра
0	1	000
1	2	110
1	3	101
0	4	100

Код Голея – двоичный совершенный циклический код с параметрами $n = 23$, $k = 12$ и $d = 7$.

Порождающий многочлен:

$$g(X) = X^{11} + X^9 + X^7 + X^6 + X^5 + X + 1.$$

Проверочный многочлен

$$h(X) = X^{12} + X^{10} + X^7 + X^4 + X^3 + X^2 + X + 1$$

реализуется на основе регистров сдвига. Пример реализации кода приведен на рис. 7.4.

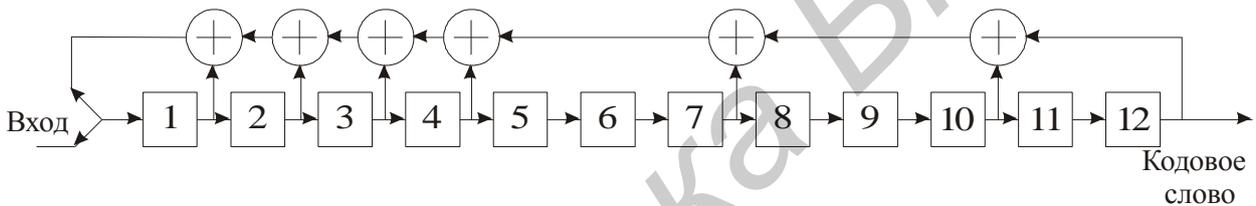


Рис. 7.4. Реализация кода Голея

Коды Боуза–Чоудхури–Хоквингема (БЧХ)

Для любых целых чисел m и $t < (n/2)$ существует двоичный код БЧХ длиной $n = 2^m - 1$, исправляющий все комбинации из t_0 и меньшего числа ошибок и содержащий не более mt_0 проверочных символов в каждом слове.

Порождающие многочлены БЧХ-кодов представляют собой НОК минимальных многочленов $m_i(X)$ делителей двучлена $X^{2^m-1} - 1$:

$$g(X) = \text{НОК}[m_1(X) m_3(X) \dots M_{2^{t_0}-1}].$$

Коды БЧХ – квазисовершенные коды, т. е. лучшие из двоичных кодов.

Коды Рида–Соломона

Составляют подкласс недвоичных БЧХ-кодов. Параметры РС-кодов:

- длина кодового слова $N = q - 1 = (2^k - 1)$ q -ичных символов, где $k = 1, 2, 3, \dots, n - 1$;
- количество информационных символов в кодовом слове $K = N - 2t$;
- минимальное кодовое расстояние $d_{min} = N - K + 1$;
- кодовая скорость $R = K / N$;

– количество исправляемых ошибок в кодовом слове t ,

$$t = \left\lfloor \frac{1}{2}(D_{min} - 1) \right\rfloor = \left\lfloor \frac{1}{2}(N - K) \right\rfloor.$$

Эти коды позволяют исправлять все ошибки, группирующиеся в пачки длиной не более t q -ичных символов.

7.10. Мажоритарно декодируемые коды

В основе функционирования этих кодов лежит одно правило – правило ортогональных проверок. Для каждого символа произвольного кодового слова выполняется проверка с помощью многочлена $h(X)$. Суть проверки состоит в том, что ни один из символов кодового слова не должен входить в систему составленных уравнений два и более раз. При этом при декодировании вычисляется синдром и составляется таблица соответствия синдромов образцам исправляемых ошибок. Если синдром равен нулю, то символ передан правильно.

7.11. Декодирование циклических кодов

Методы декодирования циклических кодов разделяются на два больших класса: алгебраические и неалгебраические. В основе алгебраических методов лежит решение систем уравнений, определяющих места расположения и значения ошибок. При неалгебраических методах та же цель достигается с использованием известных структурных понятий и свойств кодов, позволяющих определить места и значения тех же ошибок более простыми путями. Среди алгоритмов жесткого декодирования выделяют алгоритм табличного декодирования по минимуму расстояния. Рассмотрим данный алгоритм на примере блочного кода (5.2). Данный код образует четыре кодовых слова:

00000 01011 10101 11110.

В этих словах первые два символа информационные, три последующих – проверочные. Минимальное кодовое расстояние между кодовыми словами $d_{min} = 3$, следовательно, может быть исправлена любая, произвольно расположенная одиночная ошибка. В табл. 7.1 показан принцип табличного декодирования блочного кода. В столбцах расположены двоичные пятиразрядные комбинации, имеющие меньшее расстояние Хемминга до кодового слова, образующего данный столбец, чем до кодовых слов, образующих другие столбцы.

Если принята комбинация, принадлежащая первому столбцу, она отождествляется с кодовым словом 00000 и декодируется как 00.

Технически кодер реализуется в виде запоминающего устройства, по адресам которого записаны следующие комбинации: для первого столбца – 00; для второго – 01; для третьего – 10; для четвертого – 11. Современные

микросхемы памяти позволяют строить табличные декодеры для кодов с $n < (15-20)$. Приведенный алгоритм является алгоритмом декодирования по минимальному расстоянию Хемминга, т. е. алгоритмом максимального правдоподобия для ДСК.

Синдромное декодирование является достаточно простым методом жесткого декодирования циклических кодов. Рассмотрим данный алгоритм в общем виде. Для всех вариантов исправляемых выбранным кодом ошибок вычисляются синдромы и составляется таблица соответствия синдромов образцам исправляемых ошибок. При этом под синдромом понимается следующее. Если вектор S определяется координатами S_{j-k} , то такой вектор называют синдромом комбинации B_j .

$$S_{j-k} = \sum_{i=0}^k h_i b_{j-i} = 0, \text{ при } k \leq j \leq n-1. \quad (7.12)$$

Таблица 7.1

Табличное декодирование блокового кода

№ п/п	Кодовые слова			
	00000	01011	10101	11110
1	00001	01010	10100	11111
2	00010	01001	10111	11100
3	00100	01111	10001	11010
4	01000	00011	11101	10110
5	10000	11011	00101	01110
6	11000	10011	01101	00110
7	10010	11001	00111	01100

Ниже приведена таблица соответствия синдромов. Если код циклический, то синдром вычисляется с помощью регистров сдвига путем деления комбинации в виде многочлена $B(X)$ на многочлен $g(X)$. Остаток от деления $r(X)$, имеющий степень, меньшую, чем $n - k$, и есть искомым синдромом $S(X)$. Дальнейшие операции по декодированию выполняются с помощью запоминающего устройства, адресами которого являются синдромы, а в информационной части хранятся соответствующие им векторы ошибок. Далее, сложив принятую комбинацию и вычисленный вектор ошибок, определяем наиболее вероятно принятое слово и соответствующий ему блок информационных символов.

Таблица 7.2
Соответствие синдромов

№ п/п	Вектор ошибки	Синдром
1	00000	000
2	00001	001
3	00010	010
4	00100	100
5	01000	011
6	10000	101
7	11000	110
8	10010	111

7.12. Сверточные коды

Сверточный код в отличие от блочного описывается тремя целыми числами – n , k , K . Отношение k/n имеет тот же смысл, что и для блочных кодов, однако n не определяет длину кодового слова или блока. Целое число K указывает число разрядов k -кортежа в кодирующем регистре сдвига. Его называют длиной кодового ограничения. Важная особенность сверточных кодов состоит в том, что кодер имеет память – n -кортежи, получаемые в результате кодирования, являются функцией не только одного входного k -кортежа, но и предыдущих $K - 1$ входных k -кортежей.

После кодирования каждое кодовое слово состоит из двоичных кодовых символов, называемых канальными символами, канальными битами, или битами кода. В отличие от битов входного сообщения кодовые символы не являются независимыми.

Обычный сверточный кодер показан на рис. 7.5. Он реализован на K -разрядном регистре сдвига с n сумматорами по модулю 2, где K – длина кодового ограничения – количество k -битовых сдвигов, после которых один информационный бит может повлиять на выходной сигнал кодера. В каждый момент времени на место первых k разрядов перемещаются k новых бит; все биты в регистре смещаются на k разрядов вправо, и выходные данные n сумматоров последовательно дискретизируются, формируя выходной код.

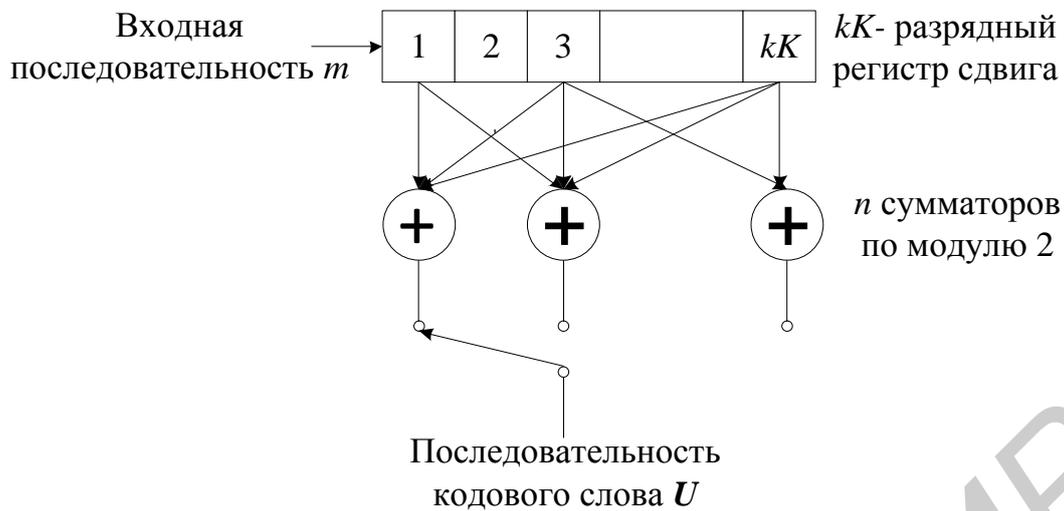


Рис. 7.5. Сверточный кодер с длиной кодового ограничения K и степенью кодирования k/n

Вторым примером сверточного кодера является кодер, изображенный на рис. 7.1, и древовидная диаграмма кода, изображенная на рис. 7.2. Для рассматриваемого сверточного кода решетчатая диаграмма приведена на рис. 7.6.

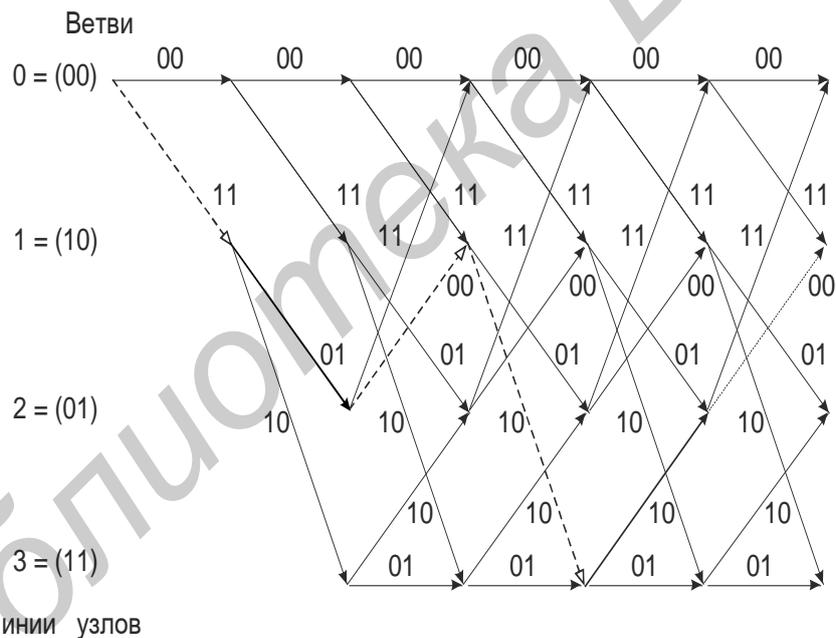


Рис. 7.6. Решетчатая диаграмма сверточного кода

Линии узлов решетки представляют состояние регистра: $0 = (00)$, $1 = (10)$, $2 = (01)$, $3 = (11)$. Для рассматриваемого кода имеется четыре линии узлов. На рисунке сплошная линия перехода из одного узла в другой соответствует ветви (двум выходным символам), генерируемой после поступления входного бита), а штриховая – ветви, которая генерируется при поступлении входного бита 1.

В структуре решетчатой диаграммы, как и в структуре древовидной, имеет место периодичность повторения состояния после достижения глубины ветвления K (в нашем случае $K = 3$). Так, из узла 0: при входном бите 0

происходит переход в узел 0, при этом на выходе кодера генерируется ветвь 00; при входном бите 1 происходит переход в узел 1, при этом на выходе кодера генерируется ветвь 11. Из узла 1: при входном бите 0 происходит переход в узел 2, при этом проходит ветвь 01; при входном бите 1 происходит переход в узел 3, при этом генерируется ветвь 10.

Путь на решетчатой диаграмме при исходном состоянии регистра 000 и входной последовательности 10110 отмечен жирной линией. Вышеназванной входной последовательности соответствует последовательность кодированных символов 1101001010.

Решетчатая диаграмма, как и древовидная, однозначно связывает информационную последовательность бит, последовательность состояний кодера и последовательность кодированных символов на его выходе.

Известны три основных алгоритма декодирования сверточных кодов: пороговое декодирование, декодирование по алгоритму Витерби и последовательное декодирование. Пороговое декодирование не нашло широкого практического применения, хотя его модификация как многопороговое декодирование позволяет получить существенный энергетический выигрыш.

При декодировании по Витерби каждой последовательности информационных бит соответствует свой путь на решетчатой диаграмме выбранного кода. Задача декодера сверточного кода, реализующего алгоритм Витерби, сводится к тому, чтобы найти на решетчатой диаграмме разрешенный путь, наиболее близкий по какой-либо приемлемой метрике к пути, соответствующему принятому на фоне помех последовательности кодированных символов и по нему принять решения о значении переданных информационных бит. Поскольку здесь анализируются все разрешенные пути и в результате анализа выбирается путь с минимальной метрикой как наиболее правдоподобный, то решающий указанную задачу декодер (декодер Витерби) является декодером максимального правдоподобия.

Существует несколько реализаций алгоритмов последовательного декодирования сверточных кодов. Наиболее широкое применение в системах связи нашел алгоритм последовательного декодирования, предложений Фано еще до появления алгоритма Витерби. В отличие от декодера Витерби, этот алгоритм при приемлемой сложности реализации позволяет декодировать сверточные коды с длиной кодового ограничения $K > 9$. Алгоритм Фано позволяет декодировать и коды со значениями $K < 9$, но поскольку он не является алгоритмом максимального правдоподобия, то по помехоустойчивости при прочих равных условиях уступает алгоритму Витерби.

В отличие от алгоритма Витерби, где на каждом шаге производится обновление метрик всех выживших путей, в алгоритме Фано обновляется метрика только того пути, который оценивается как наиболее вероятный. Такой подход к декодированию может рассматриваться как метод проб и ошибок для поиска правильного пути на диаграмме. Поиск осуществляется последовательно, так что в каждый момент времени происходит формирование только одного пути. Декодер может двигаться по выбранному пути вперед или назад и менять предыдущие решения. Однако основное правило при реа-

лизации алгоритма последовательного декодирования состоит в том, что декодер не должен двигаться ни вперед, ни назад по пути, достоверность которого ниже некоторого текущего порога. Все разрешенные переходы, исходящие из текущего узла, должны сравниваться с принятой ветвью и упорядочиваться в соответствии с их вероятностью. Декодер «движется» по наиболее вероятному пути из каждого узла до тех пор, пока значение его достоверности превышает значение текущего порога.

7.13. Основные принципы каскадного кодирования

На практике нередко возникают задачи, когда необходимо использовать наиболее мощные коды (обеспечение высокой помехоустойчивости при дефиците энергетики радиолиний и т.д.). Применение в таких случаях более длинных блочных или сверточных кодов приводит к недопустимому усложнению самого декодера. Форми предложил схему каскадного (многоуровневого) кодирования. На практике наиболее широко используют двухуровневую схему. Структурная схема двухуровневого каскадного кодирования представлена на рис. 7.7.

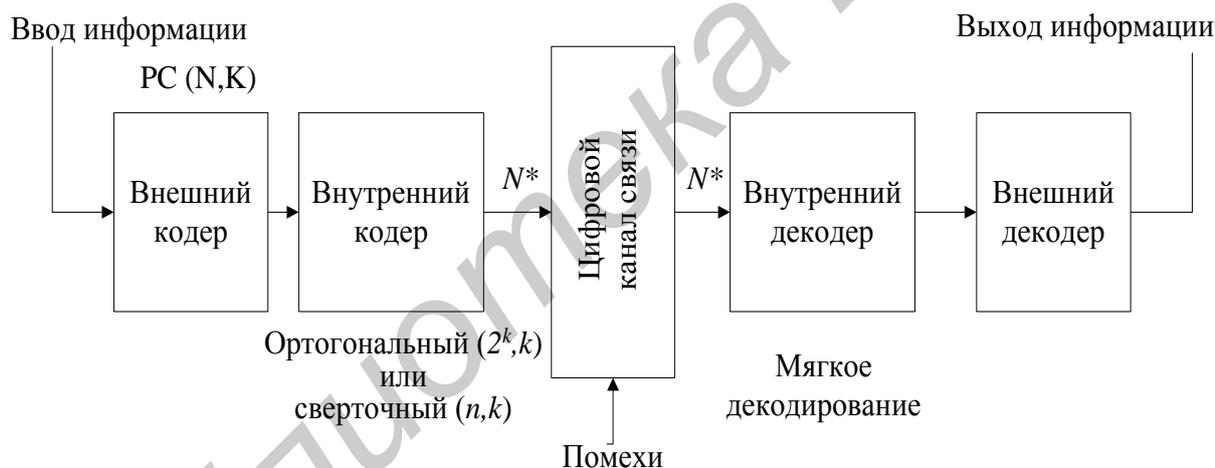


Рис. 7.7. Схема двухуровневого каскадного кодирования

Здесь используются два кода: внешний код, обычно блочный и недвоичный, и внутренний. В качестве внешних кодов обычно используются q -ичный код ($q = 2^k$) Рида–Соломона. Такие коды имеют максимальное кодовое расстояние для фиксированных N и K и относительно просто декодируются. В качестве внутренних кодов используются циклические, ортогональные и сверточные коды.

Перед кодированием внешним кодом двоичная информация разбивается на последовательные группы по k символов каждая. Такая группа представляет собой один q -ичный символ. Далее последовательности из K q -ичных символов кодером ставится в соответствие кодовое слово РС-кода дли-

ной N q -ичных символов или длиной N k двоичных символов. Далее сообщение кодируется внутренним кодом.

При использовании внутреннего кодирования каждой указанной выше группе k двоичных символов ставится в соответствие кодовое слово внутреннего кода длиной n двоичных символов. Длина полученного кодового слова каскадного кода будет равна $N^* = N n$ двоичных символов; количество информационных бит – $K^* = Kk$, скорость каскадного кода будет равна

$$R^* = R_{\text{внешний}} / R_{\text{внутр}} = Kk / Nn. \quad (7.13)$$

При использовании в качестве внутренних кодов ортогональных m -ичных кодов каждой указанной выше группе из k двоичных символов ставится в соответствие кодовое слово ортогональных m -ичных кодов. Так как при использовании ортогональных m -ичных кодов требования к полосе пропускания растут пропорционально m , то и применение таких кодов возможно только в том случае, когда имеется запас по полосе пропускания.

При использовании сверточных кодов РС длина полученного кодового слова каскадного кода будет равна $N^* = Nn_0$, количество информационных бит составит $K^* = Kk_0$, а скорость кода

$$R^* = R_{\text{внешний}} / R_{\text{внутр}} = Kk_0 / Nn_0. \quad (7.14)$$

Декодирование информации осуществляется в обратном порядке.

7.14. Сверточные турбокоды

Разработка турбокодов явилась заметным достижением в теории и технике помехоустойчивого кодирования, так как позволила вплотную приблизиться к пределу Шеннона. Для таких кодов получены следующие параметры: при $R=1/2$ в канале связи с БГШ $P_{\text{ош}}=10^{-5}$ требуется соотношение $E_b/N_0=0,7$ дБ, что на (1,5–2) дБ меньше, чем при использовании лучших каскадных кодов при $R=1/2$.

Турбокоды бывают следующего вида:

- сверточные турбокоды (СТК), которые формируются путем параллельного соединения двух и более сверточных кодеров;
- блочные турбокоды (БТК), которые формируются путем параллельного соединения двух и более блочных кодеров.

Кодер эффективного СТК строится параллельным соединением двух или нескольких рекурсивных систематических сверточных кодов (РССК) с относительно небольшой длиной кодового слова – $K=3-5$. Схема кодера представлена на рис. 7.8.

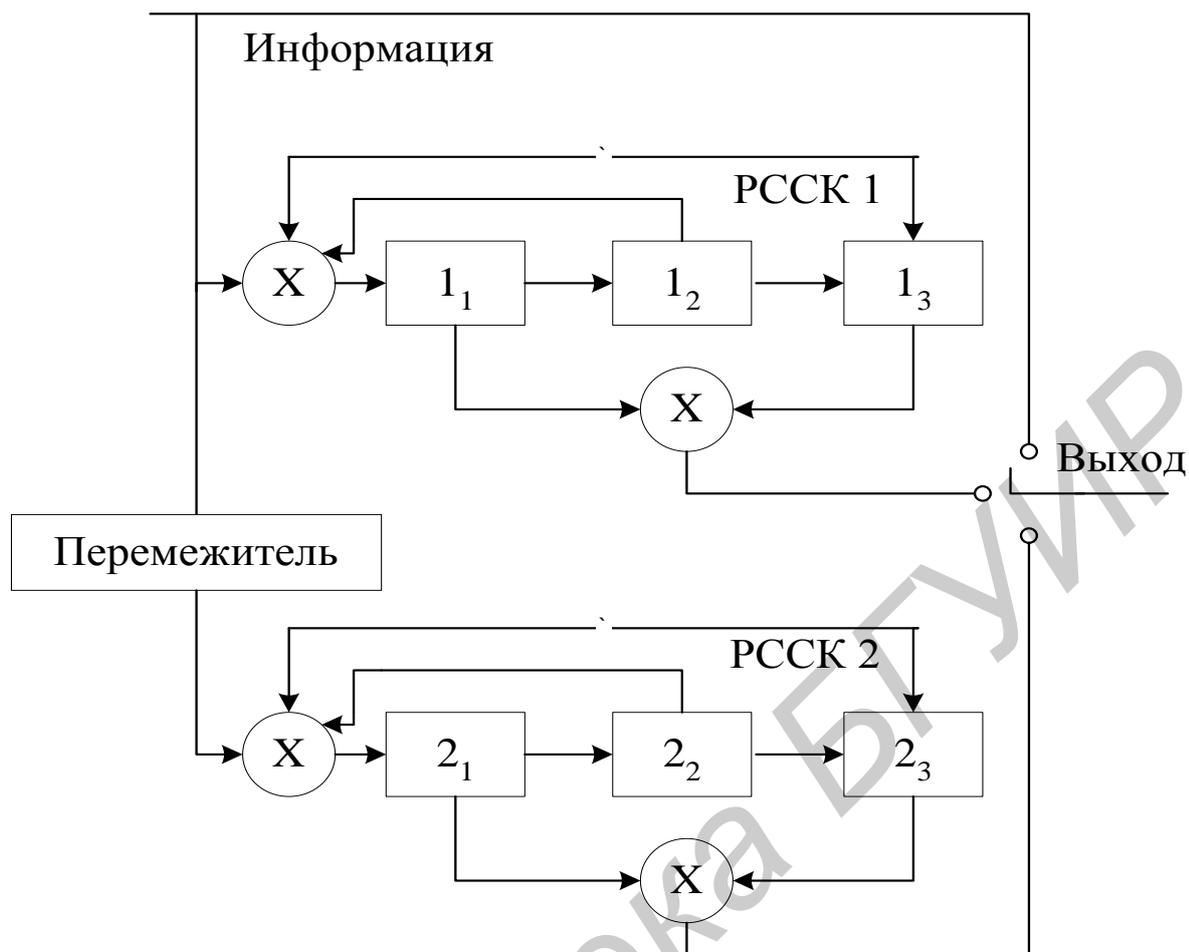


Рис. 7.8. Кодер систематического сверточного турбокода с $K = 4$

Каждый кодер РССК по умолчанию работает со скоростью $R = 1/2$. Значение символа на систематическом выходе верхнего РССК совпадает со значением входного бита, а на втором выходе верхнего регистра формируется проверочный символ. Для осуществления операции перемежения выделяется некоторый блок из k информационных бит, который перед кодированием записывается в матричную память и считывается из нее на вход нижнего РССК по псевдослучайному закону. Систематический выход нижнего регистра не используется, а второй его выход формирует проверочный бит. В результате на выходе получаем: последовательность a_k с систематического выхода верхнего регистра (она же совпадает с входной последовательностью) и последовательно два проверочных символа с верхнего x_k и нижнего y_k регистров. Кодовая скорость сформированной таким образом последовательности будет равна $R = 1/3$. Если проверочные символы снимать с регистров через такт (в одном такте – с первого, во втором – со второго), то скорость можно повысить до $R = 1/2$.

Анализ результатов экспериментальных исследований СТК показал, что структура перемежителя существенно не влияет на эффективность кодирования.

Декодер СТК, так же как и кодер, состоит из двух декодеров (рис. 7.9).

В декодерах 1 и 2 используется алгоритм для учета особенностей рекурсивных кодов. Это алгоритм работает на основе критерия максимальной апостериорной вероятности (Maximum A Posteriori) и позволяет формировать так называемые мягкие решения: мягкий вход – мягкий выход (Soft Input – Soft Output, SISO). Информация, полученная после декодирования вторым декодером, поступает на вход первого декодера, что повышает эффективность декодирования. Таких операций может быть две–три. Таким образом, используется модульная конструкция декодера (рис. 7.10).

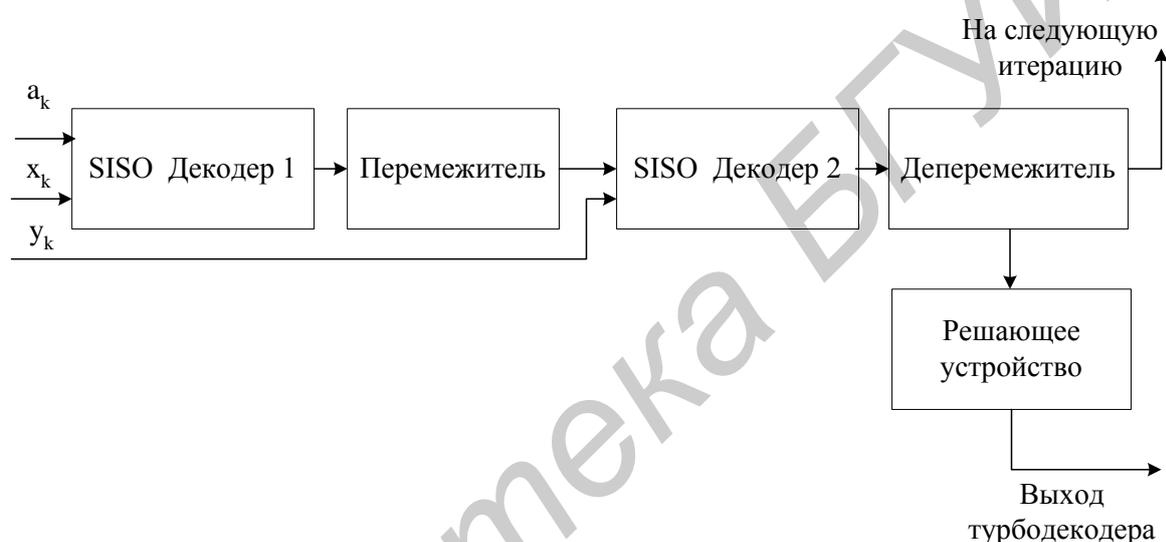


Рис. 7.9. Схема декодера СТК (одна итерация)

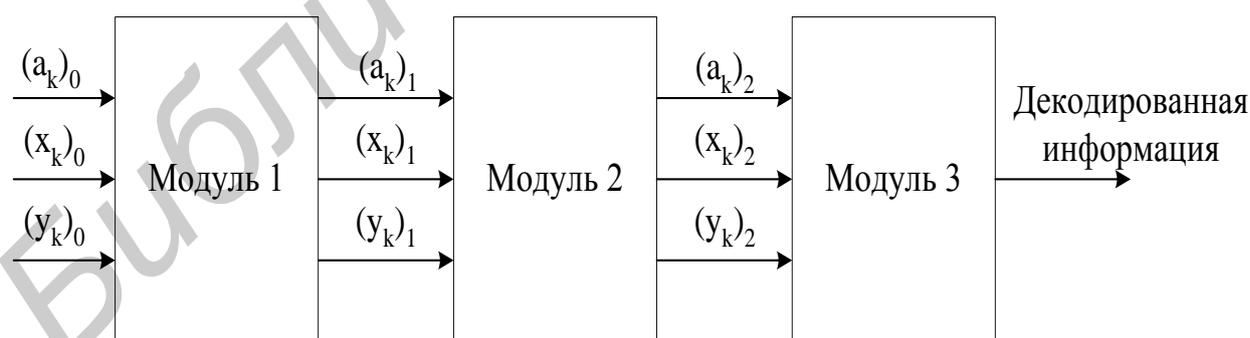


Рис. 7.10. Модульная схема итеративного декодера СТК

7.15. Блочные турбокоды

Блочные турбокоды в сочетании с алгоритмом SISO с несколькими итерациями позволяют приблизиться к пределу Шеннона. БТК более эффективны, чем сверточные при $R > 0,7$. БТК могут формироваться как последовательным, так и параллельным соединением двух или нескольких блоковых кодов. Последовательное построение блоковых кодов – турбокоды произведения (ТРС) – является более эффективным. Структура такого турбо-кода представлена на рис. 7.11.

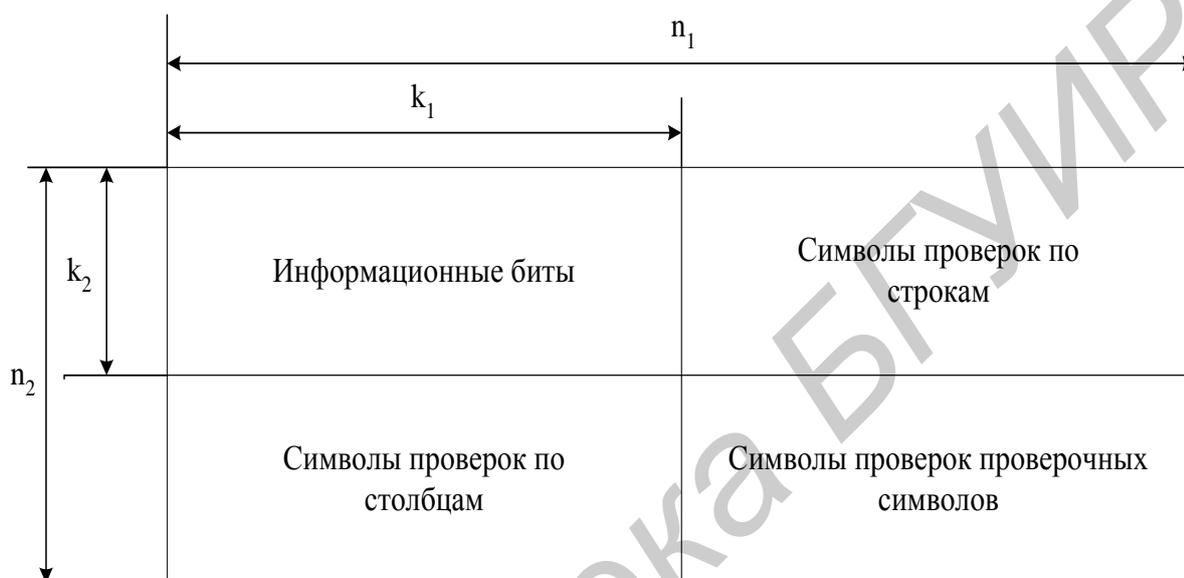


Рис. 7.11. Структура ТРС, сформированного из блоков кодов

Исходный массив информационных бит последовательно, например, по строкам, записывается в виде матрицы. Каждая строка матрицы представляет собой кодовое слово некоторого кода длиной n_1 и содержит k_1 информационных символов и соответственно $(n_1 - k_1)$ проверочных. Этот код по аналогии с каскадным можно назвать внутренним кодом. Каждый столбец матрицы также представляет собой кодовое слово такого или другого блокового кода длины n_2 и содержит k_2 информационных и $(n_2 - k_2)$ проверочных символов. Это внешний код.

Скорость ТРС определяется как $R = R_1 R_2 = k_1 k_2 / n_1 n_2$. Расстояние Хемминга для ТРС равно произведению расстояния составляющих кодов: $d = d_1 d_2$. В качестве исходных блоковых кодов часто выбирают коды Хемминга (расширенные или укороченные) или коды проверки на четность. Описанный метод кодирования обеспечивает независимость строк и столбцов, поэтому введение более сложного перемежения не приводит к какому-либо заметному положительному результату.

Структурная схема кодера ТРС практически не отличается от схемы кодера каскадного типа: сначала информационные биты кодируются внешним кодом, а затем кодовые слова записываются в матричную память (например по строкам), считываются по строкам и кодируются внутренним кодом.

Структурная схема декодера ТРС также не отличается от декодера каскадного типа. Однако на первой итерации декодер внутреннего кода по каждому декодированному символу на основе критерия максимума апостериорной вероятности формирует мягкие решения для декодера внешнего кода. На второй итерации на вход декодера внутреннего кода подаются мягкие решения, достоверность которых уточнена после первой итерации и по строкам, и по столбцам. Окончательное решение о каждом переданном бите информации выносится после заданного числа итераций внешним декодером.

Основное отличие декодера турбокода от других декодеров состоит в том, что декодер турбокодов вычисляет мягкие решения для декодируемых символов на всех итерациях их декодирования (кроме последней), т.е. декодирование производится по принципу SISO. Асимптотически выигрыш от кодирования с ростом отношения сигнал/шум при мягком декодировании может быть оценен по следующей формуле:

$$G = 10 \lg(R \times d_{min}), \text{ дБ.} \quad (7.15)$$

8. ПЕРЕМЕЖЕНИЕ ДАННЫХ

8.1. Общие сведения о перемежении данных

Перемежитель, или интерлевер (от англ. Interleaver, Interleaving) – фундаментальный компонент помехоустойчивого кодирования, предназначенный для борьбы с групповыми ошибками. Суть перемежения состоит в том, что производится изменение порядка следования символов информационной последовательности таким образом, что стоящие рядом символы оказываются разделенными несколькими другими. По сути дела, перемежение обеспечивает преобразование групповых ошибок в канале связи в одиночные.

Перемежители подразделяются на блочные и свёрточные. Блочные устройства работают с двумерными массивами информации, запись данных в которых осуществляется по столбцам, а считывание – по строкам. На приёме запись и считывание осуществляются в обратном порядке.

Перемежитель является одним из основных элементов турбо-кодов, который меняет порядок символов входного сигнала или передает блок данных на вход следующего кодирующего устройства (рис. 8.1). Перемежение выполняется процессором в соответствии с алгоритмом перемешивания. Объединение перемеженных блоков данных с входными используется для снижения частоты ошибочных бит в рассматриваемом канале. Процесс перемежения увеличивает разнородность в данных, поступающих в кодер, что приводит к исправлению ошибок в декодере (на приемной стороне) с помощью алгоритма исправления, если символы искажаются при передаче.

Основная цель перемежителя состоит в том, чтобы предложить каждому кодеру некоррелированную или случайную версию информации, в результате чего паритетные биты каждого кодера становятся независимыми.

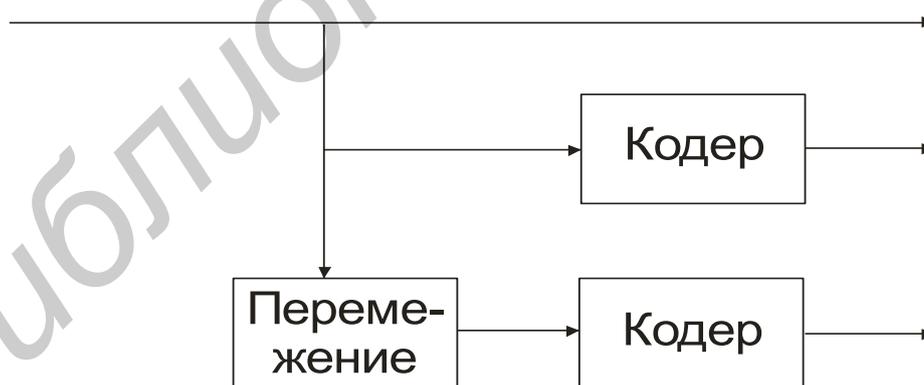


Рис. 8.1. Структурная схема устройства турбокодирования с перемежителем

Блок перемежения функционирует на основе матрицы, строки которой формируются последовательностью входных символов. Затем определенное число бит в зависимости от алгоритма перемежения образует блок. Из перемежителя данные передаются в блок кодера с помощью последовательного

вывода столбцов матрицы. В результате сигналы в расположенных рядом строках матрицы в потоке достоверных сигналов различаются номерами.

В соответствии с идеальной ситуацией количество столбцов и строк надо выбрать так, чтобы расстояние (разница в индексах символов входной последовательности) между соседними битами после перемежителя было наибольшим (для более эффективного исправления ошибок в декодере).

Из-за неконтролируемого увеличения числа строк образуется сигнальная задержка в схеме. В результате необходимо ограничивать размеры перемежителя с целью обеспечения задержки сигналов в допустимых пределах. С другой стороны, уменьшение размеров матрицы ограничивает разделение сигналов на большое расстояние, а следовательно, снижает эффективность коррекции ошибок в декодере.

Существуют два основных метода перемежения: однородное (равномерное) и неоднородное (псевдонеравномерное). В случае однородного перемежения расстояние между двумя любыми соседними битами будет одинаковым. Это не является положительным качеством, поскольку результирующий выходной сигнал (на выходах обоих кодеров) характеризуется относительно невысоким минимальным расстоянием.

Неоднородное перемежение позволяет получить наилучшее «максимальное разбрасывание» данных и «максимальное упорядочивание» выходной последовательности информации. Это означает, что избыточность, вносимая двумя элементарными кодами, равномерно распределена в выходной последовательности турбокодера. Минимальное расстояние между символами в неоднородном перемежении при больших значениях блока увеличивается по сравнению с однородным перемежением. Основная проблема для неоднородного перемежения заключается в достижении достаточной неоднородности. К тому же условное перемежение требует существенного объема памяти в кодере. Матрицы условного перемежения также требуют компенсации задержек, которые ограничивают их применение в соответствии с современными требованиями.

Помимо турбокодов перемежители являются основным компонентом в кодах Рида–Соломона, в мобильных протоколах передачи данных (3GPP technical specification 36.212), в LDPC-кодах. Использование в системе перемежителя повышает её латентность, так как для обработки необходимо принять весь перемешанный блок данных.

8.2. Кодирование информации в каналах связи на базе стандарта TETRA

Рассмотрим, как обеспечивается применение перемежения информации для повышения помехоустойчивости передачи в реальных каналах связи. В каналах радиосвязи на базе стандарта TETRA для защиты от ошибок используется помехоустойчивое канальное кодирование сигнала, которое осуществ-

ляется путем введения в состав передаваемого сигнала достаточно большого объема избыточной информации (рис. 8.2).

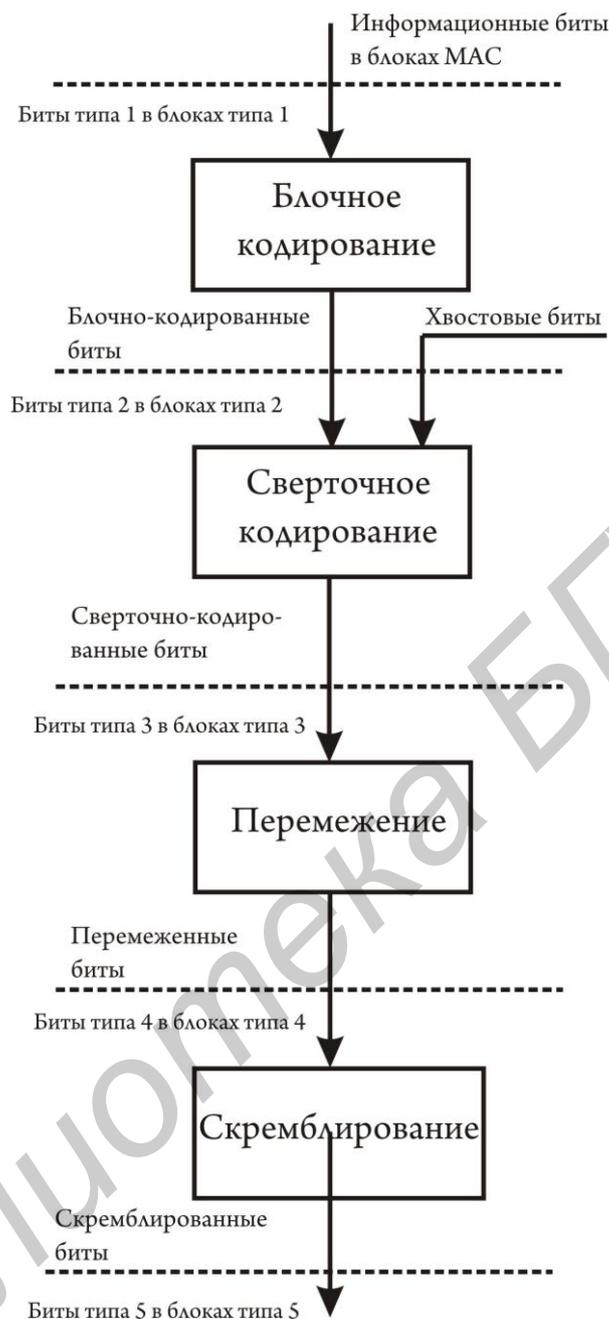


Рис. 8.2. Общая схема канального кодирования в стандарте TETRA

В стандарте TETRA канальное кодирование реализуется в виде четырех процедур:

- D – блочного кодирования;
- E – сверточного кодирования;
- F – перемежения;
- G – скремблирования.

При блочном кодировании входная информация разбивается на блоки по k символов, которые преобразуются по определенному закону в n -символьные блоки, где $n > k$. Блочное кодирование предназначено в ос-

новном для обнаружения одиночных и групповых ошибок в канале связи и в определенных случаях для их исправления.

При сверточном кодировании каждый символ входной информационной последовательности, состоящий из k бит, преобразуется в n -битовый символ выходной последовательности, причем $n > k$. Сверточное кодирование является мощным средством борьбы с одиночными ошибками, хотя и не обеспечивает их обнаружения.

При перемежении производится изменение порядка следования символов информационной последовательности таким образом, что стоящие рядом символы оказываются разделенными несколькими другими. Перемежение обеспечивает преобразование групповых ошибок в канале связи в одиночные.

Скремблирование состоит в преобразовании входной информационной последовательности в выходную путем ее побитного сложения по модулю 2 со специально формируемой шифровальной последовательностью. Скремблирование используется для определенной защиты передаваемой информации, а также для аутентификации абонентов.

Стандарт TETRA описывает процесс канального кодирования как последовательное преобразование данных на четырех уровнях, при этом входные данные обозначаются как биты типа 1 в блоках типа 1, данные после блочного кодирования – биты типа 2 в блоках типа 2 и т. д. После блочного кодирования к битам, закодированным блочным кодом, добавляются так называемые «хвостовые» биты.

При перемежении обеспечивается преобразование $K3$ бит входной последовательности, полученной в результате сверточного кодирования, в $K4$ бит выходной последовательности, причем $K3 = K4$, т. е. перемежение не вносит в сигнал избыточность, а только производит перестановку бит в информационном блоке.

Таким образом, в системах связи, работающих на базе данного стандарта, перемежение используется для преобразования групповых ошибок, возникающих в канале связи из-за наличия глубоких замираний сигнала в условиях многолучевого распространения, в одиночные, с которыми легче бороться с помощью блочного и сверточного кодирования.

В стандарте TETRA применяется 2 вида перемежения: блочное и перемежение по N блокам.

8.3. Блочное перемежение

Обозначим блочный перемежитель как (K, J) , где K означает количество бит во входном информационном блоке, а J – количество бит, на которое разносятся соседние для входного блока биты. При блочном перемежении соответствие бит выходного блока битам входной последовательности, т. е. $b4(k) = b3(i)$, где $i = 1, 2, \dots, K - 1$, производится по следующему правилу:

$$k = 1 + ((J \times i) \bmod K). \quad (8.1)$$

Работу схемы блочного перемежителя можно представить как проведение последовательной построчной записи входной информационной последовательности в матрицу, в которой длина строки соответствует J (число столбцов $(K \operatorname{div} J + 1)$), а затем считывание записанной информации по столбцам. Оператор $\operatorname{div} J$ означает результат целочисленного деления.

Блочное перемежение с различными параметрами преобразования используется в каналах SCH/HD, SCH/HU, SCH/F, BNCH, STCH, BSCH и TCH/S.

8.4. Перемежение по N блокам

Перемежение по N блокам применяется в логических каналах передачи данных типа TCH/2,4 и TCH/4,8, информация по которым передается блоками, длиной 432 бита.

При данной схеме перемежения блочная перестановка бит дополняется так называемым *диагональным перемежением*. Его особенностью является то, что оно производится одновременно для нескольких последовательных блоков, т. е. может осуществляться перестановка битов из одного блока в другой. Количество блоков N , участвующих в перемежении, может быть равным 1, 4 или 8 ($N = 1$ – вырожденный случай, при котором перемежение осуществляется в пределах одного блока). При этом в результате перемежения образуется $M + N - 1$ блоков по 432 бита, где M – целое число, т. е. процедура перемежения по N блокам может увеличивать общее количество информационных блоков.

Таким образом, при перемежении по N блокам преобразование входной информации производится в 2 этапа: 1-й этап – диагональное перемежение с увеличением общего количества блоков, 2-й этап – блочное перемежение в каждом из полученных блоков.

Первоначально M блоков $B3(1), B3(2), \dots, B3(M)$, участвующих в преобразовании, с помощью диагонального перемежителя преобразуются в $M + N - 1$ блок $B3'(1), B3'(2), \dots, B3'(M + N - 1)$. Если обозначить k -й бит блока $B3'(m)$ как $b3'(m, k)$, где $k = 1, 2, \dots, 432$, а $m = 1, 2, \dots, M + N - 1$, то

$$b3'(m, k) = b3(m - j, j + 1 + (i \times N)) \quad \text{при} \quad 1 \leq m - j \leq M, \quad (8.2)$$

$$b3'(m, k) = 0 \quad \text{при} \quad 1 > m - j \quad \text{и} \quad m - j > M, \quad (8.3)$$

при этом j и k определяются следующими выражениями:

$$j = (k - 1) \operatorname{div}(432 / N), \quad (8.4)$$

$$i = (k - 1) \operatorname{mod}(432 / N). \quad (8.5)$$

Практически это означает формирование $M + N - 1$ блоков, каждый из которых состоит из $1/M$ частей следующих друг за другом M блоков (при $M = 2$ это половина блока, при $M = 3$ – $1/3$ блока и т. д.). Для первых и последних блоков, количество которых соответствует увеличению общего числа блоков, свободные разряды заполняются нулями. При $M = 2$ нулями заполняется вторая половина первого и первая половина последнего блока, при $M = 3$ нулями заполняются первые две трети первого блока, первая треть второго блока, первая треть предпоследнего блока и первые две трети последнего блока. При этом следует учитывать, что предварительно в каждом из блоков биты также переставляются путем сборки в блоке сначала бит с номерами $1, 1+M, 1+2M, \dots$, затем $2, 2+M, 2+2M, \dots$ и т. д.

На втором этапе производится блочное перемежение бит в каждом из полученных блоков в соответствии с правилом

$$b_4(m, i) = b_3'(m, k), \quad (8.6)$$

$$\text{где } i = 1 + ((103k) \operatorname{mod} 432), \quad (8.7)$$

$$k = 1, 2, \dots, 432. \quad (8.8)$$

9. ФОРМИРОВАНИЕ ЦИФРОВЫХ ПОТОКОВ

9.1. Вводные замечания

Цифровые потоки – это последовательности 0 и 1, передаваемые по линии связи. Нули и единицы могут нести информацию о речи, тексте, изображении и т.д. При этом скорости потоков будут, естественно, отличаться: для текста – 50...100 бит/с, для компьютерных данных – 200 бит/с и выше, для речи – 64 кбит/с, для подвижной «картинки» – более 100 Мбит/с.

Как же «строить» цифровые системы передачи? Сколько цифровых потоков можно объединять и направлять в одну линию связи – провод в электрическом кабеле, ствол в радиорелейной или спутниковой линиях, волоконный световод в оптическом кабеле? Можно ли стандартизировать скорости передачи?

Начнем с того, что узлы различных систем передачи должны быть однотипными, или унифицированными.

Цифровые системы передачи создают во всем мире; коммуникации связи не знают государственных границ. Каждая страна должна выпускать аппаратуру, согласовывая ее со стандартами, принятыми в других странах. Государства должны договориться, на каких принципах строить аппаратуру. С этой целью создан межгосударственный орган – Международный союз электросвязи (МСЭ). Он рекомендует строить цифровые системы передачи по иерархическому принципу.

Примером иерархического построения системы является календарь. Иерархия календаря состоит в следующем. За единицу измерения выбраны сутки. Семь суток объединяются в неделю. Из четырех или четырех с половиной недель образуется месяц. Три месяца составляют квартал. Четыре квартала – это год. Годы складываются в десятилетия и века, а века – в тысячелетия. При необходимости эту иерархию можно продолжить и «вниз» от суток: сутки состоят из 24 часов, час – из 60 мин и т.д.

Иерархия, рекомендованная для цифровых систем передачи, чем-то похожа на иерархию календаря. Прежде всего необходимо было выбрать некоторую единицу измерения – «элементарную» скорость цифрового потока, единую для всех стран и предприятий, выпускающих аппаратуру систем передачи, и позволяющую измерять скорость суммарных цифровых потоков. Такая «единичная» скорость во всем мире – скорость передачи оцифрованной речи, равная 64 кбит/с. Выбор этой величины в качестве единицы объединения цифровых потоков связан, скорее, с традициями, нежели с какими-то другими соображениями.

Канал, в котором информация передается со скоростью 64 000 бит/с, получил название основного цифрового канала (ОЦК). Возможности любой цифровой системы передачи оцениваются числом организованных с ее помощью именно таких стандартных каналов.

9.2. Цифровые иерархии

9.2.1. Плезиохронная цифровая иерархия

Исторически появившаяся первой плезиохронная цифровая иерархия (ПЦИ) имеет европейскую, североамериканскую и японскую разновидности (табл. 9.1).

Для цифровых потоков ПЦИ применяют соответствующие обозначения. Для североамериканской и японской ПЦИ применяется обозначение T (иногда DS), для европейской ПЦИ – E. Цифровые потоки первого уровня обозначаются соответственно T-1 и E-1, второго T-2 и E-2 и т.д.

Таблица 9.1

Иерархия цифровых потоков

Уровень иерархии	Европа		Северная Америка		Япония	
	Скорость, Мбит/с	Коэф. мультиплекс.	Скорость, Мбит/с	Коэф. мультиплекс.	Скорость, Мбит/с	Коэф. мультиплекс.
0	0,064	-	0,064	-	0,064	-
1	2,048	30	1,544	24	1,544	24
2	8,448	4	6,312	4	6,312	4
3	34,368	4	44,736	7	32,064	5
4	139,264	4	-	-	97,728	3

Чем выше ступень иерархии, тем мощнее цифровой поток, т. е. тем выше его скорость. К системам передачи, стоящим в самом низу иерархической лестницы, относится цифровой поток E1 (30 стандартных основных цифровых каналов). У подобных систем передачи скорость цифрового потока относительно невысока (около 2 Мбит/с для E1), что ограничивает сферу их применения.

9.2.2. Структура цифрового потока E1

Цифровой поток преобразуется в блоки стандартной логической структуры – циклы. Цикловая структура обеспечивает работу процедур мультиплексирования и демultipлексирования («вставки» и «вырезки» конкретного основного цифрового канала), передачу управляющей информации, а также встроенную диагностику по параметру ошибок в цифровой системе передачи. Существует три основных варианта цикловой структуры E1: неструктурированный поток, поток с цикловой структурой и с цикловой и сверхцикловой структурой.

Неструктурированный поток E1 используется в сетях передачи данных и не имеет цикловой структуры, т. е. деления на каналы (основные цифровые каналы).

Поток E1 с цикловой структурой предусматривает деление на 32 канала ОЦК по 64 кбит/с в форме деления на каналные интервалы (Time Slot – TS) от 0 до 31. Цикловая структура приведена на рис. 9.1. Для каждого каналного интервала в составе цикла отводится 8 бит, таким образом длина цикла равна 256 бит, что при заданной скорости передачи E1 составляет 125 мкс (длительность одного цикла). Тридцать каналных интервалов (1–15 и 17–31) используются для передачи информации (например голоса). Нулевой каналный интервал отводится под передачу сигнала цикловой синхронизации FAS (Frame Alignment Signal), шестнадцатый – для передачи служебной информации.

Структура FAS представлена на рис. 9.2. Различаются четные и нечетные циклы. В TS0 нечетных циклов передается сигнал FAS (на рисунке – первая строчка), который включает в себя последовательность цикловой синхронизации 0011011 и один служебный бит, зарезервированный под задачи международного использования. В TS0 четных циклов передается сигнал NFAS, не содержащий кодовой последовательности цикловой синхронизации.

В составе сигнала NFAS передается бит Si, зарезервированный под задачи международного использования, бит A, используемый для передачи сигналов о неисправностях, а также пять служебных бит Sn4, Sn5, Sn6, Sn7, Sn8, используемые для передачи сигналов сетевого управления первичной сети E1, диагностики и дополнительных процедур контроля ошибок.

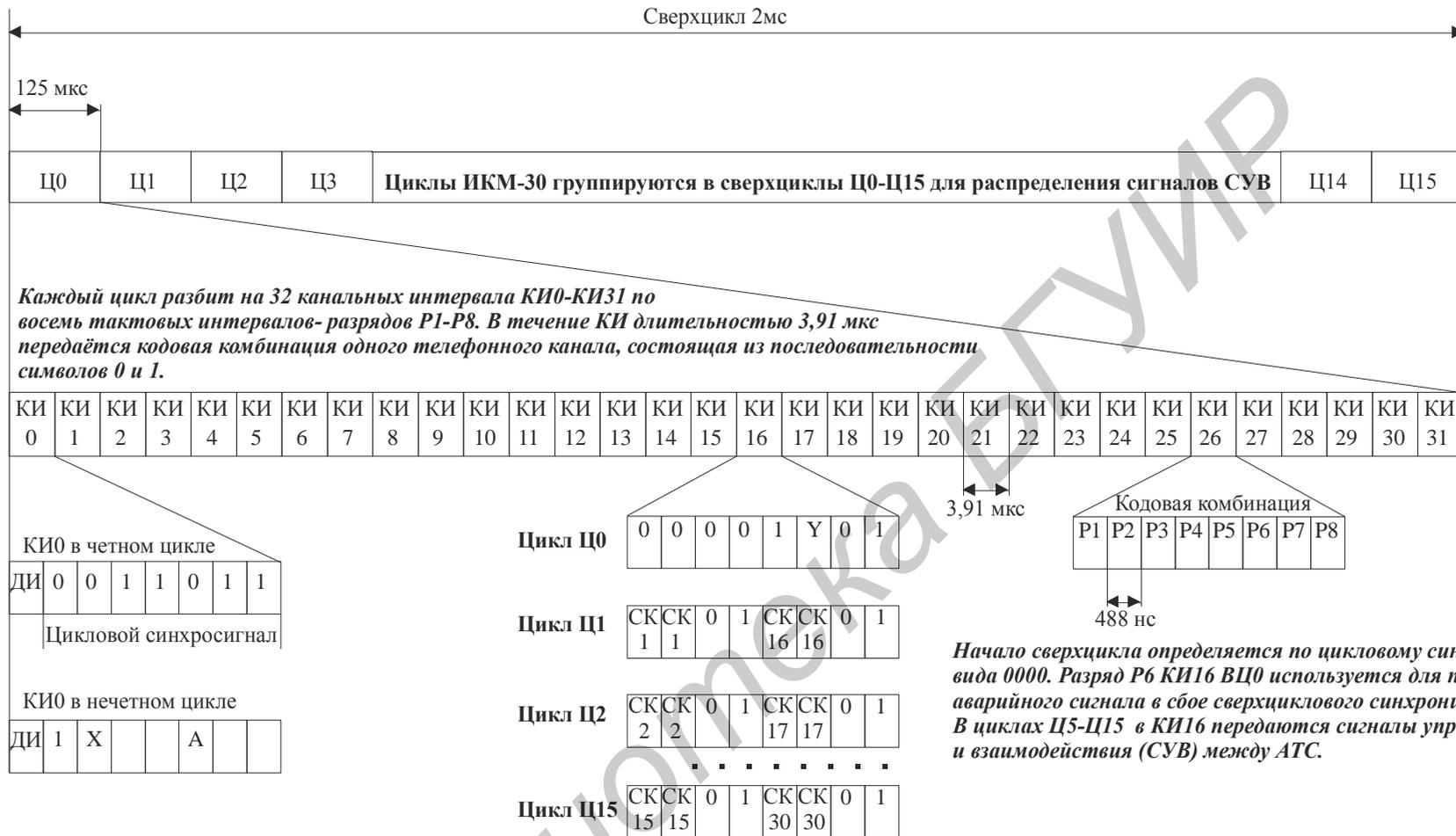
В отечественной терминологии вариант потока E1 с цикловой структурой получил название ИКМ-31. Он используется в ряде систем передачи данных, а также в некоторых приложениях ОКС7, ISDN и В-ISDN.

Сверхцикловая структура потока E1. В этом случае поток E1 имеет дополнительно к цикловой структуре еще и сверхцикловую структуру (MFAS – Multi Frame Alignment Signal). В отечественной терминологии такой вариант цикловой структуры E1 получил название ИКМ-30. При этом 16 циклов объединяются в сверхцикл размером 4096 бит и длительностью 2 мс. Когда идет передача/прием информации в виде сверхциклов MFAS, индивидуальная информация FAS каждого цикла теряет значимость. Необходимо рассматривать всю информацию FAS – 16 циклов.

Первый цикл содержит информацию MFAS о сверхцикле в шестнадцатом каналном интервале, а остальные 15 интервалов используются для передачи сигнальной информации. Структура MFAS показана на рис. 9.3.

Процедуры контроля ошибок передачи. Использование избыточного кода CRC-4. Передача потока E1 с цикловой структурой ИКМ-30 имеет важный механизм – процедуру встроенной диагностики параметров

Библиотека БГУИР



Начало цикла определяется по цикловому синхросигналу вида 0011011, передаваемому в разрядах Р2-Р8 КИ0 ЦЦ.

Первый разряд КИ0 отводится для передачи дискретной информации, разряд КИ3 НЦ – для передачи сигналов аварии о потери цикловой синхронизации "Х".

Прием значения 0 соответствует нормальному состоянию, а 1 – аварийному.

В тактовом интервале Р2 постоянно передается 1 (в отличие от Р2 в цикловом синхросигнале), что необходимо для проверки в процессе поиска ЦСС. Остальные разряды могут быть использованы для служебной информации.

Так, в первых модификациях АЦО шестой разряд используется для передачи сигнала контроля остаточного затухания А

Рис. 9.1 Структура потока Е1

Библиотека БГУИР

Time slot 0	1	...	31				
BITS							
1	2	3	4	5	6	7	8
Si	0	0	1	1	0	1	1
Si	1	A	Sn	Sn	Sn	Sn	Sn

Рис. 9.2. Структура нулевого канального интервала потока E1

Первая строка содержит сигнал FAS.

Вторая строка содержит сигнал NFAS.

Sn = биты, зарезервированные под задачи национального использования.

Si = биты, зарезервированные под задачи международного использования.

Последовательность FAS = 0011011.

ошибки. Для этой цели используется биты Si в составе циклового заголовка FAS и NFAS.

Процедура использует сверхцикловую структуру в 16 циклов, как показано на рис. 9.3, и механизм расчета параметра ошибки по контрольному избыточному коду CRC-4 (полином $X(4) + X + 1$).

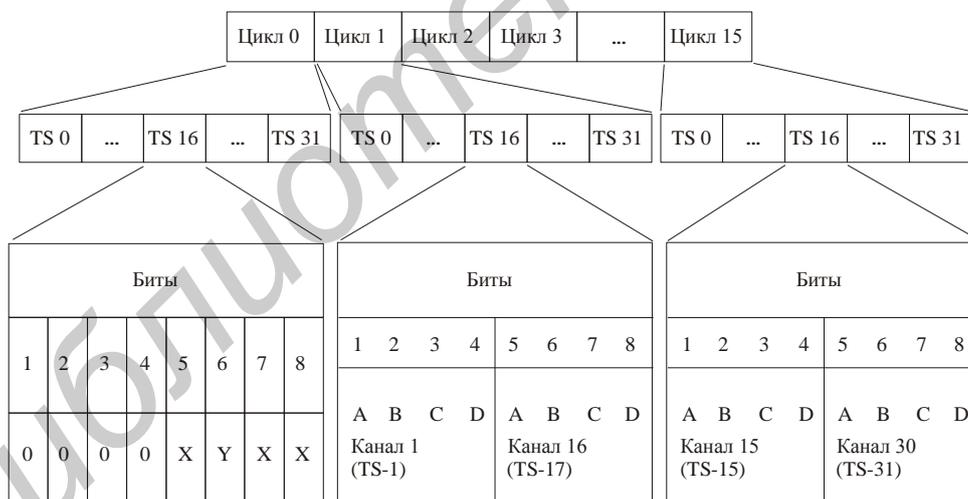


Рис. 9.3. Структура сверхцикла E1

Сигнал MFAS = 0000XYXX

X – запасные биты (1 обычно не используется);

Y – удаленная неисправность MFAS (равно 1 в случае, если потеряна сверхцикловая синхронизация).

Принцип CRC-4 базируется на простом математическом расчете, производимом в каждом сверхцикле данных. Оборудование передачи E1 произ-

водит расчет суммы CRC-4 и включает результаты суммы в сигнал следующего сверхцикла. Оборудование приемника принимает сигнал и производит аналогичный расчет и сравнение полученной суммы и переданной в следующем сверхцикле. Если в двух полученных суммах имеется расхождение, генерируется сигнал ошибки CRC-4. В настоящее время в оборудование и системы самодиагностики цифровых каналов ИКМ встраивается функция анализа по CRC-4. Часто возникает вопрос о целесообразности проведения анализа по BER, если система и так анализирует параметр ошибки по CRC-4.

Отвечая на этот вопрос, необходимо учесть два основных принципа использования CRC-4. Во-первых, каждая ошибка CRC-4 не обязательно связана с ошибкой одного бита информации. Несколько битовых ошибок в одном сверхцикле дадут только одну ошибку CRC-4 для блока. Во-вторых, несколько битовых ошибок могут компенсировать друг друга в смысле значения суммы CRC-4. CRC-4 является удобным методом контроля ошибок в процессе сервисного мониторинга при работающем канале, когда практически невозможно измерить реальные параметры ошибок по битам, поскольку невозможно добиться синхронизации тестовой последовательности. Тем не менее, в режиме с отключением канала необходимо проводить измерения ошибок по битам, поскольку результаты таких измерений более точны.

Рассмотрим механизм передачи информации о CRC-4. Для передачи этой информации используется сверхцикловая структура. Однако сверхцикл CRC-4 не обязательно связан со сверхциклом MFAS. Каждый сверхцикл может быть разбит на 2 подцикла (SMF). В табл. 9.2 ниже они отмечены как SMF#1 и SMF#2 и содержат по 8 циклов каждый. Соответственно отмечены по 4 бита, используемые для передачи CRC-4 каждого сверхцикла (биты C). Биты CRC-4 вычисляются методом деления двоичной суммы содержания 8 последовательных циклов на тестовый полином, остаток от деления в виде 4-битового сообщения вставляется в следующий сверхцикл для передачи в потоке E1.

Структура сверхцикла CRC-4 включает несколько сообщений сверхцикловой структуры CRC-4. В четных циклах MFAS битами S_i передается сообщение сверхцикловой синхронизации CRC-4 – комбинация 001011, которая используется на приемной стороне для синхронизации по CRC. Кроме этого, в составе сверхцикла SMF передается его главная информация – сообщение C1C2C3C4. В качестве дополнительного сообщения, передаваемого в направлении передачи при обнаружении ошибки по CRC-4, используется 2-битовое сообщение E1E2. Каждый бит E связан с одним из принятых сверхциклов SMF (в составе сверхцикла MFAS используется два сверхцикла SMF, поэтому сообщение использует два бита E). Когда оборудование приемника получает информацию об ошибке CRC-4, оно генерирует бит E для сообщения передатчику о принятой ошибке.

Мониторинг по CRC производится в режиме реального времени непосредственно после установления циклового синхронизма. В этом случае возникает цикловая синхронизация по CRC-4 и сверхциклам SMF по кодовой последовательности 001011. Эта синхронизация достаточно устойчивая, по-

теря цикловой синхронизации по CRC-4 отображается только после того, как более чем 914 сообщений CRC в секунду не соответствует ожидаемым значениям.

			Канальный интервал 0								
			Биты								
Сверхцикл	Подцикл	Цикл	Бит 1	Бит 2	Бит 3	Бит 4	Бит 5	Бит 6	Бит 7	Бит 8	
	FAS	1	C1	0	0	1	1	0	1	1	
	NFAS	2	0	1	A	Sn4	Sn5	Sn6	Sn7	Sn8	
	SMF#1	3	C1	0	0	1	1	0	1	1	
		4	0	1	A	Sn4	Sn5	Sn6	Sn7	Sn8	
		5	C3	0	0	1	1	0	1	1	
		6	0	1	A	Sn4	Sn5	Sn6	Sn7	Sn8	
		7	C4	0	0	1	1	0	1	1	
		8	0	1	A	Sn4	Sn5	Sn6	Sn7	Sn8	
		SMF#2	9	C1	0	0	1	1	0	1	1
			10	1	1	A	Sn4	Sn5	Sn6	Sn7	Sn8
	11		C2	0	0	1	1	0	1	1	
	12		1	1	A	Sn4	Sn5	Sn6	Sn7	Sn8	
	13		C3	0	0	1	1	0	1	1	
	14		E1	1	A	Sn4	Sn5	Sn6	Sn7	Sn8	
	15		C4	0	0	1	1	0	1	1	
	16		E2	1	A	Sn4	Sn5	Sn6	Sn7	Sn8	

Рис. 9.4. Структура подциклов SMF:

SMF#1 – подцикл 1;

SMF#2 – подцикл 2;

c1, c2, c3, c4 – биты CRC;

001011 – сигнал сверхцикловой синхронизации CRC;

Биты E1 и E2 идентифицируют ошибку по CRC (E=1 – нет ошибки, E=0 – ошибка по CRC).

Sn = биты, зарезервированные под задачи национального использования и для передачи управляющей информации сетевого уровня

Сетевой уровень потока E1. Стандартизация систем передачи E1 охватывает также третий, сетевой уровень, где осуществляются процедуры управления первичной сетью. При работе процедур управления они широко используют сигналы о неисправностях, генерируемые в современных цифровых системах передачи, а также сигналы о возникновении ошибок, фиксируемые встроенными средствами диагностики. Эта информация собирается в узлах системы управления и обрабатывается. Таким образом, сетевой уровень E1 включает в себя набор определенных служебных сигналов и сообще-

ний, используемых системой управления первичной сетью. Такие сообщения делятся на три категории:

- сообщения о возникновении ошибок в системе передачи;
- сообщения о неисправностях, возникающих в системе передачи;
- сообщения, используемые для реконфигурации первичной сети и восстановления плана синхронизации.

Последняя категория сообщений сетевого уровня E1, получивших название сообщений SSM (System Synchronization Messages – сообщения в системах синхронизации), будет отдельно рассматриваться в главе, посвященной проблемам построения и эксплуатации современных систем синхронизации. Здесь же мы рассмотрим первые две категории сообщений.

Сообщение о возникновении ошибок в системе передачи E1 использует сообщения E1E2. Действительно, сообщения, передаваемые битами E, служат подтверждением возникновения блоковой ошибки CRC и могут служить критерием качества цифровой системы передачи. Система управления анализирует значения бит E и собирает информацию о возникающих в системе передачи E1 ошибках.

Сообщения о неисправностях в системе передачи передаются циклами NFAS, а также битами MFAS в случае, если поток E1 имеет сверхцикловую структуру. Как было описано в предыдущем подразделе, в состав четных циклов NFAS входят биты, зарезервированные под задачи национального использования – бит A и биты Sn4, Sn5, Sn6, Sn7, Sn8. Именно эти биты используются для передачи различных сообщений о неисправностях в цифровой системе передачи. Биты Sn образуют своего рода канал управления, ресурсы которого используются сетевым уровнем E1. Помимо битов Sn для передачи сообщений о неисправностях могут использоваться биты XXУХ MFAS.

Эти биты используются главным образом для передачи сигналов о неисправностях в сверхцикловой структуре E1 (так, бит У непосредственно и определяется как индикатор неисправности MFAS на удаленном конце). Бит A (иногда он называется также Sn3) представляет собой бит оперативного сигнала о неисправности. В случае возникновения существенной неисправности, требующей оперативного вмешательства, бит A становится равным единице. Такую существенную неисправность называют RDI (Remote Defect Indication – индикация дефекта на удаленном конце).

В случае возникновения так называемого «неоперативного» сигнала неисправности, система передачи генерирует NFAS с инверсией бита Sn4 с 0 на 1. Рекомендации ITU-T устанавливают следующие возможные причины генерации такого сигнала:

- в случае, если параметр ошибки BER в FAS становится хуже, чем $10(-3)$;
- в случае неисправности в цепи питания кодека;
- в случае, если потерял входной сигнал или имеет место сбой цикловой синхронизации.

Неоперативный сигнал о неисправности дает возможность получения информации о значительном увеличении параметра ошибки на стороне передатчика. Оборудование приемника обычно имеет установленные пороговые значения для генерации сигнала Sn4. В случае увеличения параметра ошибки более порога приемник генерирует сигнал «неоперативной» неисправности в направлении передатчика. При получении этого сигнала система управления может перевести передачу на резервный канал E1, за счет чего достигается высокое качество связи.

Биты Sn5, Sn6, Sn7, Sn8 образуют служебный канал передачи данных емкостью 2 кбит/с, который может использоваться для передачи сигналов о неисправностях. В последнее время в связи с развитием систем управления (в частности платформы TMN) возможности этого канала используются довольно широко. Речь здесь идет именно о канале, поскольку в этом случае важно не абсолютное значение бит Sn, а последовательность сигналов, генерируемых в этих битах. Каждый бит образует так называемый «вертикальный протокол», т. е. сообщение о неисправности передается не одним, а несколькими последовательными битами Sn.

Тип передаваемых сообщений и алгоритм их генерации устанавливается национальными стандартами, производителями оборудования или специальными требованиями (например операторов ведомственных сетей). Наибольшее распространение получил стандарт ETS 300-233, определяющий использование бит Sn в мультимплексах PRI ISDN. Генерируемые и принимаемые оборудованием сообщения могут успешно использоваться системами управления, которые работают на принципах анализа именно этих сообщений. Таким образом, сообщения о неисправностях представляют собой базу для создаваемых систем управления, их количество непосредственно определяет максимальный уровень интеллектуальности системы управления и перечень параметров, доступных для контроля сети.

Использование битов Sn дает возможность разграничить степень ответственности различных операторов. В качестве иллюстрации рассмотрим несколько примеров (рис. 9.2), где представлены два варианта использования битов Sn, позволяющих определить сторону, ответственную за ухудшение качества передачи.

9.2.3. Мультиплексирование цифровых потоков

К системам передачи, стоящим в самом низу иерархической лестницы, относится аппаратура ИКМ-30 (30 – число организуемых основных цифровых каналов), использующая цифровой поток первого уровня E1). У подобных систем передачи сравнительно невысокая скорость передачи (около 2 Мбит/с), что делает их пригодными для организации связи между АТС по обычным городским и сельским линиям связи, образующим довольно обширную сеть магистралей.

Скорость передачи по линиям связи более высокой иерархии увеличивается до 8 Мбит/с (цифровой поток E2). В этом случае происходит мультиплексирование (объединение 4 потоков E1. Аппаратура, осуществляющая это объединение, называется по числу образованных каналов – ИКМ-120. Скорость потока на выходе этой аппаратуры 8,448 Мбит/с.

Для образования высокоскоростных потоков объединяют цифровые потоки четырех систем ИКМ-120 (поток E3). В результате скорость передачи в линии возрастает до 34,368 Мбит/с. Число каналов в новой системе равно 480, поэтому она получила название ИКМ-480.

Поступая далее аналогичным образом, получаем при слиянии четырех потоков систем передачи ИКМ-480 суммарный цифровой поток со скоростью 139,264 Мбит/с (цифровой поток E4). Это уже аппаратура ИКМ-1920. На этом иерархия цифровых систем передачи не заканчивается. Можно продолжать укрупнять потоки и дальше.

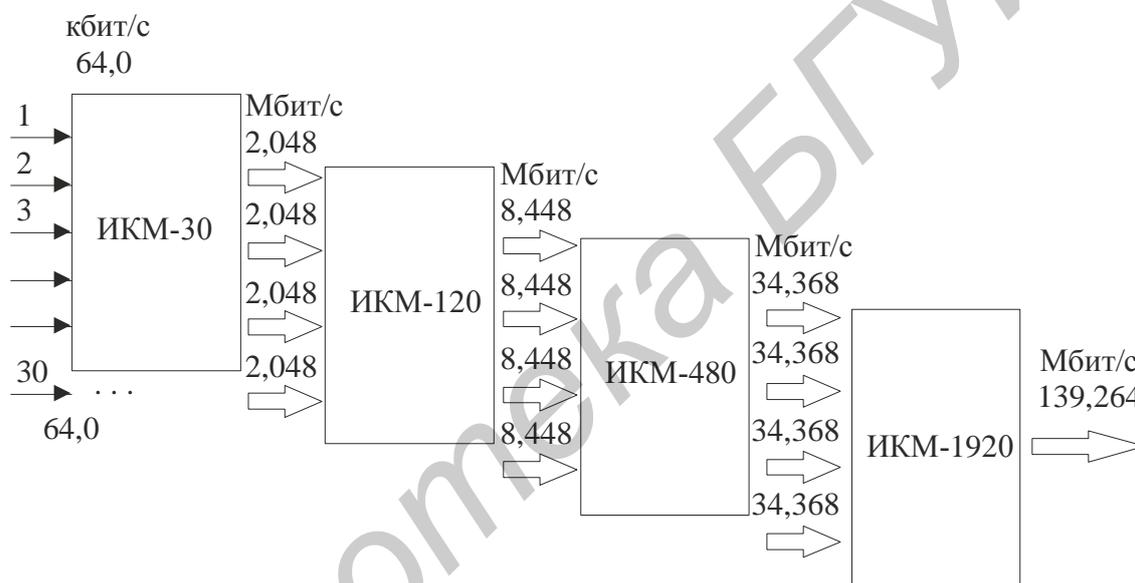


Рис. 9.5. Иерархия плезиохронных цифровых систем передачи

Принципы синхронизации остаются неизменными для систем передачи всех ступеней иерархии, сколько бы их не было: точно так же выделяются из цифрового потока тактовые импульсы и точно так же для обеспечения синхронной (а если точнее, синфазной) работы мультиплексоров и демultipлексоров посылаются в линию комбинации импульсов цикловой синхронизации. Правда, некоторые отличия все же есть. О них и пойдет речь дальше.

Дело в том, что в системах передачи, начиная со второй ступени иерархии (это аппаратура ИКМ-120, ИКМ-480, ИКМ-1920 и т.д.), объединение потоков выполняется по принципу чередования битов (на рис. 9.6) дан пример объединения двух потоков). Таких потоков – четыре, и скорость каждого из них 2,048 Мбит/с. Четыре «двери» мультиплексора передающей станции поочередно открываются и пропускают в линию по одному биту из каждого цифрового потока. Разумеется, что они должны успеть это сделать за время,

пока данные биты не успели смениться следующими. Затем все снова повторяется.

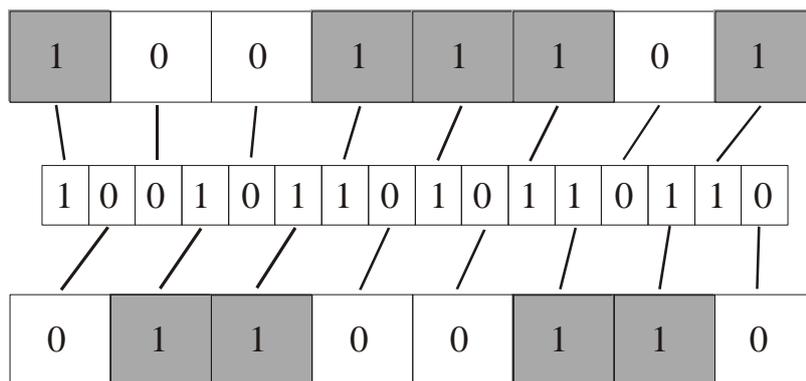


Рис. 9.6. Мультиплексирование двух потоков по принципу чередования бит

Понятно, что объединение потоков становится возможным только за счет укорочения в 4 раза длительности передаваемых импульсов, т. е. фактически за счет уменьшения в 4 раза времени передачи каждого из них. Но как же в этом случае ввести в цифровой поток сигнал цикловой синхронизации, ведь места-то для него нет? Вероятно, путь только один – укоротить информационные импульсы еще чуть-чуть. Пусть они немного «потеснятся», тогда в цикле передачи появятся «пустые» временные интервалы, в которые и можно будет вставлять синхросигнал.

Вот как это делается практически. Приходящие на вход системы передачи биты из четырех информационных потоков записываются в ячейки памяти ЗУ, а затем считываются с них и направляются в линию. Зачем так делать? Казалось бы, ничего не изменилось, только усложнилась аппаратура. Но это не так. Поскольку шины записи и считывания ЗУ независимы друг от друга, становится возможным записывать биты с одной скоростью, а считывать – с другой, чуть чаще. «Прочитали» содержимое ячеек памяти быстрее – вот и появилась во времени «дырка» для вставки синхроимпульсов (рис. 9.7). Если импульсы считывания не «отстают» и не «убегают вперед», а «идут» весьма стабильно, то в каждом потоке регулярно появляются «пустые» интервалы. В системе передачи ИКМ-120 таким «пустым» интервалом, не несущим никакой информации, является во всех потоках каждый 33-й интервал. При объединении потоков в линию поочередно посылаются импульсы каждого из них, а так как на указанных «пустых» интервалах ни в одном из потоков никаких информационных импульсов нет, то в общем потоке периодически образуются «дырки» шириной в четыре интервала. В них-то и «вставляют» синхроимпульсы, а также другую служебную информацию. Напомним, что строгая периодичность синхросигнала – это одно из важнейших свойств, используемое для его распознавания.

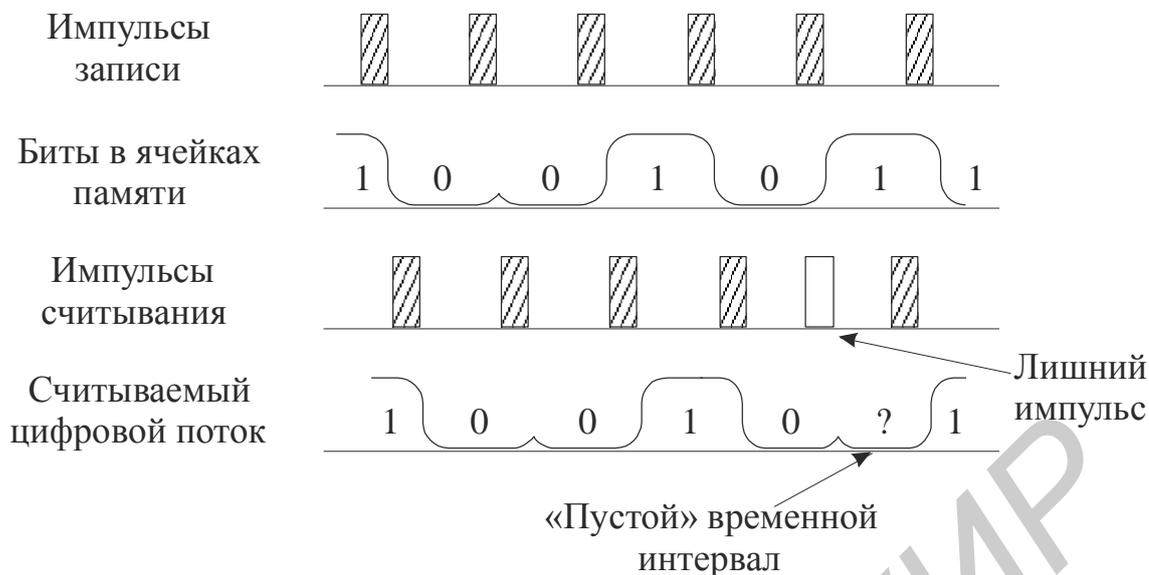


Рис. 9.7. Формирование «пустых» интервалов в цифровом потоке

Совсем иная картина будет наблюдаться, если местный генератор окажется не очень стабильным. В этом случае главные «часы» цифровой системы передачи (тактовые импульсы) могут «отставать» или «убегать вперед» по сравнению с их нормальным «ходом». В свою очередь, это будет вызывать смещение во времени «пустых» интервалов в каждом цикле передачи, и, значит, нарушится строгая периодичность их повторения. На каком-то этапе может произойти полный сбой в работе системы синхронизации и, как следствие, – всей аппаратуры в целом. Чтобы такого не случилось, местные «часы» (тактовые импульсы) нужно систематически «подводить». Подобная процедура реализована практически во всех современных системах передачи высших (начиная со второй) иерархий и называется «согласованием», а иногда «выравниванием», скоростей цифровых потоков, или скоростей следования тактовых импульсов записи и считывания.

Как же все происходит? Специальное устройство следит за взаимным положением импульсов записи и считывания. Пусть расстояние между соседними парами этих импульсов постепенно начинает уменьшаться. Значит, местный генератор ускорил свой бег и импульсы считывания начали следовать быстрее. Как только контролируемый интервал уменьшится до критической величины, наш строгий «контролер» подаст сигнал тревоги: «пустой» интервал появился раньше времени. Поскольку ему еще не время появляться, другое устройство введет в этот пустой интервал ложный импульс, не несущий никакой информации. Все происходит так же, как и в случае с нашими часами, когда, подводя их вперед, мы добавляем потерянные секунды. Вот и здесь мы тоже добавляем как бы потерянный импульс. Так достигается согласование, или выравнивание, скоростей записи и считывания цифровых потоков, которое в данном случае называется *положительным* (рис. 9.8).

Почему же обязательно нужно вставлять ложный импульс, не лучше ли взять да и притормозить чуть-чуть генератор тактовых импульсов? Дело в

том, что тактовые импульсы разных цифровых потоков могут, в принципе, и не совпадать точно друг с другом, а генератор – один на всю систему передачи. Представим, что будет, если начать его непрерывно подстраивать то под один цифровой поток, то под другой. Тут единственный путь – вставлять по мере необходимости в каждый из потоков ложные импульсы.

В американской технической литературе описанная выше процедура согласования скоростей называется коротко одним словом: «staffing», что по-русски произносится как «стаффинг», а переводится как «вставка».

После того как в цифровой поток введен ложный импульс, нужно передать на приемную станцию команду: «Внимание! Произошло согласование скоростей». Она служит сигналом для ликвидации на приеме ложного импульса. Такой командой может служить, например, посылка по служебному каналу единичного бита. В качестве служебного канала можно использовать один из «законных» пустых интервалов, не занятый синхроимпульсом. Итак, если на приемной стороне в служебном интервале объединенного цифрового потока появляется 1, это означает, что из выделенного низкоскоростного потока нужно исключить очередной импульс – он ложный.

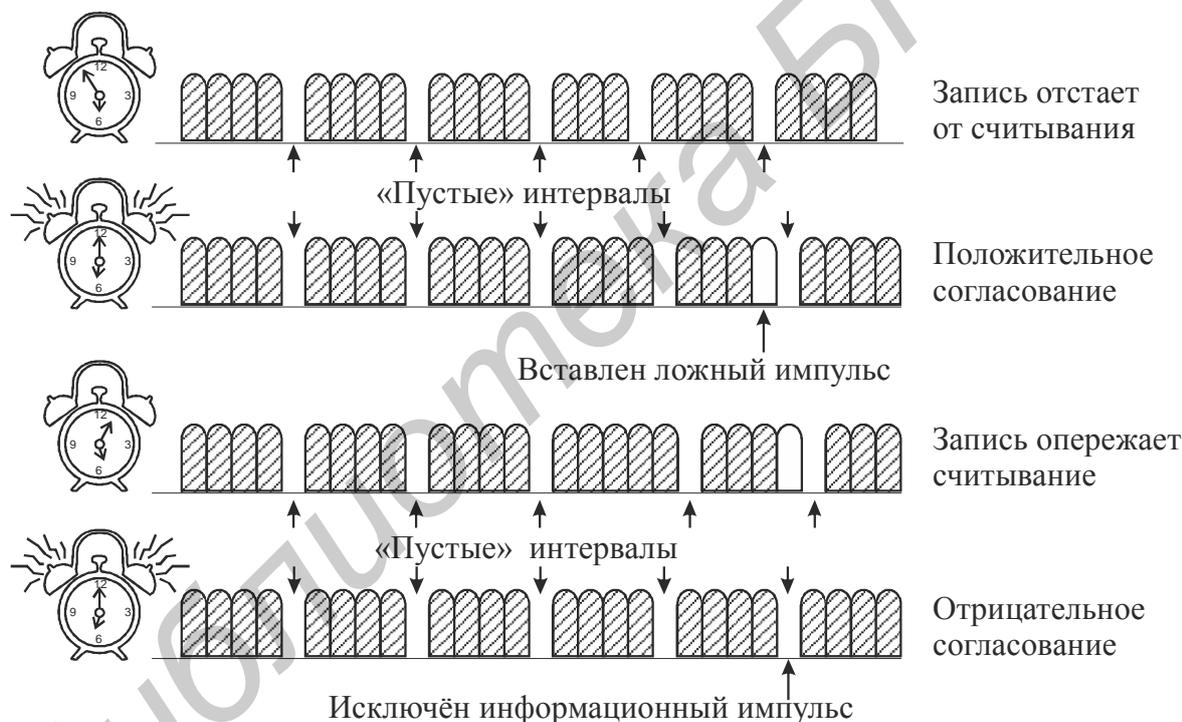


Рис. 9.8. Согласование скоростей цифровых потоков

Посылать по линии команду, состоящую всего из одного бита, является крайне неэффективным. Под воздействием помех 1 может превратиться в 0, а 0 в 1, и тогда случится непоправимое – информация будет декодирована неправильно. Поэтому для большей надежности команду согласования скоростей многократно дублируют, например, посылая ее 3 раза. В данном случае она будет иметь вид 111. Теперь, если в ней после воздействия помех останется только одна 1, команда все равно будет воспринята. Комбинация же

000 говорит о том, что согласование скоростей не производилось и пока все идет нормально.

До сих пор речь шла о том, что местный генератор может только «убежать вперед». Но с таким же успехом он может и «отставать», вырабатывая импульсы считывания реже, чем необходимо. Может случиться так, что в цифровом потоке уже должен появиться «пустой» интервал, а тактовые импульсы из-за замедленной их скорости до сих пор еще не считали из ЗУ предшествующий ему информационный импульс. Что делать в таком случае? Придется исключить из цифрового потока этот бит и предоставить временной интервал «по расписанию» для передачи очередной порции служебной информации (скажем, синхроимпульса). Только так можно согласовать, или выровнять, скорости тактовых импульсов записи и считывания. Такое согласование получило название отрицательного (см. рис. 9.8).

Если местные «часы» системы передачи (тактовые импульсы) подвоятся и в ту, и в другую сторону, то одной команды: «Внимание! Произошло согласование скоростей» будет мало. Нужно еще сообщить на приемную станцию, какое согласование произошло: положительное или отрицательное, вставлен ложный импульс или исключен информационный. Для этой цели вводят команду «Вид согласования», посылая по другому служебному каналу 1 при положительном согласовании и 0 – при отрицательном. Для надежности ее также повторяют 3 раза. Комбинация 111 во втором служебном канале (организованном также за счет части «пустых» интервалов) будет воспринята как сигнал о том, что в цифровой поток вставлен ложный импульс, а комбинация 000 – что из потока «вырезан» информационный бит. Устройства распознавания команд выполнены таким образом, что они сработают даже в том случае, если в командах «выживут» всего по одному биту, а остальные «погибнут» в борьбе с помехами.

Исключенный при передаче информационный бит также посылают по третьему служебному каналу, причем для верности тоже повторяют 3 раза. Итак, приемник цифровой системы передачи по первой команде (комбинация 111) узнает, что произошло согласование, по второй команде поймет, что нужно или ликвидировать ложный импульс (комбинация 111), или восстановить пропущенный информационный (комбинация 000), а по информации, взятой из третьего служебного канала, определит, какой бит пропущен – 1 (комбинация 111) или 0 (комбинация 000).

Объединение потоков с выравниванием скоростей получило название плезиохронного (почти синхронного), а существующая иерархия скоростей передачи цифровых потоков, а значит, и систем передачи типа ИКМ – плезиохронной цифровой иерархией (в англоязычном написании Plesiochronous Digital Hierarchy – PDH).

Плезиохронная цифровая иерархия была разработана в начале 80-х годов прошлого столетия. На системы передачи данной иерархии возлагались большие надежды. Однако она оказалась очень негибкой: чтобы вводить в цифровой поток высокоскоростной или выводить из него низкоскоростные потоки, необходимо полностью «расшивать», а затем снова «сшивать» высо-

коскоростной поток. Это требует установки большого числа мультиплексо-ров и демультиплексо-ров. Ясно, что выполнять эту операцию часто весьма дорого. На рис. 9.8 показана операция выделения потока со скоростью 2 Мбит/с из PDH потока со скоростью 140 Мбит/с.

В этом случае пришлось один поток со скоростью 140 Мбит/с демультиплексировать в четыре потока со скоростями 34 Мбит/с; затем один поток в 34 Мбит/с – в четыре потока 8 Мбит/с и только после этого «расшить» один поток 8 Мбит/с на четыре потока со скоростями 2 Мбит/с. Только таким сложным путем можно вывести или ввести поток пользователя в PDH-системах передачи.

Недостатком систем передачи плезиохронной цифровой иерархии является также то, что при нарушении синхронизации группового сигнала восстановление синхронизации первичных цифровых потоков происходит многоступенчатым путем, а это занимает довольно много времени.

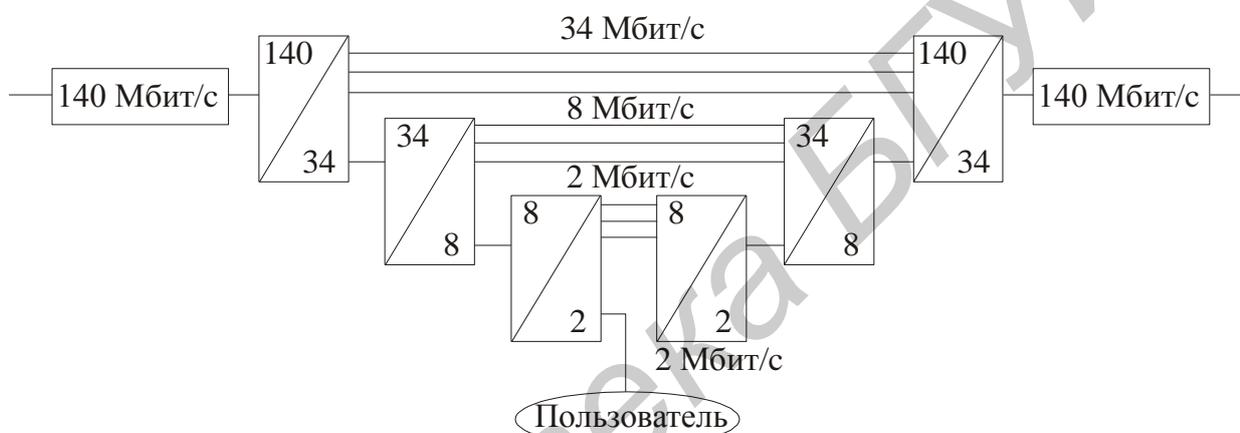


Рис. 9.9. Выделение сигнала со скоростью 2 Мбит/с из плезиохронного цифрового потока 140 Мбит/с

В настоящее время среди систем передачи PDH «выживают» только системы первого уровня иерархии, снабженные новой аппаратурой так называемого гибкого мультиплексирования, которая обеспечивает кроссовые соединения каналов 64 кбит/с; выделение и ввод отдельных каналов 64 кбит/с в любом наборе; пользовательские интерфейсы от двухпроводных окончаний для телефона до окончаний базового доступа в цифровую сеть с интеграцией услуг; видеоконференц-связь и многое другое.

Но самое главное, что заставило уже в середине 80-х годов XX в. искать новые подходы к построению цифровых иерархий систем передачи, это почти полное отсутствие возможностей автоматически контролировать состояние сети связи и управлять ею. А без этого создать надежную сеть связи с высоким качеством обслуживания практически невозможно. Все эти факторы и побудили разработать еще одну цифровую иерархию.

9.2.4. Синхронная цифровая иерархия

Синхронные транспортные модули. Новая цифровая иерархия была задумана как скоростная информационная автострада для транспортирования цифровых потоков с разными скоростями. В этой иерархии объединяются и разъединяются потоки со скоростями 155,520 Мбит/с и выше. Поскольку способ объединения потоков был выбран синхронный, то данная иерархия получила название синхронной цифровой иерархии (Synchronous Digital Hierarchy – SDH).

Для передачи цифрового потока в SDH генерируются синхронные транспортные модули.

Важной особенностью SDH является то, что в заголовках, помимо маршрутной информации, наличествуют данные, позволяющие обеспечить управление всей сетью в целом, обеспечивать дистанционные переключения в мультиплексорах, реализовывать эффективность эксплуатации сети и сохранять качество на должном уровне.

Для транспортирования цифрового потока со скоростью 155 Мбит/с создается синхронный транспортный модуль (Synchronous Transport Module) STM-1. Его упрощенная структура дана на рис. 9.10.



Рис. 9.10. Структура синхронного транспортного модуля

Модуль представляет собой фрейм (рамку) $9 \times 270 = 2430$ байт. Кроме передаваемой информации (называемой в литературе полезной нагрузкой), он содержит в 4-й строке указатель (Pointer, PTR), определяющий начало записи полезной нагрузки.

Чтобы определить маршрут транспортного модуля, в левой части рамки записывается секционный заголовок (Section Over Head – SOH). Нижние $5 \times 9 = 45$ байт (после указателя) отвечают за доставку информации в то место сети, к тому мультиплексору, где этот транспортный модуль будет переформировываться. Данная часть заголовка так и называется: секционный заголовок мультиплексора (MSOH). Верхние $3 \times 9 = 27$ байт (до указателя) представ-

ляют собой секционный заголовок регенератора (RSOH), где будут осуществляться восстановление потока, «поврежденного» помехами, и исправление ошибок в нем.

Один цикл передачи включает в себя считывание в линию такой прямоугольной таблицы. Порядок передачи байт – слева направо, сверху вниз (так же, как при чтении текста на странице). Продолжительность цикла передачи STM-1 составляет 125 мкс, т.е. он повторяется с частотой 8 кГц. Каждая клеточка соответствует скорости передачи 8 бит×8 кГц = 64 кбит/с. Значит, если тратить на передачу в линию каждой прямоугольной рамки 125 мкс, то за секунду в линию будет передано $9 \times 270 \times 64$ кбит/с = 155520 кбит/с, т.е. 155 Мбит/с.

Для создания более мощных цифровых потоков в SDH-системах формируется следующая скоростная иерархия (табл. 9.3): 4 модуля STM-1 объединяются путем побайтового мультиплексирования в модуль STM-4, передаваемый со скоростью 622,080 Мбит/с; затем 4 модуля STM-4 объединяются в модуль STM-16 со скоростью передачи 2488,320 Мбит/с; наконец, 4 модуля STM-16 могут быть объединены в высокоскоростной модуль STM-64 (9953,280 Мбит/с).

Таблица 9.3

Синхронная цифровая иерархия

Уровень иерархии	Тип синхронного транспортного модуля	Скорость передачи данных, Мбит/с
1	STM-1	155,520
2	STM-4	622,080
3	STM-16	2488,320
4	STM-64	9953,280

Мультиплексирование происходит по следующей схеме: сначала каждые 4 модуля STM-1 с помощью мультиплексоров с четырьмя входами объединяются в модуль STM-4, затем четыре модуля STM-4 мультиплексируются таким же 4-входовым мультиплексором в модуль STM-16. Следует отметить, что существует мультиплексор на 16 входов, с помощью которого можно одновременно объединить 16 модулей STM-1 в модуль STM-16.

Формирование модуля STM-1. В сети SDH применены принципы контейнерных перевозок. Подлежащие транспортировке сигналы предварительно размещаются в стандартных контейнерах (Container – C). Все операции с контейнерами производятся независимо от их содержания, чем и достигается прозрачность сети SDH, т. е. способность транспортировать различные сигналы, в частности, сигналы PDH.

Наиболее близким по скорости к первому уровню иерархии SDH (155,520 Мбит/с) является цифровой поток со скоростью 139,264 Мбит/с, образуемый на выходе аппаратуры плезиохронной цифровой иерархии ИКМ-

1920 (использует поток E4). Его проще всего разместить в модуле STM-1. Для этого поступающий цифровой сигнал сначала «упаковывают» в контейнер (т. е. размещают на определенных позициях его цикла), который обозначается С-4.

Рамка контейнера С-4 содержит 9 строк и 260 однобайтовых столбцов. Добавлением слева еще одного столбца – маршрутного или трактового заголовка (Path Over Head – ПОН) – этот контейнер преобразуется в виртуальный контейнер VC-4.

Наконец, чтобы поместить виртуальный контейнер VC-4 в модуль STM-1, его снабжают указателем (PTR), образуя таким образом административный блок AU-4 (Administrative Unit), а последний помещают непосредственно в модуль STM-1 вместе с секционным заголовком SOH (рис. 9.11 и 9.12).

Синхронный транспортный модуль STM-1 можно загрузить и плезиохронными потоками со скоростями 2,048 Мбит/с. Такие потоки формируются аппаратурой ИКМ-30, они широко распространены в современных сетях. Для первоначальной «упаковки» используется контейнер С12. Цифровой сигнал размещается на определенных позициях этого контейнера. Путем добавления маршрутного, или транспортного, заголовка (ПОН) образуется виртуальный контейнер VC-12. Виртуальные контейнеры формируются и расформируются в точках окончания трактов.

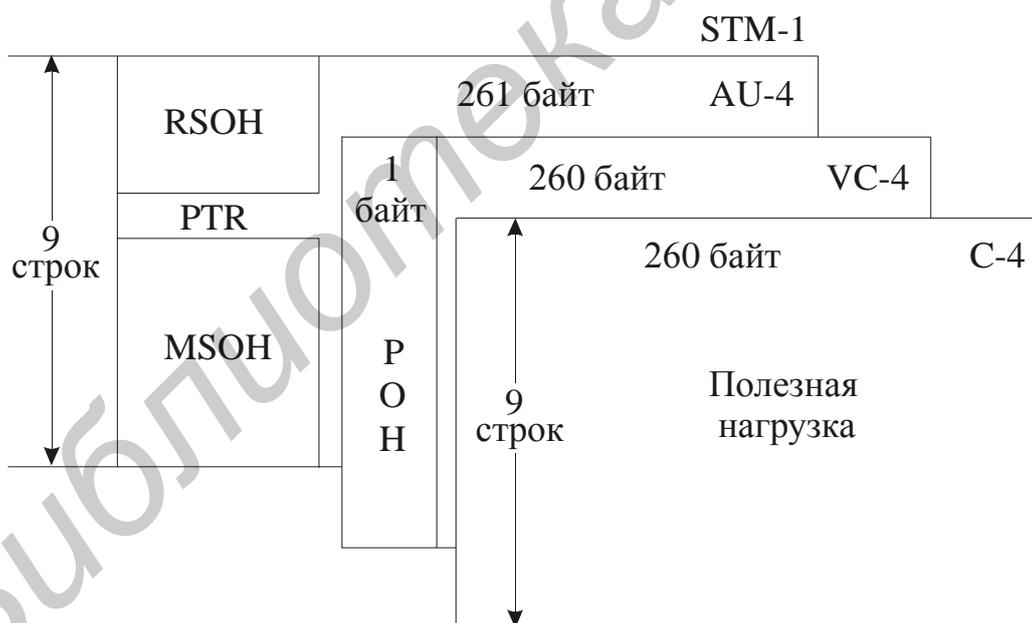


Рис. 9.11. Преобразования потока E4 в STM-1

В модуле STM-1 можно разместить 63 виртуальных контейнера VC-12. При этом поступают следующим образом. Виртуальный контейнер VC-12 снабжают указателем (PTR) и образуют тем самым транспортный блок TU-12 (Tributary Unit). Теперь цифровые потоки разных транспортных блоков можно объединять в цифровой поток 155,520 Мбит/с (см. рис. 9.12). Сначала три транспортных блока TU-12 путем мультиплексирования объединяют в груп-

пу транспортных блоков TUG-2 (Tributary Unit Group), затем семь групп TUG-2 мультиплексируют в группы транспортных блоков TUG-3, а три группы TUG-3 объединяют вместе и помещают в виртуальный контейнер VC-4. Далее путь преобразований известен.

На рис. 9.12 показан также способ размещения в STM-1 трех потоков от аппаратуры плезиохронной цифровой иерархии ИКМ-480 (34,368 Мбит/с).

Плезиохронные цифровые потоки всех уровней размещаются в контейнерах C с использованием процедуры выравнивания скоростей (положительного, отрицательного и двухстороннего).

Наличие большого числа указателей (PTR) позволяет совершенно четко определить местонахождение в модуле STM-1 любого цифрового потока со скоростями 2,048; 34,368 и 139,264 Мбит/с. Выпускаемые промышленностью мультиплексоры ввода-вывода (Add/Drop Multiplexer-ADM) позволяют ответвлять и добавлять любые цифровые потоки.

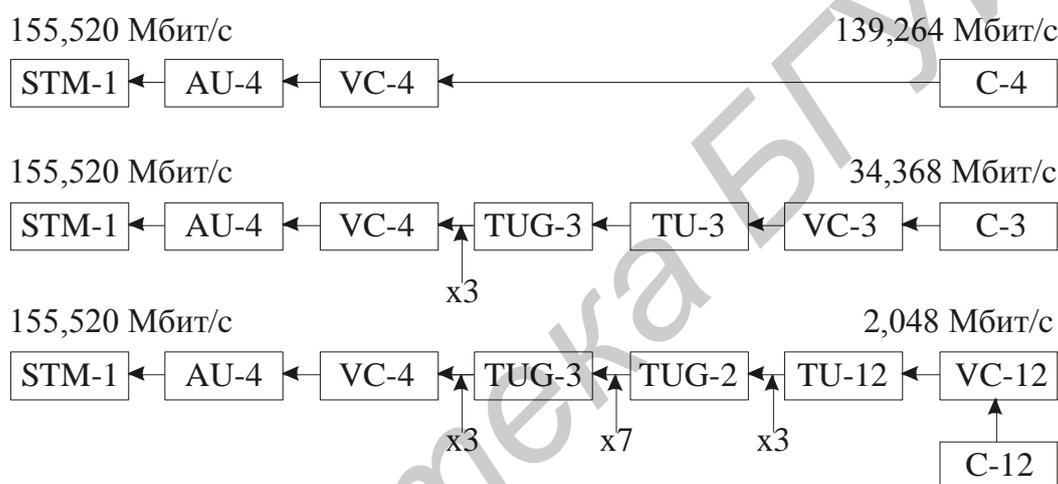


Рис. 9.12. Размещение потоков E1, E3 и E4 в контейнере STM-1

Важной особенностью аппаратуры SDH является то, что в трактовых и сетевых заголовках помимо маршрутной информации создается много информации, позволяющей обеспечить наблюдение и управление всей сетью в целом, дистанционные переключения в мультиплексорах по требованию клиентов, осуществлять контроль и диагностику, своевременно обнаруживать и устранять неисправности, реализовать эффективную эксплуатацию сети и сохранить высокое качество предоставляемых услуг.

9.2.5. Методы асинхронной передачи

В последнее десятилетие к передаче информации стали предъявляться более широкие требования. Одному и тому же абоненту могут быть переданы различные по характеру сообщения: движущиеся изображения (видеотелефон, видеоконференция); компьютерные данные (файлы); электронная почта; информация из системы дистанционного обучения (в том числе мультимедийная).

тимедийная); фильмы по кабельному телевидению и др. Причем источники этой информации являются, как правило, асинхронными. Информация от одних источников может поступать непрерывно, от других – время от времени. Скорость поступления информации от различных источников различная. Так, речевой поток поступает со скоростью 64 кбит/с, а передача движущегося изображения требует скорости от 1,5 до 100 Мбит/с.

Для согласования всех этих различных требований в 1980–1990-х гг. была предложена новая технология передачи, получившая название моды асинхронной передачи (Asynchronous Transfer Mode – АТМ). Асинхронный режим передачи (Asynchronous Transfer Mode, АТМ, МАП в русской транскрипции) – сетевая технология, предоставляющая возможность передавать данные, оцифрованную видео- и аудиоинформацию, а также изображения.

Пользователь сам выбирает тип сети: сети с коммутацией каналов применялись для передачи информации, когда задержки сигнала нежелательны; сети с коммутацией пакетов применялись в случае, когда некоторые задержки допустимы. АТМ объединяет в себе преимущества первой и второй технологий.

В основе технологии АТМ лежит передача данных в виде ячеек фиксированной длины (53 байта) главным образом из одной локальной сети в другую с помощью глобальной сети. Эта технология предполагает запись любого вида информации в ячейки (Cells) фиксированной длины. Ячейки содержат (рис. 9.13) полезную информацию и заголовок (Header).

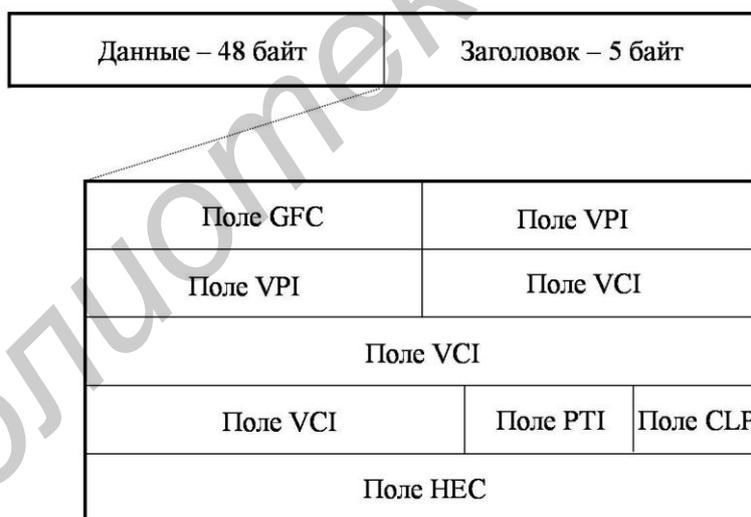


Рис. 9.13. Структура АТМ-ячейки

Поле GFC (Generic Flow Control) занимает первые 4 байта и предназначено для управления потоком между АТМ-сетью и пользователями.

Поле VPI (Virtual Path Identifier) определяет виртуальный маршрут, представляющий собой набор каналов, по которым передается информация (длина 8 бит или 12 бит, в зависимости от используемого интерфейса).

Второй частью идентификатора соединения является поле VCI (Virtual Channel Identifier). Имеет длину 16 бит и предназначено для идентификации соединения двух АТМ-станций.

Поле PTI (Payload Type Identifier – идентификатор полезной нагрузки) имеет длину 3 бита и определяет тип информации, передаваемой АТМ ячейкой (управляющая или данные).

Значение в поле CLP (Cell Loss Priority – приоритет ячейки) имеет длину 1 бит и определяет относительный приоритет данных, передаваемых в ячейке.

В поле HEC (Header Error Control), имеющее длину 8 бит, записывается результат контроля ошибок в заголовке, вычисленный с помощью циклического избыточного кода (CRC).

Цифровая информация от источников сообщений заполняет ячейки. Поскольку ячейки имеют фиксированную длину, то нет необходимости отделять их друг от друга (т. е. определять их начало и конец) с помощью служебной информации. Если у источника отсутствует потребность в передаче информации, то передаются пустые ячейки. Небольшие объемы информации, появляющиеся через большие временные интервалы, могут быть собраны для наполнения ячеек либо могут передаваться в отдельных ячейках. В последнем случае в полупустые ячейки добавляется «наполнитель». Ячейки формируются источниками по мере потребности. В случае непрерывной передачи (речь, видеоконференция и т. п.) ячейки следуют через строго определенное время.

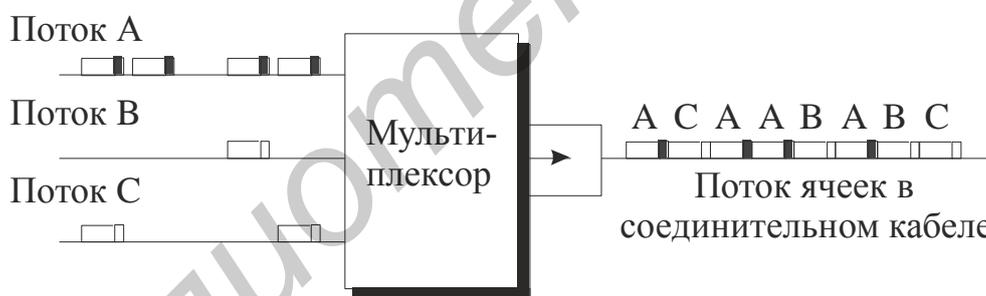


Рис. 9.14. Структура АТМ-ячейки

Потоки ячеек от различных источников могут быть объединены с помощью временного мультиплексирования (рис. 9.14).

Преимуществом МАП-ячеек является то, что их очень легко обрабатывать при прохождении через коммутатор. Прочитав идентификатор канала в заголовке ячейки, коммутатор переправляет ее из одного порта в другой вне зависимости от той информации, которая находится в ячейке.

Коммутатор может переключать целые группы виртуальных каналов, не проводя анализа информации по каждому каналу в отдельности. Для этого несколько виртуальных каналов, проходящих по одному и тому же направлению на каком-либо участке сети связи, объединяют в виртуальный путь. VCI и VPI образуют уникальный индивидуальный адрес маршрута на каждом отдельном участке сети. Изменение идентификаторов может происхо-

дить в каждом промежуточном коммутаторе. Назначение маршрута передачи ячеек может быть осуществлено оператором сети или сигнальной системой.

Технология АТМ имеет свою эталонную модель по аналогии с 7-уровневой моделью взаимодействия открытых систем (OSI – Open System Interconnection).

Основу эталонной модели АТМ составляют физический уровень, уровень АТМ и уровень адаптации АТМ (рис. 9.15). В АТМ фактически физический уровень не определяется – при создании данной технологии изначально была предусмотрена возможность ее работы с различными средами.

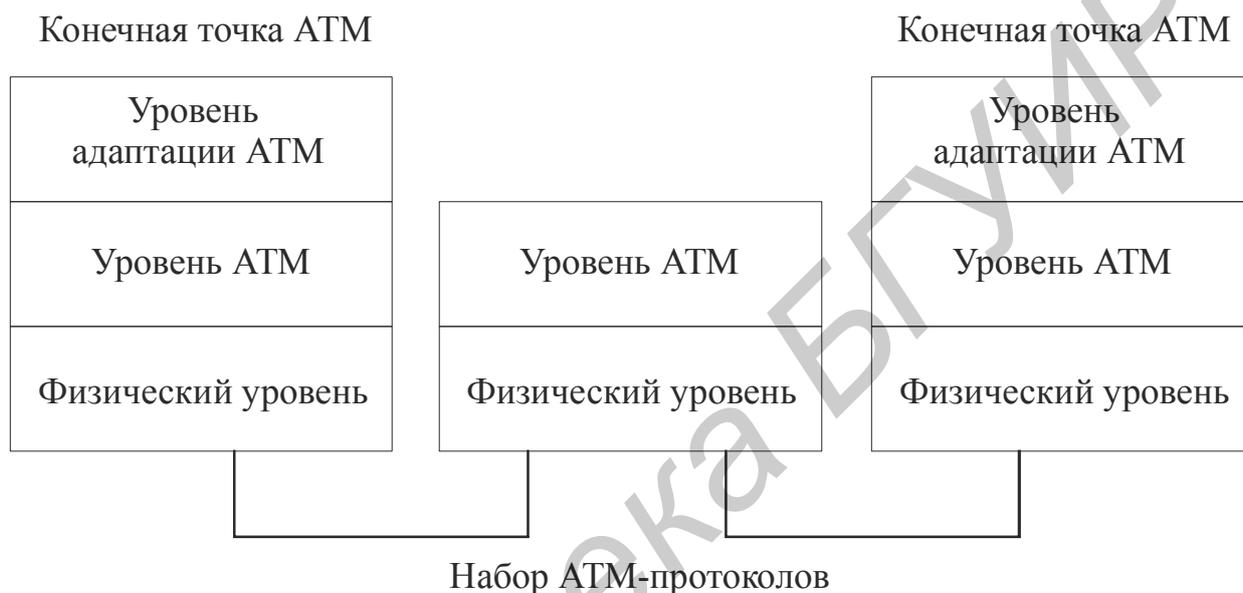


Рис. 9.15. Эталонная модель АТМ

Технология АТМ хорошо согласуется с технологией SDH: МАП-ячейки могут быть помещены в синхронный транспортный модуль STM-1. Для этого они сначала «упаковываются» в виртуальный контейнер VC-4, а в заголовке (РОН) этого контейнера отмечается начало записи МАП-ячеек (рис. 9.16). Затем контейнер, как обычно, помещается в модуль STM-1.

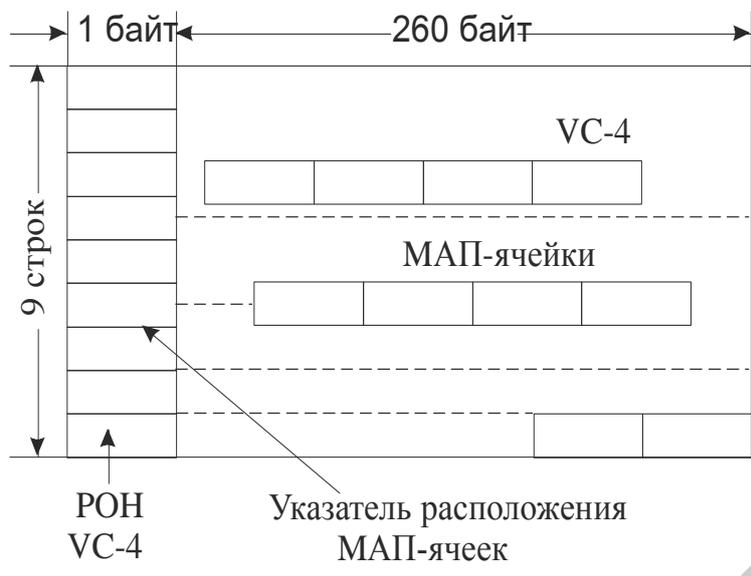


Рис. 9. 16. Размещение потоков АТМ-ячеек в контейнере STM-1

Библиотека БГУИР

10. МОДУЛЯЦИЯ В ОСНОВНОЙ (ИНФОРМАЦИОННОЙ) ПОЛОСЕ ЧАСТОТ (BASEBAND MODULATION)

10.1. Основные принципы модуляции

Для начала определим понятие «модуляция». В наиболее широком понимании модуляция это процесс изменения по заданному закону во времени величин, характеризующих какой-либо регулярный физический процесс.

Одна из особенностей информации как объекта заключается в ее взаимосвязи с материальными объектами. Информация непосредственно связана с событиями, происходящими с реальными объектами, и таким образом, наличие той или иной информационной нагрузки в процессах, происходящих с материальными объектами, может, в общем случае, трактоваться как процесс информационной модуляции данных объектов. Таким образом, информация вне материальных объектов не существует и ее передача возможна только посредством использования материального носителя. Применительно к технике электросвязи в качестве носителей выступают сигналы. В качестве данных сигналов используется электромагнитное поле в направляющих системах (проводниках) или в пространстве.

Следует различать два вида модуляции, используемых в электросвязи. К первому относится так называемая BaseBand-модуляция, или модуляция в основной полосе частот. Отличительная особенность данного вида модуляции заключается в том, что скорость изменения во времени сигнала-переносчика соизмерима со скоростью изменения информации. В этом случае сигнал-переносчик может ассоциироваться непосредственно с информацией. Например, голосовое сообщение, носителем которого являются звуковые волны, непосредственно ассоциируется с воспринимаемым звуком. Данное свойство позволяет называть этот вид модуляции также модуляцией в информационной полосе частот.

Ко второму виду модуляции относится модуляция, использующая в качестве переносчика информационной составляющей физического носителя скорость изменения которого во времени значительно превосходит скорость изменения информации. Данный вид информации получил название полосовой модуляции. Характерным примером полосовой модуляции являются радиосигналы, у которых сигнал-переносчик представляет собой гармоническое колебание, скорость изменения которого значительно выше модулирующей информационной составляющей.

Модуляция в основной полосе частот может рассматриваться как первичная или «получающаяся сама собой» (информация как бы самостоятельно присоединяется к носителю), а полосовая модуляция – как вторичная, т. е. результат воздействия первично модулированных носителей информации на другие материальные объекты.

Применительно к технике цифровой радиосвязи информация преобразуется в наборы символов, которые ассоциируются с конкретными сигналами (например, символ логической «1» связывается с высоким уровнем сигнала, а

символ логического «0» – с низким уровнем), таким образом происходит первичная модуляция или модуляция в основной полосе частот электрического сигнала.

Особенность систем электросвязи заключается в необходимости передачи информации на достаточно большие расстояния. Данная особенность накладывает определенные требования на сигналы-переносчики информации.

Применительно к сигналам, модулированным в основной полосе частот, к основным требованиям можно отнести:

1. Отсутствие постоянной составляющей в сигнале-переносчике информации.
2. Нечувствительность к смене «полярности» используемой линии передачи.
3. Простота и надежность восстановления информации на приемной стороне (высокая помехоустойчивость, простота процесса синхронизации и др.).

В системах цифровой связи сигналы с модуляцией в основной полосе частот часто называют кодами линии или линейными кодами (не следует путать с линейными помехоустойчивыми кодами).

Коды линии можно условно разделить на две группы: сигналы с импульсной модуляцией (ИМ) и сигналы с импульсно-кодовой модуляцией (ИКМ).

К группе сигналов с импульсной модуляцией относятся сигналы, сформированные в соответствии с методами, изложенными ниже.

Амплитудно-импульсная модуляция (АИМ). Сигнал в этом случае представляет собой последовательность импульсов постоянной длительности, следующих с фиксированной частотой. Информационная составляющая закладывается в амплитуду данных импульсов.

Частотно-импульсная модуляция (ЧИМ). Сигнал представляет собой последовательность импульсов постоянной длительности и амплитуды. Информационная составляющая закладывается в частоту повторения данных импульсов.

Широтно-импульсная модуляция (ШИМ). Сигнал представляет собой последовательность импульсов постоянной амплитуды с фиксированной частотой повторения. Информационная составляющая закладывается в длительность данных импульсов.

Фазоимпульсная модуляция (ФИМ), или импульсно-позиционная модуляция (ИПИМ)/Pulse-position modulation (PPM). Сигнал представляет собой последовательность импульсов фиксированной амплитуды и длительности. Информационная составляющая закладывается в изменение временного положения импульсов относительно опорного (синхронизирующего импульса).

Примечание. Указанные выше методы импульсной модуляции могут использоваться для переноса как непрерывной, так и дискретной (цифровой информации).

К группе сигналов с импульсно-кодовой модуляцией относятся сигналы, в которых информационная составляющая представляется сменой определенных кодовых последовательностей элементарных импульсов. При этом каждому информационному символу соответствует своя последовательность импульсов. Например, если информационные – это n -разрядные двоичные числа, то данные

последовательности, как правило, представляют собой комбинацию из n импульсов, представляющих логические символы «1» и «0».

В настоящее время в системах цифровой связи наиболее широкое распространение получили методы модуляции в основной полосе частот, основанные на ИКМ. Именно данные сигналы чаще всего называют кодами линии. Рассмотрим их более подробно.

10.2. Коды линии

10.2.1. Код без возвращения к нулю (NRZ)

Код без возвращения к нулю (NRZ – non-return-to-zero) является двоичным кодом линии, в котором символ логической «1» кодируется одним уровнем, а символ логического «0» – другим уровнем. При кодировании кодом NRZ отсутствует третье (промежуточное, нейтральное состояние).

Код NRZ не относится к числу кодов, обеспечивающих тактовую самосинхронизацию. Иными словами, для обеспечения тактовой синхронизации необходима реализация специальных мер по выделению тактовой частоты, поскольку в спектре сигнала с NRZ отсутствуют компоненты, представляющие тактовые импульсы.

Существует несколько разновидностей кода без возвращения к нулю: униполярный код без возвращения к нулю (UNRZ); биполярный код без возвращения к нулю (BNRZ); код без возвращения к нулю с инверсией (NRZI).

При кодировании UNRZ символ логической «1» кодируется высоким уровнем, а символ логического «0» – низким уровнем.

При кодировании BNRZ символ логической «1» кодируется высоким уровнем, а символ логического «0» – высоким уровнем другой полярности.

При кодировании NRZI при передаче логического «0» не происходит смены состояния кода, а при передаче логической «1» происходит смена состояния на противоположное.

На рис. 10.1 приведен пример кодирования последовательности двоичных символов различными разновидностями кода NRZ.

10.2.2. Манчестерский код (Manchester code)

Манчестерский код является бинарным кодом линии, в котором во время передачи информационного символа происходит по крайней мере одно изменение состояния. Данный код может рассматриваться как фазовая манипуляция некоторого колебания в виде меандра потоком информационных символов. Чаще всего в качестве «несущего» колебания используется сигнал тактового генератора. Если поток данных и сигнал тактового генератора представлены в коде BNRZ,

то результат кодирования манчестерским кодом эквивалентен перемножению этих кодов (см. рис. 10.2).

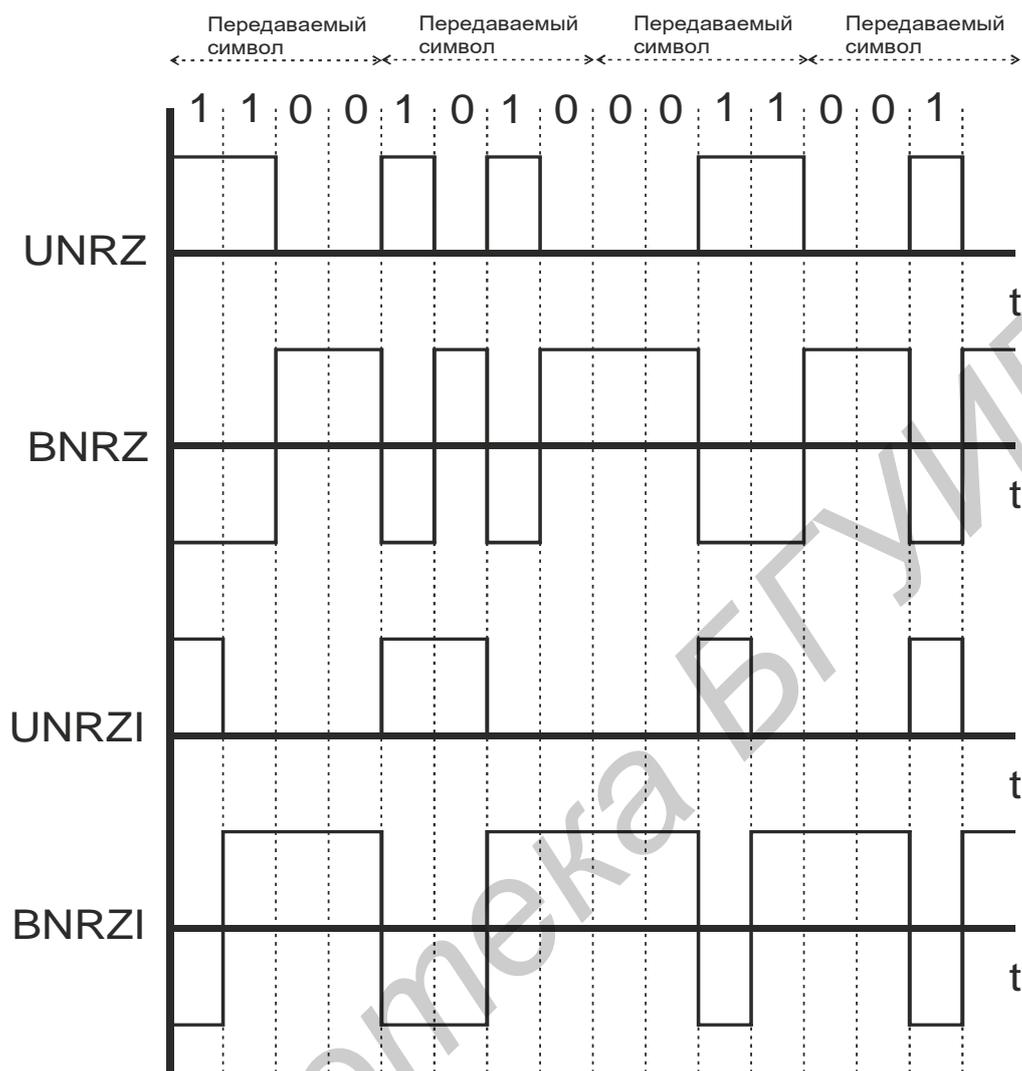


Рис. 10.1. Пример кодирования передаваемых символов кодом NRZ

Линейный манчестерский код не содержит постоянной составляющей и является самосинхронизирующимся. Данный линейный код нашел широкое распространение в сетях Ethernet и в системах радиочастотной идентификации RFID.

10.2.3. Код с возвращением к нулю (RZ)

Код с возвращением к нулю (RZ – return-to-zero) является двоичным кодом линии, в котором символ логической «1» кодируется одним уровнем, а символ логического «0» другим уровнем. При кодировании кодом NRZ вводится третье промежуточное, нейтральное состояние (нуль). Переход от одного символа к другому осуществляется только через промежуточное состояние. Промежуточное состояние вводится для реализации возможности самосинхронизации. Как и код

NRZ, код с возвращением к нулю может быть униполярным и биполярным (см. рис. 10.3).

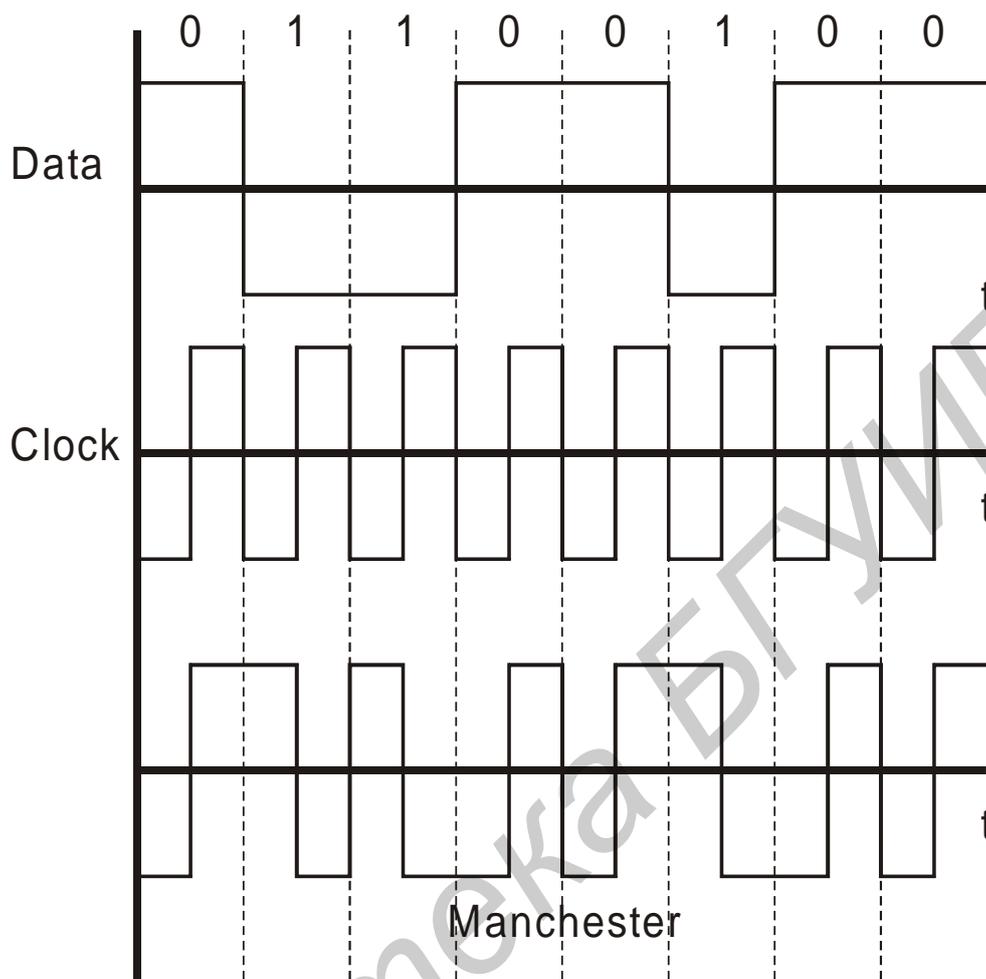


Рис. 10.2. Пример кодирования передаваемых символов манчестерским кодом

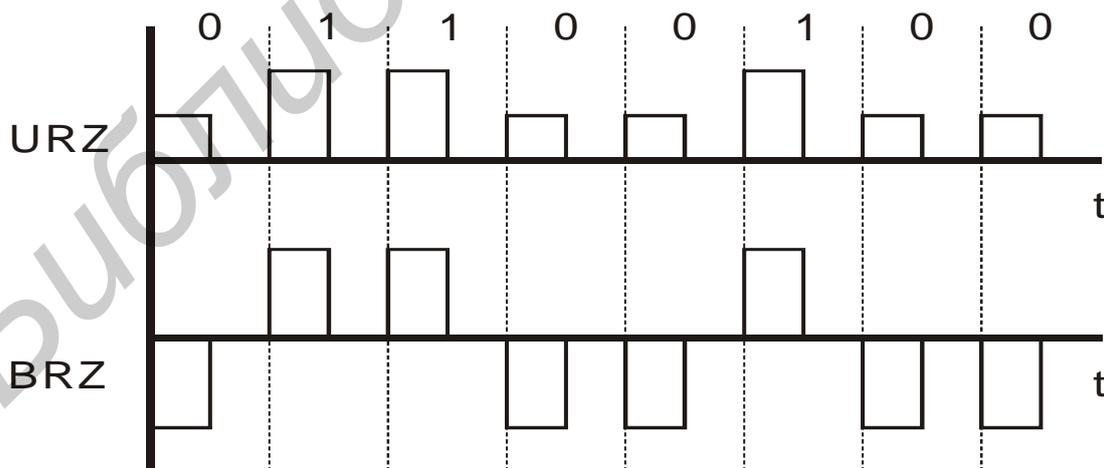


Рис. 10.3. Пример кодирования передаваемых символов кодом RZ

10.2.4. Код с чередующейся инверсией (АМІ)

Код с чередующейся полярностью импульсов (ЧПИ) – Alternate Mark Inversion (АМІ) – биполярный код, представляющий одну из разновидностей троичного кода: нулям соответствует отсутствие импульса, а единицам – попеременно чередующиеся прямоугольные импульсы отрицательной и положительной полярности. Так как постоянная составляющая импульсной последовательности равна нулю, то возможна передача по линиям, содержащим разделительные трансформаторы. Преимуществом данного кода является простота преобразования в двоичный код.

АМІ-код использует следующие представления символов (рис. 10.4):

- символы логического «0» представляются нулевым напряжением (0 В);
- символы логической «1» представляются поочередно значениями $-U$ или $+U$ (В).

АМІ-код обладает хорошими синхронизирующими свойствами при передаче серий единиц и сравнительно прост в реализации. Недостатком кода является ограничение на плотность (количество последовательно появляющихся символов) нулей в потоке данных, поскольку длинные последовательности нулей ведут к потере синхронизации.

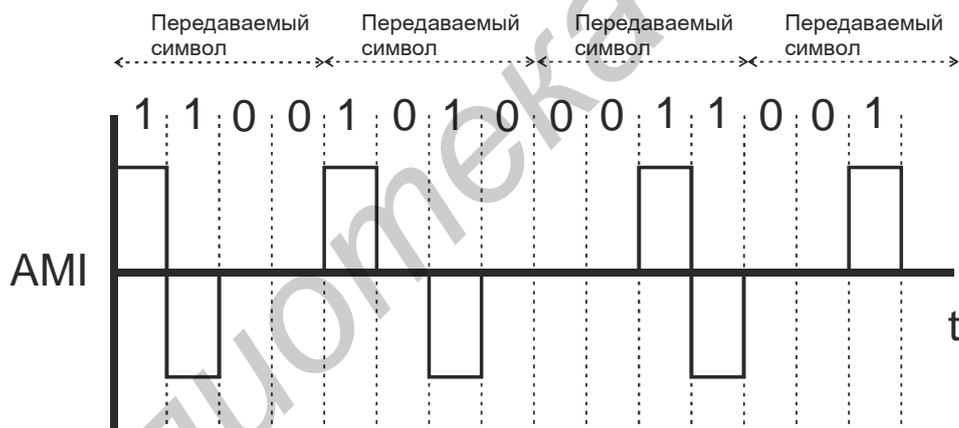


Рис. 10.4. Пример кодирования передаваемых символов кодом АМІ

10.2.5. Биполярный код высокой плотности порядка n (HDB- n)

Наиболее существенным недостатком кода с АМІ является необходимость передачи по линейному тракту сигнала с длинными сериями нулей, что может нарушить нормальную работу регенераторов, так как затрудняется процесс выделения тактовой частоты. С целью устранения этого недостатка было разработано несколько модификаций кода с АМІ, наибольшее распространение среди которых получил код HDB- n (наибольшее распространение получила версия HDB-3, принятая за стандартную для цифровых линий связи в Европе).

Принцип построения кода HDB- n такой же, как и кода с AMI, до тех пор, пока между двумя единичными символами не появится более $n + 1$ следующих подряд нулевых символов. При этом каждая последовательность из $n + 1$ нулевых символов (00...00) в исходном двоичном коде заменяется одной из двух последовательностей из $n + 1$ символов вида $B0...0V$ или $00...0V$, где B – импульс, по полярности противоположной предыдущему импульсу (отвечает правилу кодирования AMI), V (Violation) – импульс, по полярности соответствующий предыдущему импульсу (нарушающий правило кодирования AMI). Выбор замены ($B0...0V$ или $00...0V$) осуществляется таким образом, чтобы число импульсов B между двумя последовательно расположенными импульсами V было нечетным. На рис. 10.5 представлен пример кодирования в соответствии с правилами преобразования в HDB-3.

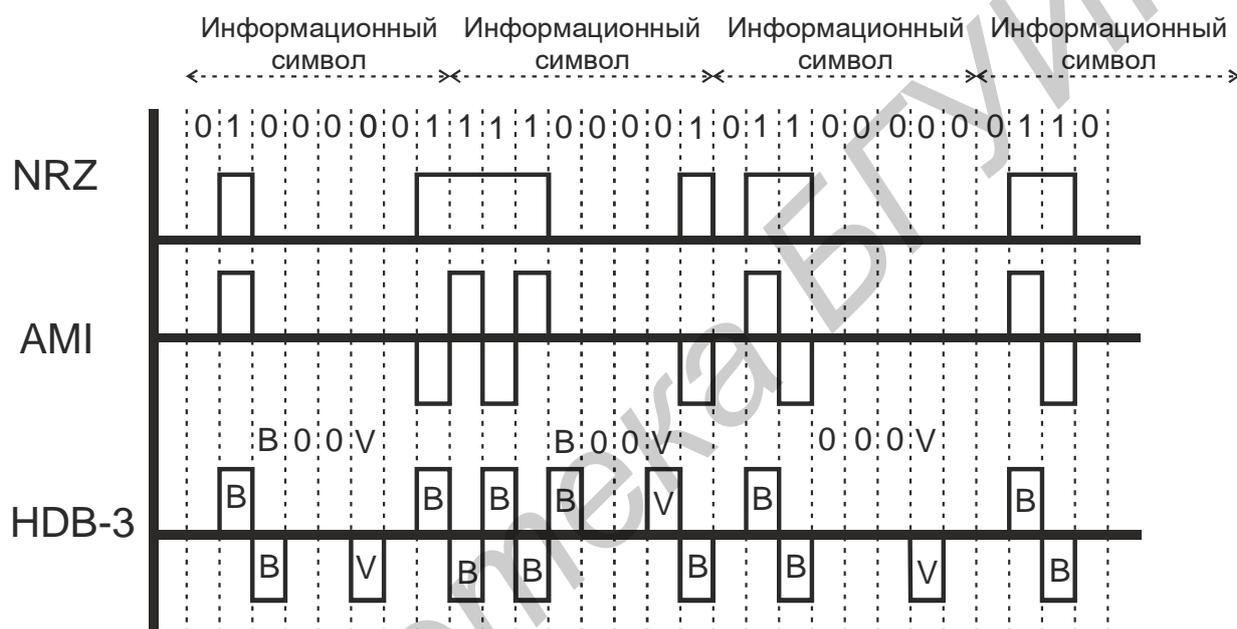


Рис. 10.5. Пример кодирования передаваемых символов кодом HDB-3

Использование двух замещающих последовательностей обеспечивает чередование полярности символов V , появляющихся в различных местах линейного кода, что, в свою очередь, позволяет устранить влияние этих символов на среднее значение, которое так же, как и у кода AMI, оказывается равным нулю.

11. ПОЛОСОВАЯ МОДУЛЯЦИЯ/ДЕМОДУЛЯЦИЯ (модуляция/демодуляция радиосигналов)

11.1. Основные принципы полосовой модуляции

Для передачи информации на расстояние применяются сигналы, эффективно излучаемые с помощью антенных устройств, обладающие способностью распространяться в виде свободных радиоволн в среде, разделяющей отправителя и получателя информации. Такими сигналами являются высокочастотные колебания. Передаваемая информация должна быть тем или иным способом заложена в высокочастотное колебание, называемое несущим. Несущее колебание необязательно должно являться гармоническим. В качестве несущего колебания могут использоваться квазигармонические («почти» или очень похожие на гармонические) колебания. Центральная частота ω_0 несущего колебания выбирается в зависимости от ряда технических и экономических факторов. Но в любом случае частота ω_0 должна быть велика по сравнению с наивысшей частотой спектра передаваемого сообщения и шириной спектра $\Delta\omega_c$ самого несущего колебания.

Это объясняется тем, что для неискаженной передачи сообщения через радиотехнические цепи, а также для устранения искажений, обусловленных распространением радиоволн, необходимо, чтобы ширина спектра радиосигнала $\Delta F_{s\ pb}$ (определяется шириной спектра сообщения ΔF_i и шириной спектра несущего колебания ΔF_c) была мала по сравнению с $f_0 = \omega_0 / 2\pi$; чем меньше отношение $\Delta F_{s\ pb} / f_0$, тем меньше проявляется несовершенство характеристик системы. Поэтому чем выше требуемая скорость передачи информации, и, следовательно, шире спектр радиосигнала $\Delta F_{s\ pb}$, тем выше должна быть несущая частота. Как правило, выполняется неравенство $\Delta F_{s\ pb} / f_0 \ll 1$.

Любой радиосигнал можно поэтому трактовать как «узкополосный» процесс даже при передаче «широкополосных» сообщений.

Основной принцип полосовой модуляции заключается в переносе информационного сообщения в несущий сигнал и последующей его передаче по физическому каналу. Полосовая демодуляция – это обратный процесс извлечения информационного сообщения из принятого полосового (радиосигнала).

Искажения передаваемого радиосигнала в физическом канале могут приводить к ошибкам в процессе демодуляции. Основная цель полосовой модуляции – передача информационных сообщений с высокой скоростью и минимизацией искажений передаваемых сообщений.

В общем случае модулированный радиосигнал может содержать информацию, закодированную в закон изменения амплитуды $A(t)$, частоты $f(t)$ или фазы $\Theta(t)$ несущего колебания.

Таким образом, модулированный радиосигнал может быть представлен следующим образом:

$$s(t) = A(t) \cos[2\pi(f_0 + f(t))t + \Theta(t) + \varphi_0] = A(t) \cos[2\pi f_0 t + \psi(t)], \quad (11.1)$$

где $\psi(t) = 2\pi f(t)t + \Theta(t) + \varphi_0$ – фазовый набег относительно гармонического несущего колебания $\cos[2\pi f_0 t]$. Следует отметить, что законы изменения амплитуды, частоты и фазы $A(t)$, $f(t)$, $\Theta(t)$ могут комбинировать в себе как закон изменения соответствующих параметров под воздействием передаваемого сообщения, так и законы изменения данных параметров в квазигармоническом несущем колебании.

Разложив косинус суммы, можно представить радиосигнал как комбинацию синфазной и квадратурной компонент:

$$s(t) = S_I(t) \cos[2\pi f_0 t] - S_Q(t) \sin[2\pi f_0 t], \quad (11.2)$$

где $S_I(t) = A(t) \cos[\psi(t)]$ и $S_Q(t) = A(t) \sin[\psi(t)]$ – синфазная и квадратурная компонента модулированного радиосигнала соответственно. Следует отметить, что рассмотренные представления радиосигнала представляют комбинированную амплитудно-угловую модуляцию.

Выбор метода модуляции (метода наложения передаваемой информации на параметры несущего колебания) определяется различными требованиями, предъявляемыми к радиосигналам. К данным требованиям относятся:

1. Спектральная эффективность – отношение скорости передачи информации к необходимой полосе частот, занимаемой радиосигналом [(бит/с)/Гц]. (Необходимая ширина полосы частот: ширина полосы частот, которая достаточна при данном классе излучения для обеспечения передачи сообщений с необходимой скоростью и качеством при определенных условиях).

2. Помехоустойчивость – минимальное отношение мощности полезного сигнала к мощности помехи (шума), обеспечивающее демодуляцию с качеством не хуже заданного (для систем цифровой радиосвязи принято характеризовать минимальным отношением энергии приходящейся на один бит передаваемой информации к спектральной плотности шума на входе демодулятора, обеспечивающее вероятность ошибки, не более заданной).

3. Компактность спектральной плотности – характеризуется уровнем мощности, проникающей в соседний частотный канал.

4. Постоянство огибающей – характеризуется глубиной амплитудной модуляции радиосигнала и накладывает существенные ограничения на возможность использования нелинейного усиления (например, усилителей, работающих в режиме С) в радиотрактах с целью повышения их энергетической эффективности.

11.2. Краткий обзор методов цифровой (дискретной) модуляции

Краткая характеристика и принятые международные сокращения основных видов цифровой модуляции приведены в табл. 11.1.

Основные методы цифровой модуляции

Аббревиатура	Альтернативная аббревиатура	Описание
1	2	3
Частотная манипуляция (Frequency Shift Keying)		
BFSK	FSK	Бинарная (двухпозиционная) частотная манипуляция (Binary Frequency Shift Keying)
MFSK		M-позиционная частотная манипуляция (M-ary Frequency Shift Keying)
Фазовая манипуляция (Phase Shift Keying)		
BPSK	PSK	Бинарная (двухпозиционная) фазовая манипуляция (Binary Phase Shift Keying)
QPSK	4PSK	Квадратурная фазовая манипуляция (Quadrature Phase Shift Keying)
OQPSK	SQPSK	Квадратурная фазовая манипуляция со сдвигом (Offset QPSK, Staggered QPSK)
$\pi/4$ -QPSK		$\pi/4$ -Квадратурная фазовая манипуляция ($\pi/4$ -Quadrature Phase Shift Keying)
MPSK		M-позиционная фазовая манипуляция (M-ary Phase Shift Keying)
Модуляция с непрерывной фазой (Continuous Phase Modulations – CPM)		
SHPM		Одноиндексная фазовая модуляция (Single-h (modulation index) Phase Modulation)
MHPM		Многоиндексная фазовая модуляция (Multi-h Phase Modulation)
MSK	FFSK	Минимальная манипуляция. Быстрая частотная манипуляция (Minimum Shift Keying. Fast FSK)
GMSK		Гауссовская минимальная манипуляция (Gaussian Minimum Shift Keying)
Амплитудная и амплитудно-фазовая модуляция (Amplitude and Amplitude-Phase Modulations)		
ASK		Амплитудная манипуляция (Amplitude Shift Keying)
OOK	ASK	Амплитудная манипуляция с пассивной паузой (Binary On-Off Keying)
MASK	MAM	M-позиционная амплитудная манипуляция (M-ary ASK, M-ary Amplitude Modulation)
QAM		Квадратурная амплитудная модуляция (Quadrature Amplitude Modulation)

Для систематизации существующей классификации методов модуляции на рис. 11.1 приведена схема, позволяющая определить взаимосвязь тех или иных модуляционных схем, широко используемых в современных телекоммуникационных системах.

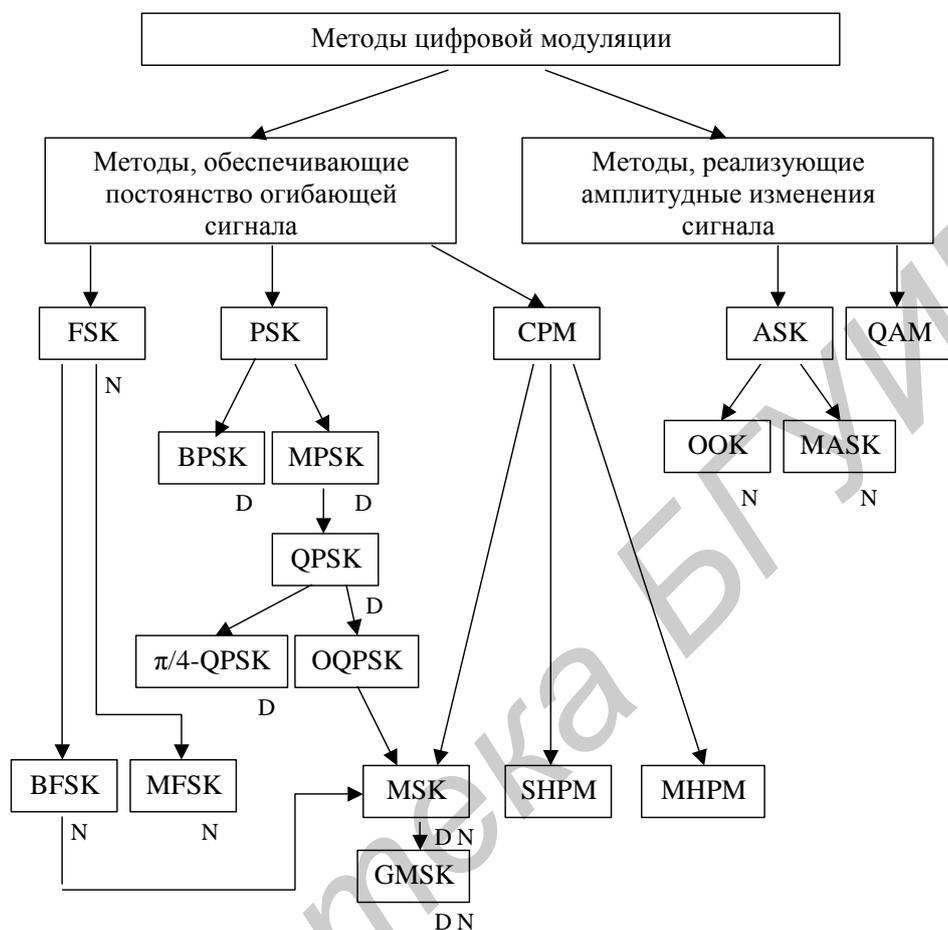


Рис. 11.1. Классификация методов цифровой модуляции

Следует отметить, что некоторые из видов модуляции, в зависимости от исходных предпосылок могут быть получены (классифицированы) больше чем из одной «родительской» ветви (схемы). Кроме того, методы, где может быть применено дифференциальное (относительное) кодирование, отдельно маркированы символом D, аналогично методы модуляции, которые могут обеспечивать некогерентную демодуляцию, маркированы символом N. Следует иметь в виду, что принципы когерентной демодуляции применимы к любой модуляционной схеме.

Все виды модуляции можно разделить на две большие категории: методы модуляции, обеспечивающие постоянство огибающей радиосигнала, и методы модуляции, приводящие к амплитудным изменениям сигнала. В каждой из указанных категорий можно выделить базовые методы модуляции. Это FSK (частотная манипуляция), PSK (фазовая манипуляция) и CPM (непрерывная фазовая модуляция) для категории модуляционных схем, обеспечивающих постоянство огибающей радиосигнала и ASK (амплитудная мани-

пуляция) и QAM для категории модуляционных схем, приводящих к амплитудным изменениям сигнала.

Остальные модуляционные схемы являются производными от базовых и получаются на основе базовых методов путем их видоизменения, комбинации и наложения ограничений на параметры и характеристики.

Охарактеризуем кратко основные методы модуляции.

BFSK – бинарная (двухпозиционная) частотная манипуляция. Сигнал представляет собой непрерывное гармоническое колебание, закон изменения частоты которого представляет прямоугольное колебание в соответствии с передаваемыми символами (один символ модуляции кодирует один бит передаваемого цифрового потока).

MFSK – M-позиционная частотная манипуляция. Сигнал представляет непрерывное гармоническое колебание, закон изменения частоты которого представляет M-уровневый прямоугольный процесс в соответствии с передаваемыми символами (один символ модуляции соответствует группе из нескольких бит передаваемого цифрового потока).

BPSK – бинарная фазовая манипуляция. Сигнал представляет непрерывное гармоническое колебание с постоянной частотой, фаза которого может принимать одно из двух заданных значений в соответствии с передаваемыми символами (один символ модуляции кодирует один бит передаваемого цифрового потока).

QPSK – квадратурная фазовая манипуляция. Сигнал представляет собой непрерывное гармоническое колебание с постоянной частотой, фаза которого может принимать одно из четырех значений в соответствии с передаваемыми символами (один символ модуляции кодирует два бита передаваемого цифрового потока).

OQPSK – квадратурная фазовая манипуляция. Разновидность QPSK, исключая одновременную смену символов в квадратурных составляющих совокупного радиосигнала.

$\pi/4$ -QPSK – разновидность QPSK, обеспечивающая сдвиг фазы между сигналами, представляющими различные символы, кратный $\pi/4$. Может рассматриваться как комбинация двух сигналов с $\pm\pi/2$ -BPSK квадратурных несущих колебаний.

M-PSK – M-позиционная фазовая манипуляция. Сигнал представляет непрерывное гармоническое колебание с постоянной частотой, фаза которого может принимать одно из M заданных значений в соответствии с передаваемыми символами (один символ модуляции кодирует несколько бит передаваемого цифрового потока).

SHPM – непрерывная фазовая модуляция с постоянным индексом модуляции. Модуляция осуществляется импульсами специальной формы без разрывов во временной области (сглаженных по тому или иному закону).

MHPM – непрерывная фазовая модуляция безразрывными импульсами специальной формы при которой индекс циклически принимает одно из M возможных значений.

MSK – минимальная манипуляция. Частотная манипуляция с минимальным сдвигом и непрерывной фазой. Под минимальным сдвигом понимается такое значение девиации частоты, при которой фазовый набег сигнала за длительность передаваемого символа составляет $\pm\pi/2$ относительно несущего колебания.

GMSK – разновидность MSK. Манипуляция осуществляется импульсами гауссовской формы.

ASK – амплитудная манипуляция. Сигнал представляет собой гармоническое колебание с постоянной частотой и фазой, закон изменения амплитуды которого определяется передаваемыми символами.

OOK – разновидность амплитудной манипуляции. Особенность данного вида модуляции заключается в пассивной паузе при передаче одного из двух возможных передаваемых символов.

MASK – M-позиционная амплитудная манипуляция. Изменение амплитуды радиосигнала осуществляется по одному из M возможных законов (состояний). Один символ модуляции кодирует несколько бит передаваемого цифрового потока.

QAM – квадратурная фазовая манипуляция. Передаваемые символы кодируются одним из нескольких возможных состояний амплитуды и начальной фазы гармонического колебания с постоянной частотой. Один символ модуляции кодирует несколько бит передаваемого цифрового потока.

11.3. Представление радиосигнала в комплексном виде.

Аналитический сигнал

11.3.1. Комплексные числа

Комплексные числа (иногда называемые мнимыми числами) не являются числами в элементарном смысле слова, применяемыми при подсчетах и измерениях. Они составляют класс математических объектов, определяемый описанными ниже свойствами.

Каждому комплексному числу \vec{c} можно поставить в соответствие единственную пару (a, b) действительных чисел a и b и обратно.

Каждое комплексное число \vec{c} может быть записано в виде суммы $\vec{c} = a + jb$ действительного числа a и чисто мнимого числа jb , мнимая единица j определяется условием $j^2 = -1$. Действительные числа $a = Re\{\vec{c}\}$ и $b = Im\{\vec{c}\}$ соответственно называются действительной частью и мнимой частью комплексного числа \vec{c} . Два комплексных числа $\vec{c} = a + jb$ и $\vec{c}^* = a - jb$, имеющие одинаковые действительные и противоположные мнимые части, называются сопряженными комплексными числами. Два комплексных числа равны в том и только том случае, если соответственно равны их действительные и мнимые части.

Любое действительное число можно представить точкой на действительной оси (это «одномерные» числа). Комплексное число тоже можно представить точкой, только на плоскости (см. рис. 11.2). Эту плоскость называют комплексной плоскостью, она дает нам возможность изображать комплексные числа, имеющие как действительную, так и мнимую части.

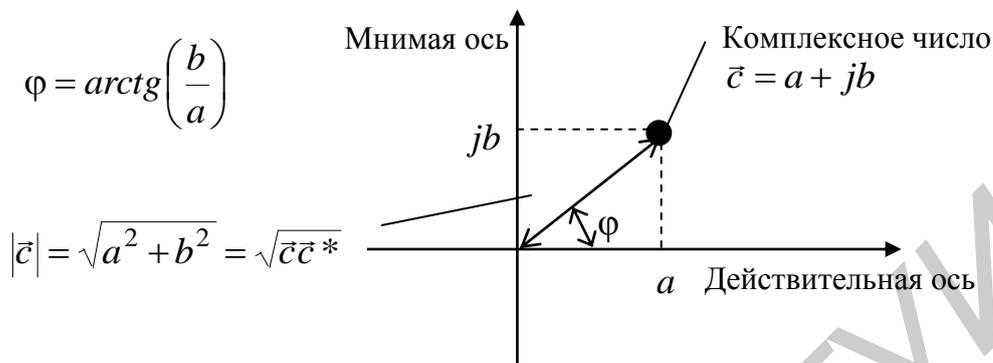


Рис. 11.2. Графическое представление комплексного числа

В соответствии с рис. 11.2 мы можем использовать геометрические свойства прямоугольного треугольника для получения нескольких различных способов представления комплексного числа \vec{c} .

Комплексное число \vec{c} может быть записано несколькими различными способами, показанными в табл. 11.2.

Таблица 11.2

Представление комплексного числа

Название	Математическое выражение
Алгебраическая (прямоугольная форма)	$\vec{c} = a + jb$
Тригонометрическая форма	$\vec{c} = \vec{c} [\cos(\varphi) + j \sin(\varphi)]$
Экспоненциальная (полярная) форма	$\vec{c} = \vec{c} e^{j\varphi}$
Модуль-аргумент	$\vec{c} = \vec{c} \angle \varphi$

Экспоненциальная форма записи непосредственно вытекает из тригонометрической на основании разложения в ряд функций $\cos(x)$ и $\sin(x)$:

$$\begin{aligned}
\cos(x) &= 1 + \sum_{n=1}^{\infty} \frac{x^{2n}}{(2n)!} (-1)^n ; \\
\sin(x) &= \sum_{n=1}^{\infty} \frac{x^{2n-1}}{(2n-1)!} (-1)^{n+1} ; \\
e^{jx} &= 1 + \sum_{n=1}^{\infty} \frac{(jx)^n}{n!} = 1 + \sum_{n=1}^{\infty} \frac{x^{2n}}{(2n)!} (-1)^n + j \sum_{n=1}^{\infty} \frac{x^{2n-1}}{(2n-1)!} (-1)^{n+1} = \\
&= \cos(x) + j \sin(x) \quad (\text{формула Эйлера})
\end{aligned} \tag{11.3}$$

Математические операции над комплексными числами подчиняются тем же правилам, что и операции над действительными числами.

$$\begin{aligned}
\vec{c}_1 &= a_1 + jb_1; \quad \vec{c}_2 = a_2 + jb_2; \\
\vec{c}_1 \pm \vec{c}_2 &= a_1 \pm a_2 + j(b_1 \pm b_2); \\
\vec{c}_1 \vec{c}_2 &= (a_1 + jb_1)(a_2 + jb_2) = a_1 a_2 + j^2 b_1 b_2 + j(a_1 b_2 + a_2 b_1) = \\
&= a_1 a_2 - b_1 b_2 + j(a_1 b_2 + a_2 b_1) = |\vec{c}_1| |\vec{c}_2| e^{j(\varphi_1 + \varphi_2)}; \\
\frac{\vec{c}_1}{\vec{c}_2} &= \frac{\vec{c}_1 \vec{c}_2^*}{\vec{c}_2 \vec{c}_2^*} = \frac{\vec{c}_1 \vec{c}_2^*}{|\vec{c}_2|^2} = \frac{|\vec{c}_1| e^{j\varphi_1} |\vec{c}_2| e^{-j\varphi_2}}{|\vec{c}_2|^2} = \frac{|\vec{c}_1|}{|\vec{c}_2|} e^{j(\varphi_1 - \varphi_2)}.
\end{aligned}$$

11.3.2. Комплексное представление радиосигнала

Рассмотрим радиосигнал как комбинацию синфазной и квадратурной компонент:

$$s(t) = S_I(t) \cos[2\pi f_0 t] - S_Q(t) \sin[2\pi f_0 t], \tag{11.4}$$

где $S_I(t) = A(t) \cos[\psi(t)]$ и $S_Q(t) = A(t) \sin[\psi(t)]$

– синфазная и квадратурная компонента модулированного радиосигнала, соответственно.

Определим некоторую функцию $\vec{S}(t)$ через синфазную и квадратурную составляющую следующим образом:

$$\vec{S}(t) = S_I(t) + jS_Q(t).$$

Данная функция несет всю информацию об изменениях амплитуды, частоты и фазы модулированного радиосигнала.

Умножим определенную выше функцию на комплексную экспоненту $e^{j\omega_0 t}$:

$$\begin{aligned}\vec{S}(t)e^{j\omega_0 t} &= (S_I(t) + jS_Q(t))e^{j\omega_0 t} = (S_I(t) + jS_Q(t))(\cos(\omega_0 t) + j\sin(\omega_0 t)) = \\ &= [S_I(t)\cos(\omega_0 t) - S_Q(t)\sin(\omega_0 t)] + j[S_I(t)\sin(\omega_0 t) + S_Q(t)\cos(\omega_0 t)].\end{aligned}$$

Действительная часть функции $\vec{S}(t)e^{j\omega_0 t}$ полностью определяет модулированный радиосигнал

$$\operatorname{Re}\{\vec{S}(t)e^{j\omega_0 t}\} = S_I(t)\cos(\omega_0 t) - S_Q(t)\sin(\omega_0 t),$$

мнимая часть $\operatorname{Im}\{\vec{S}(t)e^{j\omega_0 t}\} = S_I(t)\sin(\omega_0 t) + S_Q(t)\cos(\omega_0 t)$ является сопряженным по Гильберту отображением модулированного радиосигнала.

Преобразование Гильберта – это процедура, выполняемая над действительным сигналом $s(t)$ и дающая новый действительный сигнал $s_{ht}(t)$ (сопряженный по Гильберту). При этом $s_{ht}(t)$ представляет собой сдвинутую по фазе на 90° версию сигнала $s(t)$.

Отличие нового сигнала $s_{ht}(t)$ от исходного наиболее заметно при анализе их Фурье-образов (преобразований Фурье). А именно: все компоненты $s_{ht}(t)$ с положительными частотами равны компонентам $s(t)$ с положительными частотами, сдвинутым по фазе на минус 90° , а все компоненты $s_{ht}(t)$ с отрицательными частотами равны компонентам $s(t)$ с отрицательными частотами, сдвинутым по фазе на $+90^\circ$.

Комплексную функцию времени $s_a(t) = \vec{S}(t)e^{j\omega_0 t}$ принято называть аналитическим сигналом, а функцию $\vec{S}(t)$ – комплексной огибающей радиосигнала.

Спектральные особенности сопряженных по Гильберту сигналов обуславливают отсутствие компонент с отрицательными частотами в спектре аналитического сигнала. Действительно, для области отрицательных частот выполняется соотношение $S_{ht}(j\omega) = jS(j\omega)$, а для области положительных частот $S_{ht}(j\omega) = -jS(j\omega)$, где $S_{ht}(j\omega)$ и $S(j\omega)$ – комплексные спектральные функции (спектры) сопряженного по Гильберту сигнала $s_{ht}(t)$ и исходного сигнала $s(t)$. Тогда спектр аналитического сигнала может быть записан следующим образом:

$$\begin{aligned}
S_a(j\omega) &= \int_{-\infty}^{\infty} s_a(t) e^{-j\omega t} dt = \int_{-\infty}^{\infty} [s(t) + js_{ht}(t)] e^{-j\omega t} dt = \\
&= S(j\omega) + jS_{ht}(j\omega) = \begin{cases} S(j\omega) + j(-jS(j\omega)); & \omega > 0 \\ S(j\omega) + j(jS(j\omega)); & \omega < 0 \end{cases} = \\
&= \begin{cases} S(j\omega) + S(j\omega); & \omega > 0 \\ S(j\omega) - S(j\omega); & \omega < 0 \end{cases} = \begin{cases} 2S(j\omega); & \omega > 0; \\ 0; & \omega < 0. \end{cases}
\end{aligned}$$

Таким образом, модулированный радиосигнал может быть представлен комплексной функцией времени (аналитическим сигналом), сохраняющей все особенности исходного сигнала. Поскольку множитель $e^{j\omega_0 t}$, входящий в экспоненциальную форму записи аналитического сигнала, несет информацию только о центральной частоте исходного радиосигнала (не включает в себя законы информационной модуляции и особенности формы несущего колебания), то при решении широкого круга задач формирования и обработки модулированных радиосигналов достаточно рассмотрения только комплексной огибающей $\vec{S}(t)$, включающей в себя все особенности реального сигнала.

11.3.3. Замечания по поводу отрицательных частот

Понятие отрицательной частоты служит состоятельным и мощным инструментом при анализе и синтезе сигналов. Использование отрицательной частоты становится обязательным при представлении действительных сигналов в комплексной записи.

Трудности понимания идеи отрицательной частоты сродни тому оценению, которое ощущалось в среде средневековых математиков, когда они впервые столкнулись с отрицательными числами. Вплоть до тринадцатого века отрицательные числа рассматривались как фиктивные, потому что числа обычно использовались для счета и измерения. Соответственно, отрицательные числа просто не имели смысла (как может существовать нечто, которое меньше, чем ничто?). Идея вычитания шести из четырех должна была казаться бессмысленной. Историки математики предполагают, что отрицательные числа впервые были введены в оборот в Италии. Как свидетельствует история, около 1200 г. итальянский математик Фибоначчи работал над некой финансовой проблемой, единственное допустимое решение которой включало отрицательные числа. Отважный Лео писал: «В отношении этой проблемы я показал, что она неразрешима, если только не допустить, что первый человек имел долг».

Современное понимание отрицательных чисел связано с направлением. назад от нуля в том смысле, что положительные числа показывают направление вперед от нуля. Таким образом, понятие отрицательного числа вполне

состоятельно, если мы определяем его соответствующим образом. Как бы ни были для нас привычны отрицательные числа, отрицательная частота остается трудным и сомнительным понятием для многих.

Отрицательная частота представляет собой вполне состоятельное понятие, если мы правильно определяем ее по отношению к тому, что мы называем положительной частотой.

Как нам уже известно, гармоническое колебание с круговой частотой ω_0 в комплексном виде может быть представлено функцией следующего вида $e^{j\omega_0 t}$. Геометрическая интерпретация данной функции представляет собой вращающуюся по единичной окружности в направлении против часовой стрелки точку на комплексной плоскости (см. рис. 11.3). Проекция данной точки на действительную ось является функцией сигнала $\cos(\omega_0 t)$. Рассмотрим поведение комплексной экспоненты в случае отрицательного значения круговой частоты. В этом случае направление вращения точки, представляющей комплексную экспоненту, изменится на противоположное. При этом проекция точки на действительную ось не изменится и останется прежней ($\cos(|\omega_0|t)$), изменению подвергнется только проекция точки на мнимую ось ($\sin(\omega_0 t)$, для $\omega_0 > 0$ сменится на $-\sin(|\omega_0|t)$, для $\omega_0 < 0$).

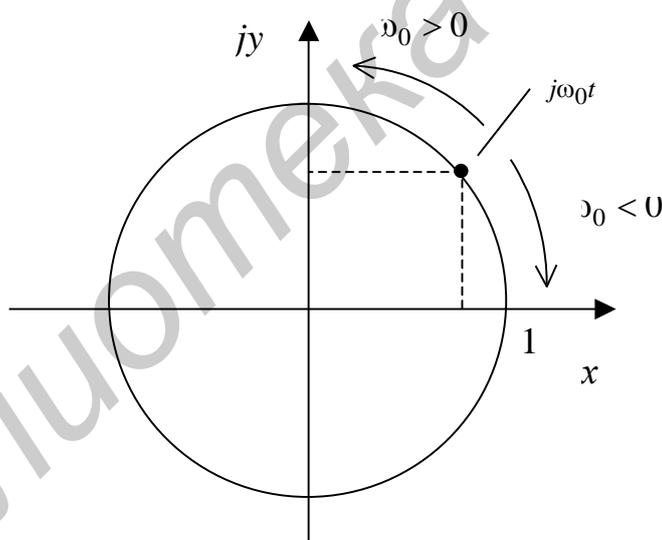


Рис. 11.3. Интерпретация отрицательной частоты

Поскольку физически реализуемым сигналом является проекция комплексной экспоненты на действительную ось, то смысл в отрицательной частоте отсутствует (так же, как отсутствовал смысл отрицательных чисел для средневековых математиков). Но это применимо только для нулевой точки отсчета. Рассмотрим линейную комбинацию колебаний $e^{j\omega_0 t} + e^{-j\omega_0 t}$. Данное колебание есть не что иное, как $2\cos(\omega_0 t)$. Осуществим линейное изменение системы координат в виде сдвига точки отсчета частоты на значение Ω , в результате этого получим

$$(e^{j\omega_0 t} + e^{-j\omega_0 t}) e^{j\Omega t} = e^{j(\omega_0 + \Omega)t} + e^{j(-\omega_0 + \Omega)t} \quad (11.5)$$

Действительная часть полученной комплексной функции соответствует физическому колебанию, подчиняющемуся следующему закону:

$$\begin{aligned} \operatorname{Re}\left\{(e^{j\omega_0 t} + e^{-j\omega_0 t}) e^{j\Omega t}\right\} &= \operatorname{Re}\left\{e^{j(\omega_0 + \Omega)t} + e^{j(-\omega_0 + \Omega)t}\right\} = \\ &= \operatorname{Re}\left\{\cos[(\omega_0 + \Omega)t] + j\sin[(\omega_0 + \Omega)t]\right\} + \\ &+ \operatorname{Re}\left\{\cos[(-\omega_0 + \Omega)t] + j\sin[(-\omega_0 + \Omega)t]\right\} = \\ &= \cos[(\omega_0 + \Omega)t] + \cos[|-\omega_0 + \Omega|t]. \end{aligned} \quad (11.6)$$

Как можно заметить, «абсолютно» эквивалентные в нулевой точке отсчета колебания $e^{j\omega_0 t}$ и $e^{-j\omega_0 t}$ совершенно не эквивалентны и дают совершенно разный результат, если пронаблюдать их «со стороны».

Таким образом, отрицательное значение частоты можно определить как направление вращения вектора сигнала.

Понятие отрицательной частоты является крайне важным в теории сигналов, поскольку его использование позволяет осуществлять синтез сигналов с повышенной спектральной эффективностью, а также является основополагающим при так называемой квадратурной обработке сигналов.

Мы определили отрицательную частоту в частотной области. Если коллеги хотят заявить, что отрицательная частота не существует во временной области, мы не будем спорить. Однако наше определение отрицательной частоты в частотной области является однозначным, совместимым с действительными сигналами и очень полезным, и мы будем его придерживаться.

11.3.4. Заключительные замечания

Может возникнуть вопрос: зачем необходимо комплексное (более «сложное») представление сигналов? Чтобы ответить на данный вопрос, еще раз обратим внимание на некоторые особенности, присущие комплексному представлению модулированных радиосигналов.

1. Составной частью комплексного описания (аналитического сигнала) является комплексная экспонента $e^{j\omega_0 t}$, которая несет информацию только о центральной частоте модулированного радиосигнала.

2. Вся информация о законах модуляции содержится непосредственно в комплексной огибающей аналитического сигнала $\bar{S}(t)$.

3. Поскольку спектр (образ Фурье) аналитического сигнала $s_a(t)$ не содержит компонент с отрицательными частотами и повторяет физический

спектр реального радиосигнала $s(t)$, то спектр комплексной огибающей $\bar{S}(t)$ полностью копирует спектр модулированного радиосигнала $s(t)$ при условии перемещения его центральной частоты в нуль.

4. Спектр комплексной огибающей $\bar{S}(t)$ содержит как компоненты с отрицательными, так и компоненты с положительными частотами. При этом отрицательная область частот необязательно является симметричной копией положительной области частот относительно нуля (область отрицательных частот становится информативной, а не просто «формальной»).

Указанные особенности аналитического сигнала и комплексной огибающей позволяют:

- осуществлять непосредственный синтез спектра итогового радиосигнала с последующим формированием комплексной огибающей методами преобразования Фурье и переносом ее на рабочую частоту квадратурным преобразователем частоты, что позволяет исключать информационное дублирование прилегающих снизу и сверху к центральной частоте полос частот модулированного радиосигнала;

- осуществлять основную обработку радиосигналов (включая демодуляцию) в области нулевой частоты путем получения комплексной огибающей методами квадратурного преобразования частоты;

- широко использовать цифровые методы, реализуемые специализированными процессорами, в задачах синтеза, анализа и обработки радиосигналов;

- повысить «информационную» эффективность дискретных преобразований Фурье (количество информативных спектральных отсчетов увеличивается в два раза).

Таким образом, комплексное описание радиосигналов позволяет существенно упростить формирование и обработку сигналов с использованием цифровых методов, разработать унифицированные алгоритмы модуляции и демодуляции сигналов, осуществлять более «гибкое» управление спектральными, корреляционными и другими характеристиками формируемого сигнала.

11.4. Прямой и обратный переход между радиосигналом и его комплексным представлением

Рассмотрим радиосигнал как комбинацию синфазной и квадратурной компонент (11.4).

Комплексное представление такого сигнала (аналитический сигнал) определяется следующим образом:

$$\bar{s}(t) = (S_I(t) + jS_Q(t))e^{j\omega_0 t} = s(t) + js_{ht}(t). \quad (11.7)$$

где $s_{ht}(t)$ – рассмотренный ранее сигнал, сопряженный по Гильберту с исходным сигналом $s(t)$.

Таким образом, для получения комплексного представления радиосигнала достаточно иметь преобразователь Гильберта, формирующий мнимую составляющую аналитического сигнала. По сути, преобразователь Гильберта является широкополосным фазовращателем, обеспечивающим сдвиг каждой частотной компоненты исходного сигнала на 90° . Физическая реализация подобного фазовращателя затруднена, поэтому для получения действительной и мнимой составляющих аналитического сигнала, сдвинутых в частотной области относительно исходного сигнала, используют квадратурные преобразователи частоты. При этом сдвиг в частотной области обычно осуществляется в область нулевой частоты, что существенно облегчает дальнейшие операции над полученным комплексным представлением сигнала. Структурная схема квадратурного преобразователя частоты представлена на рис. 11.4. Возможная ошибка установки частоты квадратурных опорных колебаний учтена фазовым набегом $\varphi(t)$.

В результате операции квадратурного преобразования частоты имеем

$$\begin{aligned} x(t) &= \{S_I(t)\cos[\omega_0 t] - S_Q(t)\sin[\omega_0 t]\} \cos(\omega_0 t + \varphi(t)) = \\ &= S_I(t)\cos[\varphi(t)] + S_Q(t)\sin[\varphi(t)]; \end{aligned} \quad (11.8)$$

$$\begin{aligned} y(t) &= \{S_I(t)\cos[\omega_0 t] - S_Q(t)\sin[\omega_0 t]\} [-\sin(\omega_0 t + \varphi(t))] = \\ &= -S_I(t)\sin[\varphi(t)] + S_Q(t)\cos[\varphi(t)]. \end{aligned} \quad (11.9)$$

Для сравнения рассмотрим операцию сдвига в область нулевых частот аналитического сигнала:

$$\begin{aligned} (S_I(t) + jS_Q(t))e^{j\omega_0 t} e^{-j(\omega_0 t + \varphi(t))} &= (S_I(t) + jS_Q(t))e^{-j\varphi(t)} = \\ &= (S_I(t) + jS_Q(t))(\cos(\varphi(t)) - j\sin(\varphi(t))) = \\ &= S_I(t)\cos(\varphi(t)) + S_Q(t)\sin(\varphi(t)) + j[-S_I(t)\sin(\varphi(t)) + S_Q(t)\cos(\varphi(t))]. \end{aligned} \quad (11.10)$$

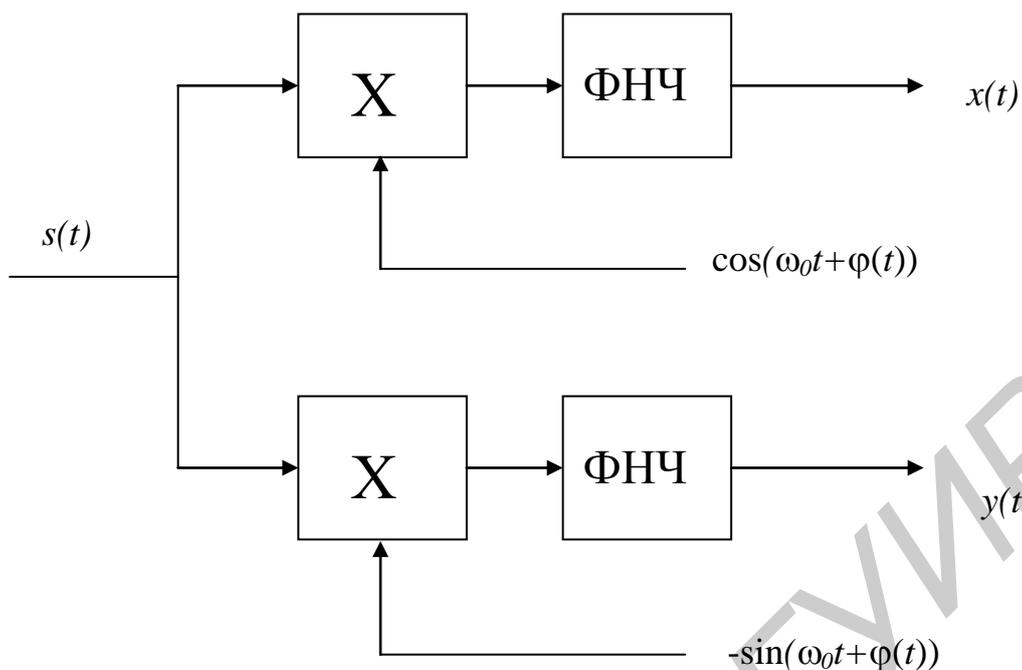


Рис. 11.4. Квадратурный преобразователь частоты

Сопоставляя полученные выражения, можно сделать вывод, что квадратурный преобразователь формирует на своем выходе действительную и мнимую компоненты комплексной огибающей исходного аналитического сигнала (его комплексного представления):

$$(S_I(t) + jS_Q(t))e^{-j\varphi(t)} = x(t) + jy(t). \quad (11.11)$$

При этом наличие фазовой ошибки при преобразовании частоты приводит к фазовому запаздыванию полученного комплексного представления исходного радиосигнала на величину данной ошибки. Данное обстоятельство, как правило, требует принятия дополнительных мер по синхронизации (компенсации ошибки) опорных колебаний квадратурного преобразователя частоты с несущим колебанием радиосигнала с точностью до фазы $\varphi(t) = 0$. Вопросы фазовой синхронизации при демодуляции будут рассмотрены позже. Следует заметить, что при осуществлении квадратурного преобразования частоты обязательным условием является опережение опорным колебанием в канале, формирующем действительную компоненту аналитического сигнала, опорного колебания в канале, формирующем мнимую компоненту аналитического сигнала, на фазовый угол, равный $\pi/2$.

Обратное преобразование аналитического сигнала осуществляется квадратурным модулятором (рис. 11.5), реализующим перенос комплексной огибающей из области нулевой частоты в область рабочей частоты ω_0 .

Принцип формирования радиосигнала с заданной комплексной огибающей поясняется следующим соотношением:

$$\operatorname{Re}\{(S_I(t) + jS_Q(t))e^{j\omega_0 t}\} = S_I(t)\cos[2\pi f_0 t] - S_Q(t)\sin[2\pi f_0 t] = s(t). \quad (11.12)$$

По аналогии с квадратурным преобразованием частоты принципиальным моментом при формировании радиосигнала является опережение опорным колебанием в синфазном канале (I) опорного колебания в квадратурном канале (Q) на фазовый угол, равный $\pi/2$.

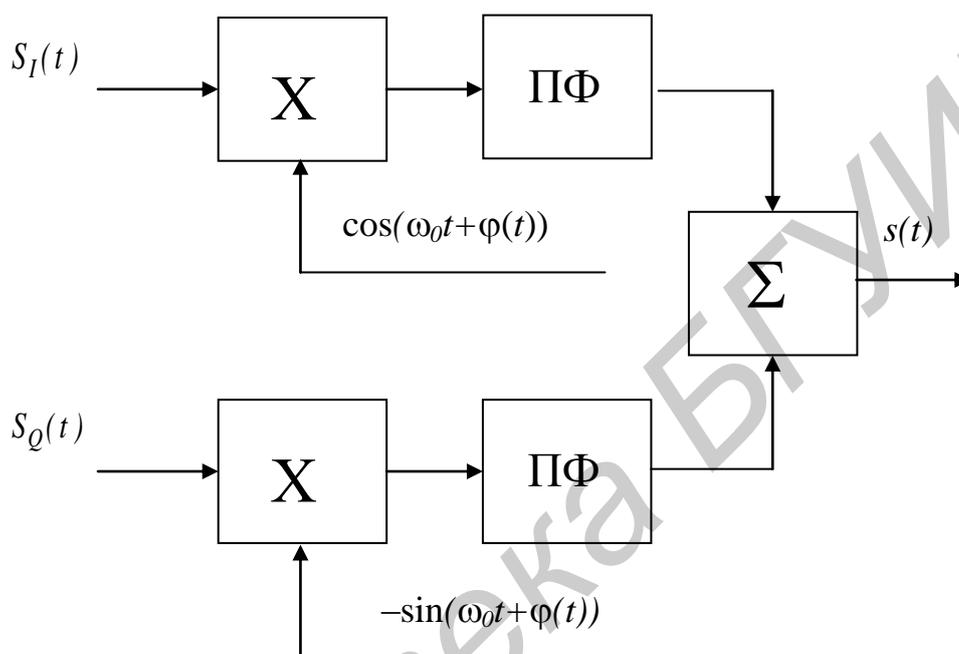


Рис. 11.5. Квадратурный модулятор

11.5. Квазиоптимальное принятие решений при приеме сигналов с дискретной модуляцией

11.5.1. Общая методология синтеза решающих правил при квазиоптимальном приеме сигналов с дискретной модуляцией

Предположим, что сигнал зависит от нескольких параметров, часть из которых подлежит определению, т.е. одни являются существенными, а остальные несущественными. К первым относятся параметры $\Lambda = [\lambda_1, \dots, \lambda_k]$, возможные значения которых известны и образуют область L . Ко вторым относятся параметры $\Phi = [\varphi_1, \dots, \varphi_l]$, возможные значения которых образуют область F . Сигнал может быть записан в виде $s(t, \Lambda, \Phi)$.

Сигнал $s(t)$ комбинируется с шумом $n(t)$, образуя случайный процесс $x(t)$. Процесс $x(t)$ наблюдается в моменты времени t_1, \dots, t_n . Выборочные значения $X = [x_1, \dots, x_n]$ образуют пространство входов решающего устройства Γ .

Предполагаются известными статистика шума $n(t)$ и статистика параметров Λ и Φ . По наблюдаемой выборке $X = [x_1, \dots, x_n]$ необходимо решить, какие значения имеют параметры $\Lambda = [\lambda_1, \dots, \lambda_k]$.

Решающее устройство формирует на выходе решения $D = [d_1, \dots, d_k]$, образующие область возможных решений G .

Будем считать, что когда принимается решение D , а параметры сигнала имеют значение Λ , решающее устройство терпит ущерб, величина которого описывается заданной функцией потери $W(D, \Lambda)$.

Введем решающую функцию (решающее правило) $\Delta(D|X)$ как условную плотность распределения вероятности случайной величины D при условии, что наблюдаются выборочные данные X . Обычно используется вариант неслучайной решающей функции следующего вида

$$\Delta(D|X) = \delta[D - \Lambda^*(X)] = \delta[d_1 - \lambda_1^*(X)] \delta[d_2 - \lambda_2^*(X)] \dots \delta[d_k - \lambda_k^*(X)];$$

$$\Lambda^* = [\lambda_1^*, \dots, \lambda_k^*].$$

Компоненты $\lambda_1^*(x_1, \dots, x_n), \dots, \lambda_k^*(x_1, \dots, x_n)$ называют обычно оценками параметров $\lambda_1, \dots, \lambda_k$.

Мерой качества оценки Λ является так называемый средний риск, который определяется следующим образом.

Образуется условная плотность вероятности случайной величины D при условии, что параметры сигнала имеют значение Λ :

$$P(D|\Lambda) = \int_{\Gamma} \Delta(D|X) P(X|\Lambda) dX, \quad (11.13)$$

где $P(X|\Lambda)$ – функция правдоподобия вектора параметров Λ (многомерная плотность распределения вероятности выборочных значений x_1, \dots, x_n наблюдаемого процесса $x(t)$ при условии конкретных значений $\lambda_1, \dots, \lambda_k$ вектора параметров Λ).

Совместная плотность распределения вероятности векторов D и Λ записывается следующим образом:

$$P(D, \Lambda) = P(D|\Lambda) P(\Lambda), \quad (11.14)$$

где $P(\Lambda) = P_{ap}(\Lambda)$ – априорная (известная статистика параметров Λ) плотность распределения вектора параметров Λ .

Средний риск, по определению, равен средним потерям, присущим выбранному решающему правилу, т.е.

$$\begin{aligned}
R &= \int \int_{GL} dD d\Lambda [W(D, \Lambda) P(D, \Lambda)] = \\
&= \int \int \int_{\Gamma GL} dX dD d\Lambda [W(D, \Lambda) \Delta(D|X) P(X|\Lambda) P_{ap}(\Lambda)]. \quad (11.15)
\end{aligned}$$

Задавая различные функции потерь, можно получать различные правила построения оценок параметров сигнала. Однако, какой бы вид не имели эти правила, все они будут основаны на операциях с апостериорной вероятностью $P_{ac}(\Lambda) = kP(X|\Lambda)P_{ap}(\Lambda)$ вектора параметров Λ . При этом функция правдоподобия существенных параметров $P(X|\Lambda)$ получается усреднением по несущественным параметрам

$$P(X|\Lambda) = \int_F P(X|\Lambda, \Phi) P_{ap}(\Phi) d\Phi, \quad (11.16)$$

где $P(X|\Lambda, \Phi)$ – плотность распределения вероятности X при условии заданных конкретных значений Λ и Φ ; $P_{ap}(\Phi)$ – априорная плотность распределения вероятности вектора Φ .

Применительно к решению задачи квазиоптимальной демодуляции сигналов с дискретной модуляцией наиболее подходящей является простая функция потерь, имеющая следующий вид:

$$W(D, \Lambda) = 1 - \delta(D - \Lambda) = 1 - \delta(d_1 - \lambda_1) \delta(d_2 - \lambda_2) \dots \delta(d_k - \lambda_k). \quad (11.17)$$

Минимизируем средний риск для простой функции потерь:

$$\begin{aligned}
R &= 1 - \int_{\Gamma} dX \int_G dD [\Delta(D|X) P(X|D) P_{ap}(D)] = \\
&= 1 - \int_{\Gamma} dX \int_G dD [\delta[D - \Lambda^*(X)] P(X|D) P_{ap}(D)] = \\
&= 1 - \int_{\Gamma} dX [P(X|\Lambda^*) P_{ap}(\Lambda^*)]. \quad (11.18)
\end{aligned}$$

Минимум среднего риска соответствует максимуму $P(X|\Lambda^*) P_{ap}(\Lambda^*)$, который достигается, если в качестве вектора оценок Λ^* (в общем случае $\Lambda^*(X)$) выбирать то значение, которое при любом значении X максимизирует апостериорную вероятность $P_{ac}(\Lambda)$.

В соответствии с вышеизложенным средний риск для простой функции потерь минимизируется следующим правилом построения оценки параметров Λ : в качестве оценок $\lambda_1^*(x_1, \dots, x_n), \dots, \lambda_k^*(x_1, \dots, x_n)$ выбирают значения,

обращающие в максимум апостериорную вероятность $P_{ac}(\lambda_1, \dots, \lambda_k) = kP(X|\lambda_1, \dots, \lambda_k)P_{ap}(\lambda_1, \dots, \lambda_k)$.

В случае, если априорная вероятность вектора параметров постоянна на интервале существенного изменения функции правдоподобия, максимум апостериорной вероятности соответствует максимуму функции правдоподобия $P(X|\lambda_1, \dots, \lambda_k)$.

Оценки по максимуму правдоподобия находятся как решение системы уравнений правдоподобия:

$$\frac{\partial}{\partial \lambda_1} \ln(P(X|\lambda_1, \dots, \lambda_k)) = 0,$$

$$\frac{\partial}{\partial \lambda_k} \ln(P(X|\lambda_1, \dots, \lambda_k)) = 0.$$

В выражении для среднего риска

$$R = 1 - \int_{\Gamma} dX \int_G dD [\Delta(D|X)P(X|D)P_{ap}(D)] \quad (11.19)$$

интеграл в правой части представляет среднюю по априорной вероятности вероятность правильного решения. Минимальный средний риск при этом равен минимально возможной средней вероятности неправильного решения (вероятности ошибки).

Необходимо отметить, что полученные правила построения оценок существенных параметров сигнала могут быть записаны и для случая непрерывного наблюдения путем осуществления соответствующих предельных переходов.

11.5.2. Синтез функции правдоподобия для канала связи с аддитивным белым гауссовским шумом

11.5.2.1. Белый гауссовский шум

Рассмотрим стационарный случайный процесс $n(t)$, функция автокорреляции которого равна дельта-функции, умноженной на некоторую постоянную величину $N_0/2$:

$$R_n(\tau) = \frac{N_0}{2} \delta(\tau). \quad (11.20)$$

Как известно, дельта-функция равна нулю всюду, за исключением точки $\tau = 0$, где $\delta(0) = \infty$, причем интеграл от дельта-функции по любому интервалу, содержащему особую точку $\tau = 0$, равен единице.

Отсюда следует, что рассматриваемый случайный процесс характеризуется тем, что значения $n(t)$ в любые два сколь угодно близкие моменты времени некоррелированы. Поэтому такой процесс $n(t)$ можно назвать абсолютно случайным процессом.

Спектральная плотность (распределение мощности по полосе частот) абсолютно случайного процесса определяется на основании теоремы Винера–Хинчина следующим образом:

$$\begin{aligned} S_n(\omega) &= \int_{-\infty}^{\infty} R_n(\tau) e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} \frac{N_0}{2} \delta(\tau) e^{-j\omega\tau} d\tau = \frac{N_0}{2} \int_{-\infty}^{\infty} \delta(\tau) e^{-j\omega\tau} d\tau = \\ &= \frac{N_0}{2} e^{-j\omega \cdot 0} = \frac{N_0}{2}. \end{aligned} \quad (11.21)$$

Таким образом, спектральная плотность абсолютно случайного процесса постоянна при всех частотах.

Случайный процесс $n(t)$, обладающий равномерным спектром в очень широком диапазоне частот, является «белым шумом» по аналогии с белым светом, имеющим в видимой части равномерный сплошной спектр.

Для белого шума формула спектральной плотности дает физически непонятный результат: дисперсия (средняя мощность) такого шума

$$\sigma_n^2 = \int_{-\infty}^{\infty} S_n(\omega) d\omega = \infty. \quad \text{Этот результат объясняется тем, что белый шум следует}$$

рассматривать как идеализацию, так как реальные процессы всегда имеют энергетический спектр, убывающий с частотой, и, следовательно, ограниченной средней мощностью. Белый шум является полезной математической идеализацией, применимой в тех случаях, когда время корреляции шума много меньше всех существенных постоянных времени системы, на которую воздействует шум, или, иначе, когда в пределах амплитудно-частотной характеристики системы спектральную плотность воздействующего реального шума можно приближенно считать постоянной.

Укажем, что если белый шум $n(t)$ представляет случайное напряжение, то величина N_0 имеет размерность $[N_0] = [\text{В}^2/\text{Гц}] = [\text{В}^2 \cdot \text{с}]$.

Определим многомерную плотность распределения вероятности белого гауссовского шума на интервале времени $[0, T]$. Для этого определим выборочные значения белого шума в момент времени t_i следующим образом:

$$n_i = \frac{1}{\Delta t} \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} n(t) dt, \quad (11.22)$$

где Δt – интервал дискретизации во времени. При этом реализация непрерывного случайного процесса на интервале времени $[0, T]$ представляется

набором выборочных значений n_0, \dots, n_M определенных для моментов времени $t_0 = 0, \dots, t_k = k\Delta t, \dots, t_M = M\Delta t = T$. Определим элементы автокорреляционной матрицы вектора выборочных значений n_0, \dots, n_M :

$$\begin{aligned} \langle n_i n_j \rangle &= \left\langle \frac{1}{\Delta t} \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} n(x) dx \frac{1}{\Delta t} \int_{t_j - \Delta t/2}^{t_j + \Delta t/2} n(y) dy \right\rangle = \\ &= \left\langle \frac{1}{(\Delta t)^2} \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} \int_{t_j - \Delta t/2}^{t_j + \Delta t/2} n(x) n(y) dx dy \right\rangle. \end{aligned} \quad (11.23)$$

где $\langle \cdot \rangle$ – операция усреднения по ансамблю реализаций. Ввиду линейности операций усреднения и интегрирования по времени очередность данных операций может быть изменена, тогда

$$\langle n_i n_j \rangle = \frac{1}{(\Delta t)^2} \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} \int_{t_j - \Delta t/2}^{t_j + \Delta t/2} \langle n(x) n(y) \rangle dx dy. \quad (11.24)$$

Заметим, что $\langle n(x) n(y) \rangle$ по определению есть автокорреляционная функция случайного процесса $n(t)$:

$$\langle n(x) n(y) \rangle = R_n(x - y) = \frac{N_0}{2} \delta(x - y).$$

С учетом этого перепишем выражение для элементов автокорреляционной матрицы выборочных значений n_0, \dots, n_M :

$$\begin{aligned} \langle n_i n_j \rangle &= \frac{1}{(\Delta t)^2} \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} \int_{t_j - \Delta t/2}^{t_j + \Delta t/2} \frac{N_0}{2} \delta(x - y) dx dy = \\ &= \begin{cases} \frac{1}{(\Delta t)^2} \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} \frac{N_0}{2} dx; & i = j \\ 0; & i \neq j \end{cases} = \begin{cases} \frac{1}{(\Delta t)^2} \frac{N_0}{2} \Delta t; & i = j \\ 0; & i \neq j \end{cases} = \begin{cases} \frac{N_0}{2\Delta t}; & i = j; \\ 0; & i \neq j. \end{cases} \end{aligned} \quad (11.25)$$

Автокорреляционная матрица выборочных значений n_0, \dots, n_M содержит ненулевые элементы, расположенные только на главной диагонали. Это позволяет сделать заключение о взаимной некоррелированности выборочных значений n_0, \dots, n_M .

Пусть мгновенные значения исходного случайного процесса $n(t)$ подчиняются гауссовскому (нормальному) закону распределения вероятности,

тогда ввиду линейности операции дискретизации по времени выборочные значения n_0, \dots, n_M также будут распределены в соответствии с нормальной плотностью распределения вероятности.

$$W(n_i) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{-\frac{(n_i)^2}{2\sigma_n^2}\right\},$$

где $\sigma_n^2 = \langle n_i^2 \rangle = \frac{N_0}{2\Delta t}$ – диагональный элемент автокорреляционной матрицы выборочных значений шума.

В случае гауссовских случайных величин некоррелированность автоматически предполагает независимость, на основании этого многомерная плотность распределения вероятности совокупности выборочных значений записывается следующим образом:

$$W(n_0, \dots, n_M) = \prod_{i=0}^M W(n_i) = \left(\frac{1}{\sqrt{2\pi \frac{N_0}{2\Delta t}}} \right)^{M+1} \exp\left\{-\frac{1}{N_0} \sum_{i=0}^M n_i^2 \Delta t\right\}. \quad (11.26)$$

Осуществив предельный переход при $\Delta t \rightarrow 0$ и $M \rightarrow \infty$, получим выражение для функционала распределения вероятности белого гауссовского шума на интервале времени $[0, T]$:

$$F(n(t)) = h_0 \exp\left\{-\frac{1}{N_0} \int_0^T [n(t)]^2 dt\right\}, \quad (11.27)$$

где h_0 – неопределенная (бесконечно малая) константа, не зависящая от статистических свойств случайного процесса.

11.5.2.2. Комплексное представление полосовых случайных процессов

Комплексное представление полосовых сигналов в подразд. 11.3 касается детерминированных сигналов. В этом подразделе рассмотрим комплексное представление полосовых стационарных случайных процессов.

Предположим, что $n(t)$ является реализацией стационарного в широком смысле случайного процесса $N(t)$ с нулевым средним и спектральной плотностью мощности $S_n(f)$. Примем, что спектральная плотность мощности равна нулю вне интервала частот, группирующихся около частот $\pm f_0$,

где f_0 – центральная частота. Случайный процесс $N(t)$ называется узкополосным случайным процессом, если ширина его полосы частот Δf намного меньше f_0 . С учетом этого условия реализация процесса $n(t)$ может быть представлена в одной из трех форм, а именно:

$$n(t) = a(t)\cos[2\pi f_0 t + \Theta(t)]; \quad (11.28)$$

$$n(t) = n_I(t)\cos[2\pi f_0 t] - n_Q(t)\sin[2\pi f_0 t]; \quad (11.29)$$

$$n(t) = \operatorname{Re}\{\bar{n}_z(t)e^{j2\pi f_0 t}\}; \quad (11.30)$$

$$\bar{n}_z(t) = n_I(t) + jn_Q(t), \quad (11.31)$$

где $a(t)$ – огибающая, а $\Theta(t)$ – фаза вещественного сигнала, $n_I(t)$, $n_Q(t)$ – квадратурные компоненты и $\bar{n}_z(t)$ – комплексная огибающая процесса $n(t)$.

Определим автокорреляционную функцию случайного процесса $N(t)$:

$$\begin{aligned} R_n(\tau) &= \langle N(t)N(t+\tau) \rangle = \\ &= \left\langle \left(n_I(t)\cos[2\pi f_0 t] - n_Q(t)\sin[2\pi f_0 t] \right) \times \right. \\ &= \left. \left(n_I(t+\tau)\cos[2\pi f_0(t+\tau)] - n_Q(t+\tau)\sin[2\pi f_0(t+\tau)] \right) \right\rangle = \\ &= \frac{1}{2}(R_I(\tau) + R_Q(\tau))\cos[2\pi f_0 \tau] + \frac{1}{2}(R_I(\tau) - R_Q(\tau))\cos[2\pi f_0(2t+\tau)] - \\ &= \frac{1}{2}(R_{QI}(\tau) - R_{IQ}(\tau))\sin[2\pi f_0 \tau] - \frac{1}{2}(R_{QI}(\tau) + R_{IQ}(\tau))\sin[2\pi f_0(2t+\tau)], \end{aligned} \quad (11.32)$$

где $R_I(\tau)$, $R_Q(\tau)$, $R_{QI}(\tau)$ и $R_{IQ}(\tau)$ – автокорреляционные и взаимокорреляционные функции квадратурных компонент случайного процесса $n(t)$.

Стационарность случайного процесса предполагает отсутствие зависимости автокорреляционной функции от абсолютного момента времени t , поэтому

$$\begin{aligned} R_n(\tau) &= \langle N(t)N(t+\tau) \rangle = \\ &= \frac{1}{2}(R_I(\tau) + R_Q(\tau))\cos[2\pi f_0 \tau] - \frac{1}{2}(R_{QI}(\tau) - R_{IQ}(\tau))\sin[2\pi f_0 \tau], \end{aligned} \quad (11.33)$$

и, следовательно:

$$R_I(\tau) = R_Q(\tau), \quad (11.34)$$

$$R_{QI}(\tau) = -R_{IQ}(\tau). \quad (11.35)$$

Таким образом, окончательное выражение для автокорреляционной функции случайного процесса записывается следующим образом:

$$R_n(\tau) = R_I(\tau)\cos(2\pi f_0\tau) - R_{QI}(\tau)\sin(2\pi f_0\tau). \quad (11.36)$$

Автокорреляционная функция комплексной огибающей $n_z(t)$ определяется следующим образом:

$$\begin{aligned} \bar{R}_z(\tau) &= \frac{1}{2} \langle \bar{n}_z(t) \bar{n}_z^*(t+\tau) \rangle = \frac{1}{2} [R_I(\tau) + R_Q(\tau) + jR_{QI}(\tau) - jR_{IQ}(\tau)], \\ \bar{R}_z(\tau) &= R_I(\tau) + jR_{QI}(\tau). \end{aligned} \quad (11.37)$$

Таким образом, автокорреляционная функция исходного случайного процесса $n(t)$ определяется автокорреляционной функцией его комплексной огибающей $\bar{n}_z(t)$:

$$R_n(\tau) = \operatorname{Re} \left\{ \bar{R}_z(\tau) e^{j2\pi f_0\tau} \right\}. \quad (11.38)$$

Взаимокорреляционная функция квадратурных компонент полосового стационарного случайного процесса удовлетворяет условию симметрии. Кроме того, любая взаимокорреляционная функция удовлетворяет условию

$$R_{xy}(\tau) = R_{yx}(-\tau). \quad (11.39)$$

Из этих особенностей следует, что $R_{IQ}(\tau) = -R_{IQ}(-\tau)$, то есть функция $R_{IQ}(\tau)$ является нечетной и $R_{IQ}(0) = 0$. Из последнего следует, что квадратурные компоненты случайного процесса являются некоррелированными при отсутствии между ними временного сдвига.

Когда стационарный случайный процесс $N(t)$ является гауссовским, квадратурные компоненты $N_I(t)$ и $N_Q(t+\tau)$ совместно тоже гауссовские. Более того, при $\tau=0$ они статистически независимы, и, следовательно, их совместная плотность вероятности

$$W(n_I, n_Q) = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{n_I^2 + n_Q^2}{2\sigma^2} \right\} = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{\bar{n}_z \bar{n}_z^*}{2\sigma^2} \right\}, \quad (11.40)$$

где дисперсия σ^2 определяется как $\sigma^2 = R_I(0) = R_Q(0) = R_n(0)$.

Белый шум является случайным процессом, который имеет постоянную спектральную плотность в неограниченном диапазоне частот.

В вопросах, связанных с демодуляцией узкополосных сигналов на фоне шумов, математически удобно представить аддитивный шум как белый и выразить его через квадратурные компоненты. Это можно выполнить, предполагая, что сигнал и шум на приёмной стороне прошли через идеальный полосовой фильтр, имеющий полосу пропускания более широкую, чем полоса сигнала. Такой фильтр может внести пренебрежимо малые искажения в сигнал, но он исключает частотные компоненты шума вне полосы пропускания фильтра.

Белый шум, прошедший через идеальный полосовой фильтр, называют полосовым белым шумом.

Спектральная плотность мощности белого и полосового белого шума симметрична относительно точки $f = 0$, данная симметрия спектральной плотности приводит к тому, что $R_{IQ}(\tau) = 0$ для всех τ . Следовательно,

$$R_z(\tau) = R_I(\tau) = R_Q(\tau) = R_n(\tau).$$

Это означает, что квадратурные компоненты белого шума некоррелированы при любых временных сдвигах τ .

Таким образом, рассмотрение исходного случайного процесса в виде белого гауссовского шума $n(t)$ может быть заменено рассмотрением его комплексного представления $\bar{n}(t) = \bar{n}_z(t) e^{j2\pi f_0 t}$. При этом одномерная и многомерная плотность распределения вероятности комплексных выборочных значений $\bar{n}_0 = (n_{I0} + jn_{Q0}), \dots, \bar{n}_M = (n_{IM} + jn_{QM})$ может быть записана следующим образом:

$$W(\bar{n}_i) = W(n_{Ii}, n_{Qi}) = \frac{1}{2\pi \frac{N_0}{2\Delta t}} \exp \left\{ -\frac{n_{Ii}^2 + n_{Qi}^2}{2 \frac{N_0}{2\Delta t}} \right\} = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{\bar{n}_i \bar{n}_i^*}{2 \frac{N_0}{2\Delta t}} \right\}, \quad (11.41)$$

$$W(\bar{n}_0, \dots, \bar{n}_M) = \prod_{i=0}^M W(\bar{n}_i) = \left(\frac{1}{\sqrt{2\pi \frac{N_0}{2\Delta t}}} \right)^{M+1} \exp \left\{ -\frac{1}{N_0} \sum_{i=0}^M \bar{n}_i \bar{n}_i^* \Delta t \right\}. \quad (11.42)$$

Соответственно функционал распределения вероятности для случая непрерывного наблюдения можно записать

$$F(\bar{n}(t)) = h_0 \exp \left\{ -\frac{1}{N_0} \int_0^T \bar{n}(t) \bar{n}^*(t) dt \right\}. \quad (11.43)$$

Следует заметить, что приведенные выражения для многомерной плотности распределения вероятности комплексного представления белого гауссовского шума инвариантны к множителю $e^{j2\pi f_0 t}$.

11.5.2.3. Функция правдоподобия для канала связи с аддитивным белым гауссовским шумом

Оценки компонент вектора информационных параметров $\Lambda = [\lambda_1, \dots, \lambda_k]$ по методу максимума правдоподобия находятся как решение системы уравнений правдоподобия:

$$\frac{\partial}{\partial \lambda_1} \ln(P(X|\lambda_1, \dots, \lambda_k)) = 0, \quad (11.44)$$

⋮

$$\frac{\partial}{\partial \lambda_k} \ln(P(X|\lambda_1, \dots, \lambda_k)) = 0. \quad (11.45)$$

где $P(X|\lambda_1, \dots, \lambda_k) = P(X|\Lambda)$ – функция правдоподобия вектора параметров Λ (многомерная плотность распределения вероятности выборочных значений x_1, \dots, x_n наблюдаемого процесса $x(t)$ (аддитивной смеси полезного сигнала и помехи при условии конкретных значений $\lambda_1, \dots, \lambda_k$ вектора информационных параметров Λ).

Определим статистические свойства случайного процесса

$$x(t) = s(t, \Lambda) + n(t),$$

где $s(t, \Lambda)$ – полезный сигнал, являющийся известной функцией времени и вектора информационных параметров Λ ; $n(t)$ – белый гауссовский шум с известной спектральной плотностью $N_0/2$. Считаем, что информационные параметры сигнала известны. Зафиксируем момент времени и для него определим условное выборочное значение (при условии конкретного значения информационных параметров сигнала): $x = s(\Lambda) + n$.

Поскольку сигнал является известной функцией времени, то для его выборочного значения плотность распределения можно записать следующим образом:

$$W_s(s) = \delta(s - s(t, \Lambda)). \quad (11.46)$$

Плотность распределения вероятности выборочного значения шума является гауссовской:

$$W_n(n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{n^2}{2\sigma^2}\right\}. \quad (11.47)$$

Условная плотность распределения вероятности суммы двух независимых случайных величин s и n (при условии конкретного значения информационных параметров) находится следующим образом:

$$\begin{aligned} W_x(x|\Lambda) &= \int_{-\infty}^{\infty} W_s(x-y)W_n(y)dy = \int_{-\infty}^{\infty} W_s(y)W_n(x-y)dy = \\ &= \int_{-\infty}^{\infty} \delta(y-s(t,\Lambda))W_n(x-y)dy = W_n(x-s(t,\Lambda)). \end{aligned} \quad (11.48)$$

Производя аналогичные преобразования для многомерного случая вектора выборочных значений процесса $x(t)$ на интервале времени $[0, T]$, получим выражение для функции правдоподобия и ее предельного случая – функционала правдоподобия:

$$P(X|\Lambda) = \left(\frac{1}{\sqrt{2\pi \frac{N_0}{2\Delta t}}} \right)^{M+1} \exp\left\{-\frac{1}{N_0} \sum_{i=0}^M (x_i - s_i(\Lambda))^2 \Delta t\right\}, \quad (11.49)$$

$$F(n(t)|\Lambda) = h_0 \exp\left\{-\frac{1}{N_0} \int_0^T [x(t) - s(t, \Lambda)]^2 dt\right\}, \quad (11.50)$$

где $x_i = \frac{1}{\Delta t} \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} x(t) dt$ и $s_i(\Lambda) = \frac{1}{\Delta t} \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} s(t, \Lambda) dt$ – выборочные значения смеси

сигнала и шума и сигнала соответственно; $M + 1$ – число выборочных значений на рассматриваемом интервале времени $[0, T]$; Δt – шаг дискретизации во времени.

Аналогичные выражения получаются и для комплексного представления полезного сигнала и шума

$$P(\bar{X}|\Lambda) = \left(\frac{1}{\sqrt{2\pi \frac{N_0}{2\Delta t}}} \right)^{M+1} \exp\left\{-\frac{1}{N_0} \sum_{i=0}^M (\bar{x}_i - \bar{s}_i)(\bar{x}_i - \bar{s}_i)^* \Delta t\right\}. \quad (11.51)$$

Соответственно функционал распределения вероятности для случая непрерывного наблюдения можно записать

$$F(\bar{x}(t)|\Lambda) = h_0 \exp \left\{ -\frac{1}{N_0} \int_0^T (\bar{x}(t) - \bar{s}(t, \Lambda)) (\bar{x}(t) - \bar{s}(t, \Lambda))^* dt \right\}, \quad (11.52)$$

где знак \rightarrow обозначает комплексное представление соответствующего процесса, а $*$ – операцию комплексного сопряжения.

11.5.3. Обобщенные решающие правила при квазиоптимальном приеме сигналов с дискретной модуляцией

11.5.3.1. Демодуляция информационного символа в целом

Для упрощения рассмотрим случай, когда вектор информационных параметров полезного сигнала представлен только одной компонентой λ . Будем считать, что на заданном интервале времени τ_c (длительность символа модуляции) информационный параметр является постоянным и принимает одно из L известных значений (L является размером ансамбля символов модуляции). Например,

$$s(t, \lambda_k) = A_0 \cos(\omega_0 t + \lambda_k),$$

где λ_k – одно из возможных значений начальной фазы сигнала.

Конкретное значение информационного параметра соответствует конкретному значению символа передаваемой информации. Например, для случая $L = 2$ значение $\lambda_1 = 0$ и $\lambda_2 = \pi$ могут соответствовать передаче двоичной «1» и двоичного «0», соответственно, или, при $L = 32$, значения $\lambda_1 = 0$, $\lambda_2 = \pi/16$, $\lambda_3 = 2\pi/16$..., $\lambda_{32} = 31\pi/16$ могут соответствовать буквам кириллицы. Схема кодирования может быть, в общем случае произвольной и определяется видом модуляции.

Общим для всех видов модуляции является то, что каждому модулирующему символу ставится в соответствие свой конкретный несущий сигнал. Так в первом примере конкретным несущим сигналом для двоичной «1» является

$$s(t, \lambda_1) = s^1(t) = A_0 \cos(\omega_0 t), \quad (11.53)$$

а для двоичного «0»

$$s(t, \lambda_2) = s^2(t) = A_0 \cos(\omega_0 t + \pi) = -A_0 \cos(\omega_0 t + \pi). \quad (11.54)$$

Таким образом, вне зависимости от схемы модуляции (вида модуляции) полезный сигнал может быть представлен как:

$$s(t, \lambda) = \begin{cases} s^1(t) & \lambda = \lambda_1 \\ s^2(t) & \lambda = \lambda_2 \\ \vdots & \vdots \\ s^L(t) & \lambda = \lambda_L \end{cases}. \quad (11.55)$$

В соответствии с этим квазиоптимальный демодулятор должен формировать (вычислить) L значений функции правдоподобия (для каждого конкретного несущего сигнала) $s^k(t)$

$$P(\bar{X} | \lambda_k) = \left(\frac{1}{\sqrt{2\pi \frac{N_0}{2\Delta t}}} \right)^{M+1} \exp \left\{ -\frac{1}{N_0} \sum_{i=0}^M (\bar{x}_i - \bar{s}_i^k)(\bar{x}_i - \bar{s}_i^k)^* \Delta t \right\}. \quad (11.56)$$

Множитель при экспоненте и спектральная плотность мощности шума наблюдения входят во все выражения для частных функций правдоподобия и не оказывают влияние на поиск максимального значения, на основании этого функцию правдоподобия можно упростить:

$$\begin{aligned} P(\bar{X} | \lambda_k) &= \exp \left\{ -\frac{1}{N_0} \sum_{i=0}^M (\bar{x}_i - \bar{s}_i^k)(\bar{x}_i - \bar{s}_i^k)^* \Delta t \right\} = \\ &= \exp \left\{ -\sum_{i=0}^M (\bar{x}_i - \bar{s}_i^k)(\bar{x}_i - \bar{s}_i^k)^* / (2\sigma^2) \right\}. \end{aligned} \quad (11.57)$$

Множитель $\frac{1}{2\sigma^2} = \frac{\Delta t}{N_0}$ практического значения для реализации алгоритма демодуляции не имеет, но оставлен для сохранения правильной размерности в математических выражениях.

Для уменьшения вычислительной сложности поиск максимума можно вести не между непосредственными значениями функции правдоподобия, а между любыми монотонными преобразованиями от нее (в частности, ее натурального логарифма), тогда оптимальная оценка информационного параметра определяется следующим образом:

$$\begin{aligned} \lambda^* = \lambda_k &\Rightarrow \max \{ P(\bar{X} | \lambda), \lambda = \lambda_1, \dots, \lambda_L \} = \\ &= \max \left\{ -\sum_{i=0}^M (\bar{x}_i - \bar{s}_i^k)(\bar{x}_i - \bar{s}_i^k)^* / (2\sigma^2), s = s^1, \dots, s^L \right\}. \end{aligned} \quad (11.58)$$

В соответствии с полученной оценкой информационного параметра λ^* (значения λ_k , которое наиболее вероятно принимает информационный параметр в полезном сигнале, наблюдаемом совместно с шумом на интервале времени τ_c) и принятой схемой модуляции в дальнейшем производится декодирование принятого информационного символа.

11.5.3.2. Поэлементная демодуляция информационного символа

При передаче цифровых сообщений каждому информационному символу (символу модуляции, конкретному несущему сигналу, представляющему информационный символ) соответствует двоичное число, содержащее K двоичных символов d_m , где m – номер двоичного разряда, например в случае 8-позиционной частотной манипуляции:

$$\begin{aligned} d_1 d_2 d_3 &\Leftrightarrow s^i(t) = A_0 \cos(\omega_i t) = s^{d_1 d_2 d_3}(t) \\ 000 &\Leftrightarrow s^1(t) = A_0 \cos(\omega_1 t) = s^{000}(t) \\ 001 &\Leftrightarrow s^2(t) = A_0 \cos(\omega_2 t) = s^{001}(t) \\ 010 &\Leftrightarrow s^3(t) = A_0 \cos(\omega_3 t) = s^{010}(t) \\ 011 &\Leftrightarrow s^4(t) = A_0 \cos(\omega_4 t) = s^{011}(t) \\ 100 &\Leftrightarrow s^5(t) = A_0 \cos(\omega_5 t) = s^{100}(t) \\ 101 &\Leftrightarrow s^6(t) = A_0 \cos(\omega_6 t) = s^{101}(t) \\ 110 &\Leftrightarrow s^7(t) = A_0 \cos(\omega_7 t) = s^{110}(t) \\ 111 &\Leftrightarrow s^8(t) = A_0 \cos(\omega_8 t) = s^{111}(t) \end{aligned}$$

Сущность поэлементного приема (демодуляции) заключается в том, что демодулируется не значение информационного символа $d_1 d_2 d_3$ непосредственно с последующим его декодированием, а производится оценка (демодуляция) непосредственно элементов $d_m, m=1,2,3\dots$, составляющих информационный символ.

При поэлементном приеме также применима методология синтеза решающих правил на основе максимума правдоподобия.

Поскольку каждый элемент информационного символа принимает только одно из двух возможных значений, то более удобным является использование не функции правдоподобия, а так называемого отношения правдоподобия, определяемого следующим образом:

$$L_m(\vec{X}) = \frac{P(\vec{X}|d_m=1)}{P(\vec{X}|d_m=0)}, \quad (11.59)$$

где $P(\vec{X}|d_m=1)$ и $P(\vec{X}|d_m=0)$ – функции правдоподобия (в общем случае апостериорные вероятности принятия элементом d_m значений «1» и «0»), соответственно.

Принятие решения о принятом значении элемента осуществляется в соответствии со следующим правилом:

$$\begin{cases} L_m(\vec{X}) \geq L_0 \Rightarrow d_m^* = 1, \\ L_m(\vec{X}) < L_0 \Rightarrow d_m^* = 0, \end{cases} \quad (11.60)$$

где L_0 – порог принятия решения. Для случая симметричного канала (равновероятного принятия значений «1» и «0» рассматриваемым элементом d_m и одинаковых потерь при искажении передаваемых «1» и «0») значение порога принятия решения равно $L_0 = 1$.

С порогом можно сравнивать не непосредственно отношение правдоподобия, а любое монотонное преобразование от него, в частности натуральный логарифм. Тогда для симметричного канала правило принятия решения можно переписать:

$$\begin{cases} \ln(L_m(\vec{X})) \geq 0 \Rightarrow d_m^* = 1, \\ \ln(L_m(\vec{X})) < 0 \Rightarrow d_m^* = 0. \end{cases} \quad (11.61)$$

Учитывая несовместность событий присутствия на входе демодулятора конкретных несущих сигналов (одновременно передается только один символ модуляции λ_k и, соответственно, только один несущий сигнал $s^k(t)$), апостериорная вероятность (функция правдоподобия) принятия значений «1» и «0» конкретным элементом может быть определена на основании теоремы суммирования вероятностей:

$$P(\vec{X}|d_m=1) = \sum_{\forall \lambda_k \Rightarrow d_m=1} P(\vec{X}|\lambda_k), \quad (11.62)$$

$$P(\vec{X}|d_m=0) = \sum_{\forall \lambda_k \Rightarrow d_m=0} P(\vec{X}|\lambda_k). \quad (11.63)$$

Таким образом, при поэлементном приеме необходимо сформировать (вычислить) для каждого элемента d_m свое отношение правдоподобия

$$L_m(\bar{X}) = \frac{\sum_{\forall \lambda_k \Rightarrow d_m=1} P(\bar{X}|\lambda_k)}{\sum_{\forall \lambda_k \Rightarrow d_m=0} P(\bar{X}|\lambda_k)} \quad (11.64)$$

и по результатам его сравнения с пороговым значением сформировать решение d_m^* о переданном значении элемента.

Проиллюстрируем полученное решающее правило на примере передачи трех рассмотренных выше элементарных двоичных символов при условии наблюдения сигнала на фоне белого гауссовского шума с заданными характеристиками. Запишем отношение правдоподобия для второго элемента d_2 модулирующего (информационного) символа $d_1 d_2 d_3$,

$$\begin{aligned} P(\bar{X}|d_2=1) = & \exp \left\{ -\frac{\sum_{i=0}^M (\bar{x}_i - \bar{s}_i^{010})(\bar{x}_i - \bar{s}_i^{010})^*}{2\sigma^2} \right\} + \\ & + \exp \left\{ -\frac{\sum_{i=0}^M (\bar{x}_i - \bar{s}_i^{011})(\bar{x}_i - \bar{s}_i^{011})^*}{2\sigma^2} \right\} + \exp \left\{ -\frac{\sum_{i=0}^M (\bar{x}_i - \bar{s}_i^{110})(\bar{x}_i - \bar{s}_i^{110})^*}{2\sigma^2} \right\} + \\ & \exp \left\{ -\frac{\sum_{i=0}^M (\bar{x}_i - \bar{s}_i^{111})(\bar{x}_i - \bar{s}_i^{111})^*}{2\sigma^2} \right\}, \end{aligned} \quad (11.65)$$

$$\begin{aligned}
P(\vec{X} | d_2 = 0) = & \exp \left\{ - \frac{\sum_{i=0}^M (\vec{x}_i - \vec{s}_i^{000})(\vec{x}_i - \vec{s}_i^{000})^*}{2\sigma^2} \right\} + \\
& + \exp \left\{ - \frac{\sum_{i=0}^M (\vec{x}_i - \vec{s}_i^{001})(\vec{x}_i - \vec{s}_i^{001})^*}{2\sigma^2} \right\} + \exp \left\{ - \frac{\sum_{i=0}^M (\vec{x}_i - \vec{s}_i^{100})(\vec{x}_i - \vec{s}_i^{100})^*}{2\sigma^2} \right\} + \\
& + \exp \left\{ - \frac{\sum_{i=0}^M (\vec{x}_i - \vec{s}_i^{101})(\vec{x}_i - \vec{s}_i^{101})^*}{2\sigma^2} \right\}, \quad (11.66)
\end{aligned}$$

$$L_2(\vec{X}) = \frac{P(\vec{X} | d_2 = 1)}{P(\vec{X} | d_2 = 0)}, \quad (11.67)$$

где \vec{x}_i – отсчеты комплексного представления входной смеси полезного сигнала и шума; $\vec{s}_i^{d_1 d_2 d_3}$ – отсчеты комплексного представления конкретного сигнала, несущего информационный символ $d_1 d_2 d_3$; $M + 1$ – количество выборочных отсчетов приходящихся на длительность τ_c информационного символа (символа модуляции); $\sigma^2 = \frac{N_0}{2\Delta t}$ – дисперсия выборочных значений шума наблюдения.

Если представить выборочные значения соответствующих комплексных представлений процессов $\vec{x}(t)$ и $\vec{s}^{d_1 d_2 d_3}(t)$ как координаты векторов на комплексной плоскости, то полученное правило принятия решения, по сути, является правилом поиска (выбора) того из сигналов $\vec{s}^{d_1 d_2 d_3}(t)$, изображающий вектор которого максимально близок к входному процессу $\vec{x}(t)$.

11.6. Помехоустойчивость процесса демодуляции

11.6.1. Общее соотношение для вероятности ошибочного приема

Будем считать, что на заданном интервале времени τ_c (длительность символа модуляции) информационный параметр является постоянным и принимает одно из L известных значений (L является размером ансамбля символов модуляции).

Как отмечалось выше, вне зависимости от схемы модуляции (вида модуляции) полезный сигнал может быть представлен как

$$s(t, \lambda) = \begin{cases} s^1(t) & \lambda = \lambda_1, \\ s^2(t) & \lambda = \lambda_2, \\ \vdots & \vdots \\ s^L(t) & \lambda = \lambda_L, \end{cases} \quad (11.68)$$

где λ_k и $s^k(t)$ – конкретное значение информационного параметра и конкретный несущий сигнал, представляющие k -й передаваемый информационный символ.

Демодулятор формирует (вычисляет) L значений функции правдоподобия (для каждого конкретного несущего сигнала) $s^k(t)$

$$P(\bar{X} | \lambda_k) = - \sum_{i=0}^M (\bar{x}_i - \bar{s}_i^k)(\bar{x}_i - \bar{s}_i^k)^* / (2\sigma^2), \quad k = 1, \dots, L, \quad (11.69)$$

$$P(\bar{X} | \lambda_k) = \left[- \sum_{i=0}^M |\bar{x}_i|^2 - \sum_{i=0}^M |\bar{s}_i^k|^2 + 2 \sum_{i=0}^M \operatorname{Re} \left\{ \bar{x}_i (\bar{s}_i^k)^* \right\} \right] / (2\sigma^2). \quad (11.70)$$

Первые два слагаемых в выражении для функции правдоподобия, представляют собой величины, пропорциональные энергии смеси полезного сигнала с шумом и энергии k -го несущего сигнала $s^k(t)$. При условии равенства энергий несущих колебаний (случай симметричного канала связи) данные слагаемые дают одинаковый вклад в итоговые значения функций правдоподобия рассчитанных для конкретных сигналов $s^k(t)$. В соответствии с этим максимум $P(\bar{X} | \lambda_k)$ достигается при достижении максимума третьим

слагаемым $\sum_{i=0}^M \operatorname{Re} \left\{ \bar{x}_i (\bar{s}_i^k)^* \right\}$. Данное слагаемое представляет собой значение оценки ненормированного коэффициента корреляции между совокупностью выборочных значений входной смеси сигнала с шумом и k -го несущего сигнала $s^k(t)$. Формально при определении несущего сигнала $s^k(t)$, приводящего к его максимизации функции правдоподобия при конкретном наблюдении входного процесса $x(t)$ демодулятор вычисляет L значений оценки ненормированного коэффициента корреляции между совокупностью выборочных значений входной смеси сигнала с шумом и всех возможных несущих сигналов $s^k(t)$, $k = 1, \dots, L$ и находит среди них максимальное. На основании данного замечания определим среднюю вероятность ошибки приема информационного символа в целом.

Вероятность ошибки P_e^c определяется как взвешенная (с учетом априорной вероятности символов) линейная комбинация вероятностей ошибки (искажения) $P_e(k)$ каждого символа:

$$P_e^c = \frac{1}{L} \sum_{k=1}^L P_e(k). \quad (11.71)$$

Обозначим значение ненормированного коэффициента корреляции между совокупностью выборочных значений входной смеси сигнала с шумом и несущего сигналов $s^k(t)$ следующим образом:

$$Z_k = \sum_{i=0}^M \operatorname{Re} \left\{ \bar{x}_i (s_i^k)^* \right\} \Leftrightarrow \int_0^{\tau_c} \operatorname{Re} \left\{ \bar{x}(t) (\bar{s}^k(t))^* \right\} dt \Leftrightarrow \int_0^{\tau_c} x(t) s^k(t) dt. \quad (11.72)$$

Определим вероятности ошибки $P_e(k)$. Для этого сначала предположим, что передавался, например, первый символ и определим вероятность искажения при передаче первого символа $P_e(1)$.

Вычисляемые значения Z_k являются случайными величинами, поскольку в их формировании участвует шум наблюдения, входящий во входную смесь $x(t)$. Гауссовский характер шума наблюдения и линейные операции формирования значений Z_k позволяют сделать заключение, что данные случайные величины также будут подчиняться гауссовскому закону распределения вероятности.

Найдем математические ожидания случайных величин Z_k :

$$\begin{aligned} \langle Z_k \rangle &= \left\langle \int_0^{\tau_c} x(t) s^k(t) dt \right\rangle = \left\langle \int_0^{\tau_c} [n(t) + s^1(t)] s^k(t) dt \right\rangle = \\ &= \int_0^{\tau_c} s^1(t) s^k(t) dt = E_c \rho_{1k}, \end{aligned} \quad (11.73)$$

где $E_c = \int_0^{\tau_c} s^k(t) s^m(t) dt \Big|_{k=m}$ — энергия, приходящаяся на один информационный символ (для симметричного канала одинакова для всех передаваемых символов); $\rho_{1k} = \frac{1}{E_c} \int_0^{\tau_c} s^1(t) s^k(t) dt$ — нормированный коэффициент корреляции между первым и k -м несущими сигналами.

Аналогично находятся коэффициенты ковариации между случайными величинами Z_k :

$$\begin{aligned}
 cov(Z_k, Z_m) &= \langle Z_k Z_m \rangle - \langle Z_k \rangle \langle Z_m \rangle = \\
 &= \left\langle \int_0^{\tau_c} \int_0^{\tau_c} [n(t) + s^1(t)] s^k(t) [n(u) + s^1(u)] s^m(u) dt du \right\rangle - E_c^2 \rho_{1k} \rho_{1m} = \\
 &= \int_0^{\tau_c} \int_0^{\tau_c} s_k(t) s_m(u) \langle n(t) n(u) \rangle dt du = \int_0^{\tau_c} \int_0^{\tau_c} s_k(t) s_m(u) \frac{N_0}{2} \delta(t-u) dt du = \\
 &= \frac{N_0 E_c}{2} \rho_{km}, \quad (11.74)
 \end{aligned}$$

где $\rho_{km} = \frac{1}{E_c} \int_0^{\tau_c} s^k(t) s^m(t) dt$ – нормированный коэффициент корреляции между k -м и m -м несущими сигналами.

Коэффициенты $cov(Z_k, Z_m)$ образуют корреляционную матрицу многомерной случайной величины $[Z_1, Z_2, \dots, Z_L]$.

После того как таким образом определены статистические характеристики случайных величин Z_k , легко найти вероятность ошибки $P_e(1)$ или вероятность правильного приема $P_{np}(1) = 1 - P_e(1)$ первого информационного символа. Она равна вероятности того, что величина Z_1 будет иметь большее значение, чем величины остальных значений Z_2, Z_3, \dots, Z_L .

$$P_{np}(1) = P \left\{ Z_1 = \max_{k \in [1, \dots, L]} Z_k \right\} = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{Z_1} \dots \int_{-\infty}^{Z_1} W(\mathbf{Z}) dZ_2 \dots dZ_L \right] dZ_1, \quad (11.75)$$

где $W(\mathbf{Z}) = W(Z_1, Z_1, \dots, Z_L)$ – L -мерная нормальная плотность вероятности с корреляционной матрицей $cov(Z_k, Z_m)$ и средними значениями $\langle Z_k \rangle$. Следует иметь в виду, что для некоторых наборов несущих сигналов $s^k(t)$ матрица коэффициентов ковариации является сингулярной. Это означает, что невозможно определить многомерную гауссовскую плотность распределения вероятности $W(\mathbf{Z})$ ввиду однозначного определения одних случайных величин Z_k другими.

Осуществив ряд замен переменных интегрирования, произведем дальнейшую формализацию выражения для вероятности правильного приема.

$$\begin{aligned}
P_{\text{np}}(\mathbf{1}) &= \begin{cases} v_k = Z_k - \langle Z_k \rangle, \\ v_1 = Z_1 - \langle Z_1 \rangle \Rightarrow Z_1 = v_1 + \langle Z_1 \rangle, \\ dv_k = dZ_k, \\ Z_k = -\infty \Rightarrow v_k = -\infty, \\ Z_k = Z_1 \Rightarrow v_k = v_1 + \langle Z_1 \rangle - \langle Z_k \rangle \end{cases} = \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{v_1 + \langle Z_1 \rangle - \langle Z_2 \rangle} \int_{-\infty}^{v_1 + \langle Z_1 \rangle - \langle Z_L \rangle} \dots \int_{-\infty}^{v_1 + \langle Z_1 \rangle - \langle Z_L \rangle} w(\mathbf{V}) dv_2 \dots dv_L \right] dv_1. \\
P_{\text{np}}(\mathbf{1}) &= \begin{cases} u_k = \sqrt{2/(E_c N_0)} v_k \\ du_k = \sqrt{2/(E_c N_0)} dv_k \\ v_k = -\infty \Rightarrow u_k = -\infty \\ v_k = v_1 + \langle Z_1 \rangle - \langle Z_k \rangle \Rightarrow u_k = u_1 + \sqrt{2/(E_c N_0)} (\langle Z_1 \rangle - \langle Z_k \rangle) \end{cases} = \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{u_1 + \sqrt{2/(E_c N_0)} (\langle Z_1 \rangle - \langle Z_2 \rangle)} \int_{-\infty}^{u_1 + \sqrt{2/(E_c N_0)} (\langle Z_1 \rangle - \langle Z_L \rangle)} \dots \int_{-\infty}^{u_1 + \sqrt{2/(E_c N_0)} (\langle Z_1 \rangle - \langle Z_L \rangle)} w_{\text{н}}(\mathbf{U}) du_2 \dots du_L \right] du_1 = \\
&= \langle Z_1 \rangle - \langle Z_k \rangle = E_c (1 - \rho_{1k}) = \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{u_1 + \sqrt{2E_c/N_0} (1 - \rho_{12})} \int_{-\infty}^{u_1 + \sqrt{2E_c/N_0} (1 - \rho_{1L})} \dots \int_{-\infty}^{u_1 + \sqrt{2E_c/N_0} (1 - \rho_{1L})} w_{\text{н}}(\mathbf{U}) du_2 \dots du_L \right] du_1,
\end{aligned} \tag{11.77}$$

где $w(\mathbf{V})$ – L -мерная гауссовская плотность распределения вероятностей с корреляционной матрицей, состоящей из элементов $\text{cov}(Z_k, Z_m)$ и нулевым вектором средних значений; $w_{\text{н}}(\mathbf{U})$ – L -мерная гауссовская плотность распределения вероятностей с нормированной корреляционной матрицей, состоящей из элементов ρ_{km} и нулевым вектором средних значений.

Таким образом, вероятность правильного приема первого символа будет равна

$$\begin{aligned}
P_{np}(1) &= 1 - P_e(1) = \\
&= \frac{1}{(2\pi)^{L/2} |\boldsymbol{\rho}|^{1/2}} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{u_1 + \sqrt{2E_c/N_0}(1-\rho_{12})} \dots \right. \\
&\dots \left. \int_{-\infty}^{u_1 + \sqrt{2E_c/N_0}(1-\rho_{1L})} \exp\left\{-\frac{1}{2} \mathbf{u}^T \boldsymbol{\rho}^{-1} \mathbf{u}\right\} du_2 \dots du_L \right) du_1,
\end{aligned} \tag{11.78}$$

где $|\boldsymbol{\rho}|$ – определитель нормированной корреляционной матрицы; $\boldsymbol{\rho}^{-1}$ – матрица, обратная нормированной корреляционной матрице.

Аналогичным образом могут быть получены выражения для вероятности правильного приема $P_{np}(k)$ всех остальных символов и вероятности ошибочного приема для всех символов $P_e(k) = 1 - P_{np}(k)$, определяющие общую вероятность ошибки приема символов $P_e^c = \frac{1}{L} \sum_{k=1}^L P_e(k)$.

11.6.2. Случай ортогональных сигналов

Рассмотрим случай ортогональных несущих сигналов. Для таких сигналов нормированная корреляционная матрица представляет собой единичную матрицу $\rho_{rm} = 0, \forall k \neq m$. В этом случае вероятность правильного приема, например, условно первого символа равна

$$P_{np}(1) = 1 - P_e(1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u_1^2/2} \left[\prod_{k=2}^L \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_1 + \sqrt{2E_c/N_0}} e^{-u_k^2/2} du_k \right] du_1 = \tag{11.79}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(F(x + \sqrt{2E_c/N_0}) \right)^{L-1} dx,$$

где $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$ – функция распределения вероятности нормальной случайной величины с нулевым математическим ожиданием и единичной дисперсией.

Следовательно, вероятность искажения первого символа равна

$$P_e(1) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(F(x + \sqrt{2E_c/N_0}) \right)^{L-1} dx. \tag{11.80}$$

Полученное выражение справедливо для любого из L используемых информационных символов. Поэтому $P_e(k) = P_e(1)$ для $k = 1, 2, \dots, L$. Соответственно, вероятность ошибки может быть определена следующим образом:

$$\begin{aligned}
 P_e^c &= \frac{1}{L} \sum_{k=1}^L P_e(k) = \frac{1}{L} L P_e(1) = P_e(1) = \\
 &= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(F(x + \sqrt{2E_c/N_0}) \right)^{L-1} dx.
 \end{aligned}
 \tag{11.81}$$

Полученная вероятность ошибочного приема соответствует вероятности ошибочного приема информационного символа (символа модуляции), она удобна при передаче символов соответствующих, например, буквам алфавита или конкретным командам. Если передается последовательность независимых двоичных элементов (бит), группы из которых формируют информационный символ, более удобной мерой качества является вероятность ошибочного приема отдельного бита. Если при ортогональных несущих сигналах произошла ошибка, то может быть выбрано с одинаковой вероятностью любое из $L - 1$ неправильных решений.

11.6.3. Переход от вероятности ошибки на символ модуляции к вероятности ошибки на элемент (бит) модуляции

Пусть информационный символ формируется группой из k бит, тогда общее количество информационных символов равно $L = 2^k$. Если произошла ошибка при приеме символа, то вероятность того, что исказятся ровно m из k бит, составляющих символ, равна

$$P_e^{m/k} = \frac{C_k^m}{\sum_{j=1}^k C_k^j} = \frac{C_k^m}{2^k - 1} = \frac{C_k^m}{L - 1},
 \tag{11.82}$$

где $C_k^m = \frac{k!}{m!(k-m)!}$ — число возможных сочетаний из k элементов по m ,

т. е. наборов из m элементов, выбранных из данных k элементов, наборы, отличающиеся только порядком следования элементов (но не составом), считаются одинаковыми.

Среднее количество искаженных бит при искажении символа модуляции определяется следующим образом:

$$m_{\text{cp}} = \sum_{i=1}^k iP_e^{i/k} = \frac{\sum_{i=1}^k iC_k^i}{\sum_{j=1}^k C_k^j} = \frac{k2^{k-1}}{2^k - 1}. \quad (11.83)$$

Таким образом, средняя вероятность искажения одного бита при ошибочном приеме символа модуляции равна

$$P_e^{\text{бит}} = \frac{m_{\text{cp}}}{k} = \frac{2^{k-1}}{2^k - 1}. \quad (11.84)$$

Для упрощения понимания изложенных выше соображений рассмотрим пример определения условной вероятности искажения конкретного бита при искажении принятого информационного символа.

Правильная комбинация : 000

Ошибочные комбинации :

001	<i>ошибка в одном символе</i>
010	<i>ошибка в одном символе</i>
011	<i>ошибка в двух символах</i>
100	<i>ошибка в одном символе</i>
101	<i>ошибка в двух символах</i>
110	<i>ошибка в двух символах</i>
111	<i>ошибка в трех символах</i>

В случае ошибочного приема символа 000 имеется 7 возможных равновероятных ошибочных комбинаций, из которых три комбинации отличаются от правильной только в одном элементе, три – в двух элементах и только одна искажает все элементы. Следовательно, можно определить вероятность искажения одного, двух и трех элементов:

- вероятность искажения одного элемента – 3/7;
- вероятность искажения двух элементов – 3/7;
- вероятность искажения трех элементов – 1/7.

Обратим внимание на то, что искажение одного элемента из трех соответствует вероятности ошибки 1/3; искажение двух элементов из трех соответствует вероятности ошибки 2/3; искажение трех элементов из трех возможных соответствует вероятности ошибки 3/3.

Таким образом, средняя вероятность ошибки с учетом вероятностей искажения одного, двух и трех элементов составляет $1/3 \times 3/7 + 2/3 \times 3/7 + 3/3 \times 1/7 \cong 0,571 \cong (1 \times 3/7 + 2 \times 3/7 + 3 \times 1/7)/3$.

При этом среднее количество искаженных символов $1 \times 3/7 + 2 \times 3/7 + 3 \times 1/7 = 1,714$.

Рассчитанное таким образом значение вероятности ошибки соответствует значению, определяемому в соответствии с $\frac{2^{k-1}}{2^k - 1}$ при $k = 3$.

С учетом вероятности ошибочного приема информационного символа, состоящего из k бит (элементов) имеем вероятность ошибки приема отдельного бита

$$P_e^b = \frac{2^{k-1}}{2^k - 1} P_e^c = \frac{2^{k-1}}{2^k - 1} \left[1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(F(x + \sqrt{2E_c/N_0}) \right)^{L-1} dx \right] = \quad (11.85)$$

$$= \frac{2^{k-1}}{2^k - 1} \left[1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(F(x + \sqrt{2kE_b/N_0}) \right)^{2^k - 1} dx \right],$$

где $E_b = E_c/k$ – энергия, приходящаяся на один бит передаваемой информации.

На рис. 11.6 и 11.7 приведены зависимости вероятности ошибочного приема одиночных бит от отношения энергии, приходящейся на один бит передаваемой информации, к спектральной плотности мощности шума наблюдения E_b/N_0 и от отношения энергии, приходящейся на один информационный символ, к спектральной плотности мощности шума наблюдения E_c/N_0 при различной длине информационного символа $k = \log_2 L$.

Беглый взгляд на приведенные зависимости оставляет противоречивое ощущение: в соответствии с одними зависимостями лучшей помехоустойчивостью (меньшей вероятностью ошибочного приема при одинаковом отношении сигнал/шум) обладают схемы модуляции с большим количеством информационных символов $L = 2^k$, а в соответствии с другими – наоборот. Если здесь противоречие?

Во-первых, следует обратить внимание на то, что на обоих рисунках кривые для случая бинарной модуляции ($k = 1$) совпадают. Различия возникают для так называемых многопозиционных методов модуляции. Чтобы разобраться с данными различиями проанализируем полученные зависимости подробнее.

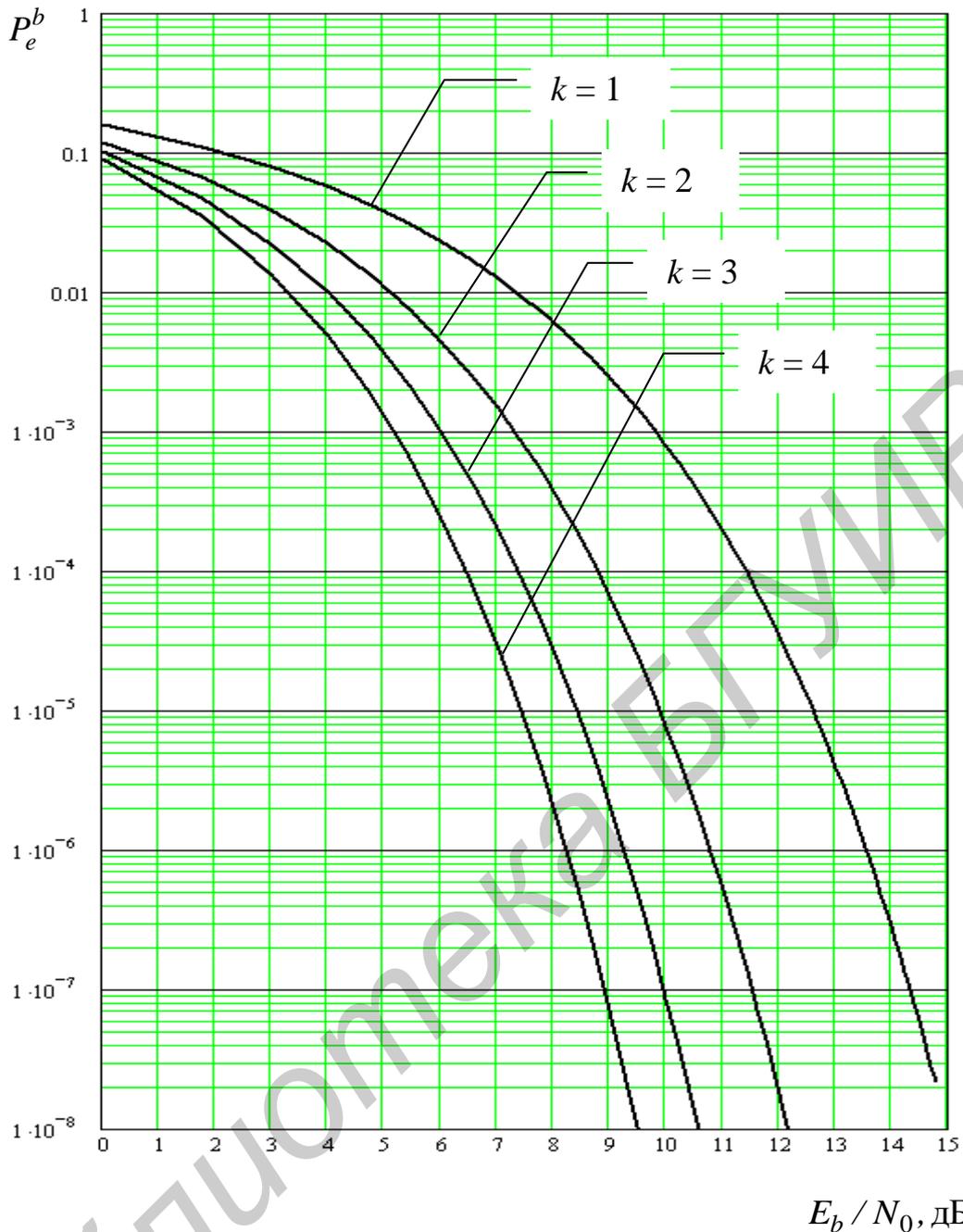


Рис. 11.6. Зависимость вероятности ошибки от отношения сигнал/шум

Энергия, приходящаяся на один информационный символ, $E_c = P_s \tau_c$, численно равна произведению мощности несущего данный символ сигнала P_s на его длительность τ_c . В случае, когда информационный символ представляет один передаваемый элемент (бит), вся эта энергия затрачивается на передачу данного бита, т. е. $E_c = E_b$ и $\tau_c = \tau_b$. В случае, когда информационный символ представляет k элементарных бит, энергия E_c распределяется между несколькими элементами и $E_b = E_c/k$, при этом возрастает в k раз количество переданных элементов (бит) за интервал времени τ_c , т. е. одна и та же энергия E_c затрачивается на перенос значительно большего количества

информации, что закономерно ухудшает качество (помехоустойчивость) данного переноса. В то же время, если сравнивать различные схемы модуляции при фиксированной энергии, приходящейся на один элементарный бит E_b , увеличение количества элементов, которые представляет информационный символ, приводит к увеличению энергии несущего сигнала $E_c = kE_b$, что при, например, фиксированной длительности τ_c эквивалентно повышению мощности P_s , т. е. кажущееся повышение помехоустойчивости (рис. 11.6) с ростом величины k эквивалентно ухудшению чувствительности демодулятора.

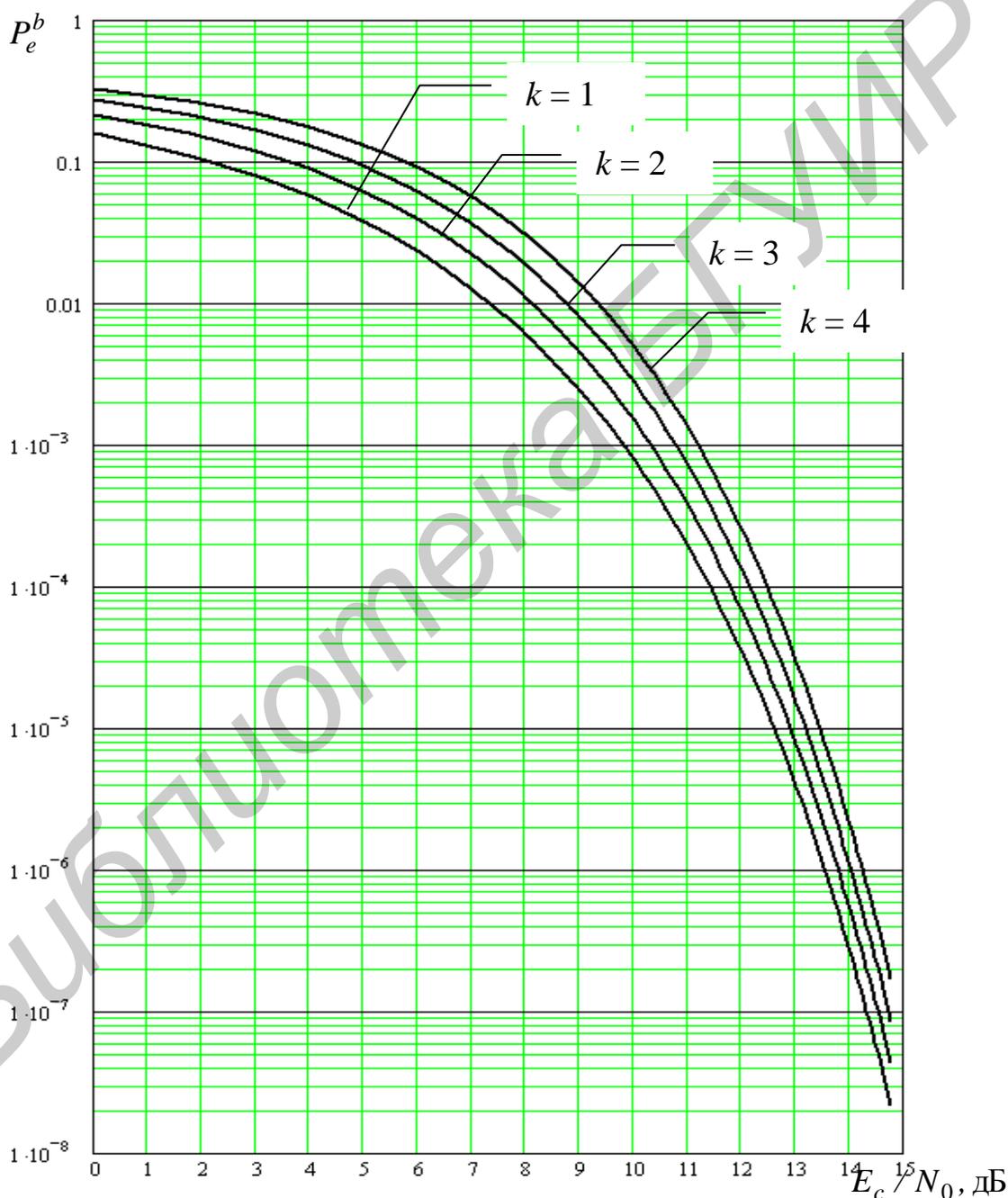


Рис. 11.7. Зависимость вероятности ошибки от отношения сигнал/шум

11.7. Основные методы дискретной модуляции

11.7.1. Фазовая манипуляция (PSK)

11.7.1.1. Бинарная фазовая манипуляция (BPSK)

Бинарная фазовая манипуляция – вид дискретной модуляции, при которой радиосигнал в общем виде может быть представлен следующим образом:

$$s(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) \cos[\omega_0 t + \varphi_0 + C_k \Theta], \quad (11.86)$$

где A_0 , ω_0 , φ_0 – амплитуда, центральная частота и начальная фаза несущего колебания; Θ – принятая девиация фазы; $C_k = \pm 1$ – последовательность, соответствующая передаваемым информационным символам «0» и «1» (обычно +1 соответствует «0», а –1 представляет «1»); $g_F(t) = 0 \quad \forall t \notin [0, \tau_b]$ – формирующий импульс; τ_b – длительность информационного символа (бита); k – текущий номер передаваемого информационного символа.

Примечание. Длительность формирующего импульса может превосходить длительность информационного символа, при этом происходит наложение передаваемых символов друг на друга. Для упрощения понимания основных процедур при демодуляции сигналов далее будем считать, что длительность формирующего импульса соответствует длительности модулирующего символа.

Формирующий импульс в общем случае может иметь произвольную форму (прямоугольную, треугольную, пилообразную и пр.) $g_F(t)$. Обычно форма импульса выбирается исходя из требований к форме спектра формируемого радиосигнала или прямоугольной. Влияние длительности и формы формирующего импульса будет рассмотрено позже в подразд. 11.8, посвященном спектральным характеристикам модулированных радиосигналов.

Пример временной функции, соответствующей радиосигналу при условии девиации фазы, равной $\pi/2$, и прямоугольной формы формирующего импульса, приведен на рис. 11.8.

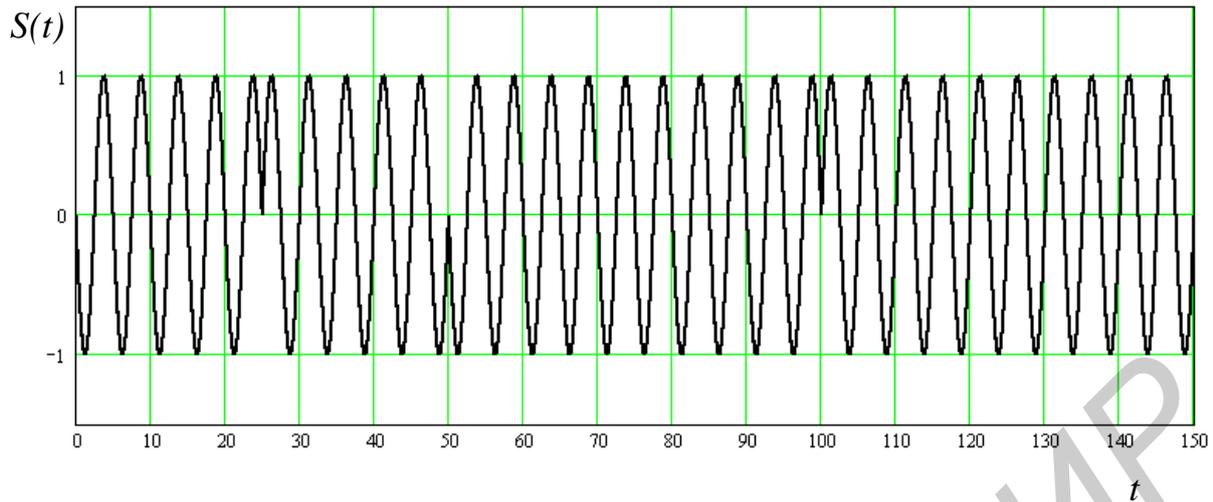


Рис. 11.8. Форма радиосигнала с BPSK

Комплексное представление радиосигнала с BPSK может быть определено следующим образом:

$$s(t) = A_0 \left\{ \begin{array}{l} \sum_{k=0}^{\infty} C_k g_F(t - k\tau_b) \cos[\Theta] \cos[\omega_0 t] - \\ - \sum_{k=0}^{\infty} C_k g_F(t - k\tau_b) \sin[\Theta] \sin[\omega_0 t] \end{array} \right\} =$$

$$= S_I(t) \cos[\omega_0 t] - S_Q(t) \sin[\omega_0 t]; \quad (11.87)$$

$$S_I(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) \cos[C_k \Theta]; \quad (11.88)$$

$$S_Q(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) \sin[C_k \Theta]; \quad (11.89)$$

$$\vec{s}(t) = [S_I(t) + jS_Q(t)] e^{j\omega_0 t} = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) \left\{ \begin{array}{l} \cos[\omega_0 t + C_k \Theta] + \\ + j \sin[\omega_0 t + C_k \Theta] \end{array} \right\} =$$

$$= A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) e^{jC_k \Theta} e^{j\omega_0 t} . \quad (11.90)$$

Комплексная огибающая сигнала с бинарной фазовой манипуляцией полностью определяется его квадратурной и синфазной компонентами и представляется следующим выражением:

$$\vec{S}(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) e^{jC_k \Theta} = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) (\cos[C_k \Theta] + j \sin[C_k \Theta]) . \quad (11.91)$$

Векторная диаграмма сигнала с бинарной фазовой манипуляцией представлена на рис. 11.9.

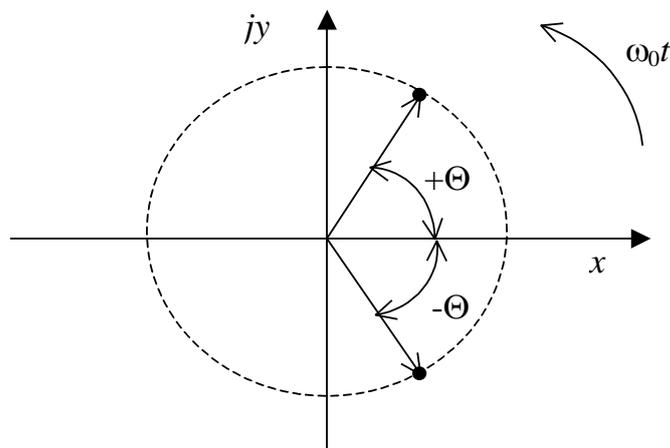


Рис. 11.9. Векторная диаграмма сигнала с $\pm\Theta$ -BPSK

При квазиоптимальном приеме решение (решающее правило при демодуляции) принимается на основании наблюдения входной смеси сигнала и шума на интервале времени, соответствующем длительности τ_b , с учетом этого для интервала наблюдения на k -м интервале модуляции комплексная огибающая может быть переписана в следующем виде:

$$\begin{aligned} \vec{S}(t) &= A_0 g_F(t - k\tau_b) e^{jC_k \Theta} = A_0 g_F(t - k\tau_b) [\cos(C_k \Theta) + j \sin(C_k \Theta)] = \\ &= A_0 g_F(t - k\tau_b) [\cos(\Theta) + j C_k \sin(\Theta)]. \end{aligned} \quad (11.92)$$

А соответствующий ей радиосигнал будет иметь следующий вид:

$$s(t) = A_0 g_F(t - k\tau_b) [\cos(\Theta) \cos(\omega_0 t) - C_k \sin(\Theta) \sin(\omega_0 t)]. \quad (11.93)$$

Таким образом, сигнал, несущий информацию о передаче двоичной «1», можно записать как

$$s^1(t) = A_0 g_F(t - k\tau_b) [\cos(\Theta) \cos(\omega_0 t) + \sin(\Theta) \sin(\omega_0 t)], \quad (11.94)$$

и сигнал несущий информацию о передаче двоичного «0»:

$$s^0(t) = A_0 g_F(t - k\tau_b) [\cos(\Theta) \cos(\omega_0 t) - \sin(\Theta) \sin(\omega_0 t)]. \quad (11.95)$$

Функционал правдоподобия, соответствующий передаче двоичной «1»:

$$\begin{aligned}
P(\vec{X} | 1) &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\vec{X}(t) - \vec{S}^1(t)] \times [\vec{X}(t) - \vec{S}^1(t)]^* dt \right\} = \\
&= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\vec{X}(t) - A_0 g_F(t - k\tau_b) [\cos(\Theta) - j \sin(\Theta)]] \times \right. \\
&\quad \left. \times [\vec{X}(t) - A_0 g_F(t - k\tau_b) [\cos(\Theta) - j \sin(\Theta)]]^* dt \right\} \quad (11.96)
\end{aligned}$$

и передаче двоичного «0»:

$$\begin{aligned}
P(\vec{X} | 0) &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\vec{X}(t) - \vec{S}^0(t)] \times [\vec{X}(t) - \vec{S}^0(t)]^* dt \right\} = \\
&= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\vec{X}(t) - A_0 g_F(t - k\tau_b) [\cos(\Theta) + j \sin(\Theta)]] \times \right. \\
&\quad \left. \times [\vec{X}(t) - A_0 g_F(t - k\tau_b) [\cos(\Theta) + j \sin(\Theta)]]^* dt \right\},
\end{aligned}$$

где $\vec{X}(t)$ – комплексная огибающая входной смеси полезного сигнала и шума.

Решение о значении передаваемого информационного символа C_k на k -м шаге модуляции формируется на основании сравнения рассчитанных значений функционала (функции для дискретного времени наблюдения) правдоподобия.

$$P(\vec{X} | 0) > P(\vec{X} | 1) \Rightarrow C_k^* = 1 \quad (\text{передавался «0»}) \quad (11.97)$$

$$P(\vec{X} | 0) < P(\vec{X} | 1) \Rightarrow C_k^* = -1 \quad (\text{передавалась «1»}) \quad (11.98)$$

При демодуляции сигнала с BPSK особое внимание следует обратить на точность восстановления опорного колебания, используемого при квадратурном преобразовании частоты, для получения комплексной огибающей входной смеси полезного сигнала и помехи (рис. 11.10).

При квадратурном преобразовании частоты выполняются следующие соотношения:

$$(X_I(t) + jX_Q(t))e^{-j\varphi(t)} = I(t) + jQ(t), \quad (11.99)$$

где $X_I(t)$ и $X_Q(t)$ – синфазная и квадратурная компоненты комплексной огибающей входной смеси полезного сигнала и шума; $\varphi(t)$ – ошибка опорного колебания квадратурного преобразователя частоты относительно несущего колебания полезного сигнала. Для правильного принятия решений при

демодуляции необходимо выполнение условия $\varphi(t) \rightarrow 0$ (условие когерентной демодуляции). Вопросы компенсации фазовой ошибки $\varphi(t)$ будут рассмотрены позже.

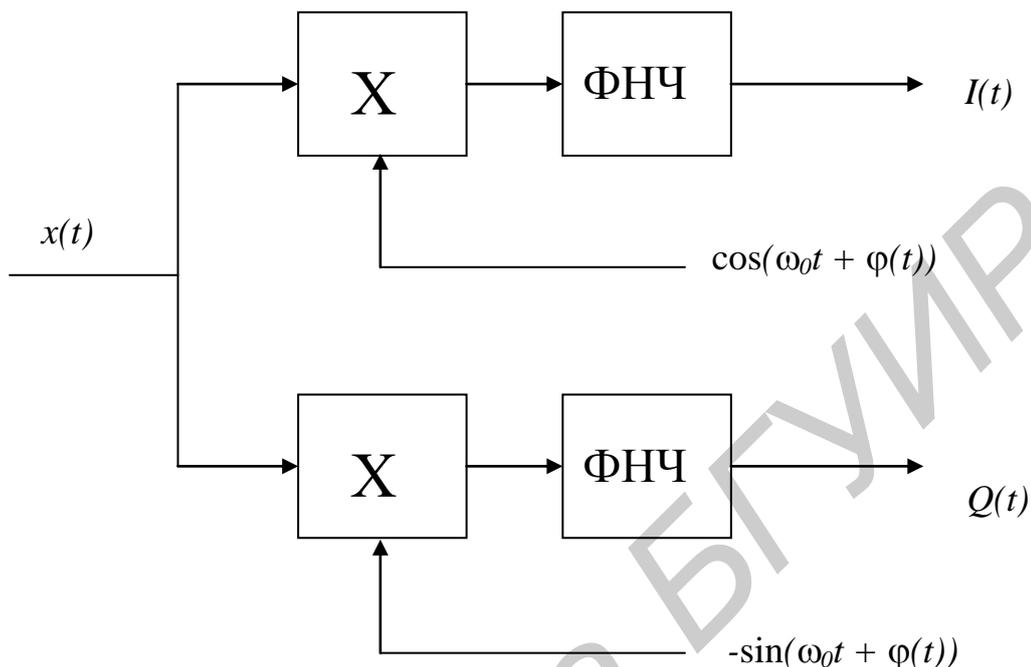


Рис. 11.10. Квадратурный преобразователь частоты.

Рассмотрим вероятность ошибки. Определим элементы нормированной корреляционной матрицы ρ сигнала с BPSK.

$$\rho_{km} = \frac{1}{E_b} \int_0^{\tau_b} s^k(t) s^m(t) dt = \frac{1}{E_b} \int_0^{\tau_b} s^k(t) s^m(t) dt; \quad (11.100)$$

$$\rho_{11} = \rho_{22} = \frac{1}{E_b} \int_0^{\tau_b} [A_0 \cos(\Theta) \cos(\omega_0 t) \mp A_0 \sin(\Theta) \sin(\omega_0 t)]^2 dt = \quad (11.101)$$

$$= \frac{1}{E_b} \frac{A_0^2 \tau_b}{2} = 1,$$

где $E_b = \frac{A_0^2 \tau_b}{2}$.

$$\begin{aligned}
\rho_{12} = \rho_{21} &= \frac{1}{E_b} \int_0^{\tau_b} [A_0 \cos(\Theta) \cos(\omega_0 t) - A_0 \sin(\Theta) \sin(\omega_0 t)] \times \\
&\times [A_0 \cos(\Theta) \cos(\omega_0 t) + A_0 \sin(\Theta) \sin(\omega_0 t)] dt = \\
&= -\frac{1}{E_b} A_0^2 \int_0^{\tau_b} [(\cos(\Theta) \cos(\omega_0 t))^2 - (\sin(\Theta) \sin(\omega_0 t))^2] dt = \\
&= \frac{1}{E_b} \frac{A_0^2 \tau_b}{2} \cos(2\Theta) = \cos(2\Theta) \dots; \\
\rho &= \begin{pmatrix} 1 & \cos(2\Theta) \\ \cos(2\Theta) & 1 \end{pmatrix}. \tag{11.103}
\end{aligned}$$

Вероятность ошибочного приема информационного символа (бита) можно определить на основании общего выражения для ошибки приема, полученного ранее:

$$P_e^b = 1 - \frac{1}{(2\pi)^{|\rho|^{1/2}}} \int_{-\infty}^{\infty} \int_{-\infty}^{u_1 + \sqrt{2E_b/N_0}(1-\rho_{12})} \exp\left\{-\frac{1}{2} \mathbf{u}^T \boldsymbol{\rho}^{-1} \mathbf{u}\right\} du_2 du_1. \tag{11.104}$$

Рассмотрим бинарную фазовую манипуляцию с изменением фазы на $\pm\pi/2$. Если внимательно присмотреться к сигналам, несущим различные информационные символы $s^1(t)$ и $s^0(t)$, то можно заметить, что они содержат одинаковые слагаемые, представляющие немодулированное несущее колебание $-A_0 g_F(t - k\tau_b) \cos(\Theta) \cos(\omega_0 t)$ и не зависящие от передаваемой информации (т. е. не являющиеся носителями информации). Наличие данных слагаемых приводит к неоправданному расходу энергии радиосигнала и, как следствие, к понижению помехоустойчивости его приема.

Наиболее целесообразным с точки зрения повышения энергетической эффективности метода модуляции является случай использования девиации фазы равной $\Theta = \pi/2$ ($\pm\pi/2$ -BPSK или просто BPSK). При этом неинформативная составляющая радиосигнала $A_0 \cos(\pi/2) \cos(\omega_0 t) = 0$ и вся энергия расходуется на передачу информационных символов.

В этом случае комплексная огибающая радиосигнала на интервале k -го шага модуляции равна

$$\begin{aligned}
\vec{S}(t) &= A_0 g_F(t - k\tau_b) e^{jC_k \frac{\pi}{2}} = A_0 g_F(t - k\tau_b) \left[jC_k \sin\left(\frac{\pi}{2}\right) \right] = \\
&= jA_0 g_F(t - k\tau_b) C_k = \begin{cases} jA_0 g_F(t - k\tau_b), C_k = 1 \Leftrightarrow \langle\langle 0 \rangle\rangle, \\ -jA_0 g_F(t - k\tau_b), C_k = -1 \Leftrightarrow \langle\langle 1 \rangle\rangle, \end{cases} \tag{11.105}
\end{aligned}$$

а нормированная корреляционная матрица будет иметь следующий вид:

$$\rho = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \quad (11.106)$$

Функционал правдоподобия, соответствующий передаче двоичной «1»:

$$\begin{aligned} P(\bar{X}|1) &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\bar{X}(t) - \bar{S}^1(t)] \times [\bar{X}(t) - \bar{S}^1(t)]^* dt \right\} = \\ &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\bar{X}(t) + jA_0 g_F(t - k\tau_b)] \times [\bar{X}(t) + jA_0 g_F(t - k\tau_b)]^* dt \right\} \end{aligned} \quad (11.107)$$

и передаче двоичного «0»

$$\begin{aligned} P(\bar{X}|0) &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\bar{X}(t) - \bar{S}^0(t)] \times [\bar{X}(t) - \bar{S}^0(t)]^* dt \right\} = \\ &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\bar{X}(t) - jA_0 g_F(t - k\tau_b)] \times [\bar{X}(t) - jA_0 g_F(t - k\tau_b)]^* dt \right\}, \end{aligned} \quad (11.108)$$

где $\bar{X}(t)$ – комплексная огибающая входной смеси полезного сигнала и шума.

Решение о значении передаваемого информационного символа C_k на k -м шаге модуляции формируется на основании сравнения рассчитанных значений функционала (функции для дискретного времени наблюдения) правдоподобия:

$$P(\bar{X}|0) < P(\bar{X}|1) \Rightarrow C_k^* = 1 \quad (\text{передавался «0»});$$

$$P(\bar{X}|0) > P(\bar{X}|1) \Rightarrow C_k^* = -1 \quad (\text{передавалась «1»}).$$

Поскольку нормированная корреляционная матрица сигнала с $\pi/2$ -BPSK является сингулярной (вырожденной, $|\rho| = 0$), общее выражение для нахождения вероятности ошибки непосредственно применено быть не может. Это вызвано тем, что корреляционные интегралы находятся в следующей зависимости.

$$Z_k = \sum_{i=0}^M \operatorname{Re} \left\{ \bar{x}_i (s_i^k)^* \right\} \Leftrightarrow \int_0^{\tau_c} \operatorname{Re} \left\{ \bar{x}(t) (s^k(t))^* \right\} dt \Leftrightarrow \int_0^{\tau_c} x(t) s^k(t) dt, \quad k = 0, 1.$$

Однозначно определяют друг друга лишь $Z_1 = -Z_2$. В этом случае совместная плотность распределения вероятности случайных величин отличается от гауссовской и может быть записана следующим образом:

$$W(Z_1, Z_2) = \delta(Z_1 + Z_2) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[Z_1 - \langle Z_1 \rangle]^2}{2\sigma^2}\right\}, \quad (11.109)$$

где $\sigma^2 = \frac{N_0 E_b}{2}$ и $\langle Z_1 \rangle = E_b$.

$$\begin{aligned} P_e^b &= 1 - \int_{-\infty}^{\infty} \int_{-\infty}^{Z_1} \delta(Z_1 + Z_2) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[Z_1 - \langle Z_1 \rangle]^2}{2\sigma^2}\right\} dZ_1 dZ_2 = \\ &= 1 - \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[Z + \langle Z_1 \rangle]^2}{2\sigma^2}\right\} dZ = \\ &= 1 - \int_{-\infty}^0 \frac{1}{\sqrt{\pi N_0 E_b}} \exp\left\{-\frac{[Z + E_b]^2}{N_0 E_b}\right\} dZ. \end{aligned} \quad (11.110)$$

Вспользуемся стандартной и модифицированной функцией ошибок соответственно:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad \text{и} \quad \operatorname{erfc}(x) = 1 - \operatorname{erf}(x), \quad (11.111)$$

тогда выражение для вероятности ошибочного приема можно переписать так:

$$P_e^b = \frac{1}{2} \operatorname{erfc} \sqrt{\frac{E_b}{N_0}}. \quad (11.112)$$

Необходимо отметить, что $\pm\pi/2$ -BPSK метод модуляции обладает наивысшей помехоустойчивостью. На рис. 11.11 изображены зависимости вероятности ошибочного приема от отношения E_b/N_0 для сигнала с $\pi/2$ -BPSK и сигнала с двухпозиционной модуляцией, использующего ортогональные несущие сигналы с одинаковой энергией (бинарная ортогональная модуляция). Высокая помехоустойчивость объясняется обеспечением максимально возможного расстояния между векторами, представляющими сигналы, несущие двоичные «1» и «0».

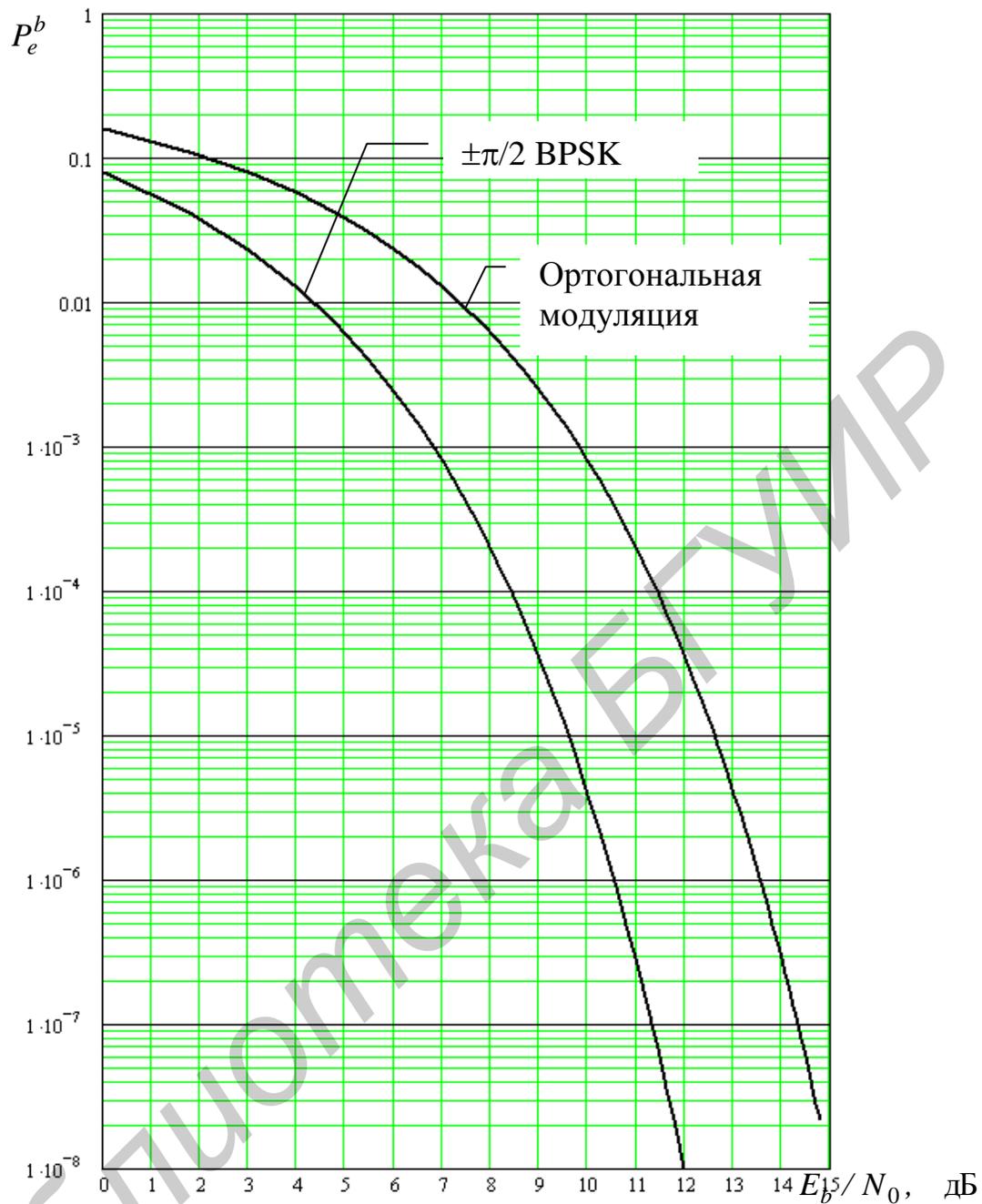


Рис. 11.11. Сравнение помехоустойчивости сигналов с $\pm\pi/2$ -BPSK и бинарной ортогональной модуляцией

11.7.1.2. M -ичная фазовая манипуляция (M-PSK)

M -ичная фазовая манипуляция – вид дискретной модуляции, при которой радиосигнал в общем виде может быть представлен следующим образом:

$$s(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \cos[\omega_0 t + \phi_0 + \Theta(k)], \quad (11.113)$$

где A_0 , ω_0 , ϕ_0 – амплитуда, центральная частота и начальная фаза несущего колебания; $\Theta(k)$ – фазовый набег соответствующий передаваемой информации; $g_F(t) = 0 \quad \forall t \notin [0, \tau_c]$ – формирующий импульс; τ_c – длительность информационного символа; k – текущий номер передаваемого информационного символа. Фазовый набег $\Theta(k)$ постоянен на интервале времени, совпадающем с интервалом действия информационного символа и принимает одно из M известных значений Θ_m (обычно $\Theta_m = \frac{2\pi(m-1)}{M}$).

Пример временной функции, соответствующей радиосигналу 8-PSK при прямоугольной форме формирующего импульса, приведен на рис. 11.12.

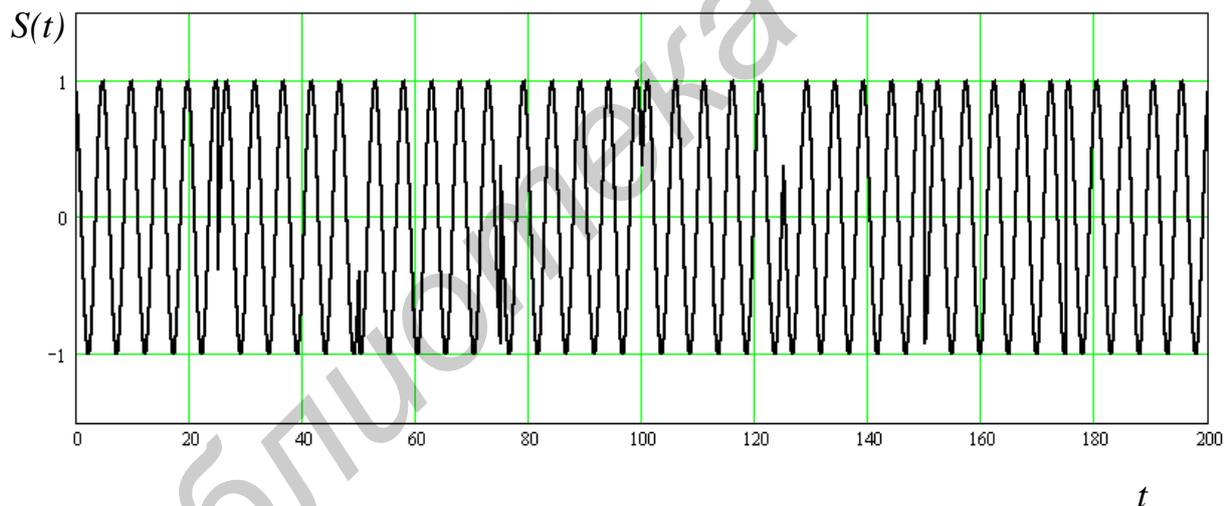


Рис. 11.12. Форма радиосигнала с 8-PSK

Комплексное представление радиосигнала с M -ичной PSK может быть определено следующим образом:

$$s(t) = A_0 \left\{ \begin{array}{l} \sum_{k=0}^{\infty} g_F(t - k\tau_c) \cos[\Theta(k)] \cos[\omega_0 t] - \\ - \sum_{k=0}^{\infty} g_F(t - k\tau_c) \sin[\Theta(k)] \sin[\omega_0 t] \end{array} \right\} =$$

$$= S_I(t) \cos[\omega_0 t] - S_Q(t) \sin[\omega_0 t], \quad (11.114)$$

$$S_I(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \cos[\Theta(k)], \quad (11.115)$$

$$S_Q(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \sin[\Theta(k)], \quad (11.116)$$

$$\begin{aligned} \vec{s}(t) &= [S_I(t) + jS_Q(t)] e^{j\omega_0 t} = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \left\{ \begin{array}{l} \cos[\omega_0 t + \Theta(k)] + \\ + j \sin[\omega_0 t + \Theta(k)] \end{array} \right\} = \\ &= A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) e^{j\Theta(k)} e^{j\omega_0 t}. \end{aligned} \quad (11.117)$$

Комплексная огибающая сигнала с многопозиционной фазовой манипуляцией полностью определяется его квадратурной и синфазной компонентами и представляется следующим выражением:

$$\vec{S}(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) e^{j\Theta(k)} = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) (\cos[\Theta(k)] + j \sin[\Theta(k)]). \quad (11.118)$$

Векторная диаграмма сигнала («созвездие») с восьмипозиционной фазовой манипуляцией представлена на рис. 11.13. Информационные (модулирующие символы) Θ_m условно закодированы двоичным трех-разрядным кодом Грея для уменьшения вероятности ошибочного приема P_e^b . Поскольку несущие сигналы различных модуляционных символов являются коррелированными, наиболее вероятными являются ошибки восприятия соседних информационных символов, например 111 и 110 или 111 и 101. Приведенная схема кодирования символов обеспечивает искажение только одного информационного элемента (бита) при подобных наиболее вероятных ошибках.

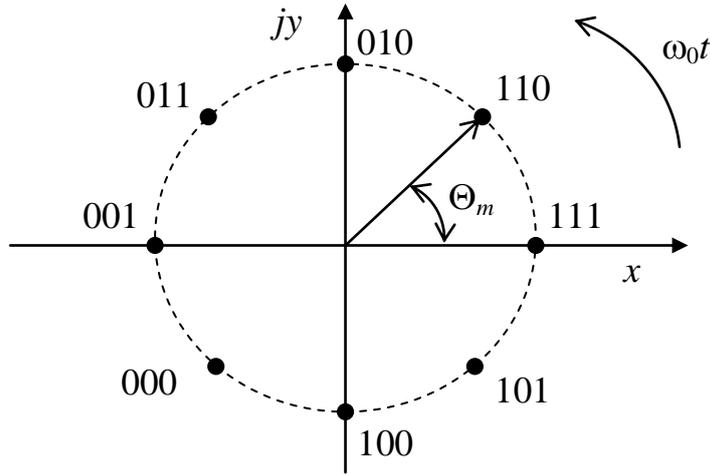


Рис. 11.13. Векторная диаграмма сигнала с 8-PSK

При квазиоптимальном приеме решение принимается на основании наблюдения входной смеси сигнала и шума на интервале времени, соответствующем длительности τ_c , с учетом этого для интервала наблюдения на k -м интервале модуляции комплексная огибающая может быть переписана в следующем виде:

$$\vec{S}(t) = A_0 g_F(t - k\tau_c) e^{j\Theta(k)} = A_0 g_F(t - k\tau_c) [\cos(\Theta(k)) + j \sin(\Theta(k))], \quad (11.119)$$

а соответствующий ей радиосигнал:

$$s(t) = A_0 g_F(t - k\tau_c) [\cos(\Theta(k)) \cos(\omega_0 t) - \sin(\Theta(k)) \sin(\omega_0 t)]. \quad (11.120)$$

Поскольку параметр $\Theta(k)$ может принимать одно из возможных значений $\Theta_m = \frac{2\pi(m-1)}{M}$, при демодуляции необходимо вычислить M значений функции (функционала) правдоподобия

$$P(\vec{X} | m) = \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_c} [\vec{X}(t) - \vec{S}^m(t)] \times [\vec{X}(t) - \vec{S}^m(t)]^* dt \right\} =$$

$$= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_c} [\vec{X}(t) - A_0 g_F(t - k\tau_c) [\cos(\Theta_m) + j \sin(\Theta_m)]] \times \right.$$

$$\left. \times [\vec{X}(t) - A_0 g_F(t - k\tau_c) [\cos(\Theta_m) + j \sin(\Theta_m)]]^* dt \right\}, \quad (11.121)$$

где $\vec{X}(t)$ – комплексная огибающая входной смеси полезного сигнала и шума.

Решение о значении передаваемого информационного символа $\Theta(k)$ на k -м шаге модуляции формируется на основании сравнения (нахождения максимального из рассчитанных значений) полученных величин функционала (функции для дискретного времени наблюдения) правдоподобия: $\Theta^*(k) = \Theta_m \Leftrightarrow \max \{P(\bar{X}|k), k = 1, 2, \dots, M\}$.

Правило принятия решения о значении передаваемого информационного символа может быть получено и без непосредственного формирования (вычисления) значений функционала правдоподобия. Для этого формализуем процедуру поиска максимума $\max\{P(\bar{X}|k), k = 1, 2, \dots, M\}$ следующим образом:

$$\left. \frac{\partial}{\partial \Theta} P(\bar{X}|\Theta) = \frac{\partial}{\partial \Theta} \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_c} \left([\bar{X}(t) - \bar{S}(t, \Theta)] \times \right. \right. \right. \Bigg|_{\Theta = \Theta^*} = 0. \quad (11.122)$$

Тогда правило принятия решения определится процедурой поиска такого значения Θ_m информационного параметра сигнала, которое наиболее «близко» к полученной оценке Θ^* или $\Theta^*(k) = \min \{|\Theta^* - \Theta_m|, m = 1, \dots, M\}$.

Поиск точки экстремума функционала правдоподобия

$$\left. \frac{\partial}{\partial \Theta} \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_c} \left([\bar{X}(t) - \bar{S}(t, \Theta)] \times \right. \right. \right. \Bigg|_{\Theta = \Theta^*} = 0$$

равноценен поиску точки экстремума функционала следующего вида

$$\left. \frac{\partial}{\partial \Theta} \int_0^{\tau_c} \left([\bar{X}(t) - \bar{S}(t, \Theta)] \times [\bar{X}(t) - \bar{S}(t, \Theta)]^* \right) dt \right|_{\Theta = \Theta^*} = 0.$$

Представим полезный сигнал в следующем виде: $\bar{S}(t, \Theta) = |\bar{S}(t, \Theta)| e^{j\Theta}$, учитывая, что при рассматриваемом виде модуляции $|\bar{S}(t, \Theta)| = |\bar{S}(t)|$ не зависит от передаваемых информационных символов, имеем

$$\left. \int_0^{\tau_c} \left(\bar{X}(t) e^{-j\Theta} - \bar{X}^*(t) e^{j\Theta} \right) dt \right|_{\Theta = \Theta^*} = 0.$$

Поскольку $\vec{X}(t) = X_I + jX_Q$, окончательно получим

$$\cos(\Theta^*) \int_0^{\tau_c} X_Q(t) dt = \sin(\Theta^*) \int_0^{\tau_c} X_I(t) dt$$

или

$$\Theta^* = \operatorname{arctg} \left(\frac{\int_0^{\tau_c} X_Q(t) dt}{\int_0^{\tau_c} X_I(t) dt} \right). \quad (11.123)$$

Полученное выражение определяет принцип работы квадратурного фазового детектора.

Точность восстановления опорного колебания, используемого при квадратурном преобразовании частоты для получения комплексной огибающей входной смеси полезного сигнала и помехи, оказывает существенное влияние на правильность принятия решений. При демодуляции необходимо выполнение условия $\varphi(t) \rightarrow 0$ (условие когерентной демодуляции).

Определим элементы нормированной корреляционной матрицы ρ несущих информационных символов сигналов.

$$\rho_{km} = \frac{1}{E_c} \int_0^{\tau_c} s^k(t) s^m(t) dt = \frac{1}{E_c} \int_0^{\tau_c} s^k(t) s^m(t) dt; \quad (11.124)$$

$$\begin{aligned} \rho_{km} &= \frac{1}{E_c} \int_0^{\tau_c} [A_0 \cos(\Theta_k) \cos(\omega_0 t) - A_0 \sin(\Theta_k) \sin(\omega_0 t)] \times \\ &\times [A_0 \cos(\Theta_m) \cos(\omega_0 t) + A_0 \sin(\Theta_m) \sin(\omega_0 t)] dt = \\ &= \frac{1}{E_c} \frac{A_0^2 \tau_c}{2} \cos(\Theta_k - \Theta_m) = \cos(\Theta_k - \Theta_m), \end{aligned} \quad (11.125)$$

где $E_c = \frac{A_0^2 \tau_c}{2}$ – энергия, приходящаяся на один информационный символ,

$$\rho = \begin{pmatrix} 1 & \cos(\Theta_1 - \Theta_2) & \dots & \cos(\Theta_1 - \Theta_M) \\ \cos(\Theta_2 - \Theta_1) & 1 & \dots & \cos(\Theta_2 - \Theta_M) \\ \vdots & \vdots & \dots & \vdots \\ \cos(\Theta_M - \Theta_1) & \cos(\Theta_M - \Theta_2) & \dots & 1 \end{pmatrix}.$$

Нормированная корреляционная матрица сигнала с М-PSK является сингулярной (вырожденной) с $|\rho| = 0$. Общее выражение для нахождения вероятности ошибки непосредственно применено быть не может.

Для нахождения вероятности ошибки рассмотрим подробнее правило принятия решения о передаваемом символе подробнее. Итак, демодулятор осуществляет поиск максимальной из рассчитанных величин

$$P(\bar{X}|m) = \exp\left\{-\frac{1}{N_0} \int_0^{\tau_c} [\bar{X}(t) - \bar{S}^m(t)] \times [\bar{X}(t) - \bar{S}^m(t)]^* dt\right\}. \quad (11.126)$$

Преобразуем данное выражение.

$$P(\bar{X}|m) = \exp\left\{-\frac{1}{N_0} \int_0^{\tau_c} [\bar{X}(t)\bar{X}^*(t) + \bar{S}^m(t)(\bar{S}^m(t))^* - 2\operatorname{Re}\{\bar{S}^m(t)\bar{X}^*(t)\}] dt\right\}.$$

Поиск максимума $P(\bar{X}|m)$ по m соответствует поиску максимума $\operatorname{Re}\{\bar{S}^m(t)\bar{X}^*(t)\}$, поскольку $\int_0^{\tau_c} \bar{S}^m(t)(\bar{S}^m(t))^* dt = E_c$ и $\int_0^{\tau_c} \bar{X}(t)\bar{X}^*(t) dt$ не зависят от передаваемого информационного символа.

Таким образом, корреляционный демодулятор проецирует принимаемый сигнальный вектор $\bar{X} = [X_I, X_Q]$ на M возможных сигнальных векторов $\bar{S}^m = [S_I^m, S_Q^m]$. Решение принимается в пользу сигнала с наибольшей проекцией.

Корреляционный детектор эквивалентен фазовому детектору, который определяет фазу принимаемого сигнала $\bar{X}(t)$ и выбирает сигнальный вектор $\bar{S}^m(t)$, фаза которого ближе всего к фазе $\bar{X}(t)$.

Подтвердим данное утверждение. Корреляционный демодулятор вычисляет значения функций (функционалов) правдоподобия для всех возможных сигналов, несущих различные информационные символы. Если внимательно присмотреться к структуре функционала, то можно увидеть операцию вычисления среднего (с точностью до постоянного множителя $1/\tau_c$) во времени расстояния между вектором входного процесса $\bar{X}(t)$ и вектором конкретного несущего сигнала $\bar{S}^m(t)$:

$$\int_0^{\tau_c} [\bar{X}(t) - \bar{S}^m(t)] \times [\bar{X}(t) - \bar{S}^m(t)]^* dt = \int_0^{\tau_c} |\bar{X}(t) - \bar{S}^m(t)|^2 dt =$$

$$= \int_0^{\tau_c} \left[\left(\operatorname{Re}\{\bar{X}(t)\} - \operatorname{Re}\{\bar{S}^m(t)\} \right)^2 + \left(\operatorname{Im}\{\bar{X}(t)\} - \operatorname{Im}\{\bar{S}^m(t)\} \right)^2 \right] dt. \quad (11.127)$$

Решение о передаваемом символе принимается путем поиска минимального из всех возможных определенных расстояний. Геометрически это интерпретируется следующим образом (рис. 11.14). Из рисунка видно, что определение минимального расстояния между каждой из точек, представляющих возможные передаваемые полезные сигналы, и точкой, представляющей входную смесь сигнала и шума, эквивалентно определению факта попадания точки, представляющей смесь сигнала и шума, в область принятия того или иного решения. Нахождение точки, представляющей смесь сигнала и шума в области принятия конкретного решения, может быть реализовано путем сравнения фазы вектора входной смеси и фазы несущего сигнала, представляющего конкретный информационный символ. Кроме того, из рисунка видно, что при демодуляции сигнала с M-PSK отсутствует необходимость амплитудного нормирования входной смеси сигнала и шума.

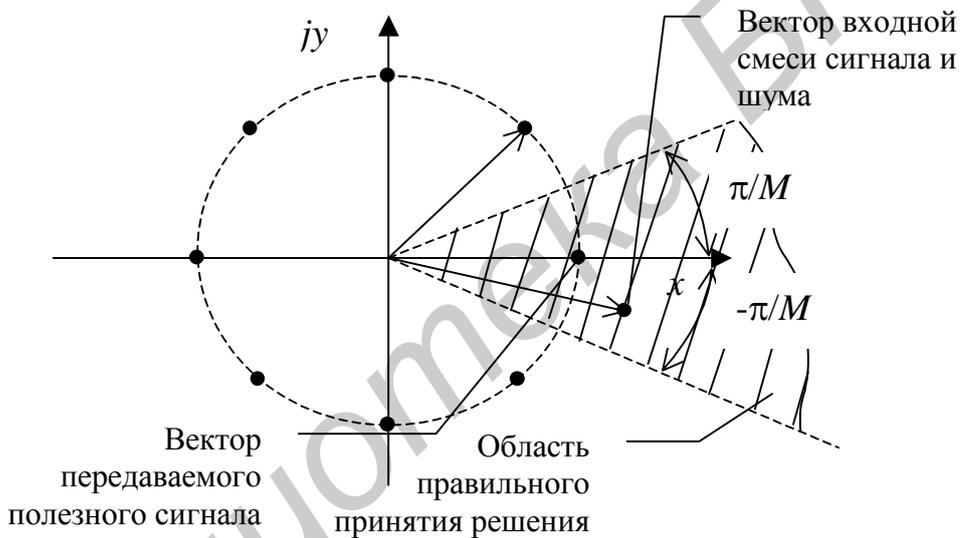


Рис. 11.14. Геометрическая интерпретация принятия решения

Вероятность ошибки при приеме равновероятных информационных символов может быть определена следующим образом:

$$P_e^c = \frac{1}{M} \sum_{k=1}^M P_e(k) = P_e(k) = 1 - \int_{-\pi/M}^{\pi/M} W(\Delta\Theta_x) d(\Delta\Theta_x), \quad (11.128)$$

где $P_e(k)$ – вероятность ошибочного приема k -го символа; $W(\Delta\Theta_x)$ – плотность распределения вероятности разности фаз $\Delta\Theta_x$ входной смеси сигнала и помехи и присутствующего в ней полезного сигнала.

На основании вышеизложенного определим вероятность ошибки при приеме сигнала с M-PSK.

Поскольку фаза входного процесса $\bar{X}(t)$ равна $\Theta_x = \arctg\left(\frac{X_Q}{X_I}\right)$, определим плотность распределения вероятности фазы через статистические характеристики вектора $\bar{X}(t)$.

Рассмотрим случай, когда фаза передаваемого сигнала $\bar{S}^1(t)$ равна $\Theta_1 = 0$. Для упрощения записи будем считать, что формирующий импульс имеет прямоугольную форму $g_F(t) = 0, t \notin [0, \tau_c]$ и $g_F(t) = 1, t \in [0, \tau_c]$. Следовательно, вектор переданного сигнала на интервале k -го шага модуляции равен $\bar{S} = A_0 e^{j\Theta(k)} = \sqrt{E_c} (\cos[\Theta_1] + j \sin[\Theta_1]) = A_0 + j0$.

Здесь для получения общего энергетического соотношения используется амплитудный множитель комплексной огибающей $\sqrt{E_c} = \sqrt{\frac{A_0^2}{2} \tau_c}$, учитывающий энергетический параметр несущего колебания A_0 и учитывается операция интегрирования на интервале времени τ_c . Вектор принимаемого сигнала имеет следующие компоненты:

$$\bar{X} = X_I + jX_Q = (\sqrt{E_c} + n_I) + jn_Q.$$

Поскольку квадратурные составляющие шума n_I и n_Q являются совместно гауссовскими случайными величинами с нулевыми средними, то квадратурные компоненты принимаемого процесса X_I и X_Q тоже являются совместно гауссовскими случайными величинами со средними значениями $\langle X_I \rangle = \sqrt{E_c}$, $\langle X_Q \rangle = 0$ и дисперсиями $\sigma_I^2 = \sigma_Q^2 = \sigma^2 = \frac{N_0}{2}$.

Следовательно, совместная плотность распределения вероятности квадратурных компонент входного процесса будет равна

$$W(X_I, X_Q) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{(X_I - \sqrt{E_c})^2 + X_Q^2}{2\sigma^2}\right\}. \quad (11.129)$$

Используя правила преобразования статистических характеристик, можно перейти к плотности распределения вероятности модуля

$$V_x = \sqrt{X_I^2 + X_Q^2} \text{ и фазы вектора входного воздействия } \Theta_x = \arctg\left(\frac{X_Q}{X_I}\right).$$

$$W(V_x, \Theta_x) = \frac{V_r}{2\pi\sigma^2} \exp\left\{-\frac{V_x^2 + E_c - 2\sqrt{E_c}V_x \cos(\Theta_x)}{2\sigma^2}\right\}. \quad (11.130)$$

Проинтегрировав по области возможных значений модуля вектора входного воздействия $[0, \infty)$, получим следующую плотность распределения фазы вектора входного воздействия:

$$W(\Theta_x) = \int_0^{\infty} W(V_x, \Theta_x) dV_x = \frac{1}{2\pi} e^{-\frac{E_c}{N_0} \sin^2(\Theta_x)} \int_0^{\infty} V_x e^{-\left(V_x - \sqrt{\frac{E_c}{N_0}}\right)^2 / 2} dV_x. \quad (11.131)$$

На рис. 11.15 изображены кривые плотности распределения вероятности фазы вектора входного сигнала (плотности распределения вероятности $W(\Delta\Theta_x)$, разности фаз $\Delta\Theta_x$ входной смеси сигнала и помехи и присутствующего в ней полезного сигнала) для различных отношений E_c/N_0 .

Как видно из рисунка, плотность распределения вероятности «концентрируется» около истинного значения $\Theta_1 = 0$ по мере увеличения отношения сигнал/шум E_c/N_0 .

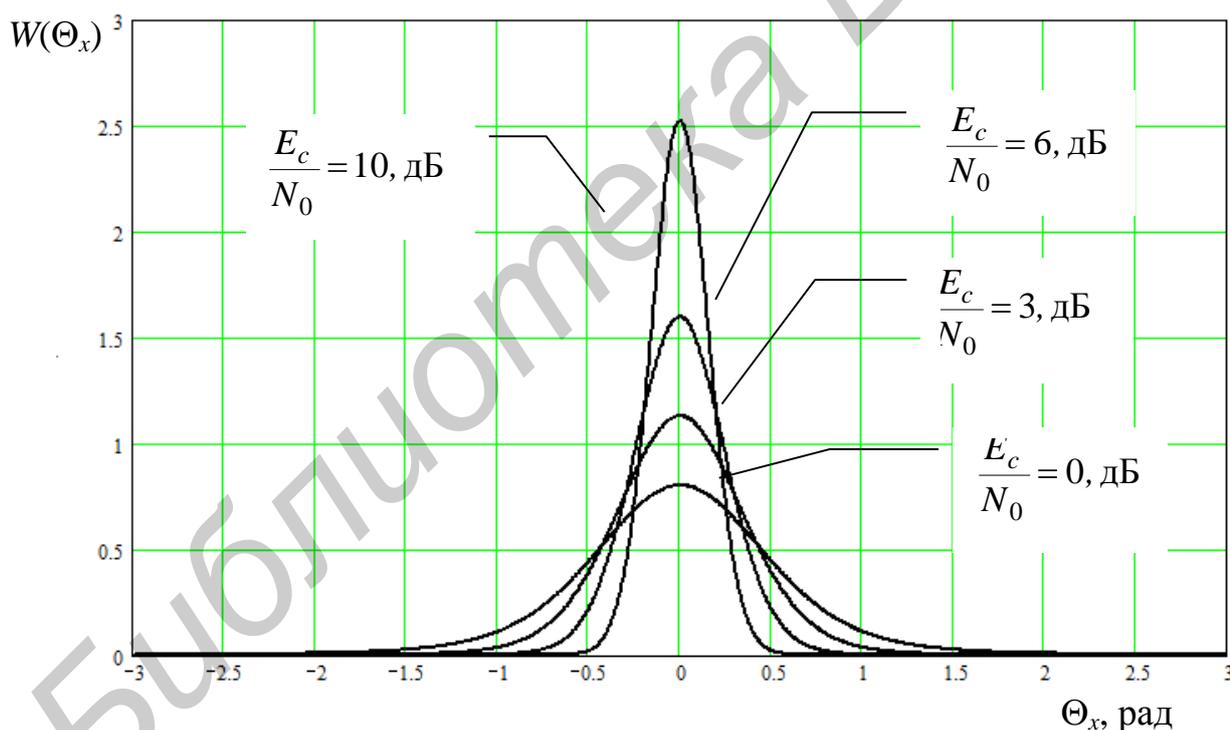


Рис. 11.15. Плотность распределения вероятности фазы вектора входного воздействия

Вероятность ошибочного приема символа находится в соответствии с приведенным выше выражением:

$$P_e^c = 1 - \int_{-\pi/M}^{\pi/M} W(\Delta\Theta_x) d(\Delta\Theta_x) = 1 - \int_{-\pi/M}^{\pi/M} W(\Theta_x) d(\Theta_x), \quad (1.132)$$

а верхняя граница (все возможные ошибки при приеме информационного символа равновероятны) вероятности ошибочного приема конкретного элемента (бита) передаваемого сообщения будет определяться следующим образом:

$$P_e^b = \frac{2^{k-1}}{2^k - 1} P_e^c, \quad k = \log_2(M). \quad (11.133)$$

На рис. 11.16 приведены зависимости вероятности ошибочного приема передаваемых бит информации от отношения энергии, приходящейся на один передаваемый бит, к спектральной плотности мощности шума E_b/N_0 для многопозиционной фазовой манипуляции.

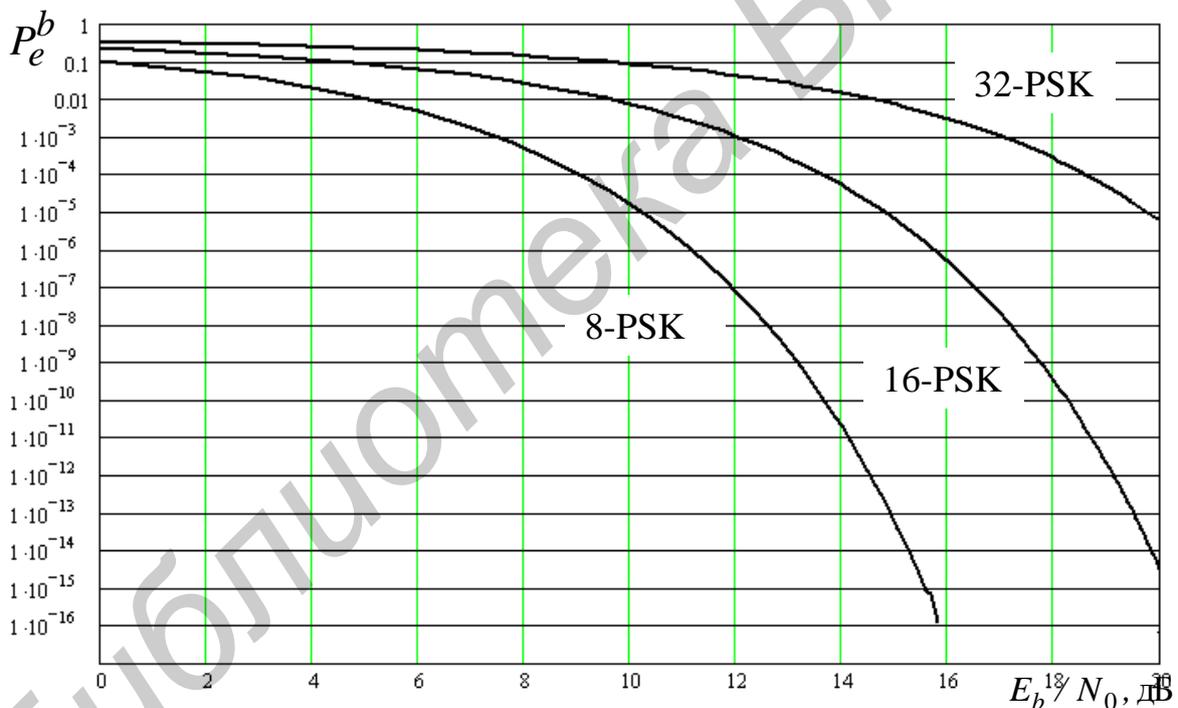


Рис. 11.16. Вероятность ошибочного приема при многопозиционной фазовой манипуляции

В реальной ситуации при достаточно высоких отношениях E_b/N_0 наиболее вероятными являются ошибки, связанные с выбором соседних значений фазы Θ_{i-1} и Θ_{i+1} относительно фазы передаваемого сигнала Θ_i . В этом случае при использовании кода Грея для кодирования состояний фазы Θ_m данные ошибки приводят к искажению только одного из k элементов

(бит), определяющих передаваемый информационный символ, и вероятность ошибки конкретного бита может быть примерно оценена следующим образом:

$$P_e^b = \frac{1}{k} P_e^c, \quad k = \log_2(M). \quad (11.134)$$

11.7.1.3. Квадратурная фазовая манипуляция и квадратурная фазовая манипуляция со сдвигом (QPSK и OQPSK)

Квадратурная фазовая манипуляция и квадратурная фазовая манипуляция со сдвигом являются частными случаями 4-позиционной фазовой манипуляции. Это вид дискретной модуляции, при которой радиосигнал в общем виде может быть представлен следующим образом.

Сигнал с QPSK:

$$s(t) = \frac{A_0}{2} g_F(t - k\tau_c) \left\{ \sum_{k=0}^{\infty} C_k^I \cos[\omega_0 t + \varphi_0] - \sum_{k=0}^{\infty} C_k^Q \sin[\omega_0 t + \varphi_0] \right\}, \quad (11.135)$$

сигнал с OQPSK:

$$s(t) = \frac{A_0}{\sqrt{2}} \left\{ \begin{array}{l} \sum_{k=0}^{\infty} C_k^I g_F(t - k\tau_c) \cos[\omega_0 t + \varphi_0] - \\ - \sum_{k=0}^{\infty} C_k^Q g_F(t - (k - \frac{1}{2})\tau_c) \sin[\omega_0 t + \varphi_0] \end{array} \right\}, \quad (11.136)$$

где $A_0/\sqrt{2}$, ω_0 , φ_0 – амплитуда, центральная частота и начальная фаза квадратурных несущих колебаний; $C_k^I = \pm 1$ и $C_k^Q = \pm 1$ – синфазная и квадратурная последовательности, соответствующие передаваемым информационным символам «0» и «1» (обычно +1 соответствует «0», а –1 представляет «1»); $g_F(t) = 0 \quad \forall t \notin [0, \tau_c]$ – формирующий импульс; τ_c – длительность символа модуляции; k – индекс текущего шага модуляции.

Пример временных функций, соответствующих радиосигналам с QPSK и OQPSK при прямоугольной форме формирующего импульса, приведен на рис. 11.17 и 11.18. На данных рисунках также приведены временные диаграммы, соответствующие основным компонентам рассматриваемых сигналов и поясняющие принцип их формирования.

Как видно из рисунков, исходный поток данных $D(t)$ разбивается на два параллельных потока, при этом синфазный поток $D_I(t)$ содержит условно нечетные биты исходного потока данных, а квадратурный $D_Q(t)$ – условно четные. Длительность элементов синфазного и квадратурного потоков со-

ответствует длительности символа модуляции $\tau_c = 2\tau_b$, равной удвоенной длительности информационного элемента (бита) исходного потока $D(t)$. Результирующий радиосигнал получается путем суммирования квадратурных компонент, модулированных потоками $D_I(t)$ и $D_Q(t)$, в соответствии с $\pm\pi/2$ -BPSK схемой модуляции.

Отличие метода модуляции OQPSK от QPSK заключается во временном сдвиге на τ_b одного модулирующего потока, например $D_Q(t)$, относительно другого. Данный сдвиг делает невозможным одновременную смену состояний модулирующих потоков и, как следствие, исключает скачки фазы в результирующем радиосигнале на $\pm\pi$, что уменьшает паразитную амплитудную модуляцию и межсимвольные переходные искажения при прохождении радиосигнала через избирательные цепи.

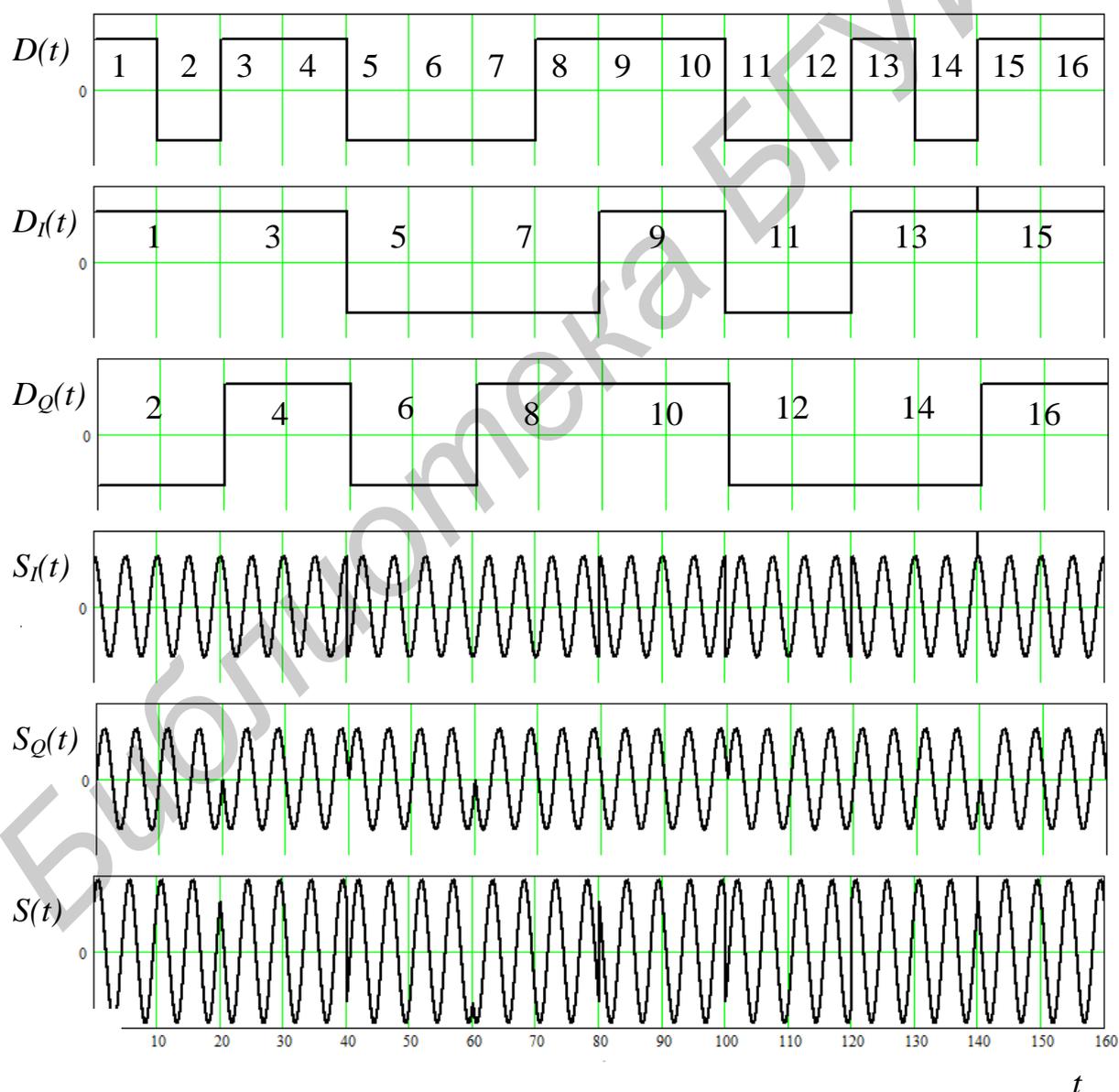


Рис. 11.17. Форма радиосигнала с QPSK:

$D(t)$ – поток передаваемых данных;

$D_I(t)$ – синфазный поток передаваемых данных;
 $D_Q(t)$ – квадратурный поток передаваемых данных;
 $S_I(t)$ – синфазная компонента радиосигнала;
 $S_Q(t)$ – квадратурная компонента радиосигнала;
 $S(t)$ – радиосигнал с QPSK

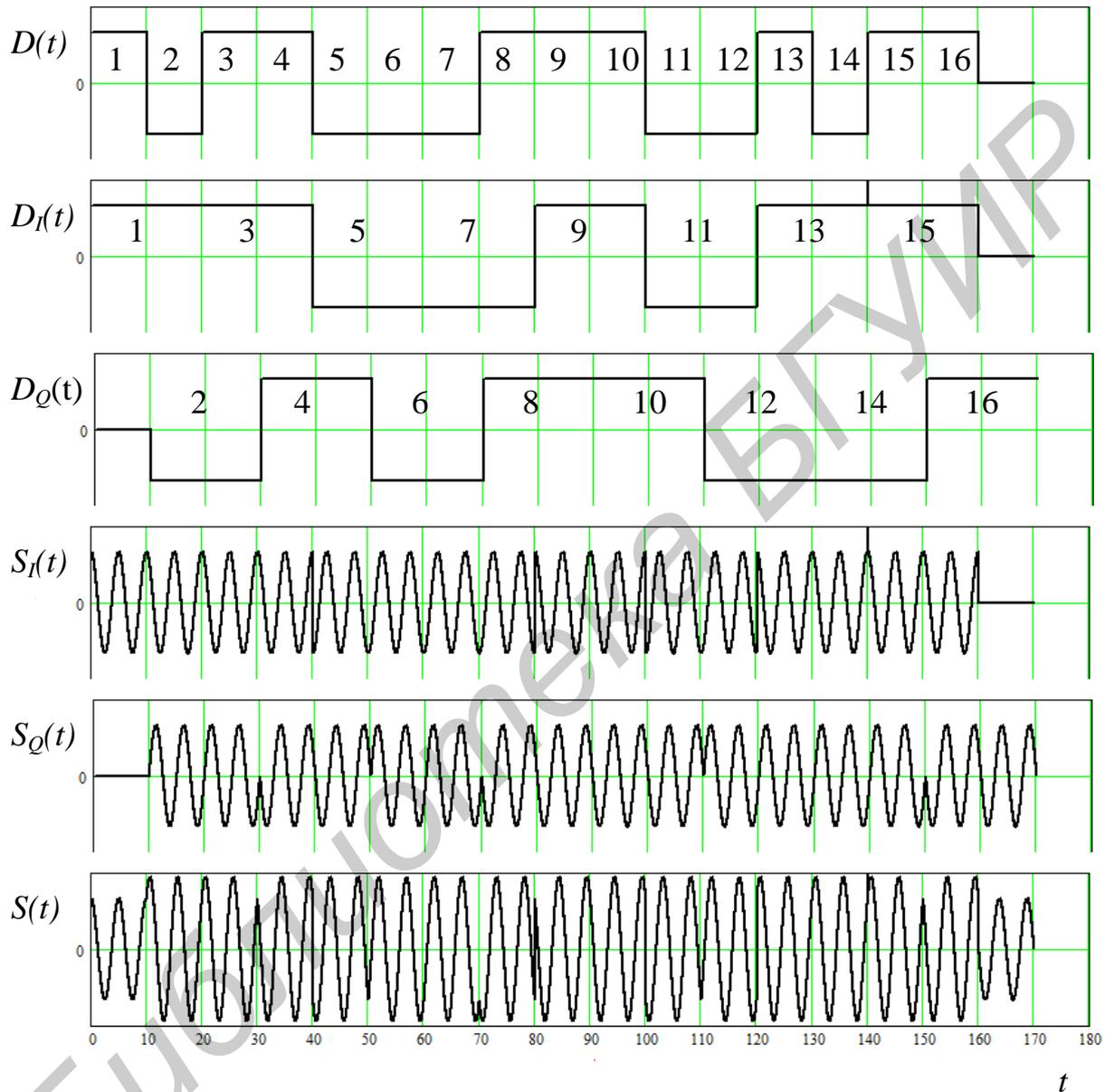


Рис. 11.18. Форма радиосигнала с OQPSK:

$D(t)$ – поток передаваемых данных;
 $D_I(t)$ – синфазный поток передаваемых данных;
 $D_Q(t)$ – квадратурный поток передаваемых данных;
 $S_I(t)$ – синфазная компонента радиосигнала;
 $S_Q(t)$ – квадратурная компонента радиосигнала;
 $S(t)$ – радиосигнал с QPSK

Комплексное представление радиосигналов с QPSK и OQPSK может быть определено следующим образом.

Для сигналов с QPSK:

$$s_{qpsk}(t) = \frac{A_0}{\sqrt{2}} \left\{ \begin{array}{l} \sum_{k=0}^{\infty} C_k^I g_F(t - k\tau_b) \cos[\omega_0 t] - \\ - \sum_{k=0}^{\infty} C_k^Q g_F(t - k\tau_b) \sin[\omega_0 t] \end{array} \right\} =$$

$$= S_I^{qpsk}(t) \cos[\omega_0 t] - S_Q^{qpsk}(t) \sin[\omega_0 t] \quad (11.137)$$

$$S_I^{qpsk}(t) = \frac{A_0}{\sqrt{2}} \sum_{k=0}^{\infty} g_F(t - k\tau_b) C_k^I \quad (11.138)$$

$$S_Q^{qpsk}(t) = \frac{A_0}{\sqrt{2}} \sum_{k=0}^{\infty} g_F(t - k\tau_b) C_k^Q \quad (11.139)$$

$$\bar{s}_{qpsk}(t) = [S_I(t) + jS_Q(t)] e^{j\omega_0 t} = \frac{A_0}{\sqrt{2}} \sum_{k=0}^{\infty} g_F(t - k\tau_b) (C_k^I + jC_k^Q) e^{j\omega_0 t} \quad (11.140)$$

Для сигналов с OQPSK:

$$s_{oqpsk}(t) = \frac{A_0}{\sqrt{2}} \left\{ \begin{array}{l} \sum_{k=0}^{\infty} C_k^I g_F(t - k\tau_b) \cos[\omega_0 t] - \\ - \sum_{k=0}^{\infty} C_k^Q g_F(t - (k - \frac{1}{2})\tau_b) \sin[\omega_0 t] \end{array} \right\} =$$

$$= S_I^{oqpsk}(t) \cos[\omega_0 t] - S_Q^{oqpsk}(t) \sin[\omega_0 t] \quad (11.141)$$

$$S_I^{oqpsk}(t) = \frac{A_0}{\sqrt{2}} \sum_{k=0}^{\infty} g_F(t - k\tau_b) C_k^I \quad (11.142)$$

$$S_Q^{oqpsk}(t) = \frac{A_0}{\sqrt{2}} \sum_{k=0}^{\infty} g_F(t - (k - \frac{1}{2})\tau_b) C_k^Q \quad (11.143)$$

$$\bar{s}_{oqpsk}(t) = [S_I(t) + jS_Q(t)] e^{j\omega_0 t} = \frac{A_0}{\sqrt{2}} \sum_{k=0}^{\infty} \left(\begin{array}{l} C_k^I g_F(t - k\tau_b) + \\ + jC_k^Q g_F(t - (k - \frac{1}{2})\tau_b) \end{array} \right) e^{j\omega_0 t} \quad (11.144)$$

Комплексная огибающая сигнала полностью определяется его квадратурной и синфазной компонентами и представляется следующими выражениями:

$$\vec{S}_{qpsk}(t) = \frac{A_0}{\sqrt{2}} \sum_{k=0}^{\infty} g_F(t - k\tau_b) (C_k^I + jC_k^Q), \quad (11.145)$$

$$\vec{S}_{oqpsk}(t) = \frac{A_0}{\sqrt{2}} \sum_{k=0}^{\infty} \left(C_k^I g_F(t - k\tau_b) + jC_k^Q g_F(t - (k - \frac{1}{2})\tau_b) \right). \quad (11.146)$$

Векторная диаграмма сигнала с квадратурной фазовой манипуляцией представлена на рис.11.19.

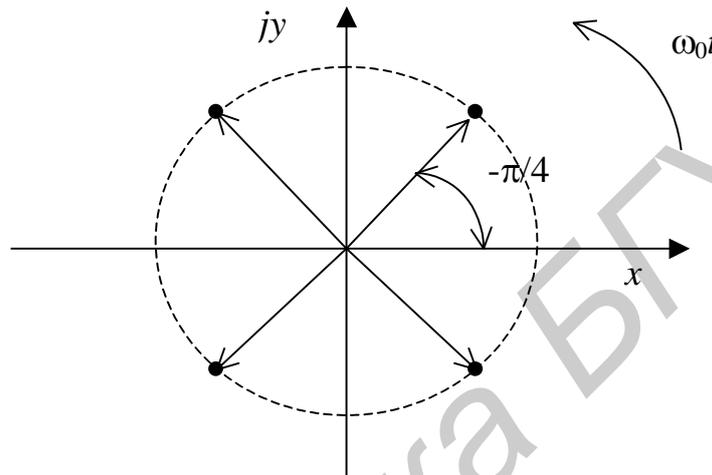


Рис. 11.19. Векторная диаграмма сигнала с QPSK/OQPSK

Поскольку сигналы с QPSK и OQPSK по своей сути являются сигналами с 4-позиционной фазовой манипуляцией, то при их демодуляции возможно использование решающего правила, описанного выше в подразделе, посвященном M-PSK методу модуляции.

Кроме этого, поскольку сигналы с QPSK и OQPSK являются линейной комбинацией биортогональных сигналов с $\pm\pi/2$ -BPSK, демодуляция может осуществляться с параллельным использованием двух демодуляторов сигналов с $\pm\pi/2$ -BPSK со сдвинутыми на $\pi/2$ опорными колебаниями.

При демодуляции сигналов с BPSK особое внимание следует обратить на точность восстановления опорного колебания. Для правильного принятия решений при демодуляции необходимо, чтобы фазовая ошибка восстановления опорного колебания стремилась к нулю $\varphi(t) \rightarrow 0$ (условие когерентной демодуляции).

Нормированная корреляционная матрица ρ сигналов с QPSK и OQPSK является сингулярной:

$$\rho = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}.$$

Общее выражение для нахождения вероятности ошибки непосредственно применено быть не может. Но поскольку сигналы с QPSK и OQPSK являются линейной комбинацией биортогональных сигналов с $\pm\pi/2$ -BPSK, то при параллельной демодуляции данных сигналов отсутствуют переходные помехи (взаимное влияние компонент с $\pm\pi/2$ -BPSK друг на друга). При этом вероятность одновременно правильного приема в обоих каналах (вероятность правильного приема символа, состоящего из 2-х элементов $C_k^I = \pm 1$ и $C_k^Q = \pm 1$) будет равна

$$P_{np}^c = (1 - P_e^{b \text{ bpsk}})^2 = \left[1 - \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \right]^2. \quad (11.147)$$

Соответственно вероятность искажения 2-битового символа будет определяться как $P_e^c = 1 - P_{np}^c$.

Вероятность ошибки при принятии конкретного бита равно вероятности принятия одного из трех возможных ошибочных решений (своеобразная верхняя граница):

$$\begin{aligned} P_e^b &= \frac{2^{2-1}}{2^2 - 1} P_e^c = \frac{2}{3} \left[1 - (1 - P_e^{b \text{ bpsk}})^2 \right] = \\ &= \frac{2}{3} \left[1 - \left(1 - 2P_e^{b \text{ bpsk}} + (P_e^{b \text{ bpsk}})^2 \right) \right] = \frac{2}{3} 2P_e^{b \text{ bpsk}} \left(1 - \frac{1}{2} P_e^{b \text{ bpsk}} \right) = \\ &= \frac{4}{3} \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \left(1 - \frac{1}{4} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \right) = \\ &= \frac{2}{3} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \left(1 - \frac{1}{4} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \right). \end{aligned} \quad (11.148)$$

Если учесть, что вероятность возникновения ошибки сразу в двух каналах демодулятора значительно меньше вероятности ошибки только в одном канале, то вероятность ошибки конкретного бита примерно равна

$$P_e^b \approx \frac{1}{2} P_e^c = \frac{1}{2} \left[1 - \left(1 - P_e^{bpsk} \right)^2 \right] = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \left(1 - \frac{1}{4} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \right), \quad (11.149)$$

что практически совпадает с вероятностью ошибки при приеме сигнала с $\pm\pi/2$ -BPSK (особенно при $E_b/N_0 \gg 1$).

На рис. 11.20 приведены зависимости вероятности ошибки P_e^b и P_e^c при демодуляции сигналов с QPSK и OQPSK от отношения E_b/N_0 , для сравнения помехоустойчивости на рисунке приведена зависимость P_e^b при демодуляции сигнала с $\pm\pi/2$ -BPSK.

Необходимо отметить, что при одинаковых скорости модуляции и мощности модулированного сигнала энергии, приходящиеся на один бит передаваемой информации, для сигналов с $\pm\pi/2$ -BPSK и QPSK совпадают.

11.7.1.4. Явление обратной работы. Относительная (дифференциальная) фазовая манипуляция (DPSK, DQPSK)

Демодуляция сигналов PSK предполагает, что демодулятор располагает совершенной оценкой фазы несущей. На практике фаза несущей, как правило, определяется по принятому сигналу путем использования нелинейных операций, которые приводят к неоднозначности фазы. Так при бинарной фазовой манипуляции сигнал часто подвергается возведению в квадрат для снятия модуляции, затем образованный сигнал с удвоенной частотой фильтруется и делится по частоте на 2, для того чтобы получить оценку частоты несущей и фазы $\varphi(t)$. Эти операции приводят к неоднозначности фазы восстановленного несущего колебания на π (180°).

Аналогично при приеме квадратурной фазовой манипуляции принимаемый сигнал возводится в четвертую степень, чтобы снять цифровую модуляцию, а затем четвертая гармоника частоты несущей фильтруется и делится на 4 для того, чтобы выделить компоненту несущей. Эти операции приводят к появлению компоненты с частотой несущей, содержащей оценку фазы несущей $\varphi(t)$ с точностью до постоянного слагаемого, равного $\pm 90^\circ$ или 180° .

Аналогичные явления проявляются и при восстановлении несущего колебания при приеме сигналов с многопозиционной фазовой манипуляцией. Следовательно, мы не имеем точной оценки фазы несущей в демодуляторе.

Подобные ошибки при восстановлении фазы несущего колебания приводят к ошибкам, проявляющимся в инверсии передаваемых символов на выходе демодулятора (двоичная «1» воспринимается как двоичный «0» и наоборот). Данное явление получило наименование «обратная работа». Опасность обратной работы заключается в том, что под действием помех может происходить неконтролируемый самопроизвольный перескок фазы восстановленного несущего колебания в пределах неоднозначности ее оцен-

ки. Например, при приеме сигнала с $\pm\pi/2$ -BPSK фаза восстановленного несущего колебания может произвольно перескакивать со значения $\varphi(t)$ на значение $\varphi(t) + \pi$ и обратно, что является причиной неоднозначности восприятия демодулированных символов.

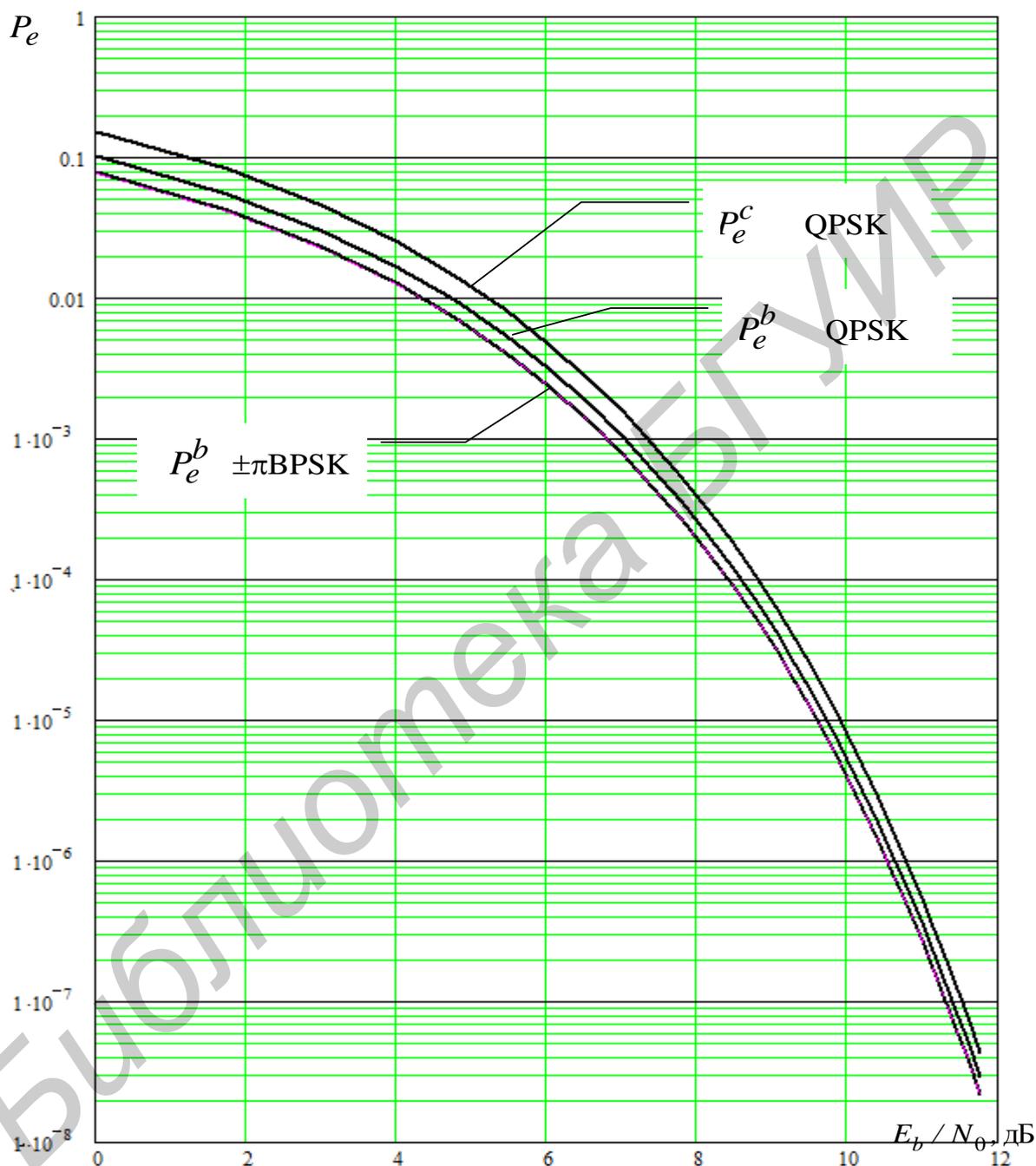


Рис. 11.20. Характеристики помехоустойчивости сигналов с QPSK и OQPSK

Проблема влияния неоднозначности фазы, возникающей при оценке фазы несущей $\varphi(t)$, обусловлена использованием так называемых методов модуляции, использующих абсолютные значения фазы для кодирования пе-

редаваемых символов и может быть преодолена путём использования дифференциальной PSK (DPSK) вместо абсолютной PSK. При дифференциальной (относительной) фазовой манипуляции кодирование информации осуществляется посредством определения фазового набега (скачка) в зависимости от передаваемого символа на текущем шаге модуляции относительно предыдущего шага модуляции, а не самой абсолютной фазы, как при обычной PSK. Например, при бинарной PSK информационный символ «1» передается всегда с дополнительным сдвигом, равным π , относительно предыдущего значения фазы (вне зависимости от значения предыдущего символа), символ «0» передается без сдвига фазы, т.е. при передаче только одних «1» происходит непрерывное изменение фазы на π от символа к символу, напротив, при передаче одних «0» никаких скачков фазы сигнала при переходах от символа к символу происходить не будет.

Аналогично могут быть закодированы символы при передаче с использованием квадратурной фазовой манипуляции. Например, паре символов «00» соответствует дополнительный скачок фазы относительно предыдущего шага модуляции, равный 0; паре «01» – $+\pi/2$; паре «10» – минус $\pi/2$ и паре «11» – π . Следует заметить, что приведенные схемы кодирования передаваемой информации не являются окончательными и приведены только для иллюстрации метода дифференциальной манипуляции.

Сигнал с дифференциально-кодированной фазовой манипуляцией (DPSK) позволяет использовать демодуляцию, которая не требует оценки фазы несущей. Вместо этого принимаемый сигнал на заданном сигнальном интервале (шаге модуляции) сравнивается по фазе с принятым сигналом на предыдущем сигнальном интервале.

Рассмотрим подробнее общий случай многопозиционной дифференциальной фазовой манипуляции. В этом случае радиосигнал в общем виде может быть представлен следующим образом:

$$s(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \cos(\omega_0 t + \varphi_k), \quad (11.150)$$

$$\varphi_k = \varphi_{k-1} + \Delta\Theta,$$

где $\Delta\Theta$ – скачок фазы, принимающий одно из возможных из M значений $\Delta\Theta_m$ в соответствии с передаваемыми символами (обычно

$\Delta\Theta_m = \frac{2(m-1)\pi}{M}$). На рис. 11.21 приведен пример временной функции, соответствующей радиосигналу с DPSK для случая $M = 2$.

Комплексное представление радиосигналов с DPSK может быть записано следующим образом:

$$s(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \{ \cos[\varphi_k] \cos[\omega_0 t] - \sin[\varphi_k] \sin[\omega_0 t] \} =$$

$$= S_I(t) \cos[\omega_0 t] - S_Q(t) \sin[\omega_0 t] \quad (11.151)$$

$$S_I(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \cos[\varphi_k] \quad (11.152)$$

$$S_Q(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \sin[\varphi_k] \quad (11.153)$$

$$\bar{s}(t) = (S_I(t) + jS_Q(t))e^{j\omega_0 t} = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) (\cos(\varphi_k) + j \sin(\varphi_k)) e^{j\omega_0 t} \quad (11.154)$$

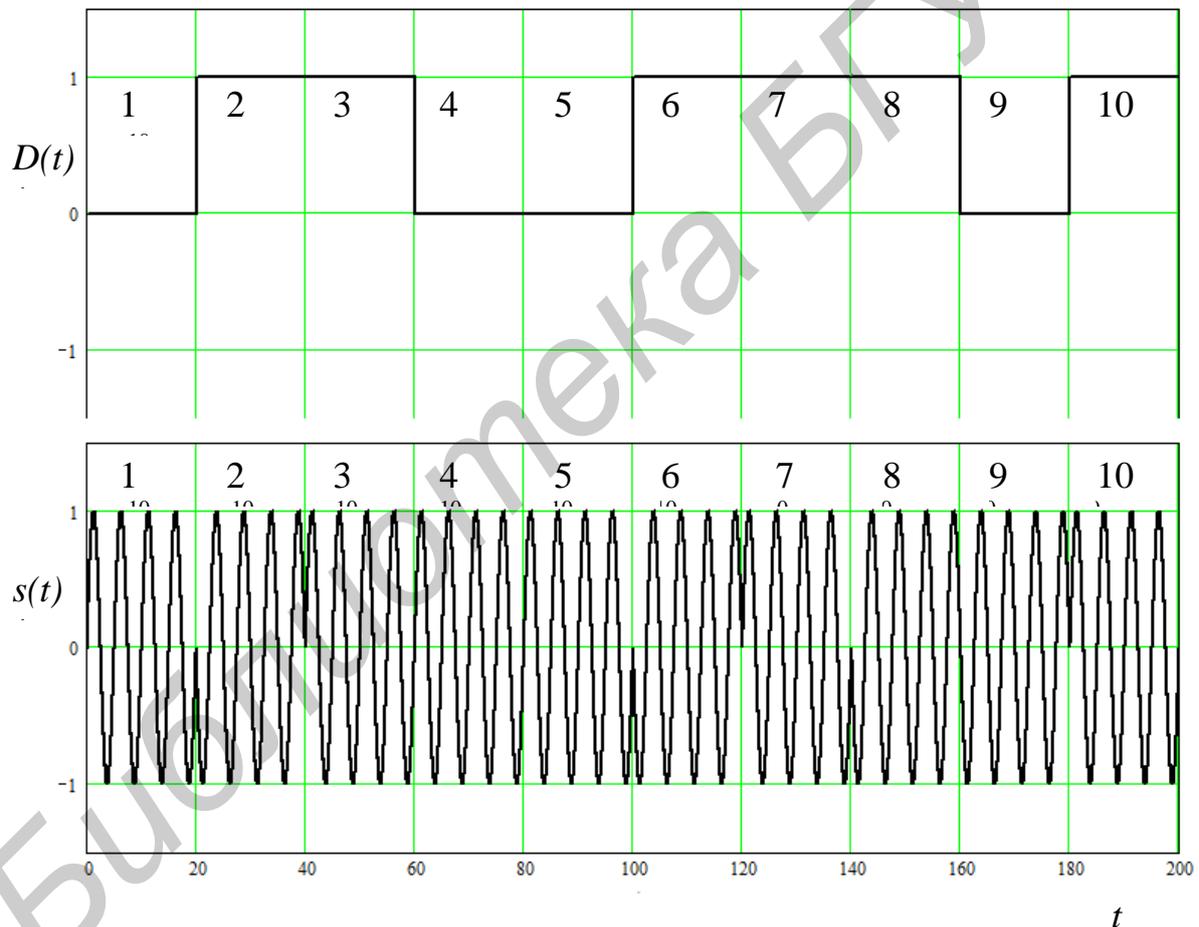


Рис. 11.21. Форма сигнала с DBPSK (2-DPSK):

$D(t)$ – поток передаваемых данных;
 $S(t)$ – радиосигнал с DBPSK

Комплексная огибающая сигнала представляется следующим выражением:

$$\vec{S}(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) (\cos(\varphi_k) + j \sin(\varphi_k)) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) e^{j\varphi_k}. \quad (11.155)$$

Как известно, оптимальный демодулятор осуществляет поиск максимальной из рассчитанных величин (значений функции или функционала правдоподобия):

$$P(\bar{X}|m) = \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_c} [\bar{X}(t) - \vec{S}^m(t)] \times [\bar{X}(t) - \vec{S}^m(t)]^* dt \right\}.$$

Сложность вычисления функционала правдоподобия для случая сигналов с DPSK заключается в том, что на каждом шаге модуляции несущие сигналы $\vec{S}^m(t)$ представляют в общем случае разные передаваемые информационные символы. Поиск максимального из M возможных значений функционала позволяет определить только факт наличия во входной смеси сигнала и шума конкретного несущего сигнала $\vec{S}^m(t)$, но никоим образом не позволяет идентифицировать передаваемый символ. Информация о передаваемом информационном символе может быть получена в результате сравнения идентифицированного несущего сигнала на предыдущем шаге модуляции $\vec{S}_{k-1}^{l*}(t)$ с идентифицированным несущим сигналом на текущем шаге модуляции $\vec{S}_k^{m*}(t)$.

Поскольку комплексная огибающая несущего сигнала на конкретном шаге модуляции равна

$$\vec{S}_k(t) = A_0 e^{j\varphi_k} = A_0 e^{j(\varphi_{k-1} + \Delta\Theta)}, \quad (11.156)$$

то для соседних шагов модуляции мы имеем

$$\vec{S}_{k-1}(t) = A_0 e^{j\varphi_{k-1}}; \quad (11.157)$$

$$\vec{S}_k(t) = A_0 e^{j\varphi_k} = A_0 e^{j(\varphi_{k-1} + \Delta\Theta)}. \quad (11.158)$$

Таким образом, идентификация переданного символа возможна на основе вычисления (определения) фазового запаздывания между идентифицированными на соседних шагах модуляции несущими сигналами:

$$\Theta^* = \Theta_m = \arg \{ \vec{S}_{k-1}^*(t) \vec{S}_k(t) \}. \quad (11.159)$$

Здесь, как и везде, символом * обозначается операция комплексного сопряжения для комплексных величин и функций (\bar{S}_k) и оценка информационного параметра для скалярных величин.

Необходимо заметить, что правильность идентификации несущих сигналов (например, ошибочная идентификация за счет фазовой ошибки восстановления опорных колебаний при квадратурном преобразовании частоты) при использовании данного решающего правила не имеет существенного значения. Важно, чтобы данная ошибка была одинаковой для рассматриваемых двух соседних шагов модуляции.

Рассмотренное выше решающее правило определяет так называемый когерентный (синхронный) корреляционный прием, при котором используются синхронные копии всех возможных несущих сигналов, представляющих передаваемые информационные символы. В случае сигналов с DPSK возможно использование более простого, так называемого автокорреляционного решающего правила, являющегося квазикогерентным.

При автокорреляционном приеме в качестве опорного колебания используется сама входная смесь полезного сигнала и шума. В этом случае функционал правдоподобия записывается следующим образом:

$$P(\bar{X}|m) = \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_c} [\bar{X}_k(t) - \bar{X}_{k-1}(t - \tau_c) e^{j\Theta_m}] \times \right. \\ \left. \times [\bar{X}_k(t) - \bar{X}_{k-1}(t - \tau_c) e^{j\Theta_m}]^* dt \right\}, \quad (11.160)$$

где $\bar{X}_k(t)$ – комплексная огибающая входной смеси сигнала и шума на текущем шаге модуляции; $\bar{X}_{k-1}(t - \tau_c)$ – задержанная на длительность символа модуляции комплексная огибающая входной смеси сигнала и шума на предыдущем шаге модуляции.

Преобразуем интеграл, стоящий под знаком экспоненты.

$$\int_0^{\tau_c} [\bar{X}_k(t) - \bar{X}_{k-1}(t - \tau_c) e^{j\Theta_m}] \times [\bar{X}_k(t) - \bar{X}_{k-1}(t - \tau_c) e^{j\Theta_m}]^* dt = \\ = \int_0^{\tau_c} [\bar{X}_k(t) \bar{X}_k^*(t) + \bar{X}_{k-1}(t) \bar{X}_{k-1}^*(t) - 2 \operatorname{Re} \{ \bar{X}_k(t) \bar{X}_{k-1}^*(t - \tau_c) e^{-j\Theta_m} \}] dt.$$

Рассмотрим интегралы вида

$$\int_0^{\tau_c} \bar{X}_k(t) \bar{X}_k^*(t) dt = \int_0^{\tau_c} (\bar{n}_k(t) + \bar{S}_k^m(t)) (\bar{n}_k(t) + \bar{S}_k^m(t))^* dt = \\ = \int_0^{\tau_c} (\bar{n}_k(t) \bar{n}_k^*(t) + \bar{S}_k^m(t) (\bar{S}_k^m(t))^* + 2 \operatorname{Re} \{ \bar{n}_k(t) (\bar{S}_k^m(t))^* \}) dt =$$

$$= E_n + E_c + 0, \quad (11.161)$$

где E_n – энергия реализации шума на интервале текущего шага модуляции; E_c – энергия, приходящаяся на один передаваемый информационный символ. Как видно, первые два слагаемых под знаком экспоненты в выражении для функционала правдоподобия на расположение максимума влияния не оказывают, поскольку не зависят от передаваемого символа. Следовательно, максимум функционала правдоподобия достигается при достижении максимума автокорреляционным интегралом и, соответственно, правило принятия решения о передаваемом информационном символе может быть сформулировано следующим образом:

$$\Theta^* = \Theta_m \leftarrow \max \left[\int_0^{\tau_c} \operatorname{Re} \left\{ \bar{X}_k(t) \bar{X}_{k-1}^*(t - \tau_c) e^{-j\Theta_m} \right\} dt, m = 1, \dots, M \right]. \quad (11.162)$$

Полученное выше правило принятия решения соответствует алгоритму работы фазового детектора, у которого в качестве опорного сигнала используется задержанная на длительность передаваемого символа копия входной смеси сигнала и шума.

Рассмотрим автокорреляционный интеграл следующего вида:

$$\begin{aligned} & \int_0^{\tau_c} \bar{X}_k(t) \bar{X}_{k-1}^*(t - \tau_c) dt = \\ & = \int_0^{\tau_c} \left(\bar{n}_k(t) \bar{n}_{k-1}^*(t) + \bar{S}_k^m(t) (\bar{S}_{k-1}^l(t))^* + \right. \\ & \quad \left. + \bar{n}_k(t) (\bar{S}_{k-1}^l(t))^* + \bar{n}_{k-1}^*(t) \bar{S}_k^m(t) \right) dt = \\ & = E_c e^{j(\Theta_m - \Theta_l)} + \sqrt{E_c} n_{k-1} e^{j(\Theta_m - \Phi_{n k-1})} + \\ & \quad + \sqrt{E_c} n_k e^{-j(\Theta_l - \Phi_{n k})} + n_k n_{k-1} e^{j(\Phi_{n k} - \Phi_{n k-1})} = \\ & = E_c e^{j(\Delta\Theta)} + n_3 e^{j\Phi_{n 3}}, \end{aligned} \quad (11.163)$$

где n_k и n_{k-1} шумовые компоненты (случайные величины), отражающие эффект шумового воздействия на выходе коррелятора от реализаций шума на интервалах текущего и предыдущего шагов модуляции; $n_3 e^{j\Phi_{n 3}}$ – эквивалентный шум на выходе коррелятора.

Из полученного выражения видно, что при достаточно больших отношениях сигнал/шум фаза сигнала на выходе коррелятора (сигнал на выходе автокорреляционного фазового детектора) соответствует приращению фазы между несущими сигналами на соседних шагах модуляции, т.е. идентифицирует передаваемый информационный символ.

Точный расчет величины вероятности ошибки для приема сигналов с DPSK достаточно сложен, поэтому ограничимся рассмотрением случая достаточно больших отношений сигнал/шум. В этом случае можно пренебречь шумовой компонентой $n_k n_{k-1} e^{j(\varphi_{nk} - \varphi_{nk-1})}$ ввиду ее малости по сравнению с компонентами $\sqrt{E_c} n_k e^{-j(\Theta_l - \varphi_{nk})}$ и $\sqrt{E_c} n_{k-1} e^{j(\Theta_m - \varphi_{nk-1})}$. Предположим, что передавался символ, кодированный нулевым приращением фазы $\Delta\Theta_m = 0 = \Delta\Theta$, тогда

$$\frac{1}{\sqrt{E_c}} \int_0^{\tau_c} \bar{X}_k(t) \bar{X}_{k-1}^*(t - \tau_c) dt = \sqrt{E_c} + \bar{n}_{k-1} + \bar{n}_k.$$

Шумовые компоненты \bar{n}_{k-1} и \bar{n}_k являются гауссовскими и некоррелированными (вообще, статистически независимыми). Определим действительную и мнимую части автокорреляционного интеграла

$$x = \sqrt{E_c} + \operatorname{Re}\{\bar{n}_{k-1} + \bar{n}_k\}, \quad y = \operatorname{Im}\{\bar{n}_{k-1} + \bar{n}_k\}.$$

Случайные величины x и y , ввиду линейности определяющих их операций и свойств комплексного представления белого гауссовского шума являются гауссовскими и независимыми друг от друга. Дисперсия случайных величин x и y равна удвоенной дисперсии синфазной и квадратурной компонент $\sigma_x^2 = \sigma_y^2 = \sigma^2 = 2\sigma_{n_I}^2 = 2\sigma_{n_Q}^2 = 2\sigma_n^2 = 2\frac{N_0}{2} = N_0$ исходных случайных величин \bar{n}_{k-1} и \bar{n}_k .

Фаза автокорреляционного интеграла (оценка приращения фазы за счет передаваемого на текущем шаге информационного символа) будет равна $\Theta^* = \operatorname{arctg}\left(\frac{y}{x}\right)$.

Далее применима методология нахождения вероятности ошибки, описанная ранее при рассмотрении качественных характеристик демодуляции сигналов с M-PSK. Отличие итогового результата для демодуляции сигналов DPSK заключается в удвоенной мощности (дисперсии) шума. Следовательно, прием сигналов с DQPSK обеспечивает такую же вероятность ошибки, как и прием сигналов с PSK одинаковой размерности, при условии увеличения отношения сигнал/шум в два раза (на 3 дБ). Данное утверждение достаточно справедливо для размерности модуляции $M \geq 4$. Для случая бинарной дифференциальной фазовой манипуляции данная оценка вероятности ошибочного приема является слишком пессимистической. Для сигналов с бинарной манипуляцией (DBPSK) применимо следующее соотношение для вероятности ошибочного приема:

$$P_e^b = \frac{1}{2} e^{-\frac{E_b}{N_0}}. \quad (11.164)$$

11.7.2. Частотная манипуляция (FSK)

11.7.2.1. Бинарная частотная манипуляция (BFSK)

Бинарная частотная манипуляция (просто частотная манипуляция) – вид дискретной модуляции, при которой радиосигнал в общем виде может быть представлен следующим образом:

$$s(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) \cos[(\omega_0 + C_k \omega_d)t + \varphi_0], \quad (11.165)$$

где A_0 , ω_0 , φ_0 – амплитуда, центральная частота и начальная фаза несущего колебания; ω_d – принятая девиация частоты; $C_k = \pm 1$ – последовательность, соответствующая передаваемым информационным символам «0» и «1» (обычно +1 соответствует «0», а -1 представляет «1»); $g_F(t) = 0 \quad \forall t \notin [0, \tau_b]$ – формирующий импульс; τ_b – длительность информационного символа (бита); k – текущий номер передаваемого информационного символа.

Формирующий импульс в общем случае может иметь произвольную форму $g_F(t)$. Чаще всего используется прямоугольная форма импульса.

Примеры временных функций, соответствующих радиосигналу с BFSK при прямоугольной форме формирующего импульса, приведен на рис. 11.22 и 11.23.

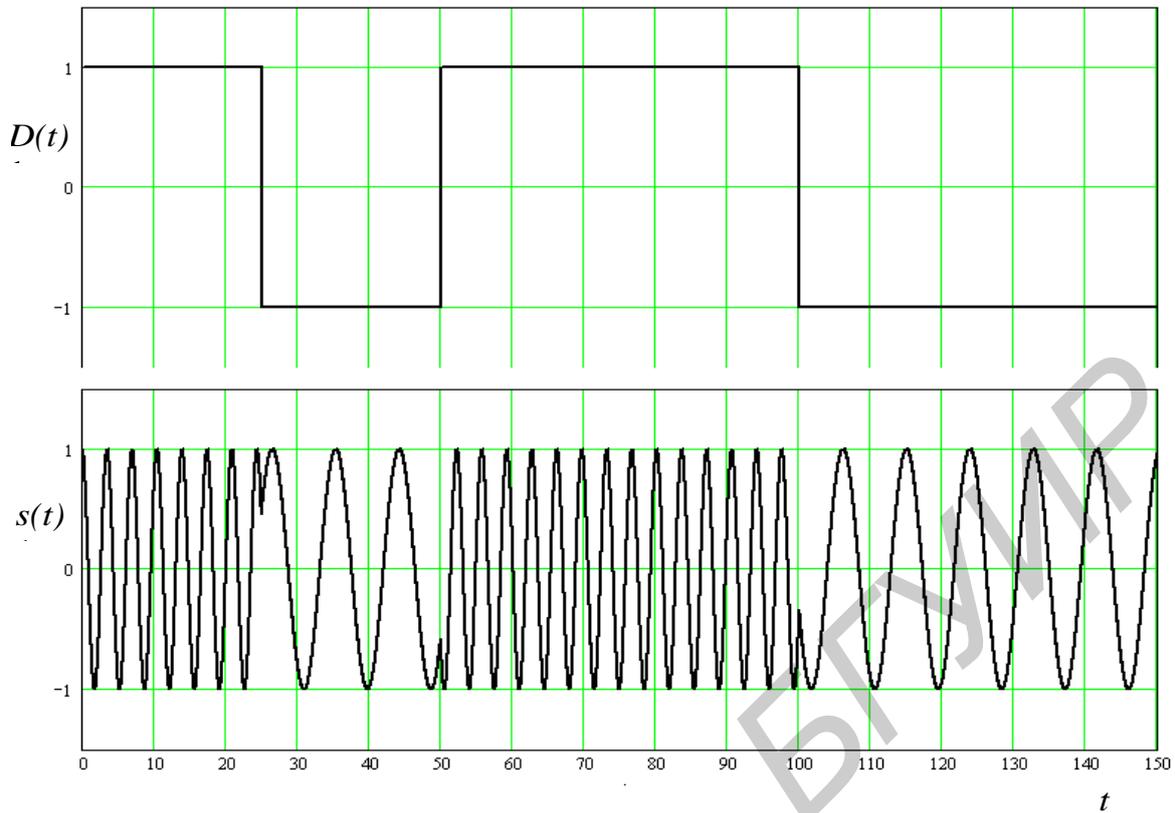


Рис. 11.22. Форма радиосигнала с BFSK (общий случай)

Комплексное представление радиосигнала с BFSK может быть определено следующим образом:

$$\bar{s}(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) e^{jC_k \omega_d t} e^{j\omega_0 t} . \quad (11.166)$$

Комплексная огибающая сигнала с бинарной частотной манипуляцией представляется следующим выражением:

$$\begin{aligned} \bar{S}(t) &= A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) e^{jC_k \omega_d t} = \\ &= A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) (\cos(\omega_d t) + j \sin(C_k \omega_d t)) . \end{aligned} \quad (11.167)$$

Синфазная и квадратурная компоненты комплексной огибающей определяются следующим образом:

$$S_I(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \cos[\omega_d t]; \quad (11.168)$$

$$S_Q(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) C_k \sin[\omega_d t]. \quad (11.169)$$

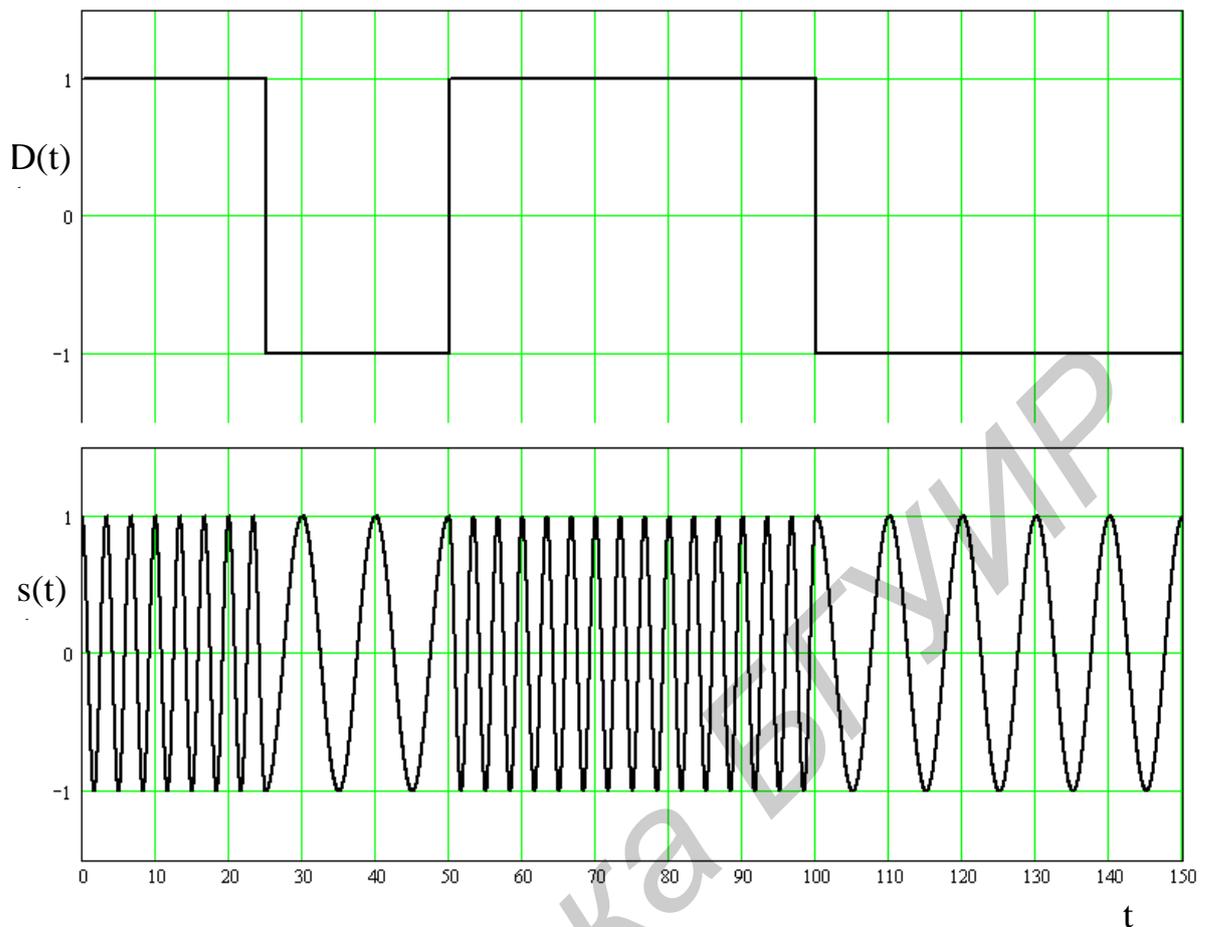


Рис.11.23. Форма радиосигнала с BFSK (без разрыва фазы)

Анализ соотношений для квадратурной и синфазной компонент сигнала с бинарной частотной манипуляцией показывает, что только половина энергии $E_c = E_b$, приходящейся на один символ (бит) модуляции, расходуется на передачу информации, вторая половина энергии не несет информационной нагрузки. Кроме того, метод наложения информационной составляющей на сигнал-носитель соответствует методу $\pm\pi/2$ -BPSK квадратурной составляющей радиосигнала. Изложенное выше позволяет предположить, что помехоустойчивость демодуляции сигнала с BFSK будет проигрывать 3 дБ помехоустойчивости демодуляции сигналов с $\pm\pi/2$ -BPSK (данное предположение будет подтверждено ниже).

Векторная диаграмма сигнала с бинарной частотной манипуляцией представлена на рис. 11.24.

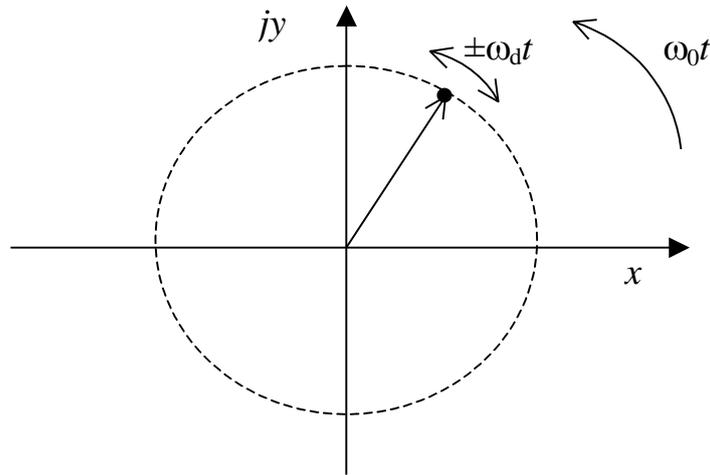


Рис. 11.24. Векторная диаграмма сигнала с BFSK

При квазиоптимальном приеме решение принимается на основании наблюдения входной смеси сигнала и шума на интервале времени, соответствующем длительности τ_b , с учетом этого для интервала наблюдения на k -м интервале модуляции комплексная огибающая может быть переписана в следующем виде:

$$\begin{aligned} \vec{S}(t) &= A_0 g_F(t - k\tau_b) e^{jC_k \omega_d t} = \\ &= A_0 g_F(t - k\tau_b) [\cos(\omega_d t) + jC_k \sin(\omega_d t)], \end{aligned} \quad (11.170)$$

а соответствующий ей радиосигнал будет определяться следующим образом:

$$\begin{aligned} s(t) &= A_0 g_F(t - k\tau_b) [\cos(\omega_d t) \cos(\omega_0 t) - C_k \sin(\omega_d t) \sin(\omega_0 t)] = \\ &= \begin{cases} A_0 g_F(t - k\tau_b) \cos[(\omega_0 - \omega_d)t], & C_k = -1 \\ A_0 g_F(t - k\tau_b) \cos[(\omega_0 + \omega_d)t], & C_k = 1 \end{cases} \quad (11.171) \\ &= A_0 g_F(t - k\tau_b) \cos[(\omega_0 + C_k \omega_d)t] \end{aligned}$$

Таким образом, сигнал, несущий информацию о передаче двоичной «1», можно записать как

$$s^1(t) = A_0 g_F(t - k\tau_b) [\cos((\omega_0 - \omega_d)t)], \quad (11.172)$$

и сигнал, несущий информацию о передаче двоичного «0»,

$$s^0(t) = A_0 g_F(t - k\tau_b) [\cos((\omega_0 + \omega_d)t)]. \quad (11.173)$$

Функционал правдоподобия, соответствующий передаче двоичной «1»,

$$\begin{aligned}
 P(\bar{X}|1) &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\bar{X}(t) - \bar{S}^1(t)] \times [\bar{X}(t) - \bar{S}^1(t)]^* dt \right\} = \\
 &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\bar{X}(t) - A_0 g_F(t - k\tau_b) (\cos(\omega_d t) - j \sin(\omega_d t))] \times \right. \\
 &\quad \left. \times [\bar{X}(t) - A_0 g_F(t - k\tau_b) (\cos(\omega_d t) - j \sin(\omega_d t))]^* dt \right\} \quad (11.174)
 \end{aligned}$$

и передаче двоичного «0»

$$\begin{aligned}
 P(\bar{X}|0) &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\bar{X}(t) - \bar{S}^0(t)] \times [\bar{X}(t) - \bar{S}^0(t)]^* dt \right\} = \\
 &= \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_b} [\bar{X}(t) - A_0 g_F(t - k\tau_b) (\cos(\omega_d t) + j \sin(\omega_d t))] \times \right. \\
 &\quad \left. \times [\bar{X}(t) - A_0 g_F(t - k\tau_b) (\cos(\omega_d t) + j \sin(\omega_d t))]^* dt \right\}, \quad (11.175)
 \end{aligned}$$

где $\bar{X}(t)$ – комплексная огибающая входной смеси полезного сигнала и шума.

Решение о значении передаваемого информационного символа C_k на k -м шаге модуляции формируется на основании сравнения рассчитанных значений функционала (функции для дискретного времени наблюдения) правдоподобия.

$$P(\bar{X}|0) > P(\bar{X}|1) \Rightarrow C_k^* = 1 \quad (\text{передавался "0"});$$

$$P(\bar{X}|0) < P(\bar{X}|1) \Rightarrow C_k^* = -1 \quad (\text{передавалась "1"})$$

или

$$\ln \left[\frac{P(\bar{X}|1)}{P(\bar{X}|0)} \right] > 0 \Rightarrow C_k^* = -1 \quad (\text{передавалась "1"}); \quad (11.176)$$

$$\ln \left[\frac{P(\bar{X}|1)}{P(\bar{X}|0)} \right] < 0 \Rightarrow C_k^* = 1 \quad (\text{передавался "0"}). \quad (11.177)$$

В итоге, осуществив необходимые преобразования, имеем:

$$\int_0^{\tau_b} X_Q(t) \sin(\omega_d t) < 0 \Rightarrow C_k^* = -1 \quad (\text{передавалась "1"}), \quad (11.178)$$

$$\int_0^{\tau_b} X_Q(t) \sin(\omega_d t) > 0 \Rightarrow C_k^* = 1 \quad (\text{передавался "0"}), \quad (11.179)$$

где $X_Q(t)$ – квадратурная компонента комплексной огибающей входной смеси сигнала и помехи.

Таким образом, в процессе демодуляции сигнала с BFSK участвует только одна из двух квадратурных компонент, вторая, не несущая информационной нагрузки, демодулятором игнорируется.

Наличие в полезном сигнале составляющей синхронной с несущей частотой и девиацией частоты позволяет осуществлять некогерентный прием сигналов с BFSK, слабо чувствительный к ошибке восстановления опорного колебания квадратурного преобразователя частоты. Схема некогерентного «автокорреляционного» демодулятора приведена на рис. 11.25.

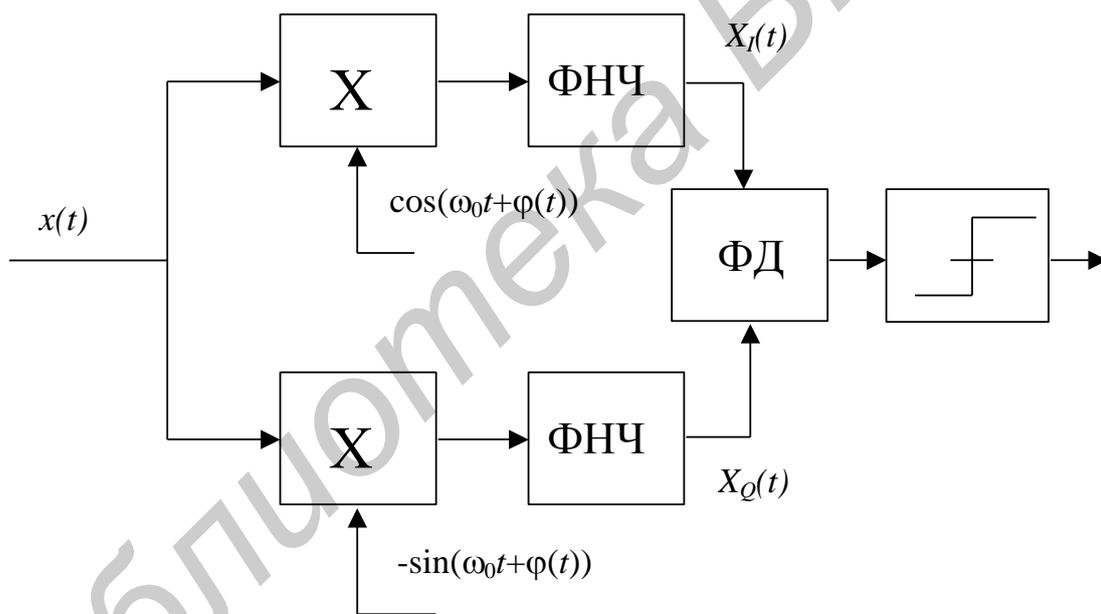


Рис. 11.25. Автокорреляционный демодулятор сигнала с BFSK

Необходимо отметить, что некогерентный прием проигрывает в помехоустойчивости когерентному (особенно при малых отношениях сигнал/шум).

Определим элементы нормированной корреляционной матрицы ρ сигнала с BFSK.

$$\rho_{km} = \frac{1}{E_b} \int_0^{\tau_b} s^k(t) s^m(t) dt = \frac{1}{E_b} \int_0^{\tau_b} s^k(t) s^m(t) dt. \quad (11.180)$$

При прямоугольной форме формирующего импульса имеем

$$\begin{aligned} \rho_{11} = \rho_{22} &= \frac{1}{E_b} \int_0^{\tau_b} [A_0 \cos[(\omega_0 \pm \omega_d)t]]^2 dt = \\ &= \frac{1}{E_b} \frac{A_0^2 \tau_b}{2} = 1, \end{aligned} \quad (11.181)$$

где $E_b = \frac{A_0^2 \tau_b}{2}$.

$$\begin{aligned} \rho_{12} = \rho_{21} &= \frac{1}{E_b} \int_0^{\tau_b} [A_0 \cos[(\omega_0 + \omega_d)t]] \times \\ &\times [A_0 \cos[(\omega_0 - \omega_d)t]] dt = \\ &= -\frac{1}{E_b} \frac{A_0^2}{2} \int_0^{\tau_b} [\cos(2\omega_d t) + \cos(2\omega_0 t)] dt \cong \\ &\cong \frac{1}{E_b} \frac{A_0^2}{2 \times 2\omega_d} \sin(2\omega_d \tau_b) = \frac{1}{2\omega_d \tau_b} \sin(2\omega_d \tau_b); \end{aligned} \quad (11.182)$$

$$\rho = \begin{pmatrix} 1 & \frac{1}{2\omega_d \tau_b} \sin(2\omega_d \tau_b) \\ \frac{1}{2\omega_d \tau_b} \sin(2\omega_d \tau_b) & 1 \end{pmatrix}. \quad (11.183)$$

Вероятность ошибочного приема информационного символа (бита) можно определить на основании общего выражения для ошибки приема, полученного ранее:

$$P_e^b = 1 - \frac{1}{(2\pi)^{|\rho|^{1/2}}} \int_{-\infty}^{\infty} \int_{-\infty}^{u_1 + \sqrt{2E_b/N_0}(1-\rho_{12})} \exp\left\{-\frac{1}{2} \mathbf{u}^T \boldsymbol{\rho}^{-1} \mathbf{u}\right\} du_2 du_1. \quad (11.184)$$

Наиболее интересным с точки зрения практического использования является ортогональная FSK, для которой $\rho_{ij} = 0 \Big|_{i \neq j}$, в этом случае

$$P_e^b = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(F(x + \sqrt{2E_c/N_0}) \right) dx = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{2N_0}} \right). \quad (11.185)$$

Сравнивая полученное выражение с выражением для вероятности ошибки приема сигнала с $\pm\pi/2$ -BPSK (рис. 11.11) $P_{e \text{ BPSK}}^b = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right)$ необходимо отметить, что метод BFSK модуляции проигрывает в помехоустойчивости методу $\pm\pi/2$ -BPSK ровно 3дБ.

В случае некогерентного приема вероятность ошибочного приема может быть оценена следующим образом:

$$P_e^b = \frac{1}{2} e^{-\frac{E_b}{2N_0}}. \quad (11.186)$$

На рис. 11.26 приведены зависимости вероятности ошибки от отношения сигнал/шум для когерентного и некогерентного приема сигналов с ортогональной BFSK, для сравнения там же приведена зависимость вероятности ошибки для приема сигнала с $\pm\pi/2$ -BPSK.

Ортогональная BFSK реализуется при выполнении условия $\sin(2\omega_d \tau_b) = 0$, отсюда вытекают требования к девиации частоты:

$$\begin{aligned} 2\omega_d \tau_b &= 4\pi f_d \tau_b = \pi n, \quad n = 1, 2, \dots, \\ f_d &= \frac{n}{4\tau_b} = \frac{n}{4} C, \quad n = 1, 2, \dots, \end{aligned} \quad (11.187)$$

где C – скорость модуляции (манипуляции) [бит/с]. При больших девиациях частоты (малых отношениях T_d/τ_b , $T_d = 1/f_d$) квазиортогональность достигается автоматически за счет множителя $\frac{1}{2\omega_d \tau_b}$, входящего в коэффициенты взаимной корреляции.

11.7.2.2. М-позиционная (многопозиционная) частотная манипуляция (М-FSK)

М-позиционная (многопозиционная) частотная манипуляция – вид дискретной модуляции, при которой радиосигнал в общем виде может быть представлен следующим образом:

$$s(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) \cos[(\omega_0 + \omega_d(t))t + \varphi_0], \quad (11.188)$$

где A_0 , ω_0 , φ_0 – амплитуда, центральная частота и начальная фаза несущего колебания; $\omega_d(t)$ – девиация частоты, соответствующая передаваемой информации; $g_F(t) = 0 \quad \forall t \notin [0, \tau_c]$ – формирующий импульс; τ_c – длительность информационного символа; k – текущий номер передаваемого информационного символа. Девиация частоты $\omega_d(t)$ постоянна на интервале времени, совпадающем с интервалом действия информационного символа и принимает одно из M известных значений ω_{dm} .

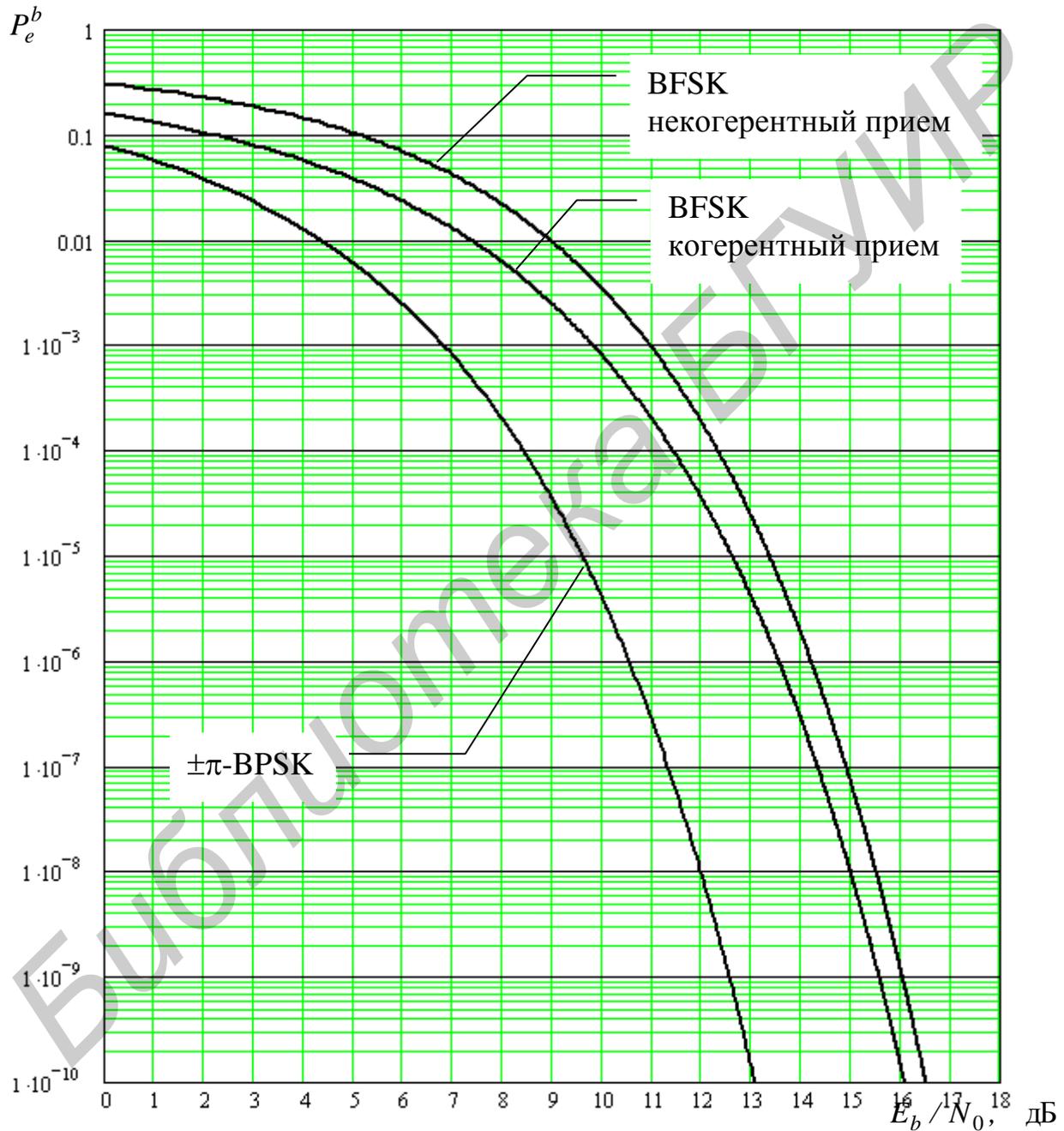


Рис. 11.26. Вероятность ошибки при приеме сигналов с BFSK

Пример временной функции, соответствующей радиосигналу 4-FSK при прямоугольной форме формирующего импульса, приведен на рис. 11.27.

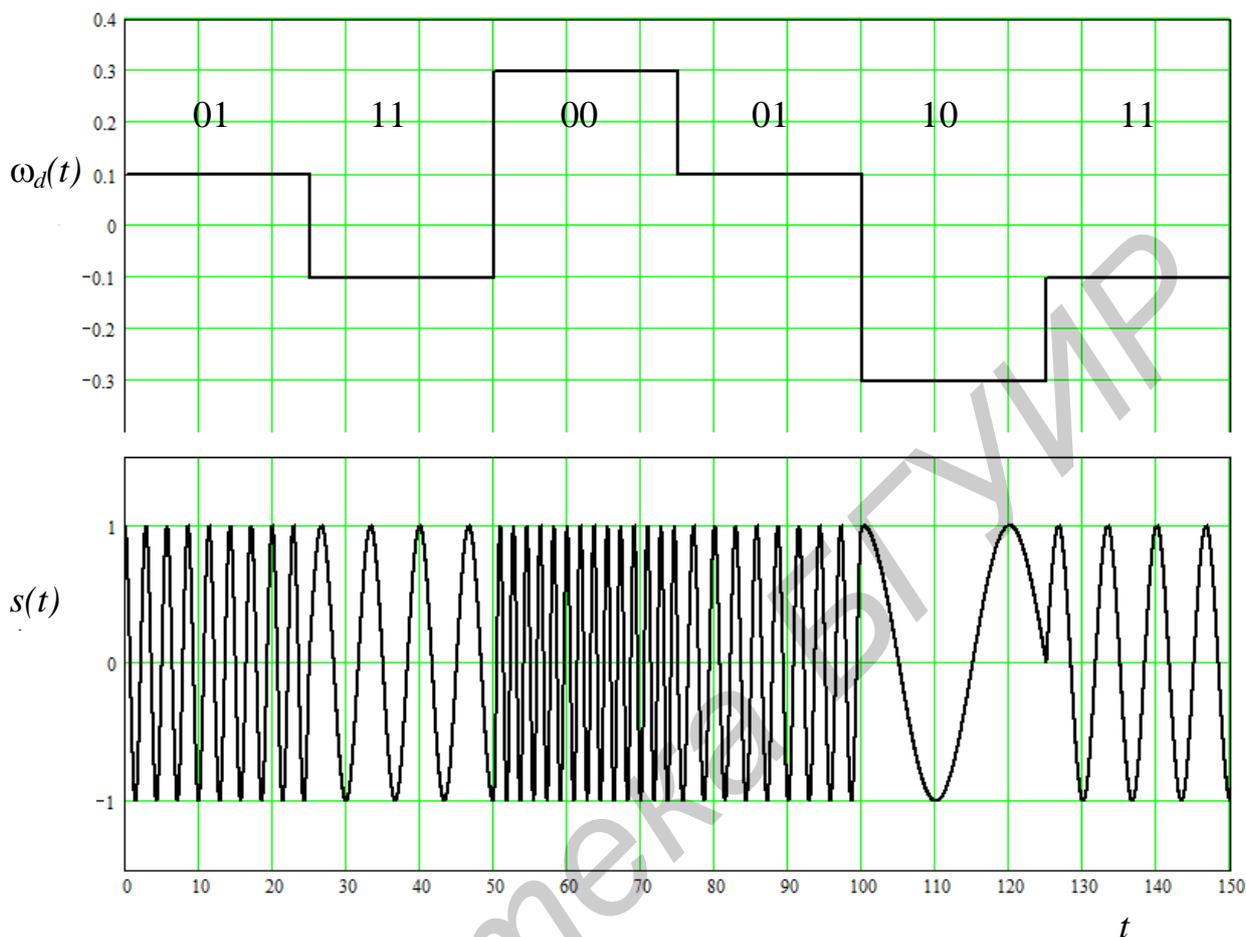


Рис. 11.27. Форма радиосигнала с 4-FSK

На рисунке введены следующие обозначения: $\omega_d(t)$ – девиация частоты, соответствующая передаваемой информации; $S(t)$ – радиосигнал с BFSK.

Комплексное представление радиосигнала с BFSK может быть определено следующим образом:

$$\vec{s}(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) e^{j\omega_d m t} e^{j\omega_0 t}. \quad (11.189)$$

Комплексная огибающая сигнала с бинарной частотной манипуляцией представляется следующим выражением:

$$\begin{aligned} \vec{S}(t) &= A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) e^{j\omega_d m t} = \\ &= A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_c) (\cos(\omega_d m t) + j \sin(\omega_d m t)). \end{aligned} \quad (11.190)$$

При квазиоптимальном приеме решение принимается на основании наблюдения входной смеси сигнала и шума на интервале времени, соответствующем длительности информационного символа τ_c , для интервала наблюдения на k -м интервале модуляции комплексная огибающая может быть переписана в следующем виде:

$$\bar{S}(t) = A_0 g_F(t - k\tau_b) e^{j\omega_{dm} t}. \quad (11.191)$$

Решение о значении передаваемого информационного символа (значении девиации частоты ω_{dm}) на k -м шаге модуляции формируется на основании сравнения рассчитанных значений функционала правдоподобия и определения того значения параметра m , которое обеспечивает его максимум.

$$P(\bar{X}|m) = \exp \left\{ -\frac{1}{N_0} \int_0^{\tau_c} [\bar{X}(t) - \bar{S}^m(t)] \times [\bar{X}(t) - \bar{S}^m(t)]^* dt \right\}. \quad (11.192)$$

Как уже неоднократно отмечалось, для симметричных каналов поиск максимума функционала правдоподобия соответствует поиску максимума среди корреляционных интегралов следующего вида:

$$\operatorname{Re} \left\{ \int_0^{\tau_c} [\bar{X}(t) (\bar{S}^m(t))^*] dt \right\}.$$

Поскольку функции $\bar{S}^m(t)$ представляют гармонические колебания с известной частотой ω_{dm} и фазой, значения данных корреляционных интегралов соответствуют откликам синхронных детекторов с опорными колебаниями вида $\cos[(\omega_0 + \omega_{dm})t]$ при когерентном детектировании или откликам обычных амплитудных детекторов, стоящих на выходе полосовых фильтров настроенных на частоты $\omega_0 + \omega_{dm}$ при некогерентном детектировании.

Определим вероятность ошибочного приема для сигнала с M-FSK при условии ортогональности несущих различные информационные символы сигналов (достаточно большом разnose между частотами ω_{dm}). В этом случае нормированная корреляционная матрица ρ имеет вид

$$\rho = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}. \quad (11.193)$$

Вероятность ошибочного приема информационного символа можно определить на основании общего выражения для ошибки приема ортогональных сигналов, полученного ранее:

$$P_e^c = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(F(x + \sqrt{2E_c/N_0}) \right)^{M-1} dx, \quad (11.194)$$

где $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$ – функция распределения вероятности нормальной случайной величины с нулевым математическим ожиданием и единичной дисперсией.

С учетом вероятности ошибочного приема информационного символа, состоящего из $k = \log_2(M)$ бит (элементов), имеем вероятность ошибки приема отдельного бита:

$$\begin{aligned} P_e^b &= \frac{2^{k-1}}{2^k - 1} P_e^c = \frac{2^{k-1}}{2^k - 1} \left[1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(F(x + \sqrt{2E_c/N_0}) \right)^{M-1} dx \right] = \\ &= \frac{2^{k-1}}{2^k - 1} \left[1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(F(x + \sqrt{2kE_b/N_0}) \right)^{2^{k-1}-1} dx \right], \end{aligned} \quad (11.195)$$

где $E_b = E_c/k$ – энергия, приходящаяся на один бит передаваемой информации. Данная оценка вероятности ошибки является слегка завышенной поскольку при правильном кодировании значений девиации частоты информационными символами вероятность искажения сразу нескольких бит составляющих символа, как правило, значительно ниже вероятности искажения только одного бита (элемента).

11.7.3. Непрерывная фазовая модуляция (CPM)

11.7.3.1. Общее представление

Обычный сигнал с частотной манипуляцией или модуляцией с частотным сдвигом (FSK) генерируется путём сдвига частоты несущей на величину $f_d I$, где $I = \pm 1, \pm 3, \dots, \pm(M-1)$. Этот вид модуляции сигналов был описан выше и по своей сути является методом модуляции без памяти (предыдущие состояния сигнала никоим образом не влияют на его текущее состояние). Переход от одной частоты к другой может быть выполнен посредством M отдельных генераторов, настроенных на необходимые частоты, и выбора одной из M частот согласно частному значению передаваемого информационного символа на сигнальном интервале длиной τ_c .

Однако такое резкое переключение с выхода одного генератора на выход другого в смежных сигнальных интервалах приводит к относительно большим долям боковых частотных составляющих вне основной полосы сигнала, и, следовательно, этот метод реализует радиосигналы с большей полосой частот, чем это необходимо для передачи сигнала. Чтобы избежать использования сигналов с большими долями боковых полос, информационный сигнал может модулировать одно несущее колебание, частота которого меняется непрерывно. Результирующий частотно-модулированный сигнал имеет в этом случае непрерывную фазу и поэтому назван частотной манипуляцией (модуляцией) с непрерывной фазой (CPFSK или CPM). Этот вид сигнала имеет память, обусловленную тем, что фазу несущей заставляют быть непрерывной и, следовательно, обеспечивают влияние предыдущих состояний сигнала на его текущее состояние.

Сигнал с непрерывной фазой может быть записан следующим образом:

$$s(t) = A_0 \cos(\omega_0 t + \Phi(t, a) + \varphi_0), \quad (11.196)$$

где

$$\Phi(t, a) = 2\pi \sum_{k=-\infty}^{\infty} h_k a_k q(t - k\tau_c);$$

$$q(t) = \int_{-\infty}^t g(\tau) d\tau,$$

где a_k – текущее значение модулирующего параметра, принимающее одно из M возможных значений $\pm 1, \pm 3, \dots, \pm(M-1)$ в зависимости от передаваемого информационного символа; $g(\tau)$ – нормированный формирующий импульс частотной манипуляции, $\int_{-\infty}^{\infty} g(\tau) d\tau = 1/2$, обладающий длительностью

$\tau_g = L\tau_c$, L – обычно натуральное число; h_k – индекс угловой модуляции (модуляция $h_k = h$ является одноиндексной – SHPM, в противном – мультииндексной – MHPM).

Если $g(t) = 0$ для $t > \tau_c$, сигнал называют сигналом с полным откликом. Если $g(t) \neq 0$ для $t > \tau_c$, модулированный сигнал называют сигналом с частичным (парциальным) откликом. Очевидно, что неограниченное число разновидностей сигналов непрерывной фазой можно генерировать выбором различных импульсов $g(\tau)$ и изменением индекса модуляции h и размера алфавита M .

Если $g(t) = 0$ для $t > \tau_c$, сигнал называют сигналом с полным откликом. Если $g(t) \neq 0$ для $t > \tau_c$, модулированный сигнал называют сигналом с частичным (парциальным) откликом. Очевидно, что неограниченное число разновидностей сигналов непрерывной фазой можно генерировать выбором различных импульсов $g(\tau)$ и изменением индекса модуляции h и размера алфавита M .

Для текущего шага модуляции ($t \in [(n-1)\tau_c, n\tau_c]$) фаза сигнала с полным откликом может быть записана:

$$\Phi(t, a) = 2\pi h_n a_n q(t - n\tau_c) + \pi \sum_{k=-\infty}^{n-1} h_k a_k = 2\pi h_n a_n q(t - n\tau_c) + \Theta_n. \quad (11.197)$$

Параметр Θ_n представляет собой накопление (память) от всех информационных символов, переданных ранее до текущего шага модуляции.

11.7.3.2. Наиболее часто используемые виды формирующего импульса частотной манипуляции

Прямоугольный формирующий импульс (Rectangular REC) иногда называют LREC, где L обозначает число, соответствующее отношению τ_g/τ_c длительности формирующего импульса к длительности передаваемого символа. Нормированный прямоугольный формирующий импульс может быть представлен следующим образом:

$$g(t) = \begin{cases} \frac{1}{2\tau_g}, & 0 \leq t \leq \tau_g \\ 0, & \text{для других } t \end{cases}. \quad (11.198)$$

В этом случае закон изменения фазы модулированного радиосигнала на интервале времени, соответствующем k -му шагу модуляции, принимает следующий вид (здесь и далее для простоты, рассматриваем случай одно индексной модуляции):

$$\begin{aligned} \Phi(t, a) &= 2\pi h \left(\sum_{i=-\infty}^{k-1} a_i \frac{1}{2} + a_k \frac{t - k\tau_g}{2\tau_g} \right) \Big|_{t \in [k\tau_g, (k+1)\tau_g]} = \\ &= \pi h \sum_{i=-\infty}^{k-1} a_i + \pi h a_k \frac{t}{\tau_g} \Big|_{t \in [0, \tau_g]} = \pi h a_k \frac{t}{\tau_g} + \Phi_{k-1} \Big|_{t \in [0, \tau_g]} \end{aligned} \quad (11.199)$$

В этом случае результирующий радиосигнал на k -м шаге модуляции представляется следующим выражением:

$$s(t) = A_0 \cos(\omega_0 t + a_k h \frac{\pi t}{\tau_g} + \Phi_{k-1}). \quad (11.200)$$

Наибольшее распространение из всех возможных вариантов реализации данной схемы модуляции получил метод минимальной манипуляции (MSK), для которого $\tau_g = \tau_c = \tau_b$, $a_k = \pm 1$ и $h = 0,5$. Данный метод манипуляции будет рассмотрен более подробно ниже.

Если в качестве формирующего выбран приподнятый косинусоидальный импульс (Raised Cosine RC или LRC), то его аналитическое выражение будет следующим:

$$g(t) = \begin{cases} \frac{1}{2\tau_g} \left(1 - \cos \left[\frac{2\pi t}{\tau_g} \right] \right), & 0 \leq t \leq \tau_g; \\ 0, & \text{для других } t. \end{cases} \quad (11.201)$$

Нормированный формирующий импульс при спектрально приподнятом косинусоидальном импульсе (Spectrally Raised Cosine SRC или LSRC) определяется следующим образом:

$$g(t) = \begin{cases} \frac{1}{\tau_g} \left(\frac{\sin \left[\frac{2\pi t}{\tau_g} \right] \cos \left[\beta \frac{2\pi t}{\tau_g} \right]}{\frac{2\pi t}{\tau_g} \left(1 - \left(\frac{4\beta t}{\tau_g} \right)^2 \right)} \right), & 0 \leq t \leq \tau_g; \quad 0 \leq \beta \leq 1; \\ 0, & \text{для других } t. \end{cases} \quad (11.202)$$

Формирующий импульс при простой частотной модуляции (Tamed Frequency Modulation TFM) определяется выражением

$$g(t) = \frac{1}{8} [g_0(t - \tau_c) + 2g_0(t) + g_0(t + \tau_c)]; \quad (11.203)$$

$$g_0(t) \approx \sin \left(\frac{\pi t}{\tau_c} \right) \left[\frac{1}{\pi t} - \frac{2 - \frac{2\pi t}{\tau_c} \operatorname{ctg} \left(\frac{\pi t}{\tau_c} \right) - \left(\frac{\pi t}{\tau_c} \right)^2}{\frac{24\pi t^3}{\tau_c^2}} \right]. \quad (11.204)$$

Гауссовский формирующий импульс (Gaussian) обычно используется совместно с обеспечением условий минимальной манипуляции ($\tau_c = \tau_b$, $a_k = \pm 1$ и $h = 0,5$), при этом схема модуляции получила наименование – гауссовская минимальная манипуляция (GMSK). В этом случае формирующий импульс представляется в виде

$$g(t) = \frac{B_b}{2} \sqrt{\frac{2\pi}{\ln(2)}} \exp\left\{-\frac{2(\pi B_b)^2 t^2}{\ln(2)}\right\}, \quad 0 \leq B_b \tau_b = B_b \tau_c \leq 1. \quad (11.205)$$

Данный метод модуляции будет рассмотрен более подробно несколько ниже.

На рис. 11.28 приведены временные функции, соответствующие рассмотренным нормированным формирующим импульсам.

Соответствующие рассмотренным формирующим импульсам фазовые отклики $q(t) = \int_{-\infty}^t g(\tau) d\tau$ приведены на рис. 11.29.

11.7.3.3. Общие принципы демодуляции сигналов с СРМ

Как было показано выше, общее правило оптимальной (квазиоптимальной) демодуляции сигналов с дискретной модуляцией основывается на расчете значений функционала (функции) правдоподобия для всех возможных состояний полезного сигнала, вызываемых процессом модуляции. Решение о значении передаваемого информационного символа формируется на основании сравнения рассчитанных значений функционала правдоподобия и определения того состояния полезного сигнала, которое обеспечивает его максимальное значение.

$$P(\bar{X}|m) = \exp\left\{-\frac{1}{N_0} \int_0^{\tau_c} [\bar{X}(t) - \bar{S}^m(t)] \times [\bar{X}(t) - \bar{S}^m(t)]^* dt\right\}, \quad (11.206)$$

где $\bar{S}^m(t)$ – m -е состояние полезного сигнала, вызванное процессом модуляции.

Особенность сигналов с СРМ заключается в наличии «памяти» сигнала о предыдущих передаваемых информационных символах. Данная память не позволяет сделать однозначное сопоставление состояний полезного сигнала и конкретных передаваемых информационных символов, кроме того, для сигналов с парциальным откликом (длительность формирующего импульса превосходит длительность информационного символа) существует явление наложения модуляционных откликов от группы соседних передаваемых символов друг на друга.

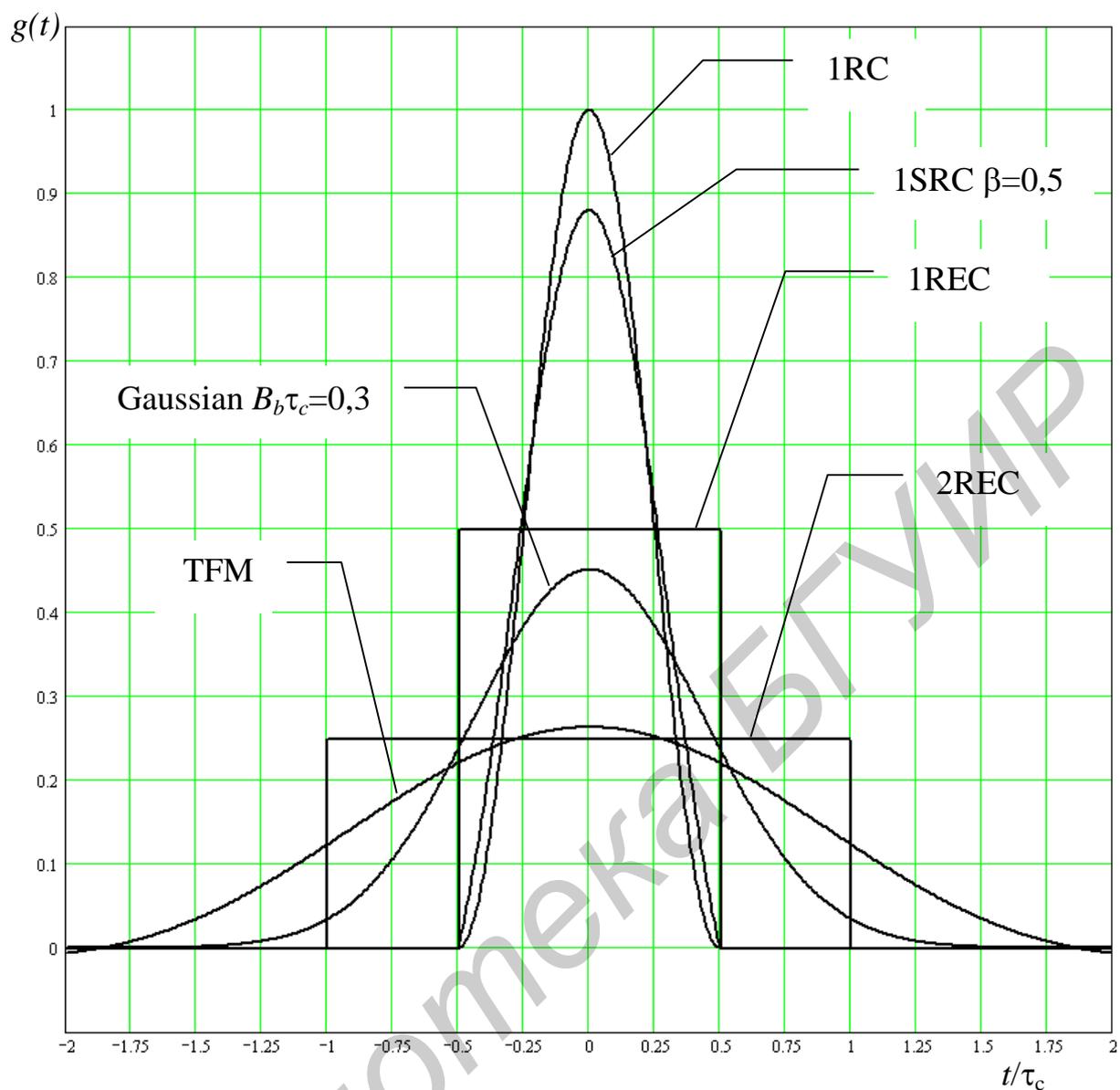


Рис. 11.28. Основные нормированные формирующие импульсы

Рассмотрим состояния полезного сигнала подробнее. Как было показано выше, сигнал с СРМ в общем виде записывается выражением (11.196).

Рассмотрим k -й шаг модуляции, для которого запишем

$$\Phi(t, a) = 2\pi \sum_{i=-\infty}^k h_i a_i q(t - i\tau_i) \quad (11.207)$$

или

$$\begin{aligned}
\Phi(t, a) &= 2\pi \sum_{i=-\infty}^{k-L} h_i a_i q(t - i\tau_c) + 2\pi \sum_{i=k-L+1}^k h_i a_i q(t - i\tau_c) = \\
&= \pi \sum_{i=-\infty}^{k-L} h_i a_i + 2\pi \sum_{i=k-L+1}^{k-1} h_i a_i q(t - i\tau_c) + 2\pi h_k a_k q(t - k\tau_c) = \\
&= \Theta_{k-L} + \Theta(t, a_{k-1}, \dots, a_{k-L+1}) + \Theta_k(t, a_k).
\end{aligned}
\tag{11.208}$$

При этом $\Theta_k(t, a_k) = 2\pi h_k a_k q(t - k\tau_c)$ представляет отклик на текущий символ модуляции (который необходимо «опознать» в процессе демодуляции); $\Theta(t, a_{k-1}, \dots, a_{k-L+1}) = 2\pi \sum_{i=k-L+1}^{k-1} h_i a_i q(t - i\tau_c)$ представляет так называемую коррелированную память сигнала или неустановившиеся (незаконченные) отклики от $L-1$ предыдущих передаваемых информационных символов; $\Theta_{k-L} = \pi \sum_{i=-\infty}^{k-L} h_i a_i$ представляет некоррелированную память или установившиеся (законченные) отклики от всех остальных предыдущих переданных символов.

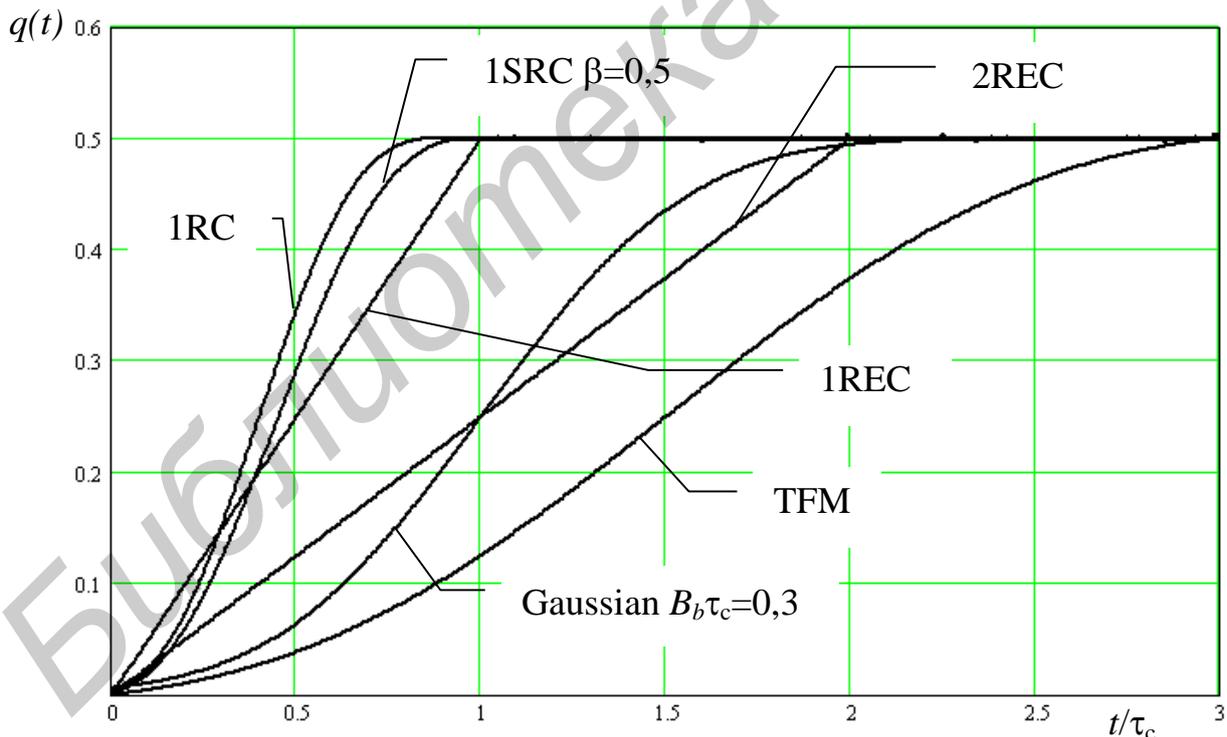


Рис. 11.29. Фазовые отклики при различных видах формирующих импульсов

Таким образом, состояния полезного сигнала для подстановки в выражение для функционала правдоподобия можно определить следующим образом:

$$s^m(t) = A_0 \cos(\omega_0 t + \Theta_k(t, a_k) + \Theta(t, a_{k-1}, \dots, a_{k-L+1}) + \Theta_{k-L}) \quad (11.209)$$

или в терминах комплексной огибающей:

$$\vec{s}^m(t) = A_0 e^{j\Theta_k(t, a_k) + \Theta(t, a_{k-1}, \dots, a_{k-L+1}) + \Theta_{k-L}} \quad (11.210)$$

По своей сути $\Theta_k(t, a_k) + \Theta(t, a_{k-1}, \dots, a_{k-L+1}) + \Theta_{k-L}$ представляет собой фазовую траекторию (путь) полезного сигнала под воздействием модулирующих информационных символов. Таким образом, задача демодулятора сводится к нахождению наиболее близкой фазовой траектории полезного сигнала из всех возможных (разрешенных) к фазовой траектории наблюдаемой смеси сигнала и шума.

Поскольку метод модуляции СРМ относится к методам дискретной модуляции, то достаточным является рассмотрение (сравнение) не непосредственно фазовых траекторий, а их «узловых точек» (значений фазовой траектории в моменты времени соответствующие моментам смены информационных символов $t_k = k\tau_c$, $k = 0, 1, 2, \dots$

Как было показано выше, поиск максимума функционала правдоподобия равноценен определению результирующей фазы принимаемой смеси сигнала и шума в момент окончания информационного символа

$$\Theta^* = \arctg \left(\frac{\int_0^{\tau_c} X_Q(t) dt}{\int_0^{\tau_c} X_I(t) dt} \right),$$

где $X_I(t)$ и $X_Q(t)$ – синфазная и квадратурная компоненты комплексной огибающей входной смеси сигнала и помехи, и поиска наиболее близкого из разрешенных фазовых состояний сигнала $\Theta_k((k+1)\tau_c, a_k) + \Theta((k+1)\tau_c, a_{k-1}, \dots, a_{k-L+1}) + \Theta_{k-L}$ и соответственно идентификации переданного символа a_k .

Необходимо отметить, что число возможных фазовых состояний сигнала в общем случае оказывается достаточно большим, например, для одноиндексной модуляции с памятью на L символов и индексом модуляции $h = m/p$ число возможных фазовых состояний равно $2pM^{L-1}$ при нечетном m и pM^{L-1} при четном m , что приводит к громоздкости и большой вычислительной сложности процедуры принятия решения о передаваемом информационном символе. Без потери оптимальности громоздкость процедуры демодуляции может быть уменьшена за счет исключения так называемых «не выживших» фазовых путей в соответствии с процедурой (алгоритмом) Витерби, определенной для декодирования циклических кодов.

11.7.3.4. Минимальная манипуляция (MSK)

Рассмотрим сигнал с РСМ для случая бинарной модуляции ($a_k = \pm 1$) с полным откликом (длительность формирующего импульса равна длительности информационного импульса $\tau_g = \tau_c = \tau_b$), прямоугольной формой формирующего импульса (1REC) и индекса модуляции, равного $h = 0,5$. В этом случае радиосигнал на k -м шаге модуляции можно записать

$$s_{msk}(t) = A_0 \cos\left(\omega_0 t + a_k \frac{\pi}{2} \frac{t}{\tau_b} + \Phi_{k-1}\right) \Big|_{t \in [0, \tau_b]} = \quad (11.211)$$

$$= A_0 \cos\left(\left(\omega_0 + a_k \frac{\pi}{2} \frac{1}{\tau_b}\right)t + \Phi_{k-1}\right) \Big|_{t \in [0, \tau_b]};$$

$$\Phi_{k-1} = \pi h \sum_{i=-\infty}^{k-1} a_i. \quad (11.212)$$

Сравним данный сигнал с сигналом бинарной частотной манипуляции, девиация частоты которой равна $\omega_d = \frac{\pi}{2\tau_b}$.

$$s_{bfsk}(t) = A_0 \cos\left(\left(\omega_0 + a_k \frac{\pi}{2} \frac{1}{\tau_b}\right)t\right) = \quad (11.213)$$

$$= A_0 \left[\cos(\omega_0 t) \cos\left(\frac{\pi}{2} \frac{1}{\tau_b} t\right) - a_k \sin(\omega_0 t) \sin\left(\frac{\pi}{2} \frac{1}{\tau_b} t\right) \right].$$

На рис. 11.30 приведены временные функции данных сигналов, модулированных одинаковой информационной последовательностью.

Как видно из рисунка, отличие двух сигналов заключается в наличии скачков фазы на π в сигнале с частотной манипуляцией.

Рассмотрим сигнал с OQPSK, у которого в качестве квадратурных несущих сигналов используются не гармонические колебания $\cos(\omega_0 t)$ и $\sin(\omega_0 t)$, а квадратурные биения вида $\cos(\omega_d t) \cos(\omega_0 t)$ и $\sin(\omega_d t) \sin(\omega_0 t)$:

$$s_{oqpsk}(t) = A_0 [I_{2k} \cos(\omega_d t) \cos(\omega_0 t) - I_{2k-1} \sin(\omega_d t) \sin(\omega_0 t)], \quad (11.214)$$

где $I_{2k} = \pm 1$ и $I_{2k-1} = \pm 1$ – потоки четных и нечетных элементов (по порядку следования) некоторого потока I , определяемого последовательностью передаваемых символов a_k . Элементы потоков I_{2k} и I_{2k-1} имеют удвоенную по сравнению с потоком I длительность и определенным образом сфазированы с колебаниями $\cos(\omega_d t)$ и $\sin(\omega_d t)$ (рис. 11.31).

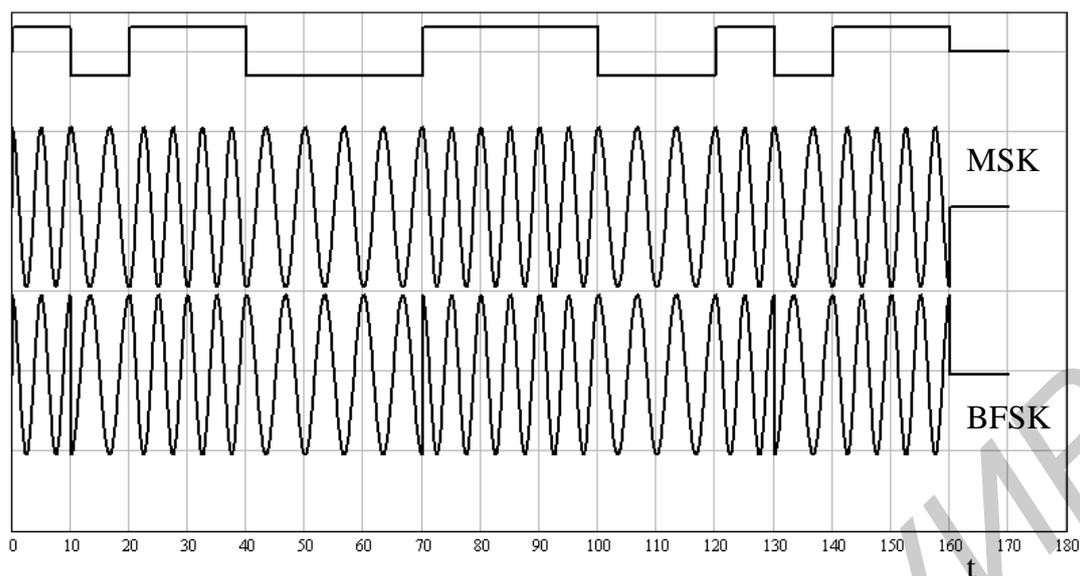


Рис. 11.30. Сравнение сигналов с MSK и BFSK

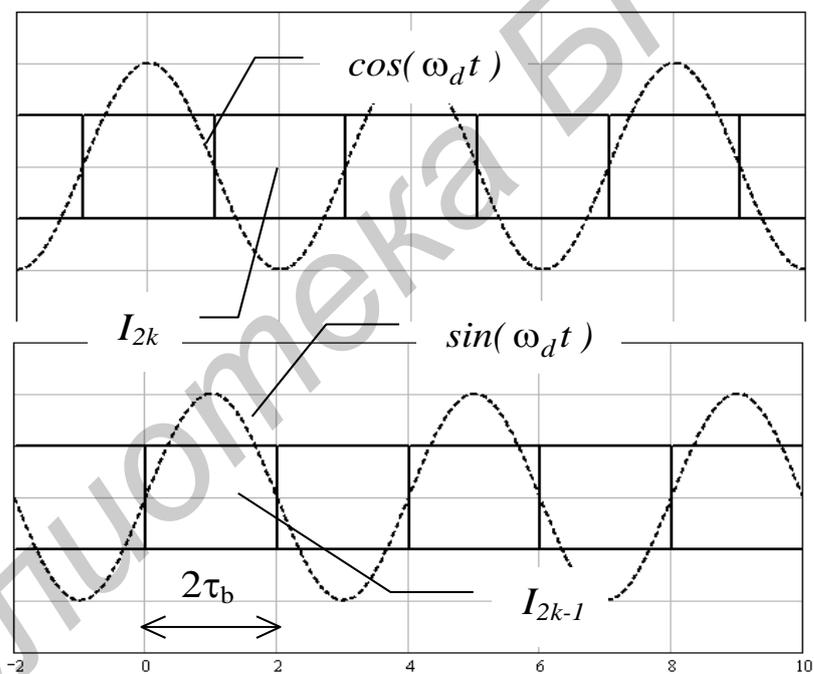


Рис. 11.31. Фазирование потоков I_{2k} и I_{2k-1}

Преобразуем выражение, отображающее сигнал с OQPSK, следующим образом:

$$s_{oqpsk}(t) = A_0 I_{2k} \left[\cos(\omega_d t) \cos(\omega_0 t) - \frac{I_{2k-1}}{I_{2k}} \sin(\omega_d t) \sin(\omega_0 t) \right], \quad (11.215)$$

или

$$s_{oqpsk}(t) = A_0 I_{2k} \left[\cos(\omega_d t) \cos(\omega_0 t) - I_{2k-1} I_{2k} \sin(\omega_d t) \sin(\omega_0 t) \right]. \quad (11.216)$$

Данный сигнал представляет собой BFSK колебание, модулированное результатом перемножения $I_{2k-1}I_{2k}$ и дополнительно манипулированное по фазе на 0 или π четными модулирующими символами I_{2k} .

Записанные выражения для сигнала с OQPSK справедливы для интервалов времени постоянства элементов I_{2k} и I_{2k-1} длительностью τ_b (соответствующего временному положению элементов исходного потока I).

Построим временные диаграммы для данного сигнала с OQPSK (рис. 11.32).

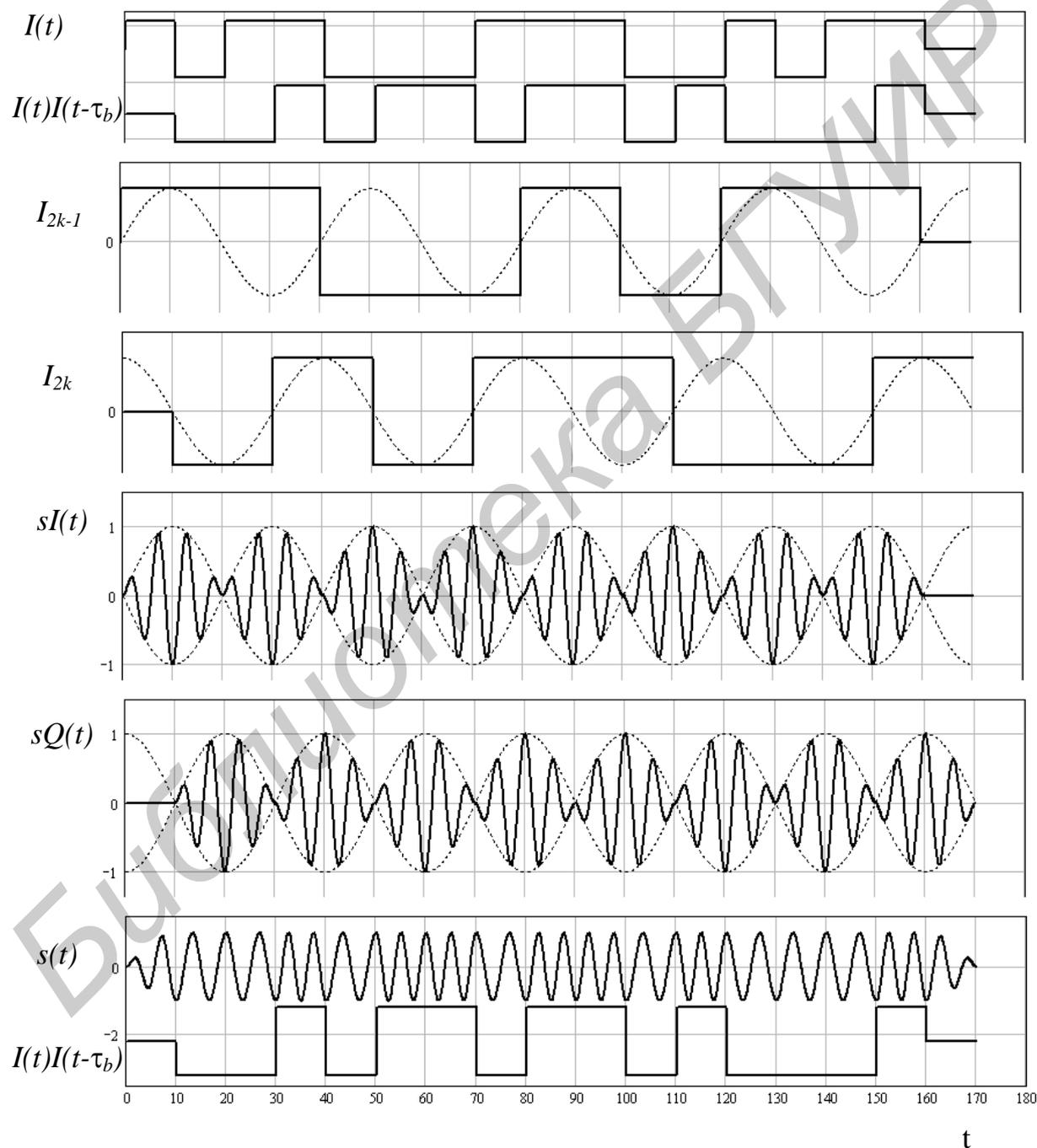


Рис. 11.32. MSK как OQPSK

Как видно из рисунка, сигнал $s_{oqpsk}(t)$ является сигналом с MSK, модулированным последовательностью, соответствующей последовательности, сформированной произведением следующего вида: $I(t)I(t - \tau_b)$.

Общеизвестно, что операция вида $I(t)I(t - \tau_b)$ соответствует процедуре преобразования кода линии вида NRZI в код линии NRZ. Таким образом, можно представить обобщенную схему формирования сигнала с MSK в виде структуры, изображенной на рис. 11.33.

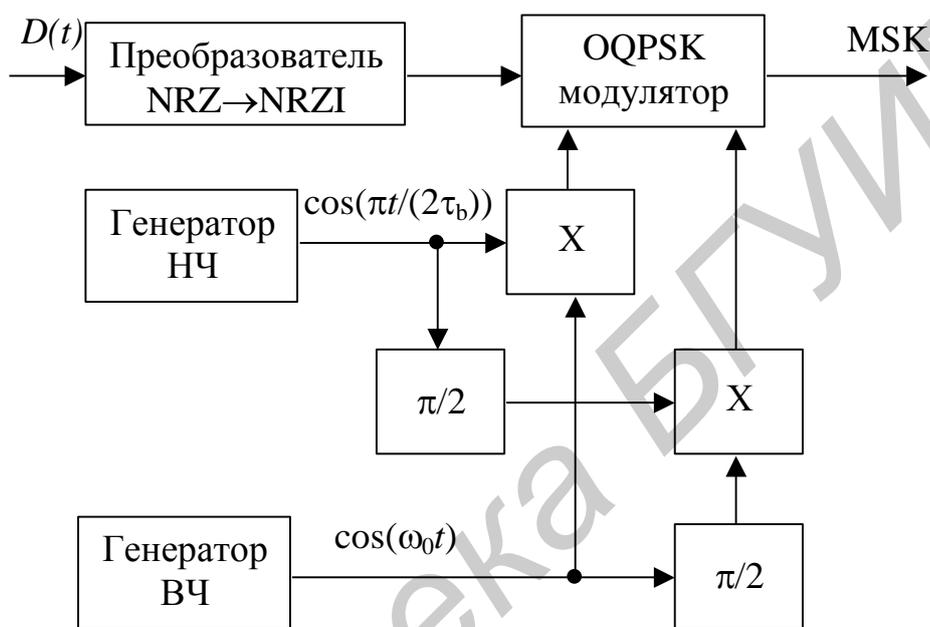


Рис. 11.33. Формирование сигнала с MSK OQPSK модулятором

Структурная схема транскодера линейных кодов NRZ-NRZI-NRZ показана на рис. 11.34, а пример кодовых последовательностей – на рис. 11.35.

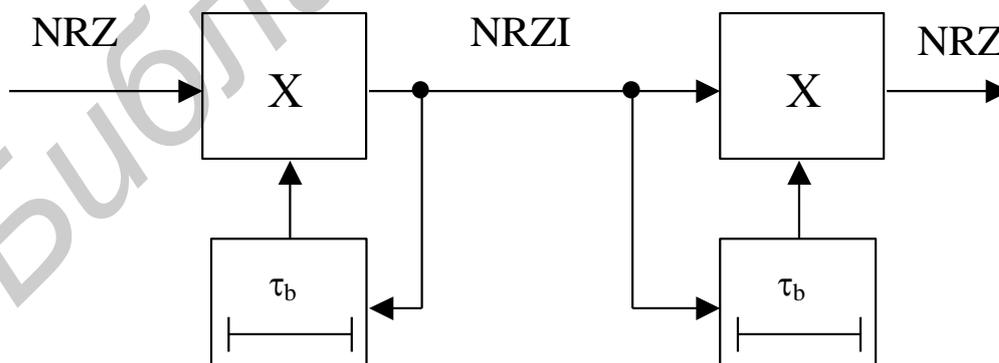


Рис. 11.34. Транскодер линейных кодов NRZ-NRZI-NRZ

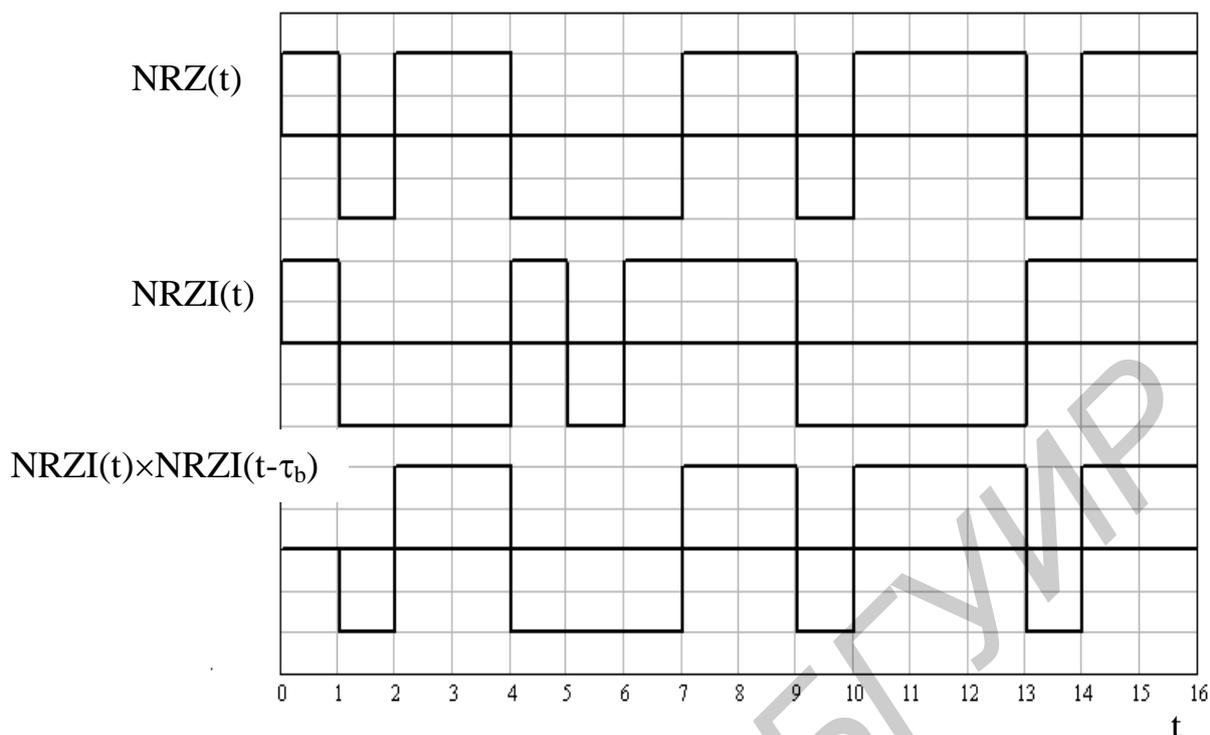


Рис.11.35. Коды линии семейства NRZ

Вероятность ошибки при когерентном приеме сигналов с MSK в точности совпадает с вероятностью ошибки при приеме сигналов с $\pm\pi/2$ -BPSK и равна

$$P_e^b = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right). \quad (11.217)$$

Поскольку сигнал с минимальной манипуляцией является сигналом CPM с полным откликом и по сути является сигналом частотной манипуляции с непрерывной фазой, то возможна некогерентная его демодуляция с использованием методов некогерентного приема сигналов с FSK.

11.7.3.5. Гауссовская минимальная манипуляция (GMSK)

Гауссовская минимальная манипуляция является одним из наиболее широко используемых методов CPM модуляции. С целью большего уменьшения величины спектральных составляющих модулированного сигнала вне основной полосы частот, по сравнению с сигналом просто минимальной манипуляции, при GMSK используется гауссовский предмодуляционный фильтр (гауссовская форма нормированного формирующего импульса частотной модуляции CPM). Обычно полоса пропускания предмодуляционного фильтра выбирается из условия $B_b < 1/\tau_b$, что соответствует случаю CPM с

неполным (парциальным) откликом. На рис. 11.36 приведены основные временные диаграммы, отражающие поведение фазового набега $\Phi(t, a)$ и мгновенной девиации частоты $\omega_d(t)$ в процессе модуляции информационными символами a_k при $B_b = 0,3/\tau_b$ и индексе модуляции $h = 0,5$, соответствующем случаю минимальной манипуляции.

Вероятность ошибки при приеме сигнала с GMSK несколько проигрывает вероятности ошибки приема сигналов с просто MSK и примерно равна

$$P_e^b \approx \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{\alpha E_b}{N_0}} \right), \quad (11.218)$$

$$\text{где } \alpha = \begin{cases} 0,68 & B_b \tau_b = 0,25 \\ 1 & B_b \tau_b \rightarrow \infty \end{cases}.$$

11.7.4. Квадратурная амплитудная модуляция (Quadrature Amplitude Modulation QAM)

Прежде чем рассматривать квадратурную амплитудную модуляцию, рассмотрим некоторые особенности передачи дискретной информации на примере сигнала с 8-PSK, сигнальное созвездие которого приведено на рис. 11.37. Предположим, что непрерывно передается только один символ, например «110».

Оптимальный демодулятор «вычисляет» расстояния между всеми разрешенными состояниями полезного сигнала и откликом от входной смеси переданного сигнала и шума

$$\vec{X} = \int_0^{\tau_c} X_I(t) dt + j \int_0^{\tau_c} X_Q(t) dt, \quad (11.219)$$

где $X_I(t)$ и $X_Q(t)$ – синфазная и квадратурная компоненты комплексной огибающей входной смеси сигнала и помехи.

Решение о передаваемом символе принимается на основании выбора ближайшего к текущему отклику от входной смеси переданного сигнала и шума разрешенного состояния сигнала. На рис. 11.37 отмечены области (сектора) принятия того или иного решения.

Необходимо отметить, что отклики от входной смеси переданного сигнала и шума концентрируются в окрестности «истинного» состояния сигнала, по крайней мере при достаточно больших отношениях сигнал/шум.

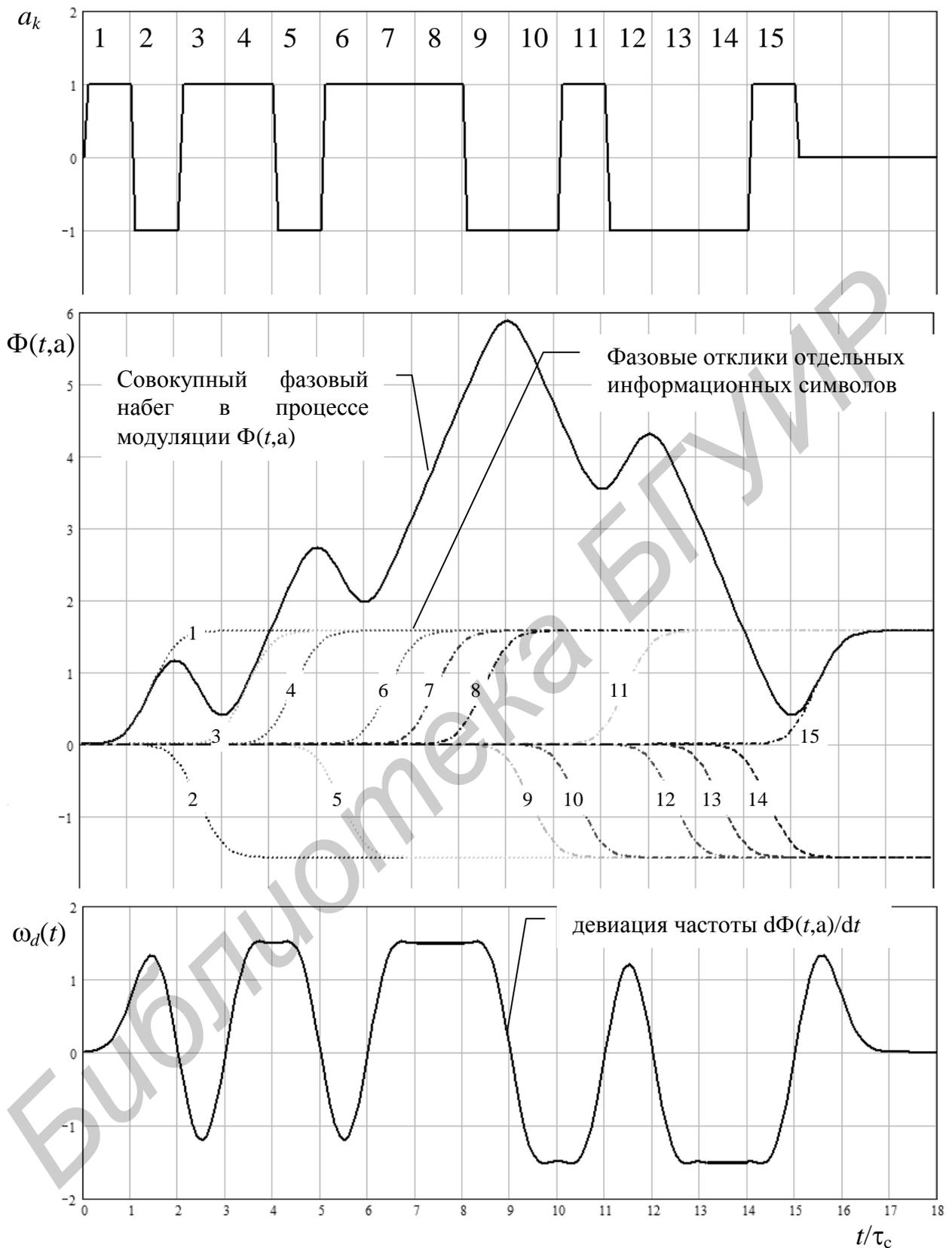


Рис. 11.36. Частотно-фазовые диаграммы сигнала с GMSK ($\tau_b = \tau_c = 1$; $B_b\tau_b = 0,3$; $h = 0,5$; $a_k = \pm 1$; $\tau_g \approx 3\tau_b$)

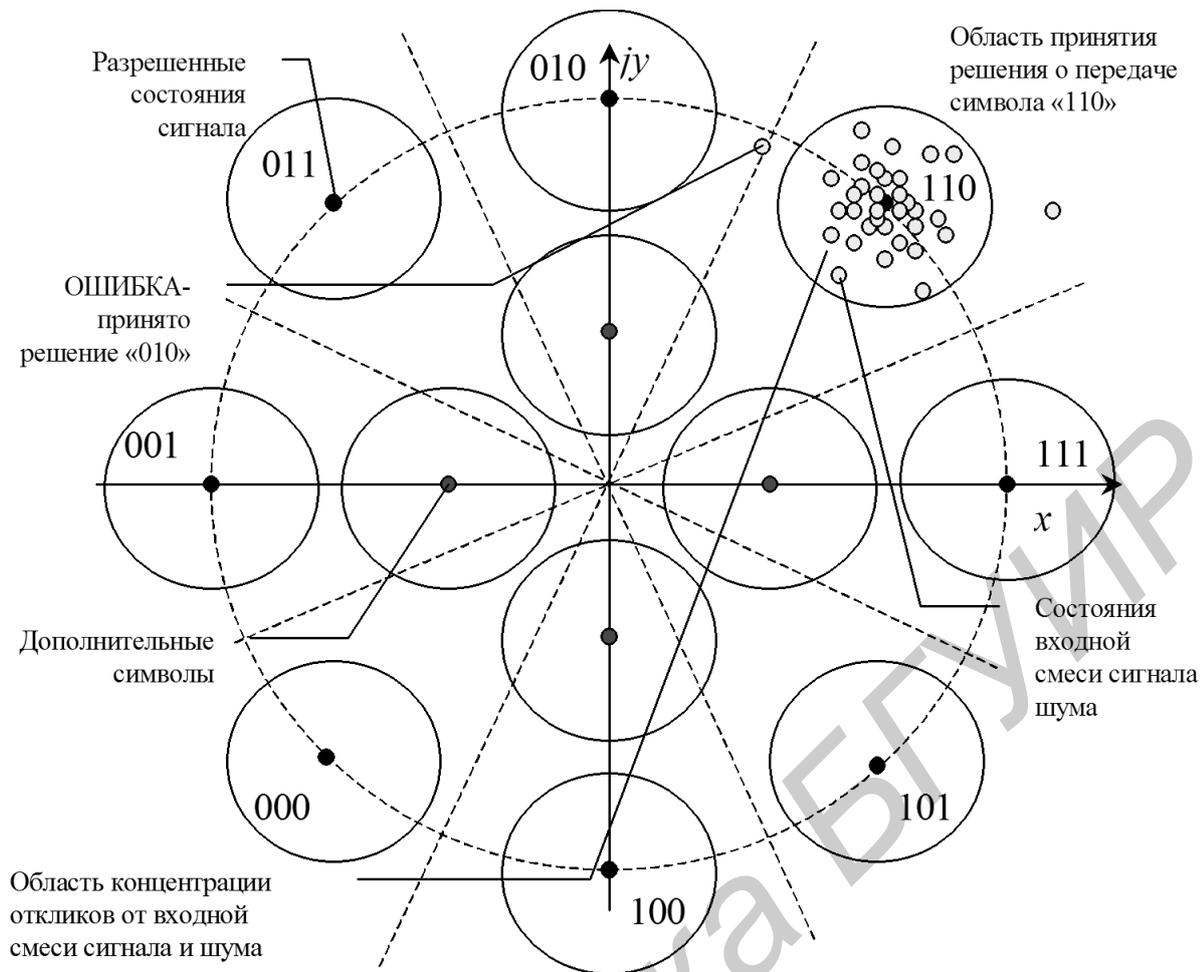


Рис. 11.37. Диаграмма принятия решений при оптимальной демодуляции сигнала с 8-PSK

Как видно из рис. 11.37, существует область в пределах окружности, ограничивающей энергетические (амплитудные) характеристики сигнала, в пределах которой возможно добавление дополнительных состояний сигнала (дополнительных информационных символов). При этом вероятность ошибки, приходящаяся на один символ, практически не увеличится, а энергия, приходящаяся на один символ, снизится, что позволит увеличить скорость передачи при фиксированной полосе частот и вероятности ошибки, либо повысить помехоустойчивость при фиксированной скорости передачи и вероятности ошибочного приема. Однако при приеме таких сигналов наряду с точностью фазовой синхронизации опорных сигналов квадратурного преобразователя частоты существенное значение приобретает амплитудное нормирование принимаемой смеси сигнала и шума к опорным сигналам (разрешенным состояниям полезного сигнала) демодулятора.

Подобное добавление информационных символов приводит к амплитудным и фазовым изменениям радиосигнала – переносчика информации. Поскольку любое амплитудно-фазовое состояние радиосигнала может быть получено амплитудной модуляцией двух квадратурных гармонических сиг-

налов, подобные сигналы получили наименование сигналов с квадратурной амплитудной модуляцией (манипуляцией) или QAM, при этом, как правило, при обозначении способа модуляции указывается число возможных состояний сигнала, например 32-QAM.

В настоящее время исследовано большое количество возможных множеств сигналов QAM с целью найти структуру, которая снизит вероятность появления ошибок при данном среднем отношении сигнал/шум. На рис. 11.38 показано несколько примеров множеств символов для $M = 4, 8$ и 16 . Обозначение (а, б, ...) соответствует сигналу, у которого а – количество сигналов во внутреннем круге; б – количество сигналов на следующем круге и т.д. В общем, правило построения «созвездия» сигнала QAM, обеспечивающего оптимальные характеристики (правило Кампопьяно–Глейзера/Camporiano–Glazer) можно сформулировать так: из бесконечного массива точек, плотно упакованных в регулярный массив или решетку, в качестве множества сигналов следует выбрать плотно упакованное подмножество из 2^k точек. В данном случае «оптимальный» означает минимальное отношение сигнал/шум, обеспечивающее заданную вероятность ошибки. В двухмерном пространстве сигналов оптимальная граница, окружающая массив сигнальных точек, стремится к окружности, а наиболее плотной является гексагональная решетка. По сравнению с квадратной кольцевая граница дает улучшение характеристик помехоустойчивости всего на 0,2 дБ.

Кодирование состояний информационными символами обычно осуществляется в соответствии с кодом Грея (соседние состояния должны отличаться друг от друга только одним элементом), что позволяет уменьшить вероятность ошибки, приходящейся на один элемент (бит) передаваемой информации. Пример подобного кодирования для «квадратной» 16-QAM приведен на рис. 11.39.

Вероятность ошибочного приема сигналов с квадратной M-QAM (кодирование кодом Грея) определяется следующим образом:

$$P_e^b = \frac{4}{\log_2(M)} \left(1 - \frac{1}{\sqrt{M}}\right) \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{3 \log_2(M)}{M-1} \frac{E_b}{2N_0}} \right). \quad (11.220)$$

В случае «неквадратной» QAM вероятность ошибки, приходящаяся на один информационный символ, удовлетворяет следующему условию:

$$P_e^c < (M-1) \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{d_{\min}^2}{4N_0}} \right), \quad (11.221)$$

где d_{\min}^2 – минимальное «расстояние» между разрешенными состояниями созвездия сигнала.

Прием сигналов с QAM осуществляется в соответствии с решающим правилом, описанным для случая многопозиционной фазовой манипуляции M-PSK, при этом каждое разрешенное состояние сигнала (информационный

символ) определяется комплексным числом, соответствующим изображающей точке на сигнальном созвездии. Кроме того, при формировании сигналов с QAM возможно использование формирующих импульсов ($g_F(t) = 0 \quad \forall t \notin [0, \tau_c]$; τ_c – длительность информационного символа) для контроля спектральных характеристик формируемого сигнала.

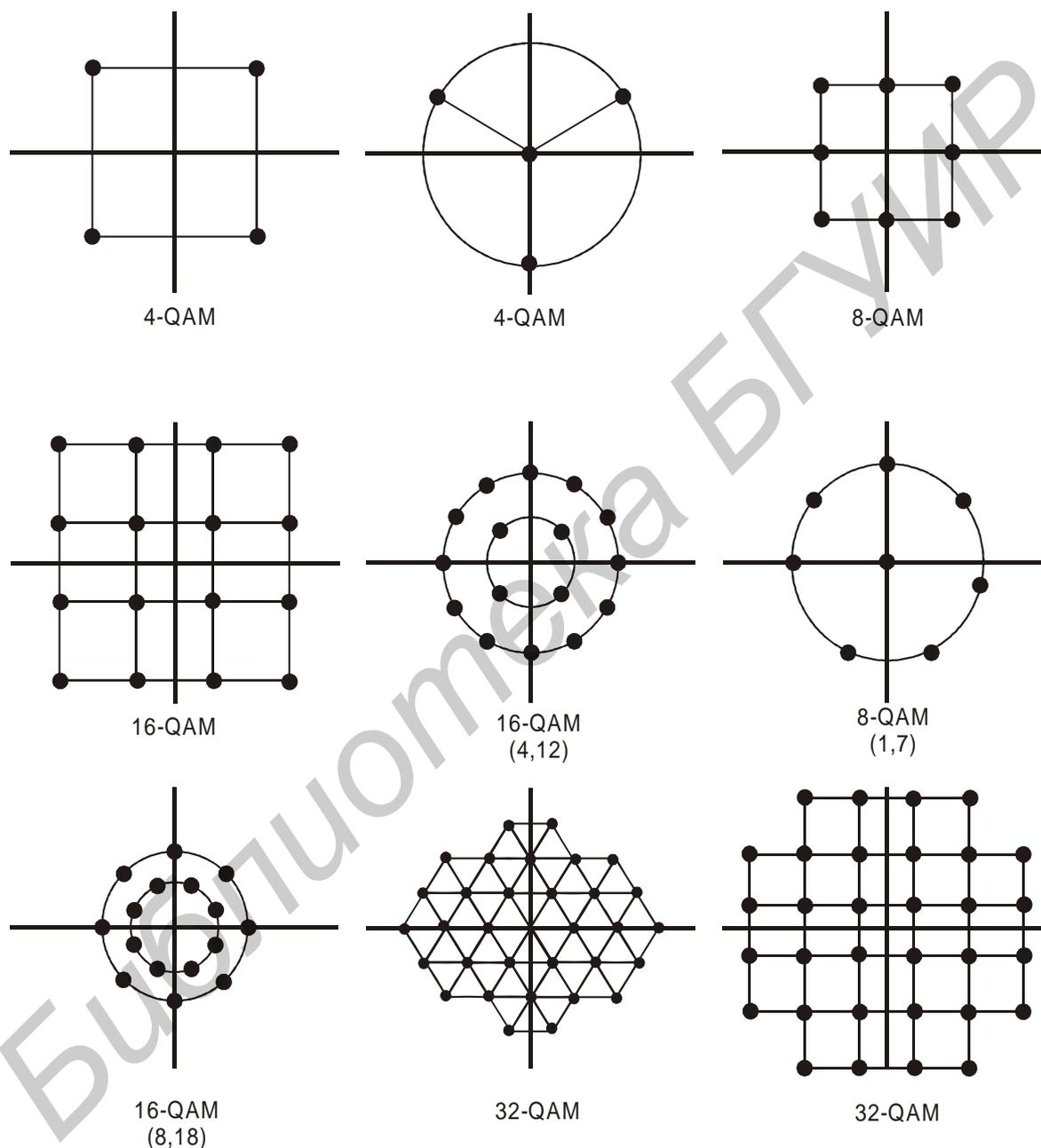


Рис. 11.38. Примеры сигнальных созвездий квадратурной амплитудной манипуляции

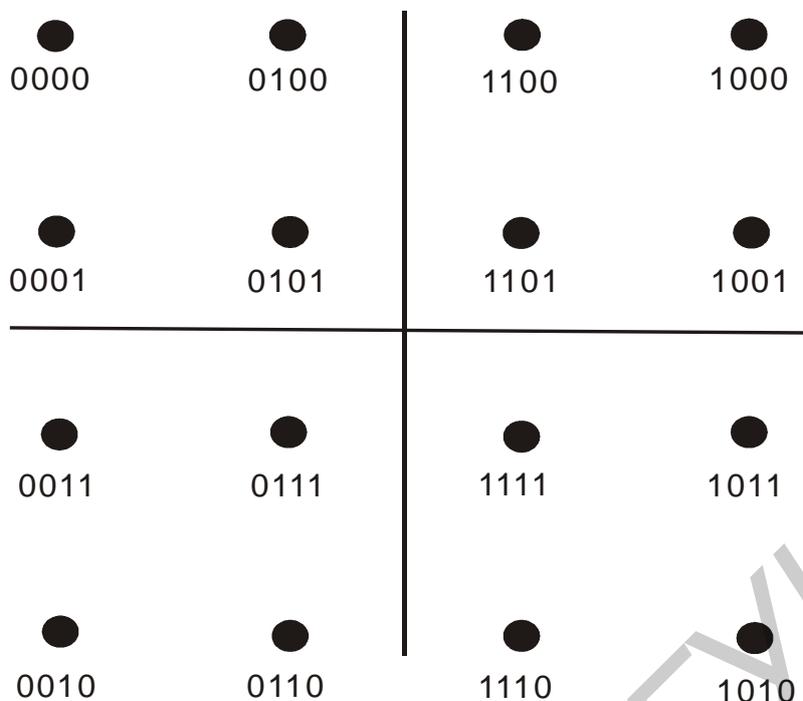


Рис. 11.39. Примеры кодирования кодом Грея состояний сигнала 16-QAM

11.8. Спектральные характеристики излучения при дискретной полосовой модуляции

11.8.1. Основные принципы определения спектральных характеристик сигналов с дискретной полосовой модуляцией

В большинстве систем цифровой радиосвязи имеющаяся в распоряжении полоса частот ограничена. Как следствие, проектировщик системы должен рассмотреть, как ограничение полосы частот канала влияет на выбор техники модуляции, используемой при передаче информации. Из этих соображений для нас важно определить спектральные характеристики радиосигналов с цифровой модуляцией, рассмотренных в подразд. 11.7.

Поскольку информационная последовательность случайная, сигнал цифровой модуляции является случайным процессом. Нас интересует спектральная плотность мощности (распределение мощности по частоте) такого процесса. Исходя из спектральной плотности мощности, можно определить полосу частот канала, необходимую для передачи модулированного сигнала.

Определим общие подходы к нахождению спектральной плотности мощности сигналов с цифровой модуляцией.

Известно, что в соответствии с теоремой Винера–Хинчина спектральная плотность мощности стационарного случайного процесса (статистические характеристики которого не изменяются во времени) связана преобразованием Фурье с его автокорреляционной функцией.

$$S(\omega) = \int_{-\infty}^{\infty} K(\tau) e^{-j\omega\tau} d\tau. \quad (11.222)$$

Автокорреляционная функция случайного процесса $x(t)$ в свою очередь определяется следующим образом:

$$K(\tau) = \langle x(t)x(t-\tau) \rangle, \quad (11.223)$$

где $\langle \cdot \rangle$ – обозначение операции статистического усреднения по ансамблю возможных реализаций случайного процесса. Для эргодических случайных процессов операция усреднения по ансамблю реализаций может быть заменена усреднением по времени одной теоретически бесконечно длинной реализации:

$$K(\tau) = \langle x(t)x(t-\tau) \rangle = \frac{1}{T} \int_0^T x(t)x(t-\tau) dt. \quad (11.224)$$

Определим некоторые полезные для последующего использования при анализе спектральных характеристик сигнала соотношения.

Спектральная плотность мощности суммы двух случайных процессов

$$\begin{aligned} S_{x+y}(\omega) &= \int_{-\infty}^{\infty} K_{x+y}(\tau) e^{-j\omega\tau} d\tau = \\ &= \int_{-\infty}^{\infty} \langle (x(t) + y(t))(x(t-\tau) + y(t-\tau)) \rangle e^{j\omega\tau} = \\ &= \int_{-\infty}^{\infty} \langle x(t-\tau) + y(t) + x(t)y(t-\tau) + x(t-\tau)y(t) \rangle e^{j\omega\tau} = \\ &= S_x(\omega) + S_y(\omega) + S_{xy}(\omega) + S_{yx}(\omega). \end{aligned} \quad (11.225)$$

равна сумме спектральных плотностей мощности исходных процессов плюс их взаимные спектральные плотности $S_{xy}(\omega)$ и $S_{yx}(\omega)$. В случае если исходные случайные процессы являются некоррелированными ($\langle x(t)y(t-\tau) \rangle = 0$ и $\langle x(t-\tau)y(t) \rangle = 0$), то спектральная плотность суммы (разности) двух случайных процессов равна сумме их спектральных плотностей.

Спектральная плотность мощности произведения двух случайных процессов

$$S_{xxy}(\omega) = \int_{-\infty}^{\infty} K_{xxy}(\tau) e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} \langle x(t)x(t-\tau)y(t)y(t-\tau) \rangle e^{-j\omega\tau} d\tau. \quad (11.226)$$

В случае если случайные процессы являются независимыми, можно сделать следующие преобразования:

$$S_{xxy}(\omega) = \int_{-\infty}^{\infty} \langle x(t)x(t-\tau)y(t)y(t-\tau) \rangle e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} \langle x(t)x(t-\tau) \rangle \langle y(t)y(t-\tau) \rangle e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} K_x(\tau) K_y(\tau) e^{-j\omega\tau} d\tau. \quad (11.227)$$

Учитывая то, что преобразование Фурье от произведения двух функций равно свертке Фурье образов этих функций, окончательно имеем

$$S_{xxy} = \int_{-\infty}^{\infty} S_x(\Omega) S_y(\omega - \Omega) d\Omega. \quad (11.228)$$

Необходимо заметить, что рассматриваемые выше спектральные плотности действительных случайных процессов являются четными функциями частоты (т.е. определены в области как положительных, так и отрицательных частот); распределение мощности реального случайного процесса определяется как удвоенная спектральная плотность в положительной области частот; условие некоррелированности выполняется всегда, если случайные процессы не пересекаются во времени или не имеют совпадающих частотных компонент (не пересекаются в частотной области).

При спектральном анализе радиосигналов с цифровой модуляцией следует различать две их разновидности.

К первым из них относятся сигналы с так называемой линейной модуляцией. К таким сигналам относятся все сигналы, у которых информационная модуляция входит линейно во временную функцию сигнала, например, $s(t) = A_0 C_k \cos(\omega_0 t)$, где $C_k = \pm 1$ по закону передаваемой информации. К сигналам данной группы формируются с использованием рассмотренных выше схем модуляции (BPSK, $\pm\pi/2$ -BPSK, QPSK, OQPSK, M-PSK, M-QAM), к ним также могут быть отнесены сигналы FSK, M-FSK при условии прямоугольной формирующей функции частотной модуляции.

Ко вторым относятся сигналы с так называемой нелинейной модуляцией. Данные сигналы формируются схемой модуляции CPM, в которых информационный параметр нелинейно входит во временную функцию модулированного радиосигнала (в частности определяет нелинейные изменения фа-

зы радиосигнала во времени). Ввиду достаточной сложности рассматривать их не будем.

11.8.2. Спектральные характеристики сигналов с линейной модуляцией

Выше мы определили, что спектральная плотность мощности случайного процесса полностью определяется его автокорреляционной функцией (11.222).

Определим автокорреляционную функцию модулированного радиосигнала через его комплексное представление:

$$K(\tau) = \frac{1}{2} \operatorname{Re} \left\langle \bar{s}(t) \bar{s}^*(t + \tau) \right\rangle, \quad (11.229)$$

где $\bar{s}(t) = \bar{S}(t)e^{j\omega_0 t}$ и $\bar{S}(t)$ комплексная огибающая модулированного радиосигнала. С учетом этого автокорреляционная функция записывается следующим образом:

$$K(\tau) = \frac{1}{2} \operatorname{Re} \left\langle \bar{S}(t) \bar{S}^*(t + \tau) e^{j\omega_0 \tau} \right\rangle = \operatorname{Re} \left\{ K_S(\tau) e^{j\omega_0 \tau} \right\}, \quad (11.230)$$

где $K_S(\tau) = \frac{1}{2} \left\langle \bar{S}(t) \bar{S}^*(t + \tau) \right\rangle$ – автокорреляционная функция комплексной огибающей.

Осуществив преобразование Фурье от автокорреляционной функции модулированного радиосигнала, получим следующее выражение для спектральной плотности:

$$\begin{aligned} S(\omega) &= \frac{1}{2} \int_{-\infty}^{\infty} S_S(\omega) (\delta(\omega - \omega_0) + \delta(\omega + \omega_0)) d\omega = \\ &= \frac{1}{2} [S_S(\omega - \omega_0) + S_S(\omega + \omega_0)] \end{aligned} \quad (11.231)$$

где $S_S(\omega)$ – спектральная плотность мощности комплексной огибающей модулированного сигнала. Таким образом, для определения спектральной плотности мощности модулированного сигнала достаточно определить автокорреляционную функцию и спектральную плотность мощности комплексной огибающей.

Для сигналов с линейной модуляцией комплексная огибающая в общем виде записывается следующим образом:

$$\bar{S}(t) = \sum_{k=-\infty}^{\infty} g_F(t - k\tau_c) I_k, \quad (11.232)$$

где $g_F(t) = 0 \quad \forall t \notin [0, \tau_c]$ – формирующий импульс; τ_c – длительность информационного символа (символа модуляции); k – текущий номер передаваемого информационного символа; \vec{I}_k – в общем случае комплексное число, представляющее информационный символ, принимающее конечный ряд известных значений и являющееся константой на интервале времени действия информационного символа.

Например,

$$\text{BPSK} \rightarrow \vec{I}_k = A_0 (\cos(C_k \Theta) + j \sin(C_k \Theta));$$

$$\text{M-PSK} \rightarrow \vec{I}_k = A_0 (\cos(\Theta_k) + j \sin(\Theta_k));$$

$$\text{QPSK} \rightarrow \vec{I}_k = \frac{A_0}{\sqrt{2}} (C_k^I + j C_k^Q);$$

$$\text{DBPSK, DQPSK} \rightarrow \vec{I}_k = A_0 (\cos(\varphi_k) + j \sin(\varphi_k));$$

$$\text{M-QAM} \rightarrow \vec{I}_k = A_m^I + j A_m^Q;$$

где C_k, C_k^I, C_k^Q – информационные символы $\{\pm 1\}$;

Θ_k – разрешенные фазовые состояния сигнала; φ_k – разрешенные фазовые переходы; A_m^I и A_m^Q амплитуды квадратурных компонент, определяющие место образа сигнала в сигнальном созвездии.

Определим автокорреляционную функцию комплексной огибающей:

$$\begin{aligned} K_S(\tau) &= \frac{1}{2} \langle \vec{S}(t) \vec{S}^*(t-\tau) \rangle = \frac{1}{2} \left\langle \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g_F(t-k\tau_c) \vec{I}_k g_F(t-m\tau_c-\tau) \vec{I}_m^* \right\rangle = \\ &= \frac{1}{2} \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g_F(t-k\tau_c) g_F(t-m\tau_c-\tau) \langle \vec{I}_k \vec{I}_m^* \rangle. \end{aligned} \quad (11.233)$$

Для упрощения будем считать, что модулирующий информационный поток не содержит избыточности. Это означает, что все элементы \vec{I}_k равновероятны и не зависят друг от друга. На основании этого предположения имеем $\langle \vec{I}_k \vec{I}_n^* \rangle = 0 \Big|_{k \neq n}$. Тогда выражение для автокорреляционной функции переписывается следующим образом:

$$K_S(\tau) = \frac{1}{2} \sum_{k=-\infty}^{\infty} g_F(t-k\tau_c) g_F(t-k\tau_c-\tau) \langle |\vec{I}_k|^2 \rangle. \quad (11.234)$$

Как видно из полученного выражения, автокорреляционная функция комплексной огибающей является функцией не только временного сдвига τ , но и функцией абсолютного времени t . Это вызвано тем, что случайный процесс $\vec{S}(t)$ является циклостационарным (или периодически стационарным) с периодом цикла, равным τ_c , т. к. функция

$\sum_{k=-\infty}^{\infty} g_F(t - k\tau_c)g_F(t - k\tau_c - \tau)$ является периодичной по переменной t с периодом τ_c . Для получения спектральной плотности такого процесса необходимо устранить зависимость от абсолютного времени путем усреднения автокорреляционной функции по времени на ее периоде. Произведя данное усреднение, получим

$$\begin{aligned} K_S(\tau) &= \frac{1}{2} \frac{1}{\tau_c} \sum_{k=-\infty}^{\infty} \int_0^{\tau_c} g_F(t - k\tau_c)g_F(t - k\tau_c - \tau) dt \langle |\bar{I}_k|^2 \rangle = \\ &= \frac{1}{2} \frac{1}{\tau_c} \sum_{k=-\infty}^{\infty} \int_{k\tau_c}^{k\tau_c + \tau_c} g_F(t)g_F(t - \tau) dt \langle |\bar{I}_k|^2 \rangle = \frac{1}{2\tau_c} \int_{-\infty}^{\infty} g_F(t)g_F(t - \tau) dt \langle |\bar{I}_k|^2 \rangle. \end{aligned} \quad (11.235)$$

Учитывая, что $\frac{1}{2} \langle |\bar{I}_k|^2 \rangle = \sigma^2$ – есть дисперсия информационных компонент \bar{I}_k (за исключением сигналов с М-QAM $\sigma^2 = A_0^2/2$), запишем окончательное выражение для автокорреляционной функции комплексной огибающей:

$$K_S(\tau) = \frac{\sigma^2}{\tau_c} \int_{-\infty}^{\infty} g_F(t)g_F(t - \tau) dt. \quad (11.236)$$

Таким образом, спектральная плотность мощности комплексной огибающей модулированного сигнала $S_S(\omega)$ полностью определяется видом формирующего импульса $g_F(t)$:

$$S_S(\omega) = |G(\omega)|^2, \quad (11.237)$$

где $G(\omega)$ – Фурье-образ (результат прямого преобразования Фурье) функции $g_F(t)$.

Окончательно имеем спектральную плотность мощности модулированного радиосигнала

$$S(\omega) = \frac{1}{2} \left[|G(\omega - \omega_0)|^2 + |G(\omega + \omega_0)|^2 \right]. \quad (11.238)$$

Таким образом, выбор соответствующего вида формирующего импульса полностью определяет спектральные характеристики итогового радиосигнала.

11.8.3. Основные виды формирующих импульсов

В настоящее время наиболее широкое распространение получили два вида формирующих импульсов.

Первый из них – прямоугольный формирующий импульс

$$g_F(t) = \begin{cases} 1, & |t| \leq \tau_c \\ 0 & \end{cases}. \quad (11.239)$$

Второй вид формирующих импульсов объединяет группу импульсов, получаемых путем пропускания прямоугольных импульсов через особый фильтр нижних частот, называемый фильтром Найквиста, или фильтром с характеристикой типа приподнятого косинуса. Амплитудно-частотная характеристика фильтра Найквиста записывается следующим образом:

$$H(f) = \begin{cases} 1, & |f| \leq \frac{1-\alpha}{2\tau_c}, \\ \frac{1}{2} \left[1 + \cos \left(\frac{\pi\tau_c}{\alpha} \left(|f| - \frac{1-\alpha}{2\tau_c} \right) \right) \right], & \frac{1-\alpha}{2\tau_c} < |f| \leq \frac{1+\alpha}{2\tau_c}, \\ 0, & \frac{1+\alpha}{2\tau_c} < |f|, \end{cases} \quad (11.240)$$

где $\alpha = 2\tau_c \Delta F$; ΔF – полоса пропускания фильтра по уровню минус 3 дБ.

Когда $\alpha = 0$, амплитудно-частотная характеристика (АЧХ) фильтра становится идеально прямоугольной, при $\alpha = 1$ АЧХ фильтра представляет собой в чистом виде приподнятую косинусоиду

$$H(f) = \begin{cases} \frac{1}{2} [1 + \cos(\pi\tau_c f)], & |f| \leq \frac{1}{\tau_c}, \\ 0, & |f| > \frac{1}{\tau_c}. \end{cases} \quad (11.241)$$

Импульсная характеристика (отклик на воздействие в виде дельта функции) такого фильтра описывается выражением

$$h(t) = \frac{\sin\left(\frac{\pi t}{\tau_c}\right) \cos\left(\frac{\pi\alpha t}{2\tau_c}\right)}{\frac{\pi t}{\tau_c} \left(1 - \frac{4\alpha^2 t^2}{\tau_c^2}\right)}. \quad (11.242)$$

Форма формирующего импульса полностью соответствует импульсной характеристике формирующего фильтра.

На рис. 11.40 приведены временные функции основных формирующих импульсов. Как видно из рисунка, длительность формирующего импульса Найквиста значительно превосходит длительность модулирующего символа, что вызывает наложение откликов от различных модулирующих символов друг на друга.

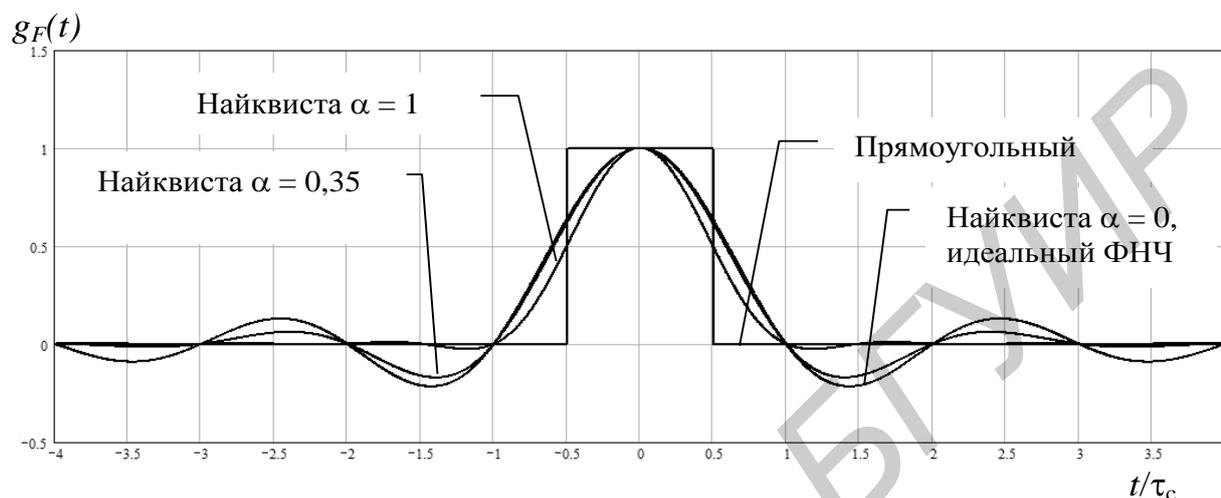


Рис. 11.40. Различные виды формирующих импульсов

Рассмотрим спектральные характеристики указанных выше формирующих импульсов (см. рис. 11.41). Как видно из рисунка, использование формирующих импульсов, отличающихся от прямоугольных, позволяет повысить спектральную эффективность метода модуляции (спектр модулированного радиосигнала соответствует спектру формирующего импульса при условии сдвига частотной оси в область частоты f_0) с 0,5 бит (символ)/Гц для случая прямоугольного формирующего импульса до практически теоретического предела 1 бит (символ)/Гц для формирующего фильтра Найквиста ($\alpha = 0$).

Кроме того, использование более «сложных» формирующих импульсов позволяет значительно (на десятки децибел) снизить излучение вне основной полосы частот.

Использование формирующих импульсов Найквиста требует организации значительной «памяти» о предыдущих символах модуляции в модулированном сигнале (в идеале формирующий импульс Найквиста имеет ненулевые значения на интервале времени $[-\infty, \infty]$).

На рис. 11.42 иллюстрируется эффект от усечения длительности формирующего импульса Найквиста всего восемью интервалами τ_c . Если внимательно рассмотреть области вне основной полосы частот, то видно, что усечение длительности формирующего импульса значительно повышает внеполосное излучение. Вместе с тем даже увеличенный уровень внеполосного излучения значительно ниже аналогичных уровней, реализуемых с по-

мощью прямоугольных формирующих импульсов. В связи с этим современные системы цифровой радиосвязи в условиях жесткого ограничения занимаемой полосы частот, как правило, используют спектрально-эффективные многопозиционные методы модуляции с формирующими импульсами (формирующими фильтрами) Найквиста со значением параметра $\alpha = 0,35$, признанным стандартным.

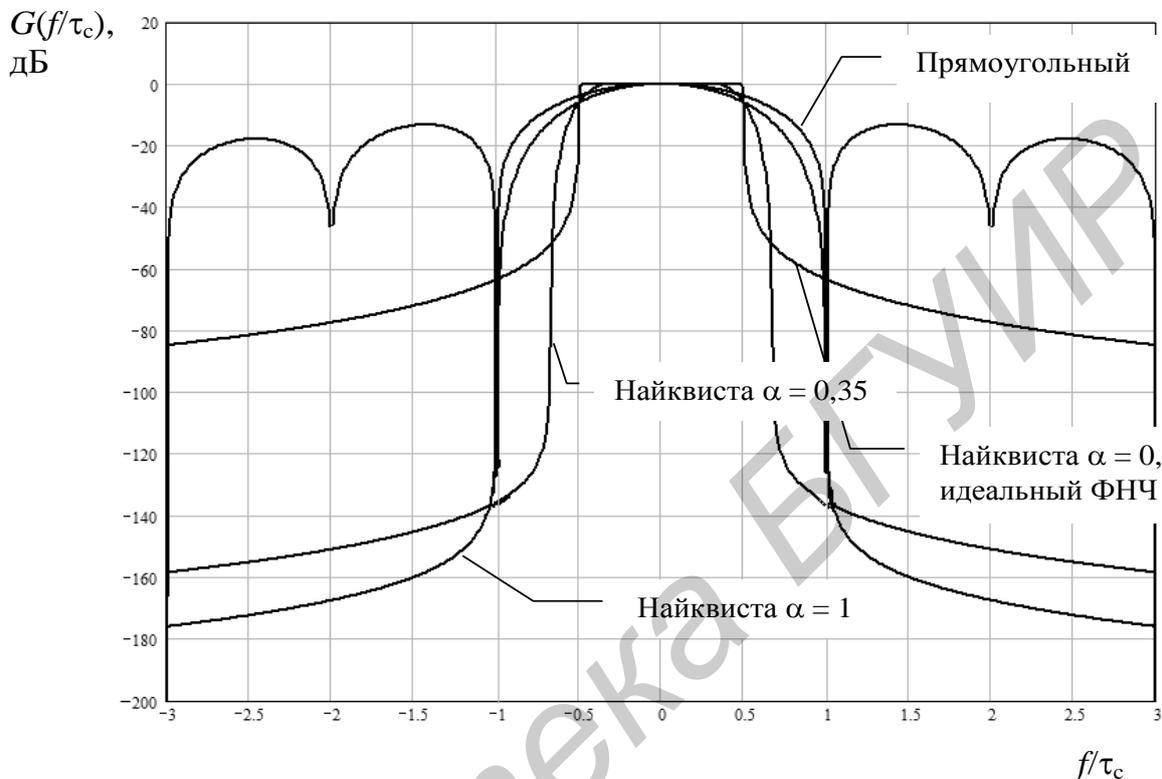


Рис. 11.41. Спектральные характеристики различных формирующих импульсов

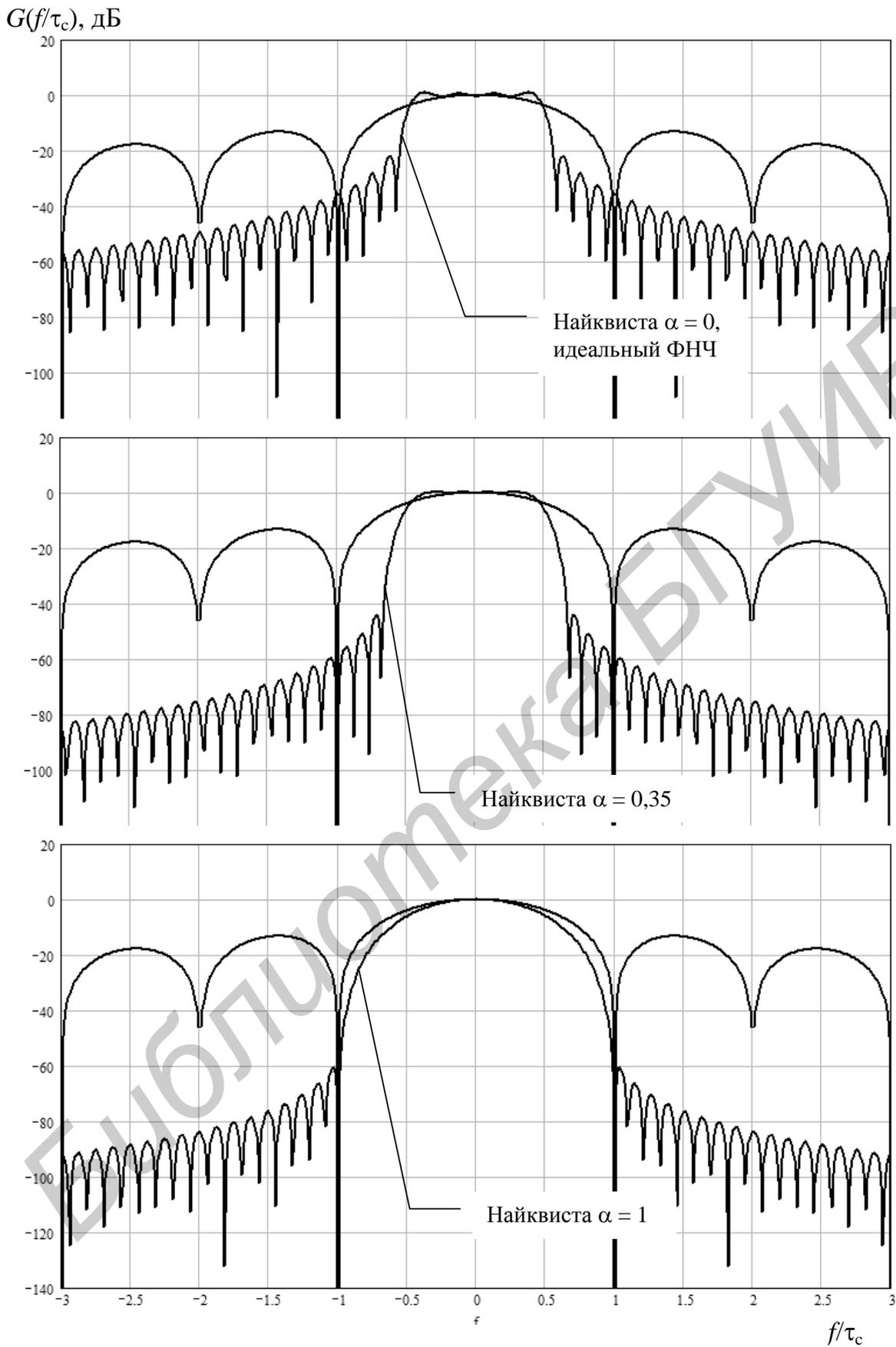


Рис. 11.42. Влияние усечения во временной области формирующего импульса на уровень излучения вне основной полосы частот

11.9. Сигнально-кодовые конструкции (Trellis Coded Modulation)

11.9.1. Место и роль сигнально-кодовых конструкций в цифровой радиосвязи

Приступая к рассмотрению основных принципов составления сигнально-кодовых конструкций, следует заметить, что принципиально любое наложение информации на физический носитель следует трактовать как сигнально-кодовую конструкцию. Таким образом, термин «сигнально-кодовая» конструкция не является самостоятельным. В связи с этим необходимо уточнение, что имеется в виду, когда используется данная терминология.

Для этого обратимся к обобщенной структуре системы передачи информации (рис. 11.43, представлена только передающая часть). Носителем любой информации является физический объект – он же источник информации. Поскольку свойства источника, в которых заложена информация, весьма разнообразны и, как правило, плохо пригодны для унифицированной передачи, информация (физический объект-носитель), поступающая от источника, преобразуется первичным преобразователем в вид, удобный для осуществления унифицированных процедур обработки и передачи. По сути, первичный преобразователь преобразует один вид физического носителя (неудобный в использовании) в другой вид носителя, который удобно использовать в дальнейшем.

Далее преобразованный физический носитель (сигнал) поступает в кодер источника. Назначение кодера источника – выделения из сигнала «чистой» информации, т.е. той информации, которую необходимо передать. Необходимость в кодере источника возникает из-за того, что сигнал обычно содержит избыточные сведения, придающие «окраску» передаваемой информации, например при передаче голосовой команды, как правило, не важен тембр и прочие особенности голоса.

В случае использования для передачи информации двоичных символов «0» и «1» кодер источника, в идеале, должен формировать на своем выходе последовательность двоичных символов, каждый из которых несет один бит информации (т.е. все элементы последовательности должны быть взаимонезависимыми и равновероятными), кроме того, общее количество информации, содержащееся в данной последовательности, должно точно соответствовать количеству информации, которое необходимо передать. Это означает, что искажение любого из символов данной последовательности наносит одинаковый ущерб передаваемой информации.

Сформированная кодером сигнала последовательность отражается модулятором (в соответствии с принятой схемой модуляции) в радиосигнал – носитель информации. Таким образом, формируется сигнально-кодовая конструкция (код кодера источника-радиосигнал модулятора). Никаких особенностей формирования сигнально-кодовой конструкции не наблюдается при использовании бинарных схем модуляции. В этом случае каждому двоичному символу ставится в соответствие свой сигнал-носитель (два символа и два сигнала). При таком соответствии не нарушается правило – расстояние меж-

ду передаваемыми символами (расстояние по Хеммингу) должно быть согласовано с расстоянием между сигналами (расстояние Евклида). То есть меньшее расстояние по Хеммингу – меньшее расстояние по Евклиду и наоборот.

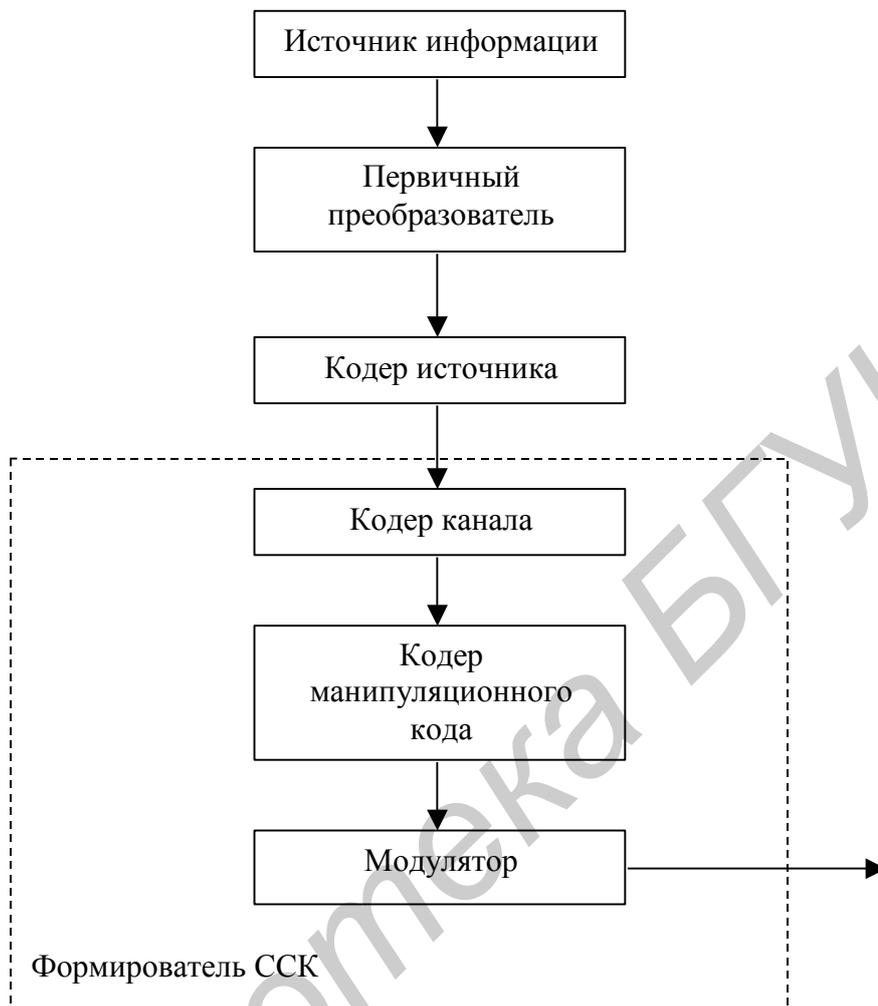


Рис. 11.43. Обобщенная структура передающей части канала передачи информации

Необходимость выполнения данного правила можно пояснить следующим образом. Как известно из предыдущего материала, вероятность того, что вместо переданного символа при демодуляции будет принято решение о передаче другого символа тем меньше, чем больше расстояние по Евклиду между сигналами, представляющими данные символы. Вместе с тем величина ущерба передаваемой информации искажением одного символа другим будет тем больше, чем больше расстояние по Хеммингу между данными символами. Из этого вытекает, что большему расстоянию по Хеммингу между символами должно соответствовать большее расстояние по Евклиду между представляющими их сигналами (большой ущерб с меньшей вероятностью).

При бинарной модуляции модулятор использует поочередно по одному элементу из сформированного кодером источника потока символов. Если

следующие друг за другом символы одинаковы, то расстояние по Хеммингу между ними равно нулю (расстояние по Хеммингу определяется как число позиций, в которых отличаются символы одинаковой длины или разрядности, например, между шестизначными символами 010110 и 110101 расстояние по Хеммингу равно трем). При этом модулятор последовательно излучает один и тот же сигнал $s^i(t)$, например, при BPSK $A_0 \cos(\omega_0 t)$, расстояние по Евклиду между последовательно излученными сигналами равно нулю (расстояние по Евклиду определяется как расстояние между точками, отображающими радиосигнал на комплексной плоскости). Сигнал $A_0 \cos(\omega_0 t)$ отображается точкой A_0 на действительной оси. В случае если следующие друг за другом символы различны, то расстояние по Хеммингу между ними равно 1. При этом модулятор последовательно излучает, например, сигнал $A_0 \cos(\omega_0 t)$ для первого символа и сигнал $-A_0 \cos(\omega_0 t)$ (точка $-A_0$ на действительной оси) для второго символа. Расстояние по Евклиду между данными сигналами равно $2A_0$. Другие ситуации невозможны.

Повышение помехоустойчивости возможно внесением изменений в схему модуляции (увеличение A_0 , увеличение энергии, приходящейся на один передаваемый бит) или применением канального кодера (помехоустойчивого кодирования). Канальный кодер вносит избыточность в «чистую» информационную последовательность, позволяющую обнаруживать и исправлять ошибки. Данная избыточность проявляется в возникновении дополнительных элементов в модулирующей последовательности. При этом возможны два изменения в процессе модуляции. Первое: сохраняется скорость модуляции (падает скорость передачи информации, вместе с информационными битами необходимо передавать еще и избыточные). Второе: сохраняется скорость передачи информации (увеличивается скорость модуляции, необходимо успевать передавать за то же время как информационные, так и избыточные биты). В первом варианте полоса частот, занимаемая радиосигналом, сохраняется, во втором – происходит ее пропорциональное увеличение. В обоих вариантах заметна существенная эффективность использования помехоустойчивого кода.

Однозначность формирования сигнально-кодовых конструкций пропадает при использовании многопозиционных схем модуляции (M-PSK, M-QAM и пр.). Необходимость использования многопозиционных методов модуляции вызывается требованиями увеличения скорости передачи информации при фиксированной (часто очень ограниченной) полосе частот, отводимой для модулированного радиосигнала. В этом случае каждый из M возможных несущих сигналов ставится в соответствие группе из $k = \log_2(M)$ символов потока, формируемого кодером источника или канальным кодером. В этом случае принципиальной становится процедура модуляционного кодирования (Mapping Scheme).

На рис. 11.44 приведены две схемы модуляционного кодирования сигнала с 8-PSK. Как видно из рисунка, произвольный выбор схемы модуляци-

онного кодирования не обеспечивает выполнения условия согласований расстояний по Хеммингу для информационных символов и по Евклиду для представляющих их сигналов.

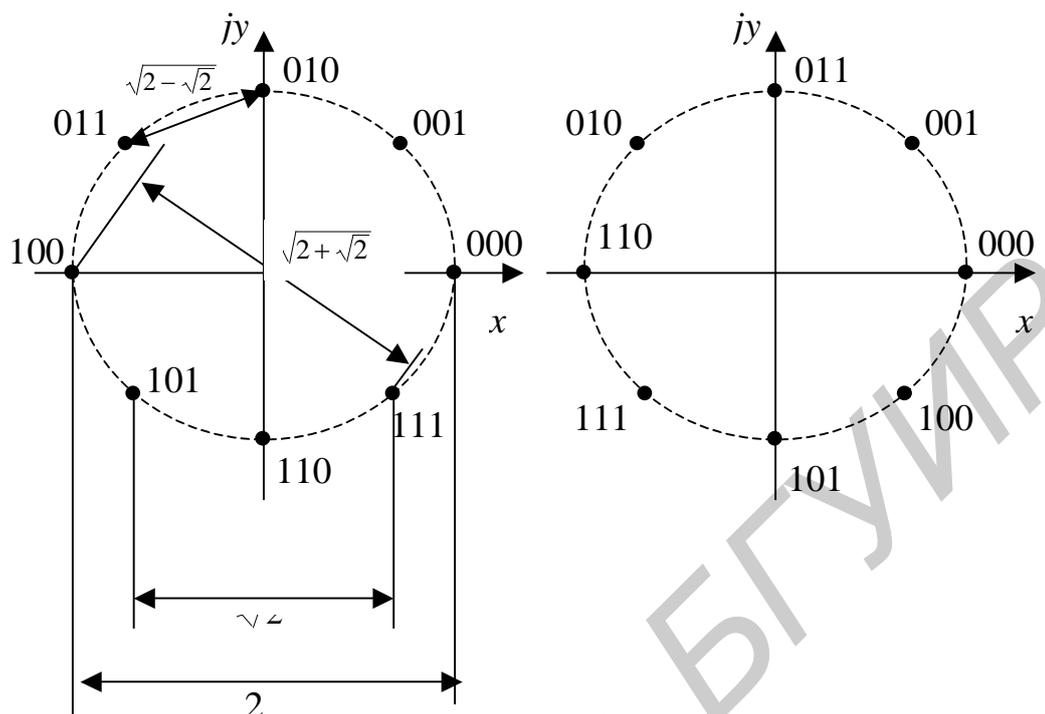


Рис. 11.44. Примеры схем модуляционного кодирования сигнала с 8-PSK

Далее рассмотрим случай передачи информации по каналу связи с ограниченной (фиксированной) полосой частот с использованием многопозиционных сигналов. Пусть реализована схема модуляции, обеспечивающая заданную скорость передачи информации при фиксированной полосе частот канала и заданной вероятности ошибки. Требуется повысить помехоустойчивость при сохранении прочих параметров качества. Одним из путей решения данной задачи является применение методов помехоустойчивого кодирования. Данные методы вносят избыточность в передаваемый информационный сигнал, позволяющую корректировать возникающие ошибки. Ширина полосы частот, занимаемая модулированным сигналом, определяется длительностью символа модуляции. Поэтому для сохранения используемой ширины полосы частот для передачи избыточной информации, заложенной помехоустойчивым кодом, необходимо использовать схемы модуляции с увеличенным числом информационных символов. Например, если в исходном варианте использовался метод модуляции 4-PSK, то при использовании кода со скоростью кодирования, равной $2/3$ (два информационных бита преобразуются в три бита помехоустойчивого кода), для сохранения занимаемой полосы частот необходимо использовать уже 8-PSK схему модуляции. Из предыдущего материала известно, что при увеличении размерности используемой схемы модуляции происходит снижение помехоустойчивости процесса демодуляции. В связи с этим возникает дилемма: компенсации потерь в помехоустойчивости при увеличении размерности схемы модуляции выигрышем

от помехоустойчивого кодирования. Другими словами, приведет ли применение помехоустойчивого кодирования к выигрышу в помехоустойчивости или окажется наоборот вредным.

Если модуляция трактуется как отдельная операция, независимая от кодирования, то требуется использование очень мощных кодов (большие кодовые ограничения сверточных кодов или большая длина блочного кода) для возмещения потерь и обеспечение некоторого достаточного выигрыша от кодирования. С другой стороны, если модуляция является частью единого процесса кодирования и она рассчитывается совместно с кодом, для увеличения минимального расстояния (евклидова расстояния) между парами кодированных сигналов потери от расширения ансамбля сигналов легко преодолеть и достигается достаточный выигрыш кодирования при относительно простых кодах. Ключ для подхода к этой совместной модуляции и кодированию заключается в изобретении эффективного метода, отображения кодовых бит (символов) в сигнальные точки так, чтобы максимизировать минимальное евклидово расстояние между парами символов.

Таким образом становится понятной необходимость синтеза специальных сигнально-кодовых конструкций, обеспечивающих оптимизацию процедуры передачи информации.

Создание методов передачи информации, в которых достигаются близкие к предельным показатели эффективности, требует согласования всех частей тракта передачи с учетом статистических свойств непрерывного канала. Это означает, что кодирование и модуляцию необходимо рассматривать как единый процесс формирования сигнала, а демодуляцию и декодирование – как процесс оптимального приема сигнально-кодированного блока в целом. Согласование модуляции и кодирования сводится к поиску такого заполнения сигнального пространства, при котором обеспечивается высокая удельная скорость (сигналы расположены достаточно плотно) и одновременно высокая помехоустойчивость (сигналы достаточно далеко друг от друга).

Декодирование модуляционного кода можно рассматривать как последний этап обработки сигнально-кодированного блока оптимальным декодером сигнально-кодированной конструкции. При этом декодер канала работает в метрике Евклида с сигналами, а не с их двоичными представлениями по правилу манипуляционного кода.

Таким образом понятие «сигнально-кодированная конструкция» является не самостоятельным, а лишь отражает подход к модуляции и кодированию как процессу объединения сигналов и кодов в единую эффективную конструкцию.

11.9.2. Согласование канала кодом Грея

Рассмотрим сигнально-кодированные конструкции, представляющие собой результат согласования известных двоичных помехоустойчивых кодов с многопозиционным ансамблем сигналов путем использования в качестве манипуляционного кода Грея.

Комбинации кода Грея получаются по следующему правилу. Кодовая комбинация исходного кода складывается по модулю 2 с такой же комбинацией, сдвинутой на один разряд вправо, при этом младший разряд сдвинутой комбинации отбрасывается (табл. 11.3).

Таблица 11.3

Пример формирования манипуляционного кода Грея

Номер состояния сигнала	Исходный двоичный код	Кодовая комбинация, кодирующая состояние сигнала по правилу Грея
0	0000	0000 – исходная комбинация
		00000 – сдвинутая комбинация
		0000 – полученная кодовая комбинация
1	0001	0001 – исходная комбинация
		00001 – сдвинутая комбинация
		0001 – полученная кодовая комбинация
2	0010	0010 – исходная комбинация
		00010 – сдвинутая комбинация
		0011 – полученная кодовая комбинация
3	0011	0011 – исходная комбинация
		00011 – сдвинутая комбинация
		0010 – полученная кодовая комбинация
4	0100	0100 – исходная комбинация
		00100 – сдвинутая комбинация
		0110 – полученная кодовая комбинация
5	0101	0101 – исходная комбинация
		00101 – сдвинутая комбинация
		0111 – полученная кодовая комбинация
6	0110	0110 – исходная комбинация
		00110 – сдвинутая комбинация
		0101 – полученная кодовая комбинация
7	0111	0111 – исходная комбинация
		00111 – сдвинутая комбинация
		0100 – полученная кодовая комбинация
8	1000	1000 – исходная комбинация
		01000 – сдвинутая комбинация
		1100 – полученная кодовая комбинация
9	1001	1001 – исходная комбинация
		01001 – сдвинутая комбинация
		1101 – полученная кодовая комбинация
10	1010	1010 – исходная комбинация
		01010 – сдвинутая комбинация
		1111 – полученная кодовая комбинация

1	2	3
11	1011	1011 – исходная комбинация 01011 – сдвинутая комбинация 0110 – полученная кодовая комбинация
12	1100	1100 – исходная комбинация 01100 – сдвинутая комбинация 1010 – полученная кодовая комбинация
13	1101	1101 – исходная комбинация 01101 – сдвинутая комбинация 1011 – полученная кодовая комбинация
14	1110	1110 – исходная комбинация 01110 – сдвинутая комбинация 1001 – полученная кодовая комбинация
15	1111	1111 – исходная комбинация 01110 – сдвинутая комбинация 1000 – полученная кодовая комбинация

Поскольку ошибки чаще происходят за счет переходов в области соседних сигналов, то кодовые блоки, соответствующие соседним сигналам, должны различаться наименьшим числом двоичных символов.

На рис. 11.45 приведен пример кодирования состояний сигнала (модуляционного кода) для 16-QAM, в скобках указан номер разрешённого состояния сигнала.

Несмотря на достаточно высокие показатели энергетической эффективности, согласование кодом Грея не является оптимальным.

Двоичные коды, оптимальные по критерию максимума хэммингова расстояния, будут оптимальны и по критерию максимума свободного евклидова расстояния, если при отображении двоичных подблоков в сигнальные точки ансамбля выполняется принцип: большему расстоянию Хэмминга соответствует большее расстояние по Евклиду.

Из рис. 11.45 видно, что наибольшее евклидово расстояние имеется, например между 0-й и 12-й точками (сигналами), в то время как расстояние по Хеммингу между определяемыми ими кодовыми комбинациями равно 2 при максимальном расстоянии для данного кода, равняющемся четырем, что говорит о неоптимальности данного варианта модуляционного кода. В связи с этим существует возможность определения более эффективных сигнально-кодовых конструкций.

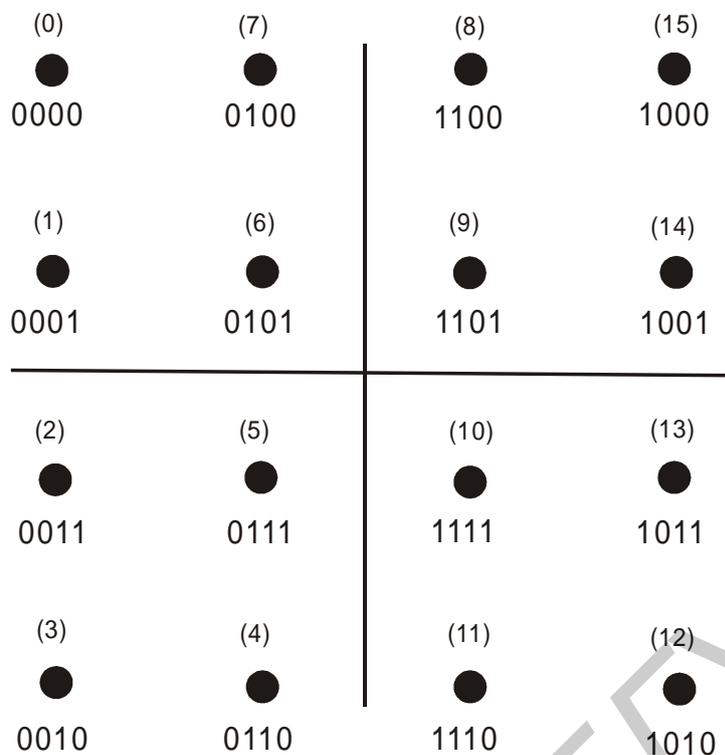


Рис. 11.45. Примеры кодирования кодом Грея состояний сигнала 16-QAM

Согласование канала кодом Грея представляет собой процедуру модуляционного кодирования и является достаточно оптимальным, в случае если исходный цифровой поток не содержит избыточности, т.е. отсутствует кодер канала и кодер источника выполняет свое назначение (оставляет «чистую» информацию). В противном случае схема манипуляционного кода Грея может не обеспечивать оптимальный разнос, обеспечивающий минимальный урон в случае возникновения ошибок.

11.9.3. Решетчатое кодирование (Trellis Coded Modulation)

11.9.3.1. Общие понятия о решетчато-кодовой модуляции

Один из возможных эффективных методов отображения кодовых бит (символов) в сигнальные точки, так чтобы максимизировать минимальное евклидово расстояние (между парами символов), был разработан в 1982 г. Унгербоеком (Ungerboeck G.). На основании анализа сигнально-кодовых на базе ансамбля 8-PSK и сверточного кода со скоростью $R = k / (k + 1)$ Унгербоек сформулировал ряд правил построения сигнально-кодовых конструкций. Сигнально-кодовые конструкции, построенные по этим правилам (Trellis-Coded Modulation – TCM), часто называют СКК Унгербоека.

По способу согласования модуляции и кодирования сигнально-кодовые конструкции Унгербоека относятся к конструкциям, полученным на основе разбиения ансамбля сигналов на вложенные подансамбли. Разбиение осу-

ществляется таким образом, что подансамбли содержат равное количество сигналов, расстояния между соседними сигналами подансамблей одинаковы, минимальные расстояния между сигналами подансамблей увеличиваются с каждым шагом разбиения.

Схемы TCM используют избыточную небинарную модуляцию плюс конечный автомат (кодер). Конечный автомат (или автомат с конечным числом состояний) – это общее название устройств, обладающих памятью о предыдущих сигналах, при этом существует ограниченное число однозначных состояний, которые может принимать автомат.

В стандартных системах, включающих модуляцию и кодирование, обычно принято отдельно описывать и реализовать детектор и декодер. Однако в системах TCM эти функции должны рассматриваться совместно. Можно добиться эффективного кодирования, не жертвуя скоростью передачи данных или не увеличивая ни ширину полосы частот, ни мощность.

При решетчатом кодировании набор сигналов многоуровневой/фазовой модуляции комбинируется со схемой решетчатого кодирования. Термин «схема решетчатого кодирования» применим к любой кодовой системе, которая обладает памятью (конечный автомат), например сверточный код.

Набор сигналов TCM наилучшим образом представляется любым набором сигналов (более чем двоичным), векторное представление которого может быть отображено на комплексной плоскости. Унгербок (Ungerboeck) показал, что при наличии шума схема TCM довольно просто может дать суммарную эффективность кодирования, близкую к 3 дБ, по сравнению с некодированной системой, а при увеличении сложности можно достичь эффективности порядка 6 дБ.

Сверточное кодирование увеличивает размер множества сигнала с 2^k до 2^{k+p} . Унгербок исследовал повышение пропускной способности, достигаемое благодаря увеличению набора сигналов, и пришел к заключению, что максимальную эффективность можно реализовать, удваивая передаваемый некодированный набор ($p = 1$). Этого можно достичь путем кодирования со степенью $k/(k+1)$ с последующим отображением групп из $k+1$ бит в набор из 2^{k+1} сигналов. Так, при реализации TCM со степенью кодирования $2/3$ набор сигналов модуляции 4-PSK заменяется набором сигналов модуляции 8-PSK, аналогично 16-QAM замещается на 32-QAM при степени кодирования $4/5$ и т.д.

В каждом из случаев TCM система синтеза сигнально-кодовых конструкций сконфигурирована таким образом, чтобы до и после кодирования средняя мощность сигнала была одинаковой. Кроме того, для обеспечения необходимой избыточности при кодировании размер набора сигналов увеличивается в два раза с $M = 2^k$ до $M' = 2^{k+1}$. Расширенный набор сигналов при фиксированной средней мощности приводит к уменьшению расстояния между соседними точками символов, как видно из рис. 11.46. В некодированной системе такое уменьшение расстояния снижает достоверность пере-

дачи. Однако вследствие избыточности, вносимой кодом, это уменьшение расстояния уже не сильно влияет на вероятность ошибки. Напротив, достоверность определяет просвет – минимальное расстояние между членами набора разрешенных кодовых последовательностей. Другими словами, когда два сигнала находятся в непосредственной близости друг от друга в сигнальном пространстве кодовой системы, их близость может и не иметь существенного значения (с точки зрения вероятности ошибки), поскольку правила кода могут не допускать перехода между двумя такими, якобы уязвимыми точками сигналов. Задача ТСМ заключается в присвоении сигнальных точек переходам в кодовой решетке таким образом, чтобы увеличить просвет между теми переходами (символами), которые вероятнее всего могут быть перепутаны.

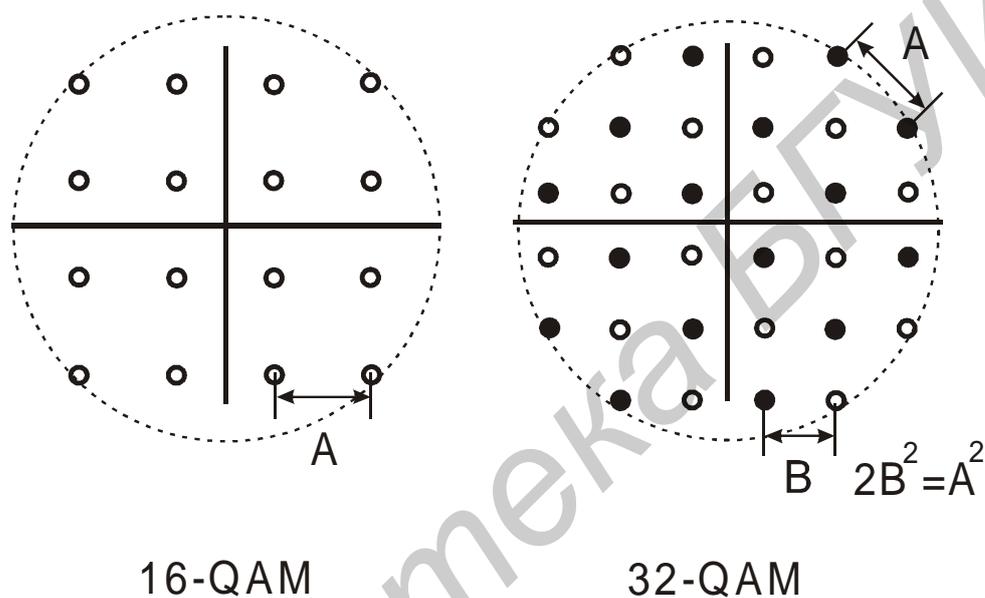


Рис. 11.46. Изменение расстояния между сигнальными точками при удвоении размерности схемы модуляции и сохранении средней мощности сигнала

11.9.3.2. Разбиение исходного ансамбля сигналов

Поскольку сигнально-кодовые конструкции Унгербоека относятся к конструкциям, полученным на основе разбиения ансамбля сигналов на вложенные подансамбли, рассмотрим данную процедуру. Разбиение осуществляется таким образом, что подансамбли, содержащие равное количество сигналов, при разбиении стремятся сохранять регулярную структуру множеств и подмножеств сигналов, при последовательном разбиении минимальные расстояния между сигналами подансамблей увеличиваются с каждым шагом разбиения, формально левая ветвь разбиения кодируется символом «0», а правая «1» для получения обозначения получаемого подмножества. На рис. 11.47 приведен пример последовательного разбиения созвездия сигнала 16-QAM на подансамбли. Из рисунка видно, что при разбиении минимальное

расстояние между сигналами подансамблей изменяется в следующем порядке: уровень $A - A_0$; уровень $B - \sqrt{2}A_0$; уровень $C - 2A_0$ и уровень $D - 2\sqrt{2}A_0$, т. е. каждый уровень разбиения увеличивает расстояние между точками, изображающими сигнал в $\sqrt{2}$ раз.

Степень, до которой происходит разбиение сигнала, зависит от характеристик сигнально-кодовой конструкции кода. В общем, процесс сигнально-кодированного отображения выполняется так, как показано на рис. 11.47.

Блок из k информационных символов делится на две группы длиной n и m , m символов кодируются в $m+p$ символов, в то время как n символов остаются не кодированными. После этого $m+p$ символов кодера используются для выбора одного из 2^{m+p} возможных подансамблей (используются только подансамбли одного уровня) в разделенном ансамбле сигналов, в то время как n символов используются для выбора одной из 2^n сигнальных точек в каждом подансамбле. Если $n=0$, все информационные символы кодируются.

11.9.3.3. Конструирование сигнально-кодовой конструкции

Конструирование сигнально-кодовой конструкции рассмотрим на базе примера, популярного в литературе по основам цифровой радиосвязи.

Постановка задачи: необходимо повысить помехоустойчивость передачи двоичных сигналов в радиоканале с фиксированной полосой частот (исходно используемая схема модуляции 4-PSK).

Из предыдущего материала известно, что максимальная эффективность от использования Trellis-Coded Modulation при сохранении занимаемой полосы частот достигается при повышении размерности схемы модуляции в два раза. На основании этого выберем для составления сигнально-кодовой конструкции метод модуляции 8-PSK (можно 8-QAM).

Произведем разбиение ансамбля сигналов 8-PSK на подмножества в соответствии со схемой, приведенной на рис. 11.49.

Для кодирования сигнально-кодовой конструкции примем следующую схему. Исходными данными для кодирования являются пары символов, кодирующие исходный сигнал 4-PSK. Один из данных символов будем использовать без преобразований (кодирования), а второй пропустим через конечный автомат (кодер) со скоростью преобразования $1/2$. Возможный вариант кодера приведен на рис. 11.50.

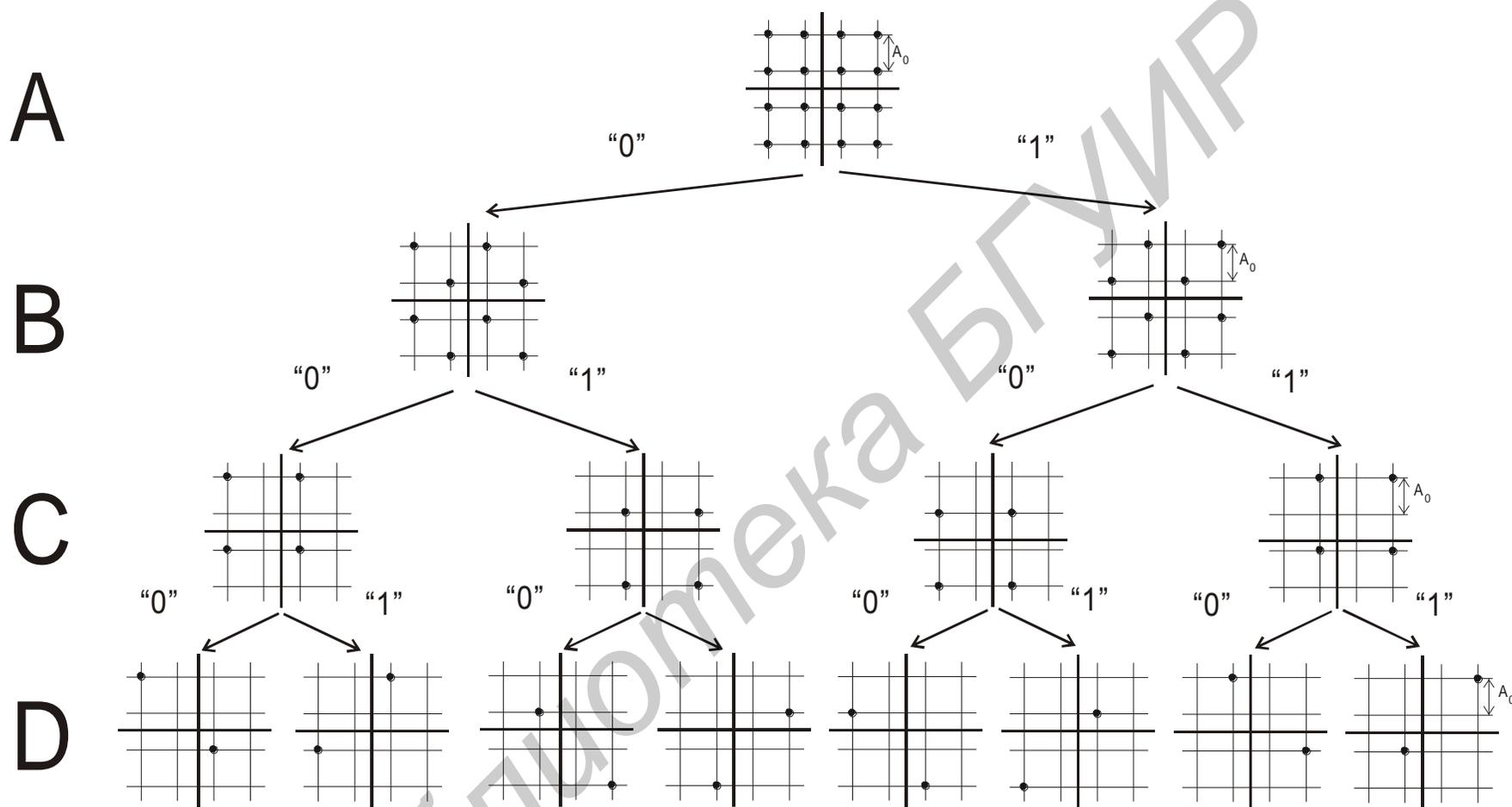


Рис. 11.47. Пример разбиения множества сигнала 16-QAM на подмножества

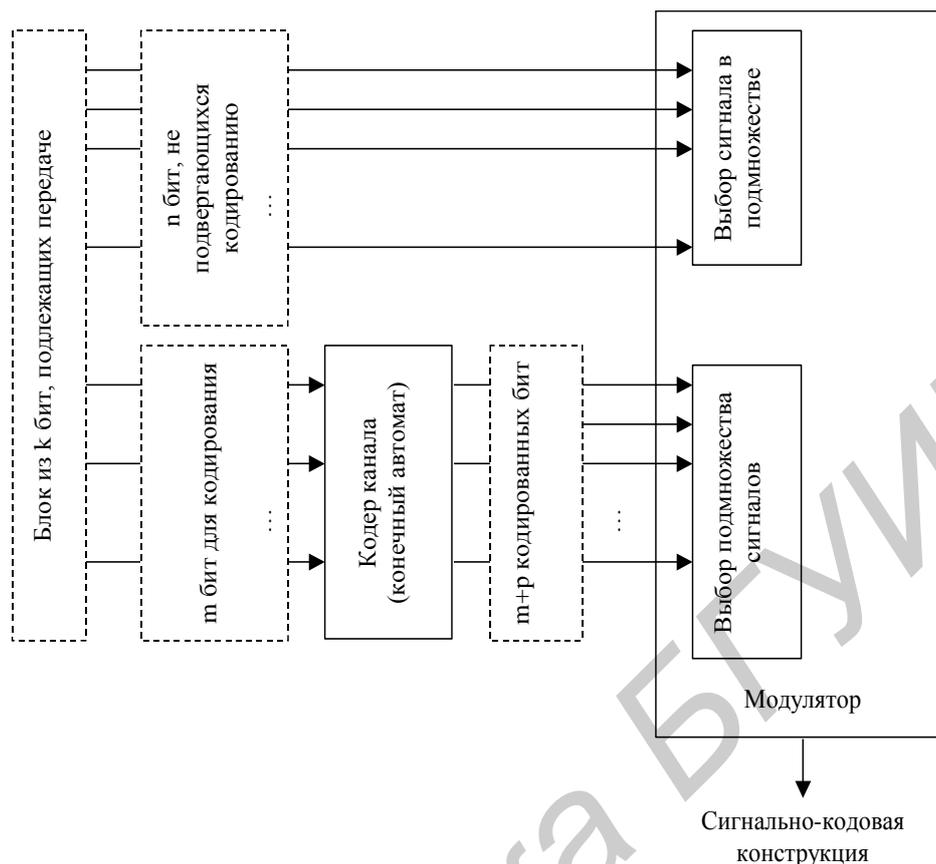


Рис. 11.48. Общая структура комбинированного кодера/модулятора

Состояния конечного автомата принято обозначать двоичными символами, хранящимися в элементах задержки, за исключением крайнего левого (входного). Приведенный на рисунке автомат может находиться в одном из четырех состояний: $a = 00$, $b = 10$, $c = 01$ и $d = 11$.

Решетчатая диаграмма переходов данного автомата приведена на рис. 11.51 (составлена на основании данных табл. 11.4).

Произведя добавление некодированного бита (введение параллельных ветвей), получаем полную решетчатую диаграмму схема помехоустойчивого кодирования, приведенную на рис. 11.52.

Выполнив наложение решетчатой диаграммы на разбитое созвездие сигнала 8-PSK, получим решетчатую диаграмму сигнально-кодвой конструкции, изображенную на рис. 11.54.

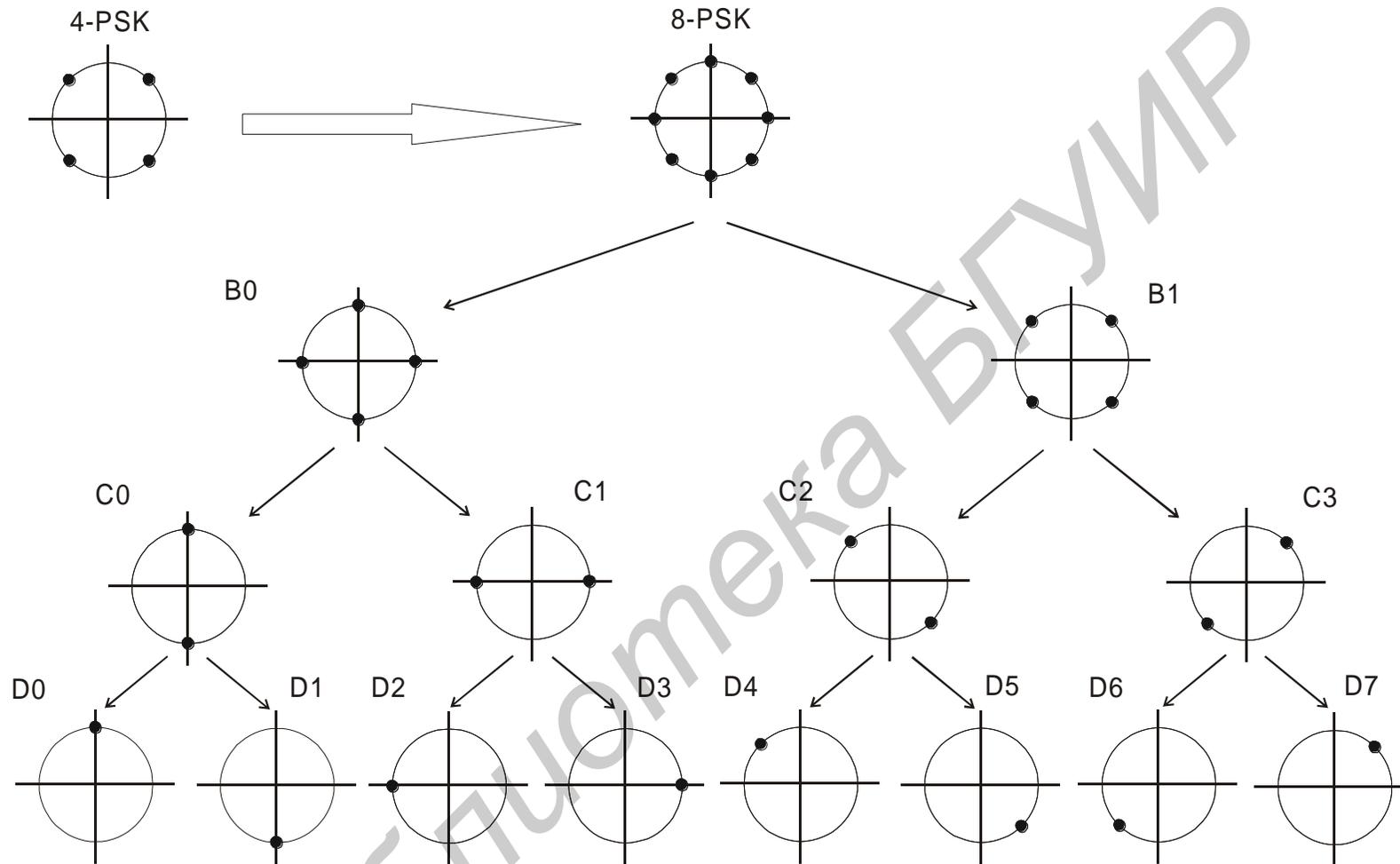


Рис. 11.49. Разбиение ансамбля сигналов 8-PSK

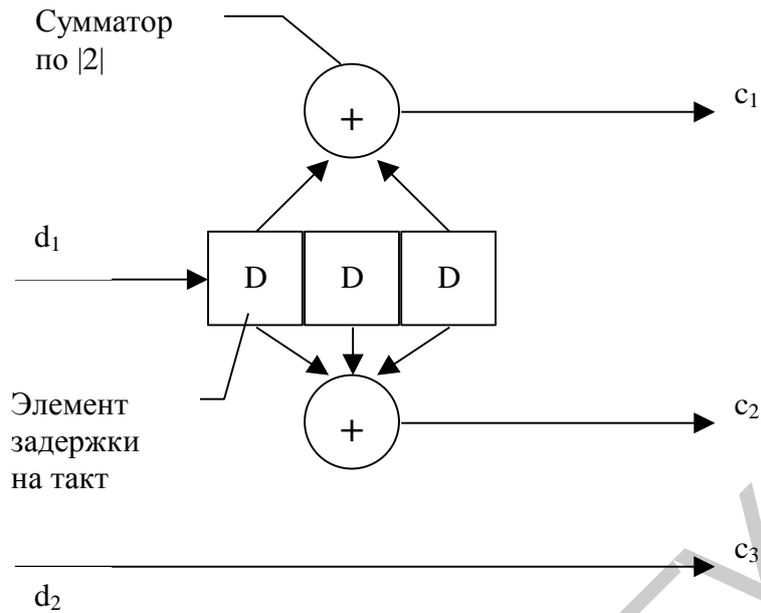


Рис. 11.50. Кодировщик автомат

Таблица 11.4

Переходы кодера

Входной символ d_1	Состояние авто- мата	Выходные сим- волы c_1c_2	Следующее со- стояние
0	00 (<i>a</i>)	00	00 (<i>a</i>)
1	00 (<i>a</i>)	11	10 (<i>b</i>)
0	10 (<i>b</i>)	01	01 (<i>c</i>)
1	10 (<i>b</i>)	10	11 (<i>d</i>)
0	01 (<i>c</i>)	11	00 (<i>a</i>)
1	01 (<i>c</i>)	00	10 (<i>b</i>)
0	11 (<i>d</i>)	10	01 (<i>c</i>)
1	11 (<i>d</i>)	01	11 (<i>d</i>)

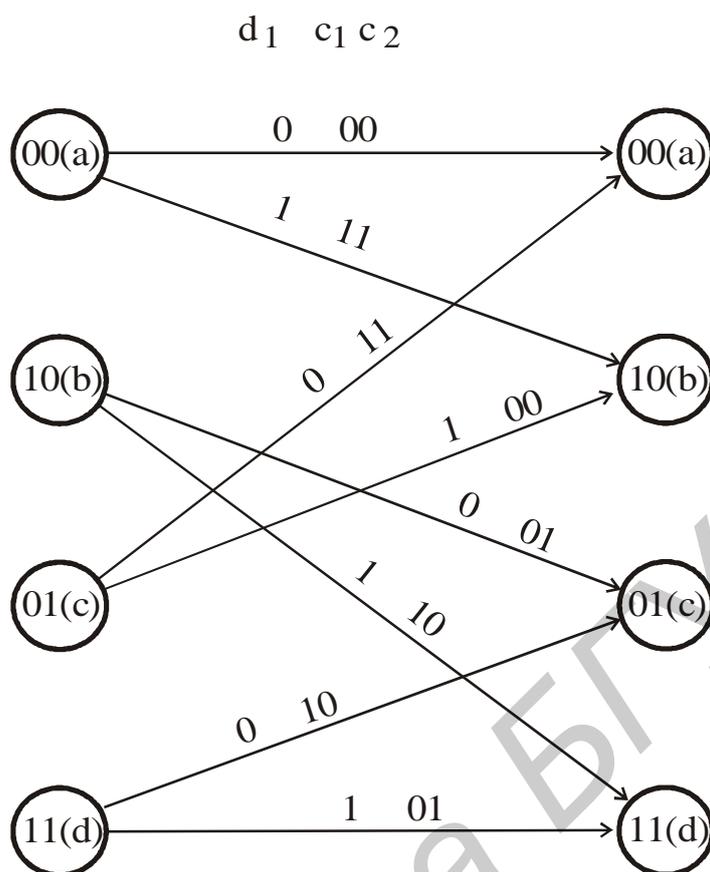


Рис. 11.51. Решетчатая диаграмма переходов кодирующего автомата

При наложении решетчатой диаграммы кодера на созвездие используемого сигнала выполняются следующие правила:

1. Параллельные переходы (когда они возникают) назначаются сигнальным точкам, разделенным максимальным расстоянием Евклида (в нашем примере точкам, принадлежащим одному из подмножеств C_0, \dots, C_3).
2. Переходы, начинающиеся (заканчивающиеся) в некотором состоянии, назначаются подмножествам более высокого уровня из одной ветви разделения (в рассматриваемом примере либо B_0 (C_0C_1), либо B_1 (C_2C_3)), имеющим следующее за максимальным расстоянием Евклида между сигнальными точками.
3. Сигнальные точки должны по возможности возникать с одинаковой частотой.

На рисунке обозначения D_0, \dots, D_7 соответствуют сигналам нижнего уровня разбиения созвездия 8-PSK и показывают, какой именно сигнал излучается при переходе сигнально-кодовой конструкции из одного состояния в другое. Состояния сигнально-кодовой конструкции обозначаются C_iC_j в соответствии с используемыми в данном состоянии подансамблями сигналов (подансамблями выходных ветвей) C_i и C_j . Кодировка сигналов D_0, \dots, D_7 определяет модуляционный код и связывает конкретные сигналы с кодом, формируемым автоматом помехоустойчивого кодирования.

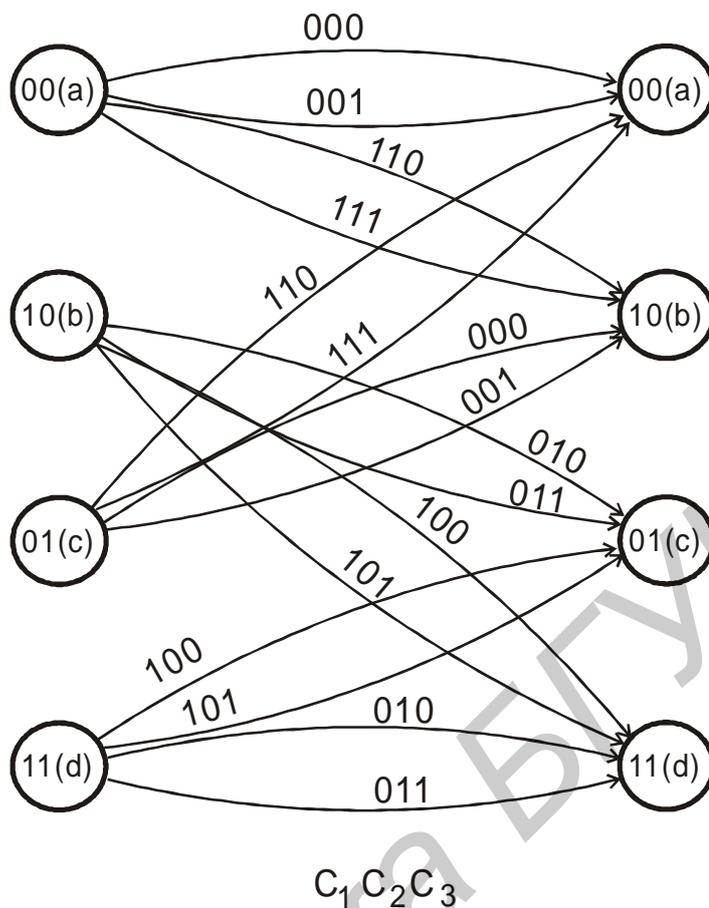


Рис. 11.52. Решетчатая диаграмма переходов помехоустойчивого кодера

Для рассматриваемого примера модуляционный код определяется следующим образом.

- 000 → D0
- 001 → D1
- 010 → D4
- 011 → D5
- 100 → D6
- 101 → D7
- 110 → D2
- 111 → D3

Два кодированных символа используются для выбора одного из четырех подобразов сигнального созвездия C₀...C₃, в то время как оставшийся информационный символ используется для выбора одной из двух точек внутри каждого подобраза.

Показанная на рис. 11.53 решетчатая диаграмма является в основе своей решёткой для свёрточного кода со скоростью $1/2$ с добавлением параллельных путей, размещающих некодированные символы d_2 . Таким образом, кодированные символы c_1c_2 используются для выбора одного из четырех подобразов, каждый из которых содержит две сигнальные точки, а некодированные символы используются для выбора одной из двух сигнальных точек внутри каждого подобраза. Заметим, что сигнальные точки внутри подобраза удалены на расстояние 2. Таким образом, расстояние Евклида между параллельными путями равно 2.

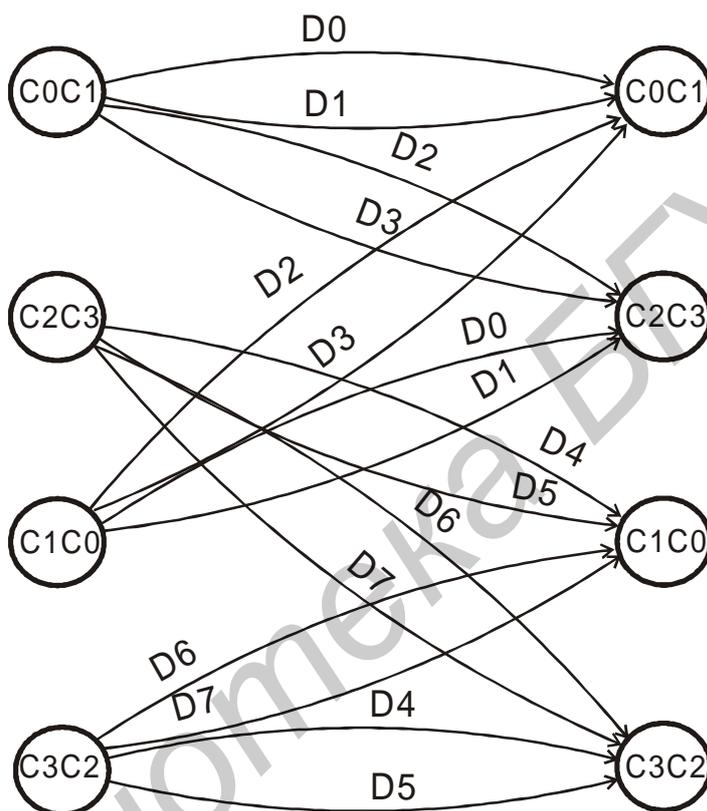


Рис. 11.53. Решетчатая диаграмма сигнально-кодовой конструкции

Анализ решетчатой диаграммы показывает, что некодированные символы d_2 всегда передаются сигналами, максимально отстоящими друг от друга (для случая 8-PSK это нормированное расстояние равно 2). Следует отметить, что в исходной схеме модуляции 4-PSK минимальное нормированное расстояние между некодированными символами равнялось $\sqrt{2}$. Таким образом, помехоустойчивость передачи некодированных символов (каждого второго из передаваемых) повысилась на $20 \lg(\sqrt{2}) = 3$ дБ. При передаче второго (кодированного) символа следует также ожидать повышения помехоустойчивости, поскольку он подвергнут помехоустойчивому кодированию со скоростью $1/2$ (каждый информационный символ передается двумя символами модуляции).

В качестве альтернативной схемы кодирования мы можем использовать кодер со скоростью $2/3$ и кодировать оба информационных символа так, как, например, показано на рис. 11.54. Это кодирование ведет к решётке с восемью состояниями и обеспечивает лучшее качество, но требует более сложную реализацию декодера.

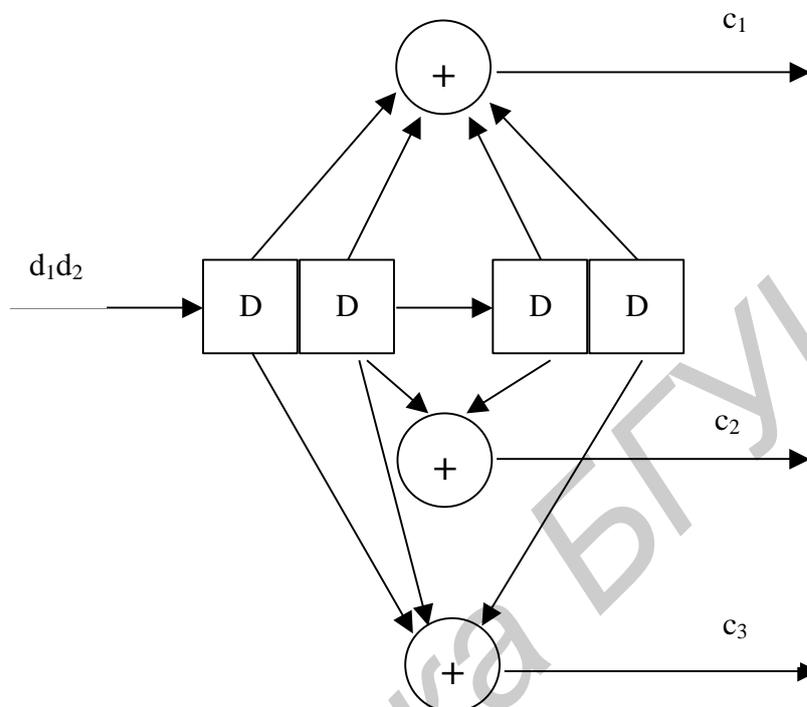


Рис. 11.54. Пример альтернативного кодирующего автомата

Демодуляция сигналов ТСМ осуществляется в соответствии с правилами и алгоритмами, применимыми для декодирования соответствующих помехоустойчивых кодов, с той лишь разницей, что в процессе декодирования (демодуляции) вместо расчета расстояний между кодовыми конструкциями в метрике Хемминга используется расстояние между сигнальными точками в метрике Евклида. В связи с этим процесс демодуляции – декодирования для сигналов с ТСМ является неразрывным и единым. Кроме того, если позволяет вычислительная мощность демодулятора, при демодуляции сигналов с ТСМ широко используются так называемые мягкие методы декодирования, основанные на отслеживании большего числа возможных траекторий СКК, в той или иной степени близких к наблюдаемой траектории входной смеси сигнала и помехи. Мягкие методы декодирования (демодуляции) используют многопороговые схемы принятия решения, позволяющие определять «степень» достоверности того или иного решения, а не только самое достоверное.

11.9.3.4. Сравнение помехоустойчивости кодированной и некодированной схем

Основные принципы сравнения помехоустойчивости кодированной (TCM) и некодированной схем модуляции проведем на основе рассмотренного ранее примера построения сигнально-кодовой конструкции.

Итак, исходная схема модуляции 4-PSK. Пусть используется набор сигналов для представления модулирующих символов, соответствующий подмножеству сигналов В1 (см. рис. 11.49). Решетчатая диаграмма данной сигнальной конструкции приведена на рис. 11.55.

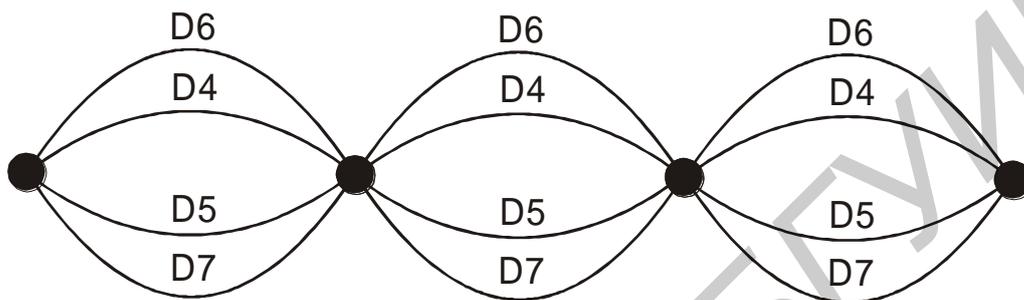


Рис. 11.55. Сигнально-кодовая конструкция некодированной 4-PSK

Некодированный сигнал 4-PSK имеет одно состояние и четыре ветви, минимальное нормированное расстояние по Евклиду между которыми составляет $\sqrt{2}$.

Решетчатая диаграмма рассматриваемого сигнала с TCM приведена на рис. 11.56.

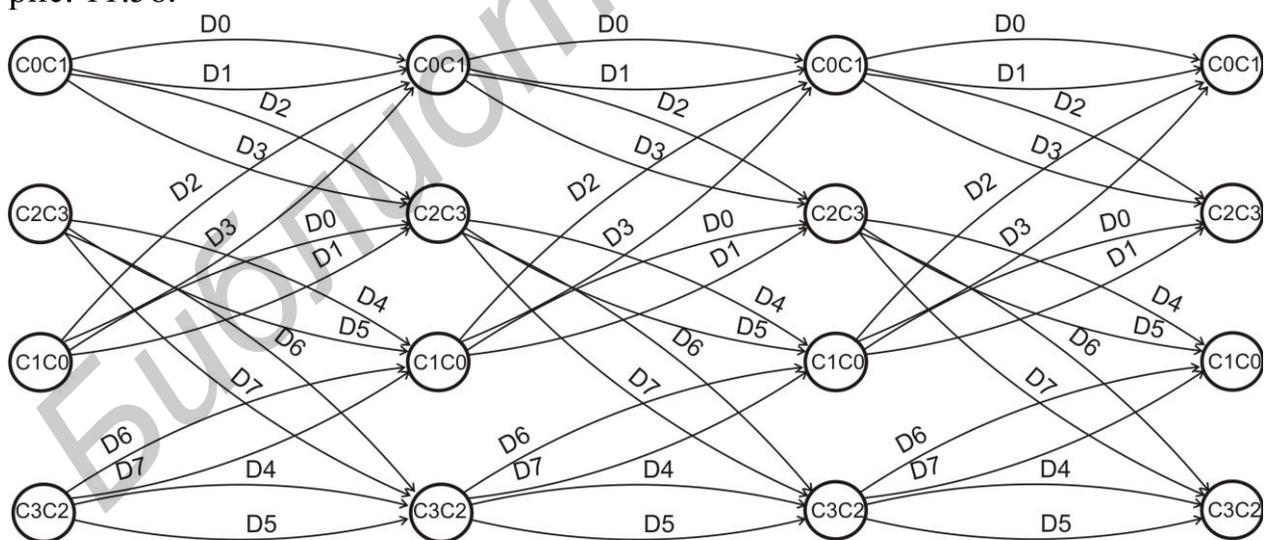


Рис. 11.56. Решетчатая диаграмма кодированного сигнала

Демодуляция некодированной конструкции осуществляется на одном переходе, следовательно, ее помехоустойчивость определяется квадратом расстояния по Евклиду между ветвями (сигналами), соединяющими одина-

ковые состояния. В рассматриваемом случае это $(\sqrt{2})^2 = 2$ (нормированное к энергии сигнала, приходящейся на символ модуляции).

Демодуляция кодированного сигнала осуществляется путем отслеживания траектории на протяжении нескольких переходов (определяется ограничением кода). Если определить минимальное расстояние между любыми двумя путями, соединяющими одинаковые состояния, на участке с более чем одним переходом, то квадрат минимального расстояния между данными траекториями будет равен 4 (квадрат расстояния между кратчайшими траекториями, например, параллельными цепочками D0-D1). Поскольку помехоустойчивость определяется квадратом минимального расстояния, то налицо выигрыш рассматриваемой кодированной схемы в 3 дБ по сравнению с некодированной. Большой выигрыш может быть получен использованием кодов с большим числом состояний (позволяют исключить параллельные ветви в решетчатой диаграмме сигнально-кодовой конструкции).

Примечание. Квадрат расстояния между двумя траекториями определяется как сумма квадратов расстояний между ветвями принадлежащими данным траекториям на каждом переходе. Например (см. рис. 11.56), для двух траекторий D0-D1-D0 и D3-D5-D2 квадрат расстояния равен $(D0-D3)^2 + (D1-D5)^2 + (D0-D2)^2$ (обозначение DI-DJ соответствует расстоянию между изображающими точками в созвездии сигнала).

11.10. Заключение

Существует большое разнообразие возможных вариантов осуществления информационной модуляции радиосигналов. Вне зависимости от выбранного способа модуляции существуют общие закономерности, присущие всем модуляционным схемам. К таким закономерностям можно отнести следующие:

1. Полоса занимаемых радиосигналом частот прямо пропорциональна символьной скорости модуляции, символ/с.

2. Вид спектральной плотности мощности модулированного сигнала в значительной мере определяется видом формирующего импульса и не зависит (для случая линейной модуляции) от конкретного вида модуляции.

3. Спектральная эффективность радиосигнала (бит/Гц)/с, может быть повышена использованием многопозиционных методов модуляции и соответствующим выбором вида и параметров формирующего импульса.

4. Помехоустойчивость демодуляции радиосигналов вне зависимости от вида модуляции определяется отношением энергии, приходящейся на один бит передаваемой информации, к спектральной плотности мощности шума наблюдения (вероятность ошибочного приема падает с ростом данного отношения).

5. Наилучшими характеристиками помехоустойчивости (при равных энергетических соотношениях) обладают модуляционные схемы, при кото-

рых наибольшее расстояние по Хеммингу между передаваемыми символами соответствует наибольшему расстоянию по Евклиду между сигналами, отображающими данные символы.

6. Большой эффективностью в каналах с ограниченной полосой пропускания обладают методы повышения помехоустойчивости, в которых помехоустойчивое кодирование согласовано с модуляционной схемой (сигнально-кодовые конструкции).

7. Оптимальный прием большинства радиосигналов с дискретной модуляцией требует точного знания фазы несущего колебания и, соответствующих схем (процедур) восстановления данной информации из принимаемого сигнала при отсутствии данной информации.

8. Поскольку решение о значении передаваемых информационных символов принимается в дискретные моменты времени, необходимым условием успешной демодуляции является наличие тактовой синхронизации демодулятора с передаваемым потоком данных.

Библиотека БГУИР

12. СИНХРОНИЗАЦИЯ

12.1. Общая задача синхронизации при приеме сигналов с дискретной модуляцией

Из предыдущих разделов известно, что наибольшая энергетическая и спектральная эффективность передачи дискретной информации достигается при использовании методов фазовой манипуляции, таких как BPSK, QPSK, M-PSK и M-QAM. Прием сигналов, модулированных в соответствии с данными методами, осуществляется по так называемой когерентной схеме. Данная схема предполагает точное знание частоты и фазы несущего колебания. В реальных условиях частота и фаза несущего колебания не известны. Это обусловлено особенностями распространения радиоволн, нестабильностью частоты опорных генераторов, изменяющимся местоположением передающего и приемного устройств, эффектом Доплера и пр.

В связи с тем что параметры несущего колебания заранее определены быть не могут, при осуществлении когерентной демодуляции сигналов необходимо осуществлять оценку частоты и фазы несущего колебания. Результаты данной оценки затем используются для компенсации возможных ошибок установки опорного колебания в демодуляторе.

Наряду с оценкой параметров несущего колебания процесс демодуляции требует тактовой синхронизации демодулятора и потока передаваемых (принимаемых символов). Данная синхронизация нужна, во-первых, для правильной временной привязки процесса генерации опорных несущих сигналов, используемых для формирования (вычисления) значений соответствующих функционалов правдоподобия; во-вторых, процедура непосредственно принятия решения о том или ином значении принимаемого символа дискретна во времени (сначала происходит энергетическое накопление результатов сравнения опорных копий несущих сигналов с принятой смесью сигнала и шума, а затем, в момент максимального накопления, обычно в конце передаваемого символа, происходит принятие решения) и, в-третьих, после принятия того или иного решения необходим сброс накопительных элементов.

Таким образом, обобщенная структура когерентного (квазикогерентного, поскольку параметры несущего сигнала не могут быть определены с идеальной точностью) демодулятора приведена на рис. 12.1.

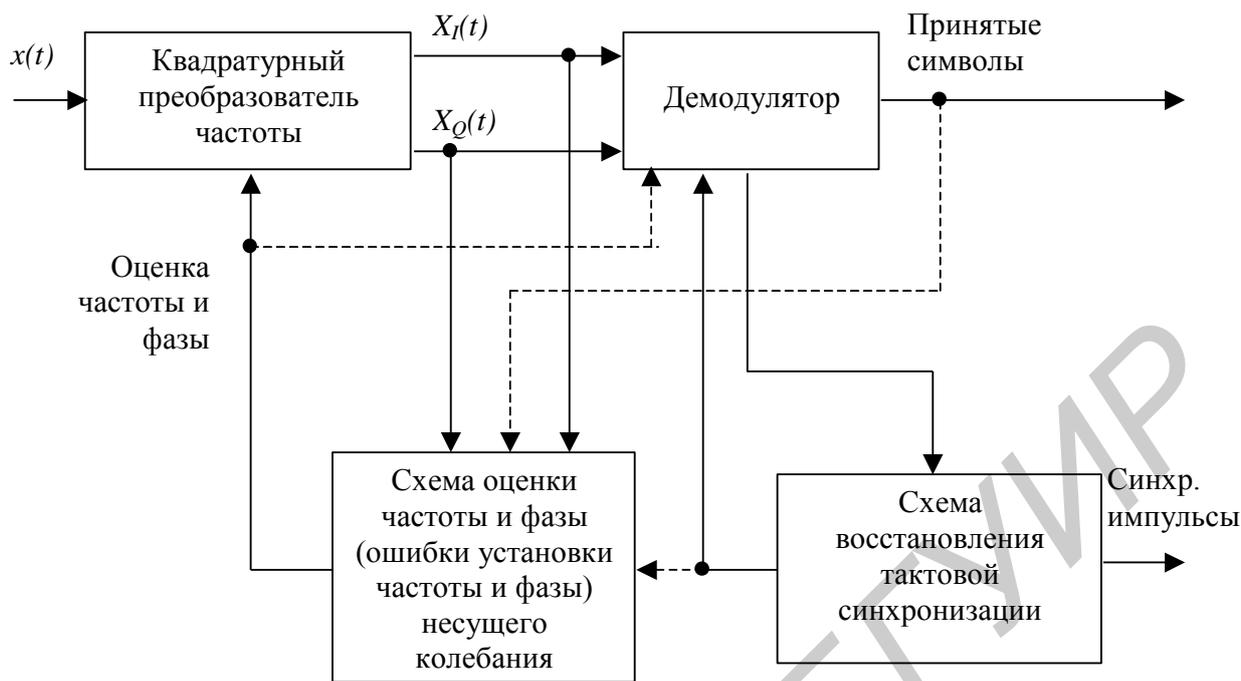


Рис. 12.1. Обобщенная структура квазикогерентного демодулятора

Квадратурный преобразователь частоты осуществляет перенос принятого радиосигнала в область нулевой частоты, при этом на его выходах формируются сигналы, представляющие действительную и мнимую части комплексной огибающей принятой смеси сигнала и шума. Демодулятор по результатам обработки комплексной огибающей входной смеси сигнала и шума формирует на своем выходе оценки принятых информационных символов. Работа демодулятора синхронизирована с потоком принимаемых информационных символов сигналами тактовой синхронизации, сформированными схемой восстановления тактовой синхронизации. Схема восстановления тактовой синхронизации в качестве входного сигнала использует промежуточные результаты обработки комплексной огибающей в демодуляторе (может непосредственно использовать комплексную огибающую). Схема оценки частоты и фазы (ошибки установки частоты и фазы) несущего колебания на основании обработки комплексной огибающей входной смеси сигнала и шума формирует оценку данных параметров, которая используется для формирования опорного колебания квадратурного преобразователя частоты либо для коррекции сигналов комплексной огибающей в демодуляторе (показано пунктирной линией). Процесс оценки частоты и фазы может сопровождаться снятием информационной модуляции (так называемая обратная связь по решению показана пунктирной линией).

Следует отметить, что точность оценки частоты и фазы (синхронизации по несущему сигналу) оказывает сильное влияние на реализуемые вероятности ошибочного приема сигналов с дискретной модуляцией.

12.2. Методы восстановления несущего колебания при полосовой демодуляции

Известны два базовых подхода для обеспечения синхронизации по несущей в приёмнике. Один сводится к замешиванию на передаче специального сигнала, называемого пилот-сигналом, который можно извлечь из принимаемого сигнала и по его частоте и фазе синхронизировать местный опорный генератор. Когда по каналу передана одновременно немодулированная несущая (пилот-сигнал) и информационный сигнал, приёмник использует замкнутую петлю фазовой автоподстройки (ФАП, PLL), чтобы захватить и отслеживать несущую. ФАП рассчитывается так, чтобы реализовать узкую полосу пропускания и тем самым обеспечить слабое воздействие частотных компонент информационного сигнала на ее работу.

Второй подход сводится к непосредственной оценке фазы несущей по модулированному сигналу. Этот подход имеет отчётливое преимущество, поскольку вся мощность передатчика направлена на передачу информационного сигнала.

Поскольку реализация первого подхода более или менее прозрачна, ограничим наше внимание вторым подходом.

Сложность оценки фазы несущего сигнала непосредственно по модулированному сигналу заключается в том, что, как правило, спектр сигнала с цифровой модуляцией (особенно при использовании схем требующих только когерентного детектирования) не содержит составляющих, представляющих несущее колебание в чистом виде. И поэтому для осуществления синхронизации по несущему колебанию, в первую очередь, необходимо получить данные составляющие свободные от информационной модуляции. В настоящее время известны два способа получения компонент, несущих информацию о фазе несущей и свободных от информационной модуляции. К первому относятся так называемые методы с обратной связью по решению. В данной группе методов результаты демодуляции информационных символов повторно модулируют принятый сигнал для снятия с него информационной составляющей. Второй способ заключается в использовании особенностей сигнального созвездия используемого метода модуляции, благодаря которым путем определенного нелинейного преобразования принятого сигнала можно осуществить «съём» информационной модуляции (методы не управляемые решением).

Рассмотрим эти методы получения информации о фазе несущего сигнала более подробно.

12.2.1. Максимально правдоподобная оценка фазы

Из предыдущего материала известно, что максимально правдоподобная оценка неизвестного параметра сигнала (в нашем случае мгновенной началь-

ной фазы $\varphi(t) = \Delta\omega(t) + \varphi_0$), где $\Delta\omega(t)$ – неизвестная ошибка установки частоты; φ_0 – неизвестная начальная фаза) определяется исходя из условия

$$\left. \frac{\partial}{\partial \varphi(t)} \ln(P(X|\varphi(t))) \right|_{\varphi(t) = \varphi^\#(t)} = 0, \quad (12.1)$$

где $\varphi^\#(t)$ – оптимальная в соответствии с методом максимального правдоподобия оценка неизвестной фазы сигнала. Здесь необходимо подчеркнуть, что значение оценки фазы сигнала производится в каждый момент времени, т.е. оценка является текущей (решается задача непрерывной фильтрации неизвестного параметра сигнала). Опустив громоздкие выкладки, запишем обобщенное правило формирования текущей оценки параметра сигнала для случая его наблюдения на фоне белого гауссовского шума в установившемся режиме.

$$L(\varphi^\#(t)) = K \frac{\partial}{\partial \varphi^\#(t)} \left[-\frac{1}{N_0} (\bar{X}(t) - \bar{S}(t, \varphi^\#(t))) \times (\bar{X}(t) - \bar{S}(t, \varphi^\#(t)))^* \right], \quad (12.2)$$

где $L(p(t)) = a_n \frac{d^n p(t)}{dt^n} + a_{n-1} \frac{d^{n-1} p(t)}{dt^{n-1}} + \dots + a_0 p(t)$ – линейный оператор, определяемый динамическими свойствами (динамикой поведения во времени) оцениваемого параметра; K – параметр, определяющий усиление в петле слежения за неизвестным параметром сигнала; $\bar{X}(t)$ – комплексная огибающая принятой смеси полезного сигнала и шума; $\bar{S}(t, \varphi^\#)$ – комплексная огибающая полезного сигнала; $*$ – операция комплексного сопряжения; $\#$ – обозначение соответствующей оценки (значение, используемое в опорном сигнале при приеме).

Учитывая, что в случае осуществления синхронизации по несущему колебанию оценке подлежит неэнергетический параметр – фаза колебания, уравнение для оценки можно преобразовать к виду

$$L(\varphi^\#(t)) = K \frac{\partial}{\partial \varphi^\#(t)} \left[\frac{2}{N_0} \operatorname{Re} \{ \bar{X}(t) \bar{S}^*(t, \varphi^\#(t)) \} \right]. \quad (12.3)$$

Рассмотрим комплексную огибающую полезного сигнала подробнее.

$$\bar{S}(t, \varphi(t)) = \bar{S}_{inf}(t) e^{j\varphi(t)}, \quad (12.4)$$

где $\bar{S}_{inf}(t)$ – компонента комплексной огибающей, представляющая информационную модуляцию; $e^{j\varphi(t)}$ – компонента, определяющая фазу несущего

колебания. С учетом этого уравнение для оценки фазы несущего колебания можно переписать так:

$$L(\varphi^\#(t)) = K \left[\frac{2}{N_0} \operatorname{Re} \left\{ -j \bar{X}(t) (\bar{S}_{inf}^\#(t))^* e^{-j\varphi^\#(t)} \right\} \right], \quad (12.5)$$

где $\bar{S}_{inf}^\#(t)$ – информационная компонента комплексной огибающей с учетом оценки передаваемых информационных символов.

Для лучшего понимания полученного результата рассмотрим пример.

Пусть $L(\varphi^\#(t)) = \frac{d}{dt} \varphi^\#(t)$ соответствует отсутствию каких-либо детерминированных проявлений в поведении отслеживаемого процесса $\varphi(t)$ во временной области; $\bar{S}(t) = \pm A_0 e^{j\varphi(t)}$ соответствует сигналу с BPSK вида $s(t) = \pm A_0 \cos(\omega_0 t + \varphi(t))$, тогда имеем

$$\begin{aligned} \frac{d}{dt} \varphi^\#(t) &= -K \left[\frac{2}{N_0} \operatorname{Re} \left\{ j \left(\pm A_0 e^{j\varphi(t)} + \bar{N}(t) \right) \left(\pm A_0^\# e^{-j\varphi^\#(t)} \right) \right\} \right] = \\ &= -K \left[\frac{2}{N_0} \operatorname{Re} \left\{ j \left((\pm A_0) (\pm A_0^\#) e^{j(\varphi(t) - \varphi^\#(t))} + \pm A_0^\# \bar{N}(t) e^{-j\varphi^\#(t)} \right) \right\} \right] = \\ &= -\frac{2K}{N_0} (\pm A_0) (\pm A_0^\#) \sin(\Delta\varphi(t)) - K \left[\frac{2}{N_0} \operatorname{Re} \left\{ j \left(\pm A_0^\# \bar{N}(t) e^{-j\varphi^\#(t)} \right) \right\} \right], \end{aligned} \quad (12.6)$$

где $\Delta\varphi(t) = \varphi^\#(t) - \varphi(t)$ – ошибка оценки фазы; $\pm A_0^\#$ – информационная компонента опорного колебания на приемной стороне; $\bar{N}(t)$ – комплексная огибающая шума наблюдения; $\frac{2K}{N_0} (\pm A_0) (\pm A_0^\#) \sin(\Delta\varphi(t))$ – компонента, отражающая рассогласование принимаемого и опорного сигнала;

$K \left[\frac{2}{N_0} \operatorname{Re} \left\{ j \left(\pm A_0^\# \bar{N}(t) e^{-j\varphi^\#(t)} \right) \right\} \right]$ – шумовая компонента в схеме оценки фазы принимаемого сигнала. В случае когда информационные компоненты принимаемого и опорного сигналов совпадают (информационные модуляции принимаемого и опорного сигналов совпадают), мы имеем

$$\frac{d}{dt} \varphi^\#(t) = -\frac{2K}{N_0} A_0^2 \sin(\Delta\varphi(t)) + K \left[\frac{2}{N_0} \operatorname{Re} \left\{ j \left(\pm A_0^\# \bar{N}(t) e^{-j\varphi^\#(t)} \right) \right\} \right] \quad (12.7)$$

или

$$\varphi^{\#}(t) = \int_0^t \left(\frac{-2K}{N_0} A_0^2 \sin(\Delta\varphi(y)) + K \left[\frac{2}{N_0} \operatorname{Re} \left\{ j \left(\pm A_0^{\#} \bar{N}(y) e^{-j\varphi^{\#}(y)} \right) \right\} \right] \right) dy. \quad (12.8)$$

Данные соотношения соответствуют алгоритму функционирования классической петли фазовой автоподстройки, изображенной на рис.12.2. (ФНЧ – фильтр нижних частот; ГУН – генератор, управляемый напряжением).

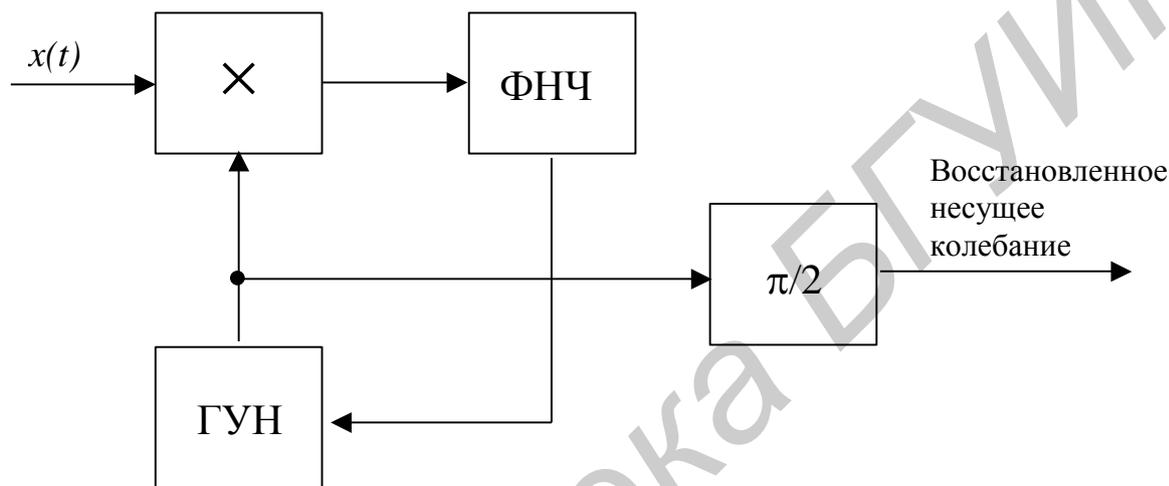


Рис. 12.2. Схема петли фазовой автоподстройки

Анализ полученных соотношений показывает, что в отсутствие ошибки синхронизации приращение оценки фазы (фазы опорного колебания) во времени равно нулю, при положительной ошибке оценка фазы начинает уменьшаться, при отрицательной ошибке – увеличиваться. При этом скорость изменения оценки фазы пропорциональна абсолютной величине ошибки. Таким образом, происходит автоматическое регулирование (синхронизация) так, чтобы величина ошибки $\Delta\varphi(t)$ поддерживалась равной нулю.

Наличие информационного рассогласования принимаемого и опорного колебаний приводит к невозможности отслеживания фазы несущего сигнала, поскольку усреднение во времени величины $(\pm A_0)(\pm A_0^{\#})\sin(\Delta\varphi(t))$ будет вне зависимости от существующей ошибки синхронизации всегда давать нулевой результат. Операция $(\pm A_0)(\pm A_0^{\#})$ (в общем случае $\bar{S}_{inf(t)}(\bar{S}_{inf}^{\#}(t))^*$) по своему существу является операцией снятия информационной модуляции с принимаемого сигнала.

Точность слежения за фазой определяется характеристиками динамики изменения фазы во времени и интенсивностью шума наблюдения. Вопросам обеспечения требуемых качественных характеристик слежения за фазой при-

нимаемого сигнала посвящено большое количество исследований, результаты которых являются отдельным предметом и нами рассматриваться не будут.

12.2.2. Методы синхронизации по несущему сигналу с обратной связью по решению

Как было показано выше, на процесс синхронизации по фазе несущего колебания существенное влияние оказывает согласование принимаемого сигнала и опорного колебания по информационной модуляции.

Правильное снятие информационной модуляции с принимаемого сигнала может быть осуществлено путем использования результатов (выходных сигналов) непосредственно демодулятора. При этом необходимо иметь в виду, что решение о передаваемых символах в процессе демодуляции принимается с запаздыванием на длительность данных символов. То есть решение принимается в конце интервала передачи конкретного символа (когда решение уже есть во входном сигнале, уже нет данного символа). В связи с этим для правильного информационного согласования принимаемого и опорного сигналов в схеме оценки частоты и фазы несущего колебания должна использоваться копия входной смеси сигнала и шума, задержанная на время запаздывания решения в демодуляторе. Следует отметить, что существенным условием нормального функционирования данного решения синхронизации по фазе несущего колебания являются медленные изменения фазы несущего колебания по сравнению со скоростью информационной модуляции (за время, соответствующее длительности информационного символа, фаза практически не изменяется). Структурная схема блока оценки частоты и фазы несущего колебания с обратной связью по решению приведена на рис. 12.3.

12.2.3. Методы синхронизации по несущему колебанию, не управляемые решением

Вместо использования схемы, управляемой решениями для получения оценки фазы, можно трактовать передаваемые данные как случайные величины и просто усреднить функцию правдоподобия по этим случайным величинам, а затем искать ее максимум. Чтобы выполнить такое усреднение, можно использовать или действительную функцию распределения вероятностей данных, если она известна, или можно предположить некоторое распределение вероятностей, которое является подходящим приближением для правильного распределения. Следует отметить, что в результате подобного усреднения вообще может пропасть информация о фазе несущего сигнала, так что подобная методика не является универсальной для большинства модуляционных схем.

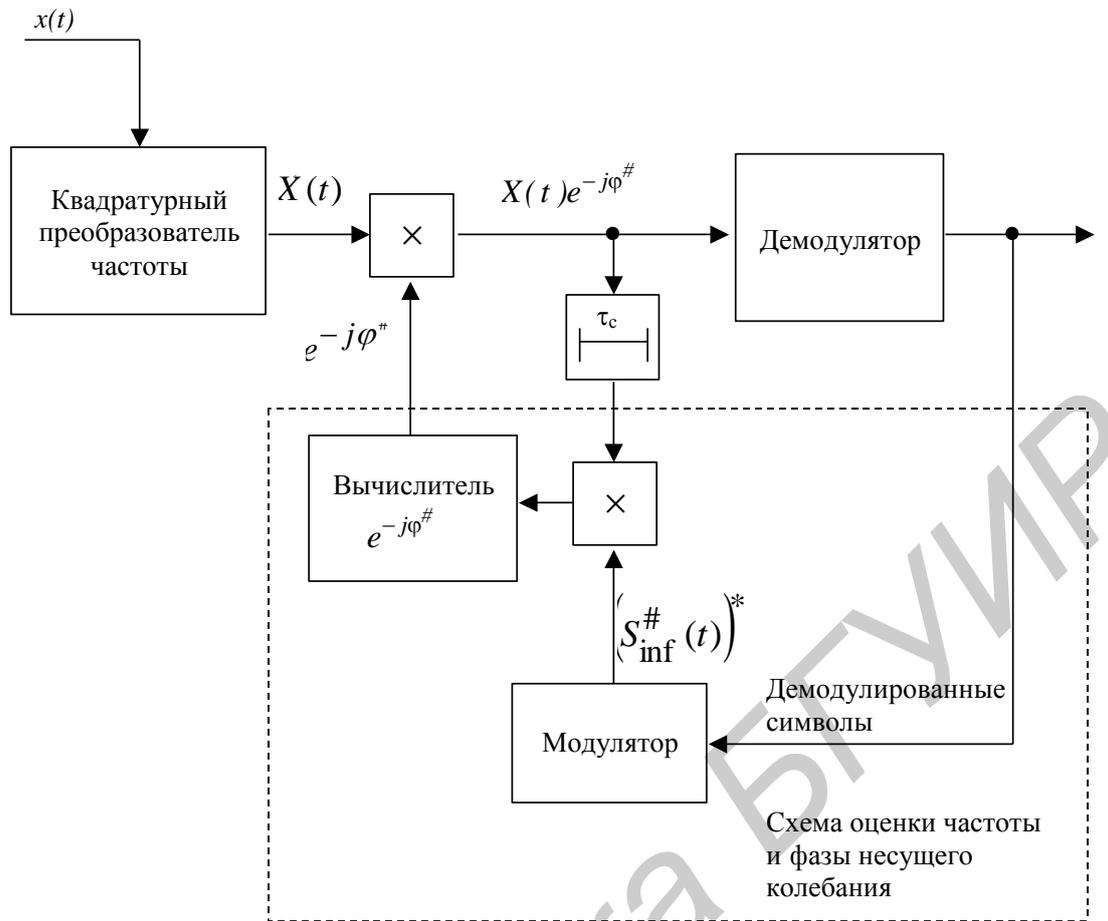


Рис. 12.3. Синхронизация по несущему колебанию с обратной связью по решению

Наряду с методами восстановления фазы несущего колебания, основанными на усреднении по ансамблю передаваемых информационных символов, широкое распространение получили «полуэвристические» методы, основанные на использовании свойств симметрии сигнальных созвездий некоторых модуляционных схем. Рассмотрим их более подробно.

12.2.3.1. Синхронизация при приеме сигнала с $\pm\pi/2$ -BPSK

Комплексная огибающая сигнала с произвольной бинарной фазовой манипуляцией полностью определяется его квадратурной и синфазной компонентами и представляется следующим выражением:

$$\vec{S}(t) = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) e^{jC_k \Theta} = A_0 \sum_{k=0}^{\infty} g_F(t - k\tau_b) (\cos[C_k \Theta] + j \sin[C_k \Theta]) . \quad (12.9)$$

Слежение за фазой несущего колебания в случае, когда используемая девиация фазы Θ отлична от величины $\pi/2$, не представляет сложности. Данное слежение осуществимо обычной петлей фазовой автоподстройки, поскольку модулированный сигнал содержит немодулированную компоненту, представляющую несущее колебание $\cos[C_k \Theta] = \cos(\Theta) = const$. В случае ко-

гда $\Theta = \pi/2$, данная компонента пропадает, что требует осуществления специальных мер по снятию информационной модуляции. Рассмотрим два наиболее «популярных» способа осуществления данной процедуры.

При $\pm\pi/2$ -BPSK комплексная огибающая для любого шага модуляции записывается следующим образом:

$$\bar{S}(t) = A_0 e^{jC_k \frac{\pi}{2}} = jA_0 C_k. \quad (12.10)$$

Напомним, что $C_k = \pm 1$, при этом знак определяется модулирующими символами.

С учетом неизвестной фазы несущего колебания имеем

$$\bar{S}(t) = A_0 e^{jC_k \frac{\pi}{2}} e^{j\varphi(t)} = jA_0 C_k e^{j\varphi(t)}. \quad (12.11)$$

Возведем комплексную огибающую в квадрат:

$$(\bar{S}(t))^2 = A_0^2 e^{jC_k \pi} e^{j2\varphi(t)} = -A_0^2 e^{j2\varphi(t)}. \quad (12.12)$$

Анализ полученного выражения показывает, что квадрат комплексной огибающей сигнала с $\pm\pi/2$ -BPSK свободен от модуляции информационными символами и содержит информацию об удвоенной фазе несущего колебания, которая может быть отслежена петлей фазовой автоподстройки (способ реализуется так называемой схемой Сифорова):

$$2 \frac{d}{dt} \varphi^\#(t) = -K \left[\frac{2}{N_0} \operatorname{Re} \left\{ j \left(A_0^2 e^{j2\varphi(t)} \right) \left(A_0^2 e^{-j2\varphi^\#(t)} \right) \right\} \right]. \quad (12.13)$$

При записи последнего выражения были опущены шумовые компоненты, определяемые результатом возведения комплексной огибающей входной смеси модулированного радиосигнала и шума в квадрат.

Второй способ использует идентичность информационной модуляции синфазной и квадратурной компонент сигнала с $\pm\pi/2$ -BPSK. Запишем комплексную огибающую принятого сигнала с учетом ее коррекции оценкой фазы несущей $\varphi^\#(t)$:

$$\begin{aligned} \bar{S}(t) &= jA_0 C_k e^{j\varphi(t) - j\varphi^\#(t)} = \\ &= A_0 \left(jC_k \cos(\varphi(t) - \varphi^\#(t)) - C_k \sin(\varphi(t) - \varphi^\#(t)) \right), \end{aligned} \quad (12.14)$$

$$S_I(t) = A_0 C_k \sin(\Delta\varphi(t)), \quad (12.15)$$

$$S_Q(t) = A_0 C_k \cos(\Delta\varphi(t)), \quad (12.16)$$

$$\Delta\varphi(t) = \varphi^\#(t) - \varphi(t). \quad (12.17)$$

Произведение квадратурных компонент дает следующий результат:

$$S_I(t)S_Q(t) = A_0^2 (C_k)^2 \sin(\Delta\varphi(t))\cos(\Delta\varphi(t)) = \frac{A_0^2}{2} \sin(2\Delta\varphi(t)). \quad (12.18)$$

В этом случае также происходит съём информационной модуляции, а результат перемножения содержит информацию об удвоенной фазе (или ошибке отслеживания фазы) несущего колебания, которая может отслеживаться обычной петлей фазовой автоподстройки (способ реализуется так называемой схемой Костаса).

Следует заметить, что схемы Костаса и Сифорова идентичны, поскольку в обеих схемах в основе оценки фазы лежит результат перемножения синфазной и квадратурной компонент принятого сигнала.

Структурные схемы Сифорова и Костаса, реализованные в области нулевой и радиочастот, приведены на рис. 12.4 и 12.5 соответственно.

12.2.3.2. Синхронизация при приеме сигнала с M-PSK

Комплексная огибающая сигнала на интервале действия одного информационного символа в случае сигнала с M-PSK определяется следующим образом:

$$\vec{S}(t) = A_0 e^{j\Theta_m}, \quad (12.19)$$

где $\Theta_m = \frac{2\pi(m-1)}{M}$ – разрешенные фазовые состояния сигнала представляющие различные информационные символы; $m = 1, \dots, M$.

С учетом неизвестной фазы несущего колебания имеем

$$\vec{S}(t) = A_0 e^{j\frac{2\pi(m-1)}{M}} e^{j\varphi(t)}. \quad (12.20)$$

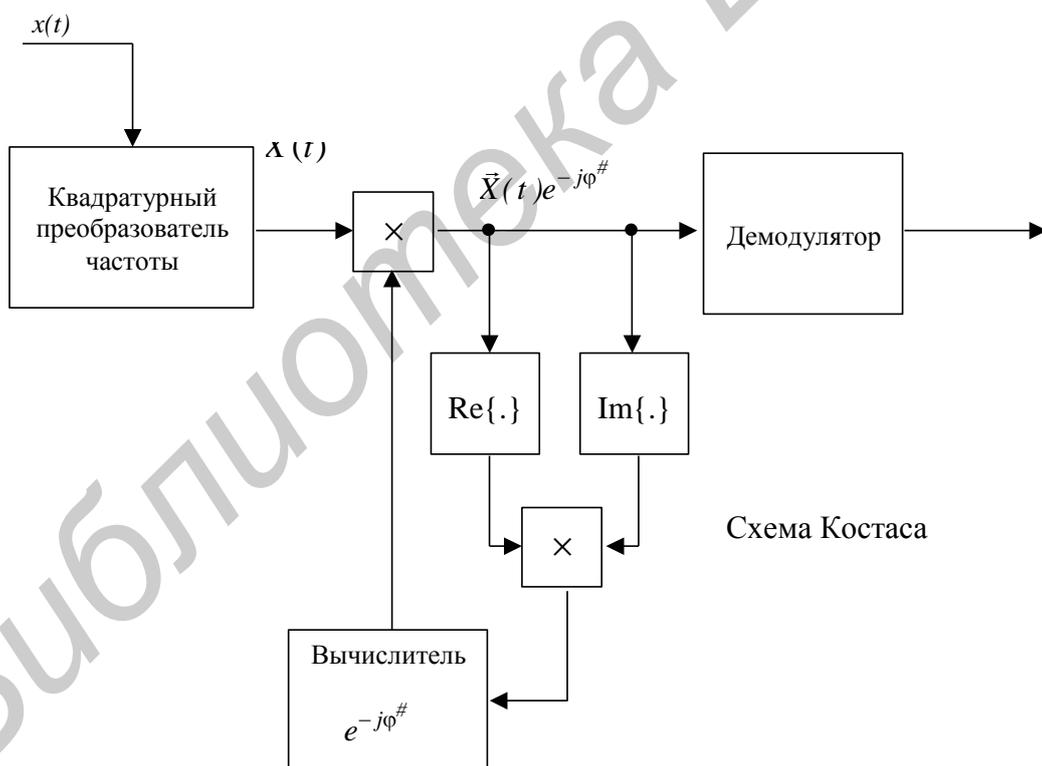
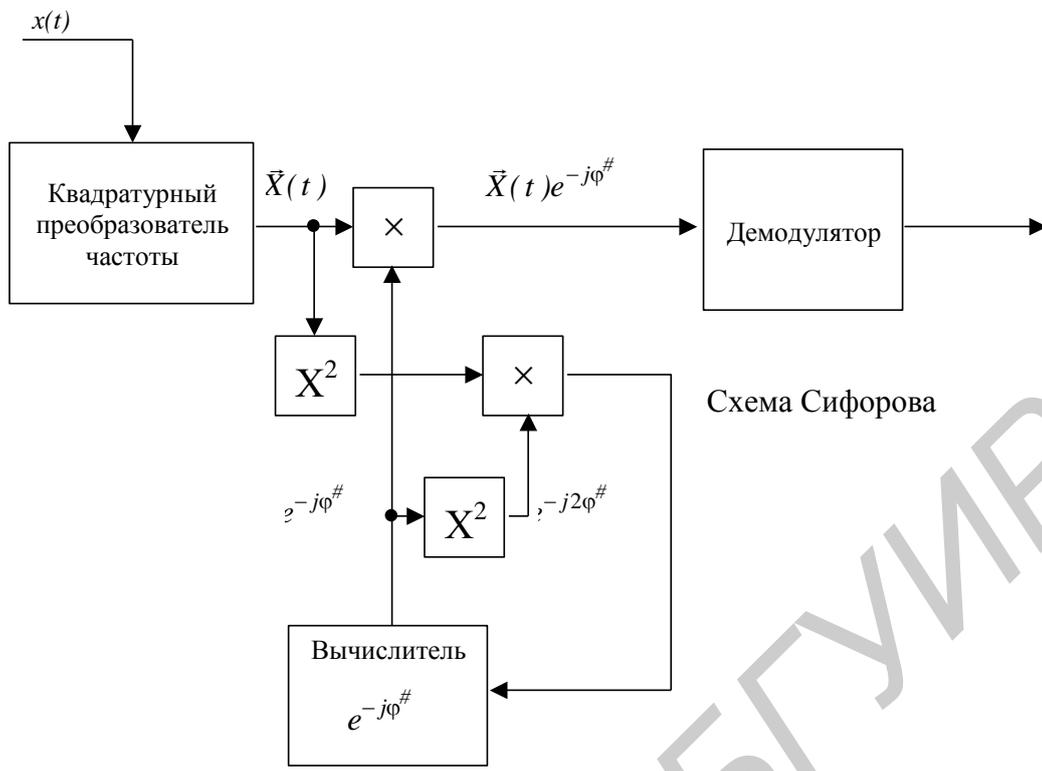


Рис. 12.4. Компенсация фазы несущего колебания при приеме сигнала с $\pm\pi/2$ -BPSK

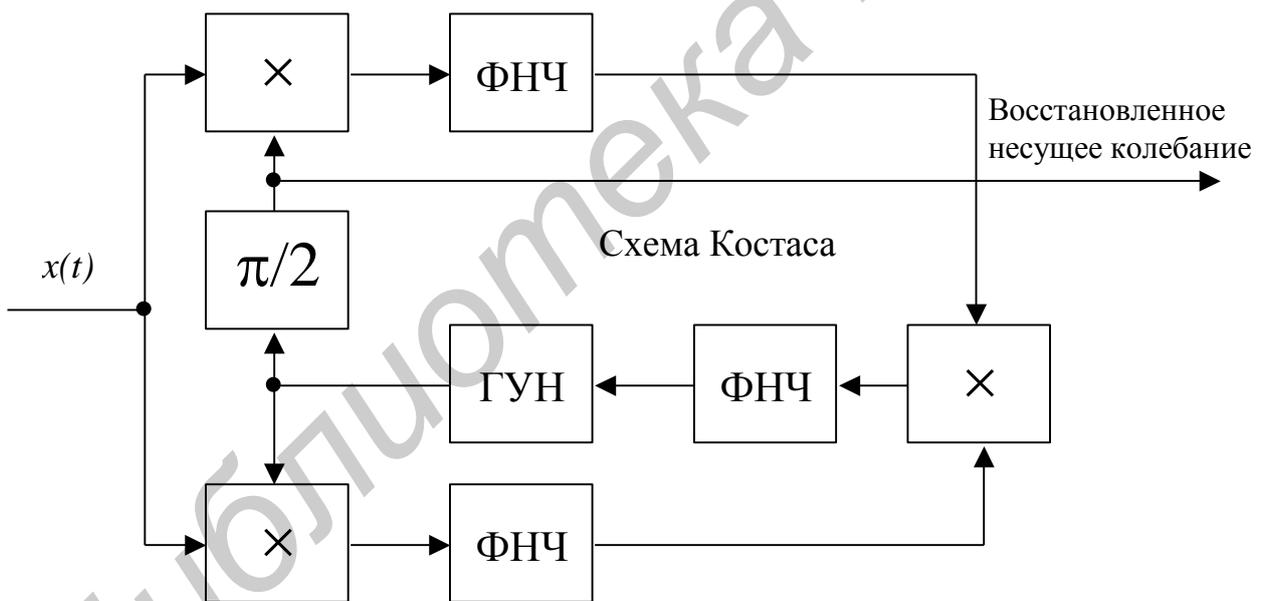
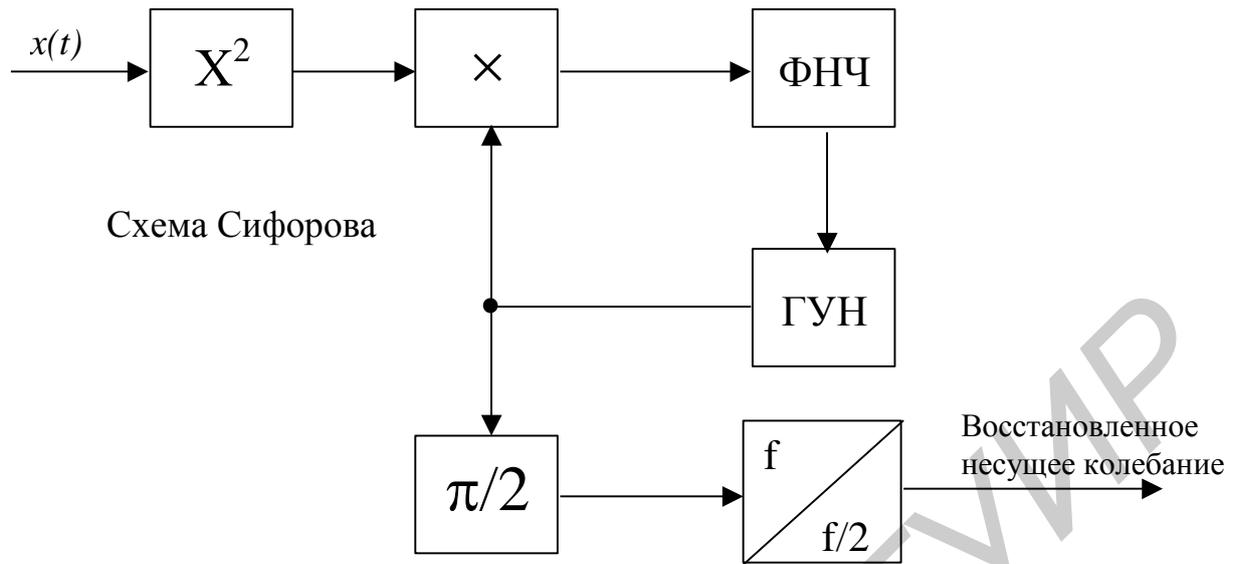


Рис. 12.5. Восстановление несущего колебания при приеме сигнала с $\pm\pi/2$ -BPSK

Возведем комплексную огибающую в M -ю степень:

$$(\vec{S}(t))^M = A_0^M e^{j2\pi(m-1)t} e^{jM\varphi(t)} = A_0^M e^{jM\varphi(t)}. \quad (12.21)$$

Таким образом, M -я степень комплексной огибающей сигнала с M -PSK свободна от модуляции информационными символами и содержит информацию о фазе несущего колебания, которая может быть отслежена петлей фазовой автоподстройки (обобщенная на случай многопозиционной фазовой манипуляции схема Сифорова). Выражение, описывающее работу данной схемы, записано ниже (шумовые компоненты при записи опущены):

$$M \frac{d}{dt} \varphi^{\#}(t) = -K \left[\frac{2}{N_0} \operatorname{Re} \left\{ j \left(A_0^M e^{jM\varphi(t)} \right) \left(A_0^M e^{-jM\varphi^{\#}(t)} \right) \right\} \right]. \quad (12.22)$$

На рис. 12.6 изображены структурные схемы слежения за фазой несущего колебания при приеме сигналов с многопозиционной фазовой манипуляцией для случаев обработки на радиочастоте и в области нулевой частоты.

12.2.3.3. Синхронизация при приеме сигнала с QPSK и OQPSK

Сигналы с квадратурной фазовой манипуляцией по своей сути являются частным случаем многопозиционной фазовой манипуляции M -PSK ($M = 4$). Для восстановления несущего колебания при их приеме возможно использование обобщенной схемы Сифорова. В качестве альтернативного варианта слежения за изменениями фазы несущего колебания может быть использован вариант схемы Костаса, расширенный на прием сигналов с 4-позиционной фазовой манипуляцией. Рассмотрим его подробнее.

Комплексная огибающая принятого сигнала с учетом ее коррекции оценкой фазы несущей $\varphi^{\#}(t)$ для сигналов с QPSK и OQPSK записывается следующим образом:

$$\begin{aligned} \bar{S}(t) &= A_0 (C_I + jC_Q) e^{j\varphi(t) - j\varphi^{\#}(t)} = \\ &= A_0 \left([C_I \cos(\Delta\varphi(t)) + C_Q \sin(\Delta\varphi(t))] + j [C_Q \cos(\Delta\varphi(t)) - C_I \sin(\Delta\varphi(t))] \right), \end{aligned} \quad (12.23)$$

$$S_I(t) = A_0 [C_I \cos(\Delta\varphi(t)) + C_Q \sin(\Delta\varphi(t))], \quad (12.24)$$

$$S_Q(t) = A_0 [C_Q \cos(\Delta\varphi(t)) - C_I \sin(\Delta\varphi(t))], \quad (12.25)$$

$$\Delta\varphi(t) = \varphi^{\#}(t) - \varphi(t), \quad (12.26)$$

$$C_I; C_Q = \pm 1. \quad (12.27)$$

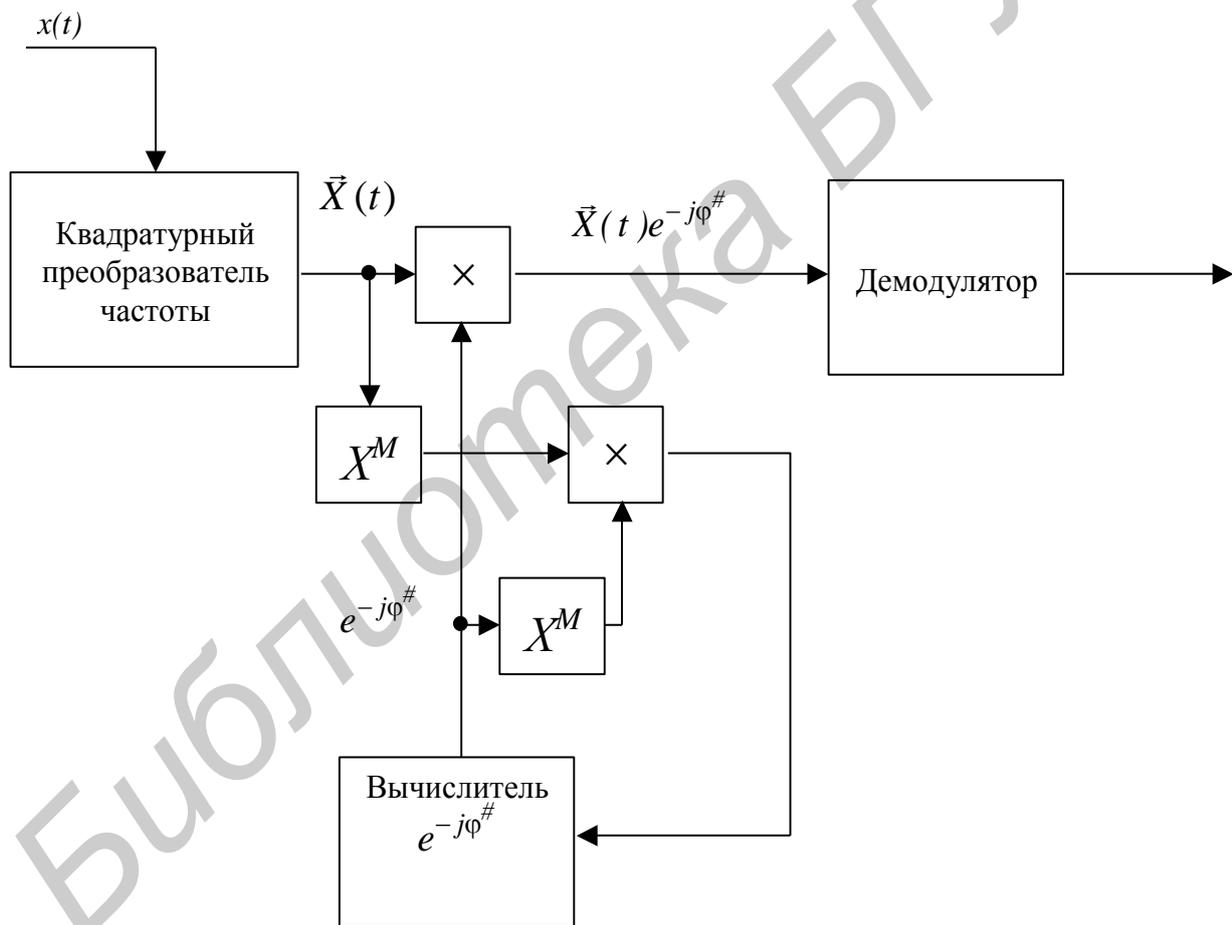
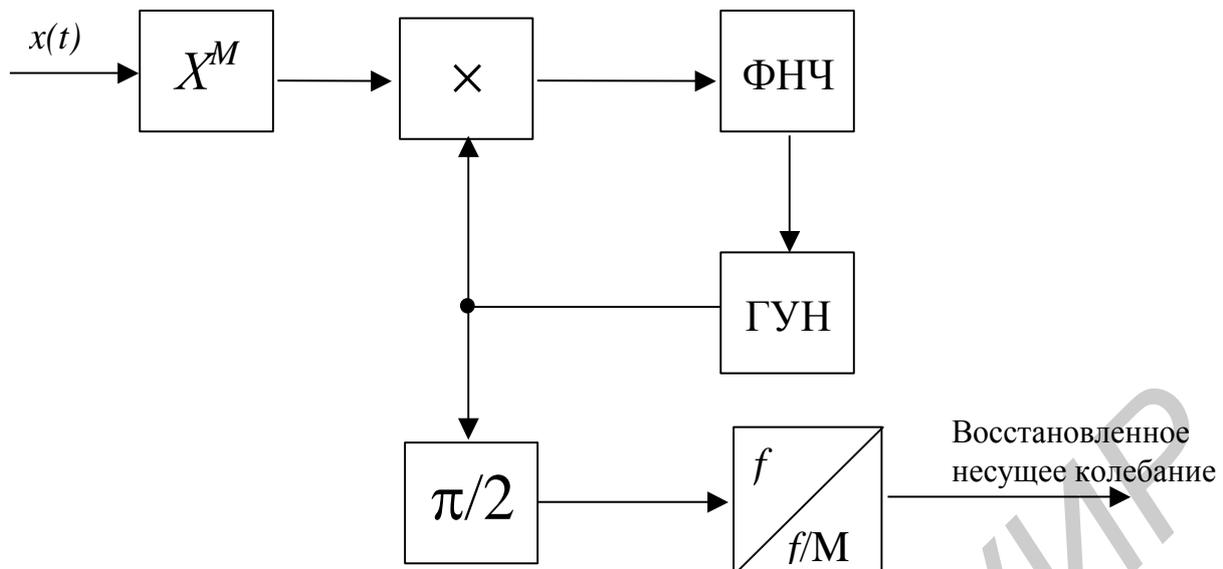


Рис. 12.6. Обобщенная схема Сифорова

Рассмотрим $\pm\pi/4$ -окрестность нулевой ошибки синхронизации ($\Delta\varphi = 0$). В данной окрестности знак синфазной составляющей $S_I(t)$ будет определяться знаком символа C_I , а знак квадратурной компоненты

$S_Q(t)$ – знаком символа C_Q или, с другой стороны, выполняются следующие равенства:

$$C_I = \text{sign}(S_I(t)), \quad (12.28)$$

$$C_Q = \text{sign}(S_Q(t)). \quad (12.29)$$

Выполнив перемножения следующего вида, получим

$$S_I(t) \text{sign}(S_Q(t)) = A_0(C_I C_Q \cos(\Delta\varphi(t)) + \sin(\Delta\varphi(t))), \quad (12.30)$$

$$S_Q(t) \text{sign}(S_I(t)) = A_0(C_I C_Q \cos(\Delta\varphi(t)) - \sin(\Delta\varphi(t))). \quad (12.31)$$

В результате окончательная операция снятия информационной модуляции может быть определена следующим образом:

$$S_I(t) \text{sign}(S_Q(t)) - S_Q(t) \text{sign}(S_I(t)) = 2A_0 \sin[\Delta\varphi(t)]. \quad (12.32)$$

Соответствующие структурные схемы приведены на рис. 12.7.

12.2.3.4. Синхронизация при приеме сигнала с QAM

В качестве примера рассмотрим синхронизацию по фазе несущего колебания сигнала 16-QAM с квадратным созвездием сигнала, изображенным на рис. 12.8. Как видно из рисунка, сигнальное созвездие содержит три группы точек, расположенных на перпендикулярных прямых. Первая группа точек 0000, 0101, 1111, 1010, 0010, 0111, 1101, 1000; вторая 0100, 1110, 0011, 1001 и третья 0001, 1011, 0110, 1100. При последовательном осуществлении амплитудного ограничения в пределах приведенной окружности и возведении в четвертую степень ограниченного сигнала, каждая из групп точек преобразуется в одну. Таким образом, вместо сигнального созвездия из 16 точек получится сигнальное созвездие, содержащее только три точки (рис. 12.9). На рисунке цифры, которыми помечены сигнальные точки, соответствуют группам точек исходного созвездия.

При равной вероятности появления символов в исходном сигнале вероятность появления сигнала, соответствующего точке 1, будет в два раза превосходить вероятность появления каждого из оставшихся двух сигналов (точки 2 и 3).

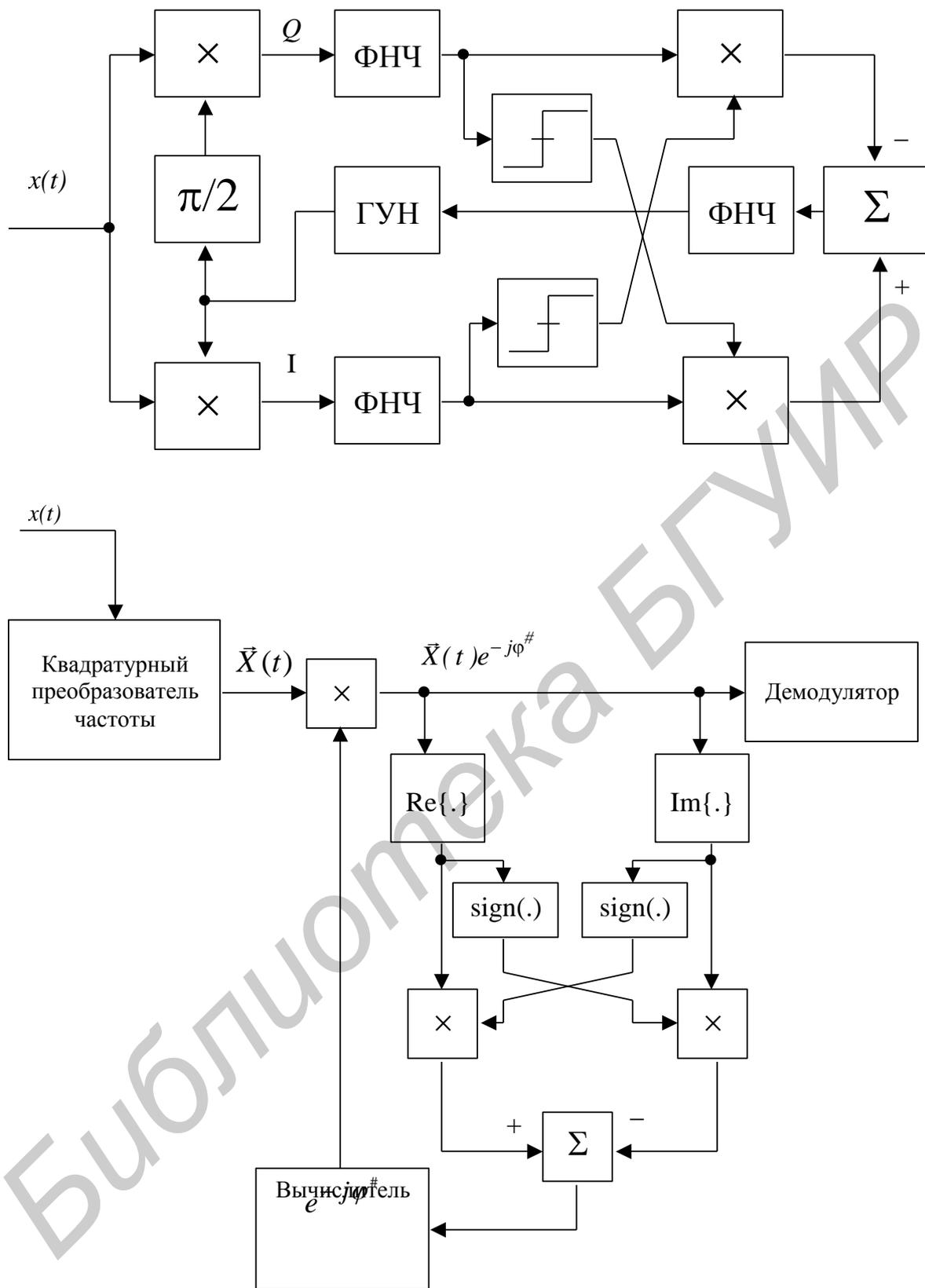


Рис. 12.7. Схема Костаса для случая приема сигналов с QPSK и OQPSK

Указанные выше два последовательных преобразования уже позволяют отслеживать фазовые изменения несущего колебания с использованием схемы фазовой автоподстройки. При этом значение отслеживаемой фазы будет

соответствовать фазовому положению точки 1. Следует отметить, что две оставшиеся точки можно рассматривать как сигнал с BPSK (девиация фазы отлична от $\pi/2$), при этом немодулированная компонента данного сигнала (несущее колебание) находится в противофазе сигнальной точке 1, что приводит к неоправданным потерям энергии сигнала и ухудшению помехоустойчивости процесса слежения за фазой. Для исключения данного обстоятельства целесообразно преобразованный сигнал подвергнуть делению частоты на четыре. В результате данного деления частоты созвездие преобразованного сигнала приобретет вид, приведенный на рис.12.10.

У сигнала с созвездием, соответствующим изображенному на рис. 12.10, практически вся энергия сосредоточена в спектральных компонентах совпадающих по фазе с первой точкой.

12.1.1.1. Заключительные замечания

В заключение необходимо отметить, что приведенные выше методы синхронизации по фазе несущего колебания обладают существенным недостатком. Данный недостаток заключается в том, что синхронизация осуществляется с точностью до постоянных заранее определенных фазовых углов. Например, для сигнала с QPSK данные фазовые ошибки могут составлять $\pm\pi/2$ и π . Это может приводить к явлению так называемой обратной работы, описанному ранее. Методы борьбы с обратной работой известны и здесь не рассматривались.

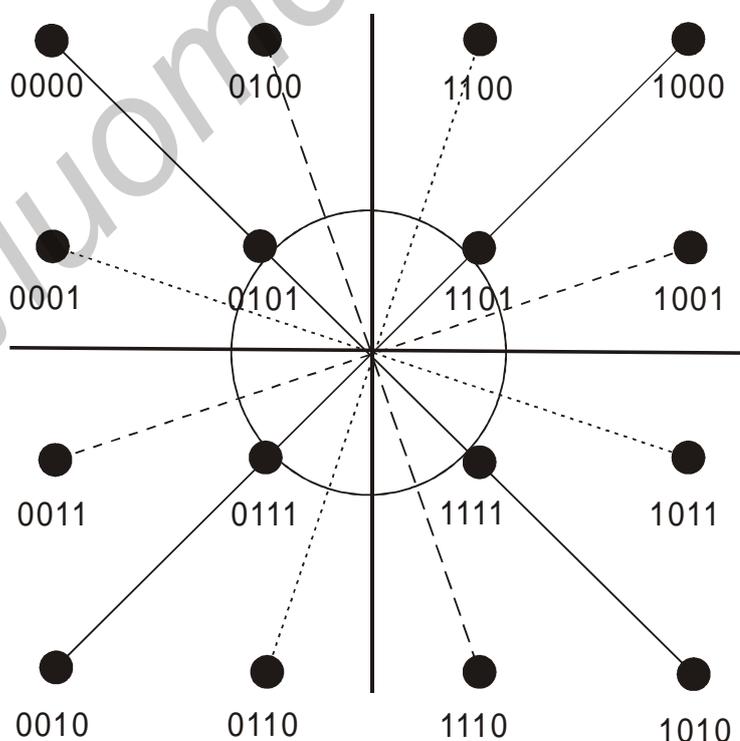


Рис. 12.8. Сигнальное созвездие 16-QAM

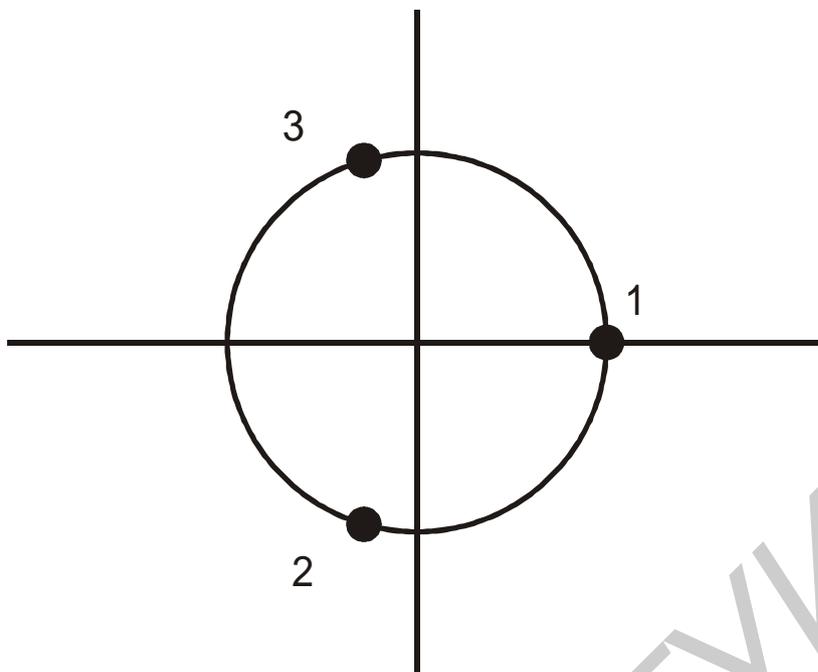


Рис. 12.9. Сигнальное созвездие 16-QAM после ограничения и возведения в четвертую степень

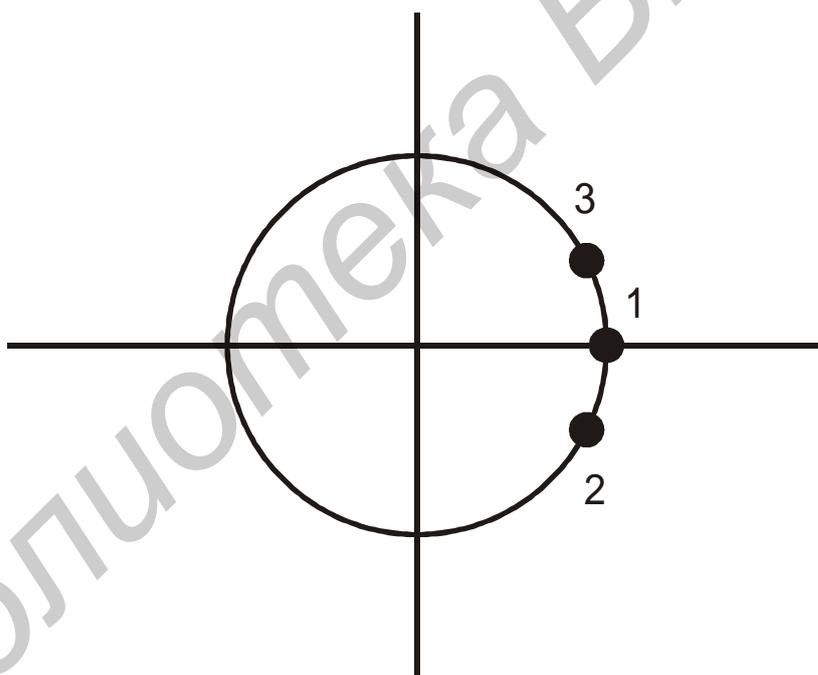


Рис. 12.10. Сигнальное созвездие преобразованного сигнала 16-QAM после деления частоты

12.2. Методы восстановления тактовой синхронизации

В цифровой системе связи выход демодулятора должен периодически стробироваться со скоростью передачи символов в точно выбранных моментах $t_m = m\tau_c + \tau$, где τ_c – длительность модулирующего символа, а τ – номинальное время задержки, которое определяется временем распространения сигнала от передатчика к приёмнику. Чтобы выполнять это периодическое

стробирование, требуется синхронизирующий сигнал в приемнике. Процесс получения такого сигнала в приёмнике обычно называется тактовой синхронизацией (синхронизацией символов), или восстановлением отсчётов времени.

Восстановление отсчётов времени – одна из наиболее ответственных функций, которые выполняются в приёмнике цифровых систем связи. Отметим, что приёмник должен знать не только частоту ($1/\tau_c$), с которой стробируются выходы согласованных фильтров или корреляторов, но также тот момент времени, в который необходимо взять отсчёты внутри каждого символьного интервала. Положение отсчётного момента внутри символьного интервала длительностью τ_c называют фазой синхронизации. Тактовую синхронизацию можно выполнить несколькими путями. В некоторых системах связи синхрогенераторы передатчика и приёмника синхронизированы по образцовому таймеру, который обеспечивает высокостабильную сетку времени (например, системы GPS, CdmaOne, CDMA2000 и UMTS). В этом случае приёмник должен оценить и скомпенсировать только относительное время задержки между переданным и принятыми сигналами.

Другой метод для достижения синхронизации символов заключается в одновременной передаче передатчиком опорной частоты, равной или кратной $1/\tau_c$, вместе с информационным сигналом. Приёмник может просто использовать узкополосный фильтр, настроенный на переданную опорную частоту, и таким образом извлечь синхронизирующий сигнал для стробирования. Этот подход прост для применения. Здесь имеется, однако, несколько неудобств. Первое заключается в том, что передатчик должен отдать некоторую часть возможной мощности для передачи опорного сигнала. Другое заключается в том, что некоторая малая часть доступной в канале полосы частот должна быть занята для передачи опорного сигнала. Несмотря на эти неудобства, этот метод часто используется в системах телефонной связи, которые используют большие полосы частот для передачи сигналов многих пользователей. В таком случае затраты на передачу синхронизирующего сигнала делятся на многих пользователей. Посредством такого деления потери в мощности передатчика и полосе частот уменьшаются пропорционально числу пользователей.

Синхронизирующий сигнал можно также извлечь из принимаемого информационного сигнала. Имеется определённое число различных методов, которые можно использовать в приёмнике, чтобы достичь самосинхронизации. В этом разделе мы обсудим как метод оценивания, управляемого решениями, так и метод оценивания, не управляемого решениями.

12.2.1. Тактовая синхронизация, управляемая решением

Обобщенное правило формирования текущей оценки времени запаздывания сигнала для случая его наблюдения на фоне белого гауссовского шума в установившемся режиме определяется следующим образом (данное правило справедливо для оценки любого параметра сигнала и уже нами использо-

валось при рассмотрении методов синхронизации по фазе несущего колебания):

$$L(\tau^\#(t)) = K \frac{\partial}{\partial \tau^\#(t)} \left[-\frac{1}{N_0} (\bar{X}(t) - \bar{S}(t, \tau^\#(t))) \times (\bar{X}(t) - \bar{S}(t, \tau^\#(t)))^* \right], \quad (12.33)$$

поскольку время запаздывания сигнала является неэнергетическим параметром, уравнение для оценки можно преобразовать к виду

$$L(\tau^\#(t)) = K \frac{\partial}{\partial \tau^\#(t)} \left[\frac{2}{N_0} \operatorname{Re} \{ \bar{X}(t) \bar{S}^*(t, \tau^\#(t)) \} \right]. \quad (12.34)$$

Рассмотрим комплексную огибающую полезного сигнала (для простоты будем рассматривать только методы линейной модуляции – PSK, QAM и др.).

$$\bar{S}(t, \varphi(t, \tau), \tau) = \bar{S}_{inf}(t - \tau) e^{j\varphi(t, \tau)}. \quad (12.35)$$

В общем случае фаза несущего колебания зависит от времени запаздывания. Мы будем считать, что синхронизация по фазе несущего колебания уже осуществлена. В этом случае комплексная огибающая имеет вид

$$\bar{S}(t - \tau) = \bar{S}_{inf}(t - \tau). \quad (12.36)$$

С учетом этого имеем следующее уравнение для оценки времени запаздывания сигнала:

$$L(\tau^\#(t)) = K \frac{\partial}{\partial \tau^\#(t)} \left[\frac{2}{N_0} \operatorname{Re} \{ \bar{X}^{cor}(t) \bar{S}_{inf}^*(t - \tau^\#(t)) \} \right], \quad (12.37)$$

где $\bar{X}^{cor}(t)$ – результат коррекции принятой смеси полезного сигнала и шума с учетом синхронизации по фазе несущего колебания.

Структурная схема, реализующая полученное уравнение для оценки времени запаздывания сигнала, приведена на рис. 12.11.

Снятие информационной модуляции с принимаемого сигнала осуществляется результатами демодуляции информационных символов. Поскольку решение о значении передаваемых символов в процессе демодуляции принимается с запаздыванием на длительность данных символов, в цепи тактовой синхронизации используется скорректированная копия смеси сигнала и шума, задержанная на длительность информационного символа.

12.2.2. Тактовая синхронизация, неуправляемая решением

Оценка параметров задержки принимаемого сигнала, неуправляемая решением, возможна путем усреднения функции правдоподобия с учетом вероятностных характеристик передаваемых символов.

Среди устройств тактовой синхронизации, не управляемых решением, наиболее широкое распространение получили нелинейные синхронизаторы. Поскольку в спектре радиосигнала, как правило, отсутствуют частотные компоненты, представляющие необходимую тактовую синхропоследовательность, устройства синхронизации этого класса генерируют данные компоненты, пропуская поступающий низкочастотный сигнал (сигнал после квадратурного преобразования частоты) через последовательность фильтра и нелинейного устройства. На рис. 12.12 приведены три примера нелинейных тактовых синхронизаторов.

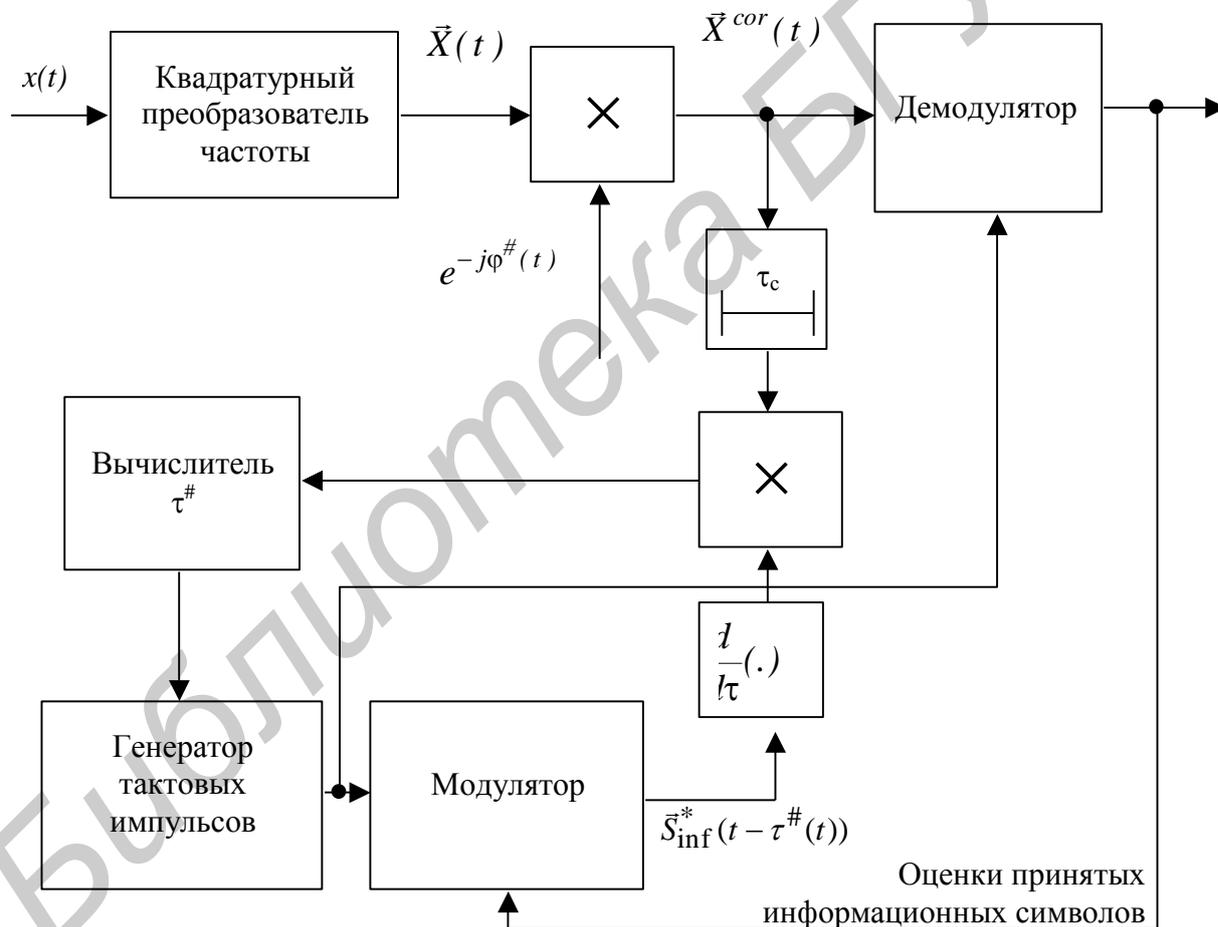
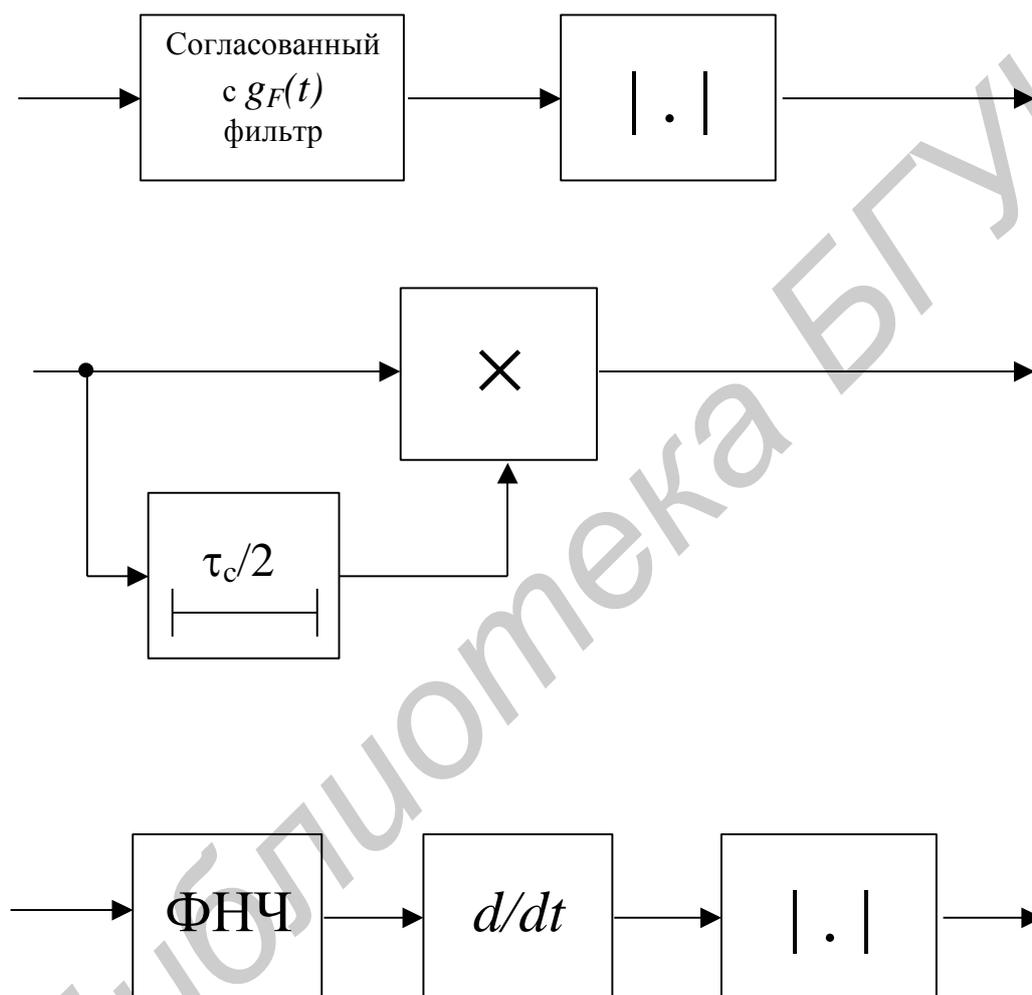


Рис. 12.11. Тактовая синхронизация с обратной связью по решению

В первом примере поступающий сигнал фильтруется с использованием согласованного формирующим импульсом $g_F(t)$ фильтра. Выходной сигнал этого фильтра представляет собой автокорреляционную функцию импульса $g_F(t)$. Например, при передаче с помощью прямоугольных формирующих

импульсов на выходе будем иметь треугольный сигнал. Затем полученная последовательность автокорреляционных импульсов выпрямляется с помощью некоторой нелинейности четного порядка, например квадратичного устройства или устройства взятия модуля. Полученный таким образом сигнал будет содержать пики положительной амплитуды, которые с точностью до временной задержки соответствуют переходам входных символов. Таким образом, сигнал с выхода нелинейности будет содержать компоненту на требуемой частоте и с требуемой фазой тактового генератора. Данная частотная составляющая может быть использована для формирования синхронизированного тактового сигнала.



к устройству выделения и формирования сигнала тактовой синхронизации

Рис. 12.12. Нелинейные тактовые синхронизаторы

Во втором примере (рис. 12.12) компонента на тактовой частоте создается посредством задержки и умножения. Длительность задержки равна половине периода передачи символа, и это значение является оптимальным, поскольку оно обеспечивает наибольшую амплитуду получаемой синхронизирующей составляющей. Как и ранее, нужная спектральная компонента мо-

жет быть использована для формирования синхронизированного тактового сигнала.

Последний пример соответствует контурному детектору. Основными операциями здесь являются дифференцирование и выпрямление (посредством использования квадратичного устройства). Если на вход поступает сигнал прямоугольной формы, дифференциатор дает положительные или отрицательные пики на всех переходах символов. При выпрямлении получаемая последовательность положительных импульсов будет формировать Фурье-компоненту, соответствующую тактовой синхронизации символов. Потенциальной проблемой данной схемы является то, что дифференциаторы обычно весьма чувствительны к широкополосному шуму. Это делает необходимым введение перед дифференциатором фильтра нижних частот, как показано на рис. 12.12.

12.3. Кадровая синхронизация

Практически все потоки цифровых данных имеют некоторую кадровую структуру. Другими словами поток данных разбит на равные группы символов. Если поток данных – это оцифрованный телесигнал, каждый пиксел в нем представляется словом из нескольких бит, которые группируются в горизонтальные растровые развертки, а затем в вертикальные растровые развертки. Компьютерные данные обычно разбиваются на слова, состоящие из некоторого числа 8-битовых байт, которые, в свою очередь, группируются в пакеты, кадры или файлы. Любая система, использующая кодирование с защитой от блочных ошибок, в качестве основы кадра должна брать длину кодового слова. Оцифрованная речь обычно передается пакетами или кадрами, неотличимыми от других цифровых данных.

Чтобы входной поток данных имел смысл для приемника, приемник должен синхронизироваться с кадровой структурой потока данных. Кадровая синхронизация обычно выполняется с помощью некоторой специальной процедуры передатчика. Данная процедура может быть как простой, так и довольно сложной, в зависимости от среды, в которой должна функционировать система.

Простейшим методом, используемым для облегчения кадровой синхронизации, является введение маркера. Маркер кадра – это отдельный символ или краткая последовательность символов, периодически вводимая передатчиком в поток данных. Приемник должен знать эту последовательность и период ее введения. Приемник, достигший синхронизации данных, сопоставляет (проверяет корреляцию) эту известную последовательность с потоком поступающих данных в течение известного периода введения. Если приемник не синхронизирован с кадровой последовательностью, корреляция будет слабой. При синхронизации приемника с кадровой структурой корреляция будет практически идеальной, повредить которую может только случайная ошибка детектирования.

Преимуществом маркера кадра является его простота. Для маркера может быть достаточно даже одного символа, если перед принятием решения, находится ли система в состоянии кадровой синхронизации, было выполнено достаточное число корреляций. Основным недостатком состоит в том, что данное достаточное число может быть очень большим; следовательно, большим может быть и время, требуемое для достижения синхронизации. Таким образом, наибольшую пользу маркеры кадров представляют в системах, непрерывно передающих данные, подобно многим телефонным и компьютерным каналам связи, и не подходят для систем, передающих отдельные пакеты, или систем, требующих быстрого получения кадровой синхронизации. Еще одним недостатком маркера кадра является то, что введенный символ (символы) может повысить громоздкость структуры потока данных.

В системах с неустойчивыми или пульсирующими передачами либо в системах с необходимостью быстрого получения синхронизации рекомендуется использовать синхронизирующие кодовые слова. Обычно такие кодовые слова передаются как часть заголовка сообщения. Приемник должен знать кодовое слово и постоянно искать его в потоке данных, возможно, используя для этого коррелятор на согласованных фильтрах. Детектирование кодового слова укажет известную позицию (обычно – начало) информационного кадра. Преимуществом подобной системы является то, что кадрковая синхронизация может достигаться практически мгновенно. Единственная задержка – отслеживание кодового слова. Недостаток – кодовое слово, выбираемое для сохранения низкой вероятности ложного детектирования, может быть длинным по сравнению с маркером кадра. Здесь стоит отметить, что сложность определения корреляции пропорциональна длине последовательности, поэтому при использовании кодового слова коррелятор может быть относительно сложным.

12.4. Заключительные замечания

Точность тактовой синхронизации и синхронизации по фазе несущего колебания оказывает существенное влияние на вероятность ошибочного приема информационных символов. Однако узкополосность процессов изменения фазы несущего колебания и времени запаздывания позволяет осуществлять эффективную синхронизацию (реализующую характеристики помехоустойчивости процесса демодуляции, близкие к потенциальным). Следует отметить, что ввиду сильной узкополосности (по сравнению с информационным трактом) устройств синхронизации для их входа в стационарный следящий режим требуется достаточно большой промежуток времени. Это особенно касается схем синхронизации управляемых решением. Для ускорения процесса входа в синхронный режим применяют специальные методы поиска (обучения) устройств синхронизации.

13. УПЛОТНЕНИЕ И МНОЖЕСТВЕННЫЙ ДОСТУП

13.1. Общие принципы осуществления множественного доступа к каналам связи

Создание любой системы электросвязи всегда подразумевает использование некоторого ресурса связи, и эффективность такой системы связи в первую очередь оценивается эффективностью использования ее ресурса.

Ресурсом связи называется совокупность физических характеристик канала связи, доступных для передачи информации.

В системе связи критерием эффективности использования ресурса связи является достижение максимальной пропускной способности (общей скорости передачи данных) в реализованных ею каналах.

Существует лишь 3 способа увеличения пропускной способности. Первые 2 непосредственно следуют из уравнения Шеннона:

$$C \leq \Delta F \log_2 \left(1 + \frac{E_b}{N_0} \right), \text{ бит/с,}$$

где ΔF – ширина полосы частот канала; $\frac{E_b}{N_0}$ – отношение энергии, приходящейся на один бит передаваемой информации, к спектральной плотности шума.

Таким образом, для увеличения пропускной способности возможно:

– за счет увеличения ΔF ;

– за счет увеличения $\frac{E_b}{N_0}$.

Третий способ увеличения C заключается в увеличении эффективности доступа к ресурсу связи. Последнее означает необходимость такого планирования распределения ресурса связи между пользователями системы связи, при котором, все физические свойства канала связи использовались бы максимально эффективно.

Распределение ресурса связи в свою очередь задается методом множественного доступа, под которым понимается способ объединения сигналов (информационных потоков) различных пользователей для совместного использования ресурса связи.

Рассмотрим наиболее общие вопросы распределения ресурса связи между многими пользователями.

Будем считать, что системе связи выделен или определен некоторый ресурс связи. Или, иными словами, определена некоторая область (пространство) в общей совокупности каналов, которую может занимать данная система связи как физический объект. Например, канальное пространство, выделенное системе связи, может быть ограничено диапазоном используемых частот, диапазоном во времени, когда разрешена передача, территорией, где

разрешена работа, разрешенными пространственными характеристиками (свойствами) используемых электромагнитных колебаний и пр.

Доступ к ресурсу связи для различных пользователей должен осуществляться на принципах исключения взаимного влияния передачи информации одних пользователей на передачу информации другими пользователями.

Исходя из этого наиболее общее условие осуществления множественного доступа (условие ортогональности выделяемого различным пользователям ресурса) может быть сформулировано следующим образом:

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} R_i(\vec{\lambda}) R_j(\vec{\lambda}) d\vec{\lambda} \Big|_{i \neq j} = 0, \quad (13.1)$$

где $R_i(\vec{\lambda})$ – функция, характеризующая ресурс, выделяемый i -му пользователю; $\vec{\lambda}$ – вектор параметров, характеризующий особенности имеющегося ресурса связи.

Определим основные параметры, характеризующие выделенный (используемый) ресурс связи.

1. Диапазон используемых частот. Интуитивно понятно, что если выделенные пользователям ресурсы $R_i(\vec{\lambda})$ характеризуются непересекающимися диапазонами частот, то интегрирование по частоте произведений ресурсов выделяемых различным пользователям будет (в идеальном случае) всегда приводить к нулевому значению. На данном свойстве ресурса связи основан метод множественного доступа FDMA (Frequency Division Multiple Access) или множественный доступ с частотным разделением. В этом случае пользователю выделяются определенные поддиапазоны закрепленной за системой связи полосы частот.

2. Время передачи. Если различные пользователи осуществляют передачу на непересекающихся интервалах времени, то в конкретный интервал времени ненулевое значение имеет ресурс только одного из абонентов и, следовательно, обеспечивается выполнение общего условия множественного доступа, поскольку ресурсы всех других абонентов в это момент времени равны нулю. Данное свойство используется для осуществления множественного доступа по методу TDMA (Time Division Multiple Access) или множественного доступа с временным разделением. Каждому пользователю выделяются периодически временные интервалы, не пересекающиеся с временными интервалами, выделяемыми другим пользователям.

3. Форма несущего колебания. В наиболее общей трактовке отмеченные выше методы реализации множественного доступа тоже используют отличия в форме несущего колебания, поскольку использование различных поддиапазонов частот (при FDMA) означает использование различных несущих колебаний (например, $\cos(\omega_1 t)$ и $\cos(\omega_2 t)$), именно отличия данных ко-

лебаний обеспечивают выполнение условия ортогональности выделяемого различным пользователями ресурса), аналогичная ситуация имеется и при реализации TDMA. В этом случае используются различные (не перекрывающиеся во времени) импульсные несущие колебания. На основании вышеизложенного необходимо уточнить, что когда идет речь об отличиях формы несущего колебания, как правило, подразумевают более «тонкие» отличия несущих колебаний и предполагается, что возможно выполнение условия ортогональности и при совпадении несущей частоты и времени передачи. Использование данной характеристики канала (ресурса связи) используется методом множественного доступа CDMA (Code Division Multiple Access) или множественного доступа с кодовым разделением. Пользователю выделяются определенные элементы набора ортогонально (либо почти ортогонально) распределенных кодов (несущих сигналов), каждый из которых использует весь диапазон частот и интервалов времени, выделенных для системы связи.

4. Пространственные характеристики. Обычно конкретные пользователи находятся в конкретных точках пространства. Соответственно внесение определенных ограничений на направление излучения и приема сигналов конкретных пользователей может приводить к выполнению условий ортогональности выделяемых ресурсов. Подобные ограничения используются при реализации технологии SDMA (Space Division Multiple Access) или пространственного множественного доступа. В этом случае с помощью антенной системы с несколькими узкими диаграммами излучения сигналы разделяются и направляются в разные стороны, данный метод допускает многократное использование одного частотного диапазона и наибольшее распространение нашел в спутниковых системах связи (своеобразный эквивалент технологии SDMA это использование различных проводных или кабельных соединений для различных пользователей).

5. Поляризационные характеристики электромагнитной волны. Известно, что поляризационные характеристики электромагнитной волны оказывают существенное влияние на преобразование энергии электромагнитного поля в энергию электрического тока, протекающего через проводники различной конфигурации и пространственной ориентации. Соответственно условие ортогональности может быть реализовано использованием различными абонентами различных (ортогональных) поляризационных характеристик электромагнитной волны. На данном свойстве ресурса связи основан метод множественного доступа PDMA (Polarization Division Multiple Access) или множественный доступ с поляризационным разделением. В данном случае разделение сигналов основывается на избирательности приемных антенн к положению вектора электрического поля (вертикальная, горизонтальная) или к направлению вращения данного вектора (вправо или влево). Данный метод наибольшее распространение нашел в радиорелейной и спутниковой связи.

При построении современных систем связи, как правило, используют несколько методов множественного доступа, в чистом виде используют крайне редко.

13.2. Множественный доступ с частотным разделением (FDMA – Frequency Division Multiple Access)

Это один из самых распространенных методов множественного доступа, применяемых в системах радиосвязи. Принцип FDMA заключается в том, что весь частотный спектр разделяется между пользователями на равные или неравные частотные полосы (рис. 13.1). Причем каналы могут быть как симметричными в обоих направлениях, так и асимметричными. Источники информации могут использовать выделенный им частотный ресурс неограниченно по времени, но при этом не должны создавать помехи соседним каналам. Чтобы избежать переходных помех, вводят специальный защитный частотный интервал между соседними каналами. Она не используется для передачи информации и поэтому снижает общую пропускную способность имеющегося канала связи.

Математически выполнение условия ортогональности используемых различными пользователями ресурсов связи может быть записано следующим образом:

$$\operatorname{Re} \left\{ \int_{-\infty}^{\infty} \mathbf{S}_k(\mathbf{t}) \mathbf{S}_m^*(\mathbf{t}) e^{j\omega_k t} e^{-j\omega_m t} dt \right\} \Big|_{k \neq m} = 0, \quad (13.2)$$

где $\mathbf{S}_k(\mathbf{t})$ и $\mathbf{S}_m(\mathbf{t})$ – комплексные огибающие сигналов различных пользователей (в общем случае могут совпадать); ω_k и ω_m – несущие частоты, используемые для сигналов различных пользователей; * – операция комплексного сопряжения. Для бесконечного интервала интегрирования при различных несущих частотах всегда выполняется условие ортогональности, так как

$$\operatorname{Re} \left\{ \int_{-\infty}^{\infty} e^{j\omega_k t} e^{-j\omega_m t} dt \right\} \Big|_{\omega_k \neq \omega_m} = 0. \quad (13.3)$$

На практике бесконечный интервал интегрирования означает выделение пользователю полосы частот с шириной, равной 0 Гц (невозможна передача информации). Поэтому условие ортогональности должно выполняться при конечных интервалах интегрирования (обычно равных времени корреляции комплексной огибающей используемых сигналов). Данное требование накладывает ограничения на минимально необходимый разнос между несущими частотами ω_k и ω_m . Обычно $\Delta\omega = |\omega_k - \omega_m| \geq \Delta\Omega$, где $\Delta\Omega$ – ширина (двусторонняя) эффективной полосы частот комплексной огибающей используемого сигнала.

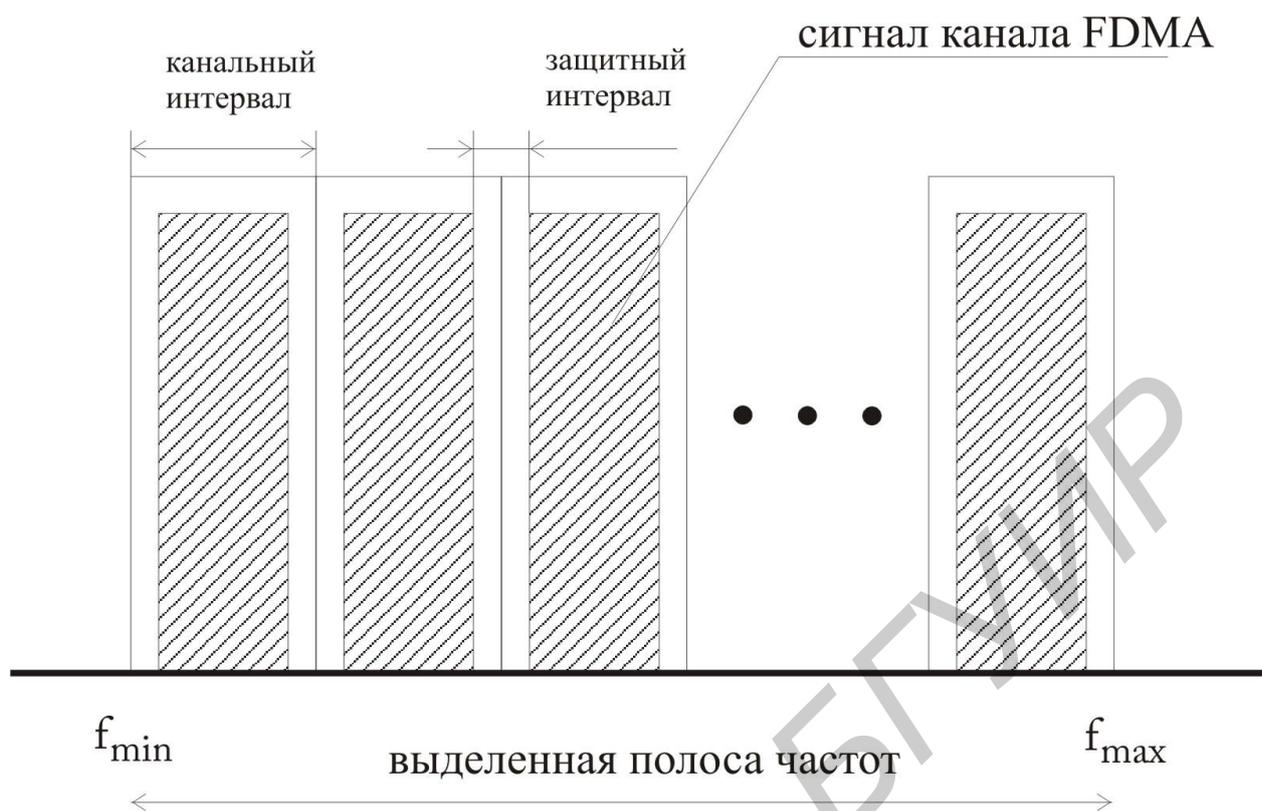


Рис. 13.1. Принцип организации FDMA

13.3. Множественный доступ с временным разделением (TDMA – Time Division Multiple Access)

Основной принцип TDMA заключается в том, что имеющийся ресурс разделяется между участниками информационного обмена на циклически повторяющиеся промежутки времени (рис. 13.2). Промежутки времени получили название «таймслот» (timeslot, TS). При этом абонент может использовать всю ширину выделенной полосы частот, но только в определенные временные отрезки. В такой ситуации главное, чтобы сигналы соседних таймслотов не накладывались друг на друга. Это может быть вызвано как слишком высокой мощностью передачи, так и помехами в канале, несовершенством используемого оборудования. Чтобы избежать подобных межслотовых помех, часто вводят специальный защитный временной интервал. Таким образом, если часть энергии одного передатчика просочится за пределы отведенного ему таймслота, то она будет оказывать воздействие лишь на защитный интервал, не несущий информации. Введение такого интервала снижает общую пропускную способность канала связи, но необходимо для поддержания заданных характеристик качества обслуживания.

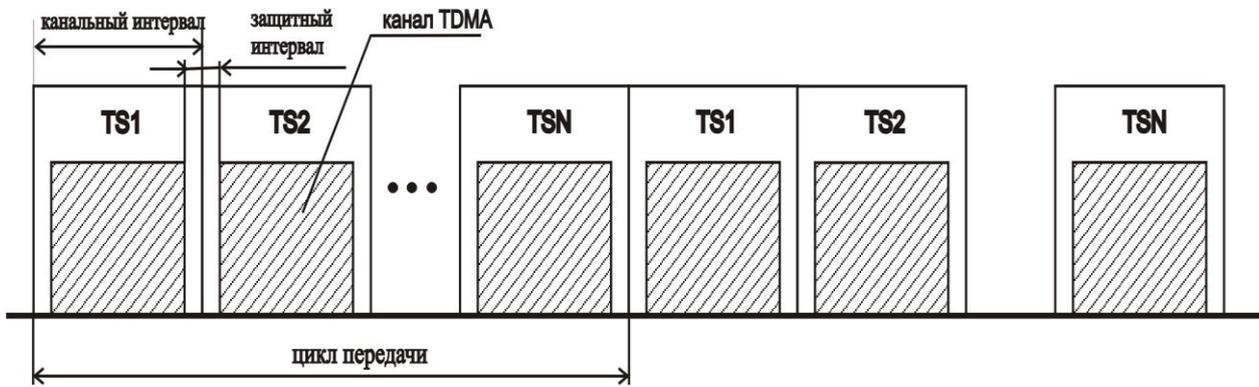


Рис. 13.2. Структура цикла TDMA

Математическое выполнение условия ортогональности используемых различными пользователями ресурсов связи может быть рассмотрено следующим образом:

$$\operatorname{Re} \left\{ \int_{-\infty}^{\infty} \mathbf{S}_k(\mathbf{t}) \mathbf{S}_m^*(\mathbf{t}) e^{j\omega_0 t} e^{-j\omega_0 t} dt \right\} \Big|_{k \neq m} = 0, \quad (13.4)$$

где $\mathbf{S}_k(\mathbf{t}) = g_F(t - k\tau) \mathbf{S}_k^i(\mathbf{t})$ и $\mathbf{S}_m(\mathbf{t}) = g_F(t - m\tau) \mathbf{S}_m^i(\mathbf{t})$ – комплексные огибающие сигналов различных пользователей;

$g_F(t) = \begin{cases} 1 & |t| \leq \tau/2 \\ 0 & |t| > \tau/2 \end{cases}$ – формирующий TDMA импульс; τ – длительность используемых интервалов передачи;

$\mathbf{S}_k^i(\mathbf{t})$ и $\mathbf{S}_m^i(\mathbf{t})$ – информационные составляющие комплексных огибающих различных пользователей (могут совпадать).

Условие ортогональности выполняется, так как

$$\int_{-\infty}^{\infty} g_F(t - k\tau) g_F(t - m\tau) dt = 0 \Big|_{k \neq m}. \quad (13.5)$$

При этом ортогональность сохраняется и при сокращении интервала интегрирования до интервала корреляции информационной составляющей, равного τ .

13.4. Метод множественного доступа с кодовым разделением (CDMA – Code Division Multiple Access)

Этот метод отличается от двух других наиболее распространенных методов разделения каналов FDMA и TDMA тем, что коды в отличие от времени и частоты не являются явным ресурсом канала связи. Несмотря на сложность реализации, данный метод используется в радиосвязи уже довольно давно, т.к. обладает некоторыми преимуществами, которых не имеют другие методы множественного доступа.

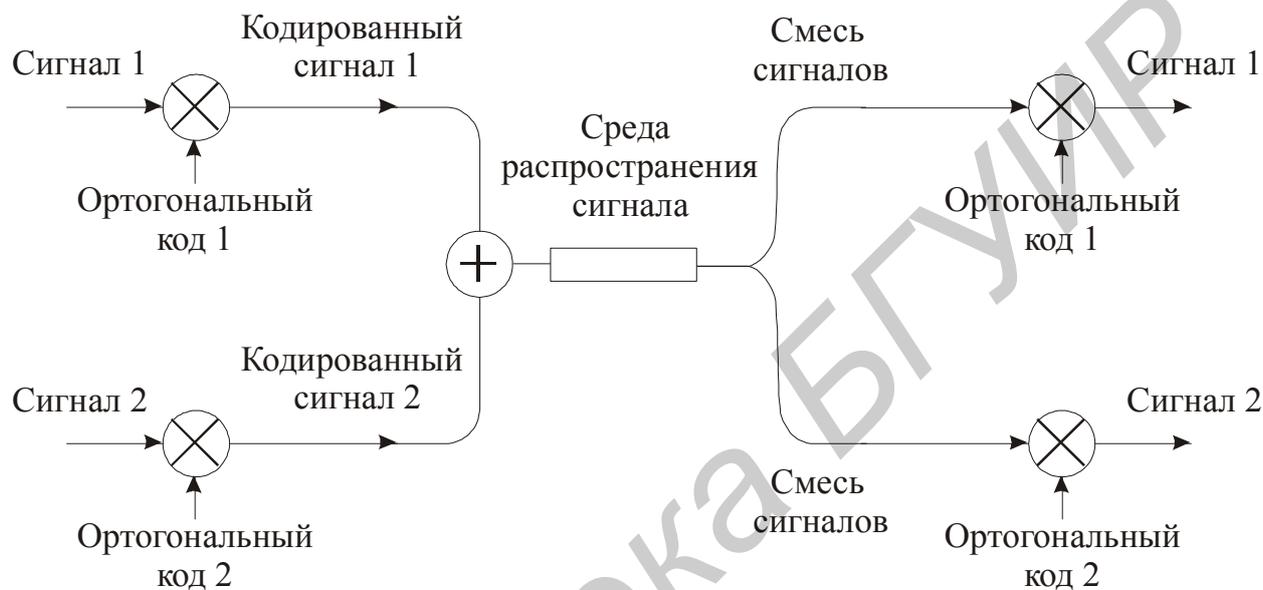


Рис. 13.3. Принцип организации CDMA

Принцип CDMA заключается в том, что каждому источнику информации назначается индивидуальный код (канальный код или код канала), при помощи которого он кодирует передаваемое сообщение. Приемник информации также знает этот код, и его задача в том, чтобы выделить закодированное сообщение нужного отправителя из всего потока других сообщений. В этом и заключается вся сложность, т. к. коды должны быть как можно меньше похожи друг на друга, даже при временном смещении сообщений. На математическом языке свойство «похожести» называется корреляцией. Таким образом, закодированные сообщения должны обладать как можно меньшей корреляцией. Этим свойством обладают ортогональные коды, взаимное скалярное произведение которых равно 0. На практике удастся получить не полностью ортогональные коды, а почти ортогональные. Если бы можно было сгенерировать бесконечное число полностью ортогональных кодов, то в одном канале связи смогут одновременно работать бесконечное число абонентов. Из-за неполной ортогональности кодов сигналы различных источников могут создавать помехи друг на друга. Причем чем выше мощности сигналов, тем ощутимее может оказаться влияние. Поэтому число одновременно

работающих абонентов и максимально возможная передаваемая мощность в системе с CDMA ограничены.

Условие ортогональности при использовании технологии CDMA определяется следующим образом:

$$\left. \operatorname{Re} \left\{ \int_{-\infty}^{\infty} \mathbf{S}_k(\mathbf{t}) \mathbf{S}_m^*(\mathbf{t}) e^{j\omega_0 t} e^{-j\omega_0 t} dt \right\} \right|_{k \neq m} = 0, \quad (13.6)$$

где $\mathbf{S}_k(\mathbf{t}) = \mathbf{g}_{\text{cdma } k}(\mathbf{t}) \mathbf{S}_k^i(\mathbf{t})$ и $\mathbf{S}_m(\mathbf{t}) = \mathbf{g}_{\text{cdma } m}(\mathbf{t}) \mathbf{S}_m^i(\mathbf{t})$ – комплексные огибающие сигналов различных пользователей; $\mathbf{g}_{\text{cdma } k}(\mathbf{t})$ – составляющая комплексной огибающей, представляющая каналный код, определенный k -му пользователю. Соответственно для обеспечения ортогональности каналные коды должны удовлетворять следующему условию:

$$\left. \operatorname{Re} \left\{ \int_{-\infty}^{\infty} \mathbf{g}_{\text{cdma } k}(\mathbf{t}) \mathbf{g}_{\text{cdma } m}^*(\mathbf{t}) dt \right\} \right|_{k \neq m} = 0. \quad (13.7)$$

13.5. Метод доступа с пространственным разнесением пользователей (SDMA – Space Division Multiple Access)

SDMA является методом доступа к каналам связи, основанным на пространственном разнесении пользователей услуг связи. В этом случае путем использования направленных антенных систем организовываются несколько пространственных областей. За счет пространственной избирательности передающих и приемных антенн в данных областях возможно одновременное использование «одинаковых» сигналов (сигналов, совпадающих по времени действия, занимаемой полосе частот, используемому каналному коду). Таким образом, возможно повышение эффективности использования ресурса связи. Метод организации пространственного множественного доступа становится наиболее привлекательным при использовании технологии Smart антенн («умных» антенн), которая позволяет с высокой точностью определять пространственное местоположение абонентов и концентрировать излучение и прием в их направлении.

13.6. Метод доступа, основанный на использовании поляризационных свойств электромагнитного излучения (PDMA – Polarization Division Multiple Access)

В результате согласованного использования поляризаций передающей и приемной антенны возможен отдельный прием электромагнитных излучений с ортогональными поляризациями (например, горизонтальной и вертикальной или с правой и левой круговыми поляризациями), что практически позволяет вдвое повысить эффективность использования выделенного ресурса связи. Данный метод множественного доступа получил широкое распространение в космических и радиорелейных линиях связи прямой видимости, поскольку при распространении радиоволн вне условий прямой видимости (за счет переотражений, дифракции, рефракции и пр.) наблюдается явление деполяризации электромагнитной волны.

13.7. Мультиплексирование с ортогональным частотным разделением (OFDM – Orthogonal Frequency Division Multiplexing)

OFDM – мультиплексирование с ортогональным частотным разделением использует большое количество близко расположенных ортогональных поднесущих колебаний. Каждое поднесущее колебание модулируется по обычной схеме линейной модуляции (например BPSK или QAM) с достаточно низкой скоростью, Большая скорость передачи сохраняется благодаря большому количеству используемых поднесущих колебаний (обычно несколько сотен).

По своей сути OFDM является частным случаем FDMA. Принципиальной особенностью здесь является синхронность используемых поднесущих частот.

Для того чтобы понять суть метода OFDM, рассмотрим следующий пример.

Пусть имеется многоканальный модулятор, формирующий на своих выходах N сигналов с $\pm\pi/2$ -BPSK. При этом каждый сигнал имеет несущее колебание вида

$$s_{ni}(t) = A_0 \cos(i\omega_0 t), \quad i = 1, 2, \dots, N. \quad (13.8)$$

Тогда сумма из N сигналов с $\pm\pi/2$ -BPSK запишется следующим образом:

$$s_{\Sigma}(t) = A_0 \left(\sum_{i=1}^N C_i \cos(i\omega_0 t) \right), \quad (13.9)$$

где $C_i = \pm 1$, модулирующие символы i -го модулятора. Обратим внимание на то, что выражение, стоящее в скобках, представляет собой разложение в ряд Фурье по гармоническим функциям совокупного сигнала $s_{\Sigma}(t)$. Аналогично можно записать представление совокупного сигнала в виде комплексного ряда Фурье. При этом модулирующие символы могут рассматриваться как изменяющиеся амплитуды спектральных компонент совокупного сигнала.

На основании вышеизложенного может быть предложена следующая схема (рис. 13.4) формирования совокупного (многоканального) сигнала $s_{\Sigma}(t)$.

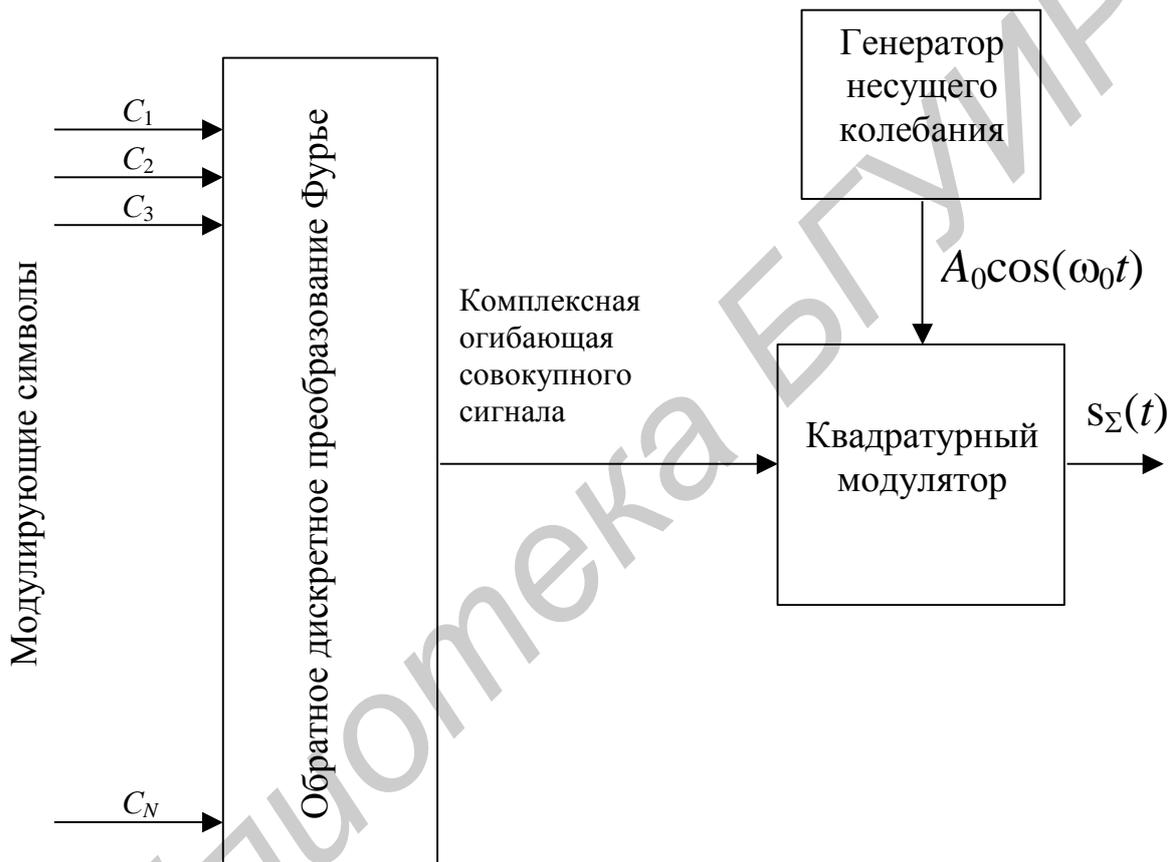


Рис. 13.4. Формирование многоканального сигнала OFDM

Модулирующие символы могут соответствовать любой линейной схеме дискретной модуляции, например, $C_i = \pm 1$ обеспечивает модуляцию поднесущих по методу $\pm\pi/2$ -BPSK; $C_i = \pm 1 \pm j1$ – QPSK или OQPSK; $C_i = \pm A \pm jB$ – QAM и т.д.

Необходимо отметить, что несимметричность исходного «спектра» C_1, \dots, C_N приводит к комплексному характеру получаемой в результате обратного преобразования Фурье временной реализации совокупного модулированного сигнала.

Основным преимуществом OFDM по сравнению со схемами, использующими одно несущее колебание, является её способность противостоять сложным условиям в канале. Например, бороться с узкополосными помехами и частотно-избирательным затуханием, вызванным многолучевым характером распространения, без использования сложных фильтров-эквалайзеров. Кроме того, OFDM обеспечивает большую гибкость и адаптивность в распределении ресурса радиоканала между различными пользователями (осуществляется простое перераспределение потоков данных между соответствующими входами обратного преобразования Фурье).

Выделение информации из совокупного сигнала OFDM осуществляется с использованием операции прямого преобразования Фурье.

Спектр сигнала с OFDM приведен на рис.13.5.

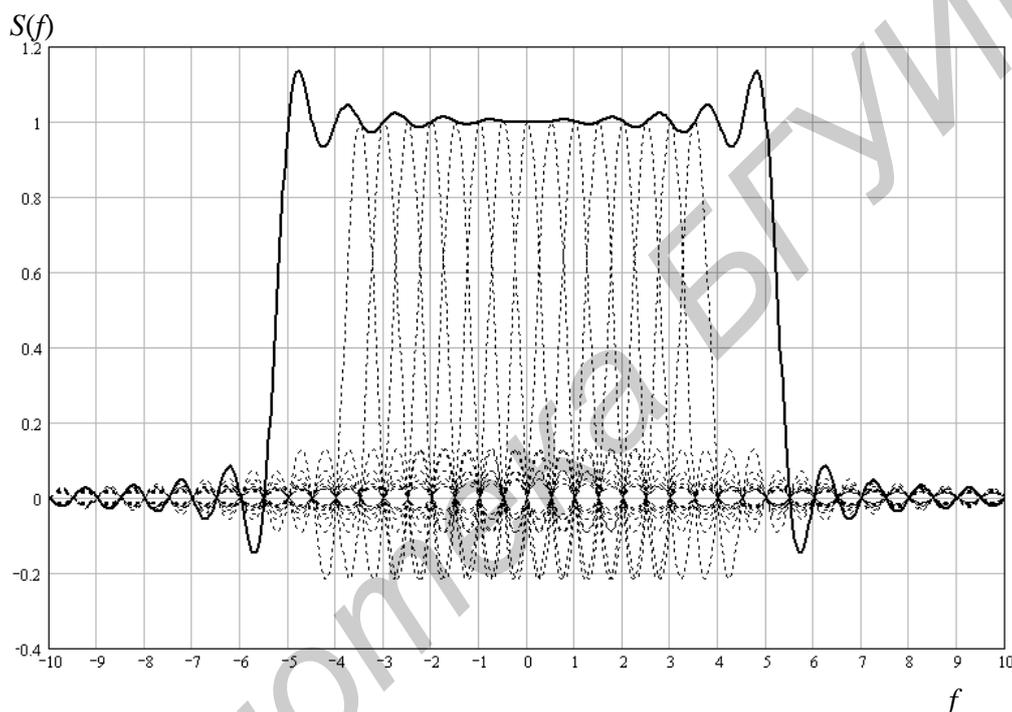


Рис. 13.5. Спектр сигнала с OFDM

ЗАКЛЮЧЕНИЕ

Актуальность и важность вопросов разработки и эксплуатации систем цифровой радиосвязи определяется их высокой востребованностью в различных отраслях экономики. Эта ситуация вызвана прежде всего требованиями высокого качества предоставляемых услуг, недостижимого в принципе в аналоговых системах, при формировании, передаче, приеме и обработке информации.

Теоретические основы цифровой радиосвязи излагаются в ряде литературных источников. Как правило, это переводные издания, многие из которых написаны научно-популярным языком и не в полной мере отражают сведения, необходимые инженеру-разработчику цифровых систем. Существует также и вторая крайность – в отдельных изданиях теоретические основы излагаются с использованием достаточно сложного математического аппарата, что затрудняет их эффективное использование для инженерной подготовки.

В данном пособии авторы постарались в какой-то мере ликвидировать пробелы существующих изданий как с точки зрения глубины и полноты излагаемого материала, так и с точки зрения его доступности. Отдельные главы пособия изложены достаточно полно (например, глава 11), некоторые главы представлены фрагментарно. В пособии есть абсолютно новый материал, ранее излагавшийся лишь в научных публикациях, например глава 8.

При написании материала авторы использовали собственный опыт и знания, а также свои разработки по созданию различных систем приема/передачи информации, обеспечению их надежного и эффективного функционирования.

ЛИТЕРАТУРА

1. Прокис, Дж. Цифровая связь: пер. с англ.; под ред. Д. Д. Кловского / Дж. Прокис. – М. : Радио и связь. 2000. – 800 с.
2. Таненбаум, Э. Компьютерные сети. – 4-е изд. / Э. Таненбаум. – СПб. : Питер, 2010. – 992 с.
3. Скляр, Б. Цифровая связь. Теоретические основы и практическое применение. – 2-е изд., испр.; пер. с англ. / Б. Скляр. – М.: Издательский дом «Вильямс», 2003. – 1104 с.
4. Волков, Л. Н. Системы цифровой радиосвязи: базовые методы и характеристики: учеб. пособие / Л. Н. Волков, Н. С. Немировский, Ю. С. Шинаков. – М.: Эко-Трендз, 2005. – 392 с.
5. Системы и сети цифровой радиосвязи : учеб. пособие / Н. И. Листопад [и др.]. – Минск: Изд-во Гревцова, 2009. – 200 с.
6. Гоноровский, И. С. Радиотехнические цепи и сигналы: учебник для вузов / И. С. Гоноровский. – 3-е изд., перераб. и доп. – М.: Сов. радио, 1977. – 608 с.
7. Хелд, Г. Технологии передачи данных. – 7-е изд. / Г. Хелд. – СПб. : Питер, К.: Издательская группа ВHV, 2003. – 720 с.
8. Шиман, Д. В. Метод псевдонеравномерного перемежения символов турбокода / Д. В. Шиман // Труды БГТУ. Сер. VI. Физ.-мат. науки и информ. – 2005. – Вып. XII. – С. 16–21.
9. Овчинников, А. М. TETRA – стандарт открытый для всех / А. М. Овчинников // Специальная техника. – 1999. – № 4. – С. 21–29.
10. Goldsmith, A. Wireless Communications / A. Goldsmith. – Cambridge University Press, 2005. – 571 p.
11. Рыжков, А. В. Синтезаторы частоты в технике радиосвязи / А. В. Рыжков, В. Н. Попов. – М. : Радио и связь, 1991. – 264 с.
12. Телекоммуникационные системы и сети: учеб. пособие. В 3 т. Т. 1: Современные технологии / Б. И. Крук, В. Н. Попантонопуло, В. П. Шувалов; под ред. проф. В. П. Шувалова. – 3-е изд., испр. и доп. – М. : Горячая линия-Телеком, 2003. – 647 с.
13. Радиопередающие устройства: учебник для вузов/ В. В. Шахгильдян [и др.]; под ред. В. В. Шахгильдяна. – 3-е изд., перераб. и доп. – М. : Радио и связь, 2003. – 560 с.
14. Проектирование радиопередатчиков / В. В. Шахгильдян [и др.]; под ред. В. В. Шахгильдяна. – М. : Радио и связь, 2000. – 656 с.

15. Столингс, В. Компьютерные системы передачи данных / В. Столингс. – 6-е изд.; пер. с англ. – М.: Издательский дом «Вильямс», 2002. – 928 с.
16. Кульгин, М. Технологии корпоративных сетей. Энциклопедия / М. Кульгин. – СПб.: Питер, 1989. – 704с.
17. Олифер, В. Г. Компьютерные сети. Принципы, технологии, протоколы. / В. Г. Олифер. – СПб.: Питер, 2000. – 672 с.
18. Олифер, В. Г. Новые технологии и оборудование IP-сетей / В. Г. Олифер, Н. А. Олифер. – СПб.: БХВ – Санкт-Петербург, 2000. – 512 с.
19. Листопад, Н. И. Моделирование и оптимизация глобальных сетей / Н. И. Листопад. – Минск : БГУ, 2000. – 156 с.
20. Устав Международного Союза Электросвязи (МСЭ). – Женева, 1992.
21. Об электросвязи: Закон Республики Беларусь от 19 июля 2005 г. № 45-3: одобр. Советом Респ. 30 июня 2005 г. – Минск, 2006. – 31 с.
22. Козел, В. М. Некоторые особенности трафика и эффективности использования частотного ресурса в сетях подвижной связи Министерства внутренних дел / В. М. Козел, К. А. Ковалев, К. Л. Горбачев // Техника СВЧ и телекоммуникационные технологии: материалы 17-й Междунар. Крымской конф., Севастополь, 10–14 сент. 2007 г. – Т. 1. – С. 309–310.
23. Recommendation ITU G.702 Digital Hierarchy bit rate.
24. Recommendation CCITT G.707 Synchronous digital hierarchy bit rate.
25. Лайонс, Р. Цифровая обработка сигналов / Р. Лайонс. – 2-е изд.; пер. с англ. – М. : ООО «Бином-Пресс», 2006. – 656 с.
26. Корн, Г. Справочник по математике: для научных работников и инженеров / Г. Корн, Т. Корн. – М. : Наука, 1974. – 832 с.
27. Миддлтон, Д. Введение в статистическую теорию связи / Д. Миддлтон; пер. с англ.; под ред. Б. Р. Левина. – М. : Сов. радио, 1962.
28. Тихонов, В. И. Статистическая радиотехника / В. И. Тихонов. – М. : Сов. радио, 1966. – 677 с.
29. Амиантов, И. Н. Избранные вопросы статистической теории связи / И. Н. Амиантов. – М. : Сов. радио, 1971. – 416 с.
30. Горяинов, В. Т. Статистическая радиотехника. Примеры и задачи: учеб. пособие для вузов / В. Т. Горяинов; под ред. В. И. Тихонова. – 2-е изд., перераб. и доп. – М. : Сов. радио, 1980. – 544 с.
31. Витерби, Э. Д. Принципы когерентной связи / Э. Д. Витерби; пер. с англ.; под ред. Б. Р. Левина. – М. : Сов. радио, 1970. – 392 с.

32. Xiong, F. Digital modulation techniques /F. Xiong; Artech House, Inc. – Boston-London, 2000. – 669 p.
33. Венцель, Е. С. Теория вероятности / Е. С. Венцель. – Изд. 4-е, стереотип. – М. : Наука, 1969. – 576 с.
34. Шеннон, К. Работы по теории информации и кибернетике / К. Шеннон; пер. с англ. под ред. Р. Л. Добрушина и О. Б. Лупанова с предисловием А. Н. Колмогорова. – М. : Издательство иностранной литературы, 1963. – 827 с.
35. Mengali, U. Synchronization techniques for digital receivers / U. Mengali, A. N. D'Andrea. – New-York, London: Plenum Press, 1997: –529 p.
36. Витерби, А. Д. Принципы цифровой связи и кодирования / А. Д. Витерби; пер. с англ.; под ред. К. Ш. Зигангирова. – М. : Радио и связь, 1982. –536 с.
37. Акулиничев, Ю. П. Теория электрической связи: учеб. пособие. / Ю. П. Акулиничев. – СПб. : Лань, 2010. – 240 с.

Учебное издание

Листопад Николай Измайлович
Козел Виктор Михайлович
Дубровский Василий Викторович и др.

**ТЕОРЕТИЧЕСКИЕ ОСНОВЫ
ЦИФРОВОЙ РАДИОСВЯЗИ**

УЧЕБНОЕ ПОСОБИЕ

Редактор *Г. С. Корбут*
Корректор *Е. Н. Батурчик*
Компьютерная верстка *Ю. Ч. Клочкевич*

Подписано в печать 11.12.2012. Формат 60x84 1/16. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. Уч.-изд. л. 20,0. Тираж экз. Заказ 693.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники»
ЛИ №02330/0494371 от 16.03.2009. ЛП №02330/0494175 от 03.04.2009.
220013, Минск, П. Бровки, 6